Honors 1: Undergraduate Math Lab*

Peter Sarnak†, Steven Miller‡, Alex Barnett§

October 2^{nd}, 2002

Courant Institute of Mathematical Sciences
New York University
New York, NY

Abstract

Proof of the Central Limit Theorem (via the Fourier Transform). Proof of the Spectral Theorem for Real Symmetric Matrices (via maximization). Spectral Gap and Families of Expanders. Lecture by Peter Sarnak; notes by Steven Miller.

1 Central Limit Theorem

Let $X : \Omega \to \mathbb{R}$ be a random variable ($x(\omega) \in \mathbb{R}$). Define the density function μ on $\mathbb{R}$ by

$$\mu[a, b] = \text{Prob}\Big(\omega : x(\omega) \in [a, b]\Big). \tag{1}$$

Thus, X gives rise to the probability measure μ on $\mathbb{R}$.
Assume

1. $X_1, X_2, \dots$ are independent identically distributed random variables (iidrv) with density μ.

*Homepage: `http://www.math.nyu.edu/~millerj/`
†E-mail: `sarnak@math.princeton.edu`
‡E-mail: `millerj@cims.nyu.edu` or `sjmiller@math.princeton.edu`
§E-mail: `barnett@nmr.mgh.harvard.edu`

2. $E(X) = \int_\Omega x(\omega)dp(\omega) = \int_{\mathbb{R}} x d\mu(x)$.

3. $\mathrm{Var}(X) = \int_\Omega x^2(\omega)dp(\omega) = \int_R x^2 d\mu(x) = 1$.

Define $S_N = \sum X_j$, and let $N(0,1)$ denote the standard Gaussian or normal distribution (mean zero, variance one).

Theorem 1.1.

$$\frac{S_N}{\sqrt{N}} \underset{in\ probability}{\longrightarrow} N(0,1), \tag{2}$$

ie,

$$Prob\Big(\frac{S_N}{\sqrt{N}} \in [a,b]\Big) \to \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{x^2}{2}} dx. \tag{3}$$

Proof: We have a random variable X, induces a probability measure μ on the real line $\mathbb{R}$. Thus, $d\mu = f(x)dx$, $f(x)$ is a nice function. As μ is a probability measure on $\mathbb{R}$, we must have $f(x) \geq 0$ and $\int_R f(x)dx = 1$.

The given assumptions about the X_is imply

1. $\int_R x f(x) dx = 0$.

2. $\int_R x^2 f(x) dx = 1$.

Definition 1.2 (Fourier Transform).

$$\hat{f}(\xi) \ = \ \int_R f(x) e^{-2\pi i x\xi} dx. \tag{4}$$

Clearly, $|\hat{f}(\xi) \leq \int_R f(x)dx \leq 1$. Further, $\hat{f}(0) = \int_R f(x)dx = 1$.

Now

$$\hat{f}'(\xi) = \int_R (2\pi i x) f(x) e^{-2\pi i x\xi} dx. \tag{5}$$

Thus, $\hat{f}'(0) = 0$ (from $E(x) = 0$).

We will assume $\hat{f}$ is continuous (although this is implied by our assumptions).

Further,

$$\hat{f}''(\xi) = -4\pi^2 \int_R x^2 f(x) e^{-2\pi i x\xi} dx. \tag{6}$$

Therefore, $\hat{f}''(0) = -4\pi^2$ (by our assumption on the variance).
Using Taylor to expand $\hat{f}$ we obtain

$$\begin{aligned} \hat{f}(\xi) &= 1 + \frac{f''(0)}{2}\xi^2 + \cdots \\ &= 1 - 2\pi^2\xi^2 + \cdots \end{aligned} \tag{7}$$

Near the origin, the above shows $\hat{f}$ looks like a concave down parabola.

Theorem 1.3 (Fourier Inversion). *If f is a nice function (so all the quantities below make sense, ie, f decays fast enough):*

$$f(x) = \int_R \hat{f}(\xi) e^{2\pi i x\xi} d\xi. \tag{8}$$

Definition 1.4. $e(y) = e^{2\pi i y}$.

Exercise 1.5. *If $\phi_\xi(x+y) = e\Big((x+y)\xi\Big)$, prove $\phi_\xi(x+y) = \phi_\xi(x)\phi_\xi(y)$. ϕ_ξ is a* ***character*** *of $(\mathbb{R}, +)$.*

Exercise 1.6. *Is there a $\psi : \mathbb{R} \to \mathbb{R}$ such that $\psi(x+y) = \psi(x) + \psi(y)$? IE, can you find a homomorphism that takes addition to addition? If yes (of course: see above!) what can you say about ψ? If we assume ψ is continuous, it must be of the form ϕ_ξ.*

Think of $\mathbb{R}$ as a vector space over $\mathbb{Q}$ (ie, the scalars are $\mathbb{Q}$). What is the dimension of $\mathbb{R}$ over $\mathbb{Q}$, and what is a basis? As the reals are uncountable and the rationals are countable, there are uncountably many basis vectors.

Any linear transformation will satisfy the desired condition! For example, choose any basis (called a ***Hamel Basis****) of $\mathbb{R}$ over $\mathbb{Q}$ (very hard! need the Axiom of Choice).*

To show every character of the reals is of the form ϕ_ξ, you need more. If you assume the character is continous, then it must be of the form ϕ_ξ.

Suppose we have random variables X and Y with measures μ and ν, with induced functions $f(x)$ and $g(x)$. If we choose X and Y independently, what is the distribution of $X + Y$?

Lemma 1.7. *The distribution of $X + Y$ clearly cannot be the sum (as that won't be a probability measure). It is $f * g$, and $f * g$ is a probability measure.*

Definition 1.8 (Convolution).

$$(f * g)(x) = \int_R f(x-y)g(y)dy \tag{9}$$

We show the convolution is a probability measure. We assume our functions f and g are nice (ie, that we may use Fubini to interchange the order of integration).

$$\begin{aligned}
\int_R (f * g)(x)dx &= \int_R \int_R f(x-y)g(y)dydx \\
&= \int_R \int_R f(x-y)g(y)dxdy \\
&= \int_R g(y)\left(\int_R f(x-y)dx\right)dy \\
&= \int_R g(y)\left(\int_R f(t)dt\right)dy \\
&= \int_R g(y)\cdot 1dy = 1. \qquad (10)
\end{aligned}$$

Exercise 1.9. $\widehat{f * g}(\xi) = \hat{f}(\xi)\cdot\hat{g}(\xi)$. *Thus, Fourier Transform converts convolution to multiplication.*

Let $Z = X + Y$. What is

$$\text{Prob}\Big(Z \in [z, z+dz]\Big) = h(z)dz. \tag{11}$$

x can be anywhere; given x, y must lie between $z-x$ and $z-x+dz$.

$$\begin{aligned}
\text{Prob}\Big(X+Y \in [z,z+dz]\Big) &= \int_{x\in\mathbb{R}} \text{Prob}\Big(X \in [x,x+dx] \text{ and} \\
&\qquad Y \in [z-x, z-x+dz]\Big) f(x)dx \\
&= \int_{x\in\mathbb{R}} f(x)g(z-x)dx = h(z), \qquad (12)
\end{aligned}$$

where the last step follows from the definition of the density functions.

By Induction, we see $X_1 + \cdots X_N$ has distribution $f * \cdots * f$ (as the random variables are iidrv).

However, we want to study $S_N = \frac{X_1 + \cdots X_N}{\sqrt{N}}$.

Definition 1.10 (FT). *Let $FT(f) = \hat{f}$; FT denotes the Fourier Transform.*

Lemma 1.11.

$$FT\left(\textit{the distribution of } \frac{X_1 + \cdots + X_N}{\sqrt{N}}\right) = \left[\hat{f}\left(\frac{\xi}{\sqrt{N}}\right)\right]^N. \tag{13}$$

We take the limit as $N \to \infty$ for **fixed** ξ. Recall we showed that $\hat{f}(\xi) = 1 - 2\pi^2\xi^2 + \cdots$. Thus, we have to study

$$\left[1 - \frac{2\pi^2\xi^2}{N} + O\left(\frac{|\xi|^3}{N^{\frac{3}{2}}}\right)\right]^N. \tag{14}$$

But as $N \to \infty$, the above goes to

$$e^{-2\pi\xi^2}. \tag{15}$$

The universality arises because *only* terms up to the second order in the Taylor Series contribute.

Exercise 1.12. *Show the Fourier Transform of the Gaussian is the Gaussian.*

Key point:

- Used Fourier Analysis to study the sum of independent identically distributed random variables, as it converts convolution to multiplication.

Exercise 1.13. *Fix g a nice, smooth, rapidly decreasing function. Consider the linear transformation A:*

$$(Af)(x) = \int_R g(x-y)f(y)dy. \tag{16}$$

Any convolution operator is diagonalized by these characters (exponentials). At a formal level,

$$
\begin{aligned}
(A\phi_\xi)(x) &= \int_R g(x-y)\phi_\xi(y)dy \\
&= \int_R g(x-y)e^{-2\pi i\xi y}dy \\
&= \int_R g(t)e^{-2\pi i\xi(x-t)}dt \\
&= e^{-2\pi ix\xi}\int_R g(t)e^{2\pi i\xi t}dt = \hat{g}(\xi)\phi_\xi(x).
\end{aligned} \tag{17}
$$

Thus, ϕ_ξ is an eigenvector of A with eigenvalue $\hat{g}(\xi)$.

2 Spectral Theorem for Real Symmetric Matrices

Let A be a real symmetric matrix acting on $\mathbb{R}^n$. Then A has an orthonormal basis $v_1, \ldots, v_n$ such that $Av_j = \lambda_j v_j$.

A simpler proof, assuming all eigenvalues are distinct, is available in the September 25^{th} lecture notes.

Write the inner or dot product $\langle v, w\rangle = v^t w$. As A is symmetric, $\langle Av, w\rangle = \langle v, Aw\rangle$.

Definition 2.1. $V^\perp = \{w : \forall v \in V, \langle w, v\rangle = 0\}$.

Lemma 2.2. *Suppose $V \subset \mathbb{R}^n$ is an* ***invariant vector subspace*** *under A (if $v \in V$, then $Av \in V$). Then $V^\perp$ is also A-invariant: $A(V^\perp) \subset V^\perp$.*

This proves the spectral theorem. Suppose we find a $v_0 \neq 0$ such that $Av_0 = \lambda_0 v_0$. Take $V = \{\mu v_0 : \mu \in \mathbb{R}\}$ for the invariant subspace.

By Lemma 2.2, $V^\perp$ is left invariant under A, and is one dimension less. Thus, by whatever method we used to find an eigenvector, we apply the same method on $V^\perp$.

Thus, all we must show is given an A-invariant subspace, there is an eigenvector.

Consider

$$
\max_{v \ with \ \langle v,v\rangle = 1} \Big\{\langle Av, v\rangle\Big\}. \tag{18}
$$

Standard fact: every continuous function on a compact set attains its maximum (not necessarily uniquely). See, for example, W. Rudin, *Principles of Mathematical Analysis.*

Let v_0 be a vector giving the maximum value, and denote this maximum value by λ_0. As $\langle v_0, v_0 \rangle = 1$, v_0 is not the zero vector.

Lemma 2.3. $Av_0 = \lambda_0 v_0$.

Clearly, if Av_0 is a multiple of v_0 it has to be λ_0 (from the definition of v_0 and λ_0).

Thus, it is sufficient to show

Lemma 2.4. $\{\mu v_0 : \mu \in \mathbb{R}\}$ *is an A-invariant subspace.*

Proof: let w be an arbitrary vector perpendicular to v_0, and ϵ be an arbitrary small real number. Consider

$$\langle A(v_0 + \epsilon w), v_0 + \epsilon w \rangle \tag{19}$$

We need to renormalize, as $v_0 + \epsilon w$ is not unit length; it has length $1 + \epsilon^2 \langle w, w \rangle$. As v_0 was chosen to maximize $\langle Av, v \rangle$ subject to $\langle v, v \rangle = 1$, after normalizing the above cannot be larger. Thus,

$$\langle A(v_0 + \epsilon w), v_0 + \epsilon w \rangle \;\; = \;\; \langle Av_0, v_0 \rangle + 2\epsilon \langle Av_0, w \rangle + \epsilon^2 \langle w, w \rangle. \tag{20}$$

Normalizing the vector $v_0 + \epsilon w$ by its length, we see that in Equation 20, the order ϵ terms must be zero. Thus,

$$\langle Av_0, w \rangle = 0; \tag{21}$$

however, this implies Av_0 is in the space spanned by v_0 (as w was an arbitrary vector perpendicular to v_0), completing our proof. □

Corollary 2.5. *Any local maximum will lead to an eigenvalue-eigenvector pair.*

The second largest eigenvector (denoted λ_1) is

$$\lambda_1 \;\; = \;\; \max_{\langle v, v_0 \rangle = 0} \frac{\langle Av, v \rangle}{\langle v, v \rangle}. \tag{22}$$

We can either divide by $\langle v, v \rangle$, or restrict to unit length vectors.

3 Applications to Graph Theory

Let G be a k-regular graph, $f : V \to \mathbb{R}$. Recall $v \sim w$ if there is an edge connecting v and w. Let A be the adjacency matrix of G, and define

$$\begin{aligned} Af(v) &= \sum_{v\sim w} f(w) \\ \langle f, g\rangle &= \sum_{v\in V} f(v)\overline{g}(v). \end{aligned} \tag{23}$$

Consider the function $f_0(v) = 1$ for all $v \in V$. Then

$$Af_0(v) = \sum_{v\sim w} f_0(w) = kf_0(v). \tag{24}$$

Thus, f_0 is an eigenfunction with eigenvalue k.

Theorem 3.1 (Expander Families). *Fix k. Suppose we have a sequence of k-regular graphs G_j with $|V_j| \to \infty$ and suppose $k - \lambda_1(G) \geq \delta > 0$, δ fixed. Then G_j is an expander family.*

Remark 3.2. *If you have an algorithm that has a random element, then one can show there is another algorithm which does what this algorithm does without having a random component. (More or less, some slight of hand).*

Suppose we have a **bipartite graph**: there are two sets of vertices I (inputs) and O (outputs). Edges run only between I and O; there are no edges between two vertices in I or between two vertices in O. Let there be n inputs and n outputs, and join each input with k outputs.

Fix $\delta_0 > 0$. We want, for any $B \subset I$, $|\partial B| \geq \delta_0 |B|$.

We give a sketch of the proof. We will show that knowledge of a spectral gap ensures that the boundary of *any* subset of I will be big. We do bipartite for simplicity.

Let $B \subset I$. Define

$$f(v) = \begin{cases} 2n - |B| & \text{for } v \in B \\ -|B| & \text{otherwise.} \end{cases} \tag{25}$$

Then

$$\sum_{v \in V} f(v) = |B| \cdot (2n - |B|) + (2n - |B|) \cdot (-|B|) = 0. \tag{26}$$

Then

$$\lambda_1 = \max_{\langle \tilde{f}, f_0 \rangle = 0} = \frac{\langle A\tilde{f}, \tilde{f} \rangle}{\langle \tilde{f}, \tilde{f} \rangle}. \tag{27}$$

In particular,

$$\frac{\langle Af_1, f_1 \rangle}{\langle f_1, f_1 \rangle} \leq \lambda_1. \tag{28}$$

Definition 3.3 (Laplacian). $\Delta = kI - A$.

Thus, $\Delta f_0 = kf_0 - Af_0 = 0 \cdot f_0$.
The eigenvalues of Δ are trivially related to the eigenvalues of A.

Remark 3.4 (Motivation for Laplacian). *On the line we have* $\frac{d^2}{dx^2}$*. A discrete version is (exercise)*

$$\frac{f(x+h) - 2f(x) + f(x-h)}{h^2}. \tag{29}$$

In the plane, we would have $\frac{d^2}{dx_1^2} + \frac{d^2}{dx_2^2}$*. Integrating by parts we have*

$$\begin{aligned} \int_\Omega (\Delta f) \cdot g dx_1 dx_2 &= -\int_\Omega \nabla f \cdot \nabla g dx_1 dx_2 \\ &= \int_\Omega f \cdot (\Delta g) dx_1 dx_2. \end{aligned} \tag{30}$$

We want to integrate by parts on a graph!

$$(\Delta F)(x) = kF(x) - \sum_{x \sim y} F(y). \tag{31}$$

Therefore

$$
\begin{aligned}
\langle \Delta F, F\rangle &= \sum_{x\in V}\left(kF(x) - \sum_{y\sim x} F(y)\right)F(x) \\
&= k\sum_{x\in V} F(x)^2 - \sum_{x\in V}\sum_{x\sim y} F(x)F(y). \qquad (32)
\end{aligned}
$$

For each edge e, orient it by e^+ and e^-. The analogue of the Laplacian becomes

$$
\begin{aligned}
\sum_e \left(F(e^+) - F(e^-)\right)^2 &= \sum_e F^2(e^+) - 2F(e^+)F(e^-) + F^2(e^-) \\
&= 2\langle \Delta F, F\rangle, \qquad (33)
\end{aligned}
$$

where the last line follows by thinking about what it means for vertices to be connected, and how often each vertex is hit.

Recall

$$
k - \lambda_1 = \min_{\langle F, f_0\rangle = 0} \frac{\langle \Delta F, F\rangle}{\langle F, F\rangle} = \min_{\langle F, f_0\rangle = 0} \frac{1}{2}\frac{||dF||^2}{\langle F, F\rangle}. \qquad (34)
$$

Plug in the function f defined above, namely,

$$
f(v) = \begin{cases} 2n - |B| & \text{for } v \in B \\ -|B| & \text{otherwise.} \end{cases} \qquad (35)
$$

If the edge runs from input to input or output to output, we get zero. The only way we get non-zero contribution is from an input to an output (or vice-versa). For our f,

$$
\frac{1}{2}\sum_{e\in E}\left|f(e^+) - f(e^-)\right|^2 = \frac{1}{2}(2n)^2 \cdot \#\{\text{edges e running from } B \text{ to } B^c\}, \quad (36)
$$

where B^c is the complement of B. We want $0 < \delta_0 = k - \lambda_1$. Divide the previous equation by $\langle F, F\rangle$, where $\langle F, F\rangle = (2n - |B|)^2 \cdot |B| + |B|^2 \cdot (2n - |B|)$ $= |B| \cdot (2n - |B|) \cdot 2n$.

Thus,

$$\delta_0 \leq \frac{\frac{1}{2}(2n)\#\{edges\}}{|B| \cdot (2n - |B|) \cdot 2n}. \tag{37}$$

Therefore,

$$\begin{aligned} \#\{edges\} &\geq \frac{\delta_0 |B| \cdot (2n - |B|)}{n} \\ &\geq \delta_0 |B|, \end{aligned} \tag{38}$$

as $|B| \leq n$.

Thus, the total number of edges is at least $\delta_0|B|$. But each vertex gets k edges. Thus,

$$|\partial B| \geq \frac{\delta_0 |B|}{k}. \tag{39}$$

4 $2\sqrt{k-1}$

Let G be a k-regular connected graph. A is the adjacency matrix. Biggest eigenvalue is $\lambda_0 = k$. Eigenvalues cannot be smaller than $-k$. How big is the gap between k and λ_1?

Theorem 4.1 (Alon-Boppana). *Fix k. Take any sequencce of graphs where the number of vertices $|G| \to \infty$. Then*

$$\overline{lim}_{|G|\to\infty} \lambda_1(A_G) \geq 2\sqrt{k-1}. \tag{40}$$

Remark 4.2. *In 1-dimension, with probability one the drunk returns home; same in 2-dimensions. He escapes with finite probability in 3 and higher dimensions!*

Consider a 3 regular graph with many vertices. Go to v, look locally. If there are no short circuits, know what it looks like locally: it will look like a tree. (This is p-adic hyperbolic geometry).

Let T be the infinite tree where each vertex is connected to three other vertices (and a vertex cannot be connected to itself). Suppose a drunk is walking on a tree. The only way he can get back is to exactly undo what he's done.

Consider the following operator: consider an infinite dimensional Hilbert space $l_2(V)$, the set of all $f : V \to \mathbb{R}$ such that $\sum_v |f(v)|^2 < \infty$. This space is infinite dimensional (for each v, take the function $f_v(w) = 1$ if $w = v$ and 0 otherwise).

Using $|ab| \leq \frac{a^2+b^2}{2}$,

$$\langle f, g \rangle = \sum_v f(v)g(v) \tag{41}$$

exists (and is our inner product). We define

$$Af(v) = \sum_{w \sim v} f(w). \tag{42}$$

It is not obvious that there are any eigenvectors (and, in fact, there are no eigenvectors!). There is still a notion of spectrum. We will show the spectrum of this operator ($k = 3$) is $[-2\sqrt{2}, 2\sqrt{2}]$.

Note the constant function is horrendously not in this space (not even *close* to being square-integrable.