Chapter 1

Resolution and random signals

[1]The accuracy of measurements on observed signals is limited not only by practical realities, but also by certain fundamental principles. The most famous example included in this chapter is the **time-bandwidth** product in Fourier-transformation theory, called the "**uncertainty principle**."

Observed signals often look random and are often modeled by filtered random numbers. In this chapter we will see many examples of signals built from random numbers and discover how the nomenclature of statistics applies to them. Fundamentally, this chapter characterizes "**resolution**," resolution of frequency and arrival time, and the statistical resolution of signal amplitude and power as functions of time and frequency.

We will see $\sqrt{n}$ popping up everywhere. This $\sqrt{n}$ enters our discussion when we look at spectra of signals built from random numbers. Also, signals that are theoretically uncorrelated generally appear to be weakly correlated at a level of $1/\sqrt{n}$, where n is the number of independent points in the signal.

Measures of resolution (which are variously called **variance**s, **tolerance**s, uncertainties, **bandwidth**s, **duration**s, **spread**s, **rise time**s, spans, etc.) often interact with one another, so that experimental change to reduce one must necessarily increase another or some combination of the others. In this chapter we examine basic cases where such conflicting interactions occur.

To avoid confusion I introduce the unusual notation Λ where Δ is commonly used. Notice that the letter Λ resembles the letter Δ, and Λ connotes length without being confused with wavelength. Lengths on the time and frequency axes are defined as follows:

dt, df	mesh intervals in time and frequency
Δt, Δf	mesh intervals in time and frequency
ΔT ΔF	extent of time and frequency axis
ΛT, ΛF	time duration and spectral **bandwidth** of a signal

[1]A more detailed and analytic version of this chapter is on the web at http://sep.stanford.edu/sep/prof/ in my book, Earth Soundings Analysis: Processing versus Inversion.

There is no mathematically tractable and universally acceptable definition for time span ΔT and **spectral bandwidth** ΔF. A variety of defining equations are easy to write, and many are in general use. The main idea is that the time span ΔT or the frequency span ΔF should be able to include most of the energy but need not contain it all. The time duration of a damped exponential function is infinite if by duration we mean the span of nonzero function values. However, for practical purposes the time span is generally defined as the time required for the amplitude to decay to e^{-1} of its original value. For many functions the span is defined by the span between points on the time or frequency axis where the curve (or its envelope) drops to half of the maximum value. Strange as it may sound, there are certain concepts about the behavior of ΔT and ΔF that seem appropriate for "all" mathematical choices of their definitions, yet these concepts can be proven only for special choices.

1.1 TIME-FREQUENCY RESOLUTION

A consequence of Fourier transforms being built from $e^{i\omega t}$ is that narrowing a function in one domain widens it in the other domain. Scaling ω implies inverse scaling of t to keep the product ωt constant. A pure sinusoidal wave has a clearly defined frequency, but it is spread over the infinitely long time axis. At the other extreme is an impulse function (often called a delta function), which is nicely compressed to a point on the time axis but contains a mixture of all frequencies. In this section we examine how the width of a function in one domain relates to that in the other. By the end of the section, we will formalize this into an inequality:

For any signal, the time duration ΔT and the spectral bandwidth ΔF are related by

$$\Delta F \, \Delta T \quad \geq \quad 1 \tag{1.1}$$

This **inequality** is the **uncertainty principle**.

Since we are unable to find a precise and convenient analysis for the definitions of ΔF and ΔT, the inequality (1.1) is not strictly true. What is important is that rough equality in (1.1) is observed for many simple functions, but for others the inequality can be extremely slack (far from equal). Strong *inequality* arises from **all-pass filter**s. A phase shifting (all-pass) filter leaves the spectrum unchanged, and hence ΔF unchanged, but it can spread out the signal arbitrarily, increasing ΔT arbitrarily. Some people say that the **Gaussian** function has the minimum product in (1.1), but that really depends on a particular method of measuring ΔF and ΔT.

1.1.1 A misinterpretation of the uncertainty principle

It is easy to misunderstand the uncertainty principle. An oversimplification of it is to say that it is "impossible to know the frequency at any particular time." This oversimplification leads us to think about a **truncated sinusoid**, such as in Figure 1.1. We know the frequency exactly, so ΔF seems zero, whereas ΔT is finite, and this seems to violate (1.1). But what the figure shows is that the truncation of the sinusoid has broadened the frequency band. More particularly, the impulse function in the frequency domain has been convolved by the sinc function that is the Fourier transform of the truncating rectangle function.

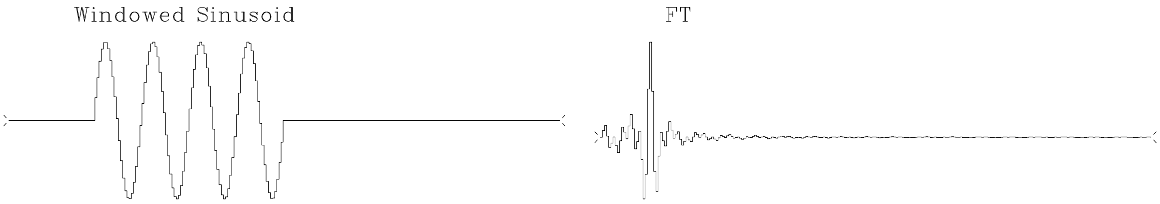

Figure 1.1: Windowed sinusoid and its Fourier transform. rnd-windcos [NR]

1.1.2 Measuring the time-bandwidth product

Now examine Figure 1.2, which contains sampled **Gaussian** functions and their Fourier transforms. The Fourier transform of a Gaussian is well known to be another Gaussian function, as the plot confirms. I adjusted the width of each Gaussian so that the widths would be about equal in both domains. The Gaussians were sampled at various values of n, increasing in steps by a factor of 4. You can measure the width dropping by a factor of 2 at each step. For those of you who have already learned about the uncertainty principle, it may seem paradoxical that the function's width is dropping in both time and frequency domains.

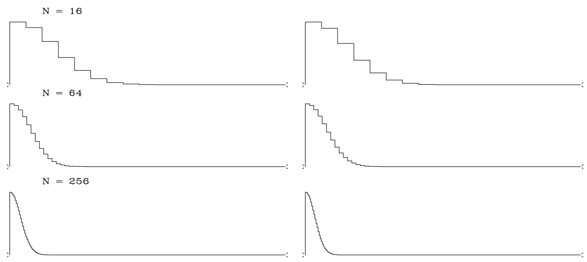

Figure 1.2: Sampled Gaussian functions and their Fourier transforms for vectors of length n = 16, 64, and 256. rnd-uncertain [NR]

The resolution of the paradox is that the physical length of the time axis or the frequency axis is varying as we change n (even though the plot length is scaled to a constant on the page). We need to associate a physical mesh with the computational mesh. In real physical space as well as in Fourier transform space, the object remains a constant size as the mesh is refined.

Let us read from Figure 1.2 values for the widths ΛF and ΛT. On the top row, where $N = 16$, I pick a width of about 4 points, and this seems to include about 90% of the area under the function. For this signal (with the widths roughly equal in both domains) it seems that $\Lambda T = \sqrt{N} dt$ and $\Lambda F = \sqrt{N} df$. Using the relation between dt and df found in equation (**??**), which says that $dt\, df = 1/N$, the product becomes $\Lambda T \Lambda F$ = 1.

We could also confirm the inequality (1.1) by considering simple functions for which we know the analytic transforms—for example, an impulse function in time. Then $\Lambda T = dt$, and the Fourier transform occupies the entire frequency band from minus to plus the Nyquist frequency $\pm .5/dt$ Hz, i.e., $\Lambda F = 1/dt$. Thus again, the product is $\Lambda T \Lambda F$ = 1.

1.1.3 The uncertainty principle in physics

The inequality (1.1) derives the name "**uncertainty principle**" from its interpretation in **quantum mechanics**. Observations of subatomic particles show they behave like waves with spatial frequency proportional to particle momentum. The classical laws of mechanics enable prediction of the future of a mechanical system by extrapolation from the currently known position and momentum. But because of the wave nature of matter, with momentum proportional to spatial frequency, such prediction requires simultaneous knowledge of both the location and the spatial frequency of the wave. This is impossible, as we see from (1.1); hence the word "uncertainty."

1.1.4 Gabor's proof of the uncertainty principle

Although it is easy to verify the uncertainty principle in many special cases, it is not easy to deduce it. The difficulty begins from finding a definition of the width of a function that leads to a tractable analysis. One possible definition uses a second moment; that is, ΛT is defined by

$$(\Lambda T)^2 \quad = \quad \frac{\int \, t^2 \, b(t)^2 \, dt}{\int b(t)^2 \, dt} \tag{1.2}$$

The spectral bandwidth ΛF is defined likewise. With these definitions, Dennis **Gabor** prepared a widely reproduced proof. I will omit his proof here; it is not an easy proof; it is widely available; and the definition (1.2) seems inappropriate for a function we often use, the **sinc** function, i.e., the FT of a step function. Since the sinc function drops off as t^{-1}, its width ΛT defined with (1.2) is infinity, which is unlike the more human measure of width, the distance to the first axis crossing.

1.1.5 My rise-time form of the uncertainty principle

A similar and possibly more basic concept than the product of time and frequency spreads is the relationship between spectral **bandwidth** and the "**rise time**" of a system-response function. The rise time ΛT of a system response is defined as follows: when we kick a physical system with an impulse function, it usually responds rapidly, rising to some maximum level, and then dropping off more slowly toward zero. The quantitative value of the rise time is generally, and somewhat arbitrarily, taken to be the span between the time of excitation and the time at which the system response is more than halfway up to its maximum.

"Tightness" (nearness to equality) in the inequality (1.1) is associated with minimum phase. "Slackness" (remoteness from equality) in the (1.1) would occur if a filter with an additional all-pass component were used. Slackness could also be caused by a decay time that is more rapid than the rise time, or by other combinations of rises and falls, such as random combinations. Physical systems generally respond rapidly compared to the rate at which they later decay. Focusing our attention on such systems, we can now seek to derive the inequality (1.1) applied to rise time and bandwidth.

The first step is to choose a definition for rise time. I have found a tractable definition of **rise time** to be

$$\frac{1}{\Lambda T} \quad = \quad \frac{\int_0^\infty \frac{1}{t}\, b(t)^2\, dt}{\int_0^\infty b(t)^2\, dt} \tag{1.3}$$

where $b(t)$ is the response function under consideration. Equation (1.3) defines ΛT by the first negative moment. Since this is unfamiliar, consider two examples. Taking $b(t)$ to be a step function, recognize that the numerator integral diverges, giving the desired $\Lambda T = 0$ rise time. As a further example, take $b(t)^2$ to grow linearly from zero to t_0 and then vanish. Then the rise time ΛT is $t_0/2$, again reasonable. It is curious that $b(t)$ could grow as $\sqrt{t}$, which rises with infinite slope at $t = 0$, and not cause ΛT to be pushed to zero.

Proof

We omit the details of the proof of the rise time form of the uncertainty principle which are found in PVI and mention instead a few insights found along with the proof. The "slack" in the uncertainty principle relates to comparing the duration of a signal ΛT to the duration of its autocorrelation ΛT_{auto}. Generally speaking, it is easy to find a long signal that has short autocorrelation. Just take an arbitrary short signal and convolve it using a lengthy all-pass filter. Conversely, we cannot get a long autocorrelation function from a short signal. A good example is the autocorrelation of a rectangle function, which is a triangle. The triangle appears to be twice as long, but considering that the triangle tapers down, it is reasonable to assert that the ΛT's are the same. Thus, we conclude that

$$\Lambda T_{\text{auto}} \quad \leq \quad \Lambda T \tag{1.4}$$

From this inequality may be proven the inequality in the uncertainty relation

$$\Lambda T \ \Lambda F \quad \geq \quad 1 \tag{1.5}$$

1.2 FT OF RANDOM NUMBERS

Many real signals are complicated and barely comprehensible. In experimental work, we commonly transform such data. To better understand what this means, it will be worthwhile to examine signals made from **random** numbers.

Figure 1.3 shows discrete Fourier transforms of random numbers. The basic conclusion to be drawn from this figure is that transforms of random numbers look like more random numbers. A random series containing all frequencies is called a "white-noise" series, because the color white is made from roughly equal amounts of all colors. Any series made by independently chosen random numbers is said to be an "independent" series. An independent series must be white, but a white series need not be independent. Figure 1.4 shows Fourier transforms of random numbers surrounded by zeros (or zero padded). Since all the vectors of random numbers are the same length (each has 1024 points, including both sides of the even

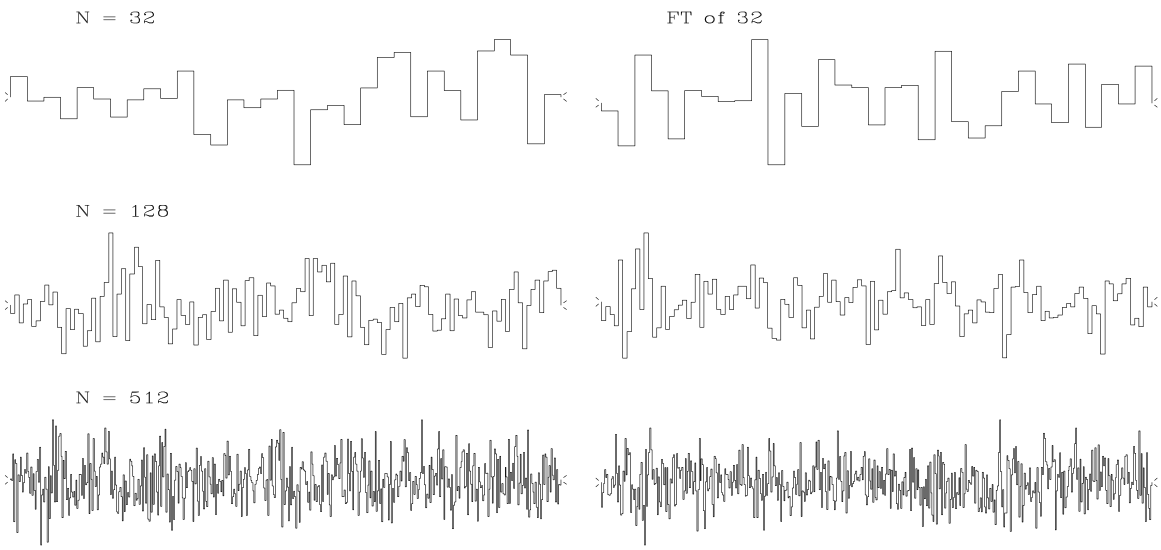

Figure 1.3: Fourier cosine transforms of vectors containing random numbers. N is the number of components in the vector. rnd-nrand [NR]

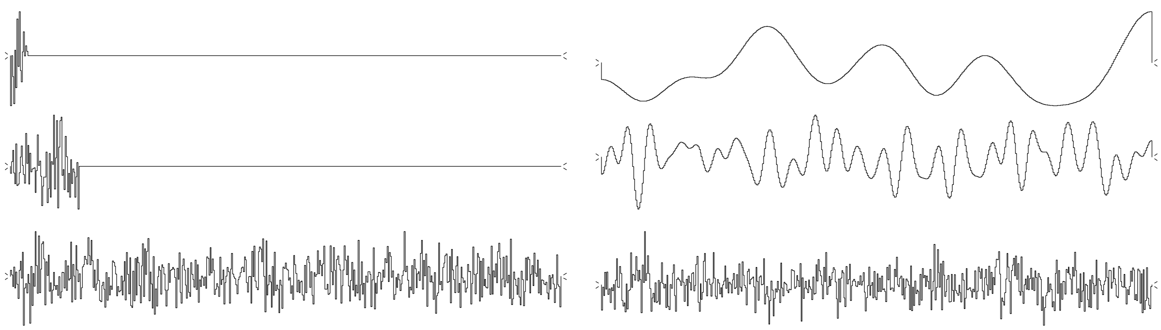

Figure 1.4: Zero-padded random numbers and their FTs. rnd-pad [NR]

function with the even part (513 points) shown), the transforms are also the same length. The top signal has less randomness than the second trace (16 random numbers versus 64). Thus the top FT is smoother than the lower ones. Although I introduced this figure as if the left panel were the time domain and the right panel were frequency, you are free to think of it the opposite way. This is more clear. With the left-hand signal being a frequency function, where higher frequencies are present, the right-hand signal oscillates faster.

1.2.1 Bandlimited noise

Figure 1.5 shows bursts of 25 random numbers at various shifts, and their Fourier transforms. You can think of either side of the figure as the time domain and the other side as the frequency domain. I like to think of the left side as the Fourier domain and the right side as the signals. Then the signals seem to be sinusoids of a constant frequency (called the “center" frequency) and of an amplitude that is modulated at a slower rate (called the “**beat**” frequency). Observe that the center frequency is related to the *location* of the random bursts, and that the beat frequency is related to the *bandwidth* of the noise burst.

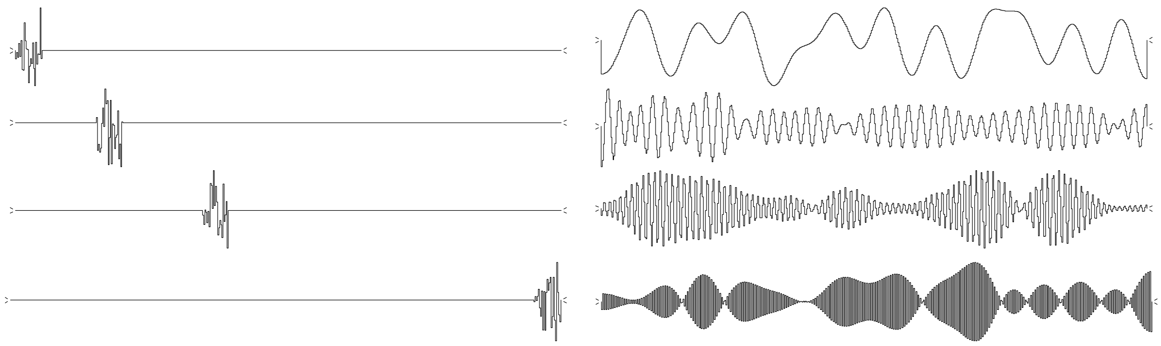

Figure 1.5: Shifted, zero-padded random numbers in bursts of 25 numbers. Left and right related by cosine transform. rnd-shift [NR]

You can also think of Figure 1.5 as having one-sided frequency functions on the left, and the right side as being the *real part* of the signal. The real parts are cosinelike, whereas the imaginary parts (not shown) are sinelike and have the same envelope function as the cosinelike part.

You might have noticed that the bottom plot in Figure 1.5, which has Nyquist-frequency modulated beats, seems to have about twice as many beats as the two plots above it. This can be explained as an end effect. The noise burst near the **Nyquist** frequency is really twice as wide as shown, because it is mirrored about the Nyquist frequency into negative frequencies. Likewise, the top figure is not modulated at all, but the signal itself has a frequency that matches the beats on the bottom figure.

1.3 TIME-STATISTICAL RESOLUTION

If we flipped a fair coin 1000 times, it is unlikely that we would get exactly 500 heads and 500 tails. More likely the number of heads would lie somewhere between 400 and 600. Or would it lie in another range? The theoretical value, called the “**mean**" or the “**expectation**," is 500. The value from our experiment in actually flipping a fair coin is called the “**sample mean**.” How much difference Λm should we expect between the sample mean $\hat{m}$ and the true mean m? Both the coin flips x and our sample mean $\hat{m}$ are *random variables*. Our 1000-flip experiment could be repeated many times and would typically give a different $\hat{m}$ each time. This concept will be formalized in section 11.3.5. as the “**variance of the sample mean**,” which is the expected squared difference between the true mean and the mean of our sample.

The problem of estimating the **statistic**al parameters of a time series, such as its mean, also appears in seismic processing. Effectively, we deal with seismic traces of finite duration, extracted from infinite sequences whose parameters can only be estimated from the finite set of values available in these seismic traces. Since the knowledge of these parameters, such as signal-to-noise ratio, can play an important role during the processing, it can be useful not only to estimate them, but also to have an idea of the error made in this estimation.

1.3.1 Ensemble

The “true value” of the mean could be defined as the mean that results when the coin is flipped n times, when n is conceived of as going to infinity. A more convenient definition of true value is that the experiment is imagined as having been done separately under identical conditions by an infinite number of people (an “**ensemble**”). The ensemble may seem a strange construction; nonetheless, much literature in statistics and the natural sciences uses the ensemble idea. Let us say that the ensemble is defined by a probability as a function of time. Then the ensemble idea enables us to define a time-variable mean (the sum of the values found by the ensemble weighted by the probabilities) for, for example, coins that change with time.

1.3.2 Expectation and variance

A conceptual average over the ensemble, or **expectation**, is denoted by the symbol E. The index for summation over the ensemble is never shown explicitly; every random variable is presumed to have one. Thus, the true mean at time t is defined as $m_x(t) = \mathrm{E}(x_t)$. The mean can vary with time:

$$m_x(t) \quad = \quad \mathrm{E}[x(t)] \tag{1.6}$$

The “**variance**” σ^2 is defined to be the power after the mean is removed, i.e.,

$$\sigma_x(t)^2 \quad = \quad \mathrm{E}\left[(x(t)-m_x(t))^2\right] \tag{1.7}$$

(Conventionally, σ^2 is referred to as the variance, and σ is called the “**standard deviation**.”)

For notational convenience, it is customary to write $m(t)$, $\sigma(t)$, and $x(t)$ simply as m, σ, and x_t, using the verbal context to specify whether m and σ are time-variable or constant. For example, the standard deviation of the seismic amplitudes on a seismic trace before correction of spherical divergence decreases with time, since these amplitudes are expected to be "globally" smaller as time goes on.

When manipulating algebraic expressions, remember that the symbol E behaves like a summation sign, namely,

$$\mathrm{E} \quad \equiv \quad (\lim N \to \infty) \quad \frac{1}{N}\sum_{1}^{N} \tag{1.8}$$

Note that the summation index is not given, since the sum is over the ensemble, not time. To get some practice with the expectation symbol E, we can reduce equation (1.7):

$$\sigma_x^2 \quad = \quad \mathrm{E}\,[(x_t - m_x)^2] \quad = \quad \mathrm{E}(x_t^2) \;-\; 2m_x\mathrm{E}(x_t) + m_x^2 \quad = \quad \mathrm{E}(x_t^2) \;-\; m_x^2 \tag{1.9}$$

Equation (1.9) says that the energy is the variance plus the squared mean.

1.3.3 Probability and independence

A random variable x can be described by a **probability** $p(x)$ that the amplitude x will be drawn. In real life we almost never know the probability function, but theoretically, if we do know it, we can compute the **mean** value using

$$m \quad = \quad E(x) \quad = \quad \int x\, p(x)\, dx \tag{1.10}$$

"**Statistical independence**" is a property of two or more random numbers. It means the samples are drawn independently, so they are unrelated to each other. In terms of probability functions, the independence of random variables x and y is expressed by

$$p(x,y) \quad = \quad p(x)\, p(y) \tag{1.11}$$

From these, it is easy to show that

$$\mathrm{E}(xy) \quad = \quad \mathrm{E}(x)\mathrm{E}(y) \tag{1.12}$$

1.3.4 Sample mean

Now let x_t be a time series made up of identically distributed random numbers: m_x and σ_x do not depend on time. Let us also suppose that they are *independently* chosen; this means in particular that for any different times t and s $(t \neq s)$:

$$\mathrm{E}(x_t x_s) \quad = \quad \mathrm{E}(x_t)\mathrm{E}(x_s) \tag{1.13}$$

Suppose we have a sample of n points of x_t and are trying to determine the value of m_x. We could make an estimate $\hat{m}_x$ of the mean m_x with the formula

$$\hat{m}_x = \frac{1}{n}\sum_{t=1}^{n} x_t \tag{1.14}$$

A somewhat more elaborate method of estimating the mean would be to take a weighted average. Let w_t define a set of weights normalized so that

$$\sum w_t = 1 \tag{1.15}$$

With these weights, the more elaborate estimate $\hat{m}$ of the mean is

$$\hat{m}_x = \sum w_t \, x_t \tag{1.16}$$

Actually (1.14) is just a special case of (1.16); in (1.14) the weights are $w_t = 1/n$.

Further, the weights could be *convolved* on the random time series, to compute *local* averages of this time series, thus smoothing it. The weights are simply a filter response where the filter coefficients happen to be positive and cluster together. Figure 1.6 shows an example: a **random walk** function with itself smoothed locally.

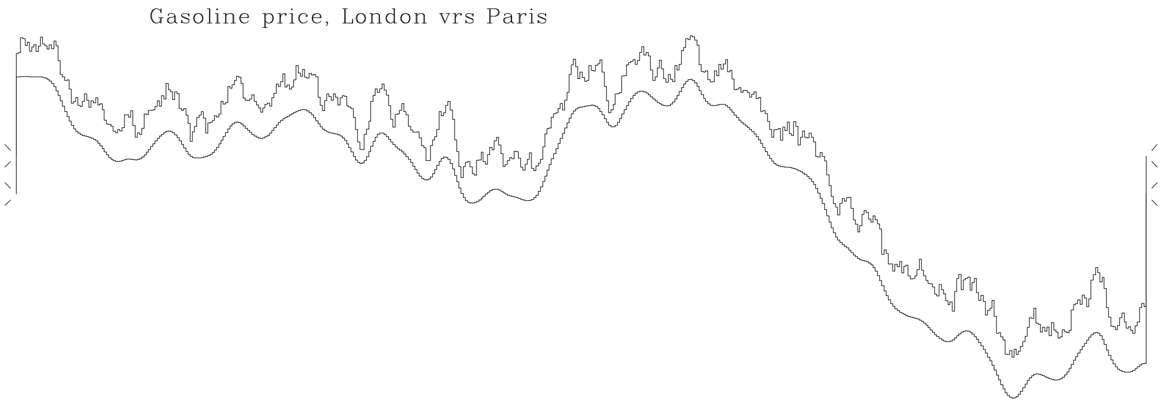

Figure 1.6: Random walk and itself smoothed (and shifted downward). rnd-walk [NR]

1.3.5 Variance of the sample mean

Ordinarily I like to squeeze all the equations out of my books, but I include them here because they show why random fluctions so often involve the square root of the number of variables. When you hear, for example, that a survey of the voters has a 3% accuracy, you can deduce that the surveyors spoke with about $1000 = 1/.03^2$ voters.

Formally now, we calculate how far the estimated mean $\hat{m}$ is likely to be from the true mean m for a sample of length n. This difference is the **variance of the sample mean** and is given by $(\Lambda m)^2 = \sigma_{\hat{m}}^2$, where

$$\sigma_{\hat{m}}^2 = \mathrm{E}[(\hat{m}-m)^2] \tag{1.17}$$

$$= \mathrm{E}\left\{[(\sum w_t x_t) - m]^2\right\} \tag{1.18}$$

Now use the fact that $m = m\sum w_t = \sum w_t m$:

$$\sigma_{\hat{m}}^2 = \mathrm{E}\left\{\left[\sum_t w_t(x_t - m)\right]^2\right\} \tag{1.19}$$

$$= \mathrm{E}\left\{\left[\sum_t w_t(x_t - m)\right]\left[\sum_s w_s(x_s - m)\right]\right\} \tag{1.20}$$

$$= \mathrm{E}\left[\sum_t \sum_s w_t w_s (x_t - m)(x_s - m)\right] \tag{1.21}$$

The step from (1.20) to (1.21) follows because

$$(a_1 + a_2 + a_3)(a_1 + a_2 + a_3) \quad = \quad \text{sum of} \begin{bmatrix} a_1a_1 & a_1a_2 & a_1a_3 \\ a_2a_1 & a_2a_2 & a_2a_3 \\ a_3a_1 & a_3a_2 & a_3a_3 \end{bmatrix} \tag{1.22}$$

The expectation symbol E can be regarded as another summation, which can be done after, as well as before, the sums on t and s, so

$$\sigma_{\hat{m}}^2 = \sum_t \sum_s w_t\, w_s\, \mathrm{E}\left[(x_t - m)(x_s - m)\right] \tag{1.23}$$

If $t \neq s$, since x_t and x_s are independent of each other, the expectation $\mathrm{E}[(x_t - m)(x_s - m)]$ will vanish. If $s = t$, then the expectation is the variance defined by (1.7). Expressing the result in terms of the Kronecker delta, δ_{ts} (which equals unity if $t = s$, and vanishes otherwise) gives

$$\sigma_{\hat{m}}^2 = \sum_t \sum_s w_t\, w_s\, \sigma_x^2 \delta_{ts} \tag{1.24}$$

$$\sigma_{\hat{m}}^2 = \sum_t w_t^2 \sigma_x^2 \tag{1.25}$$

$$\sigma_{\hat{m}} = \sigma_x \sqrt{\sum_t w_t^2} \tag{1.26}$$

For n weights, each of size $1/n$, the **standard deviation of the sample mean** is

$$\Lambda m_x = \sigma_{\hat{m}_x} = \sigma_x \sqrt{\sum_{t=1}^{n} \left(\frac{1}{n}\right)^2} = \frac{\sigma_x}{\sqrt{n}} \tag{1.27}$$

This is the most important property of random numbers that is not intuitively obvious. Informally, the result (1.27) says this: given a sum y of terms with random polarity, whose theoretical mean is zero, then

$$y \quad = \quad \underbrace{\pm 1 \pm 1 \pm 1 \cdots}_{n \text{ terms}} \tag{1.28}$$

The sum y is a random variable whose standard deviation is $\sigma_y = \sqrt{n} = \Lambda y$. An experimenter who does not know the mean is zero will report that the mean of y is $\mathrm{E}(y) = \hat{y} \pm \Lambda y$, where $\hat{y}$ is the experimental value.

If we are trying to estimate the mean of a random series that has a time-variable mean, then we face a basic dilemma. Including many numbers in the sum in order to make Λm small conflicts with the possibility of seeing m_t change during the measurement.

The "**variance of the sample variance**" arises in many contexts. Suppose we want to measure the storminess of the ocean. We measure water level as a function of time and subtract the mean. The storminess is the variance about the mean. We measure the storminess in one minute and call it a sample storminess. We compare it to other minutes and other locations and we find that they are not all the same. To characterize these differences, we need the *variance of the sample variance* $\sigma^2_{\hat{\sigma}^2}$. Some of these quantities can be computed theoretically, but the computations become very cluttered and dependent on assumptions that may not be valid in practice, such as that the random variables are independently drawn and that they have a Gaussian probability function. Since we have such powerful computers, we might be better off ignoring the theory and remembering the basic principle that a function of random numbers is also a random number. We can use simulation to estimate the function's mean and variance. Basically we are always faced with the same dilemma: if we want to have an accurate estimation of the variance, we need a large number of samples, which limits the possibility of measuring a time-varying variance.

EXERCISES:

1 Suppose the mean of a sample of random numbers is estimated by a triangle weighting function, i.e.,

$$\hat{m} \quad = \quad s \sum_{i=0}^{n} (n-i)\, x_i$$

Find the scale factor s so that $\mathrm{E}(\hat{m}) = m$. Calculate Λm. Define a reasonable ΛT. Examine the uncertainty relation.

2 A random series x_t with a possibly time-variable mean may have the mean estimated by the feedback equation

$$\hat{m}_t \quad = \quad (1-\epsilon)\hat{m}_{t-1} + b x_t$$

a. Express $\hat{m}_t$ as a function of $x_t, x_{t-1}, \ldots$, and not $\hat{m}_{t-1}$.

b. What is ΛT, the effective averaging time?

c. Find the scale factor b so that if $m_t = m$, then $\mathrm{E}(\hat{m}_t) = m$.

d. Compute the random error $\Lambda m = \sqrt{\mathrm{E}(\hat{m} - m)^2}$. (HINT: Λm goes to $\sigma\sqrt{\epsilon/2}$ as $\epsilon \to 0$.)

e. What is $(\Lambda m)^2 \Lambda T$ in this case?

1.4 SPECTRAL FLUCTUATIONS

Recall the basic model of time-series analysis, namely, random numbers passing through a filter. A sample of input, filter, and output amplitude spectra is shown in Figure 1.7. From the spectrum of the output we can guess the spectrum of the filter, but the figure shows there are some limitations in our ability to do so. Let us analyze this formally.

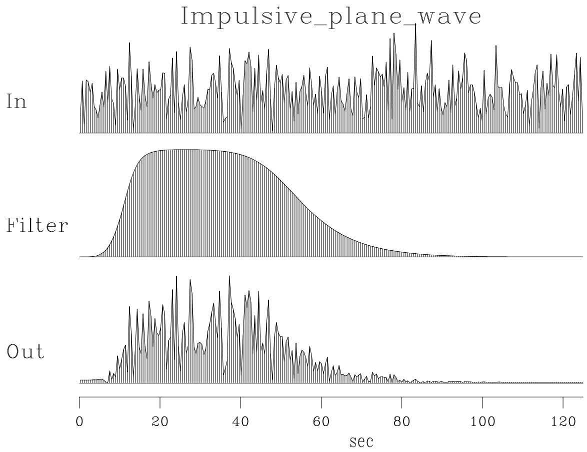

Figure 1.7: Random numbers into a filter. Top is a spectrum of random numbers. Middle is the spectrum of a filter. Bottom is the spectrum of filter output. rnd-filter [NR]

Observations of sea level over a long period of time can be summarized in terms of a few statistical averages, such as the mean height m and the variance σ^2. Another important kind of statistical average for use on geophysical time series is the "**power spectrum**." Many mathematical models explain only statistical averages of data and not the data itself. To recognize certain **pitfall**s and understand certain fundamental limitations on work with power spectra, we first consider the idealized example of random numbers.

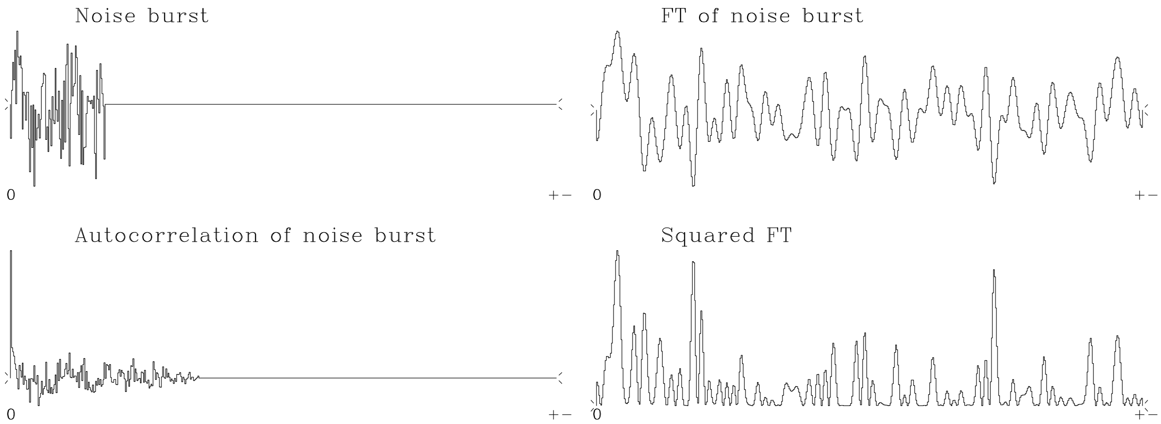

Figure 1.8: Autocorrelation and spectra of random numbers. rnd-auto [NR]

Figure 1.8 shows a signal that is a burst of noise; its Fourier transform, and the transform squared; and its inverse transform, the autocorrelation. Here the FT squared is the same as the more usual FT times its complex conjugate—because the noise-burst signal is even, its FT is real.

Notice that the **autocorrelation** has a big spike at zero lag. This spike represents the correlation of the random numbers with themselves. The other lags are much smaller. They represent the correlation of the noise burst with itself shifted. Theoretically, the noise burst is *not* correlated with itself shifted: these small fluctuations result from the finite extent of the noise sample.

Imagine many copies of Figure 1.8. Ensemble averaging would amount to adding these other autocorrelations or, equivalently, adding these other spectra. The fluctuations aside the central lobe of the autocorrelation would be destroyed by ensemble averaging, and the fluctuations in the spectrum would be smoothed out. The **expectation of the autocorrelation** is that it is an impulse at zero lag. The **expectation of the spectrum** is that it is a constant, namely,

$$\mathrm{E}[\hat{S}(Z)] \quad = \quad S(Z) \quad = \quad \text{const} \tag{1.29}$$

1.4.1 Paradox: large n vs. the ensemble average

Now for the paradox. Imagine $n \to \infty$ in Figure 1.8. Will we see the same limit as results from the **ensemble average**? Here are two contradictory points of view:

- For increasing n, the fluctuations on the nonzero autocorrelation lags get smaller, so the autocorrelation should tend to an impulse function. Its Fourier transform, the spectrum, should tend to a constant.
- On the other hand, for increasing n, as in Figure 1.3, the spectrum does not get any smoother, because the FTs should still look like random noise.

We will see that the first idea contains a false assumption. The autocorrelation does tend to an impulse, but the fuzz around the sides cannot be ignored—although the fuzz tends to zero amplitude, it also tends to infinite extent, and the product of zero with infinity here tends to have the same energy as the central impulse.

To examine this issue further, let us discover how these autocorrelations decrease to zero with n (the number of samples). Figure 1.9 shows the autocorrelation samples as a function of n in steps of n increasing by factors of four. Thus $\sqrt{n}$ increases by factors of two. Each autocorrelation in the figure was normalized at zero lag. We see the sample variance for nonzero lags of the autocorrelation dropping off as $\sqrt{n}$. We also observe that the ratios between the values for the first nonzero lags and the value at lag zero roughly fit $1/\sqrt{n}$. Notice also that the fluctuations drop off with lag. The drop-off goes to zero at a lag equal to the sample length, because the number of terms in the autocorrelation diminishes to zero at that lag. A first impression is that the autocorrelation fits a triangular envelope. More careful inspection, however, shows that the triangle bulges upward at wide offsets, or large values of k (this is slightly clearer in Figure 1.8). Each of these observations has an analytic explanation found in **PVI**.

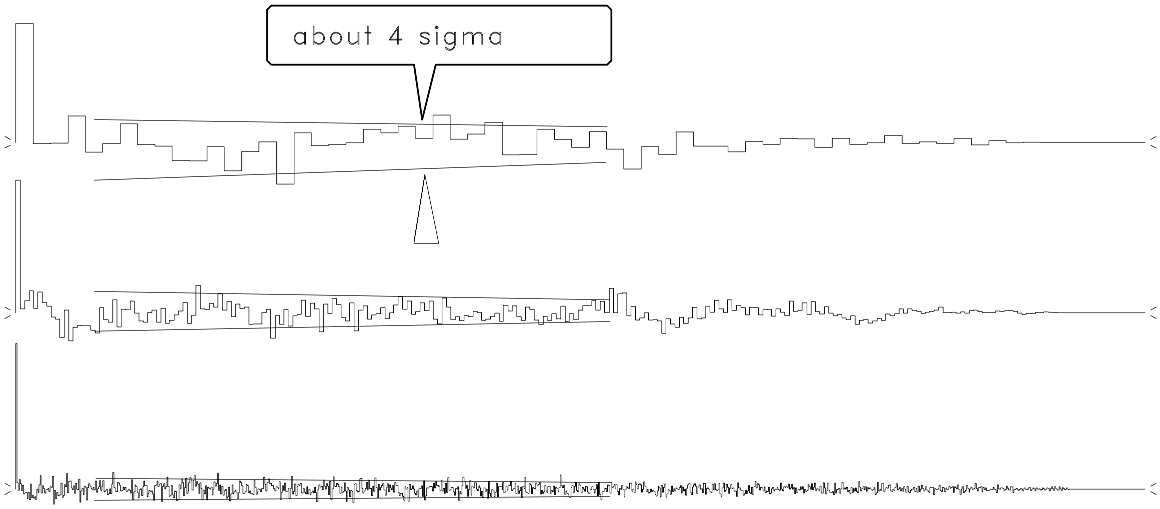

Figure 1.9: Autocorrelation as a function of number of data points. The random-noise-series (even) lengths are 60, 240, 960. rnd-fluct [NR]

1.4.2 An example of the bandwidth/reliability tradeoff

Letting n go to infinity does not take us to the expectation $\hat{S} = \sigma^2$. The problem is, as we increase n, we increase the **frequency resolution** but not the **statistical resolution** (i.e., the fluctuation around $\hat{S}$). To increase the statistical resolution, we need to simulate ensemble averaging. There are two ways to do this:

1. Take the sample of n points and break it into k equal-length segments of n/k points each. Compute an $S(\omega)$ for each segment and then average all k of the $S(\omega)$ together. The variance of the average spectrum is equal to the variance of each spectrum (σ_x^2) *divided* by the number of segments, and so the fluctuation is substantially reduced.

2. Form $S(\omega)$ from the n-point sample. Replace each of the $n/2$ independent amplitudes by an average over its k nearest neighbors. This could also be done by tapering the autocorrelation.

The second method is illustrated in Figure 1.10. This figure shows a **noise burst** of 240 points. Since the signal is even, the burst is effectively 480 points wide, so the autocorrelation is 480 points from center to end: the number of samples will be the same for all cases. The spectrum is very rough. Multiplying the autocorrelation by a triangle function effectively smooths the spectrum by a sinc-squared function, thus reducing the **spectral resolution** ($1/\Lambda F$). Notice that ΛF is equal here to the width of the sinc-squared function, which is inversely proportional to the length of the triangle (ΛT_{auto}).

However, the first taper takes the autocorrelation width from 480 lags to 120 lags. Thus the spectral fluctuations ΛS should drop by a factor of 2, since the count of terms s_k in $S(\omega)$ is reduced to 120 lags. The width of the next weighted autocorrelation width is dropped from 480

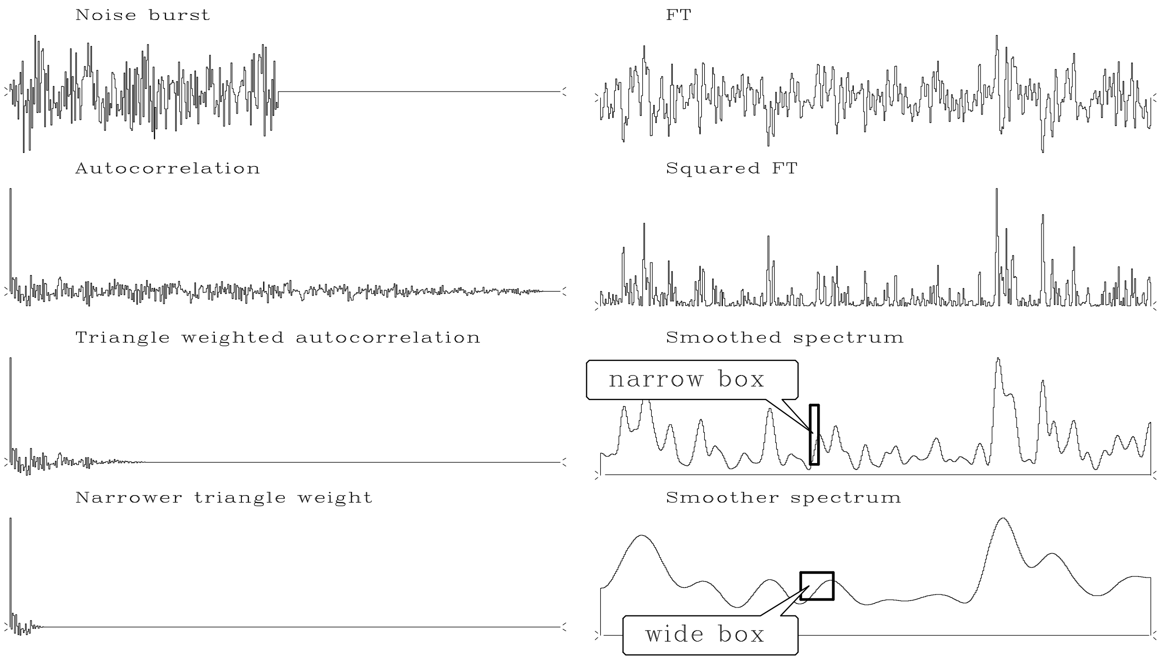

Figure 1.10: Spectral smoothing by tapering the autocorrelation. ΔT is constant and specified on the top row. Successive rows show ΔF increasing while ΔS decreases. The width of a superimposed box roughly gives ΔF, and its height roughly gives ΔS. rnd-taper [NR]

to 30 lags. Spectral roughness should consequently drop by another factor of 2. In all cases, the *average* spectrum is unchanged, since the first lag of the autocorrelations is unchanged. This implies a reduction in the relative spectral fluctuation proportional to the square root of the length of the triangle ($\sqrt{\Delta T_{\mathrm{auto}}}$).

Our conclusion follows:

The trade-off among **resolution**s of time, frequency, and spectral amplitude is

$$\Delta F \, \Delta T \left(\frac{\Delta S}{S} \right)^2 \quad > \quad 1 \tag{1.30}$$

1.4.3 Spectral estimation

In Figure 1.10 we did not care about **spectral resolution**, since we knew theoretically that the spectrum was white. But in practice we do not have such foreknowledge. Indeed, the random factors we deal with in nature rarely are white. A widely used model for naturally occurring random functions, such as microseism, or sometimes reflection seismograms, is **white noise** put into a filter. The spectra for an example of this type are shown in Figure 1.7. We can see that smoothing the envelope of the power spectrum of the output gives an estimate of the spectrum of the filter. But we also see that the estimate may need even more smoothing.

1.5 CROSSCORRELATION AND COHERENCY

With two time series we can see how crosscorrelation and coherency are related.

1.5.1 Correlation

"**Correlation**" is a concept similar to cosine. A cosine measures the angle between two vectors. It is given by the dot product of the two vectors divided by their magnitudes:

$$c \quad = \quad \frac{(\mathbf{x}\cdot\mathbf{y})}{\sqrt{(\mathbf{x}\cdot\mathbf{x})(\mathbf{y}\cdot\mathbf{y})}} \tag{1.31}$$

This is the **sample normalized correlation**.

Formally, the **normalized correlation** is defined using x and y as zero-mean, scalar, random variables instead of sample vectors. The summation is thus an **expectation** instead of a dot product:

$$c \quad = \quad \frac{\mathrm{E}(xy)}{\sqrt{\mathrm{E}(x^2)\,\mathrm{E}(y^2)}} \tag{1.32}$$

A practical difficulty arises when the **ensemble averaging** is simulated over a sample. The problem occurs with small samples and is most dramatically illustrated when we deal with a sample of only one element. Then the sample correlation is

$$\hat{c} \quad = \quad \frac{xy}{|x|\,|y|} \quad = \quad \pm 1 \tag{1.33}$$

regardless of what value the random number x or the random number y should take. For any n, the sample correlation $\hat{c}$ scatters away from zero. Such scatter is called "bias." The topic of bias and variance of coherency estimates is a complicated one, but a rule of thumb seems to be to expect bias and variance of $\hat{c}$ of about $1/\sqrt{n}$ for samples of size n. Bias, no doubt, accounts for many false "discoveries," since cause-and-effect is often inferred from correlation.

1.5.2 Coherency

The concept of "**coherency**" in time-series analysis is analogous to correlation. Taking x_t and y_t to be time series, we find that they may have a mutual relationship which could depend on time delay, scaling, or even filtering. For example, perhaps $Y(Z) = F(Z)X(Z) + N(Z)$, where $F(Z)$ is a filter and n_t is unrelated noise. The generalization of the correlation concept is to define coherency by

$$C \quad = \quad \frac{\mathrm{E}\left[X\left(\frac{1}{Z}\right)Y(Z)\right]}{\sqrt{\mathrm{E}(\overline{X}X)\,\mathrm{E}(\overline{Y}Y)}} \tag{1.34}$$

Correlation is a real scalar. ***Coherency*** is a complex function of frequency; it expresses the frequency dependence of correlation. In forming an estimate of coherency, it is always essential to simulate ensemble averaging. Note that if the ensemble averaging were to be omitted, the coherency (squared) calculation would give

$$|C|^2 \quad = \quad \overline{C}C \quad = \quad \frac{(\overline{\overline{X}Y})(\overline{X}Y)}{(\overline{X}X)(\overline{Y}Y)} \quad = \quad 1 \tag{1.35}$$

which states that the coherency squared is unity, independent of the data. Because correlation scatters away from zero, we find that coherency squared is biased away from zero.

Index

all-pass filter, 2
autocorrelation, 14
autocorrelation, random noise, 13

bandwidth, 1, 4
beat, 7

coherency, 17, 18
correlation, 17, 18
correlation
    normalized, 17
    sample normalized, 17

duration, 1

ensemble, 8
ensemble average, 14
ensemble averaging, 17
expectation, 8, 17
expectation of the autocorrelation, 14
expectation of the spectrum, 14

filter
    all-pass, 2
frequency resolution, 15

Gabor, 4
Gaussian, 2, 3

index, 19
inequality, 2

mean, 8, 9
mean, time-variable, 12

noise burst, 15
Nyquist, 7

pitfall, 13
power spectrum, 13
probability, 9
PVI, 14

quantum mechanics, 4

random, 5
random numbers, FT, 5
random walk, 10
resolution, 1, 16
rise time, 1, 4, 5

sample mean, 8
sinc, 4
spectral bandwidth, 2
spectral resolution, 15, 16
spread, 1
standard deviation, 8
standard deviation of the sample mean, 11
statistic, 8
statistical independence, 9
statistical resolution, 15

time-bandwidth, 1
time-bandwidth product, 3
tolerance, 1
truncated sinusoid, 2

uncertainty principle, 1, 2, 4

variance, 1, 8
variance of the sample mean, 8, 10
variance of the sample variance, 12

white noise, 16