Chapter 4

Resolution

In locating an earthquake or a petroleum drilling site there will be an uncertainty in location, say $(\Delta x, \Delta y, \Delta z)$ caused by measurement errors and the physical size of the target. In measuring a voltage there will be a measuring accuracy Δv. The frequency of useful seismic waves will have a bandwidth $\Delta\omega$. The time at which an earthquake occurs will have an uncertainty given by the duration of shaking Δt. A telescope of diameter Δd has at best a resolving power measured by a certain angular range $\Delta\theta$. It is often desirable to make measurements in such a way as to reduce the quantities Δx, Δy, Δz, Δv, $\Delta\omega$, Δt, Δd, and $\Delta\theta$ to values as small as possible. These measures of resolution (which are called variances, tolerances, uncertainties, bandwidths, durations, spreads, spans, etc.) sometimes interact with one another in such a way that any experimental modification which reduces one must necessarily increase another or some combination of the others. The purpose of this chapter is to discuss some of the commonly occurring situations where such conflicting interactions occur.

In this chapter we use Δt to denote the time duration of a signal. We use τ to denote the amount of time which passes between sample points. In other chapters, Δt is synonymous with τ, the sample interval.

4.1 TIME-FREQUENCY RESOLUTION

The famous "uncertainty principle" of quantum mechanics resulted from observations that subatomic particles behave like waves with wave frequency proportional to particle momentum. The classical laws of mechanics enable prediction of the future of a mechanical system by extrapolation from presently known position and momentum. But because of the wave nature of matter with momentum proportional to frequency, such prediction requires simultaneous knowledge of both the location and the frequency of a wave. A sinusoidal wave has a perfectly clearly determined frequency, but it is spread over the infinitely long time axis. At the other extreme is a delta function, which is nicely compressed to a point on the time axis but contains a mixture of all frequencies. A mathematical analysis of the uncertainty principle is thus an analysis relating functions to their Fourier transforms.

Such an analysis begins by definitions of time duration and spectral bandwidth. The time duration of a damped exponential function is infinite if by duration you mean the span of nonzero function values. However, for nearly all practical purposes the time span is chosen 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) drop to half of the maximum value. The main idea is that the time span Δt or the frequency span $\Delta\omega$ should be able to include most of the total energy but need not contain all of it. The precise definition of Δt and $\Delta\omega$ is somewhat arbitrary and may be chosen to simplify analysis. The general statement is that for any function the time duration Δt and the spectral bandwidth $\Delta\omega$ are related by

$$\Delta\omega \; \Delta t \quad \geq \quad 2\pi \tag{4.1}$$

Although it is easy to verify (4.1) in many special cases, it is not very easy to deduce (4.1) as a general principle. This has, however, been done by D. Gabor. He chose to define Δt and $\Delta\omega$ by second moments.

A similar and perhaps more basic concept than the product of time and frequency spreads is the relationship between spectral bandwidth and *rise time* of a system response function. The *rise time* Δt of a system response is also defined somewhat arbitrarily, often as the time span between the time of excitation and the time at which the system response is half its ultimate value. In principle, a broad frequency response can result from a rapid decay time as well as from a rapid rise time. *Tightness* in the inequality (4.1) may be associated with situations in which a certain rise time is quickly followed by an equal decay time. *Slackness* in the inequality (4.1) may be associated with increasing inequality between rise time and decay time. Slackness could also result from other combinations of rises and falls such as random combinations. Many systems respond very rapidly compared to the rate at which they subsequently decay. Focusing our attention on such systems, we can now seek to derive the inequality (4.1) applied to rise time and bandwidth. The first step is to choose a definition for rise time. The choice is determined not only for clarity and usefulness but also by the need to ensure tractability of the subsequent analysis. I have found a reasonable definition of rise time to be

$$\Delta t \quad = \quad \frac{\int_0^\infty b(t)^2 \, dt}{\int_0^\infty \frac{1}{t}\, b(t)^2 \, dt} \tag{4.2}$$

where $b(t)$ is the response function under consideration. The numerator is just a normalizing factor. The denominator says we have defined Δt by the first negative moment. For example, if $b(t)$ is a step function, then the denominator integral diverges, giving the desired $\Delta t = 0$ rise time. If $b(t)^2$ grows linearly from zero to t_0 and then vanishes, the rise time Δt is $t_0/2$, again a reasonable definition.

Although the Z transform method is a great aid in studying situations where divergence (as $1/t$) plays a key role, it does have the disadvantage that it destroys the formal identity between the time domain and the frequency domain. Presumably this disadvantage is not fundamental since we can always go to a limiting process in which the discretized time domain tends to a continuum. In order to utilize the analytic simplicity of the Z transform we now consider the dual to the rise-time problem. Instead of a time function whose square vanishes identically at negative time we now consider a spectrum $\overline{B}(1/Z)B(Z)$ which vanishes at negative frequencies. We measure how fast this spectrum can rise after $\omega = 0$. We will find this to be related to the time duration Δt of the complex time function b_t. More precisely, we will now define the lowest significant frequency component $\omega = \Delta\omega$ in the spectrum analogously to (4.2) to be

$$\Delta\omega \quad = \quad \frac{\int_{-\infty}^{\infty} \overline{B}B\, d\omega}{\int_{-\infty}^{\infty} \frac{1}{\omega}\overline{B}B\, d\omega} \tag{4.3}$$

Without loss of generality we can assume that the spectrum has been normalized so that the numerator integral is unity. In other words, the zero lag of the autocorrelation of b_t is $+1$. Then

$$\frac{1}{\Delta\omega} \quad = \quad \int_{-\infty}^{\infty} \frac{1}{\omega}\overline{B}B\, d\omega \tag{4.4}$$

Now we recall the bilinear transform which gives us various Z transform expressions for $(-i\omega)^{-1}$. The one we ordinarily use is the integral $(\ldots 1, 1, 1, 0.5, 0, 0, 0\ldots)$. The pole right on the unit circle at $Z = 1$ causes some nonuniqueness. Because $1/i\omega$ is an imaginary odd frequency function we will take the desired expansion to be the odd function of time given by

$$(-i\omega)^{-1} \quad = \quad \frac{(\cdots - Z^{-2} - Z^{-1} + 0 + Z + Z^2 + \cdots)}{2} \tag{4.5}$$

Converting (4.4) to an integral on the unit circle in Z transform notation we have

$$\begin{aligned} \frac{1}{\Delta\omega} &= \int_{-\infty}^{+\infty} -i\frac{1}{-i\omega}\overline{B}B\, d\omega \\ &= \frac{-i}{2\pi}\int_{-\pi}^{+\pi} \frac{1}{2}(\cdots - Z^{-2} - Z^{-1} + Z + Z^2 + \cdots)\,\overline{B}\left(\frac{1}{Z}\right)\, B(Z)\, d\omega \end{aligned} \tag{4.6}$$

But since this integral selects the coefficient of Z^0 of its argument we have

$$\frac{1}{\Delta\omega} \quad = \quad \frac{-i}{2}\left[(r_{-1} - r_1) + (r_{-2} - r_2) + \cdots\right] \tag{4.7}$$

where r_t is the autocorrelation function of b_t. This may be further expressed as

$$\frac{1}{\Delta\omega} \quad = \quad \sum_{t=1}^{\infty} \mathrm{Im}(-r_t) \quad \leq \quad \sum_{t=1}^{\infty} |r_t| \tag{4.8}$$

The sum in (4.8) is like an integral representing area under the $|r_t|$ function. Imagine the $|r_t|$ function replaced by a rectangle function of equal area. This would define a

Δt_{auto} for the $|r_t|$ function. Any autocorrelation function satisfies $|r_t| < r_0$ and we have normalized $r_0 = 1$. Thus, we extend the inequality (4.8) by

$$\frac{1}{\Delta\omega} \quad = \quad \sum_{t=1}^{\infty} |r_t| < \Delta t_{\mathrm{auto}} \tag{4.9}$$

Finally, we must relate the duration of a time function Δt to the duration of its autocorrelation Δt_{auto}. Generally speaking, it is easy to find a long time function which has short autocorrelation. Just take an arbitrary short time function and convolve it by a long and tortuous all-pass filter. The new function is long, but its autocorrelation is short. If a time function has n nonzero points, then its autocorrelation has only $2n-1$ nonzero points. It is obviously impossible to get a long autocorrelation function out of a short time function. It is not even fair to say that the autocorrelation must lie under some tapering function. To construct a time function with as long an autocorrelation as possible, the best thing to do is to concentrate the energy in two lumps, one at each end of the time function. Even from this extreme example, we see that it is not unreasonable to assert that

$$\Delta t \quad \geq \quad \Delta t_{\mathrm{auto}} \tag{4.10}$$

inserting into (4.9) we have the uncertainty relation

$$\Delta t \Delta\omega \quad \geq \quad 1 \tag{4.11}$$

The more usual form of the uncertainty principle uses the frequency variable $f = 2\omega$ and a different definition of Δt, namely *time duration* rather than *rise time*. It is

$$\Delta t \Delta f \quad \geq \quad 1 \qquad (\Delta t \text{ is duration}) \tag{4.12}$$

The choice of a 2π scaling factor to convert *rise time* to *duration* is indicative of the approximate nature of the inequalities.

EXERCISES:

1 Consider $B(Z) = [1 - (Z/Z_0)^n]/(1 - Z/Z_0)$ in the limit Z_0 goes to the unit circle. Sketch the time function and its squared amplitude. Sketch the frequency function and its squared amplitude. Choose Δf and Δt.

2 A time series made up of two frequencies may be written as

$$b_t \quad = \quad A\cos\omega_1 t + B\sin\omega_1 t + C\cos\omega_2 t + D\sin\omega_2 t$$

Given ω_1, ω_2, b_0, b_1, b_2, b_3 show how to calculate the amplitude and phase angles of the two sinusoidal components.

3 Consider the frequency function graphed below. Describe the time function in rough terms indicating the times corresponding to $1/f_1$, $1/f_2$, $1/f_3$. Try to avoid algebraic calculation. Sketch an approximate result.

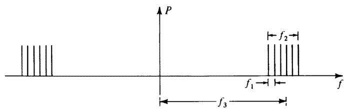

Figure 4.1: Exercise 4.1.3. c4-E4-1-3 [NR]

PROBLEM FOR RESEARCH

Can you find a method of defining $\Delta\omega$ and Δt of one-sided wavelets in such a way that for minimum-phase wavelets only the uncertainty principle takes on the equality sign?

4.2 TIME-STATISTICAL RESOLUTION

If you flipped a coin 100 times, it is possible that you would get exactly 50 "heads" and 50 "tails". More likely it would be something between 60–40 and 40–60. Typically, how much deviation from 50 would you expect to see? The average (mean) value should be 50, but some other value is almost always obtained from a random sample. The other value is called the *sample mean.* We would like to know how much difference to expect between the sample mean and the true mean. The average squared difference is called the *variance of the sample mean.* For a very large sample, the sample mean should be proportionately much closer to the true mean than for a smaller sample. This idea will lead to an uncertainty relation between the probable error in the estimated mean and the size of the sample. Let us be more precise.

The "true value" of the mean could be defined by flipping the coin n times and conceiving of n going to infinity. A more convenient definition of "true value" is that the experiment could be conceived of as having been done separately under identical conditions by an infinite number of people (an ensemble). Such an artifice will enable us to define a time-variable mean for coins which change with time.

The utility of the concept of an ensemble is often subjected to serious attack both from the point of view of the theoretical foundations of statistics and from the point of view of experimentalists applying the techniques of statistics. Nonetheless a great body of geophysical literature uses the artifice of assuming the existence of an unobservable ensemble. The advocates of using ensembles (the Gibbsians) have the advantage over their adversaries (the Bayesians) in that their mathematics is more tractable (and more explainable). So, let us begin!

A conceptual average over the ensemble, called an 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 may be defined as

$$m_t \quad = \quad E(x_t) \tag{4.13}$$

If the mean does not vary with time, we may write

$$m = E(x_t) \qquad \text{(all } t\text{)} \tag{4.14}$$

Likewise, we may be interested in a property of x_t called its *variance* which is a measure of variability about the mean defined by

$$\sigma_t^2 = E\left[(x_t - m_t)^2\right] \tag{4.15}$$

The x_t random numbers could be defined in such a way that σ or m or both is either time-variable or constant. If both are constant, we have

$$\sigma^2 = E\left[(x_t - m)^2\right] \tag{4.16}$$

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

$$E = (\lim n \to \infty) \frac{1}{n} \sum_1^n \tag{4.17}$$

Notice that the summation index is not given, since the sum is over the ensemble, not time.

Now let x_t be a time series made up from (identically distributed, independently chosen) random numbers in such a way that m and σ do not depend on time. Suppose we have a sample of n points of x_t and are trying to determine the value of m. We could make an estimate $\hat{m}$ of the mean m with the formula

$$\hat{m} = \frac{1}{n} \sum_{t=1}^{n} x_t \tag{4.18}$$

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

$$\sum w_t = 1 \tag{4.19}$$

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

$$\hat{m} = \sum w_t x_t \tag{4.20}$$

Actually (4.18) is just a special case of (4.20) where the weights are $w_t = 1/n$; $t = 1, 2, \ldots, n$.

Our objective in this section is to determine how far the estimated mean $\hat{m}$ is likely to be from the true mean m for a sample of length n. One possible definition of this excursion Δm is

$$(\Delta m)^2 = E\left[(\hat{m} - m)^2\right] \tag{4.21}$$

$$= E\left\{\left[\left(\sum w_t x_t\right) - m\right]^2\right\} \tag{4.22}$$

Now utilize the fact that $m = m \sum w_t = \sum w_t m$

$$
\begin{aligned}
(\Delta m)^2 &= E \left\{ \left[\sum_t w_t (x_t - m) \right]^2 \right\} &(4.23)\\
&= E \left\{ \left[\sum_t w_t (x_t - m) \right] \left[\sum_s w_s (x_s - m) \right] \right\} &(4.24)\\
&= E \left[\sum_t \sum_s w_t w_s (x_t - m)(x_s - m) \right] &(4.25)
\end{aligned}
$$

Now the expectation symbol E may be regarded as a summation sign and brought inside the sums on t and s.

$$
(\Delta m)^2 = \sum_t \sum_s w_t w_s \, E \left[(x_t - m)(x_s - m) \right] \tag{4.26}
$$

By the randomness of x_t and x_s the expectation on the right, that is, the sum over the ensemble, gives zero unless $s = t$. If $s = t$, then the expectation is the variance defined by (4.16). Thus we have

$$
\begin{aligned}
(\Delta m)^2 &= \sum_t \sum_s w_t w_s \sigma^2 \sigma_{ts} &(4.27)\\
&= \sum_t w_t^2 \sigma^2 &(4.28)
\end{aligned}
$$

or

$$
\Delta m = \sigma \left(\sum_t w_2^2 \right)^{1/2} \tag{4.29}
$$

Now let us examine this final result for n weights each of size $1/n$. For this case, we get

$$
\Delta m = \sigma \left[\sum_1^n \left(\frac{1}{n} \right)^2 \right]^{1/2} = \frac{\sigma}{(n)^{1/2}} \tag{4.30}
$$

This is the most important property of random numbers which is not intuitively obvious. For a zero mean situation it may be expressed in words: "n random numbers of unit magnitude add up to a magnitude of about the square root of n."

When one is trying to estimate the mean of a random series which has a time-variable mean, one faces a basic dilemma. If one includes a lot of numbers in the sum to get Δm small, then m may be changing while one is trying to measure it. In contrast, $\hat{m}$ measured from a short sample of the series might deviate greatly from the true m (defined by an infinite sum over the ensemble at any point in time). This is the basic dilemma faced by a stockbroker when a client tells him, "Since the market fluctuates a lot I'd like you to sell my stock sometime when the price is above the mean selling price."

If we imagine that a time series is sampled every τ seconds and we let $\Delta t = n\tau$ denote the length of the sample then (4.30) may be written as

$$
(\Delta m)^2 \, \Delta t = \sigma^2 t \tag{4.31}
$$

It is clearly desirable to have both Δm and Δt as small as possible. If the original random numbers x_t were correlated with one another, for example, if x_t were an approximation to a continuous function, then the sum of the n numbers would not cancel to root n. This is expressed by the inequality

$$(\Delta m)^2 \, \Delta t \quad \geq \quad \sigma^2 t \tag{4.32}$$

The inequality (4.32) may be called an uncertainty relation between accuracy and time resolution.

In considering other sets of weights one may take a definition of Δt which is more physically sensible than τ times the number of weights. For example, if the weights w_t are given by a sampled gaussian function as shown in Figure 4.2 then Δt could be taken as the separation of half-amplitude points, $1/e$ points, the time span which includes 95 percent of the area, or it could be given many other "sensible" interpretations. Given a little slop in the definition of Δm and Δt, it is clear that the inequality of (4.32) is not to be strictly applied.

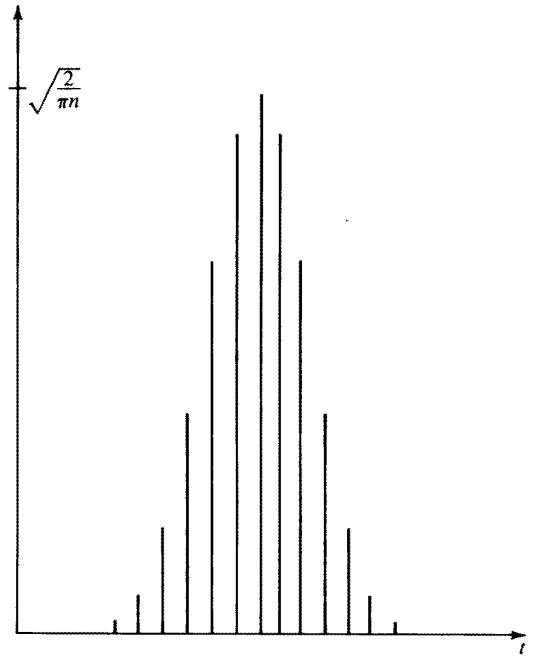

Figure 4.2: Binomial coefficients tend to the gaussian function. Plotted are the coefficients of Z^t in $(.5 + .5Z)^{20}$. $\boxed{\text{c4-4-1}}$ [NR]

Given a sample of a zero mean random time series x_t, we may define another series y_t by $y_t = x_t^2$. The problem of estimating the variance $\sigma^2 = p$ of x_t is identical to the problem of estimating the mean m of y_t. If the sample is short, we may expect an error Δp in our estimate of the variance. Thus, in a scientific paper one would like to write for the mean

$$\begin{aligned} m &= \hat{m} \pm \Delta m && (4.33) \\ &= \hat{m} \pm \sigma/\sqrt{n} && (4.34) \end{aligned}$$

but since the variance σ^2 often is not known either, it is necessary to use the estimated $\hat{\sigma}$, that is

$$m \quad = \quad \hat{m} \pm \sigma/\sqrt{n} \tag{4.35}$$

Of course (4.35) really is not right because we really should add something to indicate additional uncertainty due to error in $\hat{\sigma}$. This estimated error would again have an error, ad infinitum. To really express the result properly, it is necessary to have a probability density function to calculate all the $E(x^n)$ which are required. The probability function can be either estimated from the data or chosen theoretically. In practice, for a reason given in a later section, the gaussian function often occurs. In the exercise it is shown that

$$\Delta p \quad = \quad p\sqrt{\frac{2}{n}} \tag{4.36}$$

Since $\Delta t = n\tau$, by squaring we have

$$\left(\frac{\Delta p}{p}\right)^2 \frac{\Delta t}{2\tau} \quad \geq \quad 1 \tag{4.37}$$

The inequality applies if the random numbers x_t are not totally unpredictable random numbers. If x_t is an approximation to a continuous function, then it is highly predictable and there will be a lot of slack in the inequality.

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

$$c \quad = \quad \frac{(\mathbf{x} \cdot \mathbf{y})}{[(\mathbf{x} \cdot \mathbf{x})(\mathbf{y} \cdot \mathbf{y})]^{1/2}}$$

Correlation is the same sort of thing, except x and y are scalar random variables, so instead of having a vector subscript their subscript is the implicit ensemble subscript. Correlation is defined

$$c \quad = \quad \frac{E(xy)}{[E(x^2)\, E(y^2)]^{1/2}}$$

In practice one never has an ensemble. There is a practical problem when the ensemble average is simulated by averaging over a sample. The problem arises with small samples and is most dramatically illustrated for a sample with only one element. Then the sample correlation is

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

regardless of what value the random number x or the random number y should take. In fact, it turns out that the sample correlation $\hat{c}$ will always scatter away from zero.

No doubt this accounts for many false "discoveries". 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 $\hat{c}$ on the order of $1/\sqrt{n}$ for samples of size n.

EXERCISES:

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

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

Find the scale factor s so that $E(\hat{m}) = m$. Calculate Δm. Define a reasonable Δt. 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 Δt, the effective averaging time?

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

(d) Compute the random error $\Delta m = [E(\hat{m} - m)^2]^{1/2}$ [answer goes to $\sigma(\epsilon/2)^{1/2}$ as ϵ goes to zero].

(e) What is $(\Delta m)^2 \Delta t$ in this case?

3 Show that

$$(\Delta P)^2 \quad = \quad \frac{1}{n}[E(x^4) - \sigma^4]$$

4 Define the behavior of an independent zero-mean-tie series x_t by defining the probabilities that various amplitudes will be attained. Calculate $E(x_i)$, $E(x_i^2)$, $(\Delta P)^2$. If you have taken a course in probability theory, use a gaussian probability density function for x_i. HINT:

$$P(x) \quad = \quad \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/2\sigma^2}$$

and

$$\int_0^\infty x^{2n} e^{-ax^2}\, dx \quad = \quad \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2^{n+1}a^n}\sqrt{\frac{\pi}{a}}$$

4.3 FREQUENCY-STATISTICAL RESOLUTION

Observations of sea level for 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 such geophysical time series is the *power spectrum*. Some mathematical models explain only statistical averages of data and not the data themselves. In order to recognize certain pitfalls and understand certain fundamental limitations on work with power spectra, we first consider an idealized example.

Let x_t be a time series made up of independently chosen random numbers. Suppose we have n of these numbers. We can then define the data sample polynomial $X(Z)$

$$X(Z) \quad = \quad x_0 + x_1 Z + x_2 Z^2 + \cdots + x_{n-1} Z^{n-1} \tag{4.38}$$

We can now make up a power spectral estimate $\hat{R}(Z)$ from this sample of random numbers by

$$\hat{R}(Z) \quad = \quad \frac{1}{n} \overline{X}\left(\frac{1}{Z}\right) X(Z) \tag{4.39}$$

The difference between this and our earlier definition of spectrum is that a *power* spectrum has the divisor n to keep the expected result from increasing linearly with the somewhat arbitrary sample size n.

The definition of power spectrum is the expected value of $\hat{R}$, namely

$$R(Z) \quad = \quad E[\hat{R}(Z)] \tag{4.40}$$

It might seem that a practical definition would be to let n tend to infinity in (4.39). Such a definition would lead us into a pitfall which is the main topic of the present section. Specifically, from Figure 4.3 we conclude that $\hat{R}(Z)$ is a much fuzzier function than $R(Z)$, so that

$$R(Z) \quad \neq \quad \lim_{n\to\infty} \hat{R}(Z) \tag{4.41}$$

To understand why the spectrum is rough, we identify coefficients of like powers of

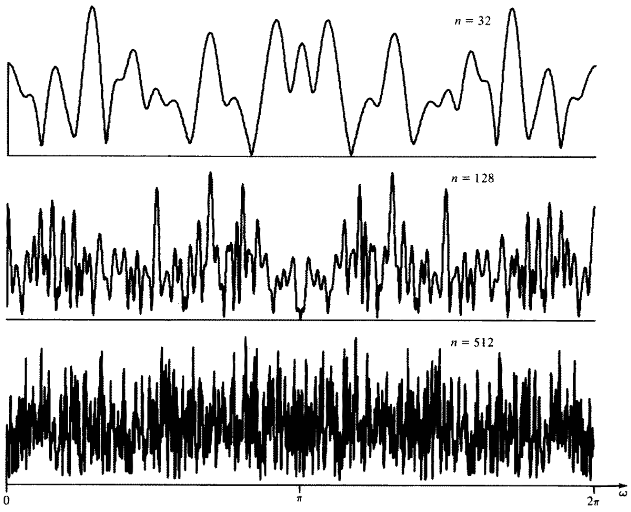

Figure 4.3: Amplitude spectra $\hat{R}(Z)^{1/2}$ of samples of n random numbers. These functions seem to oscillate over about the same range for $n = 512$ as they do for $n = 32$. As n tends to infinity we expect infinitely rapid oscillation. $\boxed{\text{c4-4-2}}$ [NR]

Z in (4.39). We have

$$\hat{r}_k \quad = \quad \frac{1}{n} \sum_{t=0}^{n-k-1} \bar{x}_t x_{t+k} \qquad k \quad = \quad -n+1 \;\text{ to }\; n-1 \tag{4.42}$$

enabling us to write (4.39) for real time series $x_t = \bar{x}_t$ as

$$\hat{R} \quad = \quad \hat{r}_0 + 2\sum_{k=1}^{n-1} \hat{r}_k \cos k\omega \tag{4.43}$$

Let us examine (4.43) for large n. To do this, we will need to know some of the statistical properties of the random numbers. Let them have zero mean $m = E(x_t) = 0$ and let them have known constant variance $\sigma^2 = E(x_t^2)$ and recall our assumption of independence which means that $E(x_t x_{t+s}) = 0$ if $0 \neq s$. Because of random fluctuations, we have learned to expect that $\hat{r}_0$ will come out to be σ^2 plus a random fluctuation component which decreases with sample size as $1/\sqrt{n}$, namely

$$\hat{r}_0 \quad = \quad \sigma^2 \pm \frac{\sigma^2}{\sqrt{n}} \tag{4.44}$$

Likewise, $\hat{r}_1$ should come out to be zero but the definition (4.42) leads us to expect a fluctuation component

$$\hat{r}_1 \quad = \quad \pm\frac{n-1}{n}\frac{\sigma^2}{\sqrt{n}} \tag{4.45}$$

For the kth correlation value $k > 1$ we expect a fluctuation of order

$$\hat{r}_k \quad = \quad \pm\frac{n-k}{n}\frac{\sigma^2}{\sqrt{n}} \tag{4.46}$$

An autocorrelation of a particular set of random numbers is displayed in Figure 4.4. Now one might imagine that as n goes to infinity the fluctuation terms vanish and

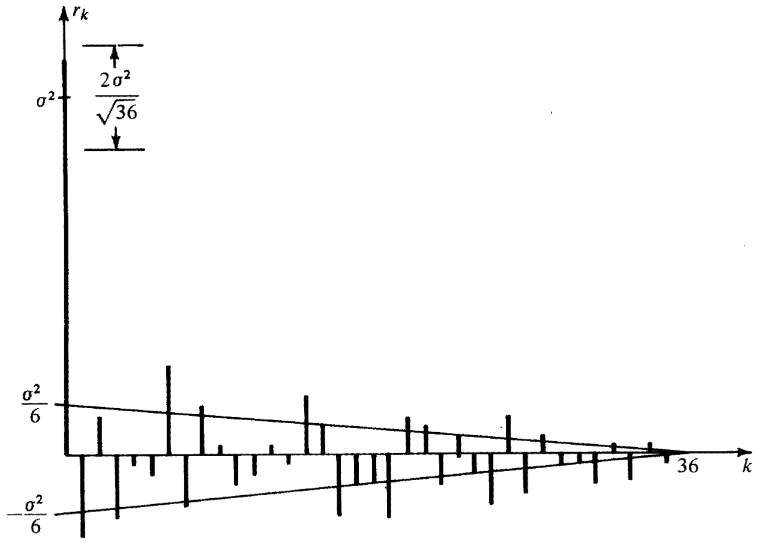

Figure 4.4: Positive lags of autocorrelation of 36 random numbers. $\boxed{\text{c4-4-3}}$ [NR]

(4.39) takes the limiting form $\hat{R} = \sigma^2$. Such a conclusion is false. The reason is that although the individual fluctuation terms go as $1/\sqrt{n}$ the summation in (4.43) contains n such terms. Luckily, these terms are randomly canceling one another so the sum does not diverge as $\sqrt{n}$. We recall that the sum of n random signed numbers of unit magnitude is expected to add up to a random number in the range $\pm\sqrt{n}$. Thus the sum (4.43) adds up to

$$\hat{a} \quad \approx \quad \left(1 \pm \frac{\sqrt{n}}{\sqrt{n}}\right) \sigma^2 \quad = \quad (1 \pm 1)\,\sigma^2 \tag{4.47}$$

This is the basic result that a power spectrum estimated from the energy density of a sample of random numbers has a fluctuation from frequency to frequency and from sample to sample which is as large as the expected spectrum.

It should be clear that letting n go to infinity does not take us to the theoretical result $\hat{R} = \sigma^2$. The problem is that, as we increase n, we increase the frequency resolution but not the statistical resolution. 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 $R(\omega)$ for each segment and then add all k of the $R(\omega)$ together, or (2) form $R(\omega)$ from the n-point sample. Of the $n/2$ independent amplitudes, replace each one by an average over its k nearest neighbors. Whichever method, (1) or (2), is used it will be found that $\Delta f = 0.5\ k/n\tau$ and $(\Delta p/p)^2 =$ inverse of number of degrees of freedom averaged over $= 1/k$. Thus, we have

$$\Delta f \left(\frac{\Delta p}{p}\right)^2 \quad = \quad \frac{.5}{n\tau}$$

If some of the data are not used, or are not used effectively, we get the usual inequality

$$\Delta f \left(\frac{\Delta p}{p}\right)^2 \quad \geq \quad \frac{1}{2n\tau}$$

Thus we see that, if there are enough data available (n large enough), we can get as good resolution as we like. Otherwise, improved statistical resolution is at the cost of frequency resolution and vice versa.

We are right on the verge of recognizing a resolution tradeoff, not only between Δf and Δp but also with $\Delta t = n\tau$, the time duration of the data sample. Recognizing now that the time duration of our data sample is given by $\Delta t = n\tau$, we obtain the inequality

$$\Delta f\ \Delta t\ \left(\frac{\Delta p}{p}\right)^2 > \frac{1}{2} \tag{4.48}$$

The inequality will be further interpreted and rederived from a somewhat different point of view in the next section.

In time-series analysis we have the concept of coherency which is analogous to the concept of correlation defined in Sec. 4-2. There we had for two random variables x and y that

$$c \quad = \quad \frac{E(xy)}{[E(x^2)\,E(y^2)]^{1/2}}$$

Now if x_t and y_t are time series, they may have a relationship between them which depends 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

$$C \quad = \quad \frac{E\left[X\left(\frac{1}{Z}\right) Y(Z)\right]}{[E(\overline{X}X)\,E(\overline{Y}Y)]^{1/2}}$$

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

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

which states that the coherency squared is +1 independent of the data. Because correlation scatters away from zero we find that coherency squared is biased away from zero.

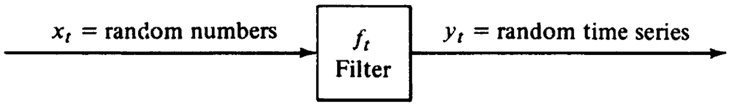

Figure 4.5: Model of random time series generation. $\boxed{\text{c4-4-4}}$ [NR]

4.4 TIME-FREQUENCY-STATISTICAL RESOLUTION

Many time functions are not completely random from point to point but become more random when viewed over a longer time scale. A popular mathematical model embodying this concept is to make a so-called *stationary time series* by putting random numbers into a filter as depicted in Figure 4.5. The input x_t may be *independent random numbers* or *white light.* [The two terms mean nearly the same thing in practice but the first term is the stronger; it means that x_t is in no way related to x_s if $t \neq s$, whereas white light means that $E(x_t x_s) = 0$ if $t \neq s$.] The output random time series y_t may vary rather slowly from point to point if f_t is a low-pass filter. This is the usual case when we are modeling continuous time functions. The random time series may be called a *stationary* random time series if neither the filter nor any property of the random numbers (such as m or σ) vary with time. Stationarity is often assumed even wwhere it cannot be strictly true.

This model will be useful later when we consider the problem of predicting a future point on y_t from knowledge of past values. Now we will use the model to examine the estimation of the spectrum of y_t given a sample of n points of y_t. To begin with, we have a very precise meaning for the spectrum of y_t. We have

$$Y(Z) \quad = \quad F(Z)\,X(Z) \tag{4.49}$$

and its conjugate

$$\overline{Y}\left(\frac{1}{Z}\right) \quad = \quad \overline{F}\left(\frac{1}{Z}\right)\overline{X}\left(\frac{1}{Z}\right) \tag{4.50}$$

Multiplying (4.49) by (4.50) we get

$$\overline{Y}\left(\frac{1}{Z}\right) Y(Z) \quad = \quad \overline{X}\left(\frac{1}{Z}\right) X(Z)\,\overline{F}\left(\frac{1}{Z}\right) F(Z) \tag{4.51}$$

but, from the previous section, we learned that $E(\overline{X}X) = \sigma^2$. Considering σ^2 to be unity, we see that the expected power spectrum of the output Y is the energy spectrum of the filter F. The overall situation is depicted in Figure 4.6. The interesting question is how well can we estimate the spectrum when we start with an n-point sample of y_t. We will describe three computationally different methods, all having the same fundamental limitations.

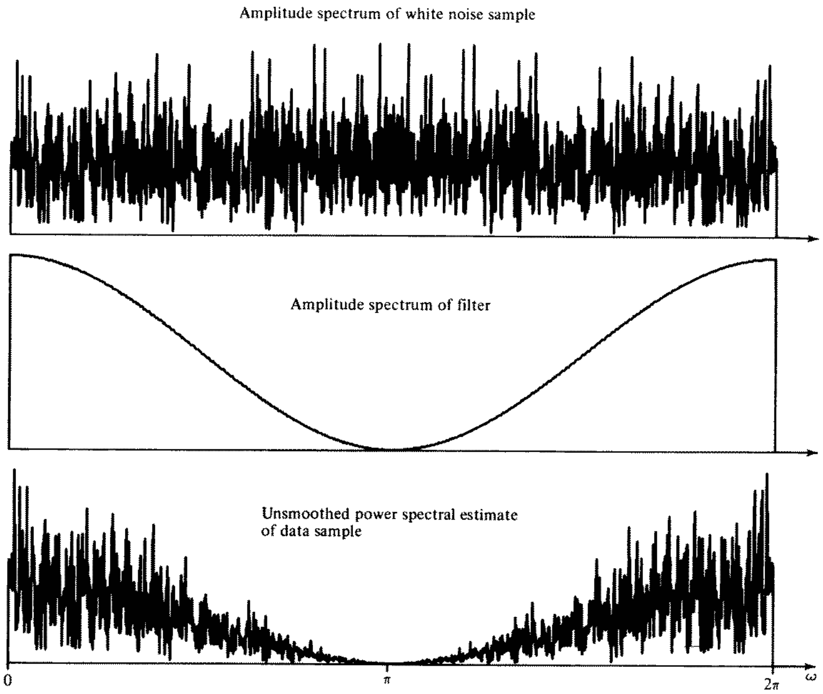

Figure 4.6: Spectral estimation. $\boxed{\text{c4-4-5}}$ [NR]

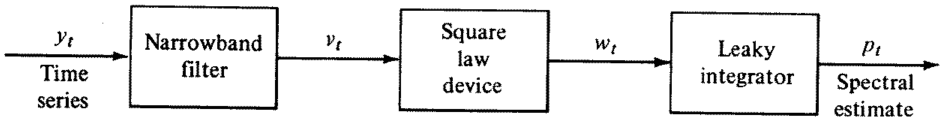

Figure 4.7: Spectral estimate of a random series. $\boxed{\text{c4-4-6}}$ [NR]

The first method uses a bank of filters as shown in Figure 4.7. When random numbers excite the narrowband filter, the output is somewhat like a sine wave. It differs in one important respect. A sine wave has constant amplitude, but the output of a narrowband filter has an amplitude which swings over a range. This is illustrated in Figure 4.8. If the bandwidth is narrow, the amplitude changes slowly. If the impulse response of the filter has duration Δt_{filter}, then the output amplitude at time t will be randomly related to the amplitude at time $t+\Delta t_{\text{filter}}$. Thus, in statistical averaging, it is not the number of time points but the number of intervals Δt_{filter} which enhance the reliability of the average. Consequently, the decay time of the integrator $\Delta t_{\text{integrator}}$ will generally be chosen to be greater than $\Delta t_{\text{filter}} = 1/\Delta f$. The variability Δp of the output p decreases as $\Delta t_{\text{integrator}}$ increases. Since v_t has independent values over time spans of about $\Delta t_{\text{filter}} = 1/\Delta f$, then the "degrees of freedom" smoothed over can be written $\Delta t_{\text{integrator}}/\Delta t_{\text{filter}} = \Delta f \Delta t_{\text{integrator}}$. The variability $\Delta p/p$ is proportional to the inverse square root of the number of degrees of freedom, and so we get

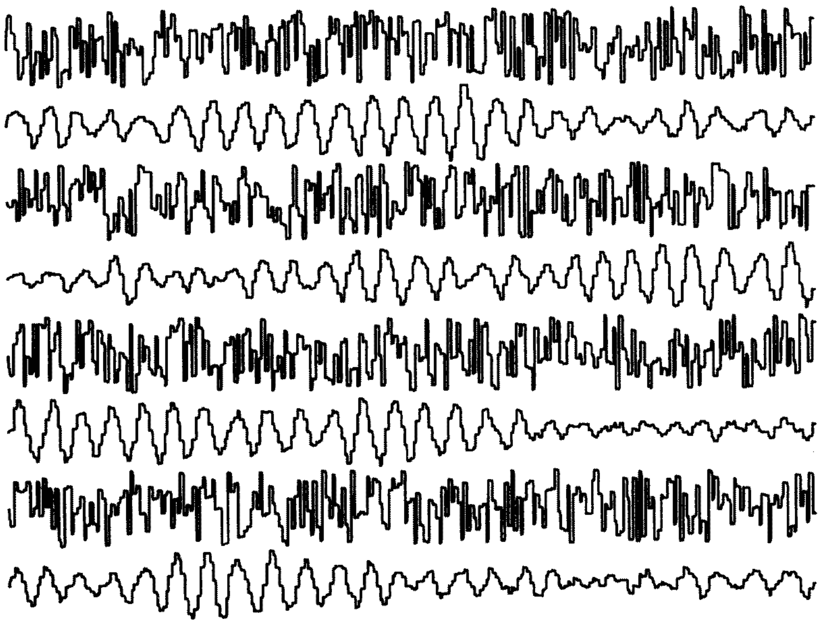

Figure 4.8: 1024 random numbers before and after narrowband filtering. The filter was $(1-Z)/((1-Z/Z_0)\,(1-Z/\overline{Z}_0))$ where $Z_0 = 1.02\, e^{i\pi/5}$. $\boxed{\text{c4-4-7}}$ [NR]

$$\left(\frac{\Delta p}{p}\right)^2 \quad = \quad \frac{1}{\Delta f\ \Delta t_{\text{integrator}}}$$

or, introducing the usual inequality,

$$\Delta t\, \Delta f\ \left(\frac{\Delta p}{p}\right)^2 \quad > \quad 1 \tag{4.52}$$

The inequality (4.52) indicates the three-parameter uncertainty which is fundamental to estimating power spectra of random functions. Two other methods of estimating the spectrum of y_t from a sample of length n are exactly the same as the methods described in Sec. 4-3 as ways of estimating the spectrum of white light. In fact, (4.52) turns out to be the same as (4.48).

The usual interpretation is that to attain a frequency resolution of Δf and a relative accuracy of $\Delta p/p$ a time sample of duration at least $\Delta t > 1/[\Delta f\ (\Delta p/p)^2]$ will be required. Although this sort of interpretation is generally correct, it will break down for highly resonant series recorded for a short time. Then the data sample may be predictable an appreciable distance off its ends so that the effective Δt is somewhat (perhaps appreciably) larger than the sample length.

EXERCISES:

1 It is popular to taper the ends of a data sample so that the data go smoothly to zero at the ends of the sample. Choose a weighting function and discuss in a semiquantitative fashion its effect on Δt, Δf, and $(\Delta p/p)^2$.

2 Answer the question of Exercise 1, where the autocorrelation function is tapered rather than the data sample.

4.5 THE CENTRAL-LIMIT THEOREM

The central-limit theorem of probability and statistics is perhaps the most important theorem in the fields of probability and statistics. A derivation of the central limit theorem explains why the gaussian probability function is so frequently encountered in nature; not just in physics but also in the biological and social sciences. No experimental scientist should be unaware of the basic ideas behind this theorem. Although the result is very deep and is even today the topic of active research, we can get to the basic idea quite easily.

One way to obtain random integers from a known probability function is to write integers on slips of paper and place them in a hat. Draw one slip at a time. After each drawing replace the slip in the hat. The probability of drawing the integer i is given by the ratio a_i of the number of slips containing the integer i divided by the total number of slips. Obviously the sum over i of a_i must be unity. Another way to get random integers is to throw one of a pair of dice. Then all a_i equal zero except $a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = \frac{1}{6}$. The probability that the integer i will occur on the first drawing and the integer j will occur on the second drawing is $a_i a_j$. If you draw two slips or throw a pair of dice, then the probability that the sum of i and j equals k is readily seen to be

$$c_k \quad = \quad \sum_i a_i a_{k-i} \tag{4.53}$$

Since this equation is a convolution, we may look into the meaning of the Z transform

$$A(Z) \quad = \quad \cdots a_{-1} Z^{-1} + a_0 + a_1 Z + a_2 Z^2 + \cdots \tag{4.54}$$

In terms of Z transforms the probability that i plus j equals k is simply the coefficient of Z^k in

$$C(Z) \quad = \quad A(Z)\,A(Z) \tag{4.55}$$

Obviously, if we add n of the random numbers, the probability that the sum of them equals k is given by the coefficient of Z^k in

$$G(Z) \quad = \quad [A(Z)]^n \tag{4.56}$$

The central-limit theorem of probability says that as n does to infinity the polynomial $G(Z)$ goes to a special form, almost regardless of the specific polynomial $A(Z)$. The specific form is such that a graph of the coefficients of $G(Z)$ comes closer and closer to fitting under the envelope of the bell-shaped gaussian function. Let us see why this happens. Our development will lack a mathematical rigor because the theorem is not always true. There are pathological A functions which do not result in G tending to gaussian. Despite the fact that some of the pathological functions sometimes turn up in applications, we will not take the time here to look at such instances.

Consider the size of $A(Z)$ for real ω. If $\omega = 0$, the sum of the terms of $A(Z)$ may be visualized in the complex plane as a sum of vectors $a_k e^{i\omega k}$ all pointing in the positive real direction. If $\omega \neq 0$ the vectors point in different directions. This is shown in Figure 4.9.

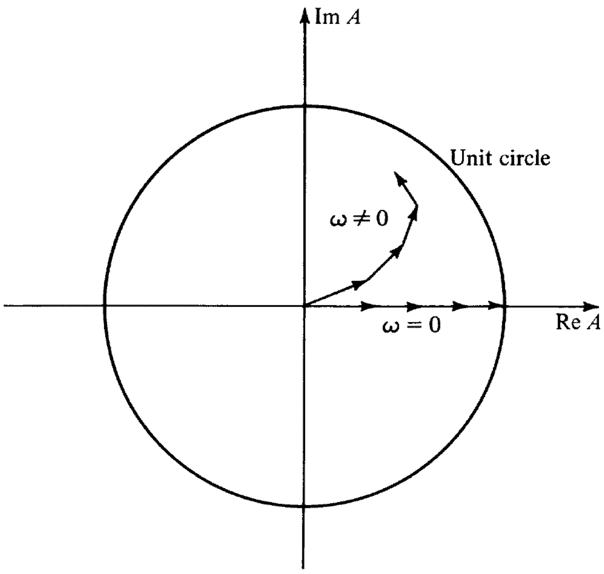

Figure 4.9: The complex numbers $a_k e^{i\omega k}$ added together. c4-4-8 [NR]

In raising $A(e^{i\omega})$ to the nth power, the values of ω of greatest concern are those near $\omega = 0$ where A is largest—because in any region where A is small A^n will be extremely small. Near $\omega = 0$ or $Z = 1$ we may expand $A(Z)$ in a power series in ω

$$A(e^{i\omega}) \quad = \quad A|_0 + \left.\frac{\partial A}{\partial \omega}\right|_0 \omega + \left.\frac{\partial^2 A}{\partial \omega^2}\right|_0 \frac{\omega^2}{2!} + \cdots \tag{4.57}$$

Note that the coefficients of this power series are proportional to the moments m_i of the probability function; that is

$$A(e^{i\omega}) = \sum_k a_k e^{ik\omega} \tag{4.58}$$

$$A(1) = \sum_k a_k = m_0 = 1 \tag{4.59}$$

$$\frac{\partial A}{\partial \omega} = \sum_k ika_k e^{ik\omega} \tag{4.60}$$

$$\left.\frac{\partial A}{\partial \omega}\right|_0 = \sum_k ika_k = im_1 \tag{4.61}$$

$$\left.\frac{\partial^2 A}{\partial \omega^2}\right|_0 = -\sum k^2 a_k = -m_2 \tag{4.62}$$

When we raise $A(Z)$ to the nth power we will make the conjecture that only the given first three terms of the power series expansion will be important. (This assumption clearly fails if any of the moments of the probability function are infinite.) Thus, we are saying that as far as G is concerned the only important things about A are its mean value $m = m_1$ and its second moment m_2. If this is really so, we may calculate G by replacing A with any function B having the same mean and same second moment as A. We may use the simplest function we can find. A good choice is the so called binomial probability function given by

$$B = \frac{Z^m(Z^\sigma + Z^{-\sigma})}{2} \tag{4.63}$$

$$= \frac{e^{i(m+\sigma)\omega} + e^{i(m-\sigma)\omega}}{2} \tag{4.64}$$

Let us verify its first moment

$$\frac{\partial B}{\partial \omega} = \frac{i\left[(m+\sigma)\, e^{i(m+\sigma)\omega} + (m-\sigma)\, e^{i(m-\sigma)\omega}\right]}{2} \tag{4.65}$$

$$\left.\frac{\partial B}{\partial \omega}\right|_0 = im \tag{4.66}$$

Now let us verify its second moment

$$\left.\frac{\partial^2 B}{\partial \omega^2}\right|_0 = -\frac{(m+\sigma)^2 + (m-\sigma)^2}{2} \tag{4.67}$$

$$\left.\frac{\partial^2 B}{\partial \omega^2}\right|_0 = -(m^2 + \sigma^2) \tag{4.68}$$

Hence, σ should be chosen so that

$$m_2 = m^2 + \sigma^2 \tag{4.69}$$

Of course, we cannot expect that m and σ will necessarily turn out to be integers; therefore (4.63) will not necessarily be a Z transform in the usual sense. It does not really matter; we simply interpret (4.63) as saying:

1. The probability of drawing the number $m + \sigma$ is one-half.
2. The probability of $m - \sigma$ is one-half.
3. The probability of any other number is zero.

Now, raising $(Z^\sigma + Z^{-\sigma})$ to the nth power gives a series in powers of Z^σ whose coefficients are symmetrically distributed about Z to the zero power and whose magnitudes are given by the binomial coefficients. A sketch of the coefficients of $B(Z)^n$ is given in Figure 4.10.

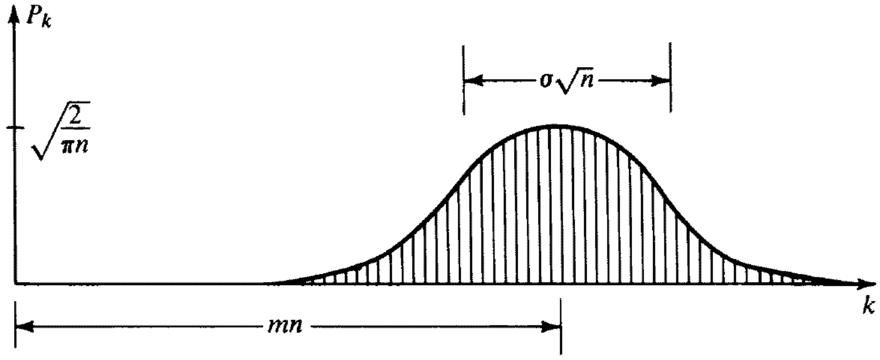

Figure 4.10: Coefficients of binomial raised to power $\boxed{\text{c4-4-9}}$ [NR]

We will now see how, for large n, the binomial coefficients asymptotically approach a gaussian. Approaching this limit is a bit tricky. Obviously, the sum of n random integers will diverge as $\sqrt{n}$. Likewise the coefficients of powers of Z in $(\frac{1}{2} + \frac{1}{2}Z)^n$ individually get smaller while the number of coefficients gets larger. We recall that in time series analysis we used th substitution $Z = e^{i\omega\,\Delta t}$. We commonly chose $\Delta t = 1$, which had the meaning that data points were given at integral points on the time axis. In the present probability theory application of Z transforms, the choice $\Delta t = 1$ arises from our original statement that the numbers chosen randomly from the slips of paper were integers. Now we wish to add n of these random numbers together; and so, it makes sense to rescale the integers to be integers divided by $\sqrt{n}$. Then we can make the substitution $Z = e^{i\omega\,\Delta t} = e^{i\omega/\sqrt{n}}$. The coefficient of Z^k now refers to the probability of drawing the number $k/\sqrt{n}$. Raising $(Z^\sigma + Z^{-\sigma})/2$ to the nth power to find the probability distribution for the sum of n

$$
\begin{aligned}
[B(Z)]^n &= \left(\frac{Z^\sigma + Z^{-\sigma}}{2}\right)^n \\
&= \left(\cos\frac{\sigma\omega}{\sqrt{n}}\right)^n
\end{aligned}
$$

Using the first term of the series expansion for cosine we have

$$[B(Z)]^n \approx \left(1 - \frac{\sigma^2\omega^2}{2n}\right)^n$$

Using the well-known fact that $(1 + x/n)^n \to e^x$, we have for large n

$$[B(Z)]^n \approx e^{-\sigma^2\omega^2/2} \tag{4.70}$$

The probability that the number t will result from the sum is now found by inverse Fourier transformation of (4.70). The Fourier transform of the gaussian (4.70) may be looked up in a table of integrals. It is found to be the gaussian

$$p(t) \quad = \quad \frac{1}{\sigma\sqrt{2\pi}}\, e^{-t^2/2\sigma^2}$$

4.6 CONFIDENCE INTERVALS

It is always important to have some idea of the size and influence of random errors. It is often important to be able to communicate this idea to others in the form of a statement such as

$$m \quad = \quad \hat{m} \pm \sigma/\sqrt{n}$$

In a matter of any controversy you may be called upon to define a probability that the true mean lies in your stated interval; in other words, what is your *confidence* that m lies in the *interval*

$$\hat{m} - \Delta m < m < \hat{m} + \Delta m$$

Before you can answer questions about probability, it is necessary to make some assumptions and assertions about the probability functions which control your random errors. The assertion that errors are independent of one another is your most immediate hazard. If they are not, as is often the case, you may be able to readjust the numerical value of n to be an estimate of the number of independent errors. We did something like this in time series analysis when we took n to be not the number of points on the time series but the number of intervals of length Δt_{filter}. The second big hazard in trying to state a confidence interval is the common assumption that, because of the central-limit theorem and for lack of better information, the errors follow a gaussian probability function. If in fact the data errors include blunders which arise from human errors or blunders from transient electronic equipment difficulties, then the gaussian assumption can be very wrong and can lead you into serious errors in geophysical interpretation. Some useful help is found in the field of nonparametric statistics.

To begin with, it is helpful to rephrase the original question into one involving the median rather than the mean. The median m_1 is defined as that value which is expected to be less than half of the population and greater than the other half. In many—if not most—applications the median is a ready, practical substitute for the arithmetic mean. The median is insensitive to a data point, which, by some blunder, is near infinity. In fact, median and mean are equal when the probability function is symmetrical. For a sample of n numbers $(x_i,\ i = 1, 2, \ldots, n)$, the median m_1 may be estimated by reordering the numbers from smallest to largest and then selecting the number in the middle as the estimate of the median $\hat{m}_1$. Specifically, let the recordered x_i be denoted by x'_i where $x'_i \leq x'_{i+1}$. Then we have $\hat{m}_1 = x'_{n/2}$. Now it turns out that without knowledge of the probability density function for the random variables x_i we will still be able to compute the probability that the true median m_1 is contained in the interval

$$x'_{n/2-\alpha\sqrt{n}} < m_1 < x'_{n/2+\alpha\sqrt{n}} \tag{4.71}$$

For example, set $\alpha = 1$ and $N = 100$, the assertion is that we can now calculate the probability that the true median m_1 lies between the 40th and the 60th percentile of our data. The trick is this: Define a new random variable

$$y \quad = \quad \text{step } (x - m_1) \tag{4.72}$$

The step function equals $+1$ if $x > m_1$ and equals 0 if $x < m_1$. The new random variable y takes on only values of zero and one with equal probability; thus we know its probability function even though we may not know the probability function for the random variable x. Now define a third random variable s as

$$s \quad = \quad \sum_{i=1}^{n} y_i$$

Since each y_i is zero or one, then s must be an integer between zero and n. Further more, the probability that s takes the value j is given by the coefficient of Z^j of $(\frac{1}{2}+Z/2)^n$. Now the probability that s lies in the interval $n/2-\alpha\sqrt{n} < s < n/2+\alpha\sqrt{n}$ is readily determined by adding the required coefficients of Z^j, and this probability is by definition equal to the probability that the median m_1 lies in the interval (4.71). For $\alpha = 1$ and large n this probability works out to about 95 percent.