LAPLACE'S METHOD, FOURIER ANALYSIS, AND RANDOM WALKS ON $\mathbb{Z}^d$

STEVEN P. LALLEY

1. Laplace's method of asymptotic expansion

1.1. **Stirling's formula.** Laplace's approach to Stirling's formula is noteworthy first, because it makes a direct connection with the Gaussian (normal) distribution (whereas in other approaches the Gaussian distribution enters indirectly, or not at all), and second, because it provides a general strategy for the asymptotic approximation of a large class of integrals with a large parameter. Stirling's formula states that as $n \to \infty$,

$$n! \sim \sqrt{2\pi n} n^n e^{-n}. \tag{1}$$

The symbol $\sim$ means that the ratio of the two sides approaches 1 as $n \to \infty$, equivalently, that the relative error in the approximation goes to 0. Laplace's starting point is the gamma function representation

$$n! = \int_0^\infty x^n e^{-x} \, dx, \tag{2}$$

which can be verified by induction, using an integration by parts to reduce the power x^n to x^{n-1}. The integrand achieves its max at $x = n$ (as you should check), and the value there is $n^n e^{-n}$. This already accounts for the largest factors in the Stirling approximation. Factoring this out gives

$$n! = n^n e^{-n} \int_0^\infty \left(\frac{x}{n}\right)^n e^{-(x-n)} \, dx.$$

Now the substitution $y = x/n$ will have the effect of moving the location of the maximum of the integrand from $x = n$ to $y = 1$, and leaves us with the formula

$$n! = n^n e^{-n} n \int_0^\infty y^n e^{-n(y-1)} \, dy. \tag{3}$$

Thus, to complete the proof of Stirling's formula it is enough to prove that the last integral is well-approximated by $\sqrt{2\pi}/\sqrt{n}$ for large n.

1.2. **Spikes and the bell curve.** At this point you are well-advised to fire up your favorite graphing calculator and plot the integrand in (3) for several values of n, perhaps $n = 4, 400, 40000$ and see what happens. You will see that as n gets larger the integrand *spikes* around the point $y = 1$. Now change the scaling of the $y-$axis, so that the spike is opened up, and you will see what is unmistakably the bell curve. This is the whole point of Laplace's method: for large n nearly all of the integral is accounted for by the spike near 1, and this spike looks more and more (after rescaling the argument) like the normal density, which we know how to integrate.

This is a phenomenon that holds much more generally. Let (a, b) be an open interval and let $g : (a, b) \to \mathbb{R}$ be any twice continuously differentiable function that satisfies the following conditions: for some $x_* \in (a, b)$,

$$
\begin{aligned}
g(x_*) &= 0; \\
-g''(x_*) &= 1/\sigma^2 > 0; \quad \text{and} \\
g(x) &< 0 \quad \text{for all } \; x \neq x_*.
\end{aligned}
\tag{4}
$$

We will be interested in the large-n behavior of the integral

$$
J(n) := \int_a^b e^{ng(x)}\,dx \tag{5}
$$

(Note that the integral in (3) has this form, with $g(x) = \log x - x + 1$.) For any $n \geq 1$ the integrand $g(x)$ attains its maximum *uniquely* at $x = x_*$, and at this point the value is 1. Raising the integrand to the nth power has the effect of sending everything rapidly to 0 *except* in the close vicinity of $x = x_*$; that is, the integrand once again *spikes*. The second assumption in (4) guarantees that the spike is approximately a normal curve, because the other two assumptions imply that the first two terms in the Taylor series for g around $x = x_*$ vanish, leaving the quadratic second term, which by the middle assumption is nondegenerate. Thus, in principle, Laplace's method should apply to any integral of the form (5) provided the function g satisfies the restrictions (4).

Unfortunately there are a few things that might still go wrong. Assumptions (4) ensure that $e^{ng(x)}$ will approach 0 everywhere except at $x = x_*$, but they don't guaranteee that it will do so *uniformly*, nor will that the total area in the tails of the spikes are finite. The following proposition gives relatively simple *sufficient* conditions that keep bad things from happening.

Proposition 1. *Assume in addition to conditions* (4) *that* $J_1 < \infty$, *and that for every* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that* $g(x) < -\varepsilon$ *for all* x *such that* $|x - x_*| > \delta$. *Then as* $n \to \infty$,

$$
J_n \sim \sqrt{\frac{2\pi}{n\sigma}}. \tag{6}
$$

Proof. It suffices to prove that for any $\varepsilon > 0$ the value J_n will eventually be larger than $(1 - \varepsilon)\times$ the right side and smaller than $(1 + \varepsilon)\times$ the right side. Choose $\delta > 0$ such that $g(x) < -\varepsilon$ for all $|x - x_*| > \delta$ and such that

$$
-(1+\varepsilon)^2 \frac{(x - x_*)^2}{2\sigma^2} < g(x) < -(1-\varepsilon)^2 \frac{(x - x_*)^2}{2\sigma^2} \tag{7}
$$

for all $|x - x_*| \leq \delta$. (That this is possible follows from Taylor's theorem and the assumptions (4).) Break the integral J_n into two pieces: the interval $[x_* - \delta, x_* + \delta]$ containing the spike, and the complementary interval(s). Consider the complementary intervals first:

$$
\int_{(a,b)-[x_*-\delta, x_*+\delta]} e^{ng(x)}\,dx \leq \int_{(a,b)} e^{g(x)} e^{-(n-1)\varepsilon}\,dx \leq J_1 e^{-(n-1)\varepsilon}.
$$

By assumption $J_1 < \infty$, so this piece of the integral is bounded by a constant times $e^{-n\varepsilon}$. Exponentials go to 0 much more rapidly than any polynomial, so it will suffice to show that the integral over the spike interval is of polynomial size. (Exercise: Fill in the details.)

Now let's consider the spike. In the interval $|x - x_*| \leq \delta$ the inequalities (7) hold, so in this region the integrand $e^{ng(x)}$ is trapped between two bell curves, one with variance $\sigma^2/n(1 - \varepsilon)$, the other with variance $\sigma^2/n(1+\varepsilon)$. Doing the change of variable $t = \sqrt{n}(x - x_*)$, we find that the integral

$$\int_{|x-x_*|\leq\delta} e^{ng(x)}\,dx$$

is trapped between

$$\int_{-\sqrt{n}\delta}^{\sqrt{n}\delta} e^{-t^2(1+\varepsilon)^2/2\sigma^2}\,dt/\sqrt{n} \sim \frac{\sqrt{2\pi}}{\sqrt{n}\sigma}(1+\varepsilon)$$

and

$$\int_{-\sqrt{n}\delta}^{\sqrt{n}\delta} e^{-t^2(1-\varepsilon)^2/2\sigma^2}\,dt/\sqrt{n} \sim \frac{\sqrt{2\pi}}{\sqrt{n}\sigma}(1-\varepsilon).$$

□

2. FOURIER SERIES

2.1. **Absolutely convergent Fourier series.** Let $(a_m)_{m\in\mathbb{Z}}$ be an absolutely summable sequence of complex numbers. (Absolutely summable means that $\sum |a_m| < \infty$.) The *Fourier series* of the sequence $(a_m)_{m\in\mathbb{Z}}$ is the series

(8)
$$A(\theta) = \sum_{m=-\infty}^{\infty} a_m e^{im\theta}, \quad \text{where } \theta \in \mathbb{R}.$$

This series converges uniformly for $\theta \in \mathbb{R}$, and the limit $A(\theta)$ is a uniformly continuous, $2\pi-$periodic function of θ. It is natural in many contexts to view $A(\theta)$ as a (continuous) function on the unit circle

(9)
$$\mathbb{T} = \mathbb{T}^1 = \{z \in \mathbb{C} : |z| = 1\},$$

by making the identification $\theta \leftrightarrow e^{i\theta}$. This point of view is useful for various reasons, most important of which are that (a) $\mathbb{T}$ is an abelian *group* under multiplication, and (b) functions on $\mathbb{T}$ extend, in certain circumstances, to analytic functions on domains containing the unit circle.

The usefulness of Fourier series stems from the fact that *the Fourier series of a convolution is the product of the Fourier series*. Here is a more precise statement. For any two absolutely summable sequences $(a_m)_{m\in\mathbb{Z}}$ and $(b_m)_{m\in\mathbb{Z}}$ define their convolution

(10)
$$c_m = (a * b)_m = \sum_{k=-\infty}^{\infty} a_k b_{m-k} = \sum_{k=-\infty}^{\infty} a_{m-k} b_k.$$

Observe (in other words, you supply the proof!) that the sequence $a * b$ is absolutely summable, and that

(11)
$$\sum_{m=-\infty}^{\infty} |c_m| \leq \sum_{m=-\infty}^{\infty} |a_m| \sum_{m=-\infty}^{\infty} |b_m|$$

Proposition 2. *If $C(\theta) = \sum c_m e^{im\theta}$ is the Fourier series of the convolution $c_m = (a * b)_m$ then*

(12)
$$C(\theta) = A(\theta)B(\theta).$$

Proof. Exercise. (Note that changing the order of summation is justified by the dominated convergence theorem and/or Fubini's theorem, because both $(a_m)_{m\in\mathbb{Z}}$ and $(b_m)_{m\in\mathbb{Z}}$ are assumed to be absolutely summable.) □

2.2. L^2 **theory of Fourier series.** Denote by $L^2(\mathbb{T})$ the vector space of all complex-valued, Borel measurable functions on $\mathbb{T}$ that have finite second moment with respect to the uniform distribution (i.e., normalized Lebesgue measure) on $\mathbb{T} = [-\pi, \pi]$. This space is equipped with a natural complex inner product, defined as follows: for any two functions $f, g \in L^2(\mathbb{T})$ set

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)\overline{g}(\theta)\, d\theta, \tag{13}$$

and define the L^2-norm of f to be $\|f\|_2 = \langle f, f\rangle^{1/2}$. The L^2-norm induces a metric[1] on $L^2(\mathbb{T})$, and it is a fact (see, for instance, RUDIN, *Real and Complex Analysis*, chs. 3–4) that this metric space is complete, and that the subspace $C(\mathbb{T})$ consisting of *continuous* functions is dense in $L^2(\mathbb{T})$.

Define the *complex exponential functions* e_k (also known as the *group characters*) on $\mathbb{T}$ to be the functions

$$e_m(\theta) = e^{im\theta}. \tag{14}$$

Theorem 3. *The characters $\{e_m\}_{m\in\mathbb{Z}}$ constitute an orthonormal basis of $L^2(\mathbb{T})$. In particular, (i) they are orthonormal, that is,*

$$\langle e_m, e_k \rangle = \delta(m, k), \tag{15}$$

and (ii) their linear span (the set of all finite linear combinations $\sum_{k=-N}^{N} a_k e_k$) is dense in $L^2(\mathbb{T})$.

Proof. The orthogonality relations (15) are an easy calculus exercise (which you should do). The completeness assertion (that the linear span of the exponentials is dense in L^2) is harder, and uses the nontrivial fact that the space of continuous functions $C(\mathbb{T})$ is dense in L^2. The hard step is then to show that for every continuous function f and every $\varepsilon > 0$ there is a trig polynomial $\psi = \sum_{k=-N}^{N} a_k e_k$ such that $\|f - \psi\|_\infty < \varepsilon$. This is done in section 2.4 below – see Theorem 7.) $\square$

Remark 1. The calculus exercise required to prove the orthogonality relations is really so easy that a second argument is overkill. Nevertheless, here is another way to look at it that emphasizes the role of the group structure more clearly. First, observe that the uniform distribution on the circle is invariant under rotations – that is, if U is a random angle with the uniform distribution and α is any constant then $U + \alpha$ (with addition mod 2π) is still uniformly distributed. Second, the exponential functions e_m are eigenfunctions of the rotation operators, that is, $e_m(\theta + \alpha) = e^{im\alpha} e_m(\theta)$. Now consider the inner product $\langle e_m, e_k\rangle$: this can be represented as

$$\langle e_m, e_k \rangle = E e_m(U)\overline{e_k(U)}$$

where U is uniformly distributed on $\mathbb{T}$. But U has the same distribution as $U + \alpha$, for any α, so

$$\begin{aligned}
\langle e_m, e_k \rangle &= E e_m(U)\overline{e_k(U)} \\
&= E e_m(U+\alpha)\overline{e_k(U+\alpha)} \\
&= e^{im\alpha} e^{-ik\alpha} E e_m(U)\overline{e_k(U)} \\
&= e^{i(m-k)\alpha} \langle e_m, e_k \rangle.
\end{aligned}$$

Consequently, the inner product must be 0 if $m \neq k$.

---

[1] Technically, it's only a *pseudo-metric*, because two functions that differ on a set of measure zero will have L^2-distance 0. Thus, elements of L^2 are really *equivalence classes* of square-integrable functions, where two functions are considered equivalent if they differ only on a set of measure zero.

It follows from Theorem 3 that any function $f \in L^2(\mathbb{T})$ can be arbitrarily well-approximated (in the L^2-distance) by trig polynomials. What is the best way to do this? Given a finite integer $N \geq 1$, which trig polynomial $\psi = \sum_{-N}^{N} a_k e_k$ minimizes $\|f - \psi\|_2$?

Proposition 4. *For any $f \in L^2(\mathbb{T})$ and any $N \geq 1$, the trig polynomial $\psi = \sum_{-N}^{N} a_k e_k$ that minimizes the distance $\|f - \psi\|_2$ is*

$$S_N f := \sum_{k=-N}^{N} \hat{f}_k e_k \quad \textit{where} \tag{16}$$

$$\hat{f}_k := \langle f, e_k \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta) e^{-ik\theta}\, d\theta. \tag{17}$$

Proof. Exercise! (Decompose sums of squares! The argument is essentially the same as the proof of the corresponding statement for orthogonal projections in finite-dimensional linear algebra.) $\square$

The complex numbers $\hat{f}_k$ that occur in the sum (16) are known as the *Fourier coefficients* of the function f. Since by Theorem 3 there are trig polynomials that approximate f arbitrarily closely, it follows that

$$f = L^2 - \lim_{N\to\infty} \sum_{-N}^{N} \hat{f}_k e_k == L^2 - \lim_{N\to\infty} S_N f. \tag{18}$$

This statement is often abbreviated by writing

$$f \overset{L^2}{=} \sum_{-\infty}^{\infty} \hat{f}_k e_k.$$

You should be aware, however that, unlike the series (8) with absolutely summable coefficients the series (18) does not always converge pointwise everywhere. It is a famous (and very hard – I mean $X-$rated) theorem of L. Carleson that for every function $f \in L^2$ the series (18) converges to f at *almost every θ*.

Proposition 5. *(Plancherel Formulas) For any functions $f, g \in L^2(\mathbb{T})$,*

$$\|f\|_2^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(\theta)|^2\, d\theta = \sum_{m=-\infty}^{\infty} |\hat{f}_m|^2 \tag{19}$$

and

$$\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta)\overline{g(\theta)}\, d\theta = \sum_{m=-\infty}^{\infty} \hat{f}_m \overline{\hat{g}_m}. \tag{20}$$

Proof. By (18), both f and g can be arbitrarily well-approximated (in L^2) by finite sections $S_N f$ and $S_N g$ of their Fourier series. But for finite trig series the formula (20) is an easy consequence of the orthogonality relations (15):

$$\begin{aligned}\langle S_N f, S_N g\rangle &= \langle \sum_{k=-N}^{N} \hat{f}_k e_k, \sum_{m=-N}^{N} \hat{g}_m e_m\rangle \\ &= \sum_{k=-N}^{N} \sum_{m=-N}^{N} \hat{f}_k \overline{\hat{g}_m} \langle e_k, e_m\rangle \\ &= \sum_{k=-N}^{N} \hat{f}_k \overline{\hat{g}_k}.\end{aligned}$$

Applying this for $f = g$ shows that

$$\|S_N f\|_2^2 = \sum_{k=-N}^{N} |\hat{f}_k|^2.$$

The relations (19) and (20) in general follow by taking L^2-limits and using the polarization identity for inner products. Here is how it goes: First, by the triangle inequality for the L^2-metric,

$$\|f\|_2^2 - \|S_N f\|_2^2 \leq \|f - S_N f\|_2.$$

Since $S_N f \to f$ in L^2 as $N \to \infty$, the right side of the inequality goes to 0 as $N \to \infty$, and therefore $\|f\|^2$ is the limit of the sequence $\|S_N f\|_2^2$. Thus, by the result of the preceding paragraph,

$$\|f\|^2 = \sum_{-\infty}^{\infty} |\hat{f}_k|^2.$$

The proof can now be finished by the *polarization technique*: what we have just proved implies that

$$\begin{aligned}\|f+g\|^2 &= \sum_{-\infty}^{\infty} |\hat{f}_k + \hat{g}_k|^2, \\ \|f\|^2 &= \sum_{-\infty}^{\infty} |\hat{f}_k|^2, \text{ and} \\ \|g\|^2 &= \sum_{-\infty}^{\infty} |\hat{g}_k|^2.\end{aligned}$$

But (here is the "polarization")

$$\|f+g\|^2 = \|f\|^2 + \|g\|^2 + 2\langle f, g\rangle.$$

Hence

$$\begin{aligned}2\langle f, g\rangle &= \sum_{-\infty}^{\infty} \hat{f}_k + \hat{g}_k|^2 - |\hat{f}_k|^2 - \hat{g}_k|^2 \\ &= \sum_{-\infty}^{\infty} 2\hat{f}_k \overline{\hat{g}_k}.\end{aligned}$$

$\square$

The formulas (19) and (20) can (and should!) be viewed as geometric statements about the space $L^2(\mathbb{T})$: they assert that the mapping Ψ that takes functions $f \in L^2(\mathbb{T})$ to the square-summable sequences $(\hat{f}_k)_{k\in\mathbb{Z}}$ consisting of the Fourier coefficients is an *isometry* of Hilbert spaces. Let $\ell^2(\mathbb{Z})$ be the space of all square-summable sequences $(a_m)_{m\in\mathbb{Z}}$, and define an inner product on this space by setting

$$\langle (a_m)_{m\in\mathbb{Z}}, (b_m)_{m\in\mathbb{Z}} \rangle = \sum_{-\infty}^{\infty} a_m \bar{b}_m. \tag{21}$$

With this inner product $\ell^2(\mathbb{Z})$ becomes a Hilbert space (i.e., an inner product space that is also a complete metric space with respect to the metric induced by the inner product). The Plancherel formulas state that the mapping $\Psi : L^2(\mathbb{T}) \to \ell^2(\mathbb{Z})$ is an isometry, and therefore injective. That it is also *surjective* follows because the image of Ψ includes all sequences that have at many finitely many nonzero entries (these are precisely the images of the trig polynomials), and these are clearly dense in $\ell^2(\mathbb{Z})$. Thus, for all practical purposes, the spaces $\ell^2(\mathbb{Z})$ and $L^2(\mathbb{T})$ are the same.

Every absolutely summable sequence $(a_m)_{m\in\mathbb{Z}}$ is also square-summable. Let $A(\theta)$ be the continuous function gotten by summing the Fourier series (8). Since $A(\theta)$ is continuous, it is square-integrable, and therefore in $L^2(\mathbb{T})$. Its Fourier coefficients are, by definition,

$$\hat{A}_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\theta)e^{-im\theta}\, d\theta.$$

Substituting the absolutely convergent series (8) for A in this integral and interchanging the order of integration and summation yields

$$\hat{A}_m = a_m.$$

(The interchange is permissible because by hypothesis the sequence $(a_m)_{m\in\mathbb{Z}}$ is absolutely summable, and so the dominated convergence theorem applies.) This point might by now seem obvious, but it is important:

Fourier Inversion Formula: *If $A(\theta) = \sum_{m=-\infty}^{\infty} a_m e^{im\theta}$ for some absolutely summable sequence $(a_m)_{m\in\mathbb{Z}}$ then the coefficients a_m can be recovered by*

$$a_m = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\theta)e^{-im\theta}\, d\theta \tag{22}$$

Exercise 1. Let $f(\theta) = \theta$ for $-\pi \le \theta < \pi$. (a) Calculate the Fourier coefficients of f. Note that since f is discontinuous on $\mathbb{T}$ (the discontinuity is at $e^{2\pi i}$) the Fourier coefficients cannot be absolutely summable. (b) Now apply the Plancherel formula to obtain an interesting identity. (c) How may other series can you sum by a similar method?

2.3. **Fourier Series in Dimensions $d \ge 2$.** There is a completely analogous theory for "sequences" defined on higher-dimensional integer lattices. Denote by $\mathbb{Z}^d$ the $d-$dimensional lattice, that is, the set of all $d-$tuples $m = (m_1, m_2, \dots, m_d)$ with integer entries, and by $\mathbb{T}^d$ the $d-$fold Cartesian power of the circle group $\mathbb{T}$. Both $\mathbb{Z}^d$ and $\mathbb{T}^d$ are abelian groups; the group $\mathbb{T}^d$ is the $d-$dimensional torus, and for all practical purposes can be viewed as the $d-$cube $[-\pi, \pi]^d$ with opposite faces glued together. The $d-$dimensional analogue of a sequence is an assignment $(a_m)_{m\in\mathbb{Z}^d}$ of complex numbers to the points of the lattice, i.e., a function $a : \mathbb{Z}^d \to \mathbb{C}$. The spaces $\ell^p(\mathbb{Z}^d)$ consist of the sequences show absolute pth powers are summable. For a sequence $(a_m)_{m\in\mathbb{Z}^d}$ in $\ell^1(\mathbb{Z}^d)$, the Fourier series $A(\theta)$ can be defined as

$$A(\theta) = \sum_{m\in\mathbb{Z}^d} a_m e^{i\langle m,\theta\rangle} = \sum_{m\in\mathbb{Z}^d} a_m e_m(\theta) \tag{23}$$

where $\langle m,\theta\rangle = \sum_{j=1}^d m_j\theta_j$ is just the usual dot product.

Theorem 6. *The complex exponential functions e_m, where $m\in\mathbb{Z}^d$, are an orthonormal basis of $L^2(\mathbb{T}^d)$. Consequently, for any function $f\in L^2(\mathbb{T}^d)$,*

$$f = L^2 - \lim_{n\to\infty} \sum_{m\in\mathbb{Z}^d\,:\,|m|\le n} \hat{f}_m e_m \quad \textit{where} \tag{24}$$

$$\hat{f}_m = \langle f, e_m\rangle = (2\pi)^{-d}\int_{[-\pi,\pi]^d} f(\theta)\,d\theta. \tag{25}$$

Furthermore, the Plancherel identities hold: for any functions $f,g\in L^2(\mathbb{T}^d)$,

$$\langle f,g\rangle = (2\pi)^{-d}\int_{[-\pi,\pi]^d} f(\theta)\overline{g(\theta)}\,d\theta = \sum_{m\in\mathbb{Z}^d}\hat{f}_m\overline{\hat{g}_m}. \tag{26}$$

Finally, if $(a_m)_{m\in\mathbb{Z}^d}$ is absolutely summable with Fourier series $A(\theta)$ then $A(\theta)$ is continuous on $\mathbb{T}^d$ and the Fourier coefficients can be recovered by the Fourier inversion formula

$$a_m = (2\pi)^{-d}\int_{[-\pi,\pi]^d} A(\theta)e^{-i\langle m,\theta\rangle}\,d\theta. \tag{27}$$

The proofs of the various statements are nearly identical to those used to prove the analogous statements in the one-dimensional case. Take note, however, of one important difference: the Fourier integrals in equations (25), (26), and (27) are $d-$fold multiple integrals with respect to (normalized) $d-$dimensional Lebesgue measure

$$d\theta = d\theta_1 d\theta_2\cdots d\theta_d.$$

2.4. **Approximation by Trig Polynomials.**

Theorem 7. *For any continuous function $f:\mathbb{T}\to\mathbb{C}$ and any $\varepsilon>0$ there is a trig polynomial $\psi(\theta) = \sum_{-N}^N a_k e^{ik\theta}$ such that*

$$\|f-\psi\|_\infty < \varepsilon \tag{28}$$

This is the fundamental theorem of Fourier analysis: it guarantees that the complex exponential functions e_m span a dense linear subspace of $L^2(\mathbb{T})$ (and also a dense linear subspace of $L^p(\mathbb{T})$, for any $1\circ < p < \infty$, but we will not need this fact). Theorem 7 is a direct consequence of the Stone-Weierstrass theorem, which gives sufficient conditions for an *algebra* of continuous functions to be uniformly dense in the space of all continuous functions. Unfortunately, Stone-Weierstrass doesn't provide a constructive method for *finding* good trig polynomial approximations (at least, not a *good* constructive method), nor does it point out the important role of notions from probability theory in such constructions. So in this section we will give a direct, elementary approach to Theorem 7.

Let $\hat{f}_m$ be the mth Fourier coefficient of f (as defined by (17)). These are bounded in absolute value (by $\|f\|_\infty$) but they need not be absolutely summable; they do go to zero as $|m|\to\infty$ [2], but

[2] This fact is the *Riemann-Lebesgue Lemma*. It is not needed for the proof of Theorem 7.

they do so rather slowly. For this reason, the sequence of finite sections

$$S_N f(\theta) := \sum_{k=-N}^{N} \hat{f}_k e^{ik\theta} \tag{29}$$

may fail to converge. This rules out the possibility of using the most obvious candidates for trig polynomial approximations. (This might at first be confusing: Proposition 4 asserts that $S_N f$ is the closest element of span $e_{k-N\leq k\leq N}$ in L^2-distance so it might seem that it should also be closest in sup norm distance. Not so!)

To get around this problem we will use a form of what statisticians would call *tapering*. Instead of using the Fourier coefficients straight up in (29), we will dampen the higher frequency terms, to obtain the approximation

$$D_N f(\theta) = \sum_{k=-N}^{N} \left(1 - \frac{|k|}{N}\right) \hat{f}_k e^{ik\theta}. \tag{30}$$

(This should remind you of the *Winsorized mean* in statistics.) The peculiar notation $D_N f$ is used because, as we will see, the function $D_N f$ is gotten by convolving f with the so-called *Dirichlet kernel*.

Lemma 8.

$$D_N f(\theta) = K_N * f(\theta) := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\theta - \alpha) K_N(\alpha)\, d\alpha \tag{31}$$

where

$$\begin{aligned} K_N(\theta) &= \sum_{k=-N}^{N} \left(1 - \frac{|k|}{N}\right) e^{ik\theta} \\ &= \frac{1}{N} \left| \sum_{k=1}^{N} e^{ik\theta} \right|^2 \\ &= \frac{1}{N} \left(\frac{1 - \cos N\theta}{1 - \cos\theta} \right). \end{aligned} \tag{32}$$

Proof. Direct calculation. (Not too hard – I hope I haven't made any errors.) □

Corollary 9. *For each $N \geq 1$ the Dirichlet kernel $K_N(\theta)$ is a probability density on $[-\pi, \pi]$ with respect to the uniform distribution (normalized Lebesgue measure). Moreover, as $N \to \infty$ the probability measures $K_N(\theta)\, d\theta/2\pi$ converge in distribution to the point mass at $\theta = 0$.*

Proof. This follows by using the different forms of the Dirichlet kernel recorded in equations (32). The first form shows that $K_N(\theta)$ integrates to 1 (using the orthonormality of the exponential functions). The second form shows that $K_N \geq 0$ everywhere. Finally, the third form shows that the densities K_N converge pointwise to 0 everywhere *except* at $\theta = 0$, so for large N the density K_N puts almost no mass outside $[-\varepsilon, \varepsilon]$. □

Proof of Theorem 7. Given Lemma 8 and Corollary 9 there isn't much left to prove. Lemma 8 shows that $D_N f$ is the convolution of f with the density K_N. Thus,

$$D_N f(\theta) = E f(\theta + Y_N)$$

where Y_N is a $\mathbb{T}-$valued random variable with density K_N (and addition is done modulo 2π). Corollary 9 implies that as $N \to \infty$ the random variables Y_N converge to 0 in probability. The result (28) now follows from the uniform continuity of f. $\square$

3. RANDOM WALK

3.1. **Recurrence of random walks in dimension $d = 1$.** Let $S_n = \sum_{j=1}^n \xi_j$ be the sum of n independent, identically distributed $\mathbb{Z}-$valued random variables with finite mean μ and finite positive variance σ^2, and let $p_k = P\{\xi_1 = k\}$ be the discrete probability density of the random variables ξ_j. A discrete probability density on $\mathbb{Z}$ is an absolutely summable sequence, and therefore its Fourier series is absolutely convergent. The Fourier series is just the *characteristic function* of ξ_1:

$$\varphi(\theta) := Ee^{i\theta\xi_1} = \sum_{-\infty}^{\infty} p_k e^{ik\theta}. \tag{33}$$

The characteristic function of S_n is $\varphi(\theta)^n$. Thus, by the Fourier inversion formula, the distribution of S_n can (in principle) be recovered from the characteristic function:

$$P\{S_n = k\} = (2\pi)^{-1} \int_{\mathbb{T}} \varphi(\theta)^n e^{-ik\theta}\, d\theta. \tag{34}$$

This integral is similar in form to the integral (5) studied in section 1 above. There are two important differences: first the integrands considered in section 1 were assumed to be real-valued; and second, the complex exponential factor $e^{-ik\theta}$ is not an nth power. Thus, Proposition 1 does not apply directly, at least not in all cases. Nevertheless, the *Laplace method* used to obtain Proposition 1 does apply, with some minor modifications.

Before we turn to the general case let's see what we can deduce from Proposition 1 directly. Suppose that the distribution $(p_k)_{k\in\mathbb{Z}}$ is *symmetric* about 0, that is, $p_k = p_{-k}$. Then the characteristic function $\varphi(\theta)$ is real and even (that is, $\varphi(\theta) = \varphi(-\theta)$), the mean $\mu = 0$, and so the first two terms of the Taylor series of $\log\varphi$ around zero vanish. Thus, at least in a neighborhood of the origin, the Fourier integral (34) has the form (5) when $k = 0$. Unfortunately, $\varphi(\theta)$ might not attain its maximum *uniquely* at $\theta = 0$, and in fact, if the probability distribution $(p_k)_{k\in\mathbb{Z}}$ is supported by (say) the even integers $2\mathbb{Z}$, then $\varphi(\theta)$ will take the value 1 not only at 0 but at $\theta = \pm\pi$. The following lemma characterizes those situations in which the characteristic function will have multiple maxima in $[-\pi, \pi]$.

Lemma 10. *If $(p_k)_{k\in\mathbb{Z}}$ is* not *supported by any proper subgroup $m\mathbb{Z}$ of $\mathbb{Z}$ (some $m \geq 2$) then $\varphi(\theta) = 1$ only at $\theta = 0$ in the interval $[-\pi, \pi]$. If $(p_k)_{k\in\mathbb{Z}}$ is not supported by any* coset *of a proper subgroup (that is, by a proper arithmetic progression) then $|\varphi(\theta)| < 1$ for all $\theta \neq 0$ in $[-\pi, \pi]$.*

Proof. Exercise. $\square$

Example 1. The Rademacher distribution (that is, the distribution that assigns probability 1/2 to ± 1) is not supported by a proper subgroup of $\mathbb{Z}$, but it is supported by the arithmetic progression $2\mathbb{Z} + 1$. It characteristic function

$$\varphi(\theta) = \cos\theta \tag{35}$$

assumes the value $+1$ only at $\theta = 0$, but it assumes the value -1 at $\theta = \pm\pi$. Note that in general if ξ_1 takes its values in the arithmetic progression $m\mathbb{Z} + k$, for some k relatively prime to m, then S_{nm} takes values in $m\mathbb{Z}$, and S_n can only return to 0 at integer multiples of m.

Proposition 11. *Assume that the distribution $(p_k)_{k\in\mathbb{Z}}$ of ξ_1 is symmetric about 0 and has variance $0 < \sigma^2 < \infty$. Then the random walk S_n is recurrent.*

Proof. The idea is to use Laplace's method and the Fourier inversion formula (34) to show that $P\{S_n = 0\} \geq Cn^{-1/2}$ for some $C > 0$, at least for even n. Polya's criterion for recurrence will then imply that the random walk returns to 0 infinitely often with probability 1.

Consider first the case where the distribution is not supported by any proper arithmetic progression. Then $|\varphi(\theta)| < 1$ for all $\theta \in [-\pi, \pi] - \{0\}$. Fix $\delta > 0$ and let $\varepsilon = \max |\varphi(\theta)|$ where the max is over all $\theta \in [-\pi, \pi]$ such that $|\theta| \geq \delta$. Then $\varepsilon < 1$, and so

$$\int_{[-\pi,\pi]-[-\delta,\delta]} |\varphi(\theta)|^n \, d\theta \leq \varepsilon^n.$$

This goes to 0 faster than any polynomial in n, so it will be enough to show that

$$\int_{[-\delta,\delta]} \varphi(\theta)^{2n} \, d\theta \geq C/\sqrt{2n}.$$

But if $\delta > 0$ is sufficiently small then $\varphi(\theta) > 0$ for all $\theta \in [-\delta, \delta]$, and so in this case the integral is of the form (5). Consequently, Proposition 1 implies that the integral is $\sim C/\sqrt{2n}$ with $C = 1/\sqrt{2\pi n}\sigma$ (I think!).

The general case can be handled the same way once the periodicities are accounted for. Suppose then that the distribution of ξ_1 is supported by $m\mathbb{Z}+k$ for some $m \geq 2$, and by no coarser arithmetic progression. Then S_m has distribution supported by $m\mathbb{Z}$, but by no proper arithmetic progression contained in $m\mathbb{Z}$; consequently S_m/m is integer-valued, and its distribution is contained in no arithmetic progression. Thus, the argument of the preceding paragraph implies that the random walk S_{mn}/m is recurrent. But this obviously implies that S_n is recurrent. $\square$

The hypotheses of Proposition 11 are much more stringent than necessary. In fact the following is true: any one-dimensional random walk S_n on the integers whose increments ξ_j have finite absolute first moment and mean $E\xi_1 = 0$ is recurrent. You can read the proof in the Stat 312 notes: it is based on the *Kesten-Spitzer-Whitman* theorem, which in turn is a consequence of the strong law of large numbers.

3.2. **Recurrence and transience in dimensions $d \geq 2$.** What about random walk on the higher-dimensional integer lattices $\mathbb{Z}^d$? To keep the discussion as elementary as possible let's consider only the *simple random walk* on $\mathbb{Z}^d$, which moves by choosing at each step one of the $2d$ nearest neighbors of the current state at random (each with probability $1/2d$) for its next state. If the random walk is started at the origin $S_0 = 0$ then the position at time n is $S_n = \sum_{j=1}^n \xi_j$ where the random vectors ξ_j are i.i.d. with characteristic function

$$\varphi(\theta) = Ee^{i\langle\theta,\xi_1\rangle} = d^{-1}\sum_{j=1}^{d} \cos\theta_j. \tag{36}$$

This is real-valued, but because the simple random walk has period 2 the characteristic function takes the value -1 at points θ whose entries θ_j are $\pm\pi$. Hence, it is easier to work with the random walk S_{2n}, whose step distribution has characteristic function $Ee^{i\langle S_2,\theta\rangle} = \varphi(\theta)^2$. This has the advantage that it is $\pi-$periodic, with absolute value strictly less than 1 for all nonzero $\theta \in [-\pi/2, \pi/2]^d$.

Fourier inversion gives

$$P\{S_{2n} = 0\} = (\pi)^{-d} \int_{[-\pi/2,\pi/2]^d} \varphi(\theta/2)^{2n}\, d\theta. \tag{37}$$

(You should check to see if I got the factors of 2 correct.) This is of the form (5), but the integral is a $d-$fold multiple integral, so Proposition 1 doesn't apply except in $d = 1$. What is needed is a $d-$dimensional version of Laplace's method.

Suppose, then, that $g : [-A, A]^d \to \mathbb{R}$ is a C^2-function such that

$$g(0) = 0 \quad \text{and} \quad g(x) < 0 \ \text{ for all } x \neq 0,$$

and assume that the matrix H of negative second partials $-\partial^2 g/\partial x_j \partial x_k$ (at the origin) is strictly positive definite. Set

$$J(n) = \int_{[-A,A]^d} e^{ng(x)}\, dx.$$

Proposition 12. *Under the preceding hypotheses,*

$$J(n) \sim (2\pi)^{d/2} n^{-d/2} \det H \quad \textit{as} \ \ n \to \infty. \tag{38}$$

Proof. This is almost exactly the same as the proof of Proposition 1. The hypotheses on g imply that for large n, in a neighborhood of the origin the function $e^{ng(x)}$ is trapped between two Gaussian densities whose covariance matrices are $(1 \pm \varepsilon)H^{-1}/n$ for some $\varepsilon > 0$ which may be taken arbitrarily small by choosing a small neighborhood of the origin. Integrating these Gaussian densities and letting $\varepsilon \to 0$ yields (38). □

Theorem 13. *(Polya) Simple random walk is recurrent in dimensions $d = 1, 2$ and transient in dimensions $d \geq 3$.*

Proof. Proposition 12 and the Fourier inversion formula (37) imply that

$$P\{S_{2n} = 0\} \sim C_d n^{-d/2}$$

for certain positive constants C_d that I am too lazy to work out. The series $\sum n^{-d/2}$ is summable in dimensions $d \geq 3$, but diverges for $d = 1, 2$. □

3.3. **Local limit theorem.**