with an initial temperature distribution $f(x)$, and we want to find a solution $u(x,t)$ of the heat equation with

$$u(x,0) = f(x)\,.$$

Both $f(x)$ and $u(x,t)$ are defined for $-\infty < x < \infty$, and there is no assumption of periodicity. Knowing the Fourier transform of the Gaussian is essential for the treatment we're about to give.

The idea is to take the Fourier transform of both sides of the heat equation with respect to x. The Fourier transform of the right-hand side of the equation, $\frac{1}{2}u_{xx}(x,t)$, is

$$\tfrac{1}{2}\mathcal{F}u_{xx}(s,t) = \tfrac{1}{2}(2\pi i s)^2\mathcal{F}u(s,t) = -2\pi^2 s^2\mathcal{F}u(s,t)\,,$$

from the derivative formula. Observe that the frequency variable (still using that terminology) s is now in the first slot of the transformed function and that the time variable t is just going along for the ride. For the left-hand side, $u_t(x,t)$, we do something different. We have

$$\begin{aligned}\mathcal{F}u_t(s,t) &= \int_{-\infty}^{\infty} u_t(x,t)e^{-2\pi i s x}\,dx \quad \text{(Fourier transform in } x\text{)}\\ &= \int_{-\infty}^{\infty} \frac{\partial}{\partial t}u(x,t)e^{-2\pi i s x}\,dx\\ &= \frac{\partial}{\partial t}\int_{-\infty}^{\infty} u(x,t)e^{-2\pi i s x}\,dx = \frac{\partial}{\partial t}\mathcal{F}u(s,t).\end{aligned}$$

Thus taking the Fourier transform with respect to x of both sides of the equation

$$u_t = \tfrac{1}{2}u_{xx}$$

leads to

$$\frac{\partial \mathcal{F}u(s,t)}{\partial t} = -2\pi^2 s^2\mathcal{F}u(s,t)\,.$$

This is a differential equation in t — an ordinary differential equation, despite the partial derivative symbol — and we can solve it:

$$\mathcal{F}u(s,t) = \mathcal{F}u(s,0)e^{-2\pi^2 s^2 t}\,.$$

What is the initial condition $\mathcal{F}u(s,0)$?

$$\begin{aligned}\mathcal{F}u(s,0) &= \int_{-\infty}^{\infty} u(x,0)e^{-2\pi i s x}\,dx\\ &= \int_{-\infty}^{\infty} f(x)e^{-2\pi i s x}\,dx = \mathcal{F}f(s)\,.\end{aligned}$$

Putting it all together,

$$\mathcal{F}u(s,t) = \mathcal{F}f(s)e^{-2\pi^2 s^2 t}\,.$$

We recognize (we are good) that the exponential factor on the right-hand side is the Fourier transform of the Gaussian,

$$g(x,t) = \frac{1}{\sqrt{2\pi t}}e^{-x^2/2t}\,.$$

We then have a product of two Fourier transforms,

$$\mathcal{F}u(s,t) = \mathcal{F}g(s,t)\,\mathcal{F}f(s)$$

and we invert this to obtain a convolution in the spatial domain:

$$u(x,t) = g(x,t) * f(x) = \left(\frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}\right) * f(x) \quad \text{(convolution in } x)$$

or, written out,

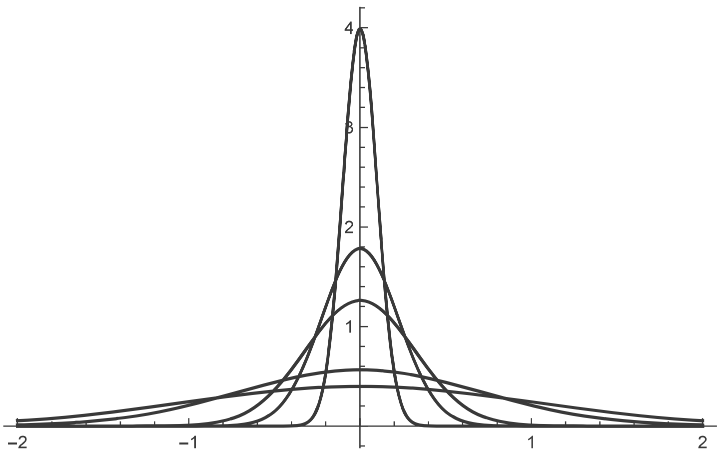
$$u(x,t) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-(x-y)^2/2t} f(y)\, dy\,.$$

It's reasonable to believe that the temperature $u(x,t)$ of the rod at a point x at a time $t > 0$ is some kind of averaged, smoothed version of the initial temperature $f(x) = u(x,0)$. That's convolution at work.

The function

$$g(x,t) = \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}$$

is called the *heat kernel* (or Green's function or the fundamental solution) for the heat equation for the infinite rod. Here are plots of $g(x,t)$, as a function of x, for $t = 1,\ 0.5,\ 0.1,\ 0.05,\ 0.01$; the more peaked curves corresponding to the smaller values of t.



You can see that the curves are becoming more concentrated near $x = 0$. However, they are doing so in a way that keeps the area under each curve 1, for

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t}\, dx &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-\pi u^2} \sqrt{2\pi t}\, du \\ &\qquad \text{(making the substitution } u = x/\sqrt{2\pi t}) \\ &= \int_{-\infty}^{\infty} e^{-\pi u^2}\, du = 1\,.\end{aligned}$$

We'll see later that the $g(x,t)$ serve as an approximation to the δ-function as $t \to 0$.

You might ask at this point: Didn't we already solve the heat equation? Is what we did in Chapter 1 related to what we did just now? Indeed we did and indeed they are: see Section 3.5.4.

3.5.3. More on diffusion; back to the cable. When last we left the problem of signals propagating down a cable, William Thomson had appealed to the heat equation to study the delay in a signal sent along a long, undersea telegraph cable. The physical intuition, as of the mid 19th century, was that charge diffused along the cable. To reconstruct part of Thomson's solution (essentially) we must begin with a slightly different setup. The equation is the same':

$$u_t = \tfrac{1}{2} u_{xx}\,,$$

so we're choosing constants as above and not explicitly incorporating physical parameters such as resistance per length, capacitance per length, etc. More importantly, the initial and boundary conditions are different.

We consider a *semi-infinite* rod, having one end (at $x = 0$) but effectively extending infinitely in the positive x-direction. Instead of an initial distribution of temperature along the entire rod, we suppose there is a source of heat (or voltage) $f(t)$ at the end $x = 0$. Thus we have the initial condition

$$u(0,t) = f(t)\,.$$

We also suppose that

$$u(x,0) = 0\,,$$

meaning that at $t = 0$ there's no temperature (or charge) in the rod. We also assume that $u(x,t)$ and its derivatives tend to zero as $x \to \infty$. Finally, we set

$$u(x,t) = 0 \quad \text{for } x < 0$$

so that we can regard $u(x,t)$ as defined for all x. We want a solution that expresses $u(x,t)$, the temperature (or voltage) at a position $x > 0$ and time $t > 0$ in terms of the initial temperature (or voltage) $f(t)$ at the endpoint $x = 0$.

The analysis of this is *really* involved. It's quite a striking formula that works out in the end, but, be warned, the end is a way off. Proceed only if interested.

First take the Fourier transform of $u(x,t)$ with respect to x (the notation $\hat{u}$ seems more natural here):

$$\hat{u}(s,t) = \int_{-\infty}^{\infty} e^{-2\pi isx} u(x,t)\,dx\,.$$

Then, using the heat equation,

$$\frac{\partial}{\partial t}\hat{u}(s,t) = \int_{-\infty}^{\infty} e^{-2\pi isx} \frac{\partial}{\partial t} u(x,t)\,dx = \int_{-\infty}^{\infty} e^{-2\pi isx} \frac{1}{2}\frac{\partial^2}{\partial x^2} u(x,t)\,dx\,.$$

We need to integrate only from 0 to ∞ since $u(x,t)$ is identically 0 for $x<0$. We integrate by parts once:

$$\begin{aligned}
&\int_0^\infty e^{-2\pi isx}\frac{1}{2}\frac{\partial^2}{\partial x^2}u(x,t)\,dx\\
&\quad=\frac{1}{2}\left(\left[e^{-2\pi isx}\frac{\partial}{\partial x}u(x,t)\right]_{x=0}^{x=\infty}+2\pi is\int_0^\infty\frac{\partial}{\partial x}u(x,t)\,e^{-2\pi isx}\,dx\right)\\
&\quad=-\tfrac{1}{2}u_x(0,t)+\pi is\int_0^\infty\frac{\partial}{\partial x}u(x,t)\,e^{-2\pi isx}\,dx\,,
\end{aligned}$$

taking the boundary conditions on $u(x,t)$ into account. Now integrate by parts a second time:

$$\begin{aligned}
\int_0^\infty\frac{\partial}{\partial x}u(x,t)\,e^{-2\pi isx}\,dx&=\left[e^{-2\pi isx}\,u(x,t)\right]_{x=0}^{x=\infty}+2\pi is\int_0^\infty e^{-2\pi ist}\,u(x,t)\,dx\\
&=-u(0,t)+2\pi is\int_0^\infty e^{-2\pi ist}\,u(x,t)\,dx\\
&=-f(t)+2\pi is\int_{-\infty}^\infty e^{-2\pi ist}\,u(x,t)\,dx
\end{aligned}$$

(we drop the bottom limit back to $-\infty$ to bring back the Fourier transform)

$$=-f(t)+2\pi is\,\hat{u}(s,t).$$

Putting these calculations together yields

$$\frac{\partial}{\partial t}\hat{u}(s,t)=-\tfrac{1}{2}u_x(0,t)-\pi isf(t)-2\pi^2s^2\hat{u}(s,t)\,.$$

This is a linear, first-order, ordinary differential equation (in t) for $\hat{u}$. It's of the general type

$$y'(t)+P(t)y(t)=Q(t)\,.$$

If you cast your mind back to courses from the dim and distant past, you will recall that to solve such an equation you multiply both sides by the integrating factor

$$e^{\int_0^t P(\tau)\,d\tau}\,,$$

which produces

$$\left(y(t)e^{\int_0^t P(\tau)\,d\tau}\right)'=e^{\int_0^t P(\tau)\,d\tau}Q(t)\,.$$

From here you get $y(t)$ by direct integration. For our particular application we have

$P(t)=2\pi^2s^2$ (that's a constant as far as we're concerned because there's no t),

$Q(t)=-\tfrac{1}{2}u_x(0,t)-\pi isf(t)$.

The integrating factor is $e^{2\pi^2s^2t}$ and we're to solve

$$(e^{2\pi^2s^2t}\hat{u}(t))'=e^{2\pi^2s^2t}\left(-\tfrac{1}{2}u_x(0,t)-\pi isf(t)\right)\,.$$

I want to carry this out so you don't miss anything.

Write τ for t and integrate both sides from 0 to t with respect to τ:

$$e^{2\pi^2s^2t}\hat{u}(s,t)-\hat{u}(s,0)=\int_0^t e^{2\pi^2s^2\tau}\left(-\tfrac{1}{2}u_x(0,\tau)-\pi isf(\tau)\right)\,d\tau\,.$$

But $\hat{u}(s,0) = 0$ since $u(x,0)$ is identically 0, so

$$\begin{aligned}\hat{u}(s,t) &= e^{-2\pi^2 s^2 t}\int_0^t e^{2\pi^2 s^2 \tau}\left(-\tfrac{1}{2}u_x(0,\tau) - \pi i s f(\tau)\right)\,d\tau\\ &= \int_0^t e^{-2\pi^2 s^2(t-\tau)}\left(-\tfrac{1}{2}u_x(0,\tau) - \pi i s f(\tau)\right)\,d\tau.\end{aligned}$$

We need to take the inverse transform of this to get $u(x,t)$. Be not afraid:

$$\begin{aligned}u(x,t) &= \int_{-\infty}^{\infty} e^{2\pi i s x}\hat{u}(s,t)\,ds\\ &= \int_{-\infty}^{\infty} e^{2\pi i s x}\Big(\int_0^t e^{-2\pi^2 s^2(t-\tau)}\left(-\tfrac{1}{2}u_x(0,\tau) - \pi i s f(\tau)\right)\,d\tau\Big)\,ds\\ &= \int_0^t\int_{-\infty}^{\infty} e^{2\pi i s x}e^{-2\pi^2 s^2(t-\tau)}\left(-\tfrac{1}{2}u_x(0,\tau) - \pi i s f(\tau)\right)\,ds\,d\tau\,.\end{aligned}$$

Appearances to the contrary, this is not hopeless. Let's pull out the inner integral for further examination:

$$\begin{aligned}&\int_{-\infty}^{\infty} e^{2\pi i s x}(e^{-2\pi^2 s^2(t-\tau)}\left(-\tfrac{1}{2}u_x(0,\tau) - \pi i s f(\tau)\right))\,ds\\ &\quad= -\tfrac{1}{2}u_x(0,\tau)\int_{-\infty}^{\infty} e^{2\pi i s x}e^{-2\pi^2 s^2(t-\tau)}\,ds - \pi i f(\tau)\int_{-\infty}^{\infty} e^{2\pi i s x}\,s\,e^{-2\pi^2 s^2(t-\tau)}\,ds\,.\end{aligned}$$

The first integral is the inverse Fourier transform of a Gaussian; we want to find $\mathcal{F}^{-1}\big(e^{-2\pi s^2(t-\tau)}\big)$. Recall the formulas

$$\mathcal{F}\Big(\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2\sigma^2}\Big) = e^{-2\pi^2\sigma^2 s^2}\,, \quad \mathcal{F}(e^{-x^2/2\sigma^2}) = \sigma\sqrt{2\pi}\,e^{-2\pi^2\sigma^2 s^2}.$$

Apply this with

$$\sigma = \frac{1}{2\pi\sqrt{(t-\tau)}}\,.$$

Then, using duality and evenness of the Gaussian, we have

$$\int_{-\infty}^{\infty} e^{2\pi i s x}e^{-2\pi s^2(t-\tau)}\,ds = \mathcal{F}^{-1}\big(e^{-2\pi s^2(t-\tau)}\big) = \frac{e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)}}\,.$$

In the second integral we want to find $\mathcal{F}^{-1}(s\,e^{-2\pi^2 s^2(t-\tau)})$. For this, note that

$$s\,e^{-2\pi^2 s^2(t-\tau)} = -\frac{1}{4\pi^2(t-\tau)}\frac{d}{ds}e^{-2\pi^2 s^2(t-\tau)}$$

and hence

$$\begin{aligned}\int_{-\infty}^{\infty} e^{2\pi i s x}\,s\,e^{-2\pi^2 s^2(t-\tau)}\,ds &= \mathcal{F}^{-1}\Big(-\frac{1}{4\pi^2(t-\tau)}\frac{d}{ds}e^{-2\pi^2 s^2(t-\tau)}\Big)\\ &= -\frac{1}{4\pi^2(t-\tau)}\mathcal{F}^{-1}\Big(\frac{d}{ds}e^{-2\pi^2 s^2(t-\tau)}\Big)\,.\end{aligned}$$

We know how to take the inverse Fourier transform of a derivative, or rather we know how to take the (forward) Fourier transform, and that's all we need to do by another application of duality. Let me remind you: we use, for a general function f,

$$\mathcal{F}^{-1}f' = (\mathcal{F}f')^- = (2\pi i x\mathcal{F}f)^- = -2\pi i x(\mathcal{F}f)^- = -2\pi i x\mathcal{F}^{-1}f\,.$$

Apply this to

$$\begin{aligned}\mathcal{F}^{-1}\left(\frac{d}{ds}e^{-2\pi^2 s^2(t-\tau)}\right) &= -2\pi i x\mathcal{F}^{-1}\big(e^{-2\pi^2 s^2(t-\tau)}\big)\\ &= -2\pi i x\frac{1}{\sqrt{2\pi(t-\tau)}}\,e^{-x^2/2(t-\tau)}\\ &\qquad\text{(from our earlier calculation, fortunately).}\end{aligned}$$

Then

$$-\frac{1}{4\pi^2(t-\tau)}\mathcal{F}^{-1}\Big(\frac{d}{ds}e^{-2\pi^2 s^2(t-\tau)}\Big) = \frac{2\pi i x}{4\pi^2(t-\tau)}\frac{e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)}} = \frac{i}{2\pi}\frac{x\,e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)^3}}\,.$$

That is,

$$\mathcal{F}^{-1}\Big(s\,e^{-2\pi^2 s^2(t-\tau)}\Big) = \frac{i}{2\pi}\frac{x\,e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)^3}}\,.$$

Finally getting back to the expression for $u(x,t)$, we can combine what we've calculated for the inverse Fourier transforms and write

$$\begin{aligned}u(x,t) &= -\tfrac{1}{2}\int_0^t u_x(0,\tau)\mathcal{F}^{-1}(e^{-2\pi s^2(t-\tau)})\,d\tau \;-\; \pi i\int_0^t f(\tau)\mathcal{F}^{-1}(s\,e^{-2\pi^2 s^2(t-\tau)})\,d\tau\\ &= -\tfrac{1}{2}\int_0^t u_x(0,\tau)\frac{e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)}}\,d\tau \;+\; \tfrac{1}{2}\int_0^t f(\tau)\,\frac{x\,e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)^3}}\,d\tau.\end{aligned}$$

We're almost there. We'd like to eliminate $u_x(0,\tau)$ from this formula and express $u(x,t)$ in terms of $f(t)$ only. This can be accomplished by a very clever, and I'd say highly nonobvious, observation. We know that $u(x,t)$ is zero for $x<0$; we have defined it to be so. Hence the integral expression for $u(x,t)$ is zero for $x<0$. Because of the evenness and oddness *in* x of the two integrands this has a consequence for the values of the integrals when x is positive. (The first integrand is even in x and the second is odd in x.) In fact, the integrals are equal!

Let me explain what happens in a general situation, stripped down, so you can see the idea. Suppose we have

$$\Phi(x,t) = \int_0^t \phi(x,\tau)\,d\tau + \int_0^t \psi(x,\tau)\,d\tau$$

where we know that: $\Phi(x,t)$ is zero for $x<0$; $\phi(x,\tau)$ is even in x; $\psi(x,\tau)$ is odd in x. Take $a>0$. Then $\Phi(-a,\tau)=0$; hence using the evenness of $\phi(x,\tau)$ and the oddness of $\psi(x,\tau)$,

$$0 = \int_0^t \phi(-a,\tau)\,d\tau + \int_0^t \psi(-a,\tau)\,d\tau = \int_0^t \phi(a,\tau)\,d\tau - \int_0^t \psi(a,\tau)\,d\tau\,.$$

We conclude that for all $a > 0$,

$$\int_0^t \phi(a,\tau) = \int_0^t \psi(a,\tau)\,d\tau\,,$$

and hence for $x > 0$ (writing x for a)

$$\begin{aligned}\Phi(x,t) &= \int_0^t \phi(x,\tau)\,d\tau + \int_0^t \psi(x,\tau)\,d\tau \\ &= 2\int_0^t \psi(x,\tau)\,d\tau = 2\int_0^t \phi(x,\tau)\,d\tau \quad \text{(either } \phi \text{ or } \psi \text{ could be used).}\end{aligned}$$

We apply this in our situation with

$$\phi(x,\tau) = -\tfrac{1}{2}u_x(0,\tau)\frac{e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)}}\,, \quad \psi(x,\tau) = \tfrac{1}{2}f(\tau)\,\frac{x\,e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)^3}}\,.$$

The result is that we can eliminate the integral with the $u_x(0,\tau)$ and write the solution — the final solution — as

$$u(x,t) = \int_0^t f(\tau)\,\frac{x\,e^{-x^2/2(t-\tau)}}{\sqrt{2\pi(t-\tau)^3}}\,d\tau\,.$$

This form of the solution was the one given by Stokes. He wrote to Thomson:

> In working out myself various forms of the solution of the equation $dv/dt = d^2v/dx^2$ [*Note*: He puts a 1 on the right-hand side instead of a 1/2, and he uses an ordinary "d".] under the condition $v = 0$ when $t = 0$ from $x = 0$ to $x = \infty$; $v = f(t)$ when $x = 0$ from $t = 0$ to $t = \infty$ I found the solution ... was ...
>
> $$v(x,t) = \frac{x}{2\sqrt{\pi}}\int_0^t (t-t')^{-3/2}e^{-x^2/4(t-t')}f(t')\,dt'\,.$$

3.5.4. Didn't we already solve the heat equation? Our first application of Fourier series (*the* first application of Fourier series) was to solve the heat equation. Let's recall the setup and the form of the solution. We heat a circle, which we consider to be the interval $0 \le x \le 1$ with the endpoints identified. If the initial distribution of temperature is the function $f(x)$, then the temperature $u(x,t)$ at a point x at time $t > 0$ is given by

$$u(x,t) = \int_0^1 g(x-y,t)f(y)\,dy\,,$$

where

$$g(x,t) = \sum_{n=-\infty}^{\infty} e^{-2\pi^2n^2t}e^{2\pi inx}\,.$$

So, explicitly,

$$u(x,t) = g(x,t) * f(x) = \int_0^1 \sum_{n=-\infty}^{\infty} e^{-2\pi^2n^2t}e^{2\pi in(x-y)}f(y)\,dy\,.$$

This was our first encounter with convolution, a convolution in the spatial variable but with limits of integration just from 0 to 1. Here $f(x)$, $g(x,t)$, and $u(x,t)$ are periodic of period 1 in x.

How does this compare to what we did for the rod? If we imagine initially heating up a circle as heating up an infinite rod by a *periodic* function $f(x)$, then shouldn't we be able to express the temperature $u(x,t)$ for the circle as we did for the rod? We will show that the solution for a circle does have the *same form* as the solution for the infinite rod by means of the remarkable identity:

$$\sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t} = \sqrt{2\pi t} \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n x}$$

Needless to say, this is *not* obvious.

As an aside, for general interest, a special case of this identity is particularly famous. The *Jacobi theta function*[10] is defined by

$$\vartheta(t) = \sum_{n=-\infty}^{\infty} e^{-\pi n^2 t},$$

for $t > 0$. It comes up in surprisingly diverse pure and applied fields, including number theory and statistical mechanics! (In the latter it is used to study partition functions.) Jacobi's identity is

$$\vartheta(t) = \frac{1}{\sqrt{t}} \vartheta\left(\frac{1}{t}\right).$$

It follows from the identity above, with $x = 0$ and replacing t by $1/2\pi t$.

We'll show later why the general identity holds. But first, assuming that it does, let's work with the solution of the heat equation for a circle and see what we get. Applying the identity to Green's function $g(x,t)$ for heat flow on the circle we have

$$g(x,t) = \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n x} = \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t}.$$

[10] Named after C. Jacobi. He contributed a staggering amount of work to mathematics, many ideas finding applications in very diverse areas. He is perhaps second only to Euler for pure analytic skill.

Regard the initial distribution of heat $f(x)$ as being defined on all of $\mathbb{R}$ and having period 1. Then

$$\begin{aligned}
u(x,t) &= \int_0^1 \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n(x-y)} f(y)\,dy \\
&= \frac{1}{\sqrt{2\pi t}} \int_0^1 \sum_{n=-\infty}^{\infty} e^{-(x-y-n)^2/2t} f(y)\,dy \\
&\qquad \text{(using the identity for Green's function)} \\
&= \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} \int_0^1 e^{-(x-y-n)^2/2t} f(y)\,dy \\
&= \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} \int_n^{n+1} e^{-(x-u)^2/2t} f(u-n)\,du \quad \text{(substituting } u = y+n\text{)} \\
&= \frac{1}{\sqrt{2\pi t}} \sum_{n=-\infty}^{\infty} \int_n^{n+1} e^{-(x-u)^2/2t} f(u)\,du \quad \text{(using that } f \text{ has period 1)} \\
&= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-(x-u)^2/2t} f(u)\,du\,.
\end{aligned}$$

Voilà! We are back to the solution of the heat equation on the line.

Incidentally, since the problem was originally formulated for heating a circle, the function $u(x,t)$ is periodic in x. Can we see that from this form of the solution? Yes, for

$$\begin{aligned}
u(x+1,t) &= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-(x+1-u)^2/2t} f(u)\,du \\
&= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-(x-w)^2/2t} f(w+1)\,dw \quad \text{(substituting } w = u-1\text{)} \\
&= \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} e^{-(x-w)^2/2t} f(w)\,dw \quad \text{(using the periodicity of } f(x)\text{)} \\
&= u(x,t)\,.
\end{aligned}$$

Now let's derive the identity

$$\sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t} = \sqrt{2\pi t} \sum_{n=-\infty}^{\infty} e^{-2\pi^2 n^2 t} e^{2\pi i n x}\,.$$

This is a great combination of many of the things we've developed to this point, and it will come up again.[11] Consider the left-hand side as a function of x, say

$$h(x) = \sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t}\,.$$

[11] It's worth your effort to go through this. The calculations in this special case will come up more generally when we do the Poisson summation formula. That formula is the basis of the sampling theorem.

This is a periodic function of period 1; it's the *periodization* of the Gaussian $e^{-x^2/2t}$. (It's even not hard to show that the series converges, etc., but we won't go through that.) What are its Fourier coefficients? We can calculate them:

$$\begin{aligned}
\hat{h}(k) &= \int_0^1 h(x)e^{-2\pi ikx}\,dx \\
&= \int_0^1 \Big(\sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t}\Big)e^{-2\pi ikx}\,dx \\
&= \sum_{n=-\infty}^{\infty} \int_0^1 e^{-(x-n)^2/2t}e^{-2\pi ikx}\,dx \\
&= \sum_{n=-\infty}^{\infty} \int_{-n}^{-n+1} e^{-u^2/2t}e^{-2\pi iku}\,du \\
&\qquad \text{(substituting } u = x-n \text{ and using periodicity of } e^{-2\pi ikx}\text{)} \\
&= \int_{-\infty}^{\infty} e^{-u^2/2t}e^{-2\pi iku}\,du\,.
\end{aligned}$$

But this last integral is exactly the Fourier transform of the Gaussian $e^{-x^2/2t}$ at $s=k$. We know how to do that. The answer is $\sqrt{2\pi t}\,e^{-2\pi^2k^2t}$.

We have shown that the Fourier coefficients of $h(x)$ are

$$\hat{h}(k) = \sqrt{2\pi t}\,e^{-2\pi^2k^2t}\,.$$

Since the function is equal to its Fourier series (really equal here because all the series converge), we conclude that

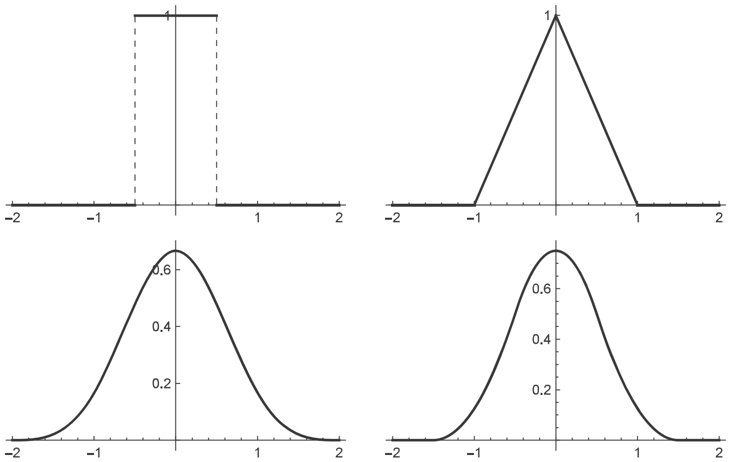
$$\begin{aligned}
h(x) &= \sum_{n=-\infty}^{\infty} e^{-(x-n)^2/2t} \\
&= \sum_{n=-\infty}^{\infty} \hat{h}(n)e^{2\pi inx} = \sqrt{2\pi t}\sum_{n=-\infty}^{\infty} e^{-2\pi^2n^2t}e^{2\pi inx}\,,
\end{aligned}$$

and there's the identity we wanted to prove.

3.6. Convolution in Action III: The Central Limit Theorem

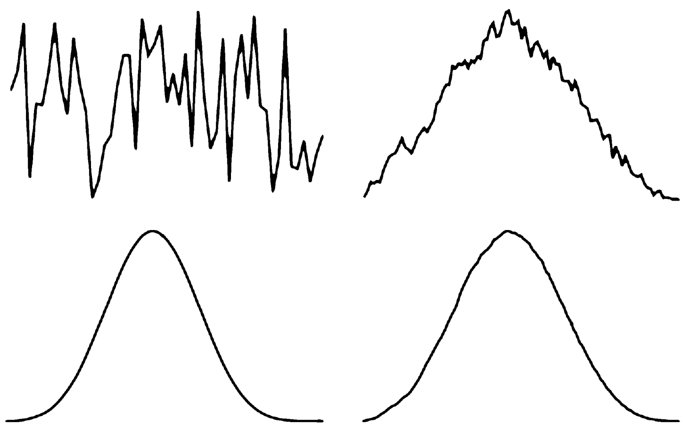
Several times we've met the idea that convolution is a smoothing operation. Let me begin with some graphical examples of this, convolving a discontinuous or rough function repeatedly with itself.

In a problem you computed, by hand, the convolution of the rectangle function Π with itself a few times. Here are plots of this, up to $\Pi * \Pi * \Pi * \Pi$, displayed clockwise from Π.



Not only are the convolutions becoming smoother, the unmistakable shape of a Gaussian is emerging. Is this a coincidence, based on the particularly simple nature of the function Π, or is something more going on?

Here is a plot of, literally, a random function $f(x)$ — the values $f(x)$ are just randomly chosen numbers between 0 and 1 — and its self-convolutions up to the four-fold convolution $f * f * f * f$, again displayed clockwise.



From seeming chaos, again we see a Gaussian emerging. This is the spookiest thing I know.

The object of this section is to explain this phenomenon, to give substance to the famous quotation:

> Everyone believes in the normal approximation, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact.
>
> G. Lippman, French Physicist, 1845–1921

The "normal approximation" (or normal distribution) is the Gaussian. The "mathematical theorem" here is the *Central Limit Theorem*. To understand the theorem and to appreciate the "experimental fact," we have to develop some ideas from probability.

3.6.1. Random variables.

In whatever field of science or engineering you pursue, you will certainly use probabilistic ideas, and you will certainly use the Gaussian. I'm going under the assumption that you probably already know some probability and probably some statistics, too, even if only in a casual way. For our present work we don't need complete generality based on hard won, abstract definitions. Just the basics will do. Every mature field settles on its fundamental notions (or tries to), sometimes clear at the outset and sometimes slow to emerge. One of the fundamental notions for probability, maybe *the* fundamental notion, is the *random variable*. It was slow to emerge, or at least a suitable definition was slow to emerge. We're going to be informal, even cavalier about this: a random variable is a number you don't know yet.[12] By that I mean that a random variable, or rather *its value*, is the numerical result of some process, like a measurement or the result of an experiment. The assumption is that you can make the measurement, you can perform the experiment, but until you do you don't know the value of the random variable. It's called "random" because a particular object to be measured is thought of as being drawn "at random" from a collection of all such objects. For example:

Random variable	Value of random variable
Height of people in US population	Height of a particular person
Length of pins produced	Length of a particular pin
Momentum of atoms in a gas	Momentum of a particular atom
Resistance of resistors off a production line	Resistance of a particular resistor
Toss of coin	0 or 1 (head or tail)
Roll of dice	The numbers that come up

A common notation is to write X for the name of the random variable and x for its value. If you thus think that a random variable X is just a function, you're right, but deciding what the *domain* of such a function should be and what mathematical structure to require of both the domain and the function demands the kind of care that we don't want to get into. As I said, this was a long time in coming. Consider, for example, Mark Kac's comment: "Independent random variables were to me (and others, including my teacher Steinhaus) shadowy and not really well-defined objects." Kac was one of the most eminent probabilists of the 20th century.

3.6.2. Probability distributions and probability density functions.

"Random variable" is the fundamental notion, but not the fundamental object of study. For a given random variable, what we're most interested in is how its values are distributed. For this it's helpful to distinguish between two types of random variables.

[12]I think this phrase to describe a random variable is due to Sam Savage in the Management Science and Engineering department at Stanford.

- A random variable is *discrete* if its values are among only a discrete set of possibilities, for us, a finite set.
  - For example, "roll the die" is a discrete random variable with values 1, 2, 3, 4, 5, or 6. "Toss the coin" is a discrete random variable with values "head" and "tail," or 0 and 1. (A random variable with values 0 and 1 is the basic random variable in coding and information theory.)
- A random variable is *continuous* if its values do not form a discrete set, for us typically filling up one or more intervals of real numbers.
  - For example, "length of a pin" is a continuous random variable since, in theory, the length of a pin can vary continuously.

For a discrete random variable we are used to the idea of displaying the distribution of values as a *histogram*. We set up bins, each bin corresponding to each of the possible values (or sometimes a range of values), we run the random process however many times we please, and for each bin we draw a bar with height indicating the *percentage* that the particular value occurs among all actual outcomes of the runs.[13] Since we plot percentages, or fractions, the total area of the histogram is 100%, or just 1.

A series of runs of the same experiment or of the same measurement will produce histograms of varying shapes.[14] We often expect some kind of limiting shape as we increase the number of runs, or we may *suppose* that the ideal distribution has some shape and then compare the actual data from a series of runs to the ideal, theoretical answer.

- The theoretical histogram is called the *probability distribution*.
- The function that describes the histogram (the shape of the distribution) is called the *probability density function*, or *pdf*, of the random variable.

Is there a difference between the probability distribution and the probability density function? No, not really — it's like distinguishing between the graph of a function and the function. Both terms are in common use, more or less interchangeably.

- The *probability* that any particular value comes up is the area of its bin in the probability distribution, which is therefore a number between 0 and 1.

If the random variable is called X and the value we're interested in is x, we write this as

$$\text{Prob}(X = x) = \text{area of the bin over } x\,.$$

Also

$$\text{Prob}(a \leq X \leq b) = \text{areas of the bins from } a \text{ to } b\,.$$

Thus probability is the percentage of the occurrence of a particular outcome, or range of outcomes, among all possible outcomes. We *must* base the definition of probability on what we presume or assume is the distribution function for a given

[13] I have gotten into heated arguments with physicist friends who insist on plotting frequencies of values rather than percentages. Fools.

[14] A run is like: "Do the experiment 10 times and make a histogram of your results for those 10 trials." A series of runs is like: "Do your run of 10 times, again. And again."

random variable. A statement about probabilities for a run of experiments is then a statement about long term trends, thought of as an approximation to the ideal distribution.[15]

One can also introduce probability distributions and probability density functions for continuous random variables. You can think of this — in fact you probably should think of this — as a continuous version of a probability histogram. It's a tricky business, however, to take a limit of the distributions for a discrete random variable, which have bins of a definite size, to produce a distribution for a continuous random variable, imagining the latter as having infinitely many infinitesimal bins.

It's easiest, and best, to define the distribution for a continuous random variable directly. In fact, we're used to this. It's another example of turning the solution of a problem into a definition. Years of bitter experience taught probabilists what properties they needed of probability distributions to make their theory work, to prove their theorems. Not so many. They started over with the following definition and haven't looked back:

- A *probability density function* is a nonnegative function $p(x)$ with area 1; i.e.,

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

Remember, x is the measured value of some experiment. By convention, we take x to go from $-\infty$ to ∞ so we don't constantly have to say how far the values extend.

In this context, over the years I have been taught to say the phrase: "Let X be a random variable *drawn* from a distribution $p(x)$ ..." and to write $X \sim p(x)$.

Here's one quick and important property of pdfs:

- If $p(x)$ is a pdf and $a > 0$, then $ap(ax)$ is also a pdf.

To show this, we have to check that the integral of $ap(ax)$ is 1. But

$$\int_{-\infty}^{\infty} ap(ax)\,dx = \int_{-\infty}^{\infty} ap(u)\frac{1}{a}\,du = \int_{-\infty}^{\infty} p(u)\,du = 1\,,$$

making the change of variable $u = ax$. We'll soon see this property in action.

- We think of a pdf as being associated with a random variable X whose values are x, and we write p_X if we want to emphasize this. The (probability) distribution of X is the graph of p_X, but, again, the terms *probability density function* and *probability distribution* are used interchangeably.
- *Probability* is defined by

$$\begin{aligned}\text{Prob}(X \le a) &= \text{area under the curve for } x \le a\\ &= \int_{-\infty}^{a} p_X(x)\,dx.\end{aligned}$$

[15] We are in dangerous territory here. Look up "Bayesian."

Also,

$$\text{Prob}(a \le X \le b) = \int_a^b p_X(x)\,dx\,.$$

For continuous random variables it really only makes sense to talk about the probability of a range of values occurring, not the probability of the occurrence of a single value. Think of the pdf as describing a limit of a (discrete) histogram: if the bins are becoming infinitely thin, what kind of event could land in an infinitely thin bin?[16]

Finally, for variable t, say, we can view

$$P(t) = \int_{-\infty}^{t} p(x)\,dx$$

as the probability function. It's also called the *cumulative probability* or the *cumulative density function*.[17] We then have

$$\text{Prob}(X \le t) = P(t)$$

and

$$\text{Prob}(a \le X \le b) = P(b) - P(a)\,.$$

According to the Fundamental Theorem of Calculus we can recover the probability density function from $P(t)$ by differentiation:

$$\frac{d}{dt}P(t) = p(t)\,.$$

In short, to know $p(t)$ is to know $P(t)$ and vice versa. You might not think this news is of any particular practical importance, but you're about to see that it is.

3.6.3. Mean, variance, and standard deviation. Suppose X is a random variable with pdf $p(x)$. The x's are the values assumed by X, so the *mean* μ of X is the weighted average of these values, weighted according to p. That is,

$$\mu(X) = \int_{-\infty}^{\infty} xp(x)\,dx\,.$$

The mean is also called the *expected value*, usually written $\mathbb{E}(X)$. "Expected value" is maybe more in keeping with the value of a random variable being determined by a measurement, as in, "What number were you expecting?"

Be careful here — the mean of X is *not* the average value of the function $p(x)$. Also, it might be that $\mu(X) = \infty$, i.e., that the integral of $xp(x)$ does not converge. This has to be checked for any particular example.

If $\mu(X) < \infty$, then we can always subtract off the mean to assume that X has mean zero. Here's what this means, no pun intended. In fact, let's do something slightly more general. What do we mean by $X - a$, when X is a random variable

[16] There's also the familiar integral identity

$$\int_a^a p_X(x)\,dx = 0$$

to contend with. In this context we would interpret this as saying that $\text{Prob}(X = a) = 0$.

[17] Cumulative density function is the preferred term because it allows for a three letter acronym: *cdf*.

and a is a constant? Nothing deep — you do the experiment to get a value of X (X is a number you don't know yet) and then you subtract a from it. What is the pdf of $X - a$? To figure that out, we have

$$\begin{aligned}\text{Prob}(X - a \leq t) &= \text{Prob}(X \leq t + a) \\ &= \int_{-\infty}^{t+a} p(x)\, dx \\ &= \int_{-\infty}^{t} p(u + a)\, du \quad \text{(substituting } u = x - a\text{)}.\end{aligned}$$

This identifies the pdf of $X - a$ as $p(x + a)$, the shifted pdf of X.[18]

Next, what is the mean of $X - a$? It must be $\mu(X) - a$. (Common sense, please.) Let's check this now knowing what pdf to integrate:

$$\begin{aligned}\mu(X - a) &= \int_{-\infty}^{\infty} xp(x + a)\, dx \\ &= \int_{-\infty}^{\infty} (u - a)p(u)\, du \quad \text{(substituting } u = x + a\text{)} \\ &= \int_{-\infty}^{\infty} up(u)\, du - a \int_{-\infty}^{\infty} p(u)\, du = \mu(X) - a\,.\end{aligned}$$

Note that translating the pdf $p(x)$ to $p(x+a)$ does nothing to the shape, or areas, of the distribution, hence does nothing to calculating any probabilities based on $p(x)$. As promised, the mean is $\mu(X) - a$. We are also happy to be certain now that "subtracting off the mean," as in $X - \mu(X)$, really does result in a random variable with mean 0. This normalization is often a convenient one to make in deriving formulas.

Continue to suppose that the mean $\mu(X)$ is finite. The *variance* $\sigma^2(X)$ is a measure of the amount that the values of the random variable deviate from the mean, *on average*, i.e., as weighted by the pdf $p(x)$. Since some values are above the mean and some are below, we weight the *square* of the differences, $(x - \mu(X))^2$, by $p(x)$ and define

$$\sigma^2(X) = \int_{-\infty}^{\infty} (x - \mu(X))^2 p(x)\, dx\,.$$

If we have normalized so that the mean is zero, this becomes simply

$$\sigma^2(X) = \int_{-\infty}^{\infty} x^2 p(x)\, dx\,.$$

The *standard deviation* is $\sigma(X)$, the square root of the variance. Even if the mean is finite, it might be that $\sigma^2(X)$ is infinite. This, too, has to be checked for any particular example.

We've just seen that we can normalize the mean of a random variable to be 0. Assuming that the variance is finite, can we normalize it in some helpful way?

[18]This is an illustration of the practical importance of going *from* the probability function *to* the pdf. We identified the pdf by knowing the probability function. This won't be the last time we do this.

Suppose X has pdf p and let a be a positive constant. Then

$$\begin{aligned}
\text{Prob}\left(\frac{1}{a}X \le t\right) &= \text{Prob}(X \le at) \\
&= \int_{-\infty}^{at} p(x)\,dx \\
&= \int_{-\infty}^{t} ap(au)\,du \quad \left(\text{making the substitution } u = \frac{1}{a}x\right).
\end{aligned}$$

This says that the random variable $\frac{1}{a}X$ has pdf $ap(ax)$. (Here in action is the scaled pdf $ap(ax)$, which we had as an example of operations on pdf's.) Suppose that we've normalized the mean of X to be 0. Then the variance of $\frac{1}{a}X$ is

$$\begin{aligned}
\sigma^2\left(\frac{1}{a}X\right) &= \int_{-\infty}^{\infty} x^2 ap(ax)\,dx \\
&= a\int_{-\infty}^{\infty} \frac{1}{a^2}u^2 p(u)\frac{1}{a}\,du \quad (\text{making the substitution } u = ax) \\
&= \frac{1}{a^2}\int_{-\infty}^{\infty} u^2 p(u)\,du = \frac{1}{a^2}\sigma^2(X)\,.
\end{aligned}$$

In particular, if we choose $a = \sigma(X)$, then the variance of $\frac{1}{a}X$ is one. This is also a convenient normalization for many formulas.

In summary:

- Given a random variable X with $\mu(X) < \infty$ and $\sigma(X) < \infty$, it is possible to normalize and assume that $\mu(X) = 0$ and $\sigma^2(X) = 1$.

You see these assumptions a lot.

3.6.4. Two examples. Let's have two leading examples of pdfs to refer to. Additional examples are in the problems.

The uniform distribution. "Uniform" refers to a random process where all possible outcomes are equally likely. In the discrete case, tossing a coin or throwing a die are examples. All bins in the ideal histogram have the same height, two bins of height $1/2$ for the toss of a coin, six bins of height $1/6$ for the throw of a single die, and N bins of height $1/N$ for a discrete random variable with N values.

For a continuous random variable, the uniform distribution is identically 1 on an interval of length 1 and zero elsewhere. We've seen such a graph before. If we shift to the interval to go from $-1/2$ to $1/2$, it's the graph of the ever versatile rectangle function: $\Pi(x)$ is now starring in yet another role, that of the uniform distribution.

The mean is 0, obviously,[19] but to verify this formally:

$$\mu = \int_{-\infty}^{\infty} x\Pi(x)\,dx = \int_{-1/2}^{1/2} x\,dx = \tfrac{1}{2}x^2\Big]_{-1/2}^{+1/2} = 0\,.$$

[19] ...the mean of the random variable with pdf $p(x)$ is *not* the average value of $p(x)$ ....

(Also, we're integrating an odd function x, over a symmetric interval.) The variance is then

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 \Pi(x)\, dx = \int_{-1/2}^{1/2} x^2\, dx = \tfrac{1}{3}x^3\Big]_{-1/2}^{+1/2} = \tfrac{1}{12}\,,$$

perhaps not quite so obvious.

The normal distribution. This whole production is about getting to Gaussians, so it seems appropriate that at some point I mention:

- The Gaussian is a pdf.

Indeed, to borrow information from earlier work in this chapter, the Gaussian

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

is a pdf with mean μ and variance σ^2. The distribution associated with such a Gaussian is called a *normal distribution*. There, it's official. But why is it "normal"? You're soon to find out.

3.6.5. Independence. An important extra property that two or more random variables may have is *independence*. The plain English description of independence is that one event or measurement doesn't influence another event or measurement. Each flip of a coin, roll of a die, or measurement of a resistor is a new event, not influenced by previous events.

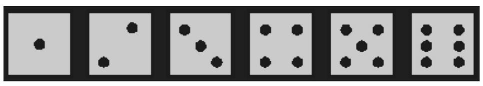
Operationally, independence implies that the probabilities multiply: if two random variables X_1 and X_2 are independent, then

$$\text{Prob}(X_1 \le a \text{ and } X_2 \le b) = \text{Prob}(X_1 \le a) \cdot \text{Prob}(X_2 \le b)\,.$$

In words, if $X_1 \le a$ occurs r percent and $X_2 \le b$ occurs s percent, then, if the events are independent, the percent that $X_1 \le a$ occurs *and* $X_2 \le b$ occurs is r percent of s percent (or the other way around).

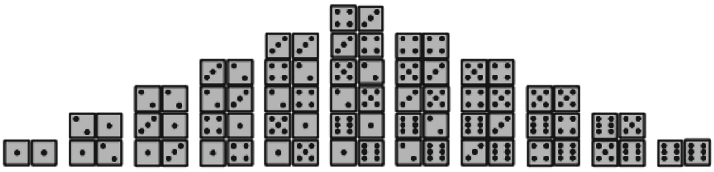
3.6.6. Convolution appears. Using the terminology we've encountered, we can begin to be more precise about the content of the Central Limit Theorem. That result — the ubiquity of the bell-shaped curve — has to do with sums of independent random variables and with the distributions of those sums. The two notions are linked through convolution.

While we'll ultimately work with continuous random variables, let's look at the discrete random variable X = "roll the die" as an example. The ideal histogram for the toss of a single die is uniform; each number 1 through 6 comes up with equal probability. We might represent it pictorially like this:



I don't mean to think just of a picture of dice here, I mean to think of the distribution as six bins of equal height 1/6, each bin corresponding to one of the six possible tosses.

What about the *sum* of the tosses of two dice? What is the distribution, theoretically, of the sums? The possible values of the sum are 2 through 12, but the values do not occur with equal probability. There's only one way of making 2 and one way of making 12, but there are more ways of making the other possible sums. In fact, 7 is the most probable sum, with six ways it can be achieved. We might represent the distribution for the sum of two dice pictorially like this:



It's triangular. Now let's see ... for the single random variable $X =$ "roll one die" we have a distribution like a rect function. For the sum, say random variables $X_1 + X_2 =$ "roll of die 1 plus roll of die 2", the distribution looks like the triangle function....

The key discovery is this:

- **Convolution and probability density functions.** The probability density function of the sum of two independent random variables is the convolution of the probability density functions of each.

What a beautiful, elegant, and useful statement! Let's see why it works.

We can get a good intuitive sense of why this result might hold by looking again at the discrete case and at the example of tossing two dice. To ask about the distribution of the sum of two dice is to ask about the probabilities of particular numbers coming up, and these we can compute directly, using the rules of probability. Take, for example, the probability that the sum is 7. Count the ways, distinguishing which throw is first:

$$\begin{aligned}
&\text{Prob}(\text{Sum} = 7)\\
&\quad = \text{Prob}(\{1 \text{ and } 6\} \text{ or } \{2 \text{ and } 5\} \text{ or } \{3 \text{ and } 4\} \text{ or } \{4 \text{ and } 3\} \text{ or } \{5 \text{ and } 2\} \text{ or } \{6 \text{ and } 1\})\\
&\quad = \text{Prob}(1 \text{ and } 6) + \text{Prob}(2 \text{ and } 5) + \text{Prob}(3 \text{ and } 4)\\
&\qquad + \text{Prob}(4 \text{ and } 3) + \text{Prob}(5 \text{ and } 2) + \text{Prob}(6 \text{ and } 1)\\
&\qquad\quad \text{(probabilities add when events are mutually exclusive)}\\
&\quad = \text{Prob}(1)\,\text{Prob}(6) + \text{Prob}(2)\,\text{Prob}(5) + \text{Prob}(3)\,\text{Prob}(4)\\
&\qquad + \text{Prob}(4)\,\text{Prob}(3) + \text{Prob}(5)\,\text{Prob}(2) + \text{Prob}(6)\,\text{Prob}(1)\\
&\qquad\quad \text{(probabilities multiply when events are independent)}\\
&\quad = 6\left(\frac{1}{6}\right)^2 = \frac{1}{6}\,.
\end{aligned}$$

The particular answer, Prob(sum = 7) = 1/6, is not important here;[20] it's the form of the expression for the solution that should catch your eye. We can write it as

$$\text{Prob}(\text{sum} = 7) = \sum_{k=1}^{6} \text{Prob}(k)\,\text{Prob}(7-k)\,,$$

which is visibly a discrete convolution of Prob with itself; it has the same form as an integral convolution with the sum replacing the integral.

We can extend this observation by introducing

$$p(n) = \begin{cases} \frac{1}{6}, & n = 1, 2, \ldots, 6\,, \\ 0, & \text{otherwise.} \end{cases}$$

This is the discrete uniform density for the random variable "throw one die." Then, by the same reasoning as above,

$$\text{Prob}(\text{sum of two dice} = n) = \sum_{k=-\infty}^{\infty} p(k)p(n-k)\,.$$

You can check that this gives the right answers, including the answer 0 for n bigger than 12 or n less than 2:

n	Prob(sum = n)
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Now let's turn to the case of continuous random variables, and in the following argument look for similarities to the example we just treated. Let X_1 and X_2 be independent random variables with probability density functions $p_1(x_1)$ and $p_2(x_2)$. Because X_1 and X_2 are independent,

$$\text{Prob}(a_1 \le X_1 \le b_1 \text{ and } a_2 \le X_2 \le b_2) = \left(\int_{a_1}^{b_1} p_1(x_1)\,dx_1\right)\left(\int_{a_2}^{b_2} p_2(x_2)\,dx_2\right).$$

Using what has now become a familiar trick, we write this as a double integral:

$$\left(\int_{a_1}^{b_1} p_1(x_1)\,dx_1\right)\left(\int_{a_2}^{b_2} p_2(x_2)\,dx_2\right) = \int_{a_2}^{b_2}\int_{a_1}^{b_1} p_1(x_1)p_2(x_2)\,dx_1\,dx_2\,;$$

[20]But do note that it agrees with what we can observe from the graphic of the sum of two dice. We see that the total number of possibilities for two throws is 36 and that 7 comes up 6/36 = 1/6 of the time.

that is,

$$\text{Prob}(a_1 \le X_1 \le b_1 \text{ and } a_2 \le X_2 \le b_2) = \int_{a_2}^{b_2} \int_{a_1}^{b_1} p_1(x_1)p_2(x_2)\, dx_1\, dx_2\,.$$

If we let a_1 and a_2 drop to $-\infty$, then

$$\text{Prob}(X_1 \le b_1 \,\text{and}\, X_2 \le b_2) = \int_{-\infty}^{b_2} \int_{-\infty}^{b_1} p_1(x_1)p_2(x_2)\, dx_1\, dx_2\,.$$

Since this holds for any b_1 and b_2, we can conclude that

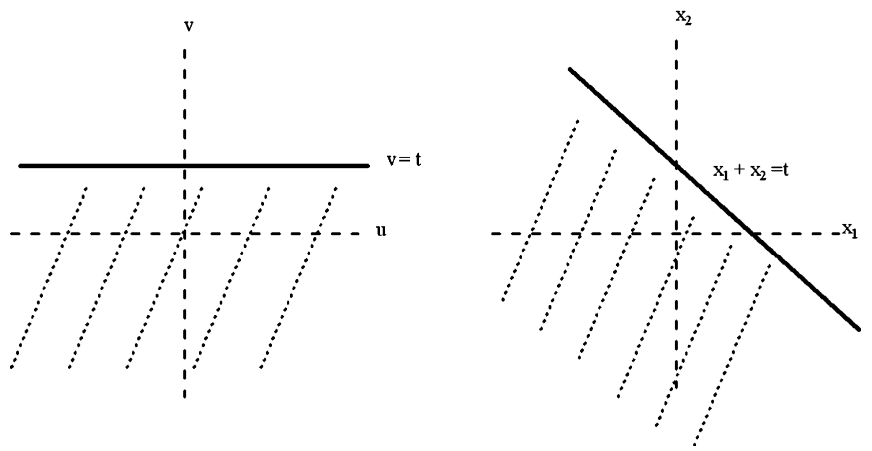
$$\text{Prob}(X_1 + X_2 \le t) = \iint_{x_1+x_2 \le t} p_1(x_1)p_2(x_2)\, dx_1\, dx_2$$

for every t. In words, the probability that $X_1 + X_2 \le t$ is computed by integrating the *joint probability density* $p_1(x_1)p_2(x_2)$ over the region in the (x_1, x_2)-plane where $x_1 + x_2 \le t$.

We're going to make a change of variable in this double integral. We let

$$x_1 = u\,,$$
$$x_2 = v - u\,.$$

Notice that $x_1 + x_2 = v$. Thus under this transformation the (oblique) line $x_1 + x_2 = t$ becomes the horizontal line $v = t$, and the region $x_1 + x_2 \le t$ in the (x_1, x_2)-plane becomes the half-plane $v \le t$ in the (u, v)-plane.



The integral then becomes

$$\iint_{x_1+x_2 \le t} p_1(x_1)p_2(x_2)\, dx_1\, dx_2 = \int_{-\infty}^{t} \int_{-\infty}^{\infty} p_1(u)p_2(v-u)\, du\, dv$$

(the convolution of p_1 and p_2 is inside!)

$$= \int_{-\infty}^{t} (p_2 * p_1)(v)\, dv\,.$$

To summarize, we now see that the probability $\text{Prob}(X_1+X_2 \le t)$ for any t is given by

$$\text{Prob}(X_1 + X_2 \le t) = \int_{-\infty}^{t} (p_2 * p_1)(v)\, dv\,.$$

Therefore the probability density function of $X_1 + X_2$ is $(p_2 * p_1)(t)$.

This extends to the sum of any finite number of random variables: if $X_1, X_2, \dots, X_n$ are independent random variables with probability density functions $p_1, p_2, \dots, p_n$, respectively, then the probability density function of $X_1 + X_2 + \cdots + X_n$ is $p_1 * p_2 * \cdots * p_n$. Cool. Cool. ... Cool.

For a single probability density $p(x)$ we'll write

$$p^{*n}(x) = (p{*}p{*}\cdots{*}p)(x) \quad (n \text{ factors of } p \text{, i.e., } n-1 \text{ convolutions of } p \text{ with itself}).$$

3.6.7. The Central Limit Theorem: The bell curve tolls for thee. People say things like, "The Central Limit Theorem[21] says the sum of n independent random variables is well approximated by a Gaussian if n is large." They mean to say that the sum is *distributed* like a Gaussian. To make a true statement we have to begin with a few assumptions — but not many — on how the random variables themselves are distributed. Call the random variables $X_1, X_2, \dots, X_n, \dots$. We assume first of all that the X's are independent. We also assume that all of the X's have the same probability density function. So for practical purposes you can think of the random variables as being the same, like making the same measurements in different trials, or throwing a die hundreds of times, recording the results, and then doing it again. Naturally, there's some terminology and an acronym that goes along with this. One says that the X's are *independent and identically distributed*, or *iid*. In particular, the X's all have the same mean, say μ, and they all have the same standard deviation, say σ. We assume that both μ and σ are finite.

Consider the sum

$$S_n = X_1 + X_2 + \cdots + X_n\,.$$

We want to say that S_n is distributed like a Gaussian as n increases. But which Gaussian? The mean and standard deviation for the X's are all the same, but for S_n they are changing with n. It's not hard to show, though, that for S_n the mean scales by n and the standard deviation scales by $\sqrt{n}$:

$$\mu(S_n) = n\mu,$$
$$\sigma(S_n) = \sqrt{n}\,\sigma\,.$$

We'll derive these later, in Section 3.6.8, so as not to interrupt the action.

To make sense of S_n approaching a particular Gaussian, we should therefore recenter and rescale the sum, say fix the mean to be zero and fix the standard deviation to be 1. We should work with

$$\frac{S_n - n\mu}{\sqrt{n}\,\sigma}$$

[21] Abbreviated, of course, as CLT.

and ask what happens as $n \to \infty$. One form of the Central Limit Theorem says that

$$\lim_{n\to\infty} \text{Prob}\left(a < \frac{S_n - n\mu}{\sqrt{n}\sigma} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx\,.$$

On the right-hand side is the Gaussian $(1/\sqrt{2\pi})e^{-x^2/2}$ with mean 0 and standard deviation 1. The theorem says that probabilities for the *normalized* sum of the random variables approach those based on this Gaussian.

We'll focus on the convergence of the pdf's for S_n, sort of an unintegrated form of the way the CLT is stated above. Let $p(x)$ be the common probability density function for the $X_1, X_2, \ldots, X_n$ (the pdf for the iid X's, for those who like to compress their terminology). We'll start by assuming already that $\mu = 0$ and $\sigma = 1$ for the X's. This means that

$$\int_{-\infty}^{\infty} xp(x)\, dx = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} x^2 p(x)\, dx = 1\,,$$

in addition to

$$\int_{-\infty}^{\infty} p(x)\, dx = 1\,,$$

which is true for every pdf.

Now, the mean of S_n is zero, but the standard deviation is $\sqrt{n}$, so we want to work $S_n/\sqrt{n}$. What is the pdf of this? We've shown that the pdf for $S_n = X_1 + \cdots + X_n$ is

$$p^{*n}(x) = (p * p * \cdots * p)(x)\,;$$

hence the probability density function for $S_n/\sqrt{n}$ is

$$p_n(x) = \sqrt{n}\, p^{*n}(\sqrt{n}\, x)\,.$$

(Careful here: It's $(p * p * \cdots * p)(\sqrt{n}\, x)$, not $p(\sqrt{n}\, x) * p(\sqrt{n}\, x) * \cdots p(\sqrt{n}\, x)$.)

We're all set to show:

- **Central Limit Theorem.** Let $X_1, X_2, \ldots, X_n, \ldots$ be independent, identically distributed random variables with mean 0 and standard deviation 1. Let $p_n(x)$ be the probability density function for

  $$S_n/\sqrt{n} = (X_1 + X_2 + \cdots + X_n)/\sqrt{n}.$$

  Then

  $$p_n(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \quad \text{as } n \to \infty\,.$$

The idea is to take the Fourier transform of p_n, which, by the convolution theorem, will essentially be the *product* of the Fourier transforms of p. Products are easier than convolutions, and the hope is to use the assumptions on p to get some information on the form of this product as $n \to \infty$. For dramatic effect, and it's very dramatic, I'm going to run through this a little quickly, supplying some further details afterward.

Begin with the Fourier transform of

$$p_n(x) = \sqrt{n}\, p^{*n}(\sqrt{n}\, x)\,.$$

We'll use the capital letter notation and write $P(s) = \mathcal{F}p(s)$. Then the Fourier transform of $p_n(x)$ is

$$P^n\left(\frac{s}{\sqrt{n}}\right) \quad \text{(ordinary } n\text{th power here).}$$

The normalization of mean zero and standard deviation allows us to do something with $P(s/\sqrt{n})$. Using a Taylor series approximation for the exponential function,[22] we have

$$\begin{aligned} P\left(\frac{s}{\sqrt{n}}\right) &= \int_{-\infty}^{\infty} e^{-2\pi i s x/\sqrt{n}}\, p(x)\, dx \\ &= \int_{-\infty}^{\infty} \left(1 - \frac{2\pi i s x}{\sqrt{n}} + \frac{1}{2}\left(\frac{2\pi i s x}{\sqrt{n}}\right)^2 + \text{small}\right) p(x)\, dx \\ &= \int_{-\infty}^{\infty} \left(1 - \frac{2\pi i s x}{\sqrt{n}} - \frac{2\pi^2 s^2 x^2}{n} + \text{small}\right) p(x)\, dx \\ &= \underbrace{\int_{-\infty}^{\infty} p(x)\, dx}_{=1} - \frac{2\pi i s}{\sqrt{n}} \underbrace{\int_{-\infty}^{\infty} x p(x)\, dx}_{=0} \\ &\qquad - \frac{2\pi^2 s^2}{n} \underbrace{\int_{-\infty}^{\infty} x^2 p(x)\, dx}_{=1} + \int_{-\infty}^{\infty} (\text{small})\, p(x)\, dx \\ &= 1 - \frac{2\pi^2 s^2}{n} + \text{small}\,. \end{aligned}$$

See how the normalizations came in:

$$\int_{-\infty}^{\infty} p(x)\, dx = 1\,, \quad \int_{-\infty}^{\infty} x p(x)\, dx = 0\,, \quad \int_{-\infty}^{\infty} x^2 p(x)\, dx = 1\,.$$

That "small" term comes from the remainder in the Taylor series for the exponential and it tends to 0 faster than $1/n$ as $n \to \infty$; see below. Using $\big(1 + x/n\big)^n \to e^x$, from calculus days, we have for large n

$$P^n\left(\frac{s}{\sqrt{n}}\right) \approx \left(1 - \frac{2\pi^2 s^2}{n}\right)^n \approx e^{-2\pi^2 s^2}\,.$$

Taking the inverse Fourier transform of $e^{-2\pi^2 s^2}$ and knowing what happens to the Gaussian, taking the limit as $n \to \infty$, taking the rest of the day off for a job well done, we conclude that

$$p_n(x) \to \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\,.$$

Catch your breath and relax.

[22] You remember that $e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$. This works when x is complex, too.

What about that "small" part? Only for those who are interested — you have to be somewhat comfortable with the remainder in Taylor series. To see more carefully what's going on with the "small" part, write

$$e^{-2\pi i(s/\sqrt{n})x} = 1 - \frac{2\pi isx}{\sqrt{n}} - \frac{2\pi s^2x^2}{n} + R_n(x).$$

The remainder $R_n(x)$ also depends on s and n and for fixed x (and s) it tends to 0 as $n \to \infty$ like $1/n^{3/2}$. Split the Fourier transform into two pieces:

$$\int_{-\infty}^{\infty} e^{-2\pi i(s/\sqrt{n})x} p(x)\,dx = \int_{|x|\leq 1} e^{-2\pi i(s/\sqrt{n})x} p(x)\,dx + \int_{|x|\geq 1} e^{-2\pi i(s/\sqrt{n})x} p(x)\,dx\,.$$

Then,

$$\begin{aligned}
&\int_{|x|\leq 1} e^{-2\pi i(s/\sqrt{n})x} p(x)\,dx + \int_{|x|\geq 1} e^{-2\pi i(s/\sqrt{n})x} p(x)\,dx \\
&\quad = \int_{|x|\leq 1} \left(1 - \frac{2\pi isx}{\sqrt{n}} - \frac{2\pi^2s^2x^2}{n} + R_n(x)\right) p(x)\,dx \\
&\qquad + \int_{|x|\geq 1} \left(1 - \frac{2\pi isx}{\sqrt{n}} - \frac{2\pi^2s^2x^2}{n} + \frac{1}{x^2}x^2R_n(x)\right) p(x)\,dx \\
&\quad = \int_{-\infty}^{\infty} \left(1 - \frac{2\pi isx}{\sqrt{n}} - \frac{2\pi^2s^2x^2}{n}\right) p(x)\,dx \\
&\qquad + \int_{|x|\leq 1} R_n(x)p(x)\,dx + \int_{|x|\geq 1} \left(\frac{1}{x^2}R_n(x)\right) x^2p(x)\,dx.
\end{aligned}$$

Since $\int_{-\infty}^{\infty} p(x)\,dx = 1$ and $R_n(x)$ is bounded for $|x| \leq 1$, we have

$$\int_{|x|\leq 1} R_n(x)p(x)\,dx = o(1), \quad n \to \infty.$$

Here $o(1)$ stands for a quantity that tends to 0 as $n \to \infty$. Likewise, since

$$\frac{1}{x^2}R_n(x) = \frac{1}{x^2}\left(e^{-2\pi i(s/\sqrt{n})x} - \left(1 - \frac{2\pi isx}{\sqrt{n}} - \frac{2\pi^2s^2x^2}{n}\right)\right)$$

is bounded for $|x| \geq 1$ and $\int_{-\infty}^{\infty} x^2p(x)\,dx = 1$, we also have

$$\int_{|x|\geq 1} \left(\frac{1}{x^2}R_n(x)\right) x^2p(x)\,dx = o(1), \quad n \to \infty.$$

Putting these together and appealing to the normalizations as earlier, we can conclude that

$$\int_{-\infty}^{\infty} e^{-2\pi i(s/\sqrt{n})x} p(x)\,dx = 1 - \frac{2\pi^2s^2}{n} + o(1).$$

That's where we ended up before, when we swept all the details into "small." All is well.

3.6.8. The mean and standard deviation of the sum of random variables. The setup for the Central Limit Theorem involves the sum

$$S_n = X_1 + X_2 + \cdots + X_n$$

of n independent random variables, all having the same pdf $p(x)$. Thus all of the X's have the same mean and the same variance:

$$\mu = \int_{-\infty}^{\infty} xp(x)\, dx\,, \quad \sigma^2 = \int_{-\infty}^{\infty} x^2 p(x)\, dx\,.$$

We needed to know that the mean and the standard deviation of S_n are

$$\mu(S_n) = n\mu\,, \quad \sigma(S_n) = \sqrt{n}\,\sigma\,.$$

Take the first of these. The pdf for $S_2 = X_1 + X_2$ is $p * p$, and hence

$$\begin{aligned}
\mu(S_2) &= \int_{-\infty}^{\infty} x(p*p)(x)\, dx \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} p(x-y)p(y)\, dy \right) dx \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} xp(x-y)\, dx \right) p(y)\, dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} (u+y)p(u)\, du \right) p(y)\, dy \quad \text{(using } u = x-y\text{)} \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} up(u)\, du + y \int_{-\infty}^{\infty} p(u)\, du \right) p(y)\, dy \\
&= \int_{-\infty}^{\infty} (\mu + y)p(y)\, dy \quad \left(\text{using } \textstyle\int_{-\infty}^{\infty} up(u)\, du = \mu \text{ and } \int_{-\infty}^{\infty} p(u)du = 1\right) \\
&= \mu \int_{-\infty}^{\infty} p(u)\, du + \int_{-\infty}^{\infty} yp(y)\, dy \\
&= \mu + \mu\,.
\end{aligned}$$

Inductively, we get $\mu(S_n) = n\mu$.

How about the variance, or standard deviation? Again let's do this for $S_2 = X_1 + X_2$. We first assume that the mean of the X's is 0, and hence the mean of S_2 is 0 as well:

$$\int_{-\infty}^{\infty} xp(x)\, dx = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} x(p*p)(x)\, dx = 0\,.$$

Then the variance of S_2 is

$$\begin{aligned}
\sigma^2(S_2) &= \int_{-\infty}^{\infty} x^2 (p*p)(x)\,dx \\
&= \int_{-\infty}^{\infty} x^2 \left(\int_{-\infty}^{\infty} p(x-y)p(y)\,dy\right) dx \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x^2 p(x-y)\,dx\right) p(y)\,dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} (u+y)^2 p(u)\,du\right) p(y)\,dy \quad (\text{using } u = x-y) \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} (u^2+2uy+y^2) p(u)\,du\right) p(y)\,dy \\
&= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} u^2 p(u)\,du + 2y\int_{-\infty}^{\infty} up(u)\,du + y^2\int_{-\infty}^{\infty} p(u)\,du\right) p(y)\,dy \\
&= \int_{-\infty}^{\infty} (\sigma^2+y^2)p(y)\,dy \ (\text{using } \textstyle\int_{-\infty}^{\infty} u^2p(u)du = \sigma^2 \text{ and } \int_{-\infty}^{\infty} up(u)du = 0) \\
&= \sigma^2 \int_{-\infty}^{\infty} p(y)\,dy + \int_{-\infty}^{\infty} y^2 p(y)\,dy \\
&= \sigma^2 + \sigma^2 \;=\; 2\sigma^2\,.
\end{aligned}$$

So the variance of S_2 is $2\sigma^2$ and the standard deviation is $\sigma(S_2) = \sqrt{2}\sigma$. Once again, inductively,

$$\sigma(S_n) = \sqrt{n}\,\sigma\,.$$

Pretty nice, really. I'll let you decide what to do if the mean is not zero at the start.

3.7. Heisenberg's Inequality

Since we've gone to the trouble of introducing some of the terminology from probability and statistics (mean, variance, etc.), I thought you might appreciate seeing another application.

Consider the stretch theorem, which reads:

- If $f(t) \rightleftharpoons F(s)$, then $f(at) \rightleftharpoons \dfrac{1}{|a|}F\left(\dfrac{s}{a}\right)$.

If a is large, then $f(at)$ is squeezed and $(1/|a|)F(s/a)$ is stretched. Conversely, if a is small, then $f(at)$ is stretched and $(1/|a|)F(s/a)$ is squeezed.

A more quantitative statement of the trade-off between the spread of a signal and the spread of its Fourier transform is related to (equivalent to) that most famous inequality in quantum mechanics, the Heisenberg uncertainty principle.

Suppose $f(x)$ is a signal with finite energy,

$$\int_{-\infty}^{\infty} |f(x)|^2\,dx < \infty\,.$$

Normalize the signal by dividing f by the square root of its energy, and thus assume that

$$\int_{-\infty}^{\infty} |f(x)|^2 \, dx = 1 \, .$$

We can then regard $|f(x)|^2$ as defining a probability density function, and it has a mean and a variance. Now, by Parseval's identity,

$$\int_{-\infty}^{\infty} |\hat{f}(s)|^2 \, ds = \int_{-\infty}^{\infty} |f(x)|^2 \, dx = 1 \, .$$

Thus $|\hat{f}(s)|^2$ also defines a probability distribution, and it too has a mean and variance. How do they compare to those of $|f(x)|^2$?

As earlier, we shift $f(x)$, or rather $|f(x)|^2$, to assume that the mean is 0. The effect on $\hat{f}(s)$ of shifting $f(x)$ is to multiply by a complex exponential, which has absolute value 1 and hence does not affect $|\hat{f}(s)|^2$. In the same manner, we can shift $\hat{f}(s)$ so that it has zero mean, and again there will be no effect on $|f(x)|^2$.

To summarize, we assume that the probability distributions $|f(x)|^2$ and $|\hat{f}(s)|^2$ each have mean 0, and we are interested in comparing their variances:

$$\sigma^2(f) = \int_{-\infty}^{\infty} x^2 |f(x)|^2 \, dx \quad \text{and} \quad \sigma^2(\hat{f}) = \int_{-\infty}^{\infty} s^2 |\hat{f}(s)|^2 \, ds \, .$$

The Heisenberg uncertainty principle states that

$$\sigma(f)\sigma(\hat{f}) \geq \frac{1}{4\pi} \, .$$

In words, this says that not both of $\sigma(f)$ and $\sigma(\hat{f})$ can be small. If one is tiny, the other has to be big enough so that their product is at least $1/4\pi$.

After all the setup, the argument to deduce the lower bound is pretty easy, except for a little trick right in the middle. It's also helpful to assume that we're working with complex-valued functions — I think the trick that comes up is actually a little easier to verify in that case. Finally, we're going to assume that $|f(x)|$ decreases rapidly enough at $\pm\infty$. You'll see what's needed. The result can be proved for more general functions via approximation arguments. Here we go:

$$\begin{aligned}
4\pi^2\sigma(f)^2\sigma(\hat{f})^2 &= 4\pi^2 \int_{-\infty}^{\infty} x^2 |f(x)|^2 \, dx \int_{-\infty}^{\infty} s^2 |\hat{f}(s)|^2 \, ds \\
&= \int_{-\infty}^{\infty} x^2 |f(x)|^2 \, dx \int_{-\infty}^{\infty} |2\pi i s|^2 |\hat{f}(s)|^2 \, ds \\
&= \int_{-\infty}^{\infty} |xf(x)|^2 \, dx \int_{-\infty}^{\infty} |\widehat{f'}(s)|^2 \, ds \\
&= \int_{-\infty}^{\infty} |xf(x)|^2 \, dx \int_{-\infty}^{\infty} |f'(x)|^2 \, dx \\
&\qquad \text{(by Parseval's identity applied to } f'(x)) \\
&\geq \left(\int_{-\infty}^{\infty} |x\overline{f(x)}f'(x)| \, dx \right)^2 \quad \text{(by the Cauchy-Schwarz inequality).}
\end{aligned}$$

Here comes the trick. In the integrand, we have $|\overline{f(x)}f'(x)|$. The magnitude of any complex number is always greater than its real part.[23] Hence

$$\begin{aligned}|x\overline{f(x)}f'(x)| &\geq x\,\mathrm{Re}\{\overline{f(x)}f'(x)\}\\ &= x\,\frac{1}{2}\big(\,\overline{f(x)}f'(x) + f(x)\overline{f'(x)}\big) = \frac{1}{2}x\frac{d}{dx}\big(\,\overline{f(x)}f(x)\big) = \frac{1}{2}x\frac{d}{dx}|f(x)|^2\,.\end{aligned}$$

Use this in the last line, above:

$$\Big(\int_{-\infty}^{\infty} |x\overline{f(x)}f'(x)|\,dx\Big)^2 \geq \Big(\int_{-\infty}^{\infty} x\frac{d}{dx}\left(\frac{1}{2}|f(x)|^2\right)\,dx\Big)^2$$

Now integrate by parts with $u = x$, $dv = \frac{d}{dx}\frac{1}{2}|f(x)|^2\,dx$. The term $[uv]_{-\infty}^{\infty}$ drops out because *we assume it does*; i.e., we assume that $x|f(x)|$ goes to zero as $x \to \pm\infty$. Therefore we're left with the integral of $v\,du$ (and the whole thing is squared). That is,

$$\Big(\int_{-\infty}^{\infty} x\frac{d}{dx}\left(\frac{1}{2}|f(x)|^2\right)\,dx\Big)^2 dx = \frac{1}{4}\Big(\int_{-\infty}^{\infty} |f(x)|^2\,dx\Big)^2 = \frac{1}{4}\,.$$

To summarize, we have shown that

$$4\pi^2\sigma(f)^2\sigma(\hat{f})^2 \geq \frac{1}{4} \quad \text{or} \quad \sigma(f)\sigma(\hat{f}) \geq \frac{1}{4\pi}\,.$$

Remark. Using the case of equality in the Cauchy-Schwarz inequality, one can show that equality holds in Heisenberg's inequality exactly for constant multiples of $f(x) = e^{-kx^2}$. Yet another spooky appearance of the Gaussian.

Is this quantum mechanics? I'm told it is. The quantum mechanics of a particle moving in one dimension that goes along with this inequality runs as follows, in skeletal form, with no attempt at motivation:

The *state* of a particle moving in one dimension is given by a complex-valued function ψ in $L^2(\mathbb{R})$, the square integrable functions on the real line. L^2 plays a big role in quantum mechanics — you need a space to work in, and L^2 is the space. Really.[24] Probabilities are done with complex quantities in this business, and the first notion is that the probability of finding the particle in the interval $a \leq x \leq b$ is given by

$$\int_a^b \psi(x)^*\psi(x)\,dx\,,$$

where in this field it's customary to write the complex conjugate of a quantity using an asterisk instead of an overline.

An *observable* is a symmetric linear operator A operating on some subset of functions (states) in $L^2(\mathbb{R})$.[25] The *average* of A in the state ψ is defined to be

$$\int_{-\infty}^{\infty} \psi(x)^*(A\psi)(x)\,dx.$$

[23] Draw a picture. The complex number is a vector, which is always longer than its x-component.

[24] It's orthogonality and all that goes with it. The inner product is $(f, g) = \int_{-\infty}^{\infty} f(x)\overline{g(x)}\,dx$. We'll come back to $L^2(\mathbb{R})$ in the next chapter.

[25] "Symmetric" means that A is equal to its transpose, which, in terms of inner products, translates to $(Af, g) = (f, Ag)$.

One important observable is the "position of the particle," and this, as it turns out, is associated to the operator "multiplication by x." Thus the average position is

$$\int_{-\infty}^{\infty} \psi(x)^*(A\psi)(x)\,dx = \int_{-\infty}^{\infty} \psi(x)^* x\psi(x)\,dx = \int_{-\infty}^{\infty} x|\psi(x)|^2\,dx\,.$$

Another important observable is *momentum*, and this is associated with the operator

$$B = \frac{1}{2\pi i}\frac{d}{dx}\,.$$

The average momentum is then

$$\begin{aligned}\int_{-\infty}^{\infty} \psi(x)^*(B\psi)(x)\,dx &= \int_{-\infty}^{\infty} \psi(x)^* \frac{1}{2\pi i}\psi'(x)\,dx \\ &= \int_{-\infty}^{\infty} \hat{\psi}(s)^* s\hat{\psi}(s)\,ds \\ &\qquad \text{(using Parseval's identity for products of functions)} \\ &= \int_{-\infty}^{\infty} s|\psi(s)|^2\,ds\,.\end{aligned}$$

The position and momentum operators do *not* commute:

$$(AB - BA)(\psi) = \frac{1}{2\pi i}\left(x\frac{d}{dx} - \frac{d}{dx}x\right)(\psi) = -\frac{1}{2\pi i}\psi\,.$$

In quantum mechanics this means that the position and momentum cannot *simultaneously* be measured with arbitrary accuracy. The Heisenberg inequality, as a lower bound for the product of the two variances, is a quantitative way of stating this.

"Uncertainty principles" have assumed an important role in Fourier analysis. Start with the fundamental paper of Donaho and Stark, Uncertainty principles and signal recovery, SIAM J. Appl. Math. **49** (1989), 906–931. Work your way forward from there.

Problems and Further Results

3.1. (a) Show that $\Pi * \Pi = \Lambda$ using the definition of convolution, i.e.,

$$(\Pi * \Pi)(x) = \int_{-\infty}^{\infty} \Pi(y)\Pi(x-y)\,dy.$$

(b) What about $\Pi_a * \Pi_a$? (Use any method you wish.)

(c) *Sercan and Aditya discuss convolution:*

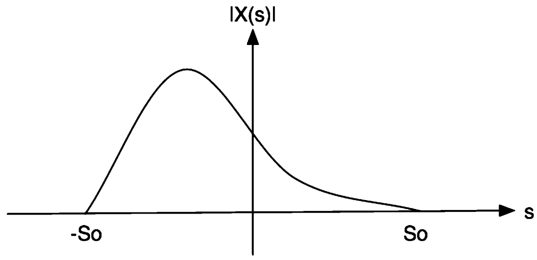
Sercan: You know, I think this problem suggests something general about how convolution spreads out a signal.

Aditya: How so?

Sercan: Well, we showed that $\Pi * \Pi = \Lambda$. I know that Λ has a different shape than Π, but notice that while $\Pi(x) = 0$ for $|x| \geq 1/2$, we have $\Lambda(x) = 0$ outside the bigger interval $|x| \geq 1$. So $\Pi * \Pi$ is more spread out.
Aditya: I agree, but do you have any more examples?
Sercan: The second example $\Pi_a * \Pi_a$ exhibits the same phenomenon.
Aditya: Is there a general pattern there?
Sercan: I'm not sure, but based on how we showed $\Pi*\Pi = \Lambda$, I think I'll conjecture that if $f(x) = 0$ for $|x| \geq a/2$, then $(f*f)(x) = 0$ for $|x| \geq a$, so $(f*f)(x)$ is twice as spread out as $f(x)$.

Is Sercan on to something or too optimistic?

3.2. Let $x(t)$ be a signal whose spectrum is identically zero outside the range $-S_0 \leq s \leq S_0$. An example of such a spectrum is shown below.



For the signal $x^3(t) * x^8(t)$, determine the range over which the spectrum is nonzero. See Problem 3.1.

3.3. *Some sample convolutions*[26]

(a) It is often useful to represent operations on signals as convolutions. For each of the following, find a function $h(t)$ such that $y(t) = (x * h)(t)$.

(i) $y(t) = \displaystyle\int_{-\infty}^{t-1} x(\tau)d\tau$.

(ii) $y(t) = \displaystyle\int_{t-T/2}^{t+T/2} x(\tau)d\tau$.

(b) Let $h(t) = \Pi(t - 1/2)$ and $x(t) = \sin(2\pi t)u(t)$, where $u(t) = 1$ for $t \geq 0$ and 0 for $t < 0$. Sketch $h(t)$ and $x(t)$ and then find $y(t) = (x * h)(t)$.

3.4. *Reversals, shifts, stretches, and convolution*

(a) If both $f(t)$ and $g(t)$ are reversed, what happens to their convolution? If one of $f(t)$ and $g(t)$ is reversed, what happens to their convolution?

(b) Show that

$$(\tau_b f) * g = \tau_b(f * g) = f * (\tau_b g)\,.$$

[26]From Raj Bhatnagar.

Use this result to deduce that if either f or g is periodic of period T, then $f * g$ is periodic of period T.

(c) Show that

$$(\sigma_a f) * g = \frac{1}{|a|}\sigma_a(f * (\sigma_{1/a}g))\,, \quad (\sigma_a f) * (\sigma_a g) = \frac{1}{|a|}\sigma_a(f * g)\,.$$

3.5. Eva and Rajiv converse about convolution:

Rajiv: You know, convolution really is a remarkable operation, the way it imparts properties of one function onto the convolution with another. Take periodicity: if $f(t)$ is periodic, then $(f * g)(t)$ is periodic with the same period as f. That was an earlier problem.

Eva: There's a problem with that statement. You want to say that if $f(t)$ is a periodic function of period T, then $(f * g)(t)$ is also periodic of period T.

Rajiv: Right.

Eva: What if $g(t)$ is also periodic, say of period R? Then doesn't $(f * g)(t)$ have two periods, T and R?

Rajiv: I suppose so.

Eva: But wouldn't this lead right to a contradiction? I mean, for example, you can't have a function with two periods, can you?

Rajiv: I think we've found a fundamental contradiction in mathematics.

Eva: Why don't we look at a simple, special case first. What happens if you convolve $\sin 2\pi t$ with itself?

Rajiv: OK, both functions have period 1 so for the convolution you get a function that's periodic of period 1, no problem.

Eva: No, you don't. Something goes wrong.

What's going on? With whom do you agree and why? What do you think about that statement: "If $f(t)$ is periodic, then $f * g$ is periodic."

3.6. *Some practice with convolution*

(a) Let $f(x) = e^{-|x|}$, $-\infty < x < \infty$. Find $(f * f)(x)$.

(b) Let $g(x) = e^{-\pi x^2}$, $-\infty < x < \infty$. Show that $(g * g)(x) = \frac{1}{\sqrt{2}}e^{-\pi x^2/2}$. From this, deduce the result of the n-fold convolution of g, i.e., $g * g * \ldots * g$ (with n factors of g).

3.7. *Convolving sincs*

(a) Show that $\operatorname{sinc} t * \operatorname{sinc} t = \operatorname{sinc} t$.

(b) A little more generally, show that $\operatorname{sinc}(pt) * \operatorname{sinc}(p't) = \frac{p''}{pp'}\operatorname{sinc}(p''s)$ where $p'' = \min\{p, p'\}$. Assume that $p, p' > 0$.

(c) Scaling equally in the two sincs, show that

$$\operatorname{sinc}(pt - a) * \operatorname{sinc}(pt - b) = \frac{1}{p}\operatorname{sinc}(pt - (a + b))\,.$$

There's still a more general formula allowing for independent scaling of the sincs. You can work that out.

3.8. *Areas multiply under convolution*

Let $h = f * g$. Show that

$$\int_{-\infty}^{\infty} h(t)\,dt = \left(\int_{-\infty}^{\infty} f(t)\,dt\right)\left(\int_{-\infty}^{\infty} g(t)\,dt\right).$$

Joelle says: "This can be used to show that the convolution of two probability distributions is a probability distribution." Is Joelle right?

3.9. *Convolution and derivative*

Use the convolution theorem and the derivative theorem $\mathcal{F}f'(s) = 2\pi is\mathcal{F}f(s)$ to show $(f * g)' = f' * g = f * g'$.

3.10. Convolution with the unit step function

$$u(x) = \begin{cases} 1, & x \geq 0\,, \\ 0, & x < 0\,, \end{cases}$$

can be used as an integrator. Show for a signal $f(x)$ that

$$\int_0^x f(y)\,dy = (u * (uf))(x)\,.$$

(uf is the product of u and f; i.e., $(uf)(x) = u(x)f(x)$.)

3.11. *Periodizing and convolution*

Let f and g be functions that are identically 0 outside the interval $(-1/2, 1/2)$. Let f_T and g_T denote the periodizations of f and g with period T (with $T \geq 1$); i.e.,

$$f_T(t) = \sum_{n=-\infty}^{\infty} f(t - nT), \quad g_T(t) = \sum_{n=-\infty}^{\infty} g(t - nT).$$

Let p be the convolution of f and g, and let q be the convolution of f_T and g_T, the latter taken as the convolution of periodic functions:

$$p(t) = \int_{-\infty}^{\infty} f(x)g(t - x)dx, \quad q(t) = \int_{-T/2}^{T/2} f_T(y)g_T(t - y)dy.$$

Note that q is periodic of period T. Show that q is a periodization of p with period T; i.e.,

$$q(t) = \sum_{n=-\infty}^{\infty} p(t - nT).$$

Hint: Work your way from the right-hand side to the left-hand side.

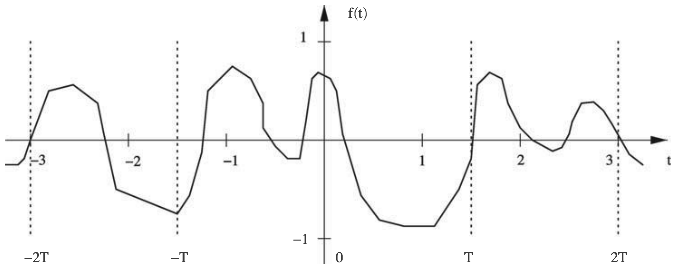
3.12. *Divided time signal*

Consider a time signal $f(t)$ with Fourier transform $F(s)$. We divide this signal into equal nonoverlapping time segments, denoting segments as

$$f_n(t) = \begin{cases} f(t), & nT < t < (n+1)T, \\ 0, & \text{otherwise,} \end{cases}$$

where n is an integer.

For an illustration, please see the figure below for division of an arbitrary time signal. The signal $f_0(t)$ is the portion of $f(t)$ in the interval $(0, T)$, $f_1(t)$ is the portion in the interval $(T, 2T)$, and so on.



(a) Find the formula for the Fourier transform of a segment function $f_n(t)$, where n is an integer, in terms of the original function's Fourier transform $F(s)$.

(b) Now let $f(t)$ be *periodic* with period T and divided equally into segment functions, $f_n(t)$, as described above. Let $F_0(s)$ be the Fourier transform of the segment $f_0(t)$. Find $F_n(s)$, the Fourier transform of other segments, in terms of $F_0(s)$.

3.13. Let $f(x)$ be a signal and for $h > 0$ let $A_h f(x)$ be the averaging operator,

$$A_h f(x) = \frac{1}{2h} \int_{x-h}^{x+h} f(y)\, dy.$$

(a) Show that we have the alternate expressions for $A_h f(x)$:

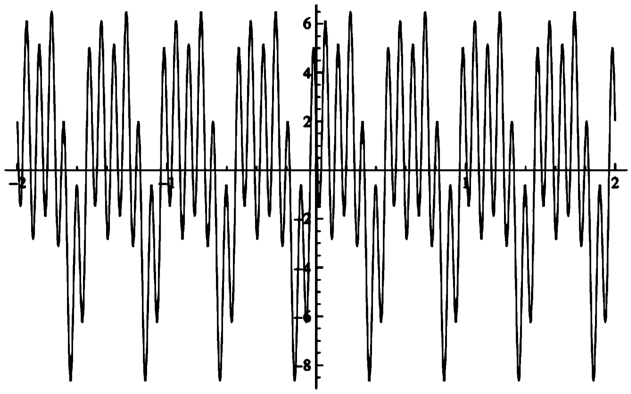
$$A_h f(x) = \frac{1}{2h} \int_{-h}^{h} f(x+y)\, dy = \frac{1}{2h} \int_{-h}^{h} f(x-y)\, dy\,.$$

(b) Express $A_h f(x)$ as a convolution and find the Fourier transform $\mathcal{F}(A_h f)(s)$.

3.14. Let $f(x)$ be a signal and for $h > 0$ let $A_h f(x)$ be the averaging operator from Problem 3.13. The signal

$$f(t) = 3 \sin 4\pi t + 2 \cos 8\pi t - 4 \sin 24\pi t$$

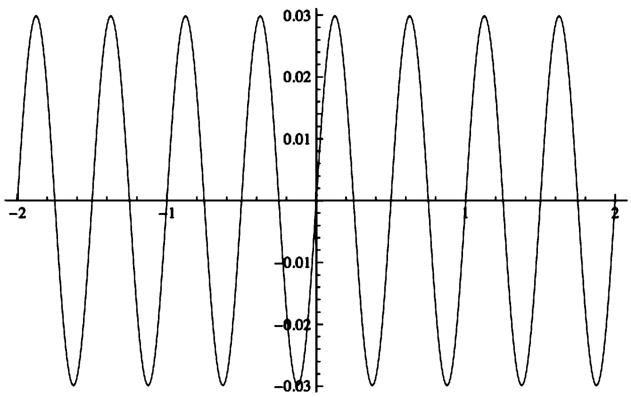
is plotted below.



It's complicated. The signal is then averaged with $h = 1/8$ to obtain

$$g(t) = A_{1/8} f(t).$$

The plot of $g(t)$, which looks like



shows only the lowest frequency of the original signal $f(t)$. Why is this?

3.15. *Frequency-hopping spread spectrum*

Per Wikipedia: "Frequency-hopping spread spectrum (FHSS) is a method of transmitting radio signals by rapidly switching a carrier among many frequency channels, using a pseudorandom sequence known to both transmitter and receiver." To an outside party without knowledge of this frequency hopping sequence, the sequence appears random; on the other hand, the intended receiver can determine the sequence with his knowledge of the seed used to generate the pseudorandom sequence. As we will see, this makes FHSS robust against eavesdropping and jamming.

Imagine that your favorite FM radio station employs FHSS. Then for instance it would transmit at 101.3 MHz for 1 minute, then switch to 88.5 MHz for 1 minute, then 94.5 MHz, 97.3 MHz, 90.1 Mhz, and so forth. If you did not have knowledge of this hopping sequence, then you would have trouble listening in! However, if you knew the seed used to generate the sequence of hops, you could anticipate each hop to the next station and dial in appropriately every minute. Because at any given time each signal only occupies a small part of the spectrum, multiple users could transmit messages simultaneously, each hopping about the spectrum.

In this MATLAB exercise, you will decode two messages that were simultaneously transmitted, each with its own pseudorandom hopping sequence. The messages $m_1(t), m_2(t)$ are bandlimited to 4 kHz (i.e., $\mathcal{F}m_1(s), \mathcal{F}m_2(s) = 0$ for $|s| > 2$ kHz) and modulated to their carrier frequency by multiplication with a cosine (see Problem 2.7). The carrier frequencies, $f_1(t), f_2(t)$, change every 1 second, but are constant for each second of transmission. Because you (as the receiver) know the pseudorandom seeds, the hopping sequences of $f_1(t)$ and $f_2(t)$ across 20 channels are known to you. These 20 channels have carrier frequencies of 12, 16, 20, . . . , 88 kHz. The transmitted message is

$$Y(t) = m_1(t)\cos(2\pi f_1(t)t) + m_2(t)\cos(2\pi f_2(t)t).$$

Download
`genSpreadSpectrum.mat`,
`genSpreadSpectrum.m`,
`decodeSpreadSpectrum.m`.

This exercise will parallel the `decodeSpreadSpectrum.m` script. The script will run as is, but your job is to complete the code inside `decodeSpreadSpectrum.m` that decodes the two messages for each scenario.

(a) For each second of transmission, show how multiplying $Y(t)$ by $\cos(2\pi f_1(t)t)$ can be used to recover $m_1(t)$ when $f_1(t) \neq f_2(t)$. What happens if $f_1(t) = f_2(t)$?

(b) Let's take a look at the frequency spectrum over time, using `spectrogram`. Does the spectrogram make sense given the channel sequence? Suppose the hopping sequence of the two messages are independent of each other. Do you see a potential problem with this? What is the probability of the two messages NOT colliding during these 10 seconds if the hopping sequences are chosen independently, without memory of its previous hops and with equal likelihood at each of the 20 channels?

(c) Decode both messages, using your knowledge of the encoding seeds. Do you hear something unexpected midway through message 1? What is the phrase, and why do we hear it?

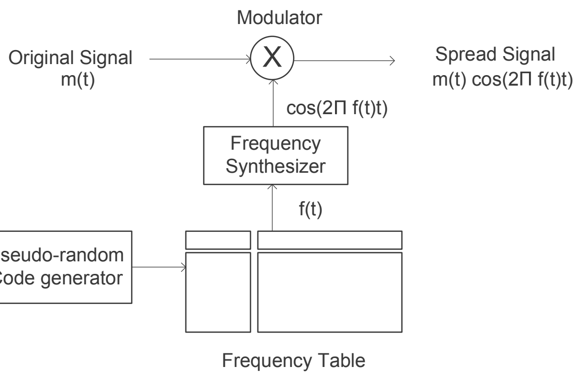
(d) *Eavesdropping (fixed channel).* Assume that an eavesdropper is listening in on a fixed channel. For instance, listen in on channel 19. Qualitatively, what do you hear?

(e) *Eavesdropping (wrong seed).* Assume that as an eavesdropper, you have the wrong seed. So while you are demodulating some frequency hopping pattern, it does not correspond to either message. Qualitatively, what do you hear?

(f) *Jamming (fixed channel).* `Jam1` is a signal designed to jam channel 11. Supposed your enemy is trying to impede your communications by transmitted noise on one of your channels. Look at the spectrogram. What do the two decoded messages sound like? What is the effect of fixed channel jamming?

(g) *Jamming (spread spectrum).* `Jam2` is a signal whose energy is spread across all channels. However, the energy required is proportional to the bandwidth (number of channels), so if we have a fixed amount of energy, the jamming energy per channel drops. Check that `Jam2` and `Jam1` have the same energy (what is the energy?). What do the two decoded messages sound like? What is the effect of spread spectrum jamming? (Is it really as bad as it looks in the spectrogram?)

Note: While we strive to make this a realistic exercise, there are several issues you should consider. First, the sampling frequency `fs` is absurdly high (200 kHz) for a voice transmission. We needed this so that we could emulate a continuous-time operation (demodulation by a cos) in MATLAB. In practice, the sampling would only happen after demodulation and low-pass filtering; this would only require an 8 kHz sampling rate. Second, the message bandwidth is usually much smaller relative to the spread across the spectrum, so it is much harder to observe the entire spectrum to track the hopping pattern (as we could determine from the spectrogram). Third, in this exercise, you could conceivably try all 500 pseudorandom seeds, but in practice you might use a 128-bit seed (as opposed to a 9-bit seed) and a predefined algorithm that operates on the seed (as opposed to a lookup table). Fourth, the message could be encrypted (such as in the voice scrambling problem earlier) prior to transmission for truly secure communcations. Finally, another interpretation of the spread spectrum jamming result is that if there is no jamming, we could transmit at much lower power ("stealth mode") to the point where the message is almost hidden by the ambient, background noise, making it much more difficult to recognize that any transmission is occurring unless you happen to be decoding correctly.



3.16. *Reproducing function*

A real function $f(t)$ is said to reproduce itself under convolution if

$$f(t/a) * f(t/b) = kf(t/c),$$

where k and c in general depend on a and b but the form of the relation is to be valid for any a and b. In words, if such a function is convolved with a similar function having the same shape (i.e., identical except for time scale and possibly amplitude), then the result also has the same shape.

(a) Find an equivalent condition in terms of $\hat{f}(s)$, the Fourier transform of $f(t)$.

(b) Assuming that $\hat{f}(0) = 1$, $\hat{f}'(0) = 0$ and that $\hat{f}''(0)$ exists but is not zero, show that $c^2 = a^2 + b^2$ and $k = |ab/c|$.

(c) Show that the function $f(t) = e^{-\pi t^2}$ reproduces itself under convolution. Are the relationships derived in (b) satisfied for $f(t)$?

3.17. *Applications of the Fourier transform to differential equations*

Let $H(t)$ be the unit step function,

$$H(t) = \begin{cases} 0, & t \leq 0, \\ 1, & t > 0. \end{cases}$$

(a) Using the Fourier transform properties, find a solution for the differential equation $x''(t) + 2x'(t) + 2x(t) = e^{-t}H(t)$. (Thus the right-hand side is a one-sided exponential decay.)

(b) The function $g(t) = e^{-a|t|}$ has the following Fourier transform:

$$G(s) = \frac{2a}{a^2 + 4\pi^2 s^2}.$$

Knowing this transform, demonstrate that one possible solution for the equation $x''(t) - x(t) = -2e^{-t}H(t)$ is

$$x(t) = \int_0^\infty e^{-|t-\tau|} e^{-\tau} d\tau.$$

3.18. *Exponential decay and a common differential equation*[27]

(a) Let $a > 0$ and let $g(t) = e^{-a|t|}$ be the two-sided exponential decay. Find $\mathcal{F}g(s)$. *Hint*: Use the one-sided exponential decay and a reversal.

(b) Let $u(t)$ be the unit step function

$$u(t) = \begin{cases} 1, & t > 0, \\ 0, & t \leq 0. \end{cases}$$

Show that a solution of the differential equation

$$x''(t) - x(t) = -2e^{-t}u(t)$$

is

$$x(t) = \int_0^\infty e^{-|t-\tau|} e^{-\tau} \, d\tau.$$

[27] From S. Arik.

3.19. *Solving the wave equation*

An infinite string is stretched along the x-axis and is given an initial displacement described by a function $f(x)$. It is then free to vibrate. The displacement $u(x,t)$ at a time $t > 0$ and at a point x on the string is described by the *wave equation*

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 u}{\partial x^2}.$$

One often includes physical constants in the equation, e.g., the speed of the wave, but these are suppressed to keep things simple.
Assume that $u(x,0) = f(x)$ and $u_t(x,0) = 0$ (zero initial velocity) and use the Fourier transform to show that

$$u(x,t) = \frac{1}{2}(f(x+t) + f(x-t)).$$

This is d'Alembert's (famous) solution to the wave equation.

3.20. *An application to quantum mechanics*

In quantum mechanics, the state of a free particle is entirely described by its wavefunction $\psi(x,t)$, which evolves according to the time-dependent Schrödinger equation:

$$\frac{\partial \psi(x,t)}{\partial t} = \frac{i}{2m}\frac{\partial^2 \psi(x,t)}{\partial x^2}.$$

(a) If the wavefunction of a particle begins in state $\phi(x)$ (i.e., $\psi(x,0) = \phi(x)$), show that

$$\psi(x,t) = \phi(x) * \sqrt{\frac{m}{2\pi i t}}e^{\frac{imx^2}{2t}}$$

for $t \geq 0$. (You've done problems like this before — treat the Schrödinger equation as a heat equation with a complex constant of proportionality.)

(b) At any given time t, we would like to interpret the square magnitude of the wavefunction, $|\psi(x,t)|^2$, to be the probability density function (pdf) associated with finding the particle at position x. Recall that a function $p(x)$ is a probability density function if

(i) $p(x) \geq 0$,

(ii) $\displaystyle\int_{-\infty}^{\infty} p(x)\,dx = 1$.

Using the result in part (a), show that if $|\phi(x)|^2$ is a pdf, then $|\psi(x,t)|^2$ is a pdf for $t \geq 0$.

3.21. *Cross correlation*

The *cross-correlation* (sometimes just called correlation) of two real-valued signals $f(t)$ and $g(t)$ is defined by

$$(f \star g)(x) = \int_{-\infty}^{\infty} f(y)g(x+y)\,dy\,.$$

Note the notation: $\star$ for cross-correlation. $(f \star g)(x)$ is often described as a measure of how well the values of g, when shifted by x, correlate with the values of f. It depends on x; some shifts of g may correlate better with f than other shifts.

To get a sense of this, think about when $(f \star g)(x)$ is positive (and large) or negative (and large) or zero (or near zero). If, for a given x, the values $f(y)$ and $g(x+y)$ are tracking each other — both positive or both negative — then the integral will be positive and so the value $(f \star g)(x)$ will be positive. The closer the match between $f(x)$ and $g(x+y)$ (as y varies) the larger the integral and the larger the cross-correlation.

In the other direction, if, for example, $f(y)$ and $g(x+y)$ maintain opposite signs as y varies (so are negatively correlated), then the integral will be negative and $(f \star g)(x) < 0$. The more negatively they are correlated, the more negative $(f \star g)(x)$.

Finally, it might be that the values of $f(y)$ and $g(x+y)$ jump around as y varies, sometimes positive and sometimes negative, and it may then be that in taking the integral the values cancel out, making $(f * g)(x)$ near zero. One might say — one does say — that f and g are uncorrelated if $(f \star g)(x) = 0$ for all x.

(a) Cross-correlation is similar to convolution, with some important differences. Show that $f \star g = f^- * g = (f * g^-)^-$. Is it true that $f \star g = g \star f$?

(b) *Cross-correlation and delays.* Show that

$$f \star (\tau_b g) = \tau_b (f \star g)\,,$$

where $(\tau_b f)(t) = f(t-b)$. Why does this make sense, intuitively? What about $(\tau_b f) \star g$?

3.22. *Autocorrelation*

The *autocorrelation* of a real-valued signal $f(t)$ with itself is defined to be

$$(f \star f)(x) = \int_{-\infty}^{\infty} f(y) f(x+y)\, dy\,.$$

This is a measure of how much the values of f are correlated, so how much they match up under shifts. For example, *ideal white noise* has uncorrelated values; thus one definition of white noise is a signal with $f \star f = 0$.

(a) Find $\Pi \star \Pi$, without much work. (Use the relationship between correlation and convolution.)

(b) Show that

$$(f \star f)(x) \le (f \star f)(0).$$

You saw this for Fourier series. To show this, use the *Cauchy-Schwarz* inequality. See Chapter 1 for a discussion of this in the context of general inner products. In the present context the Cauchy-Schwarz inequality states that

$$\int_{-\infty}^{\infty} f(t) g(t)\, dt \le \left\{ \int_{-\infty}^{\infty} f(t)^2\, dt \right\}^{1/2} \left\{ \int_{-\infty}^{\infty} g(t)^2\, dt \right\}^{1/2},$$

with equality holding if and only if $f(t)$ and $g(t)$ are proportional.

(c) Show that
$$\mathcal{F}(f \star f) = |\mathcal{F}f|^2 .$$
You also saw a version of this for Fourier series. For Fourier transforms the result is sometimes known as the *Wiener-Khintchine Theorem*, though they actually proved a more general, limiting result that applies when the transform may not exist.
The quantity $|\mathcal{F}f|^2$ is often called the power spectrum, and the result, $\mathcal{F}(f \star f) = |\mathcal{F}f|^2$, is stated for short as: "The Fourier transform of the autocorrelation is the power spectrum." Its many uses include the study of noise in electronic systems.

(d) *Correlation and radar detection.* Here's an application of the maximum value property, $(f \star f)(t) \leq (f \star f)(0)$, above. A radar signal $f(t)$ is sent out, takes a time T to reach an object, and is reflected back and received by the station. In total, the signal is delayed by a time $2T$, attenuated by an amount, say α, and subject to noise, say given by a function $n(t)$. Thus the received signal, $f_r(t)$, is modeled by
$$f_r(t) = \alpha f(t - 2T) + n(t) = \alpha(\tau_{2T} f)(t) + n(t) .$$
You know $f(t)$ and $f_r(t)$, but because of the noise you are uncertain about T, which is what you want to know to determine the distance of the object from the radar station. You also don't know much about the noise, but one assumption that is made is that the cross-correlation of f with n is constant,
$$(f \star n)(t) = C, \quad \text{a constant.}$$
You can compute $f \star f_r$ (so do that) and you can tell (in practice, given the data) where it takes its maximum, say at t_0. Find T in terms of t_0.

3.23. Compute the autocorrelation function $\Gamma_f(\tau) = (f \star f)(\tau)$ for the one-sided exponential decay function, $f(t) = e^{-at}\, u(t)$. (Here, $u(t)$ is a unit step function.)

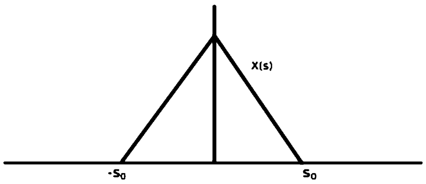
3.24. *Signal phase.*

Let $h(t)$ be a real signal with Fourier transform, $H(s)$. Write $H(s) = |H(s)|e^{i\theta(s)}$, where $\theta(s) = \angle H(s)$ is the *phase*. Assume that $\angle H(s)$ is not identically zero.

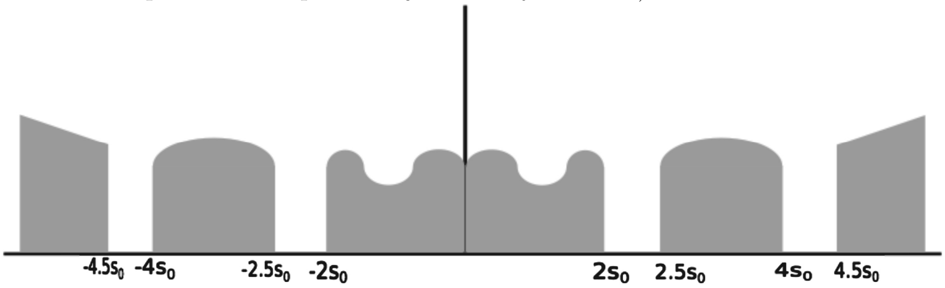
Let $x(t)$ be a real signal whose Fourier transform, $X(s)$, has a phase $\angle X(s)$. Let $x(t)$ undergo the following set of transformations, resulting in the signal $y(t)$:
$$\begin{aligned} g(t) &= h(t) * x(t) , \\ f(t) &= g(t) \star h(t) , \\ y(t) &= f(-t) . \end{aligned}$$
Let $Y(s)$ be the Fourier transform of $y(t)$. What can you say about the phase of $Y(s)$, $\angle Y(s)$, as it relates to $\angle X(s)$? Explain your work.

3.25. *Signal transmission*

You wish to transmit a secret signal $x(t)$ whose Fourier transform is shown in the figure below, centered at the origin.



Unfortunately, the entire frequency spectrum is jammed by Dr. Evil, except the regions from $2s_0$ to $2.5s_0$, $4s_0$ to $4.5s_0$, and the mirror images, $-2.5s_0$ to $-2s_0$ and $-4.5s_0$ to $-4s_0$, as shown in the figure below (the shaded regions indicate parts of the spectrum jammed by Dr. Evil):



(a) Assuming that any mathematical operation can be performed on the signal on the receiver side and assuming that you can also perform any mathematical operation on your side (the transmitter), what must you do (in mathematical terms) to the signal for it to not be affected by the jamming during the transmission? What must the receiver do to reconstruct the original signal? Your answer may be in either the time or the frequency domain.

(b) You learn that Dr. Evil has stopped jamming the frequencies he was previously jamming but is now eavesdropping on those same frequencies, i.e., the shaded frequencies in the above figure. Is your transmission network of part (a) adequate to hide the signal from Dr. Evil? If not, what modifications would you make to it in order to hide the transmission from Dr. Evil's eavesdropping? Your answer may be in either the time or the frequency domain.

3.26. *Convolution theorem for the Hartley transform*

Recall the Hartley transform that we studied previously. It is defined for a real function $f(t)$ as

$$\mathcal{H}f(s) = \int_{-\infty}^{\infty} (\cos 2\pi st + \sin 2\pi st) f(t)\, dt.$$

(a) Find the convolution theorem for the Hartley transform; i.e., find an expression for $\mathcal{H}(f * g)(s)$ in terms of $\mathcal{H}f(s)$ and $\mathcal{H}g(s)$.

Hint: You may find the following identities you derived useful:

$$\mathcal{H}f = \frac{1+i}{2}\mathcal{F}f + \frac{1-i}{2}\mathcal{F}f^- \quad \text{and} \quad \mathcal{H}f^- = \frac{1+i}{2}\mathcal{F}f^- + \frac{1-i}{2}\mathcal{F}f.$$

(It's not as simple as for the Fourier transform!)

(b) If the functions f and g have even symmetry, what can you say about $\mathcal{H}(f * g)(s)$?

3.27. *Characteristic functions*

Recall that a probability density function (pdf) satisfies the following.

(1) $p(x)$ is real valued.
(2) $p(x) \geq 0$ for all x.
(3) $\int_{-\infty}^{\infty} p(x)\,dx = 1$.

In probability theory one uses the term *characteristic function* to refer to the *Fourier transform* of a probability density function.

If $\psi(s)$ is a characteristic function corresponding to pdf $p(x)$, show that the following are also characteristic functions. Brief explanations are sufficient. The key is to show that $p_i(x) = \mathcal{F}^{-1}\psi_i(x)$ is also a pdf. If $p_i(x)$ is a pdf, then $\psi_i(s)$ is a characteristic function.

(a) $\psi_1(s) = e^{-2\pi i b s}\psi(s)$, where b is a real constant.
(b) $\psi_2(s) = \psi(as)$, where $a > 0$.
(c) $\psi_3(s) = \overline{\psi}(s)$.
(d) $\psi_4(s) = \mathrm{Re}[\psi(s)]$.
(e) $\psi_5(s) = |\psi(s)|^2$.

3.28. *Fourier transforms of probability densities*

We saw that the Gaussian is an example of a probability density whose Fourier transform is identical to the function; i.e., it satisfies $\mathcal{F}f(s) = f(s)$. In this problem we investigate if it is possible that $\mathcal{F}f(s) = af(s)$ for some complex number a.

Recall that a probability density p is a nonnegative function which integrates to 1.

(a) Does there exist a probability density p whose Fourier transform is a scaling of itself, i.e., $\mathcal{F}p(s) = ap(s)$, for $a \neq 1$?
(b) Does there exist a probability density p whose Fourier transform is the negative of itself, i.e., $\mathcal{F}p(s) = -p(s)$?

Thus, the only possible value of a is 1.

3.29. *Gaussians and convolution*

A continuous random variable X is said to be *normally distributed* with mean μ and variance σ^2 if its probability density function (pdf) is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}.$$

A common shorthand notation for this is $X \sim N(\mu, \sigma)$, read as "X is drawn from a normal distribution of mean μ and variance σ^2."

Given two independent random variables $X_1 \sim N(\mu_1, \sigma_1)$ and $X_2 \sim N(\mu_2, \sigma_2)$, find the pdf of their sum $X_1 + X_2$ by finding the convolution of their pdfs.

How are the mean and variance of $X_1 + X_2$ related to the mean and variance of X_1 and X_2?

3.30. *Infinitely divisible probability distributions*

Let $f(x)$ be a probability distribution function for a random variable X. We say that $f(x)$ is *infinitely divisible* if for every positive integer n there are n independent, identically distributed random variables $X_1, X_2, \ldots, X_n$ such that $X_1 + X_2 + \cdots + X_n$ has the same distribution as X, i.e., has pdf $f(x)$. (Note that this does *not* mean that the individual X_i's have distribution $f(x)$.)

This may seem to be a very strong condition, but many pdf's that come up in practice are infinitely divisible.

Recall that if we add two independent random variables, the distribution of the sum is the convolution of their distributions.

(a) Show that if $f(x)$ is infinitely divisible, then for every n, there exists a distribution $f_{(n)}(x)$ such that

$$\mathcal{F}f(s) = \left(\mathcal{F}f_{(n)}(s)\right)^n .$$

(b) The Gaussian distribution, with mean μ and variance σ^2, is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} .$$

Use part (a) to show that the Gaussian distribution is infinitely divisible. What are the $f_{(n)}(x)$? Is the mean and variance of $f_{(n)}(x)$ what you might expect?

Hint: You may find it easier to first consider $\mu = 0$ and $\sigma^2 = 1$ and then use the scaling and shift theorems for arbitrary μ, σ^2.

(c) The Cauchy distribution is

$$f(x) = \frac{1}{\pi\gamma\left(1 + \left(\dfrac{x - x_0}{\gamma}\right)^2\right)};$$

it involves the parameters x_0 and $\gamma > 0$. This distribution is symmetric about x_0, but neither its mean nor its variance is finite. Still, we can show that the Cauchy distribution is infinitely divisible. What are the $f_{(n)}(x)$, and how are these distributions related to the original $f(x)$?

Hint: Again, you may find it easier to first consider $x_0 = 0, \gamma = 1$. Have we seen this expression before? Look at the two-sided exponential decay.

3.31. *Plotting distributions using* MATLAB.

Suppose you roll a fair, six-sided die many times.

(a) What do you think the average of the numbers should be as the number of trials increases?

(b) Let X_i be the random variable denoting the ith roll of the die. Then the empirical average after N rolls of the die is equal to

$$\frac{1}{N}\sum_{i=1}^{N} X_i.$$

Use MATLAB to plot the distribution of the empirical average, for $N = 2$, 10, and 100.

3.32. *Approximating π: Brad learns a lesson*

Brad was doing some exercises to learn the Julia programming language. One exercise was to use Monte Carlo techniques to approximate π via the ratio of two areas. Consider the square with $-1 \leq x \leq 1$, $-1 \leq y \leq 1$, and the inscribed unit circle $x^2 + y^2 \leq 1$. Then

$$\pi = 4(\text{area of the circle})/\text{area of the square}.$$

This ratio is approximated by generating a large number of points in the square, randomly chosen and uniformly distributed, and computing

$$4(\text{number of points landing in the circle})/\text{total points in the square}.$$

Here's the code that Brad wrote, followed by the code that the Julia people suggested.

```
N=100000 # Specifies number of random points in the unit square
M=0
# Initialize the variable that counts the points in the square
 that are also in the unit circle
for k=1:N
point=rand(2)-rand(2)
# A random point (2-vector) whose entries can be between -1
 and 1.
t=dot(point,point)  # Its length, squared
if  t <=1
M=M+1
end
end
println(4M/N)

N=100000
M=0
for k=1:N
 point=[2*rand()-1,2*rand()-1]
 # A different way of generating a random point in the square.
t=dot(point,point)
if  t <=1
M=M+1
end
end
println(4M/N)
```

The function

```
rand(2)
```

returns pair of random numbers between 0 and 1. The function

```
rand()
```

returns a single random number between 0 and 1. The function

```
dot
```

returns the dot product.

A few trials of Julia's code gave the numbers 3.14084, 3.13944, 3.13608, 3.13676 whereas a few trials of Brad's code gave 3.90108, 3.89736, 3.89872, 3.89532. Even bumping up to a million points didn't help Brad; Julia's code gave 3.141872 and Brad's code gave 3.898568. What did Brad do wrong?

3.33. *An example of the Cauchy distribution*

Stand a fixed distance from a wall, hold a laser horizontally, and point it at the wall at a random angle. The locations of the burn marks on the wall, along any line, are a random variable X whose values are distributed according to

$$p(x) = \frac{1}{\pi(1+x^2)}\,.$$

This is a *Cauchy distribution* and arises in many applications.

(a) Find the distribution for the average of two (independent) series of such burn marks, i.e., $(1/2)(X_1 + X_2)$, where X_1 and X_2 are distributed as above. For this you'll likely need the Fourier transform of a two-sided decaying exponential and the fact that if $p(x)$ is the distribution for a random variable X, then $(1/a)X$ has distribution $ap(ax)$.

(b) Without any further work, what happens if we average N series of such burn marks? Note that the Central Limit Theorem does *not* apply because the mean and variance are infinite.

People's intuition is often that if they perform the same experiment many times and average the results, then they're getting close to the "actual" or "ideal" answer. What do your results above say about this intuition?

3.34. *More on convolution and probability*[28]

Let $p(x)$ be the probability density of a uniform random variable on $[0, 1]$; i.e.,

$$p(x) = \begin{cases} 1, & 0 \le x \le 1\,, \\ 0, & \text{otherwise.} \end{cases}$$

Let $p^{(n)}$ be the n-fold convolution of p given by

$$p^{(n)} = \underbrace{p * p * p * \ldots * p}_{n \text{ times}}\,.$$

[28] From A. Siripuram.

(a) Sketch a graph of $p^{(2)}(x)$.
Note that repeated convolutions keep widening the signal: $p(x)$ is nonzero on $[0,1]$, $p^{(2)}(x)$ is nonzero on $[0,2]$, and in general $p^{(n)}(x)$ is nonzero on $[0,n]$. For this problem, we are only interested in computing $p^{(n)}(x)$ on $[0,1]$, even though it is nonzero on $[0,n]$, a much larger interval.

(b) Show that

$$p^{(n)}(x) = \frac{x^{n-1}}{(n-1)!} \quad \text{for } 0 \le x \le 1\,.$$

We will use this later in the problem. Now suppose we have a container of capacity 1 unit. Water is poured into the container each second (the amount of water poured is random), and we want to determine when the container overflows. Let V_i be the amount of water added to the container in the ith second. Assume that V_i is a random variable distributed uniformly in $[0,1]$; i.e., it has a probability density(pdf) of $p(x)$. The container overflows at the Nth second if $V_1 + V_2 + \cdots + V_N > 1$. We want to find $\mathbb{E}(N)$, the expected time after which the container overflows. Now, we have

$$\begin{aligned}\mathbb{E}(N) &= \sum_{n=0}^{\infty} \text{Prob}(N > n)\\ &= 1 + \sum_{n=1}^{\infty} \text{Prob}(V_1 + V_2 + \cdots + V_n \le 1)\\ (*) \qquad &= 1 + \sum_{n=1}^{\infty} \int_0^1 \phi_n(x)dx, \ \text{where } \phi_n(x) \text{ is the pdf of } V_1 + V_2 + \cdots + V_n.\end{aligned}$$

(You are not asked to prove any of the equalities above.) You can assume that the amounts of water added in any second are independent of each other; i.e., the V_i are independent.

(c) Complete the argument from (*) to show that $\mathbb{E}(N) = e$; i.e., the container overflows on average after e seconds.

3.35. *Estimating parameters with minimum spectral samples*[29]

You are given a signal f of the form

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)^2}{2\sigma^2}}$$

and want to estimate the parameters μ and σ. The constraint is that we are only allowed to take measurements of the spectrum $\mathcal{F}f$.

Assume you are given a black box which, on input s_0, outputs $\mathcal{F}f(s_0)$. The idea is to take measurements at as few s_0 as possible to find μ and σ. For the purposes of this problem, you can assume $\mathcal{F}f(s_0)$ is given to you with infinite precision. You can also assume that μ is between -1 and 1, even though you do not know its exact value.

[29] From A. Siripuram.

(a) How will you estimate μ and σ with the procedure described above? What is the least number of measurements required?
(b) How many measurements would be needed if you were *not* given that $\mu \in (-1, 1)$?

3.36. *Benford's Law: A logarithmic central limit theorem?* (This problem is not as long as it looks!)

Benford's Law is an empirical observation about the distribution of leading digits in a list of numbers. You might think that the leading digits of naturally occurring measurements are uniformly distributed, but in a very large number of very diverse situations this is not the case. The number 1 occurs as the leading digit about 30% of the time while the number 9 occurs as the leading digit only about 5% of the time.

Benford's Law says that the leading digits follow a logarithmic distribution, that for $n = 1, 2, \ldots, 9$ the fraction of numbers with leading digit n is approximately $\log(n+1) - \log n$ (log base 10 here). Thus if $n = 1$, this is $\log 2 - \log 1 = .30$ and for $n = 9$ it is $\log 9 - \log 8 = .05$, hence the numbers 30% and 5% mentioned above.

There are many fascinating discussions of this observation and even some explanations. In recent years it has been used as a tool in discovering fraud, for peoples' tendency in fabricating lists of numbers is to think that the leading digits should be uniformly distributed, and this can be checked against what Benford's Law indicates the distribution should be. A broadcast of the NPR program Radiolab (highly recommended) had a segment on this use of Benford's Law; see their webpage for a podcast:

`http://www.radiolab.org/2009/nov/30/from-benford-to-erdos/`

See also

`http://www.washingtonpost.com/wp-dyn/content/article/2009/06/20/AR2009062000004.html`

for a recent article. The mathematician Terrence Tao has also compared several such empirical laws:

`http://terrytao.wordpress.com/2009/07/03/benfords-law-zipfs-law-and-the-pareto-distribution/`.

The Central Limit Theorem tells us that the sum of independent, identically distributed random variables will be distributed like a Gaussian. Why a log in Benford's Law? An approach to Benford's Law via Fourier series has been given in a paper by J. Boyle[30] upon which this discussion is based. In his paper, Boyle gives as motivation the following statement:

> Frequently, naturally occurring data can be thought of as products or quotients on random variables. For example . . . if a city has an initial population P_0 and grows by r_i% in year i, then the population in n years is $P_n = P_0(1+r_1)(1+r_2)\cdots(1+r_n)$, a product of a number of random variables.

[30] J. Boyle, An application of Fourier series to the most significant digit problem, Amer. Math. Monthly **101** (1994), 879–886.

The setup for the study goes as follows. We write numbers in scientific notation:

$$x = M(x) \times 10^{N(x)}$$

where the *mantissa* $M(x)$ satisfies $1 \le M(x) < 10$. The leading digit of x is then $\lfloor M(x) \rfloor$ so we want to understand the distribution of $M(x)$. For a given collection of data to say that the $M(x)$ has a log distribution is to say

$$\text{Prob}(M(x) \le M) = \log M.$$

Then

$$\text{Prob}(M \le M(x) \le M+1) = \log(M+1) - \log M.$$

In turn, $M(x)$ has a log distribution if and only if $\log M(x)$ has a uniform distribution. The goal is thus to understand when $\log M(x)$ has a uniform distribution.

If we multiply two numbers, their mantissas multiply and

$$M(xy) = \begin{cases} M(x)M(y), & 1 \le M(x)M(y) < 10, \\ \frac{M(x)M(y)}{10}, & 10 \le M(x)M(y) < 100. \end{cases}$$

If we take the log, this can be expressed more conveniently as

$$\log M(xy) = \log M(x) + \log M(y) \mod 1.$$

The following theorem of Boyle gives a justification for Benford's Law for the distribution of the leading digit when continuous random variables are multiplied:

Theorem 3.1. *Let $X_1, X_2, \ldots, X_N$ be independent, identically distributed continuous random variables on the interval $[0, 1]$. Let*

$$Z = \sum_{k=1}^{N} X_k \mod 1.$$

Then as $N \to \infty$ the probability distribution of Z approaches a uniform distribution.

Boyle goes on to extend Theorem 3.1 to quotients and powers of random variables, but we will not pursue this.

We're going to step through the proof, providing explanations for some steps and asking you for explanations in others.

The probability density functions we work with are defined on $[0, 1]$, and we assume that they are continuous and also square integrable. Moreover, since we use arithmetic mod 1 (to define Z), it is natural to assume they are periodic of period 1, and the machinery of Fourier series is then available.

Prove the following general lemma, which will be used in the proof of the theorem.

Lemma 3.1. *If $p(x)$ is a continuous, square integrable probability density on $[0, 1]$, periodic of period 1, then*
(a) $\hat{p}(0) = 1$, (b) $|\hat{p}(n)| \le 1$, $n \ne 0$.

Actually, a stronger statement than (b) holds — and that's what we'll really need, and you may use it — but we will not ask you for the proof. Namely,

$$\sup_{n\neq 0} |\hat{p}(n)| \leq A < 1$$

for some constant A. Here sup is the supremum, or least upper bound, of the Fourier coefficients for $n \neq 0$. The statement means that other than for the zeroth coefficient the $\hat{p}(n)$ stay strictly less than 1 (uniformly).

We turn to the proof of the theorem. Suppose that $p(x)$ is the common probability density function for the X_i and let $f(x)$ be the probability distribution for $Z = \sum_{k=1}^{N} X_k \mod 1$. What we (you) will show is that

$$\|f-1\|^2 = \int_0^1 (f(x)-1)^2\,dx \to 0 \quad \text{as } N \to \infty.$$

This says that $f(x)$ tends to the uniform distribution, 1, with respect to the norm on the space $L^2([0,1])$ of square integrable functions. It turns out that this is sufficient, as far as probabilities are concerned, to conclude that Z has the uniform distribution in the limit.

The steps in establishing the statement above, with prompts for you to fill in the reasons, are

$$\begin{aligned}\int_0^1 (f(x)-1)^2\,dx &= \sum_{n\neq 0} |\hat{f}(n)|^2 \quad \text{(why?)}\\ &= \sum_{n\neq 0} |\hat{p}(n)|^{2N} \quad \text{(why?)}\\ &= \sum_{n\neq 0} |\hat{p}(n)|^2\,|\hat{p}(n)|^{2N-2}\\ &\leq A^{2N-2} \sum_{n\neq 0} |\hat{p}(n)|^2 \quad \text{(why?)}.\end{aligned}$$

Now explain why

$$\int_0^1 (f(x)-1)^2\,dx \to 0 \quad \text{as } N \to \infty.$$

3.37. *Poisson transform*[31]

The *Poisson distribution* (in probability) is used to model the probability of a given number of events occurring over a period of time. For example, the number of phone calls arriving at a call center per minute is modeled using the Poisson distribution:

$$\text{Prob(number of calls} = k) = \frac{\lambda^k}{k!}e^{-\lambda}, \quad k = 0, 1, 2, \ldots,$$

where λ is the number of phone calls per minute. Now, the rate of arrivals of phone calls typically varies throughout the day, so λ itself should be considered

[31] From A. Siripuram.

to come from a random variable, Λ, say with probability density function $p(\lambda)$. In that case, the probability Prob(number of calls $= k$) is more realistically modeled by the function

$$P(k) = \int_0^\infty \frac{\lambda^k}{k!} e^{-\lambda} p(\lambda)\, d\lambda, \quad k = 0, 1, 2, \ldots .$$

We will call $P(k)$ the Poisson transform of $p(\lambda)$.

It only makes sense to consider $\lambda \geq 0$, but it is convenient for the discussion below to allow $\lambda < 0$ by declaring $p(\lambda) = 0$ when $\lambda < 0$. Then we can write

$$P(k) = \int_{-\infty}^\infty \frac{\lambda^k}{k!} e^{-\lambda} p(\lambda)\, d\lambda, \quad k = 0, 1, 2, \ldots .$$

Note that unlike the Fourier transform, the Poisson transform takes a function of a real variable and produces a function defined on the natural numbers $0, 1, 2, \ldots$.

(a) We would like to know that the Poisson transform is invertible; i.e., given the numbers $P(0), P(1), P(2), \ldots$, it is possible to find $p(\lambda)$. This is a relevant question, as the values $P(k)$ can be measured experimentally from a large set of observations.

For this, given the values $P(k)$ define the function

$$Q(s) = \sum_{k=0}^{\infty} (-2\pi i s)^k P(k).$$

Using the definition of $P(k)$ show that $Q(s)$ is the Fourier transform of $q(\lambda) = e^{-\lambda} p(\lambda)$, and so find $p(\lambda)$ in terms of $Q(s)$.

Hint: Recall the Taylor series for the exponential function

$$e^a = \sum_{k=0}^{\infty} \frac{a^k}{k!}.$$

(b) Suppose there are two independent factors contributing to the arrival rate of phone calls, so that $\Lambda = \Lambda_1 + \Lambda_2$. Recall that for the corresponding probability density functions we have

$$p = p_1 * p_2.$$

We want to find the relationship between the Poisson transforms, $P(k)$ for p, and $P_1(k), P_2(k)$ for p_1, p_2.

(i) As above, let $q(\lambda) = e^{-\lambda} p(\lambda)$ and, correspondingly, $q_1(\lambda) = e^{-\lambda} p_1(\lambda)$, $q_2(\lambda) = e^{-\lambda} p_2(\lambda)$. Show that

$$q_1 * q_2 = q \quad \text{(use the definition of convolution)}.$$

(ii) Using $Q(s) = Q_1(s) Q_2(s)$ for the Fourier transforms, deduce that

$$P(k) = \sum_{m=0}^{k} P_1(m) P_2(k - m),$$

a discrete convolution of P_1 and P_2!

3.38. *Deriving the solution of the Black-Scholes equation for financial derivatives*[32] (This problem is also not as long as it looks.)

As per Wikipedia: A call option is a financial contract between two parties, the buyer and the seller of this type of option. The buyer of the call option has the right, but not the obligation, to buy an agreed quantity of a particular commodity (the underlying) from the seller of the option at the expiration date for a certain price (the strike price). The seller is obligated to sell the commodity should the buyer so decide. The buyer pays a fee (called a premium) for this right. The buyer of a call option wants the price of the underlying instrument to rise in the future while the seller expects that it will not. A European call option allows the buyer to buy the option only on the option expiration date.

The partial differential equation to obtain a Black-Scholes European call option pricing formula is

$$(*) \qquad rC = C_t + \frac{1}{2}\sigma^2 S^2 C_{SS} + rS\, C_S \quad \text{for } t \in (0,T), S \in (0,\infty).$$

Here $C(S,t)$ is the value of the call option at time t with expiration time T, strike price K, and risk-free interest rate r. Let S be the price of the underlying call option and let the volatility of the price of the stock be σ^2. Note that $C_t = \frac{\partial C(S,t)}{\partial t}$, $C_S = \frac{\partial C(S,t)}{\partial S}$, and $C_{SS} = \frac{\partial^2 C(S,t)}{\partial S^2}$. The side conditions of this partial differential equation are

$$\begin{aligned} &C(S,T) = \max(0, S-K) \quad \text{for } S \in (0,\infty),\\ &C(0,t) = 0,\\ &C(S,t) \to S - Ke^{-r(T-t)} \quad \text{as } S \to \infty \text{ for } t \in (0,T). \end{aligned}$$

A judicious change of variables allows us to find the solution to the Black-Scholes equation using a more well-known partial differential equation, namely the heat equation on an infinite rod. First, the Black-Scholes equation is reduced to a (parabolic) differential equation

$$v_\tau = v_{xx} + (k-1)v_x - kv \quad \text{for } x \in (-\infty,\infty), \tau \in (0, T\sigma^2/2)$$

using the change of variables

$$C(S,t) = Kv(x,\tau), \quad S = Ke^x, \quad t = T - \frac{2\tau}{\sigma^2}, \quad k = 2r/\sigma^2.$$

Next, this partial differential equation is reduced to the heat equation on an infinite rod using the change of variables $v(x,\tau) = e^{\alpha x + \beta\tau}u(x,\tau)$:

$$(**) \qquad \begin{aligned} &u_\tau = u_{xx} \quad \text{for } x \in (-\infty,\infty) \text{ and } \tau \in (0, T\sigma^2/2),\\ &u(x,0) = \max(0, (e^{(k+1)x/2} - e^{(k-1)x/2})) \quad \text{for } x \in (-\infty,\infty),\\ &u(x,\tau) \to 0 \quad \text{as } x \to \pm\infty \text{ for } \tau \in (0, T\sigma^2/2), \end{aligned}$$

where $\alpha = (1-k)/2$ and $\beta = -(k+1)^2/4$. You do not need to verify that equation (*) reduces to (**) using the given change of variables.

[32] From Raj Bhatnagar.

(a) Verify that the solution of the partial differential equation (**) is given by a function $u(x,\tau)$ of the form $f(x) * \frac{1}{2\sqrt{\pi\tau}} e^{-\frac{x^2}{4\tau}}$. Find the function $f(x)$.

(b) Show that

$$u(x,\tau) = e^{(k+1)x/2+(k+1)^2\tau/4}\Phi\left(\frac{x+(k+1)\tau}{\sqrt{2\tau}}\right) - e^{(k-1)x/2+(k-1)^2\tau/4}\Phi\left(\frac{x+(k-1)\tau}{\sqrt{2\tau}}\right),$$

where $\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-w^2/2}dw$ is the cumulative distribution function of a Gaussian random variable with mean 0 and variance 1.
(*Hint*: You may use $I_\alpha = \frac{1}{\sqrt{4\pi\tau}}\int_0^\infty e^{-\frac{(x-u)^2}{4\tau}+\alpha u}du = e^{\alpha x+\alpha^2\tau}\Phi(\frac{x+2\tau\alpha}{\sqrt{2\tau}})$.)

(c) Reverse the change of variables to get the solution of the original partial differential equation (*) given by

$$C(S,t) = \Phi(w) - Ke^{-r(T-t)}\Phi(w-\sigma\sqrt{T-t}),$$

where $w = \frac{\log(S/K)+(r+\sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$. This is the Black-Scholes European call option pricing formula.
Note that the reverse change of the variables u, x, and τ to F, S, and t requires $v(x,\tau) = e^{\alpha x+\beta\tau}u(x,\tau)$, $\alpha = (1-k)/2$, $\beta = -(1+k)^2/4$, $C(S,t) = Kv(x,\tau)$, $x = \log\left(\frac{S}{K}\right)$, $\tau = \frac{\sigma^2(T-t)}{2}$, $k = 2r/\sigma^2$.

(d) What is the price $C(S,t)$ of a European call option on a stock when the stock price S is 62 euros per share, the strike price K is 60 euros per share, the continuously compounded interest rate r is 10%, the stock's volatility σ is 20% per year, and the exercise time is five months ($T = 5/12$)? What is the price at $t = 0$? (*Hint*: In MATLAB, you may use function 'qfunc(x)' to compute $Q(x) = 1 - \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty e^{-w^2/2}dw$.)
The Black-Scholes model is widely used in practice to estimate the prices of call options in financial markets. This is mainly due to the existence of a concrete, closed-form solutions for the price of the options given the parameters in the Black-Scholes formulas. The only trouble seems to be the estimation of the parameters, especially the estimation of the volatility σ from historical data. Robert Merton and Myron Scholes were awarded the Nobel Prize in Economics in 1997 to honor their contributions to option pricing!

Chapter 4

Distributions and Their Fourier Transforms

4.1. The Day of Reckoning

We've been playing a little fast and loose with the Fourier transform, applying Fourier inversion, appealing to duality, and all that. "Fast and loose" is an understatement, but it's also true that we haven't done anything wrong. All of our formulas and all of our applications have been correct, if not fully justified. Nevertheless, we have to come to terms with some fundamental questions. It will take us some time, but in the end we will have settled on a very wide class of signals with these properties:

- The allowed signals include Π's, Λ's, δ's, unit steps, ramps, sines, cosines, and other standard signals that the world's economy depends on.
- The Fourier transform and its inverse are defined for all of these signals.
- Fourier inversion works.

These are the three most important features of the development to come, but we'll also reestablish some of our specific results and as an added benefit we'll even finish off differential calculus!

4.1.1. A too simple criterion.

It's not hard to write down an assumption on a function that guarantees the existence of its Fourier transform and even implies a little more than existence.

- If $\displaystyle\int_{-\infty}^{\infty} |f(t)|\,dt < \infty$, then $\mathcal{F}f$ exists and is continuous.

The same result holds for $\mathcal{F}^{-1}f$, and the argument below applies to both $\mathcal{F}f$ and $\mathcal{F}^{-1}f$.

"Existence" means that the integral defining the Fourier transform converges. This follows from

$$\begin{aligned}|\mathcal{F}f(s)| &= \left|\int_{-\infty}^{\infty} e^{-2\pi ist} f(t)\,dt\right| \\ &\le \int_{-\infty}^{\infty} |e^{-2\pi ist}|\,|f(t)|\,dt = \int_{-\infty}^{\infty} |f(t)|\,dt < \infty\,.\end{aligned}$$

Here we've used that the magnitude of the integral is less that the integral of the magnitude.[1] There's actually something to say here, and while it's not complicated, I'd just as soon defer this and other comments on general facts on integrals to Section 4.3. Continuity is the little extra information we get beyond existence. Continuity can be deduced as follows: for any s and s' we have

$$\begin{aligned}|\mathcal{F}f(s) - \mathcal{F}f(s')| &= \left|\int_{-\infty}^{\infty} e^{-2\pi ist} f(t)\,dt - \int_{-\infty}^{\infty} e^{-2\pi is't} f(t)\,dt\right| \\ &= \left|\int_{-\infty}^{\infty} (e^{-2\pi ist} - e^{-2\pi is't}) f(t)\,dt\right| \\ &\le \int_{-\infty}^{\infty} |e^{-2\pi ist} - e^{-2\pi is't}|\,|f(t)|\,dt\,.\end{aligned}$$

As a consequence of $\int_{-\infty}^{\infty} |f(t)|\,dt < \infty$ we can take the limit as $s' \to s$ inside the integral.[2] If we do that, then $|e^{-2\pi ist} - e^{-2\pi is't}| \to 0$, and hence,

$$|\mathcal{F}f(s) - \mathcal{F}f(s')| \to 0 \quad \text{as } s' \to s\,.$$

This says that $\mathcal{F}f(s)$ is continuous.

Functions satisfying

$$\int_{-\infty}^{\infty} |f(t)|\,dt < \infty$$

are said to be in $L^1(\mathbb{R})$. One also says simply that $|f(t)|$ is *integrable*, or that $f(t)$ is absolutely integrable. The "L" is for Lebesgue, as for the square integrable functions $L^2([0,1])$ that we met in connection with Fourier series. It's Lebesgue's generalization of the Riemann integral that's intended by invoking the notation $L^1(\mathbb{R})$. The "1" refers to the fact that we integrate $|f(t)|$ to the first power. The L^1-norm of f is defined to be

$$\|f\|_1 = \int_{-\infty}^{\infty} |f(t)|\,dt\,.$$

No inner products or orthogonality for $L^1(\mathbb{R})$. If you're wondering about $L^2(\mathbb{R})$, wait till the end of the chapter.

[1] Magnitude, not absolute value, because the integral is a complex number.

[2] A general fact. We've done this before and will do it again, and I will say a little more about it in Section 4.3. But save yourself for other things and let some of these general facts ride without insisting on complete justifications — such justifications creep in everywhere once you let the rigor police back on the beat.

In this terminology the criterion says that if $f \in L^1(\mathbb{R})$, then $\mathcal{F}f$ exists and is continuous, and using the norm notation we can write

$$|\mathcal{F}f(s)| \leq \|f\|_1 \quad \text{for } -\infty < s < \infty.$$

This is a handy way to write the inequality.

While the result on existence (and continuity) holds for both $\mathcal{F}f$ and $\mathcal{F}^{-1}f$, we haven't said anything about Fourier inversion or duality, and no such statement appears in the criterion. Let's look right away at a test case.

The very first example we computed, and still an important one, is the Fourier transform of Π. We found directly that

$$\mathcal{F}\Pi(s) = \int_{-\infty}^{\infty} e^{-2\pi ist}\Pi(t)\,dt = \int_{-1/2}^{1/2} e^{-2\pi ist}\,dt = \operatorname{sinc} s\,.$$

No problem there, no problem whatsoever. Note that the criterion applies to guarantee existence, for Π is surely in $L^1(\mathbb{R})$:

$$\int_{-\infty}^{\infty} |\Pi(t)|\,dt = \int_{-1/2}^{1/2} 1\,dt = 1\,.$$

Furthermore, the transform $\mathcal{F}\Pi(s) = \operatorname{sinc} s$ is continuous, as it should be.[3] That's worth remarking on: although the signal jumps (Π has a discontinuity), the *Fourier transform* does not. Make this part of your intuition on the Fourier transform vis à vis the signal.

Many of the other examples we worked with are L^1-functions — the triangle function, the exponential decay (one- or two-sided), Gaussians — so our computations of the Fourier transforms in those cases were perfectly justifiable (and correct).

Appealing to duality we then reached the happy conclusion that $\mathcal{F}\operatorname{sinc} s = \Pi(s)$. But without invoking duality, i.e., using directly the formula,

$$\mathcal{F}\mathrm{sinc}(s) = \int_{-\infty}^{\infty} e^{-2\pi ist}\operatorname{sinc} t\,dt\,,$$

we have a problem. The sinc function *does not* satisfy the integrability criterion. It is my sad duty to inform you that

$$\int_{-\infty}^{\infty} |\operatorname{sinc} t|\,dt = \infty\,.$$

I'll give you two ways of seeing the failure of $|\operatorname{sinc} t|$ to be integrable. First, if sinc did satisfy the criterion $\int_{-\infty}^{\infty} |\operatorname{sinc} t|\,dt < \infty$, then its Fourier transform would be continuous. But its Fourier transform, which *has* to come out to be Π, is *not* continuous. Or, if you don't like that, here's a direct argument. We can find infinitely many intervals where $|\sin \pi t| \geq 1/2$; this happens when t is between $1/6$ and $5/6$, and that repeats for infinitely many intervals, for example on

---

[3]Excuse the mathematical fine point: $\operatorname{sinc} s$ is continuous because of the famous limit $\lim_{s\to 0} \sin \pi s/\pi s = 1$.

$I_n = [\frac{1}{6} + 2n, \frac{5}{6} + 2n]$, $n = 0, 1, 2, \ldots$, because $\sin \pi t$ is periodic of period 2. The I_n all have length 2/3. On I_n we have $|t| \le \frac{5}{6} + 2n$, so

$$\frac{1}{|t|} \ge \frac{1}{5/6 + 2n}$$

and

$$\int_{I_n} \frac{|\sin \pi t|}{\pi |t|}\, dt \ge \frac{1}{2\pi} \frac{1}{5/6 + 2n} \int_{I_n} dt = \frac{1}{2\pi} \frac{2}{3} \frac{1}{5/6 + 2n}\,.$$

Then

$$\int_{-\infty}^{\infty} \frac{|\sin \pi t|}{\pi |t|}\, dt \ge \sum_n \int_{I_n} \frac{|\sin \pi t|}{\pi |t|}\, dt = \frac{1}{3\pi} \sum_{n=1}^{\infty} \frac{1}{5/6 + 2n} = \infty\,.$$

Yes, $|\operatorname{sinc} t| = |\sin \pi t / \pi t|$ tends to 0 as $t \to \pm\infty$, but not fast enough to make the integral of $|\operatorname{sinc} t|$ converge.

This is the most basic example in the theory! It's not clear that the integral defining the Fourier transform of sinc exists — at least it doesn't follow from the criterion. Doesn't this bother you? Isn't it a little embarrassing that multibillion dollar industries seem to depend on integrals that don't converge?

In fact, there isn't so much of a problem with either Π or sinc. It is true that

$$\int_{-\infty}^{\infty} e^{-2\pi i s t} \operatorname{sinc} s\, ds = \begin{cases} 1, & |t| < \frac{1}{2}, \\ \frac{1}{2}, & |t| = \frac{1}{2}, \\ 0, & |t| > \frac{1}{2}. \end{cases}$$

However showing this — evaluating the improper integral that defines the Fourier transform — requires special arguments and techniques. The sinc function oscillates, as do the real and imaginary parts of the complex exponential, and integrating $e^{-2\pi i s t} \operatorname{sinc} s$ involves enough cancellations for the limit

$$\lim_{\substack{a \to -\infty \\ b \to \infty}} \int_a^b e^{-2\pi i s t} \operatorname{sinc} s\, ds$$

to exist. This is laid out for you in the problems. Thus Fourier inversion, and duality, can be pushed through in this case.[4]

The truth is that cancellations that occur in the sinc integral or in its Fourier transform are a very subtle and dicey thing. Such risky encounters are to be approached with respect, or maybe to be avoided. We'd like a more robust, trustworthy theory, and it seems that L^1-integrability of a signal is just too simple a criterion on which to build.

There are problems with other, simpler functions. Take, for example, the signal $f(t) = \cos 2\pi t$. As it stands now, this signal does not even *have* a Fourier transform — does not have a spectrum! — for the integral

$$\int_{-\infty}^{\infty} e^{-2\pi i s t} \cos 2\pi t\, dt$$

[4]And I also know full well that I have reignited the debate on how Π should be defined at the endpoints.

does not converge, no way, no how. Or an even more basic example is $f(t) = 1$ and the integral

$$\int_{-\infty}^{\infty} e^{-2\pi ist} \cdot 1\, dt\,.$$

No way, no how, again. This is no good.

Before we bury $L^1(\mathbb{R})$ as insufficient for our needs, here's one more good thing about it. There's actually an additional consequence for $\mathcal{F}f$ when $f \in L^1(\mathbb{R})$, namely:

- If $\displaystyle\int_{-\infty}^{\infty} |f(t)|\, dt < \infty$, then $\mathcal{F}f(s) \to 0$ as $s \to \pm\infty$.

This is called the *Riemann-Lebesgue lemma* and it's more difficult to prove than the statements on existence and continuity. I'll comment on it in Section 4.8.3. One might view the result as saying that $\mathcal{F}f(s)$ is at least *trying* to be integrable. It's continuous and it tends to zero as $s \to \pm\infty$. Unfortunately, the fact that $\mathcal{F}f(s) \to 0$ does not imply that it's integrable (think of sinc, again). In fact, a function in $L^1(\mathbb{R})$ need not tend to zero at $\pm\infty$; that's also discussed in Section 4.3. If we knew something, or could insist on something about the *rate* at which a signal or its transform tends to zero at $\pm\infty$, then perhaps we could push on further.

4.1.2. The path, the way. Fiddling around with $L^1(\mathbb{R})$ or substitutes, putting extra conditions on jumps — all have been used. The path to success lies elsewhere. It is well marked and firmly established, but it involves a break with the classical point of view. The outline of how all this is settled goes like this:

1. We single out a collection of functions $\mathcal{S}$ for which convergence of the Fourier integrals is assured, for which a function *and* its Fourier transform are both in $\mathcal{S}$, and for which Fourier inversion works. Furthermore, Parseval's identity holds:
$$\int_{-\infty}^{\infty} |f(x)|^2\, dx = \int_{-\infty}^{\infty} |\mathcal{F}f(s)|^2\, ds\,.$$

Perhaps surprisingly it's not so hard to find a suitable collection $\mathcal{S}$, at least if one knows what one is looking for. Also perhaps surprisingly, if not ironically, the path to finding Fourier transforms of bad functions, so to speak, starts with the *best* functions.

This much is classical; new ideas with new intentions, yes, but not new *objects*. But what comes next is definitely not classical. Parts of the theory had been anticipated first by Oliver Heaviside and were used effectively by him for solving differential equations that arise in electrical applications. He was steeped in the problems of telegraphy. The conventional formulation of Maxwell's equations of electromagnetics is due to Heaviside. No Fourier transforms, but a lot of δ's (not yet so named).

Heaviside's approach was developed, somewhat, mostly by less talented people and it remained highly suspicious. The ideas were further cultivated by the work of Paul Dirac for use in quantum mechanics, with great success and with the official naming of the Dirac δ.

Mathematicians still found much to be suspicious about, but the rest of the world was warming to the obviously important applications and to the ease with which calculations could be made. Still no generalized Fourier transform. Another influential collection of results originated in the work of Sergei Sobolev on partial differential equations, the use of so-called *weak solutions*.

Finally, in a tour de force, the ideas were expanded and crystalized by Laurent Schwartz. That's where we're headed. $\mathcal{S}$ is for Schwartz, by the way.

2. $\mathcal{S}$ forms a class of *test functions* which, in turn, serve to define a larger class of *generalized functions* or *distributions*. Distributions *operate* on test functions — you pass a distribution a test function, it returns a complex number. The operation is assumed to be linear (so we can invoke superposition when needed) and continuous (so we can take limits when needed). The model case is the δ-function, which we'll define precisely. You've probably seen δ, imprecisely, in other classes, often in connection with the term "impulse response."

"Distribution" is Schwartz's term, and it has nothing to do with probability distributions, etc. There are just so many words to go around, and sometimes they get used more than once and for different purposes. For the class $\mathcal{S}$ of test functions, one uses the term *tempered distributions* and the notation $\mathcal{T}$ for the collection of tempered distributions. The tempered distributions include, for example, L^1- and L^2-functions (which can be wildly discontinuous), the sinc function, and complex exponentials (hence periodic functions). But they include much more, like δ-functions and related objects.

Precisely because $\mathcal{S}$ was chosen to be the ideal Fourier friendly space of classical signals, the tempered distributions are well suited for Fourier methods. *We'll define the Fourier transform of a tempered distribution.*

3. The Fourier transform and its inverse will be defined so as to operate on tempered distributions, and they operate to produce distributions of the same type. The inverse Fourier transform can be applied, and the Fourier inversion theorem holds in this setting.
4. In the case when a tempered distribution comes from a function — in a way we'll explain — the Fourier transform reduces to the usual definition as an integral, when the integral makes sense, so we won't have lost anything in the process. However, tempered distributions are more general than functions, so we really will have done something new.

Our goal is to hit the relatively few main ideas in the outline above, suppressing the considerable mass of technical details. In practical terms this will enable us to introduce δ-functions and the like as tools for computation and to feel a greater measure of confidence in the range of applicability of the formulas. We're taking this path because it works, it's very interesting, *and it's easy to compute with.* I especially want you to come to believe the last point.

Vacuum tubes? We'll touch on some other approaches to defining distributions (as limits), but as far as I'm concerned they are the equivalent of vacuum tube technology. You can do distributions in other ways, and some people really love building things with vacuum tubes, but wouldn't you rather learn something more up to date?

Let us now praise famous men. Heaviside was an odd character for sure. An appreciative biography is *Oliver Heaviside: Sage in Solitude*, by Paul Nahim, published by the IEEE press. Paul Dirac was at least as odd, at least according to the title of a recent biography: *The Strangest Man: The Hidden Life of Paul Dirac, Mystic of the Atom*, by Graham Farmelo. Schwartz's autobiography (no particular oddness known to me) has now been translated into English under the title *A Mathematician Grappling with His Century.* He is an imposing figure. I don't know much about Sobolev, odd or otherwise, beyond what can be found through Wikipedia et al.

4.2. The Best Functions for Fourier Transforms: Rapidly Decreasing Functions

From our discussion of orthogonality in Chapter 1 you might remember a short sermon on the role of definitions. Ditto for the definition of convolution. This bears repeating at the commencement of this new journey: stated as a working principle, mathematics progresses more by making intelligent definitions than by proving theorems.

The hardest work is often in formulating the fundamental concepts in the right way, a way that will then make the deductions from those definitions (relatively) easy and natural. This can take awhile to sort out, and a subject might be reworked several times as it matures. When discoveries accumulate and one sees where things end up, there's a tendency to go back and change the starting point so that the trip becomes easier. I made mention of this sort of thing in connection to probability, where "random variable" ultimately emerged as the starting point. Mathematicians may be especially self-conscious about this process, but there are certainly examples in engineering where close attention to the basic definitions has shaped a field. Think of Shannon's work on information theory for a particularly striking example.

Engineers can find this tiresome, wanting to *do something* and not, so it may seem, just talk about it. "Devices don't have hypotheses" is how one of my colleagues put it. One can also have too much of a good thing. Too many trips back to the starting point to rewrite the rules can make it hard to follow the game, especially if one has already played by the earlier rules. I'm sympathetic to these concerns, and for our present work on the Fourier transform I'll try to steer a course that makes the definitions reasonable and lets us make steady forward progress. But for the wary let me also offer some honest feedback from teaching this topic this way to many engineers of many stripes. They liked it. They appreciated seeing the ideas coalesce and they liked seeing how the computations played out. Read the chapter lightly, skip around, and try to keep those things in sight. If you're more of a mind mostly to hunt for formulas, you should be able to find what you need, and that's perfectly OK.

4.2.1. Smoothness and decay. To ask how fast $\mathcal{F}f(s)$ might tend to zero, depending on what additional assumptions we might make about the function $f(x)$ *beyond* integrability, will lead to our defining rapidly decreasing functions, and this is the key. Integrability is too weak a condition on a signal $f(x)$ to get very far, but it does imply that $\mathcal{F}f(s)$ is continuous and tends to 0 at $\pm\infty$. What we're going to do is study the relationship between the *smoothness* of a function — not just continuity, but how many times it can be differentiated — and the rate at which its Fourier transform decays at infinity.

Before continuing, a word about the word "smooth." As just mentioned, when applied to a function, it's in reference to how many derivatives the function has. More often than not, and certainly in this chapter, the custom is that when one says simply that a function is smooth, one means that it has derivatives of any order. Put another way, a smooth function is, by custom, an *infinitely differentiable* function.

We'll always assume that $f(x)$ is absolutely integrable and so has a Fourier transform. Let's suppose, more stringently, the following:

- $|xf(x)|$ is integrable; i.e., $\displaystyle\int_{-\infty}^{\infty} |xf(x)|\,dx < \infty$.

Then $xf(x)$ has a Fourier transform. So does $-2\pi i x f(x)$, and its Fourier transform is

$$\begin{aligned}
\mathcal{F}(-2\pi i x f(x)) &= \int_{-\infty}^{\infty} (-2\pi i x) e^{-2\pi i s x} f(x)\,dx \\
&= \int_{-\infty}^{\infty} \left(\frac{d}{ds} e^{-2\pi i s x}\right) f(x)\,dx = \frac{d}{ds}\int_{-\infty}^{\infty} e^{-2\pi i s x} f(x)\,dx \\
&\quad \text{(switching } d/ds \text{ and the integral is justified by the integrability of } |xf(x)|) \\
&= \frac{d}{ds}(\mathcal{F}f)(s)\,.
\end{aligned}$$

This says that the Fourier transform $\mathcal{F}f(s)$ is differentiable and that its derivative is $\mathcal{F}(-2\pi i x f(x))$. When $f(x)$ is merely integrable, we know that $\mathcal{F}f(s)$ is merely continuous, but with the extra assumption on the integrability of $xf(x)$ we conclude that $\mathcal{F}f(s)$ is actually differentiable. (And its derivative is continuous. Why?)

For one more go-round in this direction, what if $|x^2 f(x)|$ is integrable? Then, by the same argument,

$$\begin{aligned}
\mathcal{F}((-2\pi i x)^2 f(x)) &= \int_{-\infty}^{\infty} (-2\pi i x)^2 e^{-2\pi i s x} f(x)\,dx \\
&= \int_{-\infty}^{\infty} \left(\frac{d^2}{ds^2} e^{-2\pi i s x}\right) f(x)\,dx \\
&= \frac{d^2}{ds^2}\int_{-\infty}^{\infty} e^{-2\pi i s x} f(x)\,dx = \frac{d^2}{ds^2}(\mathcal{F}f)(s)\,,
\end{aligned}$$

and we see that $\mathcal{F}f$ is twice differentiable. (And its second derivative is continuous.)

Clearly, we can proceed like this, and as a somewhat imprecise headline we might then announce:

- Faster decay of $f(x)$ at $\pm\infty$ leads to a greater differentiability of the Fourier transform.

Now let's take this in another direction, with an assumption on the differentiability of the signal. Suppose $f(x)$ is differentiable, that its derivative is integrable, and that $f(x) \to 0$ as $x \to \pm\infty$. I've thrown in all the assumptions I need to justify the following calculation:

$$\begin{aligned}
\mathcal{F}f(s) &= \int_{-\infty}^{\infty} e^{-2\pi isx} f(x)\,dx \\
&= \left[f(x)\frac{e^{-2\pi isx}}{-2\pi is} \right]_{x=-\infty}^{x=\infty} - \int_{-\infty}^{\infty} \frac{e^{-2\pi isx}}{-2\pi is} f'(x)\,dx \\
&\qquad \text{(integration by parts with } u = f(x),\ dv = e^{-2\pi isx}dx\text{)} \\
&= \frac{1}{2\pi is}\int_{-\infty}^{\infty} e^{-2\pi isx} f'(x)\,dx \quad \text{(using } f(x) \to 0 \text{ as } x \to \pm\infty\text{)} \\
&= \frac{1}{2\pi is}(\mathcal{F}f')(s)\,.
\end{aligned}$$

We then have

$$|\mathcal{F}f(s)| = \frac{1}{2\pi|s|}|(\mathcal{F}f')(s)| \le \frac{1}{2\pi|s|}\|f'\|_1\,.$$

The last inequality is the result: "The Fourier transform is bounded by the L^1-norm of the function." This says that $\mathcal{F}f(s)$ tends to 0 at $\pm\infty$ like $1/s$. (Remember that $\|f'\|_1$ is some fixed number here, independent of s.) Earlier we commented (without proof) that if f is integrable, then $\mathcal{F}f$ tends to 0 at $\pm\infty$, but here with the stronger assumptions, we get the stronger conclusion that $\mathcal{F}f$ tends to zero *at a certain rate*.

Let's go one step further in this direction. Suppose $f(x)$ is *twice* differentiable, that its first and second derivatives are integrable, and that $f(x)$ and $f'(x)$ tend to 0 as $x \to \pm\infty$. The same argument gives

$$\begin{aligned}
\mathcal{F}f(s) &= \int_{-\infty}^{\infty} e^{-2\pi isx} f(x)\,dx \\
&= \frac{1}{2\pi is}\int_{-\infty}^{\infty} e^{-2\pi isx} f'(x)\,dx \quad \text{(picking up on where we were before)} \\
&= \frac{1}{2\pi is}\left(\left[f'(x)\frac{e^{-2\pi isx}}{-2\pi is} \right]_{x=-\infty}^{x=\infty} - \int_{-\infty}^{\infty} \frac{e^{-2\pi isx}}{-2\pi is} f''(x)\,dx \right) \\
&\qquad \text{(integration by parts with } u = f'(x),\ dv = e^{-2\pi isx}dx\text{)} \\
&= \frac{1}{(2\pi is)^2}\int_{-\infty}^{\infty} e^{-2\pi isx} f''(x)\,dx \quad \text{(using } f'(x) \to 0 \text{ as } x \to \pm\infty\text{)} \\
&= \frac{1}{(2\pi is)^2}(\mathcal{F}f'')(s)\,.
\end{aligned}$$

Thus

$$|\mathcal{F}f(s)| = \frac{1}{|2\pi s|^2}|(\mathcal{F}f'')(s)| \le \frac{1}{|2\pi s|^2}\|f''\|_1$$

and we see that $\mathcal{F}f(s)$ tends to 0 like $1/s^2$.

The headline:

- Greater differentiability of $f(x)$, plus integrability, leads to faster decay of the Fourier transform at $\pm\infty$.

Remark on the derivative formula for the Fourier transform. The engaged reader will have noticed that in the course of our work we have rederived the derivative formula

$$\mathcal{F}f'(s) = 2\pi i s \mathcal{F}f(s)\,,$$

but here we used the assumption that $f(x) \to 0$, which we didn't use before.[5] What's up? With the assumptions we made, the derivation above via integration by parts is a natural approach. Later, when we develop the generalized Fourier transform, we'll have suitably generalized derivative formulas.

We could go on as we did above, comparing the consequences of higher differentiability, integrability, and decay, bouncing back and forth between the function and its Fourier transform. The great insight in making use of these observations is that the simplest and most useful way to coordinate *all* these phenomena is to allow for *arbitrarily great differentiability* and *arbitrarily fast decay*. We would like to have both phenomena in play. That is the crucial step, and here is the crucial definition.

4.2.2. Rapidly decreasing functions.

A function $f(x)$ is said to be *rapidly decreasing* at $\pm\infty$ if:

(1) It is infinitely differentiable.[6]

(2) For *all* integers $m \ge 0$ and $n \ge 0$,

$$\left| x^m \frac{d^n}{dx^n} f(x) \right| \to 0 \quad \text{as } x \to \pm\infty.$$

In words, *any* nonnegative power of x times *any* order derivative of f tends to zero at $\pm\infty$.[7]

Note that m and n are independent in this definition. That is, we insist that, say, the 5th power of x times the 17th derivative of $f(x)$ tends to zero and that the 100th power of x times the first derivative of $f(x)$ tends to zero, and whatever you want.

[5] In the earlier derivation we interchanged limits and integration to pull the limit of the difference quotient outside the integral. This also requires assumptions on continuity and differentiability of $f(x)$, but not on behavior at $\pm\infty$.

[6] Since differentiability implies continuity, note that all derivatives are continuous.

[7] We follow the convention that the zeroth-order derivative of a function is just the function.

The sum of two rapidly decreasing functions is again rapidly decreasing, as is their product. The latter follows easily from the former and from the product rule.

These are the functions that constitute $\mathcal{S}$. In Schwartz's honor, they are also referred to as *Schwartz functions*.

Are there any such functions? Any infinitely differentiable function that is identically zero outside some finite interval is one example, and I'll even write down a formula for one of these later. Another example is $f(x) = e^{-x^2}$. You may already be familiar with the phrase "the exponential grows faster than any power of x," and likewise with the phrase "e^{-x^2} decays faster than any power of x."[8] In fact, any derivative of e^{-x^2} decays faster than any power of x as $x \to \pm\infty$, as you can check with l'Hôpital's rule, for example. We can express this exactly as in the definition:

$$\left| x^m \frac{d^n}{dx^n} e^{-x^2} \right| \to 0 \quad \text{as } x \to \pm\infty\,.$$

There are plenty of other rapidly decreasing functions. We also observe that if $f(x)$ is rapidly decreasing, then it is in $L^1(\mathbb{R})$, and also in $L^2(\mathbb{R})$, since it's decaying so fast — faster than it needs to, to guarantee integrability. In fact, the rapid decay actually guarantees pth power integrability for any $p \geq 1$, meaning that $|f(x)|^p$ is integrable. These functions are denoted, naturally enough, by $L^p(\mathbb{R})$.

Alternative definitions. For some derivations (read: proofs) it's helpful to have other conditions that are equivalent to a function being rapidly decreasing. One such is to assume that for any nonnegative integers m and n there is a constant C_{mn} such that

$$\left| x^m \frac{d^n}{dx^n} f(x) \right| \leq C_{mn} \quad \text{as } x \to \pm\infty\,.$$

In words, the mth power of x times the nth derivative of f remains bounded for all m and n as $x \to \pm\infty$. The constant will depend on which m and n we take, as well as on the particular function, but not on x.

This looks like a weaker condition — boundedness instead of tending to 0 — but it's not. If a function is rapidly decreasing as we've already defined it, then it certainly satisfies the boundedness condition. Conversely, if the function satisfies the boundedness condition, then

$$\left| x^{m+1} \frac{d^n}{dx^n} f(x) \right| \leq C_{(m+1)n} \quad \text{as } x \to \pm\infty\,,$$

so

$$\left| x^m \frac{d^n}{dx^n} f(x) \right| \leq \frac{C_{(m+1)n}}{x} \to 0 \quad \text{as } x \to \pm\infty\,.$$

A third equivalent condition is to assume that for any nonnegative integers m and n there is a constant C'_{mn} (so a different constant than just above) such that

$$(1 + |x|)^m \left| \frac{d^n}{dx^n} f(x) \right| \leq C'_{mn}$$

for all $-\infty < x < \infty$.

---

[8] I used e^{-x^2} as an example instead of e^{-x} (for which the statement is true as $x \to \infty$) because I wanted to include $x \to \pm\infty$, and I used e^{-x^2} instead of $e^{-|x|}$ because I wanted the example to be smooth. $e^{-|x|}$ has a corner at $x = 0$.

This condition implies and is implied by the preceding boundedness condition. Here's why. If $|x|$ is big, here meaning that $|x| > A$ for some (large, fixed) A, then $(1+|x|)^m$ is about the same size as $|x|^m$, so bounds on $\left|x^m \dfrac{d^n}{dx^n} f(x)\right|$ are pretty much the same as bounds on $(1+|x|)^m \left|\dfrac{d^n}{dx^n} f(x)\right|$; if you have one, you have the other. If x is not big, here meaning that $|x| \le A$, then both $|x|^m$ and $(1+|x|)^m$ are bounded by $(1+A)^m$ so we're wanting $\left|\dfrac{d^n}{dx^n} f(x)\right|$ to stay bounded for $|x| \le A$. This follows because $\frac{d^n}{dx^n} f(x)$ is continuous, a point I want to elaborate.

- Here we are using the fact that a continuous function on a closed, bounded interval has a finite maximum and minimum on the interval. That's an important math fact. You may have heard it, and appealed to it, in studying optimization problems, for example.
- There'a a more general statement on the existence of maxima and minima, one that involves some additional terminology that will come up again and in other contexts. A set in $\mathbb{R}$ that is *closed and bounded* is called *compact*. Think of a finite union of closed, bounded intervals. The theorem is that a continuous function on a compact set has a maximum and a minimum on the set.

We'll use these alternative formulations of rapidly decreasing in some arguments. It's a matter of taste which condition one takes as the primary definition.

Just to tell you, an advantage of the final characterization is in some proofs where you might find yourself wanting to bound an integral, something involving

$$\int_{-\infty}^{\infty} \left|\frac{d^n}{dx^n} f(x)\right| \, dx \le \int_{-\infty}^{\infty} \frac{C'_{mn}}{(1+|x|)^m} \, dx \,.$$

The integrand on the right has no zeros in the denominator, and that's a good thing. If you instead stuck with the other boundedness condition in your proof, you might instead find yourself working with something like

$$\int_{-\infty}^{\infty} \left|\frac{d^n}{dx^n} f(x)\right| \, dx \le \int_{-\infty}^{\infty} \frac{C_{mn}}{|x|^m} \, dx \,.$$

For the integral on the right you would have to say something about avoiding $x = 0$. You would rather not. This will come up in Section 4.8 (which you might gloss over anyway). Just a heads-up.

4.2.3. Rapidly decreasing functions and the classical Fourier transform.

Let's start to see why Schwartz's idea was such a good one.

The Fourier transform of a function in $\mathcal{S}$ is still in $\mathcal{S}$. Let $f(x)$ be a function in $\mathcal{S}$. We want to show that $\mathcal{F}f(s)$ is also in $\mathcal{S}$. The condition involves derivatives of $\mathcal{F}f$, so what comes in are the derivative formulas for the Fourier transform and the version of the formulas for higher derivatives.

Starting with

$$2\pi i s \mathcal{F}f(s) = \Big(\mathcal{F}\frac{d}{dx}f\Big)(s)$$

and

$$\frac{d}{ds}\mathcal{F}f(s) = \mathcal{F}(-2\pi i x f(x))\,,$$

the higher-order versions of these formulas are

$$(2\pi i s)^n \mathcal{F}f(s) = \Big(\mathcal{F}\frac{d^n}{dx^n}f\Big)(s),$$
$$\frac{d^n}{ds^n}\mathcal{F}f(s) = \mathcal{F}\big((-2\pi i x)^n f(x)\big)\,.$$

Combining these formulas we find, inductively, that for all nonnegative integers m and n,

$$\mathcal{F}\Big(\frac{d^n}{dx^n}\big((-2\pi i x)^m f(x)\big)\Big) = (2\pi i s)^n \frac{d^m}{ds^m}\mathcal{F}f(s)\,.$$

Note how the roles of m and n are flipped in the two sides of the equation.

We use this last identity together with the estimate for the Fourier transform in terms of the L^1-norm of the function. Namely,

$$|s|^n \left|\frac{d^m}{ds^m}\mathcal{F}f(s)\right| = (2\pi)^{m-n}\left|\mathcal{F}\Big(\frac{d^n}{dx^n}(x^m f(x))\Big)\right| \le (2\pi)^{m-n}\left\|\frac{d^n}{dx^n}(x^m f(x))\right\|_1\,.$$

The L^1-norm on the right-hand side is finite because f is rapidly decreasing. Since the right-hand side depends on m and n, we have shown that there is a constant C_{mn} with

$$\left|s^n \frac{d^m}{ds^m}\mathcal{F}f(s)\right| \le C_{mn}\,.$$

This implies that $\mathcal{F}f$ is rapidly decreasing. Done.

Fourier inversion works for functions in $\mathcal{S}$. We first establish the inversion theorem for a *timelimited* function in $\mathcal{S}$; we had this part of the argument back in Chapter 2, but it's been awhile and it's worth seeing it again. Timelimited means that rather than $f(x)$ just tending to zero at $\pm\infty$, we suppose that $f(x)$ is *identically zero* for $|x| \ge T/2$, for some T. Shortly we'll switch to the more common mathematical term *compact support* to refer to such functions.

In this case we can periodize $f(x)$ to get a smooth, periodic function of period T. Expand the periodic function as a *converging* Fourier series. Then for $-T/2 \le t \le T/2$,

$$\begin{aligned}
f(x) &= \sum_{n=-\infty}^{\infty} c_n e^{2\pi i n x/T}\\
&= \sum_{n=-\infty}^{\infty} e^{2\pi i n x/T}\left(\frac{1}{T}\int_{-T/2}^{T/2} e^{-2\pi i n y/T} f(y)\,dy\right)\\
&= \sum_{n=-\infty}^{\infty} e^{2\pi i n x/T}\left(\frac{1}{T}\int_{-\infty}^{\infty} e^{-2\pi i n y/T} f(y)\,dy\right) \quad (f(x)=0 \text{ for } |x|\ge T/2)\\
&= \sum_{n=-\infty}^{\infty} e^{2\pi i x(n/T)}\mathcal{F}f\left(\frac{n}{T}\right)\frac{1}{T} = \sum_{n=-\infty}^{\infty} e^{2\pi i x s_n}\mathcal{F}f(s_n)\Delta s\,,
\end{aligned}$$

with $s_n = n/T$, $\Delta s = 1/T$. Our intention is to let $T \to \infty$. What we see is a Riemann sum for the integral

$$\int_{-\infty}^{\infty} e^{2\pi i x s}\mathcal{F}f(s)\,ds\,,$$

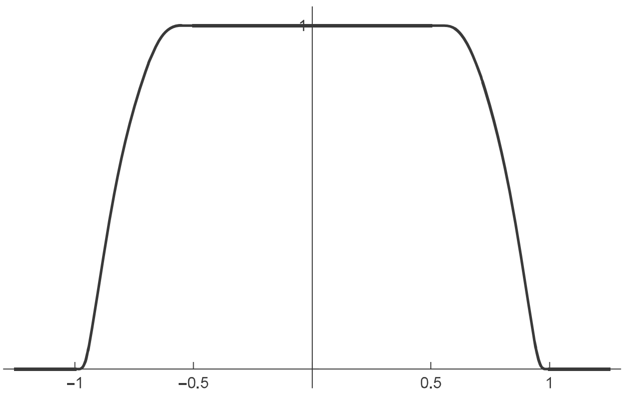
which is $\mathcal{F}^{-1}\mathcal{F}f(x)$, and the Riemann sum converges to the integral because of the smoothness of $f(x)$. I have not slipped anything past you, or the rigor police.

Thus

$$f(x) = \mathcal{F}^{-1}\mathcal{F}f(x)\,,$$

and the Fourier inversion theorem is established for timelimited functions in $\mathcal{S}$.

When f is *not* timelimited, we use *windowing*. The idea is to cut $f(t)$ off *smoothly* by multiplying it by a smooth function — the window — that's 1 over a finite interval and then decreases smoothly down to be identically 0 outside a larger interval. The interesting thing in the present context, for theoretical rather than practical use, is to make the window so smooth that the windowed function is still in $\mathcal{S}$. Some details and further applications are in Section 4.8.1, but here's the setup. To be definite, take a function $c(t)$ that is identically 1 for $-1/2 \le t \le 1/2$, that goes *smoothly* (infinitely differentiable) down to zero as t goes from $1/2$ to 1 and from $-1/2$ to -1 and is then identically 0 for $t \ge 1$ and $t \le -1$. This is a smoothed version of the rectangle function $\Pi(t)$; instead of cutting off sharply at $\pm 1/2$ we bring the function smoothly down to zero. You can certainly imagine drawing such a function, provided your imaginary pencil is infinitely sharp — enough to render curves agreeing to infinite order:



In Section 4.8.1 I'll give an explicit formula for such a window.

Now scale $c(t)$ to $c_n(x) = c(x/n)$. Then $c_n(x)$ is 1 for t between $-n/2$ and $n/2$, goes smoothly down to 0 between $\pm n/2$ and $\pm n$, and is then identically 0 for $|x| \ge n$. Next, the function $f_n(x) = c_n(x)f(x)$ is a timelimited function *in* $\mathcal{S}$. Hence the earlier reasoning shows that the Fourier inversion theorem holds for f_n and $\mathcal{F}f_n$. The window eventually moves past any given x, and consequently $f_n(x) \to f(x)$ as $n \to \infty$. Some estimates based on the properties of the window function, which I won't go through, show that the Fourier inversion theorem also holds in the limit. So far, so good.

Parseval holds for functions in $\mathcal{S}$. Actually, the more general result holds:

If $f(x)$ and $g(x)$ are complex-valued functions in $\mathcal{S}$, then

$$\int_{-\infty}^{\infty} f(x)\overline{g(x)}\,dx = \int_{-\infty}^{\infty} \mathcal{F}f(s)\overline{\mathcal{F}g(s)}\,ds\,.$$

As a special case, if we take $f = g$, then $f(x)\overline{f(x)} = |f(x)|^2$ and the identity becomes

$$\int_{-\infty}^{\infty} |f(x)|^2\,dx = \int_{-\infty}^{\infty} |\mathcal{F}f(s)|^2\,ds\,.$$

The derivation is exactly the one from Chapter 2, with the added confidence that properties of $\mathcal{S}$ justify all steps. I won't repeat it here.

Functions with compact support. Timelimited functions and windowing introduce another category of functions that also come up in the theory of distributions in their own right. These are the *infinitely differentiable functions of compact support* defined on $\mathbb{R}$, the collection of which is often denoted by $C_c^\infty(\mathbb{R})$, or sometimes by $\mathcal{C}$ if you've grown fond of script letters. They are the best functions for partial differential equations in the modern reworking of that subject, and if you take an advanced course on PDE, you will see them all over the place. We'll see them here, too.

To explain the terms, "infinitely differentiable" you already know, and that's the "∞" in $C_c^\infty(\mathbb{R})$. The capital "C" is for "continuous" — all derivatives are continuous. The lowercase, subscript "c" indicates *compact support*, which means that a function is zero in the exterior of a bounded set.

A few technical points, especially for those who have had some topology. The *support* of a function $f(x)$ is the *closure* of the set of points $\{x\colon f(x) \neq 0\}$. That's the smallest closed set containing the set $\{x\colon f(x) \neq 0\}$. This definition is independent of any smoothness property of the function.

As mentioned in Section 4.2.2, a set in $\mathbb{R}$ is said to be *compact* if it is closed and bounded. A bounded set in $\mathbb{R}$ is one that is contained in some finite interval. Thus a function $f(x)$ has compact support if the closure of the set $\{x : f(x) \neq 0\}$ is bounded. (The support is closed by definition — it's the closure of something.) Seems like a much harder way of saying that the function is zero in the exterior of a bounded set, but there are reasons to unpack the terminology.[9] In any event, a consequence of a function $f(x)$ having compact support is then that $f(x)$ is identically 0 for $|x| \geq A$, for some A (generally different A's for different functions). That's the "timelimited" condition we used earlier. The function $c_n(x)$, above, is in $\mathcal{C}$. Its support is $|x| \leq n$. (In particular, there *are* nonzero, smooth functions of compact support. A bit of a relief.)

[9] This circle of ideas belongs to the field of *general topology.* Every math major takes some course that includes some aspects of general topology, and you have to be able to toss around the terms. It's good training. One of my teachers in grad school, whose field was partial differential equations, opined that general topology ranked up there with penmanship in importance to science.

Be aware that "functions of compact support" will pop up. In mathematical practice, the property of compact support is often employed in arguments featuring integrals, and integration by parts, to replace an integral over $\mathbb{R}$ by one over a finite interval (the support) or to conclude that terms at $\pm\infty$ are zero. On the other hand, engineers might be more familiar with the term *bandlimited*, referring to a signal whose spectrum is identically zero beyond some point. To reconcile the two terms, a signal is bandlimited if its Fourier transform has compact support. This will be an important class of signals in Chapter 6 on sampling and interpolation.

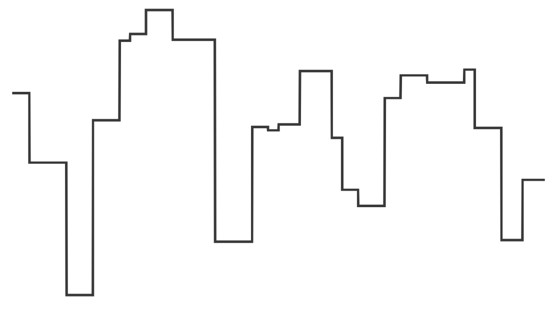
A function in $\mathcal{C}$ is certainly in $\mathcal{S}$, since by the time x reaches $\pm\infty$ the function has been identically 0 for awhile and not just deceasing to 0. So, symbolically, $\mathcal{C} \subset \mathcal{S}$. The inclusion does not go the other way. A function in $\mathcal{S}$ need not have compact support; for example, the support of the Gaussian e^{-x^2} is all of $\mathbb{R}$ since the Gaussian is never zero.

4.3. A Very Little on Integrals

This section on integrals, more of an early-chapter appendix, is not a short course on integration. It's here to provide a little, but only a little, background explanation for some of the statements on integrals made to this point. It's placed early in the chapter so you might accidentally read it. The star of this section is you.

Integrals are first defined for positive functions. In the general approach to integration (of real-valued functions) you first set out to define the integral for *nonnegative* functions. Why? Because however general a theory you're constructing, an integral is going to be some kind of limit of sums and you'll want to know when that kind of limit exists. If you work with positive, or at least nonnegative, functions, then the issues for limits will be about how big the function gets or about how big the sets are where the function is or isn't big. When the function changes sign, cancellations determine the convergence of the integral, what one of my colleagues called a conspiracy between the positive and negative values. You feel better able to analyze accumulations (for nonnegative functions) than to control conspiratorial cancellations (for functions that change sign). So you first work on defining your integral for functions $f(x)$ with $f(x) \geq 0$.

Integrals are first defined for positive step functions. In fact, having thought about it, you decide to begin with nonnegative *step functions*, functions that look like[10]



[10]These are the kinds of functions that you get, for example, from a sample-and-hold process that you've probably seen in signal processing.

Step functions are constant on intervals. They're pretty easy to work with. All you do to define an integral is add up the value of the function times the length of the interval for each of the steps.

Then you go on to use step functions to approximate other functions that you want to integrate. Take many, tiny steps. The integral is then again approximated by adding up the values of the function times the lengths of intervals, but for more general functions you find that you really need a more general concept than simply length of an interval. You invent measure theory! You base your theory of the integral on measurable sets and measurable functions. This works fine. You get more than Riemann got with his integral. You call your friend Lebesgue.

Integrals for functions that change sign. Now, backtracking, you know full well that your definition won't be too useful if you can't extend it to functions that can be both positive and negative. Here's how you do this. For any function $f(x)$ you let $f^+(x)$ be its *positive part*:

$$f^+(x) = \max\{f(x), 0\}.$$

Likewise, you let

$$f^-(x) = \max\{-f(x), 0\}$$

be its *negative part*.[11] Tricky: the "negative part" as you've defined it is actually a positive function; taking $-f(x)$ flips over the places where $f(x)$ is negative to be positive. You like that kind of thing. Then

$$f = f^+ - f^-$$

while

$$|f| = f^+ + f^- .$$

You now say that f is integrable if both f^+ and f^- are integrable — a condition which makes sense since f^+ and f^- are both nonnegative functions — and *by definition* you set

$$\int f = \int f^+ - \int f^- .$$

(For complex-valued functions you apply this to the real and imaginary parts.) You follow this approach for integrating functions on a finite interval or on the whole real line. Moreover, according to this definition $|f|$ is integrable if f is because then

$$\int |f| = \int (f^+ + f^-) = \int f^+ + \int f^-$$

and f^+ and f^- are each integrable.[12] It's also true, conversely, that if $|f|$ is integrable, then so is f. You show this by observing that

$$f^+ \leq |f| \quad \text{and} \quad f^- \leq |f|$$

and this implies that both f^+ and f^- are integrable.

[11]A different use of the notation f^- than we had before, talking about reversal. But we'll never use the negative part of a function again, so be flexible for this section.

[12]Some authors reserve the term "summable" for the case when $\int |f| < \infty$, i.e., for when both $\int f^+$ and $\int f^-$ *are finite.* They still define $\int f = \int f^+ - \int f^-$ but they allow the possibility that one of the integrals on the right may be ∞, in which case $\int f$ is ∞ or $-\infty$ and they don't refer to f as summable.