Chapter 12

Case Studies

To enable you to further explore the ways the concepts of calculus are used as analytical tools in scientific and mathematical investigations, this chapter presents four extended case studies. The four can be studied separately, although the first two and the last two are loosely linked

Stirling's Formula As an example of the way many of the ideas—Taylor series, numerical integration, reduction formulas, limits—developed in the earlier chapters of this book can be used in a tightly-reasoned argument to produce some powerful mathematical insights, in the first section we derive a famous formula approximating $n!$. This formula is then applied to the binomial probability distribution.

The Poisson Distribution Chapter 12.2 continues the probability theme by developing the Poisson distribution and using it to study the frequency of radioactive decay events.

The Power Spectrum Chapter 12.3 builds on the study of periodicity begun in chapter 7. We develop the Fourier transform, a basic tool in the sciences for detecting the relative strength of periodic components in a noisy data set.

Fourier Series Chapter 12.4 expands on some of the ideas in chapter 11. Here we develop tools for approximating functions over intervals using sums of sine and cosine terms. This is an extensively used method in a wide range of disciplines, from thermodynamics to music synthesis.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

12.1 Stirling's Formula

Factorials in probability

Given a positive integer n, we define $n!$—pronounced n *factorial*—by the rule $n! = 1 \cdot 2 \cdot 3 \cdots (n-1) \cdot n$. This is a convenient concept which occurs in a number of settings, particularly combinatorial and probabilistic ones. For instance, the probability of getting exactly n heads out of $2n$ tosses of a coin turns out to be

$$\frac{(2n)!}{2^{2n}(n!)^2}.$$

$n!$ is difficult to calculate

Unfortunately, evaluating $n!$ for values of n at all large is cumbersome at best. Although though many calculators will compute factorials, few of them can handle numbers as large as $1000!$. Even when we can evaluate $n!$, we are often as interested in the asymptotic behavior of a certain expression as much as in its exact value for specific n. For instance, using methods we develop below, it turns out that the above expression for the probability of n heads in $2n$ tosses is very close to $1/\sqrt{\pi n}$, with the approximation being more accurate the larger n is. In fact, for $n \geq 8$, the approximation is good to two places; for $n \geq 25$, the approximation gives three-place accuracy.

In his book *Methodus differentialis* (1730), the British mathematician James Stirling published the following approximation, now know as **Stirling's formula**, for the factorial operator:

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n}.$$

While the right-hand side may look much more complicated than the left, think which one you would rather evaluate for, say, $n = 100$. To see how good this approximation is, here are some comparisons:

n	$n!$	Stirling's approximation
2	2	1.9190
10	3,628,800	3,598,695.6
50	3.0414×10^{64}	3.0363×10^{64}
100	9.3326×10^{157}	9.3248×10^{157}
1000	4.02387×10^{2567}	4.02354×10^{2567}
10000	$2.84626 \times 10^{35659}$	$2.84624 \times 10^{35659}$

As an example of the way elementary ideas in calculus can be used to derive powerful and subtle results, we will outline a derivation of Stirling's

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

approximation for $n!$. You should write up your own summary of this proof, filling in the gaps in the text below. We will work in two stages. In the first stage, we will show that

$$n! \sim c\, n^{n+\frac{1}{2}}\, e^{-n}.$$

for some constant c. In the second stage we will show that this constant is actually $\sqrt{2\pi}$.

Stage One: Deriving the General Form

We first observe that

$$\ln(n!) = \ln 1 + \ln 2 + \ldots + \ln n.$$

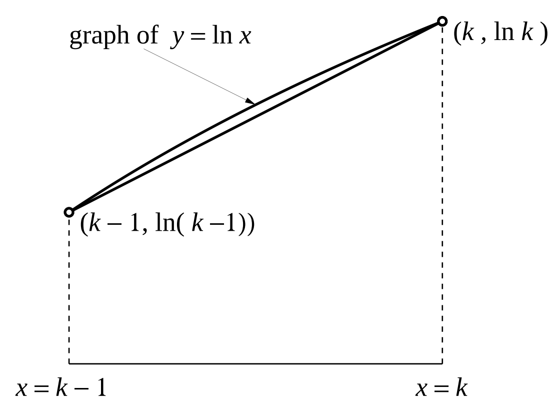
It turns out to be easier to prove things about this logarithmic form. In fact, we will deal most easily with

$$A_n = \ln 1 + \ln 2 + \ldots + \ln(n-1) + \frac{1}{2}\ln n.$$

Thus $\ln(n!) = A_n + \frac{1}{2}\ln n$. Even though $\ln 1 = 0$, it will be useful to retain the term in the expression for A_n.

We will find upper and lower bounds for A_n (and hence for $\ln(n!)$) by approximating the area under the curve $y = \ln x$ by certain inscribed and circumscribed trapezoids. We will then use these bounds to predict the asymptotic behavior of A_n for large values of n.

The upper bound: Note that if we inscribe a trapezoid under the graph of $y = \ln x$ between $x = k-1$ and $x = k$, its area will be $\frac{1}{2}(\ln(k-1) + \ln k)$. (How do we know that the straight line connecting the points $(k-1, \ln(k-1))$ and $(k, \ln k)$ will lie under the graph of $y = \ln x$?) The sum of the areas of all such trapezoids from $x = 1$ to $x = n$ is clearly less than the area under the curve $y = \ln x$ over the interval $[1, n]$.

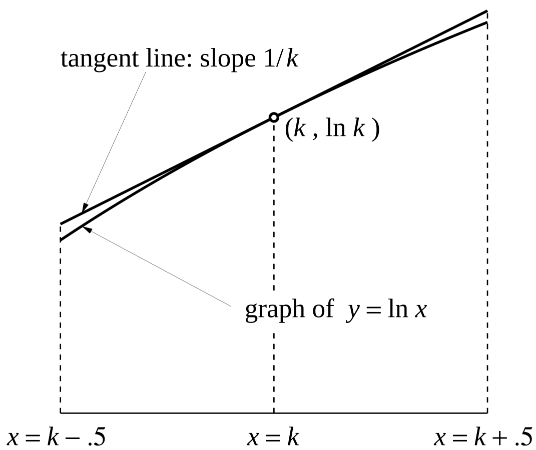


We therefore have the inequality

$$\frac{1}{2}(\ln 1 + \ln 2) + \frac{1}{2}(\ln 2 + \ln 3) + \cdots + \frac{1}{2}(\ln(n-1) + \ln n) < \int_1^n \ln x\, dx,$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

which is equivalent to $A_n < \int_1^n \ln x\,dx$.



The lower bound: On the other hand if we draw the tangent line to $y = \ln x$ at $x = k$ and form the trapezoid between $x = k - .5$ and $x = k + .5$, its area will just be $\ln k$ and will be greater than the area under the curve over the same interval. (We've used the fact—which you should check—that the area of a trapezoid equals the distance between the parallel sides times the distance between the midpoints of the other two sides.)

Adding up all such trapezoids, we get the inequality

$$\int_{\frac{3}{2}}^n \ln x\,dx < A_n.$$

Since we know that $\int \ln x\,dx = x \ln x - x$, we can evaluate these upper and lower bounds to conclude

$$n \ln n - n - \frac{3}{2} \ln \frac{3}{2} + \frac{3}{2} < A_n < n \ln n - n + 1,$$

which in turn yields

$$\left(n + \frac{1}{2}\right) \ln n - n + \frac{3}{2}\left(1 - \ln \frac{3}{2}\right) < \ln n! < \left(n + \frac{1}{2}\right) \ln n - n + 1.$$

Pause for a moment to observe that the difference

$$D_n = \left(n + \frac{1}{2}\right) \ln n - n + 1 - \ln n!$$

between the expressions on the two sides of the rightmost inequality is just the accumulated error from approximating the area under $y = \ln x$ by the inscribed trapezoids. Since the error over each interval is always positive, D_n must therefore get larger as n increases, We will need this fact shortly.

Returning to our inequalities, they can finally be rewritten as

$$\frac{3}{2}\left(1 - \ln \frac{3}{2}\right) < \ln n! - \left(n + \frac{1}{2}\right) \ln n + n < 1.$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Evaluating the constants, we thus have that for any value of n,

$$.8918 < \ln n! - \left(n + \frac{1}{2}\right) \ln n + n < 1.$$

If we exponentiate, this becomes

$$2.395 < \frac{n!}{n^{n+\frac{1}{2}} e^{-n}} < 2.719,$$

or

$$2.395\, n^{n+\frac{1}{2}} e^{-n} < n! < 2.719\, n^{n+\frac{1}{2}} e^{-n}.$$

Notice that these bounds are already quite strong and would be adequate for many estimates. Moreover, they are true for any value of n. If we are only interested in large values of n, we can do a little better. Let

These estimates are often adequate

$$\delta_n = \ln n! - \left(n + \frac{1}{2}\right) \ln n + n.$$

Then $1 - \delta_n = (n + \frac{1}{2}) \ln n - n + 1 - \ln n!$ is just the expression we called D_n a moment ago and said had to be increasing as n increases. But if $D_n = 1 - \delta_n$ is increasing, it must be true that δ_n itself is decreasing as n gets larger. We thus must have $1 > \delta_1 > \delta_2 > \ldots > \delta_n \ldots > .8918$. There must therefore be some constant $d \geq .8918$ such that $\lim_{n\to\infty} \delta_n = d$. Define the constant c by $c = e^d$. Then

$$\lim_{n\to\infty} \frac{n!}{n^{n+\frac{1}{2}} e^{-n}} = c,$$

which is what we mean when we write

$$n! \sim c\, n^{n+\frac{1}{2}}\, e^{-n}.$$

This completes stage 1. In stage 2 we will see that $c = \sqrt{2\pi}$.

Stage Two: Evaluating c

We will do this using an interesting result of a 17th century English mathematician, John Wallis, who showed that

Wallis's formula

$$\lim_{n\to\infty} \frac{2}{1} \times \frac{2}{3} \times \frac{4}{3} \times \frac{4}{5} \times \frac{6}{5} \times \frac{6}{7} \times \cdots \times \frac{2n}{2n-1} \times \frac{2n}{2n+1} = \frac{\pi}{2}.$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Suppose for the moment that we had proved Wallis's formula. We can express it in terms of factorials by noting that we can rewrite the product of the first n even numbers—$2 \times 4 \times 6 \times \ldots \times (2n)$—by factoring a 2 out of each term, leaving us

$$2 \times 4 \times 6 \times \ldots \times (2n) = 2^n\,(n!).$$

Similarly, we can take the product of the first n odd integers—$1 \times 3 \times 5 \times 7 \times \ldots \times (2n-1)$—and insert the missing even terms to get

$$\begin{aligned} 1 \times 3 \times 5 \times 7 \times \ldots \times (2n-1) &= \frac{1 \times 2 \times 3 \times 4 \times \ldots \times (2n-1) \times (2n)}{2 \times 4 \times 6 \times \ldots \times (2n)} \\ &= \frac{(2n)!}{2^n\,(n!)}. \end{aligned}$$

We can thus rewrite Wallis's formula as

$$\lim_{n\to\infty} \frac{(2^n\,n!)^4}{((2n)!)^2\,(2n+1)} = \frac{\pi}{2}.$$

If we now replace all the factorials by their corresponding expressions using Stirling's approximation, we get

$$\lim_{n\to\infty} \frac{2^{4n}c^4n^{4n+2}e^{-4n}}{c^2(2n)^{4n+1}e^{-4n}(2n+1)} = \frac{\pi}{2},$$

which, after a great deal of cancelation, reduces to

$$\lim_{n\to\infty} \frac{c^2n}{2(2n+1)} = \frac{\pi}{2}.$$

Now since

$$\lim_{n\to\infty} \frac{n}{2n+1} = \frac{1}{2},$$

this reduces to

$$\frac{c^2}{4} = \frac{\pi}{2},$$

so

$$c^2 = 2\pi,$$

and

$$c = \sqrt{2\pi},$$

as desired.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Deriving Wallis's formula

One way to derive Wallis's formula involves the integrals

$$I_k = \int_0^{\pi/2} \sin^k x\, dx.$$

Note that $I_0 > I_1 > I_2 > I_3 > \ldots$. Moreover, you should verify that

$$I_0 = \frac{\pi}{2} \qquad \text{and} \qquad I_1 = 1.$$

Using the reduction formula derived in chapter 11.5 for antiderivatives of $\sin^n x$, we have a similar reduction formula for the I_k:

$$\begin{aligned} I_k &= \int_0^{\pi/2} \sin^k x\, dx \\ &= \frac{-1}{k} \sin^{k-1} x \, \cos x \,\Big|_0^{\pi/2} + \frac{k-1}{k} \int_0^{\pi/2} \sin^{k-2} x\, dx \\ &= \frac{k-1}{k} I_{k-2}. \end{aligned}$$

This in turn leads to

$$I_k = \begin{cases} \dfrac{2n-1}{2n} \cdot \dfrac{2n-3}{2n-2} \cdots \dfrac{1}{2} \cdot \dfrac{\pi}{2} & \text{if } k = 2n \text{ is even,} \\ \dfrac{2n}{2n+1} \cdot \dfrac{2n-2}{2n-1} \cdots \dfrac{2}{3} \cdot 1 & \text{if } k = 2n+1 \text{ is odd.} \end{cases}$$

Further, note that

$$I_{2n+2}/I_{2n} = \frac{2n+1}{2n+2},$$

which has the limit 1 for large n. Since $I_{2n} > I_{2n+1} > I_{2n+2}$, it follows that I_{2n+1}/I_{2n} approaches 1 for large n. But this gives us

$$\begin{aligned} 1 &= \lim_{n\to\infty} I_{2n+1}/I_{2n} \\ &= \lim_{n\to\infty} \left(\frac{2n}{2n+1} \cdot \frac{2n-2}{2n-1} \cdots \frac{2}{3} \cdot 1 \right) \div \left(\frac{2n-1}{2n} \cdot \frac{2n-3}{2n-2} \cdots \frac{1}{2} \cdot \frac{\pi}{2} \right) \\ &= \lim_{n\to\infty} \frac{2n}{2n+1} \cdot \frac{2n}{2n-1} \cdot \frac{2n-2}{2n-1} \cdots \frac{2n-2}{2n-3} \cdots \frac{2}{3} \cdot \frac{2}{1} \cdot \frac{2}{\pi} \end{aligned}$$

If we multiply both sides of this equation by $\pi/2$, we get Wallis's formula.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Further refinements

Some refinements

Using even more careful methods of analysis, it is possible to improve on Stirling's approximation and derive approximations like

$$n! \sim \sqrt{2\pi}\, n^{n+\frac{1}{2}}\, e^{-n+\frac{1}{12n}-\frac{1}{360n^3}+\frac{1}{1260n^5}-\cdots}.$$

If we use this expression to approximate 1000!, for instance, our result is accurate for the first 24 digits.

While this approximation and Stirling's original one are good in the sense that they give more and more accurate digits the larger n gets—so that the *ratio* of $n!$ to either approximation goes to 1 as n gets large—they are bad in the sense that the *difference* between $n!$ and either approximation becomes infinite as n gets large.

The Binomial Distribution

One of the most frequently encountered concepts in probability theory is the **binomial probability distribution**. Suppose we repeat a certain experiment—flipping a penny, rolling a single die, mating a pair of fruit flies, feeding cholesterol to a lab rat—over and over. Suppose further that there is some outcome we are looking for—getting heads, rolling a 2, getting a red-eyed offspring, developing liver cancer in the rat—in each experiment. If p is the probability p of obtaining the looked-for outcome in any one experiment, denote by $P(n,k,p)$ the probability of the outcome happening exactly k times in n experiments. It turns out that

$$P(k,n,p) = \frac{n!}{k!(n-k)!}p^k(1-p)^{n-k}.$$

Example 1 How likely is it to get four 2's if we roll twelve dice? The probability of getting a 2 by throwing one die is $\dfrac{1}{6}$. Therefore the answer to the question is

$$P(12,4,\tfrac{1}{6}) = \frac{12!}{4!\,8!}\left(\frac{1}{6}\right)^4\left(\frac{5}{6}\right)^8 = .0888281$$

—we should get exactly four 2's slightly less frequently than once out of every 11 times we roll twelve dice.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Example 2 What is the probability of getting exactly 47 heads if we flip 100 pennies? Since the probability of getting heads on a single toss of a penny is $\frac{1}{2}$,

$$P(100, 47, \tfrac{1}{2}) = \frac{100!}{47!\,53!}\left(\frac{1}{2}\right)^{100} = .0665905,$$

—on the average, if we flip 100 pennies, we should get 47 heads about once out of every 15 times.

The second example demonstrates the fact that calculating binomial probabilities can get very messy very quickly. Several of the exercises are designed to show how Stirling's formula can give us quick estimates that are easy to calculate and work with.

Exercises

1. Go through the derivation in this section and find several passages that seem to you to go a bit fast or skip over details. Rewrite these sections to make them clearer and more complete.

2. Confirm the values given in the table on page 770 for the approximations of 100! and 1000! that Stirling's formula produces.

3. **Rate of growth of $n!$** Factorials get very large very rapidly. The purpose of this exercise is to develop a sense of just how rapidly $n!$ grows by comparing it to exponential functions.

Let N be some integer > 1, and consider the sequence $a_1, a_2, a_3, \ldots$ defined by

$$a_n = \frac{N^n}{n!}.$$

a) Show that $a_k = \dfrac{N}{k} a_{k-1}$, and conclude that

$$\begin{array}{lll} \text{if } k < N & \text{then} & a_{k-1} < a_k; \\ \text{if } k > N & \text{then} & a_{k-1} > a_k; \\ \text{if } k = N & \text{then} & a_{k-1} = a_k. \end{array}$$

We thus have a sequence that increases for a while:

$$a_1 < a_2 < \cdots a_{N-1} = a_N,$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

and then decreases forever after:

$$a_N > a_{N+1} > a_{N+2} > \cdots .$$

b) If $k > 2N$ show that $a_k < .5a_{k-1}$. Hence conclude that $\lim_{n\to\infty} a_n = 0$.

c) Use Stirling's approximation to show that

$$a_N \approx \frac{e^N}{\sqrt{2\pi N}}.$$

Calculate the values of this expression for $N = 10$ and $N = 100$ to get an idea of how large the sequence $\{a_n\}$ can get. This shows that *for a while*, the exponential series $\{N^n\}$ can get large much more rapidly than the series $\{n!\}$.

d) Show that $a_n < 1$ if $n > eN$. This gives an upper bound on how long it takes the factorials to catch up with the exponentials.

4. If $n \geq 5$, then $n!$ terminates in a certain number of zeroes. For instance, $5! = 120$ ends in one zero, $23! = 25852016738884976640000$ ends in four zeroes, and so on. How many zeroes are there at the end of $1000!$?

The binomial distribution

5. The formula for the binomial distribution gives us that the probability of getting exactly n heads in $2n$ flips of a coin is

$$\frac{(2n)!}{(n!)^2}\left(\frac{1}{2}\right)^{2n}.$$

Show using Stirling's formula this can be approximated by

$$\frac{1}{\sqrt{\pi n}}.$$

Use this approximation to find the probability of getting 50 heads out of 100 tosses of a coin. If you have a computer or calculator which can compute factorials, use the original binomial distribution formula to calculate the exact probability of getting 50 heads and compare the answers.

6. More generally, if we try a certain experiment n times with a probability p of success each time, the most likely number of successes is $k = np$. (Assume

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

that p is a fraction and n is such that $n \cdot p$ is an integer.) Use Stirling's approximation to show that the probability of getting exactly np successes is

$$P(n, np, p) \approx \frac{1}{\sqrt{2\pi np(1-p)}}.$$

Is this consistent with the answer to the previous exercise?

7. **One-dimensional random walk** An important class of problems, including **diffusion** and **Brownian motion** involve the long-term behavior of particles moving randomly. We will look at the simplest case of such problems. A particle starts at the origin on a line and at each stage moves one unit to the right or one unit to the left, being equally likely to do either. What can we say about where the particle will be after n steps? In this problem we will use Stirling's formula to develop some useful insights into this question.

a) Explain why the particle will be r units to the right of the origin after n steps if and only if it has moved to the right $k = (n+r)/2$ times and to the left $n - k = (n-r)/2$ times. Explain why it could never be 3 units or 7 units to the right after 100 steps.

b) Using the same symbols as in part (a), show that the probability of the particle's being exactly r units to the right after n steps is

$$\frac{n!}{k!(n-k)!}\left(\frac{1}{2}\right)^n.$$

c) Use Stirling's formula to show that this probability of being r units to the right after n steps is approximately

$$\frac{\sqrt{2}}{\sqrt{\pi n}(1+(r/n))^{\frac{1}{2}(n+r+1)}\,(1-(r/n))^{\frac{1}{2}(n-r+1)}}.$$

d) To simplify the denominator of this fraction, recall the Taylor series approximation for $\ln(1+x)$:

$$\ln(1+x) = x - \frac{x^2}{2} + \cdots.$$

Hence, if r is much smaller than n, $\ln(1 + r/n)$ can be approximated by $r/n - r^2/(2n^2)$, and $\ln(1 - r/n)$ can be approximated by $-r/n - r^2/(2n^2)$.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

By ignoring all powers of r greater than the second, conclude that

$$(1+(r/n))^{\frac{1}{2}(n+r+1)}\,(1-(r/n))^{\frac{1}{2}(n-r+1)} \approx e^{r^2/(2n)},$$

so that the probability of being r units to the right after n steps is

$$\sqrt{\frac{2}{\pi n}}\, e^{-r^2/2n}.$$

e) Explain how we can get the answer to exercise 5 as a special case of the result just obtained in part (d).

f) Using the approximation from part (d), calculate the probability that after 100 steps the particle will be *no more* than 5 units away from the starting point to either the right or the left. Remember that after 100 steps it is impossible to be an odd number of units away from the starting point. The exact probability is

$$\sum_{k=48}^{52} \frac{100!}{k!(100-k)!}\left(\frac{1}{2}\right)^{100} = .382701.$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

12.2 The Poisson Distribution

A Linear Model for α-Ray Emission

When a radioactive element decays, we know from the study of differential equations in chapter 4 that the amount $A(t)$ of radioactive material present at time t satisfies the differential equation

$$A' = -kA,$$

where $k > 0$ is the decay constant. If A_0 is the amount present at time $t = 0$, then the solution is

$$A(t) = A_0 e^{-kt}.$$

The time T it takes for a given amount of radioactive material to decay to half the starting quantity is known as the **half life** of the element. Since, by definition, $A(T) = .5\,A(0) = .5\,A_0$, we must have

$$e^{-kT} = \frac{1}{2},$$

which leads to

$$kT = \ln 2$$

and therefore

$$T = \frac{\ln 2}{k}.$$

The relation between the half-life and the decay constant

Suppose, for example, that we have a sample of polonium, which is a radioactive isotope of radium. The decay constant of polonium is $k = .500865$ % per day, and thus its half life is

$$T = \frac{\ln 2}{k} = \frac{\ln 2}{.00500865} = 138.39 \text{ days}.$$

By local linearity, $A(t)$ is closely approximated by a linear function for short intervals of time. Because polonium has a half-life of 138.39 days, a “short time” means several hours in this case. Thus, if we spend an afternoon in a laboratory studying the decay of polonium, we can assume that $A(t)$ is linear.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

When polonium decays, it produces various sorts of radiation, including α-rays ("alpha rays"). Using a scintillation counter, one can determine the number of rays emitted in given directions:

A setup like this will count a fixed percentage of the total number of α-rays emitted. Since our model of decay is linear, it follows that the number of α-rays detected should be a linear function of time. If we start counting at time $t = 0$, the number of particles observed will have a straight-line graph:

In the early 20th century, researchers like Marie Curie and Ernest Rutherford did numerous studies of the α-rays emitted by polonium. For example, in 1911, Rutherford, Geiger and Bateman counted the number of α-rays detected in a 7.5-second time period. They repeated their experiment 2608 times and detected a total of 10,097 α-rays. This is an average of

$$\frac{10097}{2608} = 3.8715 \ \alpha\text{-rays per 7.5-second period},$$

so the number of α-rays per second is

$$\frac{3.8715}{7.5} = .5162 \ \alpha\text{-rays per second}.$$

Thus the straight line in the above graph has slope .5162.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

This model of α-ray production has several problems. First, it predicts the existence of fractional α-rays, which makes no sense—the number detected is always a nonnegative integer. To remedy this, we can modify our model as follows:

Notice that the graph is now a step function. It shows that we should see a new α-ray every $1/.5162 = 1.937$ seconds. This model also has the following consequence: if we observe the number of α-rays produced in a 7.5-second interval, then we will always see 3 or 4 particles:

As the picture indicates, whether we get 3 or 4 depends on where the interval starts. Now comes the serious problem: this prediction is *inconsistent* with the experimental data collected by Rutherford and the others in 1911. For example, in 57 of the 2608 times they ran the experiment, *no* α-rays were observed, while in 139 cases, 7 α-rays were observed. Here are the complete data of the experiment:

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

number of α-rays observed n	number of occurrences N_n
0	57
1	203
2	383
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	0
13	1
14	1
Total	2608

It follows that the linear model of α-ray emission doesn't apply to time intervals of length 7.5 seconds. This is a common occurrence—a model may work nicely over a certain range, but outside of that range, its answers may be meaningless. The problem in our case comes from the random nature of radioactive decay. In fact, there are two sources of randomness to deal with: the *time* when a polonium atom decays is random, and the *direction* in which it then emits an α-ray is also random (this affects us since the scintillation counter only detects emissions in certain directions). We need to modify our model to take the randomness into account, and this is where probability enters in.

Probability Models

Randomness has structure

The basic idea of probability theory is that the outcome of a certain event can be unpredictable in the individual instance but predictable on the average. Throwing dice and tossing a coin are familiar examples. In this section, we will show how the Poisson probability distribution gives an excellent model of the α-ray experiment described above.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

The definition of probability

We will let p_n denote the probability of observing exactly n α-rays in a 7.5-second time interval. By this statement, we mean the following. Suppose we run the experiment N times, where N is large. Let N_n be the number of times we observed n α-rays. Then the ratio N_n/N is the frequency with which this outcome occurs. Now imagine N getting larger and larger. Being "predictable on the average" means that the ratios N_n/N approach a fixed number, that is, the limit $\lim_{N\to\infty} N_n/N$ exists. We then *define* this number to be the probability p_n. Thus

$$p_n = \lim_{N\to\infty} \frac{N_n}{N}.$$

For example, the data presented on page 784 were obtained from $N = 2608$ repetitions of our experiment. From the table given there, we see that 0 α-rays were observed 57 times. This means $N_0 = 57$, and thus the probability of detecting 0 α-rays is

$$p_0 \approx \frac{N_0}{N} = \frac{57}{2608} = .0218.$$

Similarly, we can approximate p_1, p_2, etc., using the data in the table. Our goal is to describe these probabilities p_0, p_1, ... Ideally, we would like to have a way of determining the numbers $p_0, p_1, p_2, \ldots$ "before the fact."

Some properties of probabilities

In any introductory course on probability, one learns certain basic principles for working with probabilities. We will give examples to illustrate some of these principles, and more examples may be found in the exercises.

For our purposes, we will be working in the following setting. There is a certain **experiment** being performed. This might consist of flipping a coin and noting which side comes up, or running a survey asking people at random their opinions about a certain TV show, or, in our case, counting the number of α-rays detected in a 7.5-second interval. Moreover, there is a **discrete** set of possible outcomes of the experiment. That is, the possible outcomes can be listed in a sequence O_1, O_2, O_3, ... In some cases, like throwing a pair of dice, this list might be finite. In other cases, like our α-ray experiment, the list might be infinite. What is ruled out are experiments like

The general context

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

choosing a person at random and measuring the person's height—there is a continuum of possible outcomes here which cannot be listed in the way we've specified. Moreover, there should be a **probability** assigned to each outcome, with the outcome O_n having probability p_n.. Finally, the possible outcomes should be **disjoint**—two different outcomes can't both result from a single experiment. Thus if we are examining the attributes of a group of people, "being male" and "having green eyes" would not be acceptable outcomes in our sense unless we somehow knew in advance that there were no green-eyed males in the group.

Knowing the probabilities $p_0, p_1, \ldots$ of the possible outcomes allows us to compute other, possibly more complicated probabilities. This brings in the concept of an **event**, which is basic to probability. In the case of our α-ray experiment, here are some examples of events:

- Detecting 3 α-rays.
- Detecting 2 or 4 α-rays.
- Detecting an odd number of α-rays.

In general, an **event** is a subcollection of the possible outcomes.

The addition rule for probabilities

Rule 1 The probability of an event is simply the sum of the probabilities of its component outcomes.

Thus, for the events just described, we have:

- The probability of detecting 3 α-rays is p_3;
- The probability of detecting 2 or 4 α-rays is $p_2 + p_4$;
- The probability of detecting an odd number of α-rays is the infinite sum

  $$p_1 + p_3 + p_5 + p_7 + \cdots$$

  (since an odd number of α-rays means that 1 or 3 or 5 or 7 etc. have been detected).

Another important property of probabilities follows directly from Rule 1:

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Rule 2 The sum of the probabilities of all possible outcomes is 1:

$$\sum_{k=0}^{\infty} p_k = 1.$$

The reason for this is that the list of outcomes was stipulated to be the list of *all possible* outcomes. Hence the event consisting of all these outcomes is bound to occur every time—its probability is 1.

A third rule we will need relates the probabilities of **independent** events. Two events are independent if the occurrence or non- occurrence of one of the events has no impact on the probability of the second event occurring. For instance, suppose we are examining a group of people. Consider the following events which may or may not occur each time we look at a person:

1. The person is female;
2. The person has green eyes;
3. The person is over 5'7" tall.

We would expect the first and second events to be independent, and also the second and third, but not the first and third.

The product rule for probabilities

Rule 3 The probability that two or more independent events all occur is the product of their separate probabilities.

Thus, for example, suppose that in our hypothetical group of people $\frac{1}{2}$ are female, $\frac{1}{8}$ are green-eyed, and $\frac{1}{3}$ are taller than 5'7". We might then expect roughly $\frac{1}{24}$ of them to be green-eyed *and* over 5'7", but we would have no particular reason to expect that $\frac{1}{6}$ of them are females taller than 5'7".

A final rule that is often useful is

The probability that something doesn't happen

Rule 4 If a certain event has a probability p of happening, then the probability that the event doesn't take place is $1 - p$.

For example, in our group of people, we would expect $\frac{2}{3}$ of them to be less than 5'7" tall, $\frac{7}{8}$ of them to have eyes colored something other than green, etc.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

The notion of a probability model

A **model** is a mathematical picture of a real-life phenomenon. We have seen that dynamical systems can be used to create models of physical situations. Another type of mathematical model is a *probability model.* In general, a **probability model** for an experiment with a finite number of outcomes is *a listing of all possible outcomes and an assignment of probabilities to each outcome so that their sum is 1.* In order that the probability model be a good picture of reality, we ask that the *probability assigned to an outcome should be the relative frequency with which that outcome would appear if the experiment were duplicated independently a large number of times.*

As an example, a probability model for one toss of a fair die consists of a list of all possible outcomes, namely 1, 2, 3, 4, 5, 6, and an assignment of a probability to each, namely $\frac{1}{6}$, $\frac{1}{6}$, $\frac{1}{6}$, $\frac{1}{6}$, $\frac{1}{6}$, $\frac{1}{6}$, respectively. We assign the number $\frac{1}{6}$ to each outcome because we expect that if the the experiment were repeated (that is, if the die were tossed) a large number of times, then any particular outcome (3, say) would occur about one sixth of the time. Another probability model for the experiment consisting of a toss of a die might be a list of all outcomes, again 1, 2, 3, 4, 5, 6, together with an assignment of the numbers $\frac{1}{2}$, 0, $\frac{1}{6}$, 0, 0, $\frac{1}{3}$ to 1, 2, 3, 4, 5, 6, respectively. This is a probability model, because the numbers we have assigned add to 1, but it certainly does not model very well the throw of a fair die.

We would like to set up a probability model for our experiment with α-rays. The outcomes are 0, 1, 2, 3, 4, ... where, for example, the number 5 labels the outcome in which we observe 5 α-rays in our 7.5-second interval. The total number of outcomes is equal to the number of α-rays that we could conceivably see in a 7.5-second interval. Since it is conceivable (but extremely unlikely) that every atom in the sample could decay and emit an α-ray in the direction of the scintillation counter in one 7.5-second interval, we could conceivably see as many α-rays as there are atoms in the sample. This number is so large that we can think of it as infinite. To have a probability model, we need to assign numbers p_0, p_1, p_2, $\ldots$ to the outcomes 0, 1, 2, $\ldots$, respectively, so that $p_0+p_1+p_2+\cdots = 1$. For the model to be reasonable, we would like each p_n to be approximately equal to the corresponding number N_n/N observed by Rutherford, Geiger, and Bateman.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

The Poisson Probability Distribution

The Poisson model of α-ray emission

To describe the probabilities p_0, p_1, ..., p_n, ... that we will observe 0, 1, ..., n, ... α-rays in a 7.5-second interval for our α-ray experiment, we use the **Poisson probability distribution**

$$p_n = \frac{\lambda^n e^{-\lambda}}{n!},$$

where λ is a number yet to be determined, and $n!$ is the familiar ***n*-factorial** function,

$$n! = \begin{cases} n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 3 \cdot 2 \cdot 1 & \text{if } n > 0, \\ 1 & \text{if } n = 0. \end{cases}$$

Thus the first few Poisson probabilities are:

$$p_0 = e^{-\lambda}, \qquad p_1 = \lambda e^{-\lambda}, \qquad p_2 = \frac{\lambda^2 e^{-\lambda}}{2}, \qquad p_3 = \frac{\lambda^3 e^{-\lambda}}{6}.$$

Note that this assignment does indeed give us a probability model, because

$$\begin{aligned} p_0 + p_1 + p_2 + p_3 + \cdots &= \frac{\lambda^0}{0!}e^{-\lambda} + \frac{\lambda}{1!}e^{-\lambda} + \frac{\lambda^2}{2!}e^{-\lambda} + \frac{\lambda^3}{3!}e^{-\lambda} + \cdots \\ &= e^{-\lambda}\left(1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots\right) \\ &= e^{-\lambda} \cdot e^{\lambda} \\ &= 1. \end{aligned}$$

(The transition from the second line to the third uses the fact that the expression in parentheses is just the Taylor series for e^{λ}.)

We will shortly derive the Poisson distribution from basic principles. For the moment, though, we will assume that the probabilities p_0, p_1, ... for α-ray emission are given by the above formulas, where we still need to choose an appropriate value for the parameter λ. The key to determining λ is the notion of **expectation**, which for us will mean the average number of α-rays observed in a 7.5-second interval.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Suppose we repeat our experiment N times. As usual, we let N_n denote the number of times exactly n α-rays were observed. Then the total number of α-rays observed in the N experiments is

$$0 \cdot N_0 + 1 \cdot N_1 + 2 \cdot N_2 + 3 \cdot N_3 + \cdots .$$

Then the "average number of α-rays observed in a 7.5-second interval" means the limit

$$E = \lim_{N\to\infty} \frac{0 \cdot N_0 + 1 \cdot N_1 + 2 \cdot N_2 + 3 \cdot N_3 + \cdots}{N}.$$

This limit is called the **expected value** or **expectation** (which explains why it is denoted E).

We claim that for the Poisson distribution, the expected value E is exactly the number λ. To see this, notice that the above limit can be written in the form

$$E = \lim_{N\to\infty} \left(0 \cdot \frac{N_0}{N} + 1 \cdot \frac{N_1}{N} + 2 \cdot \frac{N_2}{N} + 3 \cdot \frac{N_3}{N} + \cdots \right).$$

Since we defined

$$p_n = \lim_{N\to\infty} \frac{N_n}{N} ,$$

it follows that we get the following formula for the expectation:

The general formula for the expected value in a probability model

$$E = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + \cdots = \sum_{n=0}^{\infty} n p_n.$$

(Note that this equality is true for *any* probability model, not just the one we are considering)

Substituting in the values of p_n given by the Poisson distribution, we have

$$\begin{aligned} E &= 0 \cdot e^{-\lambda} + 1 \cdot \lambda e^{-\lambda} + 2 \cdot \frac{\lambda^2}{2!} e^{-\lambda} + 3 \cdot \frac{\lambda^3}{3!} e^{-\lambda} + \cdots \\ &= \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!}, \end{aligned}$$

where we pulled the common factor $\lambda e^{-\lambda}$ outside the summation, noted that the term in the summation corresponding to $n = 0$ is 0, and observed that

$$\frac{n}{n!} = \frac{n}{n(n-1) \cdot \cdots \cdot 2 \cdot 1} = \frac{1}{(n-1)!}.$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Letting $k = n - 1$, we have (again recognizing the Taylor series for e^{λ})

$$E = \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda} = \lambda.$$

This proves that the expected value is λ as claimed.

Now that we know how to interpret λ, it is easy to determine what it should be for the α-ray experiment. The data given on page 784 covered $N = 2608$ repetitions of the experiment, with

$$0 \cdot N_0 + 1 \cdot N_1 + \cdots = 10097$$

in this case. Thus

$$\frac{0 \cdot N_0 + 1 \cdot N_1 + \cdots}{N} = \frac{10097}{2608} = 3.8715$$

is an approximation of the expected value λ. However, since this is the only information about λ we have, we will let $\lambda = 3.8715$. Using this value of λ, we can then compare the frequencies predicted by the Poisson distribution to the actual data from on page 784:

number of α-rays observed n	number of occurrences N_n	probability approximation N_n/N	Poisson probability p_n	Poisson prediction $2608\,p_n$
0	57	.021855	.020827	54.3
1	203	.077837	.080632	210.3
2	383	.146855	.156083	407.1
3	525	.201303	.201426	525.3
4	532	.203398	.194955	508.4
5	408	.156441	.150953	393.7
6	273	.104677	.097402	254.0
7	139	.053297	.053870	140.5
8	45	.017254	.026070	68.0
9	27	.010352	.011214	29.2
10	10	.003834	.004341	11.3
11	4	.001533	.001528	4.0
12	0	.000000	.000492	1.3
13	1	.000383	.000146	.4
14	1	.000383	.000040	.1
Totals	2608	1	1	2608

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

The Poisson model agrees nicely with the data since for each n, N_n/N and p_n are reasonably close. Notice that we shouldn't expect perfect agreement since N_n/N is only an approximation to p_n. We would expect these approximations to get better as we take larger values of N.

Look at the last column, labelled "Poisson prediction." The numbers here are the Poisson probabilities multiplied by $N = 2608$, and they represent the "ideal" number of occurrences. This makes it easier to compare the model to the data. For example, the graph below plots the number of occurrences, both actual and predicted. The circles are the experimental data, while the line-segment graph connects the Poisson predictions.

Although the model seems to fit the data nicely, we should point out that there are statistical tests which can be used measure the fit more precisely. These tests are part of the material covered in courses in probability and statistics.

A final and very important point to make concerns the number of α-rays observed over a long period of time. Our particular Poisson model with $\lambda = 3.8715$ only works for a 7.5-second interval. What happens if we count α-rays over a longer time period? For simplicity, assume that we have a time interval of length T which is a multiple of 7.5 seconds, so that $T = 7.5\,N$ for some large integer N. We can regard this as running our 7.5-second experiment N consecutive times. Thus the ratio

$$\frac{\text{total number of }\alpha\text{-rays observed}}{N}$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

is an approximation to the expected value $\lambda = 3.8715$. It follows that

$$\text{total number of } \alpha\text{-rays observed} \approx 3.8715\,N = \frac{3.8715}{7.5} 7.5\,N = .5162\,T.$$

This shows that for large time intervals, we recover the linear model of α-ray emissions discussed on page 782. Thus our probabilistic model is consistent with what we did earlier and yet allows us to describe what happens when the linear model breaks down.

Derivation of the Poisson model

In the previous discussion we simply assumed that the α-ray probabilities were given by the Poisson distribution, and found that the Poisson probabilities agreed with the experimental data. Let's see where the Poisson formulas come from. It turns out that we can derive the Poisson probabilities p_n from the following assumptions:

- We have an extremely large number M of polonium atoms;
- Each atom has a small but equal probability of emitting an α-ray that is detected by our scintillation counter in a 7.5-second period;
- Observing an α-ray from a given atom is independent of observing an α-ray from any other atom.

Now suppose that we see an average of $\lambda = 3.8715$ α-rays in a 7.5-second period. Because the number of atoms M is large (in the Rutherford-Geiger-Bateman experiment $M > 10^{18}$), then the probability that a single fixed atom emits an α-ray detected by our scintillation counter in a given time period is very close to λ/M. The probability that the single atom does not emit a detected α-ray in the period is then $1 - \lambda/M$ (by Rule 4, page 787). Thus, the probability p_0 that none of the M atoms emits an α-ray in the 7.5-second period is $(1 - \lambda/M)^M$ (by Rule 3, page 787).

The fact that M is so large allows us to make a simplifying approximation. Recall that for any value of x, positive or negative,

$$e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n.$$

Therefore

$$p_0 = \left(1 - \frac{\lambda}{M}\right)^M \approx e^{-\lambda}.$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

You might calculate some sample values for various values of M to see how good this approximation is.

To derive the values of p_k for $k > 1$, we need a slight improvement on the estimate we just made. Let k be a relatively small (compared to the size of M) number. Then

$$\left(1 - \frac{\lambda}{M}\right)^{M-k} = \left(\left(1 - \frac{\lambda}{M}\right)^{M}\right)^{\frac{M-k}{M}}.$$

$(M-k)/M$ is essentially equal to 1

Now if M is very large compared to k—we will be thinking of values of M on the order of magnitude of 10^{18} and $k < 100$—then $(M-k)/M$ will essentially equal 1, Hence

$$\left(1 - \frac{\lambda}{M}\right)^{M-k} \approx \left(e^{-\lambda}\right)^{1} = e^{-\lambda}.$$

We can now work out p_1. Fix your attention first on a particular atom. The probability that *that* atom does emit an α-ray detected by the scintillation counter while the other $M - 1$ atoms do not is (again by Rule 3)

$$\left(\frac{\lambda}{M}\right)^{1}\left(1 - \frac{\lambda}{M}\right)^{M-1} \approx \frac{\lambda}{M}\, e^{-\lambda},$$

by the preceding approximation.

Since there are altogether M atoms which might have been responsible for the single α-ray emission, the probability that some *unspecified* atom emits an α-ray while the others do not is (Rule 1, page 786) the sum of the probability we just calculated for each of the M atoms, which is equal to M times that probability. The total probability is p_1:

$$p_1 \approx M\,\frac{\lambda}{M}\, e^{-\lambda} = \lambda\, e^{-\lambda}.$$

To work out p_2, note that the probability that each atom of some fixed pair of atoms emits an α-ray detected by the counter, and no other atoms does, is

$$\left(\frac{\lambda}{M}\right)\left(\frac{\lambda}{M}\right)\left(1 - \frac{\lambda}{M}\right)^{M-2} \approx \frac{\lambda^2}{M^2}\, e^{-\lambda},$$

using our usual approximation. Since there are $\frac{1}{2}M(M-1)$ different pairs of atoms (we can choose the first M different ways and the second $(M-1)$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

ways, but each pair gets counted twice in this scheme, so we have to divide by 2), we obtain

$$p_2 \approx \frac{M(M-1)}{2}\frac{\lambda^2}{M^2}e^{-\lambda} = \left(1 - \frac{1}{M}\right)\frac{\lambda^2}{2}e^{-\lambda} \approx \frac{\lambda^2}{2}e^{-\lambda}.$$

As one can easily imagine, the computations for p_3, p_4, ... are similar. The observant reader will note that the exact values we got for p_0, p_1 and p_2 are not the values given by the Poisson distribution. We got the Poisson probabilities only by making various approximations that were justified by the large value of M. The assumptions we have made actually lead to what is called the **binomial distribution** (see chapter 12.1), a distribution which tends to the Poisson distribution in the limit $M \to \infty$. In this case, where M is large and λ relatively small, the binomial distribution is extremely close to the Poisson distribution.

Other applications of the Poisson distribution

The Poisson distribution can be used to model many other situations that have a random element. Examples include:

- The number of chromosome interchanges caused by exposure to X-rays for a fixed interval of time.
- The number of bacteria in a given unit of area on a Petri dish.
- The number of misprints on a page in a book.
- The number of flying-bomb hits per unit area in London during World War II.

In the exercises we will explore some examples.

Exercises

Probability models

1. A fair coin is tossed. If it comes up H (heads), a fair die is rolled. If the coin comes up T, the coin is tossed again. Construct a probability model for this experiment, listing the possible outcomes and their probabilities. (Hint: the list of outcomes is $H, 1$, $H, 2$, ..., $H, 6$, TT, TH.)

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

2. Two identical fair coins are put in cup, shaken, and spilled out onto a table. Construct a probability model for this experiment.

3. a) In the disintegration of large numbers of particles of radium (Ra), it is noted that 29% of the disintegrations result in

$$Ra \longrightarrow P + A$$

and the remainder in

$$Ra \longrightarrow He^{+} + B.$$

What is a model for the disintegration of a single particle of Ra?

b) Construct a probability model for the disintegration of two particles of Ra.

The Poisson distribution

4. The purpose of this exercise is to present another way to show that the expected value E of the Poisson distribution is equal to λ. As in the text we have

$$E = 0 \cdot p_0 + 1 \cdot p_1 + 2 \cdot p_2 + 3 \cdot p_3 + \cdots = \sum_{n=0}^{\infty} n p_n.$$

The numbers np_n can be simplified as follows:

$$\begin{aligned} 0 \cdot p_0 &= 0, \\ 1 \cdot p_1 &= 1 \cdot \lambda e^{-\lambda} = \lambda \cdot e^{-\lambda} = \lambda p_0, \\ 2 \cdot p_2 &= 2 \cdot \frac{\lambda^2 e^{-\lambda}}{2} = \lambda \cdot \lambda e^{-\lambda} = \lambda p_1, \\ 3 \cdot p_3 &= 3 \cdot \frac{\lambda^3 e^{-\lambda}}{6} = \lambda \cdot \frac{\lambda^2 e^{-\lambda}}{2} = \lambda p_2. \end{aligned}$$

a) This pattern generalizes: show that

$$np_n = \lambda p_{n-1} \quad \text{for all } n > 0\ .$$

b) Use part (a) to compute the expectation E (you will need to use the fact that the sum of the probabilities is $p_0 + p_1 + p_2 + \cdots = 1$).

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

5. A model is to be constructed for the number of rain drops that fall per square foot over a short time interval. Under what conditions would a Poisson distribution be appropriate. Under what conditions would a linear model be better?

6. In analyzing flying-bomb hits in the south of London during World War II, investigators partitioned the are into 576 small sectors, each being $\frac{1}{4}$ of a square kilometer. There were 229 sectors with no hits, 211 sectors with exactly 1 hit, 93 sectors with exactly 2 hits, 35 sectors with 3 hits, 7 sectors with 4 hits, and one sector with 5 or more hits. What might lead you to expect that a Poisson distribution might be a good model for the number of hits on each sector? Fit a Poisson distribution to the data by taking λ to be the average number of hits per sector. Use this λ to compute the theoretical frequencies of 0, 1, 2, 3, 4 and 5 hits in 576 sectors.

7. A meteorite shower sprinkles a large area of the earth's surface with small meteorite hits. The average density is 5×10^{-6} hits per square meter. Set up a model assigning a probability to the number of hits per square kilometer.

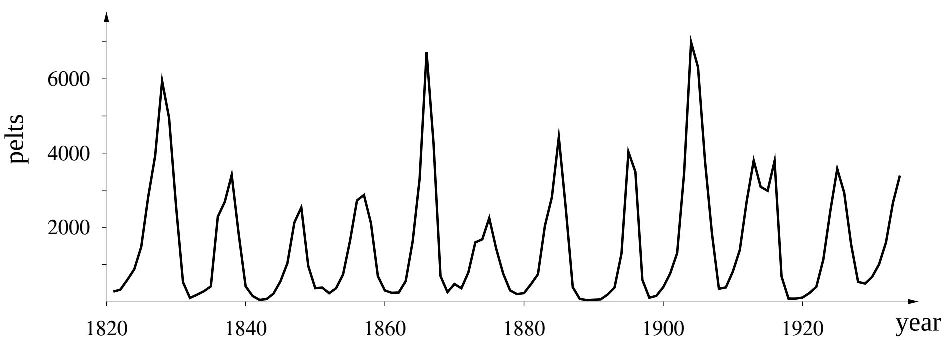
8. The central processing unit (CPU) of a laptop computer will freeze if more than ten instructions are received in a millisecond. If the average number of instructions per second received in the course of executing a large program is one per millisecond, what is the probability that the instructions received by the CPU will cause it to freeze (and, hence, the program to crash).

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

12.3 The Power Spectrum

The problem of signal + noise

This section is an application of ideas about periodic functions and integrals to the problem of separating a signal from noise. We face this problem in our daily life. Radio and television signals have noise added to them from other radio sources we can't control. The noise sounds like hissing static on a radio and looks like "snow" on a television screen. A good receiver is designed to filter out the noise while allowing the the transmitted signal to come through undistorted.



Annual harvest of lynx pelts

Scientific data and a radio broadcast have something in common: both are combinations of signal and noise. For instance, consider the annual harvest of lynx pelts by the Hudson's Bay Company. It is conceivable that the lynx population itself (the *signal*) was periodic, but various random fluctuations (the *noise*) caused the harvest (which is *signal* + *noise*) to take the form it did. If this is the case, then we should try to "filter out" the noise and find the underlying periodic signal. There is a mathematical tool to do this; it is called the **power spectrum**. We will discuss the ideas behind the power spectrum and show how it can be used to detect the underlying in noisy data.

The power spectrum filters noise to detect periodic signals

Signal + Noise

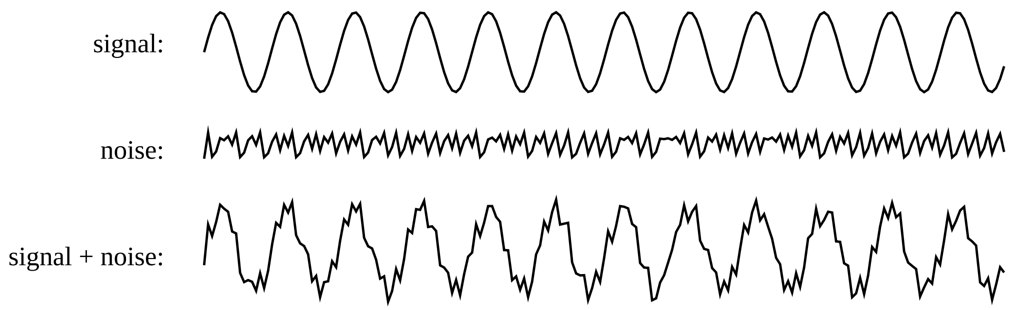
To prepare for working with the power spectrum, let's first see what happens to a periodic signal that has some noise added to it. The signal we will use is a pure sine wave. The **information** that the signal carries is the frequency of that wave. The noise will also be a function, but one whose values vary in a random fashion. It can be thought of as a combination of periodic signals of all frequencies. For this reason it is sometimes called "white noise," because

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

white light is a combination of light rays of all colors (i.e., frequencies). Here is the question we will explore: If we increase the strength of the noise, when do we lose the information contained in the original signal?

A signal with faint noise

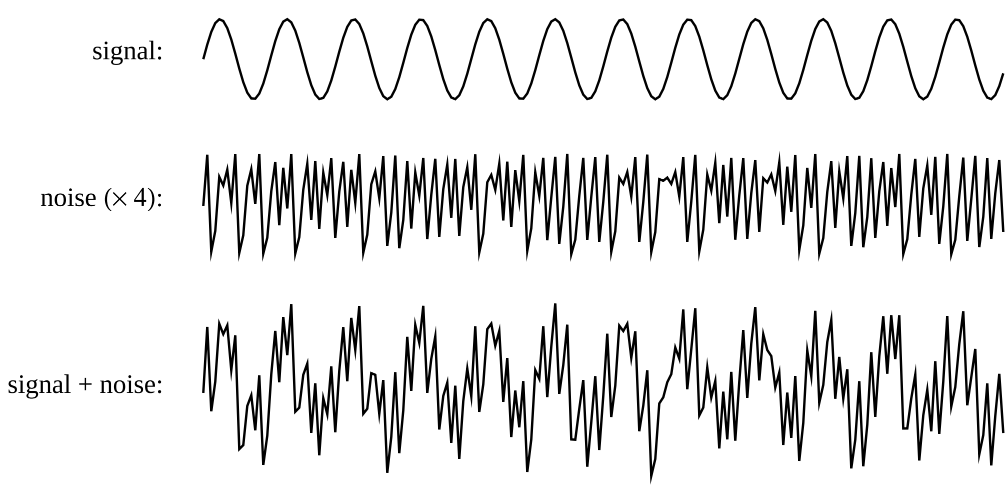
The signal and noise are shown below. As you can see, the amplitude of the signal is about 4 times as large as the amplitude of the noise. We say



that the **signal-to-noise ratio** is 4:1. The combined signal + noise is no longer a pure sine wave, of course. However, it is still recognizable as a "noisy" wave with the same frequency as the original signal. The information from the signal has not yet been lost.

The noise level becomes stronger

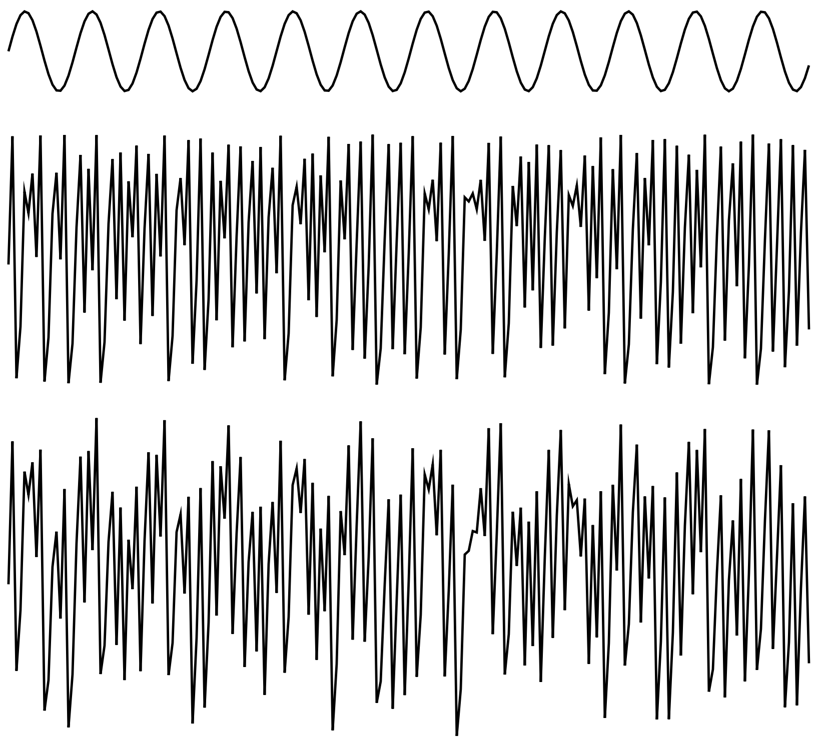
Look what happens when we increase the amplitude of the noise. In the figure below, the noise has been increased by a factor of 4, so the signal-to-noise ratio is now 1:1. The combined signal + noise is now very noisy. Would you be willing to argue that it is a wave of the same frequency as the original signal? Or would you prefer to say that it has no periodic pattern whatsoever? It appears we are close to losing the information from the original signal.



Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

The noise level becomes overwhelming

If we increase the original noise level by a factor of 10, we appear to lose the original signal altogether. The signal-to-noise ratio is now 1:2.5, and the



signal + noise appears to be as random as the noise itself. In spite of appearances, the signal is still there, and it will be detected in the power spectrum!

Detecting the Frequency of a Signal

Compare the signal to a probe whose frequency can be varied

Assume we have a signal that may be distorted by a lot of noise. We want to decide whether the signal has a periodic component; if it does, we want to determine its frequency. Our detector is based on this simple idea: *Compare the signal to a test probe of known frequency; vary the frequency of the probe until there is a positive response.* Of course, we still need to explain how the comparison is made, and what constitutes a positive response.

The test probe

Although the detector will work on a very noisy signal, like the one above, we will understand it better if we first use it to analyze a signal whose periodic nature is evident. Let the signal $S(t)$ be a pure sine wave lying above the t-axis, and suppose that t is the time measured in seconds. Our **test probe**

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

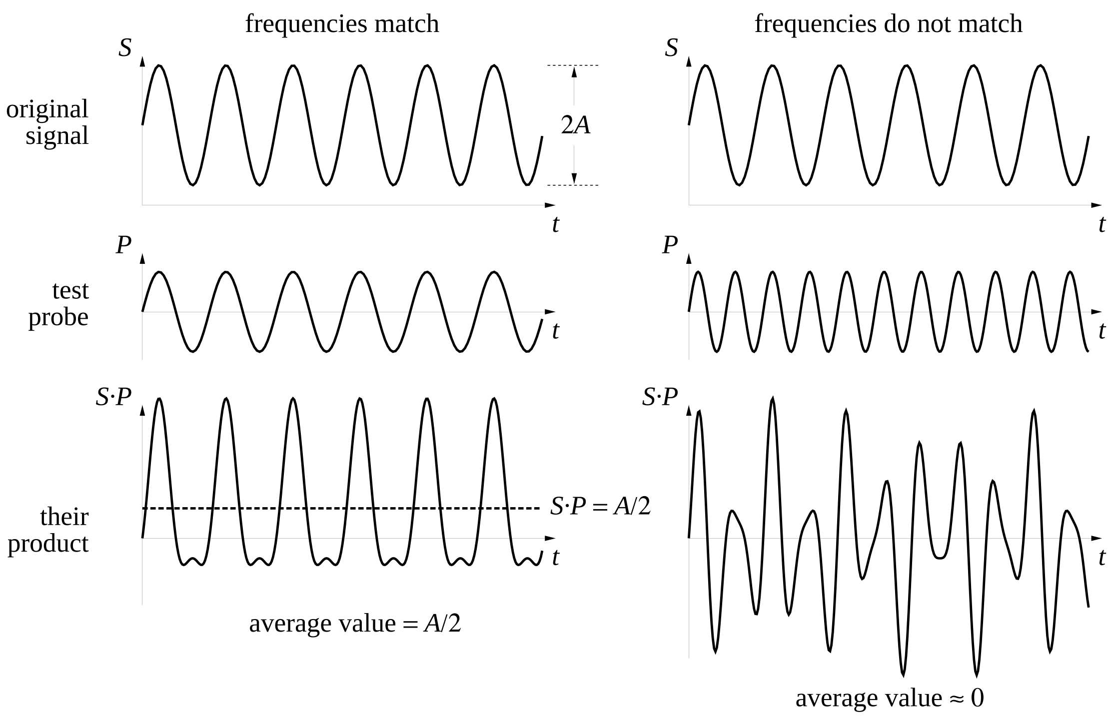
is the function

$$P(t) = \sin(2\pi\omega t)$$

whose frequency is ω cycles per second. As its graph demonstrates, the values of P are equally likely to be positive or negative.

When the signal matches the test frequency

Is the same true for the product $P(t)S(t)$? Suppose first that $S(t)$ has the same frequency as $P(t)$ (below, left). As you can see, the positive values of $P(t)$ are always multiplied by the larger values of $S(t)$. By contrast, the negative values of $P(t)$ are always multiplied by the smaller values of $S(t)$. Consequently, the positive values of $P(t)S(t)$ outweigh the negative ones. On average, the value of the product is positive. In fact, the average value of the product is half the amplitude of the original signal. Later on we will see why this is so.



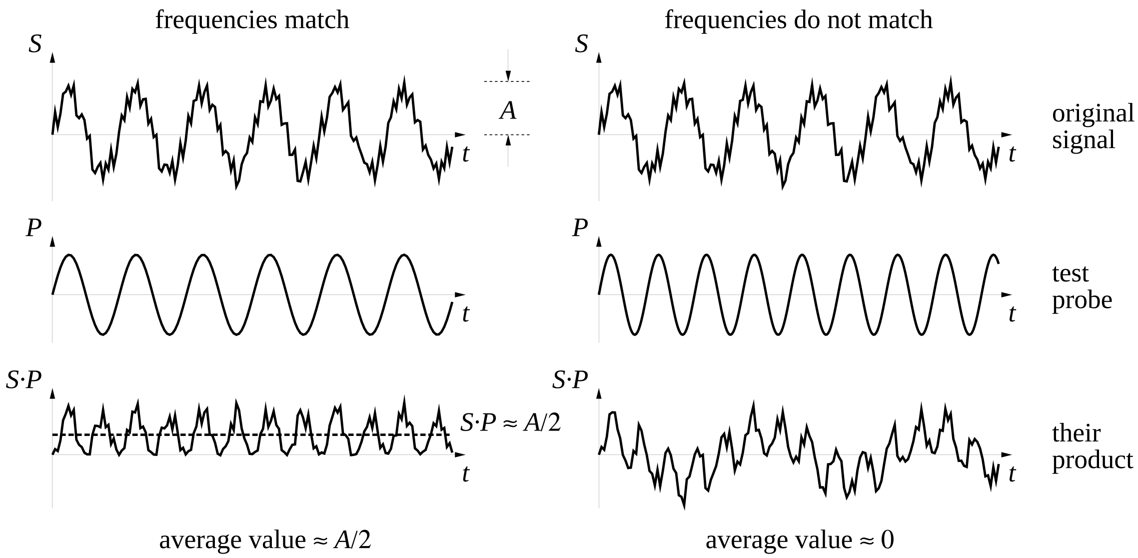
When the signal *doesn't* match the test frequency

On the right we see what happens if $S(t)$ is *not* related to $P(t)$. In that case, a large value of $S(t)$ is just as likely to multiply a positive value of $P(t)$ as a negative one. Consequently, the product $P(t)S(t)$ will have both large positive and large negative values. On average, the value of the product will be about 0.

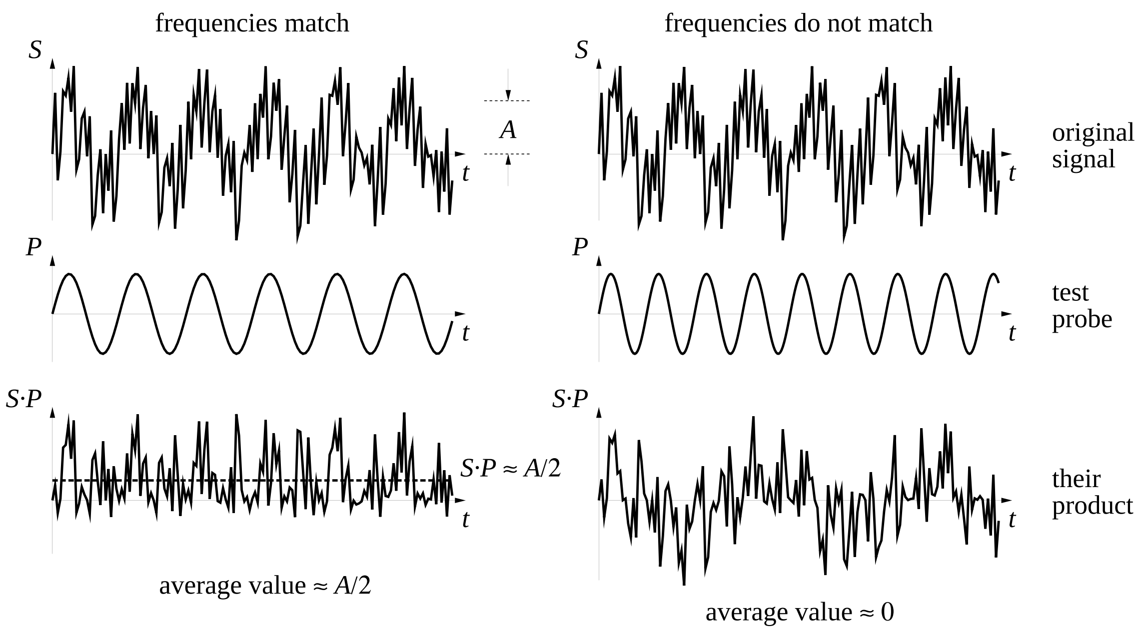
Let's use the detector on the signals we constructed on page 799. In both

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

we started with a pure sine wave and added some white noise. In the first, the signal-to-noise ratio was 4:1.



In the second the noise was stronger; the signal-to-noise ratio was 1:1.



To use the detector yourself, you have to be able to calculate the average

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

value of a function. This is discussed in chapter 6.3. The average value of $y = f(x)$ on the interval $a \le x \le b$ is

The average value of a function

$$\frac{1}{b-a}\int_a^b f(x)\,dx.$$

Our detector is the average value of the product of the signal $S(t)$ and the test probe $P(t) = \sin(2\pi\omega t)$.

Frequency detector: $D(\omega) = \dfrac{1}{b-a}\displaystyle\int_a^b S(t)\sin(2\pi\omega t)\,dt.$

Integrals with parameters define functions

Clearly, the value of the detector depends on the frequency ω of the probe P. We have tried to reflect this in the notation: the detector is a function D whose input is the frequency ω. The output of the function is calculated as an integral in which the input ω plays the role of a parameter.

This is the first time we have defined a function as an integral with a parameter. Let's see how the detector works to analyze the signal $S(t) = 3\sin(5t)$ over the interval $0 \le t \le 10$. We have

$$D(\omega) = \frac{1}{10}\int_0^{10} 3\sin(5t)\sin(2\pi\omega t)\,dt.$$

In the exercises at the end of chapter 11.3 we obtained an explicit formula for the integral of the product of two sine functions. Find that formula and check that it yields the following:

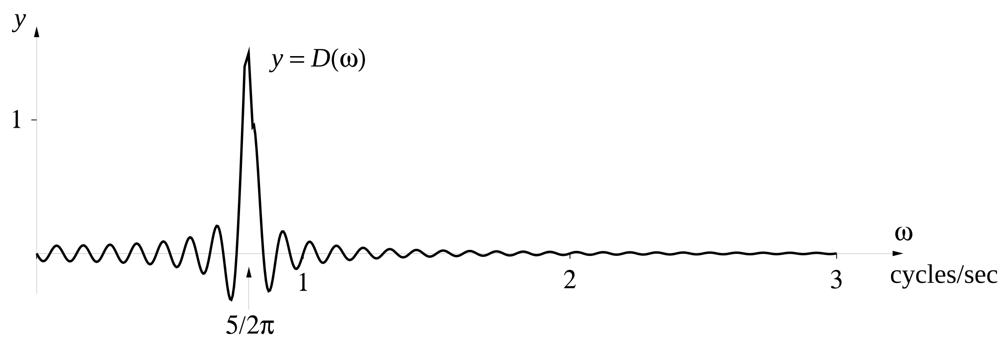
$$D(\omega) = \frac{3}{10(4\pi^2\omega^2 - 25)}(5\cos(50)\sin(20\pi\omega) - 2\pi\omega\sin(50)\cos(20\pi\omega)).$$

Notice, in your own calculations, that ω emerges as the variable on which the whole expression depends.

$D(\omega)$ peaks when ω is the frequency of the signal

The graph of $D(\omega)$ is shown on the top of the next page. You should plot it yourself, using a computer graphing utility. For most frequencies ω, the value of the detector D is close to 0. There is a single strong peak, which you can find at $\omega \approx .795$ cycles/sec. As it happens, the frequency of the signal $S = 3\sin(5t)$ is $5/2\pi = .79577\ldots$ cycles/sec! Moreover, the height of the peak is about 1.5, which is exactly half the amplitude of the signal.

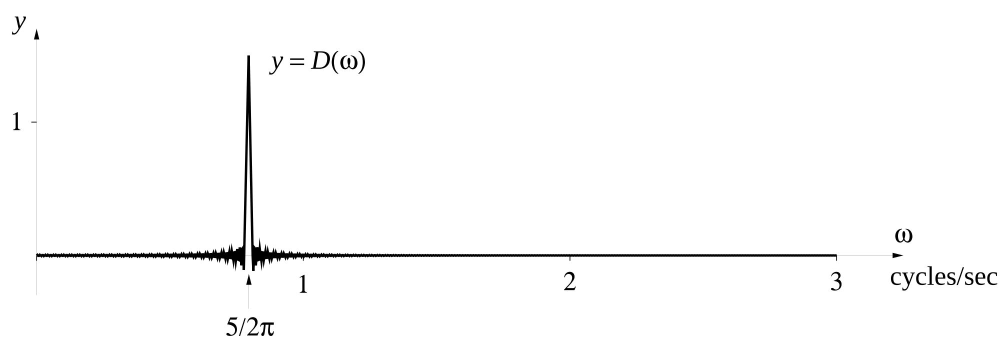
Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008



Detecting the frequency of 3sin(5*t*) on the interval $0 \le t \le 10$

The graph above tests the signal $S(t)$ when the detector is integrated over a time interval that is 10 seconds long. That is, $0 \le t \le 10$ seconds. If we repeat the test by integrating over a much larger interval, the frequency detector gives us a sharper report on the frequency of the signal. In the graph below the function $D(\omega)$ was calculated by integrating over the interval $0 \le t \le 100$ seconds.

The peak in $D(\omega)$ is sharper if the signal is tested over a longer time interval



Detecting the frequency of 3sin(5*t*) on the interval $0 \le t \le 100$

The program DETECTOR

Computation. Of course, it is rare to find a formula for $D(\omega)$ in terms of the frequency ω. For most signals $S(t)$, the best we can do is calculate the value of the integral numerically for a sequence of values of the parameter ω. The program DETECTOR, which is listed on the next page, does this. As it is written, it analyzes the function $3\sin(5t)$ on the interval $0 \le t \le 10$, and it produces the graph $D(\omega)$ at the top of this page. The "outer loop"

```
FOR j = 1 TO omegasteps  ...  NEXT j
```

plots $D(\omega)$ over the interval $0 \le \omega \le 3$, using 2^{10} equally spaced values of ω. Each $D(\omega)$ is an integral whose value is first calculated as a midpoint

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Riemann sum with 2^7 steps. The calculation is carried out by the short "inner loop"

`FOR k = 1 TO numberofsteps ... NEXT k`,

which you should recognize as an adaptation of the program RIEMANN from chapter 6.

Program: DETECTOR
To detect the frequency of a signal

```
Set up GRAPHICS
startomega = 0
endomega = 3
omegasteps = 2 ^ 10
deltaomega = (endomega - startomega) / omegasteps
twopi = 8 * ATN(1)
DEF fnf (t) = 3 * SIN(5 * t)
a = 0
b = 10
numberofsteps = 2 ^ 7
deltat = (b - a) / numberofsteps
omega = startomega
oldomega = omega
oldaccum = 0
FOR j = 1 TO omegasteps
     t = a + deltat / 2
     accum = 0
     FOR k = 1 TO numberofsteps
          deltaS = (fnf(t) * SIN(twopi * omega * t) * deltat) / (b - a)
          accum = accum + deltaS
          t = t + deltat
     NEXT k
     omega = omega + deltaomega
     Plot the line from (oldomega, oldaccum) to (omega, accum)
     oldomega = omega
     oldaccum = accum
NEXT j
```

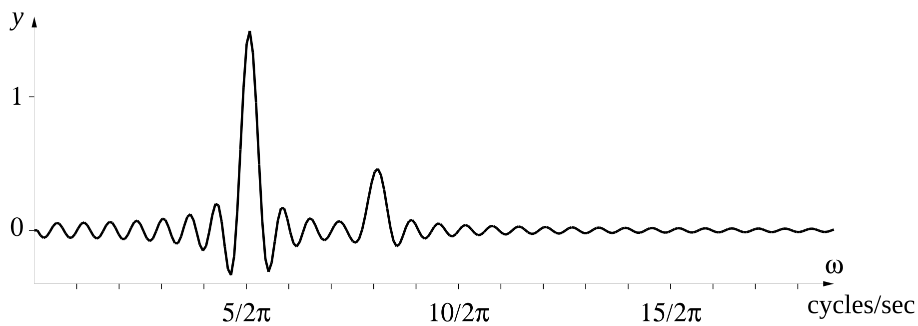
If we modify the program DETECTOR so that it analyzes the function

$$S(t) = 3\sin(5t) + \sin(8t),$$

we get the graph at the top of the next page. The scale on the ω-axis has also been modified to make it easier to read multiples of $1/2\pi$ cycles per second.

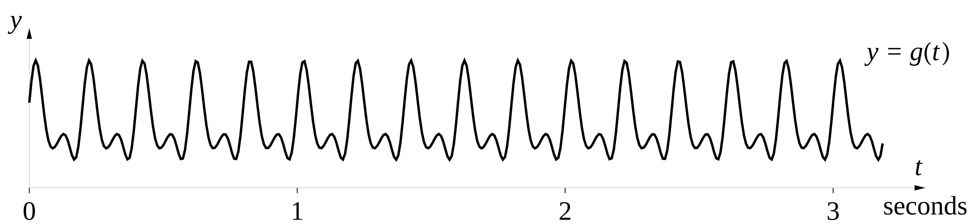
Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Notice the strongest peak is at $\omega = 5/2\pi$ cycles/sec, and $D \approx 1.5$ there. But there is now a second peak at $\omega = 8/2\pi$ cycles/sec, where $D \approx .5$. Indeed, S consists of two periodic components, one with three times the amplitude of the other. The stronger component has frequency $5/2\pi$ cycles/sec, the weaker $8/2\pi$ cycles/sec.

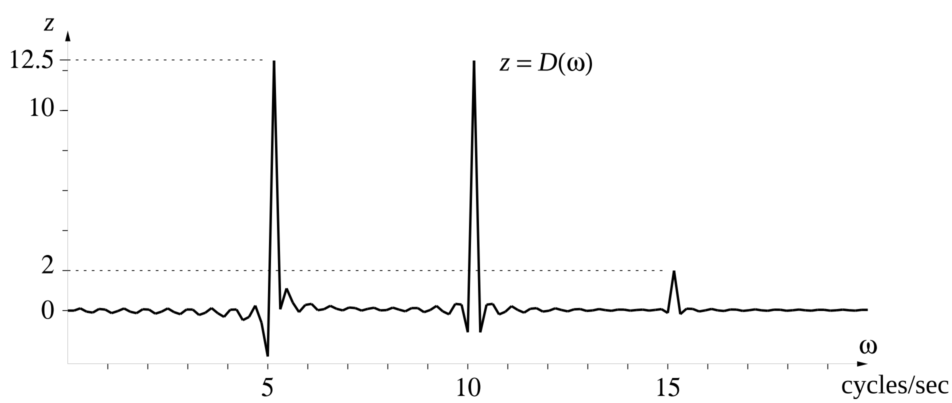


The following example first appeared in chapter 7.2. It is clear from the graph that it has a basic frequency of 5 Hz. The detector shows that it also has an equally strong component at 10 Hz and a much weaker component at 15 Hz. Can you guess a formula for $g(t)$?

A periodic signal . . .



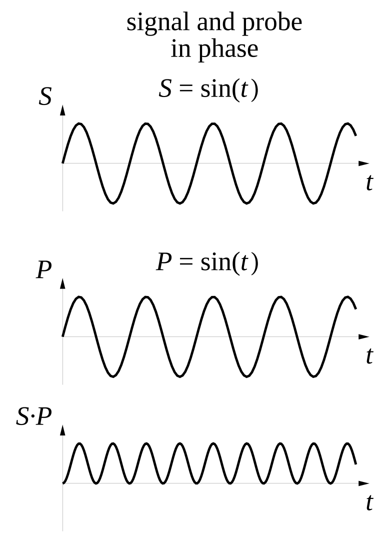
. . . and its frequency detector



The graph of $z = D(\omega)$ was produced by DETECTOR. The integral was calculated for `a = 0`, `b = 10`, and `numberofsteps = 2 ^ 9`.

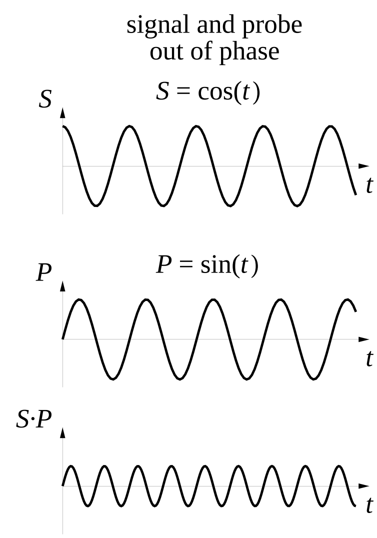
Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

The Problem of Phase



Our detector is built on the premise that, if you take the product of two functions of the same frequency, its average value will be different from 0. This is illustrated by the top three graphs on the left. The signal and the probe are both $\sin(t)$. Their product is a function that ranges between 0 and 1, and has average value 1/2. However, something quite different happens if we change the signal from $\sin(t)$ to $\cos(t)$. This doesn't change the period, but it does change the product, as you can see in the three lower graphs. The new product is centered around the t-axis; its average value is 0. Thus the detector fails to reveal that the signal has the same frequency as the probe.

A closer look at the two sets of graphs will show what has happened. In the first case, when P is positive, so is S. When P is negative, so is S. Thus, the product $S \cdot P$ is never negative; on average, its value is positive. This is what we expect.



The second case is only a little more complicated. When P is positive, S is positive only half the time; the other half it is negative. Consequently, the product $S \cdot P$ takes both positive and negative values. The same thing happens when P is negative. On average, the value of the product is 0, *even though the frequencies of P and S match.*

The problem is that their *phases* don't match. The signal $S = \cos(t)$ hits its peak $\pi/2$ seconds before the probe $P = \sin(t)$. This kind of a difference is called a **phase shift**. In the exercises for chapter 7.2, you showed that if the phase of the sine function is shifted to the left by $\pi/2$, the result is the cosine function:

$$S = \sin(t + \pi/2) = \cos(t).$$

Since $\pi/2$ radians is the same as 90°, we sometimes express this equation by saying that "the sine and the cosine are 90° out of phase."

Arbitrary phase shifts

Of course the signal could involve a phase shift of any amount φ: $S = \sin(t - \varphi)$. All these signals have the same period as the probe $P = \sin(t)$. Exercise 20 of chapter 7.2 shows what happens if this signal is tested against the probe: the average value of the product $S \cdot P$ is $\cos(\varphi)/2$. Clearly,

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

The average value varies with the phase

this depends on the size of the phase shift φ. In particular, if $\varphi = 0$ (so $S = \sin(t)$), the average value is 1/2. If $\varphi = -\pi/2$ (so $S = \cos(t)$), the average value is 0. The formula therefore agrees with what we already know for the two signals we considered as examples.

There is one more case worth glancing at: $\varphi = \pm\pi$. This is also called a phase shift of 180°. It doesn't matter whether you go forward 180° or backward; in either case $S = \sin(t \pm \pi) = -\sin(t)$. This time the average value of the product is $-1/2$.

The problem of phase . . .

The problem of phase is now be clear: The probes $P = \sin(2\pi\omega t)$ have trouble detecting the frequency of a signal that is out of phase with them. However, any phase-shifted sine function can be expressed as a sum of pure sine and cosine functions:

$$\sin(bt - \varphi) = M\sin(bt) + N\cos(bt),$$

. . . and its solution

where $M = \cos(\varphi)$ and $N = -\sin(\varphi)$. (See the exercises.) Since the sine probes P will detect $M\sin(bt)$, we need only construct a second set of probes to detect $N\cos(bt)$. The test probes we add are the cosine functions

$$P_c = \cos(2\pi\omega t).$$

We use the subscript "c" to distinguish these from the sine probes, which henceforth will be denoted P_s.

Two new detectors

We must also construct a second detector, to handle the new cosine probes. Let's take this opportunity to make a technical adjustment: we redefine a detector to be *twice* the average value of the signal and the probe. In that way, the height of the detector at a peak equals the amplitude of the signal at that frequency—rather than half the amplitude.

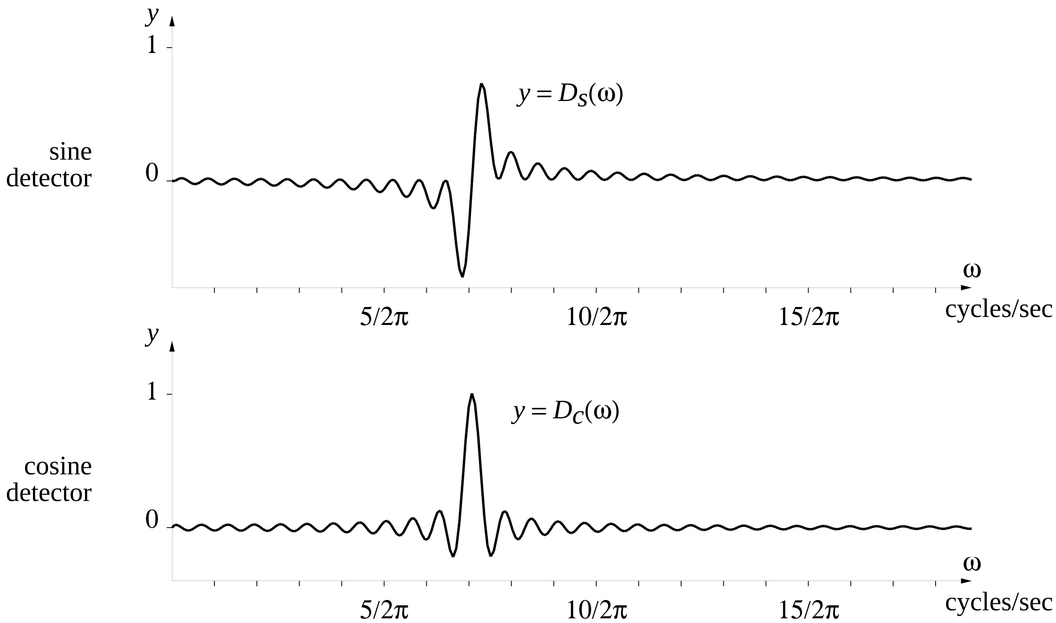
Sine detector: $$D_s(\omega) = \frac{2}{b-a}\int_a^b S(t)\sin(2\pi\omega t)\,dt.$$

Cosine detector: $$D_c(\omega) = \frac{2}{b-a}\int_a^b S(t)\cos(2\pi\omega t)\,dt.$$

The graphs of D_s and D_c

You can modify the program DETECTOR to produce the graphs of $D_s(\omega)$ and $D_c(\omega)$. You can see below how they analyze the signal $S = \cos(7t)$ over the interval $0 \le t \le 10$. The cosine detector D_c has a shape we've seen

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008



before. It has a single peak at $\omega = 7/2\pi$ cycles/sec, which is the frequency of the signal. The peak is 1 unit high, which is the amplitude of the signal. The sine detector has an unfamiliar shape. Notice first that $D_s(7/2\pi) = 0$. This confirms our earlier observation that the average value of the product of a sine and a cosine at the same frequency is 0. For values of ω slightly larger or smaller than $7/2\pi$, though, the sine detector swings relatively far from 0. This pattern is typical when a detector is analyzing a signal that is 90° out of phase with the probes.

Resonance. Try this experiment. Sit at a piano and hold all the pedals down. Then sing a note. If you sing loud enough, and hold the note long enough, one of the piano strings will start vibrating. If you stop abruptly and listen to the string, you will hear it sounding the same note you were singing. The piano has detected the frequency of your signal! It is the physical analogue of our mathematical frequency detectors. The response of the string is called **resonance**. Had you sung a lower note, a larger string would have resonated.

The physical analogue of a detector is a resonator

Resonance gives us a vivid language for describing how our detectors work. We can say a test probe "resonates" with a signal when their product is different from zero on average. The larger the average value, the stronger the resonance.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Resonance occurs all around us. Sometimes it is a nuisance—for instance, when the windows in our house rattle while a heavy truck drives by, or an air conditioner runs. Sometimes we exploit it deliberately—for instance, when we use a radio tuner as an electronic resonator to detect and amplify certain electromagnetic waves.

Detector as transform

We now have two distinct ways to describe a signal S. The function $S(t)$ is one way. It tells us how strong the signal is at each instant t. But we can also think of the signal as a mixture of sine and cosine waves of different frequencies. The detectors $D_s(\omega)$ and $D_c(\omega)$ tell us how strong the signal is at each frequency ω. That is the second way.

There is a direct connection between these two descriptions, of course. It is provided by the formulas

$$D_s(\omega) = \frac{2}{b-a}\int_a^b S(t)\sin(2\pi\omega t)\,dt \qquad D_c(\omega) = \frac{2}{b-a}\int_a^b S(t)\cos(2\pi\omega t)\,dt.$$

Integrals transform S into D_s and D_c

In effect, these formulas tell us how to *transform* the first description $S(t)$ into the second $D_s(\omega)$, $D_c(\omega)$. The transformation is so complete that even the input variable is changed—from t to ω. Look back at the formulas to see how the new variable ω is brought in.

Our detectors are essentially the same as the **Fourier sine transform** and the **Fourier cosine transform**. There is also an **inverse Fourier transform** that works in reverse: it produces $S(t)$ from the frequency data $D_s(\omega)$ and $D_c(\omega)$. The Fourier transforms are an important tool in mathematics and in science. For example, a hologram is the Fourier transform of an ordinary image. Fourier transforms and their inverses are used in photo restoration, in the enhancement of the digitized pictures sent back from cameras in space, and in filtering the signal in a stereo set.

The French mathematician Jean Baptiste Fourier (1768–1830) introduced what we call Fourier transforms and Fourier series to study the conduction of heat. Now his methods are used to study all sorts of periodic and non-periodic phenomena. They are also the foundation for the part of pure mathematics called harmonic analysis.

The Power Spectrum

A detector that ignores phase differences

The sine and cosine detectors provide enough information to reconstruct the original signal in complete detail—including phase. Often, though, they provide more detail than we want. We can use another tool—called the **power**

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

spectrum—to determine only the strength of the different frequencies that occur in a signal, without regard to their phase. The power spectrum is constructed from the two detectors in the following way:

Power spectrum: $P(\omega) = \sqrt{[D_s(\omega)]^2 + [D_c(\omega)]^2}$

To see how the power spectrum works, we'll consider the signal $S(t) = A\sin(7t - \varphi)$. This is a sine wave of frequency $\omega = 7/2\pi$ and amplitude A. Let's concentrate first on $\omega = 7/2\pi$. If there were no phase shift φ present, we would expect that

$$D_s(7/2\pi) = A \qquad D_c(7/2\pi) = 0.$$

However, because there is a phase shift, the actual values turn out to be

$$D_s(7/2\pi) = A\cos\varphi \qquad D_c(7/2\pi) = -A\sin\varphi.$$

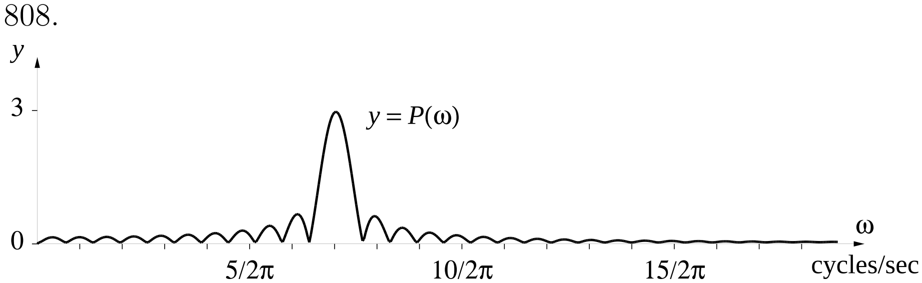
(These calculations are given as exercises.) The values of the detectors clearly depend on the phase shift. By contrast,

$$\begin{aligned} P(7/2\pi) &= \sqrt{[D_s(7/2\pi)]^2 + [D_c(7/2\pi)]^2} \\ &= \sqrt{A^2\cos^2\varphi + A^2\sin^2\varphi} \\ &= A. \end{aligned}$$

We have used the fact that $\cos^2\varphi + \sin^2\varphi = 1$ for every φ. Thus, the power spectrum does *not* depend on the phase. It tells us only the amplitude of the signal at the frequency $\omega = 7/2\pi$.

If we calculate the power spectrum over all frequencies ω, we get the graph shown at the top of the next page. The program POWER generates this graph. It was derived from the program DETECTOR. Compare the two programs, particularly the terms `deltaS` and `deltaC`. In POWER, they have been multiplied by 2, to agree with our new definition of D_s and D_c on page 808.

The program POWER



Power spectrum of $3\sin(7t - \pi/3)$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Program: POWER
The power spectrum of a signal

```
Set up GRAPHICS
startomega = 0
endomega = 3
omegasteps = 2 ^ 9
deltaomega = (endomega - startomega) / omegasteps
pi = 4 * ATN(1)
twopi = 2 * pi
DEF fnf (t) = 3 * SIN(7 * t - pi / 3)
a = 0
b = 10
numberofsteps = 2 ^ 6
deltat = (b - a) / numberofsteps
omega = startomega
oldomega = omega
oldpower = 0
FOR j = 1 TO omegasteps
     t = a + deltat / 2
     accumS = 0
     accumC = 0
     power = 0
     FOR k = 1 TO numberofsteps
          deltaS = 2 * (fnf(t) * SIN(twopi * omega * t) * deltat) / (b - a)
          accumS = accumS + deltaS
          deltaC = 2 * (fnf(t) * COS(twopi * omega * t) * deltat) / (b - a)
          accumC = accumC + deltaC
          t = t + deltat
     NEXT k
     power = SQR(accumS ^ 2 + accumC ^ 2)
     omega = omega + deltaomega
     Plot the line from (oldomega, oldpower) to (omega, power)
     oldomega = omega
     oldpower = power
NEXT j
```

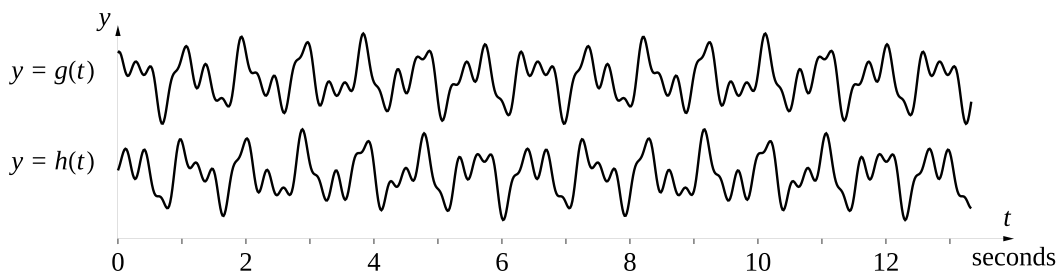
Two signals whose components differ only in phase

To see how the power spectrum detects the frequencies in a signal while overlooking the phases of the different components, consider these two signals:

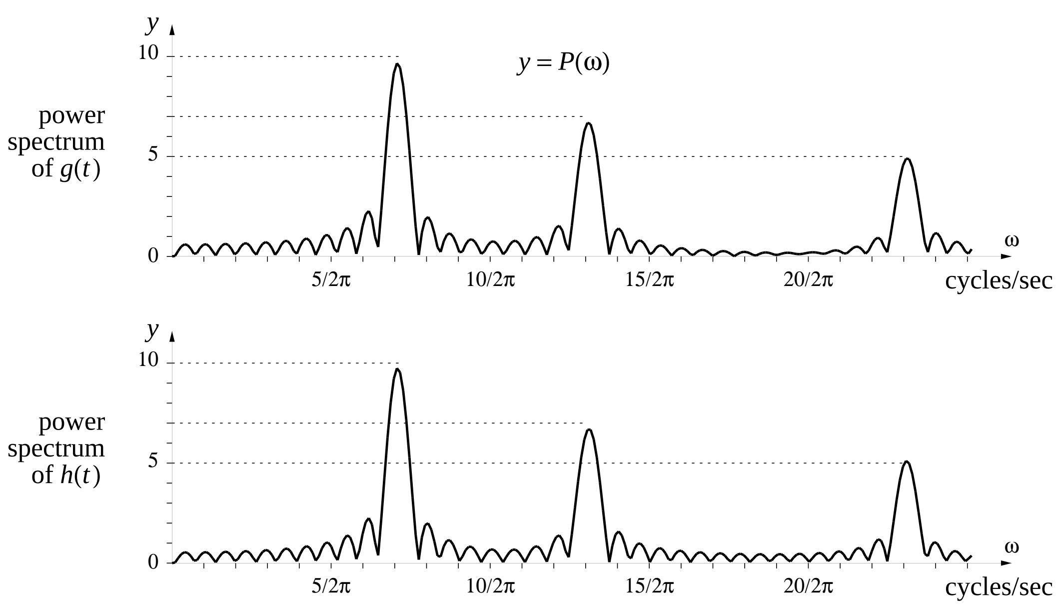
$$\begin{aligned} g(t) &= 10\sin(7t) + 7\cos(13t) + 5\cos(23t) \\ h(t) &= 10\sin(7t) + 7\cos(13t) - 5\cos(23t) \end{aligned}$$

They differ only in the sign of the last term. This is equivalent to a phase shift of 180° in that term. The graphs are drawn below (with constants

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008



added to separate them vertically). It is remarkable how different the graphs appear to be, considering how nearly alike their formulas are. You can find similarities if you look closely, though. For instance, the peaks of one graph tend to match the peaks of the other.



The power spectrum, however, has no trouble detecting the similarities between the two signals. As you can see, they indicate that the same dominant frequencies occur in g and h, and that corresponding frequencies occur with the same amplitude. We learn that the formula for g or h can be written as

$$10\sin(7t - \varphi_1) + 7\sin(13t - \varphi_2) + 5\sin(23t - \varphi_3).$$

The only thing we can't learn from the power spectrum are the three phase differences φ_1, φ_2, φ_3.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

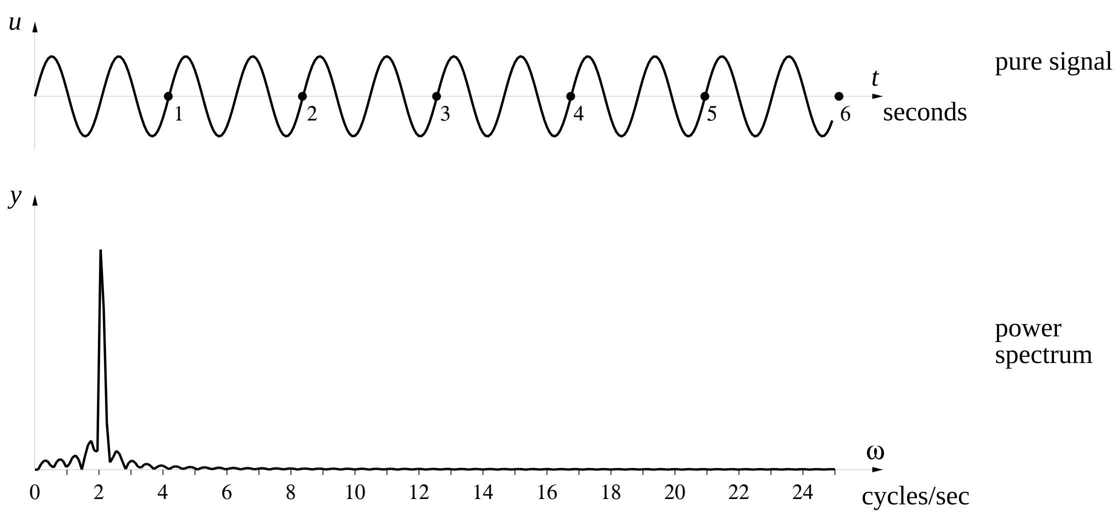
The graphs of the power spectra were drawn by POWER, using the following values:

```
endomega = 4
omegasteps = 2 ^ 8
numberofsteps = 2 ^ 7
```

These two graphs actually differ very slightly. You can see the difference most clearly near $\omega = 20/2\pi$.

Detecting a periodic wave in a noisy signal

For a final demonstration of the properties of the power spectrum, we return to the signal + noise problem that we raised at the beginning of this section. Let's see what happens to the power spectrum of a pure sine wave when we gradually gradually add noise. For simplicity, we take the frequency of the pure signal to be 2 cycles/sec. The spectrum has a single strong spike at this frequency.

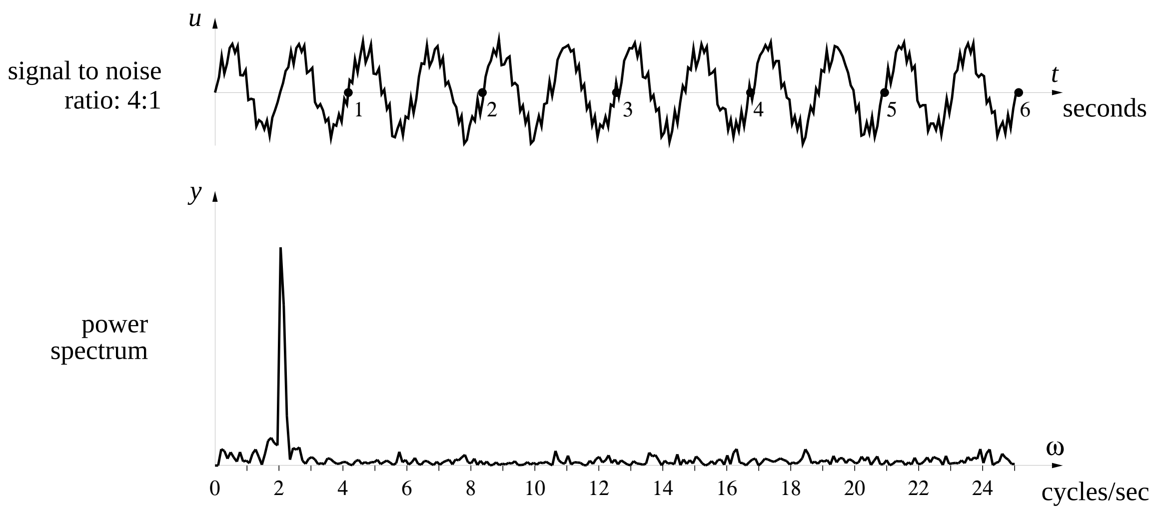


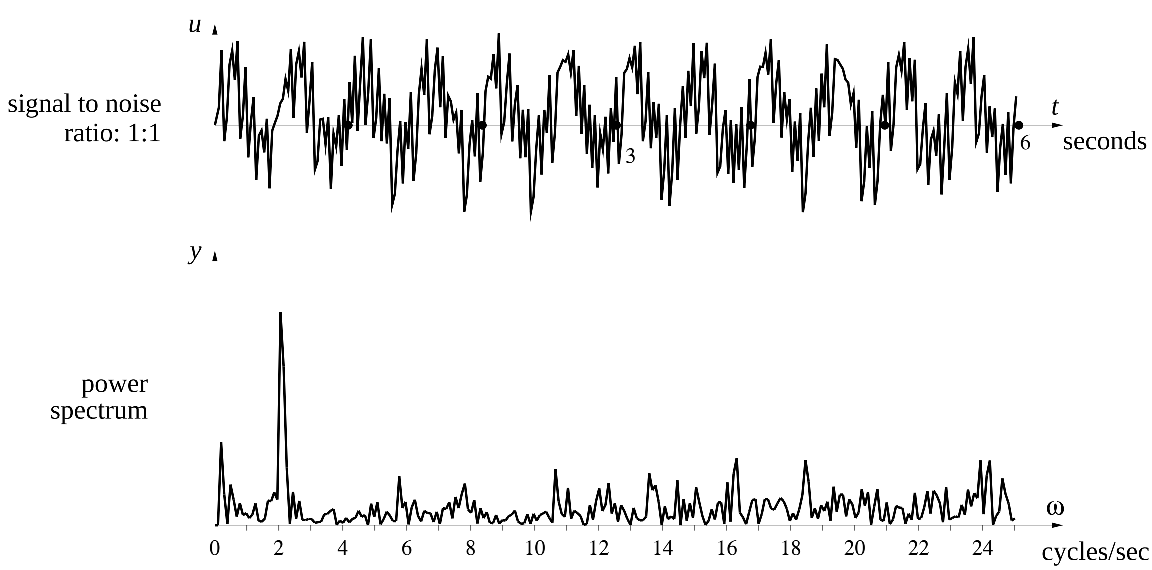
One the following pages you can see what happens as the noise level is increased. The power spectrum, which was virtually zero for all $\omega > 3$ cycles/sec, is now non-zero for almost all frequencies in the range we have graphed. In other words, the noise is a mixture of many frequencies. Notice how the height of the power graph increases with the strength of the noise. This is most noticeable in the higher frequencies. Eventually, in the final graph, we lose sight of the signal; the noise has swamped it. The signal to noise ratio is 1:2.5, meaning that the noise is $2\frac{1}{2}$ times as strong as the signal. Nevertheless, the power spectrum still shows a strong spike at $\omega = 2$

In the power spectrum, noise and signal are separated

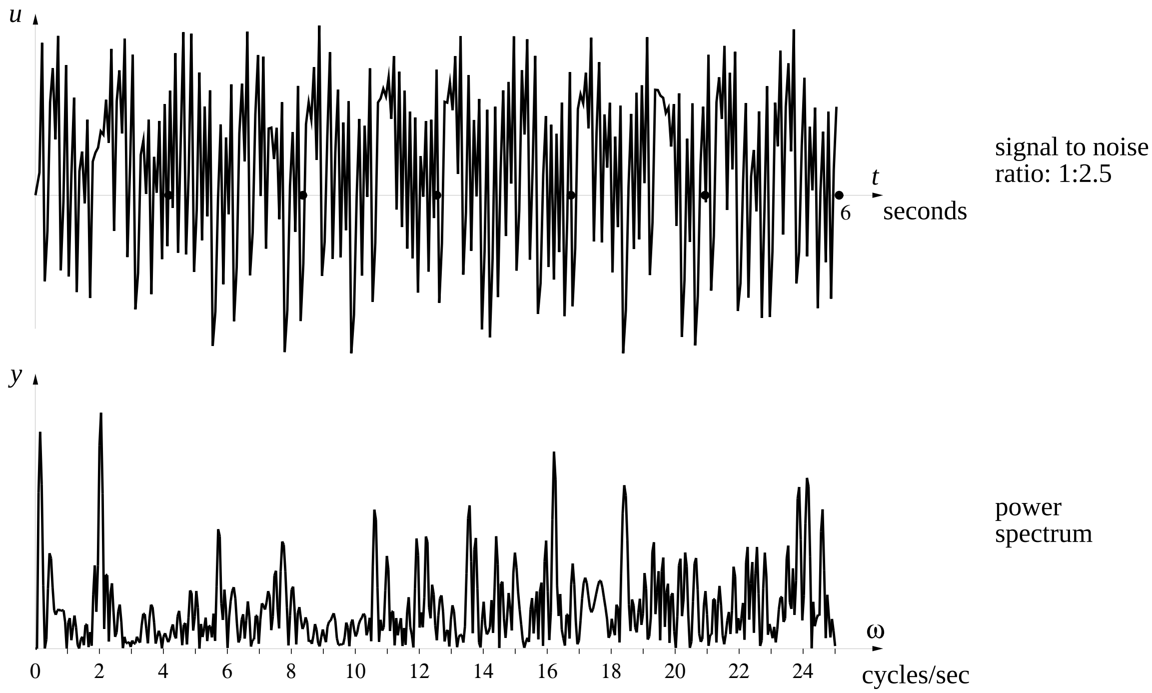
Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

cycles/sec. This corresponds to the signal. The power spectrum can still see the signal even when we can't!





Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008



Exercises

The problem of phase

1. Use the "sum of two angles formula,"

$$\sin(A + B) = \sin(A)\cos(B) + \cos(A)\sin(B),$$

to show that the circular function $\sin(bt - \varphi)$ with period $2\pi/b$ and phase difference φ can be written as a combination of pure sine and cosine functions of the same period:

$$\sin(bt - \varphi) = M \sin(bt) + N \cos(bt).$$

show that $M = \cos(\varphi)$ and $N = -\sin(\varphi)$. [Note that $M^2 + N^2 = 1$.]

2. a) Express $\sin(5t - \pi/3)$ as a sum of a pure sine function and a pure cosine function.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

b) Express $\frac{\sqrt{3}}{2}\sin(7t) + \frac{1}{2}\cos(7t)$ in the form $A\sin(bt - \varphi)$. To check your result, graph it together with the given function using a computer graphing utility.

c) Express $f(t) = \sin(t) + 2\cos(t)$ in the form $A\sin(bt - \varphi)$. Notice that the formula in exercise 1 requires that $M^2 + N^2 = 1$, but in this example $M^2 + N^2 = 5$. Therefore, first write

$$f(t) = \sqrt{5}\left(\frac{1}{\sqrt{5}}\sin(t) + \frac{2}{\sqrt{5}}\cos(t)\right).$$

The expression in parentheses has the right form. Does your result check on a computer?

3. Suppose

$$A\sin(bt - \varphi) = M\sin(bt) + N\cos(bt).$$

How are A, M, and N related?

4. The functions $\sin(t) + 2\cos(t)$ and $2\sin(t) + \cos(t)$ have the same period but differ in phase. What is the phase difference? Determine this two ways: by graphing, and by writing each expression as a single function of the form $A\sin(bt - \varphi)$.

5. Choose values for A, b, and φ so that the function

$$3\sin(2x) + 4\cos(2x) + A\sin(bx - \varphi)$$

is *identically* zero—that is, equal to 0 for every value of x.

6. Choose values of A and φ so that the function

$$\sin(x) + \sin(x + 1) + \sin(x + 2) + A\sin(x - \varphi)$$

is identically zero.

The programs DETECTOR and POWER

The purpose of these exercises is to give you experience interpreting the power spectrum of a known signal using the program POWER and modifications of DETECTOR. The first exercise asks you to construct these modifications.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

7. Modify DETECTOR to produce two new programs, SDETECTOR and CDETECTOR, which generate the sine detector and the cosine detector functions that appear on page 808.

8. a) Compare the outputs of $f(t) = \sin(t)$ and $g(t) = \cos(t)$ on POWER. Use the domain $0 \leq \omega \leq 1$. Does POWER distinguish between these functions? Would you expect it to?

b) Compare $f(t)$ and $g(t)$ using SDETECTOR. Does SDETECTOR distinguish between these functions? Would you expect it to?

c) Compare $f(t)$ and $g(t)$ using CDETECTOR. Does CDETECTOR distinguish between these functions? Is the output of $g(t)$ on CDETECTOR the same as the output of $f(t)$ on SDETECTOR?

9. a) Describe the power spectrum of the signal $S = \sin(t) + \cos(t)$. How many peaks are there, and where are they?

b) How does the spectrum of S compare with the two generated in the last question?

c) Describe the output of SDETECTOR and CDETECTOR for the signal S. Compare these outputs to the corresponding outputs for f and g in the last exercise.

10. a) Graph the function

$$h(t) = 10\sin(7t) + 7\cos(13(t) - 5\cos(23t)$$

over the domain $0 \leq t \leq 14$. Compare your result with the graph on page 812.

b) Graph the power spectrum of $h(t)$ over the frequency domain $0 \leq \omega \leq 4$. Compare your result with the text. How many peaks are there? Where are they? How high are they? Do these results agree with the amplitude and frequency information provided by the formula for $h(t)$?

11. (Continuation of the previous exercise.) Use SDETECTOR to analyze $h(t)$ over the same frequency domain. Compare the pattern near $\omega = 13/2\pi$ with the patterns generated by the sine and cosine detectors that appear on page 809. Compare the patterns near $\omega = 7/2\pi$ and near $\omega = 23/2\pi$ the same way. Would you expect the patterns near $\omega = 13/2\pi$ and $\omega = 23/2\pi$ to be similar? Are they? Are they similar to the pattern near $\omega = 7/2\pi$? Is this what you would expect?

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

12. (Continuation.) Use CDETECTOR to analyze $h(t)$. Follow the guidelines of the previous question.

A Grain of Salt

The purpose of the power spectrum is to make visible the periodic patterns contained with a given function. However, our *method of computing* the spectrum can introduce spurious information, too. It can tell us there are periods that are not really present in the function. So we must take the calculations with a grain of salt. The purpose of these exercises is to point out the spurious information, show why it arises, and how we can get rid of it.

13. Use the program POWER to graph the power spectrum of the function $\sin 2\pi x$ on the interval $0 \leq x \leq 10$. Let $0 \leq \omega \leq 3$. Set

```
numberofsteps = 100
```

but let all the other parameters keep the values they have in the program.
[Answer: The power spectrum has a single peak of height 1 at $\omega \approx 1$]

14. Now increase the domain of integration to $0 \leq x \leq 30$, and set

```
numberofsteps = 300
```

to adjust for the increase in the size of the domain. Use POWER again to graph the power spectrum. Compare this spectrum with the previous one.

15. Leave $0 \leq x \leq 30$, but restore `numberofsteps = 100`. Use POWER once again to graph the power spectrum. Compare this spectrum with the previous two.
[Answer: A new peak, of height 1, appears at $\omega \approx 7/3$.]

16. Let `numberofsteps = 50`, and calculate the power spectrum one more time. What happens?

When we reduce the number of integration steps, new peaks appear in the power spectrum. These new peaks represent *spurious* information: the function $\sin 2\pi x$ has no components whose frequencies are 2/3, 7/3, or 8/3. Let's see why this happens. We'll concentrate on $\omega = 7/3$. First, you must

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

decide whether the peak in the power spectrum at $\omega = 7/3$ comes from the sine or the cosine detector.

17. Use SDETECTOR and CDETECTOR to analyze $\sin(2\pi x)$. Take $0 \le x \le 30$, $0 \le \omega \le 3$, and set `numberofsteps = 100`. One of these detectors has the value 0 when $\omega = 7/3$. Which one?

18. According to the previous exercise, the peak in the power spectrum that is detected at $\omega \approx 7/3$ comes from the integral

$$\frac{2}{30}\int_0^{30} \sin(2\pi x)\sin\left(2\pi\tfrac{7}{3}x\right)\,dx,$$

not from the cosine integral. By using one of the sine and cosine integrals from the exercises for chapter 11.3, determine the *exact* value of this integral. Is this the value you expected to get?

The program POWER calculates the spectrum numerically. In particular, we used it to calculate

$$\int_0^{30} \sin(2\pi x)\sin\left(2\pi\tfrac{7}{3}x\right)\,dx,$$

with 100 steps. The step size is therefore $\Delta x = .3$. In the following exercises you will duplicate this numerical work "by hand."

19. Make a sketch of the graph of the function

$$h(x) = \sin(2\pi x)\sin\left(2\pi\tfrac{7}{3}x\right)$$

on an appropriate interval. What is the period of this function?

20. Determine the value of $h(x)$ at $x = 0$, .3, .6, .9, 1.2, and 1.5, and use these values to construct a Riemann sum for the integral

$$\int_0^{1.5} h(x)\,dx$$

using left endpoints and a step size of $\Delta x = .3$. Mark these values of h on the sketch you made in the previous exercise.
[Answer: The Riemann sum is $-.3(2\sin^2(2\pi/5) + 2\sin^2(\pi/5)) = -.75$.]

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

21. Evaluate the expression

$$\frac{1}{15}\int_0^{30} h(x)\,dx$$

using a left endpoint Riemann sum with a step size of $\Delta x = .3$ How can you use the previous exercise to answer this question?

[Answer: -1. Since $h(x)$ is periodic with period $x = 1.5$, the interval $[0, 30]$ contains 20 periods of h. The integral of h over $[0, 30]$ is therefore 20 times its integral over $[0, 1.5]$.]

22. Compare the values of the detector

$$\frac{2}{30}\int_0^{30} \sin(2\pi x)\sin\left(2\pi \tfrac{7}{3}x\right)\,dx,$$

you have obtained by antidifferentiation and by numerical integration.

These exercises demonstrate that the exact and computed values of the power spectrum can be quite different, essentially because the steps in a Riemann sum can pick out very special values of the integrand.

One way to deal with the problem is to increase the number of steps. How will you know if you have gone far enough? Increase in stages until the graph of the power spectrum **stabilizes**—that is, until it no longer changes when you make a further increase in the number of steps.

The *true* spectrum is the limit of the computed graphs of the spectrum

Of course, increasing the number of steps increases computer time. This creates new problems. To deal with them, however, we can switch to more efficient numerical integration methods. Simpson's rule (chapter 11.6) is the most efficient method we have covered. You should try rewriting DETECTOR using Simpson's rule to see how it improves the performance.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

12.4 Fourier Series

In chapter 10.6 we obtained polynomials which were good approximations to a function over an interval, where "good" meant minimizing the *mean squared separation* between the function and the approximating polynomials.

Difficulties with polynomial approximations

While polynomials are the most obvious approximating functions to use due to the ease with which they can be evaluated, we have seen that finding good approximating polynomials leads to several serious technical complications. The first is that we have to solve systems of equations to determine the unknown coefficients, a procedure that is very time-consuming, even for a computer if we are trying to get a polynomial of, say, degree 30. Further, if we are trying to make the approximation over even a moderately-sized interval, since we are evaluating expressions of the form x^n, we get large numbers very rapidly as x and n get large. This in turn leads to roundoff problems in the computer routines.

Another aspect of these polynomial approximations that makes them complicated is that the values of the coefficients change as we change the degree of the approximating polynomial. Thus if we determine the least squares fourth-degree approximation and then decide we want the fifth-degree approximation instead, all the coefficients have to be recalculated. Knowing what the coefficient of x^3 was in the fourth-degree approximation is no help at all in knowing what the coefficient of x^3 will be in the fifth-degree approximation.

Approximating periodic functions

There are approximating functions of another kind that avoid such difficulties. Moreover, these functions are natural ones to use when we are trying to approximate *periodic* functions. In such cases it is reasonable to take the simplest periodic functions—sines and cosine—and try to combine them to approximate more complicated periodic functions. This suggests that we want to look at functions of the form

$$\begin{aligned}\phi(x) &= a_0 + a_1\cos x + a_2\cos 2x + \cdots + a_n\cos nx \\ &\quad + b_1\sin x + b_2\sin 2x + \cdots + b_n\sin nx \\ &= a_0 + \sum_{k=1}^{n} a_k\cos kx + b_k\sin kx.\end{aligned}$$

Such a combination is called a **trigonometric polynomial of degree n**. Note that any function of this form will in fact be periodic with period 2π. More generally, if we were interested in approximating a function of period T,

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

we would want to look at trigonometric polynomials of the form

Trigonometric polynomial of degree n and period T

$$\begin{aligned}\phi(x) &= a_0 + a_1 \cos\frac{2\pi x}{T} + a_2 \cos\frac{4\pi x}{T} + \cdots + a_n \cos\frac{2n\pi x}{T} \\ &\quad + b_1 \sin\frac{2\pi x}{T} + b_2 \sin\frac{4\pi x}{T} + \cdots + b_n \sin\frac{2n\pi x}{T} \\ &= a_0 + \sum_{k=1}^{n} a_k \cos\frac{2k\pi x}{T} + b_k \sin\frac{2k\pi x}{T}.\end{aligned}$$

You should verify that this does indeed have period T.

To find the coefficients a_k and b_k of the trigonometric polynomial that best fits a period-2π function f over the interval $[c, c+2\pi]$, we proceed exactly as we did in the previous section, using the least squares criterion. That is, for a given degree n, we want to find coefficients $a_0, \ldots, a_n$ and $b_1, \ldots, b_n$ that minimize the integral

$$\int_c^{c+2\pi} (f(x) - \phi(x))^2 \, dx.$$

In practice, c is usually either 0 or $-\pi$.

Coefficients are independent of n

The solution turns out to be remarkably compact and easy to state. One of the key features of the formulas for the coefficients is that they are independent of each other and of the particular value of n being used. Thus, for example, a_3 in the 7-th degree approximation has the same value as a_3 in the 39-th degree approximation. This is a major advantage compared to the polynomial approximations over intervals that we worked with in chapter 10.

For a function f with period 2π, its least squares nth degree trigonometric polynomial approximation over a full period is

$$\phi_n(x) = a_0 + \sum_{k=1}^{n} a_k \cos kx + b_k \sin kx,$$

where

$$a_0 = \frac{1}{2\pi}\int_0^{2\pi} f(x)\,dx,$$

$$a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cdot\cos kx\,dx \quad \text{for } k = 1, 2, \ldots, n,$$

$$b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\cdot\sin kx\,dx \quad \text{for } k = 1, 2, \ldots, n.$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

The infinite series

$$a_0 + \sum_{k=1}^{\infty} a_k \cos kx + b_k \sin kx$$

with the prescribed values for a_k and b_k is called the **Fourier series** for f, and the coefficients a_k and b_k are called the **Fourier coefficients** for f. It turns out that any continuous function equals its Fourier series in the same sense we used earlier with Taylor series—for any x in the given interval, $f(x)$ is the limit as $n \to \infty$ of the nth degree approximating trigonometric polynomials evaluated at x. The derivation is straightforward, but we shall leave it to the end of this section so we can look at some examples first.

Joseph Fourier (1768–1830) was active in both politics and in mathematics. He was an advocate of the French Revolution, worked as an engineer in Napoleon's army, and served as a prefect for a while. In mathematics he was interested in the mathematics of heat conduction and developed the series that now bear his name as a tool for investigating problems in this area. His ideas initially met with considerable resistance, but eventually became a central tool in mathematics.

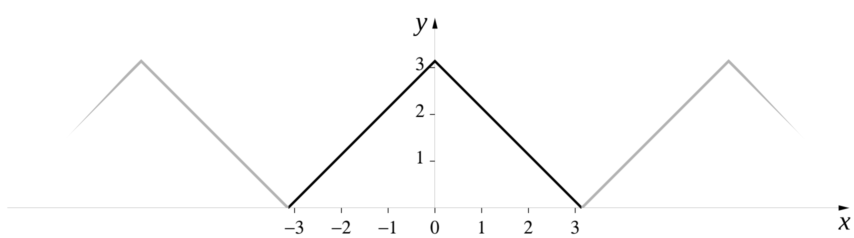
Although our formulas give the values of a_k and b_k in terms of integrals over $[0, 2\pi]$, periodicity of the integrands implies that integrations over *any* interval of width 2π gives the same values. In practice (as in the first example, immediately below), we often use $[-\pi, \pi]$ instead of $[0, 2\pi]$.

Example 1 Let's find the approximating trigonometric polynomials for

$$f(x) = \begin{cases} \pi + x & \text{if } -\pi \le x \le 0, \\ \pi - x & \text{if } 0 \le x \le \pi. \end{cases}$$

Then the graph of f simply consists of two line segments:

The graph of f is "triangular"



Now make $f(x)$ periodic over the entire x-axis by horizontal translations: $f(x) = f(x-2\pi)$. The periodic graph is shown in gray, above. We can obtain first Fourier coefficient without any calculus at all: $a_0 = (1/2\pi)\,\pi^2 = \pi/2$. It

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

is just the area of one triangle divided by 2π. The other coefficients can be evaluated with integration by parts (chapter 11.3). We have

$$\begin{aligned} a_k &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)\cos kx\,dx \\ &= \frac{1}{\pi}\int_{-\pi}^{0}(\pi+x)\cos kx\,dx + \frac{1}{\pi}\int_{0}^{\pi}(\pi-x)\cos kx\,dx. \end{aligned}$$

The first of these integrals can be evaluated as

$$\begin{aligned} \int_{-\pi}^{0}(\pi+x)\cos kx\,dx &= (\pi+x)\frac{\sin kx}{k}\Big|_{-\pi}^{0} - \int_{-\pi}^{0}\frac{\sin kx}{k}\,dx \\ &= 0 + \frac{\cos kx}{k^2}\Big|_{-\pi}^{0} \\ &= \begin{cases} \dfrac{2}{k^2} & \text{if } k \text{ is odd,} \\ 0 & \text{if } k \text{ is even.} \end{cases} \end{aligned}$$

Similarly we find

$$\begin{aligned} \int_{0}^{\pi}(\pi-x)\cos kx\,dx &= (\pi-x)\frac{\sin kx}{k}\Big|_{0}^{\pi} + \int_{0}^{\pi}\frac{\sin kx}{k}\,dx \\ &= 0 - \frac{\cos kx}{k^2}\Big|_{0}^{\pi} \\ &= \begin{cases} \dfrac{2}{k^2} & \text{if } k \text{ is odd,} \\ 0 & \text{if } k \text{ is even.} \end{cases} \end{aligned}$$

Combining these two integrals we find

$$a_k = \begin{cases} \dfrac{4}{\pi k^2} & \text{if } k \text{ is odd,} \\ 0 & \text{if } k \text{ is even.} \end{cases}$$

An analogous derivation will show that all the b_k are 0; this is left to the exercises. We can thus write down the Fourier series for f:

$$f(x) = \frac{\pi}{2} + \frac{4}{\pi}\left(\frac{\cos x}{1} + \frac{\cos 3x}{9} + \cdots + \frac{\cos(2n+1)x}{(2n+1)^2} + \cdots\right).$$

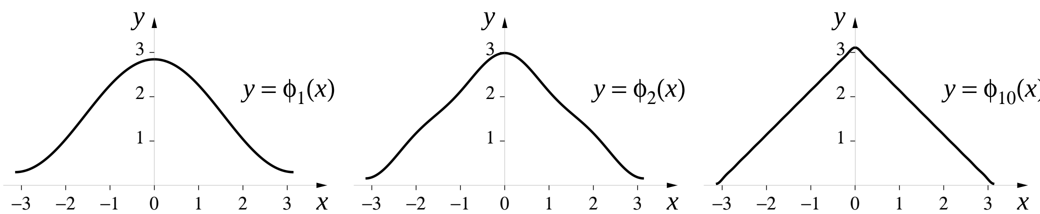
The Fourier series for f

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Let

$$\phi_n(x) = \frac{\pi}{2} + \frac{4}{\pi}\sum_{k=0}^{n} \frac{\cos{(2k+1)x}}{(2k+1)^2}.$$

Here are the graphs of $\phi_1(x)$, $\phi_2(x)$, and $\phi_{10}(x)$:

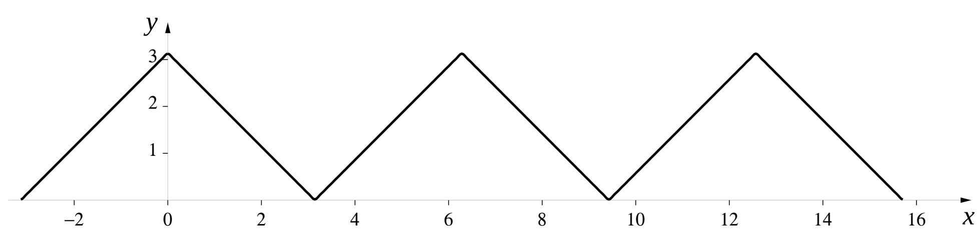


We see that $\phi_{10}(x)$ already appears to be a very good approximation to $f(x)$. If we look at the maximum separation between $f(x)$ and $\phi_n(x)$ over $[-\pi, \pi]$ for different values of n, we get the following:

n	1	2	10	50	100	1000
$\max\limits_{-\pi \le x \le \pi} \lvert f(x) - \phi_n(x) \rvert$	.298	.156	.032	.0064	.0032	.00032

Approximating a triangular wave-form

Since $\phi_n(x)$ is periodic, if we graph it over a larger interval, we get an approximation to a **triangular wave-form**. Here, for example, is the graph of $\phi_{20}(x)$ over the interval $[-\pi, 5\pi]$:



Remark 2 In addition to their use in approximating functions, Fourier series can lead to some interesting, and non-obvious, mathematical results. For instance in the preceding example, we have $f(0) = \pi$. On the other hand, we should get the same value if we set $x = 0$ in the Fourier series for f. This leads to the identity

$$\pi = \frac{\pi}{2} + \frac{4}{\pi}\left(\frac{1}{1} + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \frac{1}{81} + \cdots\right).$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

With a little rearranging, this can be rewritten as

$$\frac{\pi^2}{8} = 1 + \frac{1}{9} + \frac{1}{25} + \frac{1}{49} + \frac{1}{81} + \cdots$$

—that is, if we add up the reciprocals of the squares of all the odd integers, we get $\pi^2/8$!

The general rule for calculating Fourier series

The formulas given on page 823 for approximating functions with period 2π extend readily to approximating periodic functions of any period T. For instance, if we wanted to approximate some function f over the interval $[0, T]$, we have the following formulas. Check that when $T = 2\pi$, these equations reduce to the earlier ones. Again, there is nothing special about the interval $[0, T]$. If we had wanted to make the approximation over any other interval of length T—for example, $[-T/2, T/2]$—we simply change the limits of integration to be the endpoints of that interval.

For a function $f(t)$ with period T, its least squares nth degree trigonometric polynomial approximation over a full period is

$$\phi_n(x) = a_0 + \sum_{k=1}^{n} a_k \cos\frac{2k\pi x}{T} + b_k \sin\frac{2k\pi x}{T},$$

where

$$a_0 = \frac{1}{T}\int_0^T f(x)\,dx,$$

$$a_k = \frac{2}{T}\int_0^T f(x)\cdot\cos\frac{2k\pi x}{T}\,dx \quad \text{for } k = 1, 2, \ldots, n,$$

$$b_k = \frac{2}{T}\int_0^T f(x)\cdot\sin\frac{2k\pi x}{T}\,dx \quad \text{for } k = 1, 2, \ldots, n.$$

Fourier series for the periodic solutions of May's predator-prey model

Example 2 Consider the predator–prey model of May that we examined in chapter 7.3. Recall that there were two species, the predator y and the prey x, interacting according to the model

$$\text{prey:} \quad x' = .6\,x\left(1 - \frac{x}{10}\right) - \frac{.5\,xy}{x+1},$$

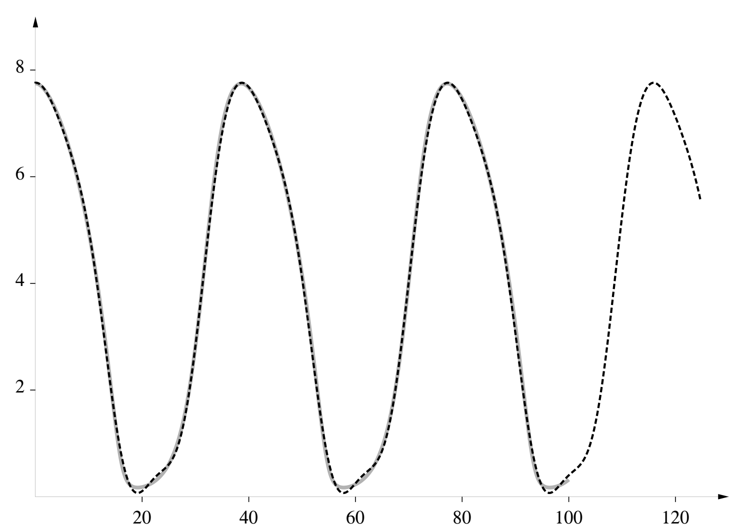
$$\text{predator:} \quad y' = .1\,y\left(1 - \frac{y}{2x}\right).$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

We discovered that the populations seemed to move towards periodic cycles, regardless of the initial conditions (although the phase of the cycles did depend on the starting values). In particular, if we begin with values on this cycle, our solution should be perfectly periodic with period, it turned out, $T = 38.6$ days. So let's start with $x = 7.75$ and $y = 2.38$, values that put us at the peak of the prey cycle. Now go to the differential equations and compute the solution numerically, storing the x-values as an array. We can then use these values to calculate all the integrals needed to find the Fourier coefficients to approximate the function $x(t)$. Here are the first 13 terms of the series:

$$\begin{aligned} x(t) = 3.7951 &+ 3.8125\cos\frac{2\pi x}{T} + .1514\cos\frac{4\pi x}{T} + .0326\cos\frac{6\pi x}{T} \\ &- .0303\cos\frac{8\pi x}{T} - .0609\cos\frac{10\pi x}{T} + .0308\cos\frac{12\pi x}{T} + \cdots \\ &+ 1.1724\sin\frac{2\pi x}{T} - .0867\sin\frac{4\pi x}{T} - .3954\sin\frac{6\pi x}{T} \\ &+ .0639\sin\frac{8\pi x}{T} - .0142\sin\frac{10\pi x}{T} + .0129\sin\frac{12\pi x}{T} + \cdots . \end{aligned}$$

Let $\phi_3(t)$ be the 7-term trigonometric polynomial whose final terms involve $\cos(6\pi t/T)$ and $\sin(6\pi t/T)$. Below we graph $\phi_3(t)$ (dashed line) and $x(t)$ (solid gray line) together. They are almost indistinguishable; we let $\phi_3(t)$ run on a little beyond $x(t)$ so you can see it's there.



Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Derivation of the formula for the Fourier coefficients

The logic behind the derivation is the same as that used in the previous subsection to find the least squares polynomial approximations. Fix n and let

$$\phi(x) = a_0 + \sum_{k=1}^{n} a_k \cos\frac{2k\pi x}{T} + b_k \sin\frac{2k\pi x}{T},$$

where now we want to choose values of the a_k and b_k to minimize the integral

$$\int_0^T (f(x) - \phi(x))^2 \, dx.$$

The value of this integral is thus a function of the undetermined coefficients $a_0, \ldots, a_n$ and $b_1, \ldots, b_n$. To find the coefficients that minimize that value we calculate the partial derivatives with respect to a_0, a_1, ... as before and set them equal to 0.

Note that

$$\frac{\partial}{\partial a_m}\phi(x) = \cos\frac{2m\pi x}{T} \quad \text{and} \quad \frac{\partial}{\partial b_m}\phi(x) = \sin\frac{2m\pi x}{T},$$

so that

$$\frac{\partial}{\partial a_m}\int_0^T (f(x) - \phi(x))^2 \, dx = \int_0^T 2(f(x) - \phi(x))\left(-\cos\frac{2m\pi x}{T}\right) dx$$

and

$$\frac{\partial}{\partial b_m}\int_0^T (f(x) - \phi(x))^2 \, dx = \int_0^T 2(f(x) - \phi(x))\left(-\sin\frac{2m\pi x}{T}\right) dx.$$

The condition that all the partial derivatives must be 0 thus leads to the equations

$$\begin{aligned}
\int_0^T 2(f(x) - \phi(x))(-1)\, dx &= 0,\\
\int_0^T 2(f(x) - \phi(x))\left(-\cos\frac{2\pi x}{T}\right) dx &= 0,\\
\int_0^T 2(f(x) - \phi(x))\left(-\cos\frac{4\pi x}{T}\right) dx &= 0,\\
&\ \vdots\\
\int_0^T 2(f(x) - \phi(x))\left(-\cos\frac{2n\pi x}{T}\right) dx &= 0,
\end{aligned}$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

and

$$\int_0^T 2(f(x)-\phi(x))\left(-\sin\frac{2\pi x}{T}\right)dx = 0,$$

$$\int_0^T 2(f(x)-\phi(x))\left(-\sin\frac{4\pi x}{T}\right)dx = 0,$$

$$\vdots$$

$$\int_0^T 2(f(x)-\phi(x))\left(-\sin\frac{2n\pi x}{T}\right)dx = 0.$$

These equations can be rewritten as

$$\int_0^T f(x)\,dx = \int_0^T \phi(x)\,dx,$$

$$\int_0^T f(x)\cos\frac{2\pi x}{T}\,dx = \int_0^T \phi(x)\cos\frac{2\pi x}{T}\,dx,$$

$$\int_0^T f(x)\cos\frac{4\pi x}{T}\,dx = \int_0^T \phi(x)\cos\frac{4\pi x}{T}\,dx,$$

$$\vdots$$

$$\int_0^T f(x)\cos\frac{2n\pi x}{T}\,dx = \int_0^T \phi(x)\cos\frac{2n\pi x}{T}\,dx,$$

and

$$\int_0^T f(x)\sin\frac{2\pi x}{T}\,dx = \int_0^T \phi(x)\sin\frac{2\pi x}{T}\,dx,$$

$$\int_0^T f(x)\sin\frac{4\pi x}{T}\,dx = \int_0^T \phi(x)\sin\frac{4\pi x}{T}\,dx,$$

$$\vdots$$

$$\int_0^T f(x)\sin\frac{2n\pi x}{T}\,dx = \int_0^T \phi(x)\sin\frac{2n\pi x}{T}\,dx,$$

The Fourier coefficients a_k and b_k that we seek appear in ϕ, and we shall obtain them by calculating the integrals on the right (the ones involving ϕ)

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

in the equations above. Since

$$\phi(x) = a_0 + \sum_{k=1}^{n} a_k \cos \frac{2k\pi x}{T} + \sum_{k=1}^{n} b_k \sin \frac{2k\pi x}{T},$$

we have (for each $m = 0, 1, \ldots, n$)

$$\begin{aligned}\phi(x) \cos \frac{2m\pi x}{T} = a_0 \cos \frac{2m\pi x}{T} &+ \sum_{k=1}^{n} a_k \cos \frac{2k\pi x}{T} \cos \frac{2m\pi x}{T} \\ &+ \sum_{k=1}^{n} b_k \sin \frac{2k\pi x}{T} \cos \frac{2m\pi x}{T},\end{aligned}$$

and (for each $m = 1, \ldots, n$)

$$\phi(x) \sin \frac{2m\pi x}{T} = \sum_{k=1}^{n} a_k \cos \frac{2k\pi x}{T} \sin \frac{2m\pi x}{T} + \sum_{k=1}^{n} b_k \sin \frac{2k\pi x}{T} \sin \frac{2m\pi x}{T}.$$

Integrating products of sines and cosines

The integrals of the expressions on the left therefore reduce to sums of integrals of various products of sines and cosines. In each sum, only *one* term yields a nonzero integral. All the values are given below. The formulas are left for you to derive in the exercises, using integration formulas from the exercises in chapter 11.3. For integers k and m, we have

$$\int_0^T \sin \frac{2k\pi x}{T} \cos \frac{2m\pi x}{T}\, dx = 0 \qquad \text{for all } k \text{ and } m;$$

$$\int_0^T \sin \frac{2k\pi x}{T} \sin \frac{2m\pi x}{T}\, dx = \begin{cases} T/2 & \text{if } k = m, \\ 0 & \text{otherwise;} \end{cases}$$

$$\int_0^T \cos \frac{2k\pi x}{T} \cos \frac{2m\pi x}{T}\, dx = \begin{cases} T/2 & \text{if } k = m \neq 0, \\ T & \text{if } k = m = 0, \\ 0 & \text{otherwise.} \end{cases}$$

For $m = 0$ we have $\cos \frac{2m\pi x}{T} = \cos 0 = 1$, so

$$\int_0^T \phi(x) \cos \frac{2m\pi x}{T}\, dx = \int_0^T a_0\, dx = a_0 \cdot T;$$

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

it follows that

$$a_0 = \frac{1}{T}\int_0^T \phi(x)\,dx.$$

For each $m = 1, 2, \ldots, n$, we have, first of all,

$$\int_0^T \phi(x)\cos\frac{2m\pi x}{T}\,dx = \int_0^T a_m\cos^2\frac{2m\pi x}{T}\,dx = a_m\cdot T/2,$$

from which it follows that

$$a_m = \frac{2}{T}\int_0^T \phi(x)\cos\frac{2m\pi x}{T}\,dx.$$

Second, we have

$$\int_0^T \phi(x)\sin\frac{2m\pi x}{T}\,dx = \int_0^T b_m\sin^2\frac{2m\pi x}{T}\,dx = b_m\cdot T/2,$$

so

$$b_m = \frac{2}{T}\int_0^T \phi(x)\sin\frac{2m\pi x}{T}\,dx.$$

The derivation is complete.

Exercises

1. Use the formulas on page 716 in chapter 11.3 to derive the following equalities; k and m are integers.

a) $\displaystyle\int_0^T \sin\frac{2k\pi x}{T}\cos\frac{2m\pi x}{T}\,dx = 0$ for all k and m.

b) $\displaystyle\int_0^T \sin\frac{2k\pi x}{T}\sin\frac{2m\pi x}{T}\,dx = \begin{cases} T/2 & \text{if } k = m, \\ 0 & \text{otherwise.}\end{cases}$

c) $\displaystyle\int_0^T \cos\frac{2k\pi x}{T}\cos\frac{2m\pi x}{T}\,dx = \begin{cases} T/2 & \text{if } k = m \neq 0, \\ T & \text{if } k = m = 0, \\ 0 & \text{otherwise.}\end{cases}$

2. Show that in the Fourier series for the triangular function discussed in the text (example 1, page 824), all the coefficients of the sine terms really are 0.

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

3. Find the Fourier series for the following functions over the interval $[-\pi, \pi]$:

a) $f(x) = x$. [Ans. $2\sum_{k=1}^{\infty} \frac{(-1)^{n-1}\sin nx}{n}$]

b) $f(x) = \pi^2 - x^2$.

c) $f(x) = \begin{cases} 0 & \text{if } -\pi \le x \le 0, \\ x^2 & \text{if } 0 \le x \le \pi. \end{cases}$

4. In May's predator–prey model, find the first seven terms of the Fourier series for the predator species, $y(t)$. Use $T = 38.6$ days and the initial conditions $x = 7.75$ and $y = 2.38$, and in example 2 in the text (page 827).

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008

Copyright 1994, 2008 Five Colleges, Inc.
DVI file created at 17:56, 25 January 2008