Chapter 21

Fourier Analysis and the Central Limit Theorem

Any theorem as important as the Central Limit Theorem deserves more than one proof. Different proofs emphasize different aspects of the problem. Our first proof was based on properties of moment generating functions. It's a nice proof, and the main idea is easily explained (for 'nice' distributions, if the moment generating function converges to the moment generating function of the standard normal, then the densities converge to the density of the standard normal). Unfortunately the proof uses some major results in Complex Analysis. We thus want to provide a proof that works under less restrictive conditions.

Sadly, the proof below won't be it. It too appeals to some black box results from Complex Analysis. As this argument still requires us to assume results beyond the scope of the book, why do we bother giving this proof? There are many reasons. The first is that it introduces **integral transforms**, specifically the **Fourier transform**. Integral transforms in general and the Fourier transform in particular are ubiquitous in higher mathematics, and it never hurts to see them. The second reason is that our statement of the Central Limit Theorem is for functions whose moment generating functions exist in a neighborhood of the origin. There are plenty of densities that have finite first, second and third moments but whose moment generating function doesn't exist. Consider for example a cousin of the Cauchy distribution,

$$f(x) = \frac{4\sin(\pi/8)}{\pi}\frac{1}{1+x^8}.$$

It shouldn't be apparent that this is a probability distribution. It is clearly non-negative and it decays rapidly enough as $|x| \to \infty$ so that the integral converges; however, it's not at all clear that it will integrate to 1, although the constant does have some nice features (it has an 8 in it, which could come from the power of x, and the normalization constant of the Cauchy distribution had a π in the denominator, which this does as well). For our purposes, *it doesn't matter!* Say we have the normalization constant wrong – who cares! There's some constant, let's call it C_8, such that $C_8/(1+x^8)$ is a probability density. While this will have finite mean, variance and third moment, the eight moment is clearly infinite, as it's

$$\int_{-\infty}^{\infty} x^8 \frac{C_8}{1+x^8} dx;$$

the integrand is essentially C_8 for $|x|$ large, and thus the integral diverges. Similarly one can show all the larger even moments blow-up, and hence the moment generating function can't converge in a neighborhood of the origin *as the moment generating function doesn't exist!*.

This example shows us that our approach to the Central Limit Theorem is too restrictive, as it eliminates many nice distributions (For example, the Cauchy distribution arises in Mandelbrot's work on fractal behavior of financial and commodities markets.). The moment generating function approach is fundamentally flawed; there's just no getting around the fact that some nice distributions don't have a moment generating function, and thus we can't do any argument that requires a moment generating function to exist! While the sums of independent Cauchy distributions do not approach normality, the cousin of it we mentioned above does. The key turns out to be having finite mean and variance.

One solution to this quandary is to study the Fourier transform of our density, which in probability is called the **characteristic function**. We'll see later that unlike the moment generating function, the characteristic function *always* exists, and is a very close analogue of the moment generating function. It has better properties (such as existence!), and is more amenable to analysis. This will allow us to adapt our previous proof. The ideas are similar, but the algebra is a little different.

The material in this chapter is thus a bit more advanced than many introductory probability courses. Most courses just don't have time to delve this deeply. While we won't prove everything we need, we'll provide enough details so that hopefully the big picture is clear, and give you a sense of some of what's waiting for you in future math classes.

21.1 Integral transforms

Given a function $K(x, y)$ and an interval I (which is frequently $(-\infty, \infty)$ or $[0, \infty)$), we can construct a map from functions to functions as follows: send f to

$$(\mathcal{K}f)(y) \ := \ \int_I f(x)K(x, y)dx.$$

As the integrand depends on the two variables x and y and we only integrate out x, the result is a function of y. Obviously it doesn't matter what letters we use for the dummy variables; other common choices are $K(t, x)$ or $K(t, s)$ or $K(x, \xi)$. We call K the **kernel** and the new function the integral transform of f.

Integral transforms are useful for studying a variety of problems. Their utility stems from the fact that the related function leads to simpler algebra for the problem at hand. We define two of the most important integral transforms, the Laplace and the Fourier transforms.

Definition 21.1.1 (Laplace Transform) *Let* $K(t, s) = e^{-ts}$. *The Laplace transform of* f, *denoted* $\mathcal{L}f$, *is given by*

$$(\mathcal{L}f)(s) \;=\; \int_0^\infty f(t)e^{-st}dt.$$

Given a function g, *its inverse Laplace transform,* $\mathcal{L}^{-1}g$, *is*

$$(\mathcal{L}^{-1}g)(t) \;=\; \lim_{T\to\infty}\frac{1}{2\pi i}\int_{c-iT}^{c+iT} e^{st}g(s)ds \;=\; \lim_{T\to\infty}\frac{1}{2\pi i}\int_{-T}^{T} e^{(c+i\tau)t}g(c+i\tau)id\tau.$$

Definition 21.1.2 (Fourier Transform (or Characteristic Function)) *Let* $K(x, y) = e^{-2\pi ixy}$. *The Fourier transform of* f, *denoted* $\mathcal{F}f$ *or* $\widehat{f}$, *is given by*

$$\widehat{f}(y) \;:=\; \int_{-\infty}^{\infty} f(x)e^{-2\pi ixy}dx,$$

where

$$e^{i\theta} \;:=\; \sum_{n=0}^{\infty}\frac{(i\theta)^n}{n!} \;=\; \cos\theta + i\sin\theta.$$

The inverse Fourier transform of g, *denoted* $\mathcal{F}^{-1}g$, *is*

$$(\mathcal{F}^{-1}g)(x) \;=\; \int_{-\infty}^{\infty} g(y)e^{2\pi ixy}dy.$$

Note other books define the Fourier transform differently, sometimes using $K(x, y) = e^{-ixy}$ *or* $K(x, y) = e^{-ixy}/\sqrt{2\pi}$.

The Laplace and Fourier transforms are related. If we let $s = 2\pi iy$ and consider functions $f(x)$ which vanish for $x \leq 0$, we see the Laplace and Fourier transforms are equal.

While we have chosen to write the Fourier transform of f by

$$\widehat{f}(y) \;=\; \int_{-\infty}^{\infty} f(x)e^{-2\pi ixy}dx,$$

other books sadly might use a different notation. Some authors use e^{-ixy} or $e^{ixy}/\sqrt{2\pi}$ instead of $e^{-2\pi ixy}$, so always check the convention when you reference a book or use a program such as Mathematica. Why are there so many different notations? It turns out that different notations lead to cleaner algebra for different problems. For our purposes, this choice leads to the simplest algebra, which is why we use it.

Given a function f we can compute its transform. What about the other direction? If we are told g is the transform of some function f, can we recover f from knowing g? If yes, is the corresponding f unique? Notice how similar these questions are to the two black-box complex analysis theorems from Chapter 20. There we knew moment generating functions and wanted to recover densities. Fortunately, the answer to both questions turns out to be 'yes', provided f and g satisfy certain nice conditions. A particularly nice set of functions to study is the Schwartz space.

Definition 21.1.3 (Schwartz space) *The Schwartz space, $\mathcal{S}(\mathbb{R})$, is the set of all infinitely differentiable functions f such that, for any non-negative integers m and n,*

$$\sup_{x \in \mathbb{R}} \left| (1+x^2)^m \frac{d^n f}{dx^n} \right| \;<\; \infty,$$

where $\sup_{x\in\mathbb{R}} |g(x)|$ is the smallest number B such that $|g(x)| \le B$ for all x (think 'maximum value' whenever you see supremum).

Whenever we define a space or a set, it's worthwhile to show that it isn't empty! Let's show there are infinitely many Schwartz functions. We claim the Gaussians $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}$ are in $\mathcal{S}(\mathbb{R})$ for any $\mu, \sigma \in \mathbb{R}$. By a change of variables, it suffices to study the special case of $\mu = 0$ and $\sigma = 1$. Clearly the standard normal, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is infinitely differentiable. Its first few derivatives are

$$\begin{aligned} f'(x) &= -x \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \\ f''(x) &= (x^2-1) \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \\ f'''(x) &= -(x^3-3x) \cdot \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \end{aligned}$$

By induction, we can show that the n^{th} derivative is a polynomial $p_n(x)$ of degree n times $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. To show f is Schwartz, by Definition 21.1.3 we must show

$$\left| (1+x^2)^m \cdot p_n(x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right|$$

is bounded. This follows from the fact that the standard normal decays faster than any polynomial. Say we want to show $|x^m e^{-x^2/2}|$ is bounded. The claim is clear for $|x| \le 1$. What about larger $|x|$? By keeping just one term of the Taylor series expansion of the exponential function, we know $(x^2/2)^k/k! \le e^{x^2/2}$ for any k, so $e^{-x^2/2} \le k! 2^k / x^{2k}$. Thus $|x^m e^{-x^2/2}| \le k! 2^k / x^{2k-m}$, and if we choose $2k > m$ then this is bounded by $k! 2^k$.

We now state the main result we need from Complex Analysis. It states precisely when the integral transform arises from a unique input. We only give the statement for the inverse Fourier transform – just stating the result for the Laplace transform requires a lot of new notation from Complex Analysis! A proof can be found in many books on Complex Analysis or Fourier Analysis (see for example [SS1, SS2]).

Theorem 21.1.4 (Inversion Theorems) *Let $f \in \mathcal{S}(\mathbb{R})$, the Schwartz space. Then*

$$f(x) \;=\; \int_{-\infty}^{\infty} \widehat{f}(y) e^{2\pi i x y}\, dy,$$

where $\widehat{f}$ is the Fourier transform of f. In particular, if f and g are Schwartz functions with the same Fourier transform, then $f(x) = g(x)$.

This interplay between a function and its transform will be very useful for us when we study probability distributions, as the moment generating function is an integral transform of the density! Recall the moment generating function is defined by $M_X(t) = \mathbb{E}[e^{tX}]$, which means

$$M_X(t) \;=\; \int_{-\infty}^{\infty} e^{tx} f(t)dt.$$

If $f(x) = 0$ for $x \leq 0$, *this is just the Laplace transform of* f. Alternatively, if we take $t = -2\pi i y$ then it's the Fourier transform of f. This is trivially related to (yet another!) generating function, the characteristic function of X.

The characteristic function. The characteristic function of a random variable X is

$$\phi(t) \;:=\; \mathbb{E}[e^{itX}].$$

Unlike the moment generating function, if X has a continuous density then the characteristic function *always* exists for all t. Note the characteristic function is essentially the Fourier transform of the density: the Fourier transform is just $\phi(-2\pi t)$.

Why does the characteristic function always exist? Remember the density is a non-negative integrable function f. We have

$$|\phi(t)| \;=\; \left|\int_{x=-\infty}^{\infty} e^{itx} f(x)dx\right| \;\leq\; \int_{x=-\infty}^{\infty} |e^{itx}|f(x)dx \;=\; \int_{x=-\infty}^{\infty} f(x)dx \;=\; 1,$$

as $|e^{itx}| = 1$. The reason the absolute value of this exponential is 1 is the **Pythagorean theorem**. We have $e^{i\theta} = \cos\theta + i\sin\theta$ for any real θ (there are many ways to see this; one way is to compare the Taylor series expansions of e^{θ}, $\cos\theta$ and $\sin\theta$, noting $i = \sqrt{-1}$). If $z = a + ib$ is a complex number, then $|z|^2 = z\overline{z}$, where $\overline{z} = a - ib$ is the **complex conjugate**; we call $|z|$ the **length**, **absolute value** or the **norm** of z. For us, we get

$$|e^{itx}|^2 \;=\; (\cos tx + i\sin tx)(\cos tx - i\sin tx) \;=\; \cos^2 tx + \sin^2 tx \;=\; 1.$$

This is a immense improvement over the moment generating function – the most important property an object can have is existence, so already we've made progress.

Furthermore, we see that the characteristic function is simply related to the moment generating function. What a difference, though, an i makes! The characteristic function and the Fourier transform are trivially related; they differ by rescaling the input by a factor of 2π. The rescaling from the characteristic function to the moment generating function is far more profound; the presence of the factor of i leads to *very* different behavior, and very different algebra.

We now see why these results from complex analysis will save the day. The inversion formulas above tell us that, if our initial distribution is nice, then knowing the integral transform of the function is the same as knowing the function; in other words, knowing the integral transform uniquely determines the distribution.

The following remark is a bit more advanced, and is meant to put these arguments into their proper context in the realm of analytic arguments. A function $f : \mathbb{R} \to \mathbb{C}$ has **compact support** if there's a finite closed interval $[a, b]$ such that for all $x \notin [a, b]$, $f(x) = 0$. Schwartz functions with compact support are extremely useful in many arguments. It can be shown that given any continuous function g on a finite closed interval $[a, b]$, there's a Schwartz function f with compact support arbitrarily close to g; i.e., for all $x \in [a, b]$, $|f(x) - g(x)| < \epsilon$. Similarly, given any such continuous function g, one can find a sum of **step functions** of intervals arbitrarily close to g in the same sense as above (a step function is a finite sum of characteristic functions of closed intervals). Often, to prove a result for step functions it suffices to prove the result for continuous functions, which is the same as proving the result for Schwartz functions. Schwartz functions are infinitely differentiable and as the Fourier Inversion formula holds, we can pass to the Fourier transform space, which is sometimes easier to study.

21.2 Convolutions and Probability Theory

An important property of the Fourier transform is that it behaves nicely under **convolution**. Remember we denote the convolution of two functions f and g by $h = f * g$, where

$$h(x) \;=\; \int_{-\infty}^{\infty} f(t)g(x-t)dt \;=\; \int_I f(x-t)g(t)dt.$$

A natural question to ask is: what must we assume about f and g to ensure that the convolution exists? For our purposes, f and g will be probability densities. Thus they're non-negative and integrate to 1. While this is all we need to ensure that $h = f * g$ integrates to 1, it's not quite enough to guarantee that $f * g$ is finite. Let's first show it integrates to 1. Since our integrand is non-negative, we're allowed to switch the order of integration. Note for each x the integral is either non-negative or positive infinity. We have

$$\begin{aligned}\int_{x=-\infty}^{\infty} (f*g)(x)dx &= \int_{x=-\infty}^{\infty}\int_{t=-\infty}^{\infty} f(t)g(x-t)dtdx \\ &= \int_{t=-\infty}^{\infty} f(t)\left[\int_{x=-\infty}^{\infty} g(x-t)dx\right]dt.\end{aligned}$$

The integral in brackets is 1. If you want, change variables and let $u = x - t$, $du = dx$. We're integrating a probability density from $-\infty$ to ∞; that's always 1. We're left with

$$\int_{x=-\infty}^{\infty} (f*g)(x)dx \;=\; \int_{t=-\infty}^{\infty} f(t)dt \;=\; 1,$$

again as the integral of a probability density from $-\infty$ to ∞ is always 1. This means our non-negative function $(f * g)(x)$ can only be zero on a set of measure (or length) 0. If you're not familiar with measure theory, no worries: here's another formulation.

It means that for any M, the length of $\{x : (f * g)(x) > M\}$ is at most $1/M$, as otherwise the integral would exceed 1.

So, this proves that for almost all x we have $(f * g)(x)$ is finite. What must we assume about f and g so that the convolution is finite for *all* x? If we assume f and g are **square-integrable**, namely $\int_{-\infty}^{\infty} f(x)^2 dx$ and $\int_{-\infty}^{\infty} g(x)^2 dx$ are finite, then $f * g$ is well-behaved everywhere. We'll see shortly how this follows from the Cauchy-Schwarz inequality, which is proved in Appendix B.6.

The Cauchy-Schwarz inequality. For complex-valued functions f and g,

$$\int_{-\infty}^{\infty} |f(x)g(x)|dx \;\leq\; \left(\int_{-\infty}^{\infty} |f(x)|^2 dx\right)^{1/2} \cdot \left(\int_{-\infty}^{\infty} |g(x)|^2 dx\right)^{1/2}.$$

Assuming f and g are square-integrable is very weak, and is met in all the standard densities we study. Even in situations where they're not square-integrable, often there are no problems. For example, if we take

$$f(x) \;=\; \begin{cases} \frac{1}{2\sqrt{x}} & \text{if } 0 < x \leq 1 \\ 0 & \text{otherwise,} \end{cases}$$

then f is integrable but not square integrable, as $\int_0^1 dx/x$ blows-up. That said, the convolution of f with itself is well-behaved. After 'some' integration, you would find

$$(f * f)(y) \;=\; \begin{cases} \pi/4 & \text{if } 0 < y \leq 1 \\ (\text{arccsc}(\sqrt{y}) - \arctan(\sqrt{y-1}))/2 & \text{if } 1 < y < 2 \\ 0 & \text{otherwise.} \end{cases}$$

We now state a wonderful result. It is because of this that the Fourier transform is so prevalent in probability. It's such an important result that we provide a full proof.

Theorem 21.2.1 (Convolutions and the Fourier Transform) *Let f, g be continuous functions on $\mathbb{R}$. If $\int_{-\infty}^{\infty} |f(x)|^2 dx$ and $\int_{-\infty}^{\infty} |g(x)|^2 dx$ are finite then $h = f * g$ exists, and $\widehat{h}(y) = \widehat{f}(y)\widehat{g}(y)$. Thus the Fourier transform converts convolution to multiplication.*

Proof: We first show $h = f * g$ exists. We have

$$\begin{aligned} h(x) &= (f * g)(x) \\ &= \int_{-\infty}^{\infty} f(t)g(x-t)dt \\ |h(x)| &\leq \int_{-\infty}^{\infty} |f(t)| \cdot |g(x-t)|dt \\ &\leq \left(\int_{-\infty}^{\infty} |f(t)|^2 dt\right)^{1/2} \left(\int_{-\infty}^{\infty} |g(x-t)|^2 dt\right)^{1/2} \end{aligned}$$

by the Cauchy-Schwarz inequality. As we're assuming f and g are square-integrable, both integrals are finite (for x fixed, as t runs from $-\infty$ to ∞ so too does $x-t$). We're not assuming f and g are densities; if we did, then the inequalities would be equalities as the densities are never negative.

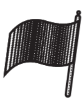
Now that we know that the convolution h exists, we can explore its properties. Let's calculate its Fourier transform. This leads to a double integral (one integral from the definition of h, and then another from the definition of the Fourier transform). The fact that we'll have two integrals suggests how we'll handle it. Usually there are two things to try with a double integral: we can change variables, or we can interchange the order of integration. We'll interchange orders; this is justified as the integrals of the absolute value is finite, and we can appeal to the Fubini Theorem (see Theorem B.2.1).

Before we change orders, however, we first cleverly **add zero** to facilitate the algebra (see §A.12 for more examples of this method). We'll see shortly an integral of $g(x-t)$ against the exponential $e^{-2\pi i x y}$. As we have g evaluated at $x-t$, we want $x-t$ in the exponential and not x. This suggests writing x as $x-t+t$. We find

$$\begin{aligned}
\widehat{h}(y) &= \int_{-\infty}^{\infty} h(x) e^{-2\pi i x y} dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t) g(x-t) e^{-2\pi i x y} dt dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(t) g(x-t) e^{-2\pi i (x-t+t) y} dt dx \\
&= \int_{t=-\infty}^{\infty} f(t) e^{-2\pi i t y}\left[\int_{x=-\infty}^{\infty} g(x-t) e^{-2\pi i (x-t) y} dx\right] dt \\
&= \int_{t=-\infty}^{\infty} f(t) e^{-2\pi i t y}\left[\int_{u=-\infty}^{\infty} g(u) e^{-2\pi i u y} dx\right] dt \\
&= \int_{t=-\infty}^{\infty} f(t) e^{-2\pi i t y}\widehat{g}(y) dt \;=\; \widehat{f}(y)\widehat{g}(y),
\end{aligned}$$

where the last line is from the definition of the Fourier transform. □

If for all $i = 1, 2, \dots$ we have f_i is square-integrable, prove for all i and j that $\int_{-\infty}^{\infty} |f_i(x) f_j(x)| < \infty$. What about $f_1 * (f_2 * f_3)$ (and so on)? Prove $f_1 * (f_2 * f_3) = (f_1 * f_2) * f_3$. Therefore convolution is associative, and we may write $f_1 * \cdots * f_N$ for the convolution of N functions. If you're stuck, we discuss this in §19.5.



It's unusual to have two operations that essentially commute. We have the Fourier transform of a convolution is the product of the Fourier transforms; as convolution is like multiplication, this is saying that using this special type of multiplication, we can switch the orders of the operations. It is rare to have two operations satisfying such a rule. For example, $\sqrt{a+b}$ typically is *not* $\sqrt{a}+\sqrt{b}$.

The following lemma is the starting point to the Fourier analytic proof of the Central Limit Theorem.

Lemma 21.2.2 *Let X_1 and X_2 be two independent random variables with densities f and g. Assume f and g are square-integrable probability densities, so $\int_{-\infty}^{\infty} f(x)^2 dx$ and $\int_{-\infty}^{\infty} g(x)^2 dx$ are finite. Then $f * g$ is the probability density for $X_1 + X_2$. More generally, if $X_1, \dots, X_N$ are independent random variables with square-integrable densities $p_1, \dots, p_N$, then $p_1 * p_2 * \cdots * p_N$ is the density for $X_1 + \cdots + X_N$. (As convolution is commutative and associative, we don't have to be careful when writing $p_1 * p_2 * \cdots * p_N$.)*

Proof: The probability of $X_i \in [x, x+\Delta x]$ is $\int_x^{x+\Delta x} f(t)dt$, which is approximately $f(x)\Delta x$ when Δx is small (as the integrand is essentially constant). The probability that $X_1 + X_2 \in [x, x + \Delta x]$ is just

$$\int_{x_1=-\infty}^{\infty} \int_{x_2=x-x_1}^{x+\Delta x - x_1} f(x_1) g(x_2) dx_2 dx_1.$$

As $\Delta x \to 0$ we obtain the convolution $f * g$, and find

$$\text{Prob}(X_1 + X_2 \in [a, b]) = \int_a^b (f * g)(z) dz. \tag{21.1}$$

We must justify our use of the word "probability" in (21.1); namely, we must show $f * g$ is a probability density. Clearly $(f * g)(z) \geq 0$ as $f(x), g(x) \geq 0$. As we are assuming f and g are square-integrable,

$$\begin{aligned}
\int_{-\infty}^{\infty} (f * g)(x) dx &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x-y) g(y) dy dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x-y) g(y) dx dy \\
&= \int_{-\infty}^{\infty} g(y) \left(\int_{-\infty}^{\infty} f(x-y) dx \right) dy \\
&= \int_{-\infty}^{\infty} g(y) \left(\int_{-\infty}^{\infty} f(t) dt \right) dy.
\end{aligned}$$

As f and g are probability densities, these integrals are 1, completing the proof. □

Remark: we really don't need to assume the densities are square-integrable. The purpose of that assumption is to make sure the density of the sum of the random variables is finite everywhere. If we're willing to allow our density to be infinite at a few places, we can drop that assumption.

This section introduced a lot of material and results, but we can now begin to see the big picture. If we take N independent random variables with densities $p_1, \dots, p_N$, then the sum has density $p = p_1 * \cdots * p_N$. While at first this equation looks frightening (what is the convolution of N exponential densities?), there's a remarkable simplification that happens. Using the Fourier transform of a convolution is the product of the Fourier transforms, we find $\widehat{p}(y) = \widehat{p}_1(y) \cdots \widehat{p}_N(y)$; in the special case when the random variables are identically distributed, this simplifies further to just $\widehat{p}_1(y)^N$. Now 'all' (and, sadly, it's a big 'all') we need to do to prove the Central Limit Theorem in the case when all the densities are equal is show that, as $N \to \infty$, $\widehat{p}_1(y)^N$ converges to the Fourier transform of something normally distributed (remember we haven't normalized our sum), and the inverse Fourier transform is uniquely determined and is normally distributed.

21.3 Proof of the Central Limit Theorem

We can now sketch the proof of the Central Limit Theorem. The version we prove is a bit more general than our earlier work. We no longer need to assume the moment generating function exists. To really grasp the nuts and bolts of this proof, we encourage you to provide the complete details to the series of problems below, each of which gives another needed input for the proof.

Theorem 21.3.1 (Central Limit Theorem) *Let $X_1, \dots, X_N$ be independent, identically distributed random variables whose first three moments are finite and whose probability density decays sufficiently rapidly. Denote the mean by μ and the variance by σ^2, let*

$$\overline{X}_N \ = \ \frac{X_1 + \cdots + X_N}{N}$$

and set

$$Z_N \ = \ \frac{\overline{X}_N - \mu}{\sigma/\sqrt{N}}.$$

Then as $N \to \infty$, the distribution of Z_N converges to the standard normal.

We highlight the key steps, but we do not provide detailed justifications (which would require several standard lemmas about the Fourier transform; see for example [SS1]). Without loss of generality, we may consider the case where we have a probability density p on $\mathbb{R}$ that has mean zero and variance one (see §20.4). We assume the density decays sufficiently rapidly so that all convolution integrals that arise below converge.

Specifically, our density p satisfies

$$\int_{-\infty}^{\infty} xp(x)dx \ = \ 0, \quad \int_{-\infty}^{\infty} x^2 p(x)dx \ = \ 1, \quad \int_{-\infty}^{\infty} |x|^3 p(x)dx \ < \ \infty. \qquad (21.2)$$

Assume $X_1, X_2, \dots$ are independent identically distributed random variables drawn from p; thus, $\text{Prob}(X_i \in [a,b]) = \int_a^b p(x)dx$. Define $S_N = \sum_{i=1}^N X_i$. Recall the standard Gaussian (mean zero, variance one) has density $\exp(-x^2/2)/\sqrt{2\pi}$.

As we are assuming $\mu = 0$ and $\sigma = 1$, we have $Z_N = \frac{(X_1+\cdots+X_N)/N}{1/\sqrt{N}} = \frac{X_1+\cdots+X_N}{\sqrt{N}}$, so $Z_N = S_N/\sqrt{N}$. We must show $S_N/\sqrt{N}$ converges in probability to the standard Gaussian:

$$\lim_{N\to\infty} \text{Prob}\left(\frac{S_N}{\sqrt{N}} \in [a,b]\right) \ = \ \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx.$$

We sketch the proof. The Fourier transform of p is

$$\widehat{p}(y) \ = \ \int_{-\infty}^{\infty} p(x) e^{-2\pi i x y} dx.$$

Clearly, $|\widehat{p}(y)| \leq \int_{-\infty}^{\infty} p(x)dx = 1$, and $\widehat{p}(0) = \int_{-\infty}^{\infty} p(x)dx = 1$.

Claim 1 *One useful property of the Fourier transform is that the derivative of $\widehat{g}$ is the Fourier transform of $2\pi i x g(x)$; thus, differentiation (hard) is converted to multiplication (easy). Explicitly, show*

$$\widehat{g}'(y) \;=\; \int_{-\infty}^{\infty} 2\pi i x \cdot g(x) e^{-2\pi i x y} dx.$$

If g is a probability density, note $\widehat{g}'(0) = 2\pi i \mathbb{E}[x]$ and $\widehat{g}''(0) = -4\pi^2 \mathbb{E}[x^2]$.

The above claim shows why it's, at least potentially, natural to use the Fourier transform to analyze probability distributions. The mean and variance (and the higher moments) are simple multiples of the derivatives of $\widehat{p}$ at zero. By Claim 1, as p has mean zero and variance one, $\widehat{p}'(0) = 0$, $\widehat{p}''(0) = -4\pi^2$. We Taylor expand $\widehat{p}$ (we do not justify that such an expansion exists and converges; however, in most problems of interest this can be checked directly, and this is the reason we need technical conditions about the higher moments of p), and find near the origin that

$$\widehat{p}(y) \quad = \quad 1 + \frac{\widehat{p}''(0)}{2} y^2 + \cdots \;=\; 1 - 2\pi^2 y^2 + O(y^3). \tag{21.3}$$

Near the origin, the above shows $\widehat{p}$ looks like a concave down parabola. There is no y term as $\widehat{p}'(0) - 0$. Here $O(y^3)$ is **big-Oh notation** for an error at most on the order of y^3; see §20.6 for more on big-Oh notation.

From §21.2, we know

- The probability that $X_1 + \cdots + X_N \in [a, b]$ is $\int_a^b (p * \cdots * p)(z)dz$.
- The Fourier transform converts convolution to multiplication. If $\text{FT}[f](y)$ denotes the Fourier transform of f evaluated at y, then we have

$$\text{FT}[p * \cdots * p](y) \;=\; \widehat{p}(y) \cdots \widehat{p}(y).$$

However, we do not want to study the distribution of $X_1 + \cdots + X_N = x$, but rather the distribution of $S_N = \frac{X_1 + \cdots + X_N}{\sqrt{N}} = x$.

Claim 2 *If $B(x) = A(cx)$ for some fixed $c \neq 0$, show $\widehat{B}(y) = \frac{1}{c}\widehat{A}\left(\frac{y}{c}\right)$.*

Claim 3 *Show that if the probability density of $X_1 + \cdots + X_N = x$ is $(p * \cdots * p)(x)$ (i.e., the distribution of the sum is given by $p * \cdots * p$), then the probability density of $\frac{X_1 + \cdots + X_N}{\sqrt{N}} = x$ is $(\sqrt{N}p * \cdots * \sqrt{N}p)(x\sqrt{N})$. By Exercise 2, show*

$$FT\left[(\sqrt{N}p * \cdots * \sqrt{N}p)(x\sqrt{N})\right](y) \;=\; \left[\widehat{p}\left(\frac{y}{\sqrt{N}}\right)\right]^N.$$

The above claims allow us to determine the Fourier transform of the distribution of S_N. It's just $\left[\widehat{p}\left(\frac{y}{\sqrt{N}}\right)\right]^N$. We take the limit as $N \to \infty$ for **fixed** y. From (21.3), $\widehat{p}(y) = 1 - 2\pi^2 y^2 + O(y^3)$. Thus we have to study

$$\left[1 - \frac{2\pi^2 y^2}{N} + O\left(\frac{y^3}{N^{3/2}}\right)\right]^N.$$

For any fixed y, we have

$$\lim_{N\to\infty}\left[1-\frac{2\pi^2y^2}{N}+O\left(\frac{y^3}{N^{3/2}}\right)\right]^N \;=\; e^{-2\pi y^2}. \tag{21.4}$$

There are two definitions of e^x (see the end of §B.3); while we normally work with the infinite sum expansion, in this case the product formulation is far more useful:

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

(you might recall this formula from compound interest). Of course, this isn't a fully rigorous proof. The problem is we don't have *exactly* the same setting as the definition, as we have the lower order error $O(y^3/N^{3/2})$. A great way to proceed rigorously is to take the logarithm of both sides of (21.4), and notice that as $N\to\infty$ the two sides are equal.

Claim 4 *Show that the Fourier transform of* $e^{-2\pi y^2}$ *at* x *is* $\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$. Hint: *This problem requires contour integration from complex analysis. If you haven't had a course in Complex Analysis, this is another black box result to return to after taking more math.*

We would like to conclude that as the Fourier transform of the distribution of S_N converges to $e^{-2\pi y^2}$ and the Fourier transform of $e^{-2\pi y^2}$ is $\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$, then the distribution of S_N equalling x converges to $\frac{1}{\sqrt{2\pi}}\,e^{-x^2/2}$. Justifying these statements requires some results from complex analysis. We refer the reader to [Fe] for the details, which completes the proof. □

The key point in the proof is that we used Fourier Analysis to study the sum of independent identically distributed random variables, as Fourier transforms convert convolution to multiplication. The universality is due to the fact that *only* terms up to the second order contribute in the Taylor expansions. Explicitly, for "nice" p the distribution of S_N converges to the standard Gaussian, independent of the fine structure of p. The fact that p has mean zero and variance one is really just a normalization to study all probability distributions on a similar scale; see Section 20.4.

The higher order terms are important in determining the *rate* of convergence in the Central Limit Theorem (see [Fe] for details and [KonMi] for an application to Benford's Law).

Here are some good problems to think about.

- Modify the proof to deal with the case of p having mean μ and variance σ^2.
- For reasonable assumptions on p, estimate the rate of convergence to the Gaussian.
- Let p_1, p_2 be two probability densities satisfying (21.2). Consider $S_N = X_1 + \cdots + X_N$, where for each i, X_1 is equally likely to be drawn randomly from p_1 *or* p_2. Show the Central Limit Theorem is still true in this case. What if we instead had a fixed, finite number of such distributions $p_1, \dots, p_k$, and for each i we draw X_i from p_j with probability q_j (of course, $q_1 + \cdots + q_k = 1$)?

21.4 Summary

There are, not surprisingly, many structural similarities with the Fourier analysis proof and the moment generating function proof. If you forgive the pun, this is to be expected. Why? The moment generating function is $M_X(t) = \mathbb{E}[e^{tX}]$ while the Fourier transform (or the characteristic function) is $\mathbb{E}[e^{-2\pi i y X}]$. The two are thus related by $t \mapsto -2\pi i y$, but what a difference that i makes! The characteristic function exists for any density, which isn't the case for the moment generating function.

This relation sheds light on Claim 1. It should now be clear why derivatives of the Fourier transform are related to the moments of the density; it's because the Fourier transform is a very close cousin of the moment generating function, where the derivatives are the moments.

There's essentially an unlimited number of things one can do in math. We can define almost anything; the question is which definitions are useful, which definitions lead to good viewpoints. In Theorem 21.2.1 we saw the Fourier transform of a convolution is the product of the Fourier transforms. When studying sums of random variables, it's hard not to try to use convolutions, as that is the most natural way to find the density. As the Fourier transform interacts well with convolutions, it shouldn't be surprising that it enters the proof.

21.5 Exercises

Problem 21.5.1 *Find sufficient conditions on f and g so that the Cauchy-Schwarz inequality holds as an equality. Try to find the weakest such conditions.*

Problem 21.5.2 *Find the Fourier transform of $f(x) = e^{-|x|}$.*

Problem 21.5.3 *Show that $\widehat{g}'(y) = \int_{-\infty}^{\infty} 2\pi i x \cdot g(x) e^{-2\pi i x y} dx$.*

Problem 21.5.4 *Prove Claim 2, that if $B(x) = A(cx)$ for some fixed $c \neq 0$, then $\widehat{B}(y) = \frac{1}{c}\widehat{A}\left(\frac{y}{c}\right)$.*

Problem 21.5.5 *Find constants C_{2m} such that $\int_{-\infty}^{\infty} \frac{C_{2m}}{1+x^{2m}} dx = 1$ for $m \in \{1, 2, 3, 4\}$. (Using techniques from Complex Analysis it's possible to find C_{2m} for all m.)*

Problem 21.5.6 *Using the Taylor series of e^x, $\cos x$ and $\sin x$, 'prove' that $e^{ix} = \cos x + i \sin x$.*

Problem 21.5.7 *Show that $\mathcal{L}^{-}1(\mathcal{L}f)(s))(x) = f(x)$.*

Problem 21.5.8 *Show that $\mathcal{F}^{-}1(\mathcal{F}f)(s))(x) = f(x)$.*

Problem 21.5.9 *Show that any infinitely differentiable functions with compact support (that is, they are only non-zero on a finite interval) are in the Schwartz space.*

Problem 21.5.10 *Explain how a function can be smooth and have compact support.*

Problem 21.5.11 *Show that the Cauchy distribtuion is not in the Schwartz space.*

Problem 21.5.12 *Use characteristic functions to show that the Cauchy distribution is strictly stable, that is, the sum of two identically distributed Cauchy distributions is a rescaled version of the original distribtution.*

Appendix E

Complex Analysis and the Central Limit Theorem

In Chapter 20 we gave a proof of the Central Limit Theorem using generating functions; unfortunately that proof isn't complete as it assumed some results from Complex Analysis. Moreover, we had to assume the moment generating function existed, which isn't always true.

We tried again in Chapter 21; we proved the Central Limit Theorem by using Fourier analysis. Instead of using the moment generating function, which can fail to even exist, this time we used the Fourier transform (also called the characteristic function), which has the very nice and useful property of actually existing! Unfortunately, here too we needed to appeal to some results from Complex Analysis.

This leaves us in a quandary, where we have a few options.

1. We can just accept as true some results from Complex Analysis and move on.
2. We can try and find yet another proof, this time one that doesn't need Complex Analysis.
3. We can drop everything and take a crash course in Complex Analysis.

This chapter is for those who like the third option. We'll explain some of the key ideas of complex analysis, in particular we'll show why it's such a different subject than real analysis. Obviously, it helps to have seen real analysis, but if you're comfortable with Taylor series and basic results on convergence you'll be fine.

It turns out that assuming a function of a real variable is differentiable doesn't mean too much, but assume a function of a complex variable is differentiable and all of a sudden doors are opening everywhere with additional, powerful facts that must be true. Obviously this chapter can't replace an entire course, nor is that our goal. We want to show you some of the key ideas of this beautiful subject, and hopefully when you finish reading you'll have a better sense of why the black-box results from Complex Analysis (Theorems 20.5.3 and 20.5.4) are true.

This chapter is meant to supplement our discussions on moment generating functions and proofs of the Central Limit Theorem. We thus assume the reader is familiar with the notation and concepts from Chapters 19 through 21.

E.1 Warnings from real analysis

The following example is one of my favorites from real analysis. It indicates why real analysis is hard, almost surely much harder than you might expect. Consider the function $g : \mathbb{R} \to \mathbb{R}$ given by

$$g(x) \;=\; \begin{cases} e^{-1/x^2} & \text{if } x \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{E.1}$$

Using the definition of the derivative and L'Hopital's rule, we can show that g is infinitely differentiable, and all of its derivatives at the origin vanish. For example,

$$\begin{aligned} g'(0) &= \lim_{h\to 0} \frac{e^{-1/h^2} - 0}{h} \\ &= \lim_{h\to 0} \frac{1/h}{e^{1/h^2}} \\ &= \lim_{k\to\infty} \frac{k}{e^{k^2}} \\ &= \lim_{k\to\infty} \frac{1}{2ke^{k^2}} \;=\; 0, \end{aligned}$$

where we used **L'Hopital's rule** in the last step ($\lim_{k\to\infty} A(k)/B(k) = \lim_{k\to\infty} A'(k)/B'(k)$ if $\lim_{k\to\infty} A(k) = \lim_{k\to\infty} B(k) = \infty$). (We replaced h with $1/k$ as this allows us to re-express the quantities above in a familiar form, one where we can apply L'Hopital's rule.) A similar analysis shows that the n^{th} derivative vanishes at the origin for all n, i.e., $g^{(n)}(0) = 0$ for all positive integer n. If we consider the Taylor series for g about 0, we find

$$g(x) \;=\; g(0) + g'(0)x + \frac{g''(0)x^2}{2!} + \cdots \;=\; \sum_{n=0}^{\infty} \frac{g^{(n)}(0)x^n}{n!} \;=\; 0;$$

however, clearly $g(x) \neq 0$ if $x \neq 0$. We are thus in the ridiculous case where the Taylor series (which converges for all x!) only agrees with the function when $x = 0$. This isn't that impressive, as the Taylor series is *forced* to agree with the original function at 0, as both are just $g(0)$.

We can learn a lot from the above example. The first is that it's possible for a Taylor series to converge for all x, but only agree with the function at one point! It's not too impressive to agree at just one point, as by construction the Taylor series *has* to agree at that point of expansion. The second, which is far more important, is that *a Taylor series does not uniquely determine a function!* For example, both $\sin x$ and $\sin x + g(x)$ (with $g(x)$ the function from equation (E.1)) have the same Taylor series about $x = 0$.

The reason this is so important for us is that we want to understand when a moment generating function uniquely determines a probability distribution. If our distribution was discrete, there was no problem (Theorem 19.6.5). For continuous distributions, however, it's much harder, as we saw in equation (19.6.5) where we met two densities that had the same moments.

Apparently, we must impose some additional conditions for continuous random variables. For discrete random variables, it was enough to know all the moments;

this doesn't suffice for continuous random variables. What should those conditions be?

Recall that if we have a random variable X with density f_X, its k^{th} moment, denoted by μ'_k, is defined by

$$\mu'_k \;=\; \int_{-\infty}^{\infty} x^k f_X(x)dx.$$

Let's consider again the pair of functions in equation (19.6.5). A nice calculus exercise shows that $\mu'_k = e^{k^2/2}$. This means that the moment generating function is

$$M_X(t) \;=\; \sum_{k=0}^{\infty} \frac{\mu'_k t^k}{k!} \;=\; \sum_{k=0}^{\infty} \frac{e^{k^2/2} t^k}{k!}.$$

For what t does this series converge? Amazingly, this series converges *only* when $t = 0$! To see this, it suffices to show that the terms do not tend to zero. As $k! \le k^k$, for any fixed t, for k sufficiently large $t^k/k! \ge (t/k)^k$; moreover, $e^{k^2/2} = (e^{k/2})^k$, so the k^{th} term is at least as large as $(e^{k/2}t/k)^k$. For any $t \neq 0$, this clearly does not tend to zero, and thus the moment generating function has a radius of convergence of zero!

This leads us to the following conjecture: *If the moment generating function converges for $|t| < \delta$ for some $\delta > 0$, then it uniquely determines a density.* We'll explore this conjecture below.

E.2 Complex Analysis and Topology Definitions

Our purpose here is to give a flavor of what kind of inputs are needed to ensure that a moment generating function uniquely determines a probability density. We first collect some definitions, and then state some useful results from complex analysis.

Definition E.2.1 (Complex variable, complex function) *Any complex number z can be written as $z = x + iy$, with x and y real and $i = \sqrt{-1}$. We denote the set of all complex numbers by $\mathbb{C}$. A complex function is a map f from $\mathbb{C}$ to $\mathbb{C}$; in other words $f(z) \in \mathbb{C}$. Frequently one writes $x = \Re(z)$ for the* ***real part****, $y = \Im(z)$ for the* ***imaginary part****, and $f(z) = u(x, y) + iv(x, y)$ with u and v functions from $\mathbb{R}^2$ to $\mathbb{R}$.*

There are many ways to write complex numbers. The most common is the definition above; however, a polar coordinate approach is sometimes useful. One of the most remarkable relations in all of mathematics is

$$e^{i\theta} \;=\; \cos\theta + i\sin\theta.$$

There are several ways to see this, depending on how much math you want to assume. One way is to use the Taylor series expansions for the exponential, sine and cosine functions. This gives another way of writing complex numbers; instead of $1 + i$ we could write $\sqrt{2}\exp(i\pi/4)$. A particularly interesting choice of θ is π, which gives $e^{i\pi} = -1$, a beautiful formula involving many of the most important constants in mathematics!

Noting $i^2 = -1$, it isn't too hard to show that

$$\begin{aligned}(a+ib)+(x+iy) &= (a+x)+i(b+y)\\ (a+ib)\cdot(x+iy) &= (ax-by)+i(ay+bx).\end{aligned}$$

The **complex conjugate** of $z = x+iy$ is $\overline{z} := x - iy$, and we define the **absolute value** (or the **modulus** or **magnitude**) of z to be $\sqrt{z\overline{z}}$, and denote this by $|z|$. This is real valued, and equals $\sqrt{x^2+y^2}$. If we were to write z as a vector, it would be $z = (x,y)$; note that in this case we see that $|z|$ equals the length of the corresponding vector.

We can write almost anything as an example of a complex function; one possible function is $f(z) = z^2 + |z|$. The question is when is such a function differentiable in z, and what does that differentiability entail. Actually, before we answer this we first need to state what it means for a complex function to be differentiable!

Definition E.2.2 (Differentiable) *We say a complex function f is* ***(complex) differentiable*** *at z_0 if it's differentiable with respect to the complex variable z, which means*

$$\lim_{h\to 0}\frac{f(z_0+h)-f(z_0)}{h}$$

exists, where h tends to zero along any *path in the complex plane. If the limit exists we write $f'(z_0)$ for the limit. If f is differentiable, then $f(x+iy) = u(x,y)+iv(x,y)$ satisfies the* ***Cauchy-Riemann equations****:*

$$f'(z) \;=\; \frac{\partial u}{\partial x}+i\frac{\partial v}{\partial x} \;=\; -i\frac{\partial u}{\partial y}+\frac{\partial v}{\partial y}$$

(one direction is easy, arising from sending $h \to 0$ along the paths $\widetilde{h}$ and $i\widetilde{h}$, with $\widetilde{h} \in \mathbb{R}$).

Here's a quick hint to see why differentiability implies the Cauchy-Riemann equations – try and fill in the details. Since the derivative exists at z_0, the key limit is independent of the path we take to the point $x_0 + iy_0$. Consider the path $x + iy_0$ with $x \to x_0$, and the path $x_0 + iy$ with $y \to y_0$, and use results from multivariable calculus on partial derivatives.

Let's explore a bit and see which functions are complex differentiable. We let $h = h_1 + ih_2$ below, with $h \to 0 + 0i$.

- If $f(z) = z$ then

$$\lim_{h\to 0}\frac{f(z+h)-f(z)}{h} \;=\; \lim_{h\to 0}\frac{z+h-z}{h} \;=\; \lim_{h\to 0} 1 \;=\; 1;$$

thus the function is complex differentiable and the derivative is 1.

- If $f(z) = z^2$ then

$$\begin{aligned}
\lim_{h \to 0} \frac{f(z+h) - f(z)}{h} &= \lim_{h \to 0} \frac{(z+h)^2 - z^2}{h} \\
&= \lim_{h \to 0} \frac{z^2 + 2zh + h^2 - z^2}{h} \\
&= \lim_{h \to 0} \frac{2zh + h^2}{h} \\
&= \lim_{h \to 0} (2z + h) \\
&= \lim_{h \to 0} 2z + \lim_{h \to 0} h \\
&= 2z + 0 \;=\; 2z.
\end{aligned}$$

  We're using the following properties of complex numbers: $h/h = 1$ and $2zh + h^2 = (2z+h)h$. Note how similar this is to the real valued analogue, $f(x) = x^2$.

- If $f(z) = \overline{z}$ then

$$\lim_{h \to 0} \frac{f(z+h) - f(z)}{h} \;=\; \lim_{h \to 0} \frac{\overline{z+h} - \overline{z}}{h}.$$

  Unlike the other limits, this one isn't immediately clear. Let's write $z = x + iy$, $h = h_1 + ih_2$ (and of course $\overline{z} = x - iy$, $\overline{h} = h_1 - ih_2$). The limit is

$$\lim_{h \to 0} \frac{x - iy + h - ih_2 - (x - iy)}{h_1 + ih_2} \;=\; \lim_{h \to 0} \frac{h_1 - ih_2}{h_1 + ih_2}.$$

  This limit does not exist; depending on how $h \to 0$ we obtain different answers. For example, if $h_2 = 0$ (traveling along the x-axis) the limit is just $\lim_{h \to 0} h_1/h_1 = 1$, while if $h_1 = 0$ (traveling along the y-axis) the limit is just $\lim_{h \to 0} -ih_2/ih_2 = -1$. Thus this function isn't complex differentiable anywhere, even though it's a fairly straightforward function to define.

If we continue to argue along these lines, we find that a function is complex differentiable if the x and y dependence is in a very special form, namely everything is a function of $z = x + iy$. In other words, we don't allow our function to depend on $\overline{z} = x - iy$. If we could depend on both, we could isolate out x (which is $z + \overline{z}$) and y (which is $(z - \overline{z})/i$). We can begin to see why being complex differentiable once implies that we're complex differentiable infinitely often, namely because of the very special dependence on x and y. Also, in the plane there's really only two ways to approach a point: from above, or from below. In the complex plane, the situation is strikingly different. There are so many ways we can move in two-dimensions, and *each* path must give the same answer if we're to be complex differentiable. This is why differentiability means far more for a complex variable than for a real variable.

To state the needed results from Complex Analysis, we also require some terminology from Point Set Topology. In particular, many of the theorems below deal with open sets. We briefly review their definition and give some examples.

Definition E.2.3 (Open set, closed set) *A subset U of $\mathbb{C}$ is an* ***open set*** *if for any $z_0 \in U$ there's a δ such that whenever $|z - z_0| < \delta$ then $z \in U$ (note δ is allowed to depend on z_0). A set C is* ***closed*** *if its* ***complement****, $\mathbb{C} \setminus C$, is open.*

The following are examples of open sets in $\mathbb{C}$.

1. $U_1 = \{z : |z| < r\}$ for any $r > 0$. This is usually called the **open ball of radius** r centered at the origin.
2. $U_2 = \{z : \Re(z) > 0\}$. To see this is open, if $z_0 \in U_2$ then we can write $z_0 = x_0 + iy_0$, with $x_0 > 0$. Letting $\delta = x_0/2$, for $z = x + iy$ we see that if $|z - z_0| < \delta$ then $|x - x_0| < x_0/2$, which implies $x > x_0/2 > 0$; U_2 is often called the open **right half-plane**.

For examples of closed sets, consider the following.

1. $C_1 = \{z : |z| \leq r\}$. Note that if we take z_0 to be any point on the boundary, then the ball of radius δ centered at z_0 will contain points more than r units from the origin, and thus C_1 isn't open. A little work shows, however, that C_1 is closed (in fact, C_1 is called the **closed ball of radius** r about the origin). We prove it's closed by showing its complement is open. What we need to do is show that, given any point in the complement, there's a small ball about that point entirely contained in the complement. I urge you to draw a picture for the following argument. If $z_0 \in \mathbb{C} \setminus C_1$ then $|z_0| > r$ (as otherwise it would be inside C_1). If we take $\delta < \frac{|z_0| - r}{2}$ then after some algebra we'll find that if $|z - z_0| < \delta$ then $z \in \mathbb{C} \setminus C_1$. Thus $\mathbb{C} \setminus C_1$ is open, so C_1 is closed.
2. $C_2 = \{z : \Re(z) \geq 0\}$. To see this set isn't open, consider any $z_0 = iy$ with $y \in \mathbb{R}$. A similar calculation as the one we did for U_2 or C_1 shows C_2 is closed.

For a set that is neither open nor closed, consider $S = U_1 \cup C_2$.

We now state two of the most important properties a complex function could have. One of the most important results in the subject is that these two seemingly very different properties are actually equivalent!

Definition E.2.4 (Holomorphic, analytic) *Let U be an open subset of $\mathbb{C}$, and let f be a complex function. We say f is* ***holomorphic*** *on U if f is differentiable at every point $z \in U$, and we say f is* ***analytic*** *on U if f has a series expansion that converges and agrees with f on U. This means that for any $z_0 \in U$, for z close to z_0 we can choose a_n's such that*

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n.$$

As alluded to above, saying a function of a complex variable is differentiable turns out to imply *far* more than saying a function of a real variable is differentiable, as the following theorem shows us.

Theorem E.2.5 *Let f be a complex function and U an open set. Then f is holomorphic on U if and only if f is analytic on U, and the series expansion for f is its Taylor series.*

The above theorem is amazing; its result seems to good to be true. Namely, as soon as we know f is differentiable once, it's infinitely (real) differentiable and f agrees with its Taylor series expansion! This is very different than what happens in the case of functions of a real variable. For instance, the function

$$h(x) \ = \ x^3 \sin(1/x) \tag{E.2}$$

is differentiable once and only once at $x = 0$, and while the function $g(x)$ from (E.1) is infinitely differentiable, the Taylor series expansion only agrees with $g(x)$ at $x = 0$. Complex analysis is a *very* different subject than real analysis!

The next theorem provides a very nice condition for when a function is identically zero. It involves the notion of a limit or accumulation point, which we define first.

Definition E.2.6 (Limit or accumulation point) *We say z is a **limit** (or an **accumulation) point** of a sequence $\{z_n\}_{n=0}^{\infty}$ if there exists a subsequence $\{z_{n_k}\}_{k=0}^{\infty}$ converging to z.*

Let's do some examples to clarify the definitions.

1. If $z_n = 1/n$, then 0 is a limit point.
2. If $z_n = \cos(\pi n)$ then there are two limit points, namely 1 and -1. (If $z_n = \cos(n)$ then *every* point in $[-1, 1]$ is a limit point of the sequence, though this is harder to show.)
3. If $z_n = (1 + (-1)^n)^n + 1/n$, then 0 is a limit point. We can see this by taking the subsequence $\{z_1, z_3, z_5, z_7, \dots\}$; note the subsequence $\{z_0, z_2, z_4, \dots\}$ diverges to infinity.
4. Let z_n denote the number of distinct prime factors of n. Then every positive integer is a limit point! For example, let's show 5 is a limit point. The first five primes are 2, 3, 5, 7 and 11; consider $N = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 = 2310$. Consider the subsequence $\{z_N, z_{N^2}, z_{N^3}, z_{N^4}, \dots\}$; as N^k has exactly 5 distinct prime factors for each k, 5 is a limit point.
5. If $z_n = n^2$ then there are no limit points, as $\lim_{n\to\infty} z_n = \infty$.
6. Let z_0 be any odd, positive integer, and set

$$z_{n+1} \ = \ \begin{cases} 3z_n + 1 & \text{if } z_n \text{ is odd} \\ z_n/2 & \text{if } z_n \text{ is even.} \end{cases}$$

   It's *conjectured* that 1 is always a limit point (and if some $z_m = 1$, then the next few terms have to be $4, 2, 1, 4, 2, 1, 4, 2, 1, \dots$, and hence the sequence cycles). This is the famous $3x + 1$ **problem**. Kakutani called it a conspiracy to slow down American mathematics because of the amount of time people spent on this; Erdös said mathematics isn't yet ready for such problems. See [Lag1, Lag2, Lag3] for some nice expositions, but be warned that this problem can be addictive!

We can now state the theorem which, for us, is the most important result from Complex Analysis. It's the basis of the black box results.

Theorem E.2.7 *Let f be an analytic function on an open set U, with infinitely many zeros $z_1, z_2, z_3, \dots$. If $\lim_{n\to\infty} z_n \in U$, then f is identically zero on U. In other words, if a function is zero along a sequence in U whose accumulation point is also in U, then that function is identically zero in U.*

Note the above is *very* different than what happens in real analysis. Consider again the function from (E.2),

$$h(x) \;=\; x^3 \sin(1/x).$$

This function is continuous and differentiable. It's zero whenever $x = 1/\pi n$ with n an integer. If we let $z_n = 1/\pi n$, we see this sequence has 0 as a limit point, and our function is also zero at 0 (see Figure E.1). It's clear, however, that this

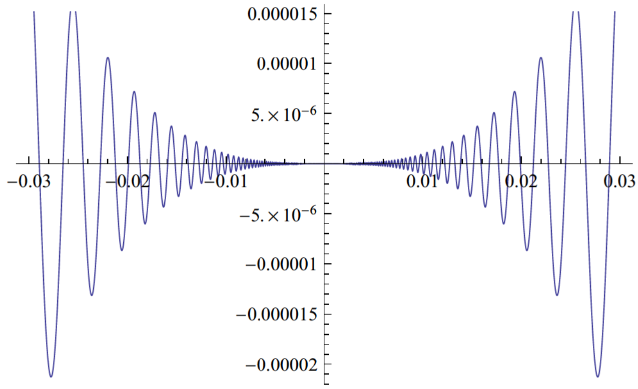


Figure E.1: Plot of $x^3 \sin(1/x)$.

function is *not* identically zero. Yet again, we see a stark difference between real and complex valued functions. As a nice exercise, show that $x^3 \sin(1/x)$ is *not* complex differentiable. It will help if you recall $e^{i\theta} = \cos\theta + i\sin\theta$, or $\sin\theta = (e^{i\theta} - e^{-i\theta})/2$.

E.3 Complex analysis and moment generating functions

We conclude our technical digression by stating a few more very useful facts. The proof of these requires properties of the **Laplace transform**, which is defined by $(\mathcal{L}f)(s) = \int_0^\infty e^{-sx} f(x)dx$. The reason the Laplace transform plays such an important role in the theory is apparent when we recall the definition of the moment generating function of a random variable X with density f:

$$M_X(t) \;=\; \mathbb{E}[e^{tX}] \;=\; \int_{-\infty}^{\infty} e^{tx} f(x)dx;$$

in other words, the moment generating function is the Laplace transform of the density evaluated at $s = -t$.

Remember that if F_X and G_Y are the cumulative distribution functions of the random variables X and Y with densities f and g, then

$$\begin{aligned} F_X(x) &= \int_{-\infty}^{x} f(t)dt \\ G_Y(y) &= \int_{-\infty}^{y} g(v)dv. \end{aligned}$$

We remind the reader of the two important results we assumed in the text (Theorems 20.5.3 and 20.5.4), which we restate below. After stating them we discuss their proofs.

Theorem E.3.1 *Assume the moment generating functions $M_X(t)$ and $M_Y(t)$ exist in a neighborhood of zero (i.e., there's some δ such that both functions exist for $|t| < \delta$). If $M_X(t) = M_Y(t)$ in this neighborhood, then $F_X(u) = F_Y(u)$ for all u. As the densities are the derivatives of the cumulative distribution functions, we have $f = g$.*

Theorem E.3.2 *Let $\{X_i\}_{i\in I}$ be a sequence of random variables with moment generating functions $M_{X_i}(t)$. Assume there's a $\delta > 0$ such that when $|t| < \delta$ we have $\lim_{i\to\infty} M_{X_i}(t) = M_X(t)$ for some moment generating function $M_X(t)$, and all moment generating functions converge for $|t| < \delta$. Then there exists a unique cumulative distribution function F whose moments are determined from $M_X(t)$ and for all x where $F_X(x)$ is continuous, $\lim_{i\to\infty} F_{X_i}(x) = F_X(x)$.*

The proof of these theorems follow from results in complex analysis, specifically the Laplace and Fourier inversion formulas. To give an example as to how the results from complex analysis allow us to prove results such as these, we give most of the details in the proof of the next theorem. We *deliberately* do not try and prove the following result in as great generality as possible!

Theorem E.3.3 *Let X and Y be two continuous random variables on $[0, \infty)$ with continuous densities f and g, all of whose moments are finite and agree. Suppose further that:*

1. *There is some $C > 0$ such that for all $c \leq C$, $e^{(c+1)t} f(e^t)$ and $e^{(c+1)t} g(e^t)$ are Schwartz functions (see Definition 21.1.3). This isn't a terribly restrictive assumption; f and g need to have decay in order for all moments to exist and be finite. As we're evaluating f and g at e^t and not t, there's enormous decay here. The meat of the assumption is that f and g are infinitely differentiable and their derivatives decay.*

2. *The (not necessarily integral) moments*

$$\mu'_{r_n}(f) \ = \ \int_0^\infty x^{r_n} f(x) dx \quad \text{and} \quad \mu'_{r_n}(g) \ = \ \int_0^\infty x^{r_n} g(x) dx$$

agree for some sequence of non-negative real numbers $\{r_n\}_{n=0}^\infty$ which has a finite accumulation point (i.e., $\lim_{n\to\infty} r_n = r < \infty$).

Then $f = g$ (in other words, knowing all these moments uniquely determines the probability density).

Proof: We sketch the proof, which is long and sadly a bit technical. Remember the purpose of this proof is to highlight why our needed results from Complex Analysis are true. Feel free to skim or skip the proof, but we urge you to read the example at the end of this section, where we return to the two densities that are causing us so much heartache. Let $h(x) = f(x) - g(x)$, and define

$$A(z) \ = \ \int_0^\infty x^z h(x) dx.$$

Note that $A(z)$ exists for all z with real part non-negative. To see this, let $\Re(z)$ denote the real part of z, and let k be the unique non-negative integer with $k \leq \Re(z) < k+1$. Then $x^{\Re z} \leq x^k + x^{k+1}$, and

$$\begin{aligned} |A(z)| \ &\leq \ \int_0^\infty x^{\Re(z)} \left[|f(x)| + |g(x)|\right] dx \\ &\leq \ \int_0^\infty (x^k + x^{k+1}) f(x) dx + \int_0^\infty (x^k + x^{k+1}) g(x) dx \ = \ 2\mu'_k + 2\mu'_{k+1}. \end{aligned}$$

Results from analysis now imply that $A(z)$ exists for all z. The key point is that A is also differentiable. Interchanging the derivative and the integration (which can be justified; see Theorem B.2.2), we find

$$A'(z) \ = \ \int_0^\infty x^z (\log x) h(x) dx.$$

To show that $A'(z)$ exists, we just need to show this integral is well-defined. There are only two potential problems with the integral, namely when $x \to \infty$ and when $x \to 0$. For x large, $x^z \log x \leq x^{\Re(z)+1}$ and thus the rapid decay of h gives

$\left|\int_1^\infty x^z(\log x)h(x)dx\right| < \infty$. For x near 0, $h(x)$ looks like $h(0)$ plus a small error (remember we're assuming f and g are continuous); thus there's a C so that $|h(x)| \le C$ for $|x| \le 1$. Note

$$\lim_{\epsilon\to 0}\int_\epsilon^1 \left|\int_0^\infty x^z(\log x)h(x)dx\right| \quad \le \quad \lim_{\epsilon\to 0} 1\int_\epsilon^1 1\cdot(-\log x)\cdot C dx.$$

The anti-derivative of $\log x$ is $x\log x - x$, and $\lim_{\epsilon\to 0}(\epsilon\log\epsilon - \epsilon) = 0$. This is enough to prove that this integral is bounded, and thus from results in analysis we get $A'(z)$ exists.

We (finally!) use our results from complex analysis. As A is differentiable once, it's infinitely differentiable and it equals its Taylor series for z with $\Re(z) > 0$. Therefore A is an analytic function which is zero for a sequence of z_n's with an accumulation point, and thus it's identically zero. This is spectacular – initially we only knew $A(z)$ was zero if z was a positive integer or if z was in the sequence $\{r_n\}$; we now know it's zero for all z with $\Re(z) > 0$. This remarkable conclusion comes from complex analysis; it's here that we use it.

We change variables, and replace x with e^t and dx with $e^t dt$. The range of integration is now $-\infty$ to ∞, and we set $\mathfrak{h}(t)dt = h(e^t)e^t dt$. We now have

$$A(z) \;=\; \int_{-\infty}^\infty e^{tz}\mathfrak{h}(t)dt \;=\; 0.$$

Choosing $z = c + 2\pi i y$ with c less than the C from our hypotheses gives

$$A(c+2\pi i y) \;=\; \int_{-\infty}^\infty e^{2\pi i t y}\left[e^{ct}\mathfrak{h}(t)\right] dt \;=\; 0.$$

Our assumptions imply that $e^{ct}\mathfrak{h}(t)$ is a Schwartz function, and thus it has a unique inverse Fourier transform. As we know this transform is zero, it implies that $e^{ct}\mathfrak{h}(t) = 0$, or $h(x) = 0$, or $f(x) = g(x)$. □

We needed the analysis at the end on the inverse Fourier transform as our goal is to show that $f(x) = g(x)$, not that $A(z) = 0$. It seems absurd that $A(z)$ could identically vanish without $f = g$, but we must rigorously show this.

What if we lessen our restrictions on f and g; perhaps one of them isn't continuous? Perhaps there's a unique continuous probability distribution attached to a given sequence of moments such as in the above theorem, but if we allow non-continuous distributions there could be additional possibilities. This topic is beyond the scope of this book, requiring more advanced results from analysis; however, we wanted to point out where the dangers lie, where we need to be careful.

After proving Theorem E.3.3, it's natural to go back to the two densities that are causing so much trouble, namely (see (19.2))

$$\begin{aligned} f_1(x) &= \frac{1}{\sqrt{2\pi x^2}}\, e^{-(\log^2 x)/2} \\ f_2(x) &= f_1(x)\left[1+\sin(2\pi\log x)\right]. \end{aligned}$$

We know these two densities have the same integral moments (their k^{th} moments are $e^{k^2/2}$ for k a non-negative integer). These functions have the correct decay; note

$$e^{(c+1)t} f_1(e^t) \ = \ e^{(c+1)t} \cdot \frac{e^{-t^2/2}}{\sqrt{2\pi e^t}},$$

which decays fast enough for any c to satisfy the assumptions of Theorem E.3.3. As these two densities are not the same, *some* condition must be violated. The only condition left to check is whether or not we have a sequence of numbers $\{r_n\}_{n=0}^{\infty}$ with an accumulation point $r > 0$ such that the $r_n{}^{\text{th}}$ moments agree. Using more results from Complex Analysis (specifically, contour integration), we can calculate the $(a+ib)^{\text{th}}$ moments. We find

$$(a+ib)^{\text{th}} \text{ moment of } f_1 \text{ is } \ e^{(a+ib)^2/2}$$

and

$$(a+ib)^{\text{th}} \text{ moment of } f_1 \text{ is } \ e^{(a+ib)^2/2} + \frac{i}{2}\left(e^{(a+i(b-2\pi))^2/2} - e^{(a+i(b+2\pi))^2/2}\right).$$

While these moments agree for $b = 0$ and a a positive integer, there's no sequence of real moments having an accumulation point where they agree. To see this, note that when $b = 0$ the a^{th} moment of f_2 is

$$e^{a^2/2} + e^{(a-2i\pi)^2/2}\left(1 - e^{4ia\pi}\right), \tag{E.3}$$

and this is never zero unless a is a half-integer (i.e., $a = k/2$ for some integer k). In fact, the reason we wrote (E.3) as we did was to highlight the fact that it's only zero when a is a half-integer. Exponentials of real or complex numbers are never zero, and thus the only way this can vanish is if $1 = e^{4ia\pi}$. Recalling that $e^{i\theta} = \cos\theta + i\sin\theta$, we see that the vanishing of the a^{th} moment is equivalent to $1 - \cos(4\pi a) - i\sin(4\pi a) = 0$; the only way this can happen is if $a = k/2$ for some k. If this happens, the cosine term is 1 and the sine term is 0.

E.4 Exercises

Problem E.4.1 *Let $f(x) = x^3\sin(1/x)$ for $x \neq 0$ and set $f(0) = 0$. (a) Show that f is differentiable once when viewed as a function of a real variable, but that it is not differentiable twice. (b) Show that f is not differentiable when viewed as a function of a complex variable z; it might be useful to note that $\sin u = (e^{iu} - e^{-iu})/2i$.*

Problem E.4.2 *If we're told that all the moments of f are finite and f is infinitely differentiable, must there be some C such that for all $c < C$ we have $e^{(c+1)t} f(e^t)$ is a Schwartz function?*