Math 182: Hidden Data in Random Matrices

Todd Kemp, APM 5202

www.math.ucsd.edu/~tkemp/182

Today: Histograms and Linear Statistics

Next: The Trace and the Large-N Limit

HW 2: Δ Regrade Window Wed Feb 12 8am → Fri Feb 14 8pm

Lab 2: Thursday 10-12, Friday 10-11 (this week)

Office Hours: Thursday 10-12, Friday 10-11 (this week)
Noise without Signal: Random Matrix Theory

To understand how to detect a signal amidst noise, we first need to understand what “noise” looks like.

\[ X = \text{random m x N matrix} \]
\[ \text{4 entries all i.i.d. } N(0,1) \]
\[ \hat{X} = \frac{1}{n} (X - \bar{X_n}I^T) \]
\[ \bar{X_n} = \frac{1}{n} (x_1 + x_2 + \cdots + x_n) \]
\[ \rightarrow \quad \text{we'll take } N \text{ large, so we can safely assume } \bar{X_n} \approx 0. \]
\[ \text{We'll typically assume } m \gg N \]
\[ \text{so work with this one.} \]
\[ C = \hat{X} \hat{X}^T \quad \text{or} \quad \hat{C} = \hat{X}^T \hat{X} \]
\[ \text{m x m} \quad \text{N x N} \]
\[ \text{whichever is smaller; same non-zero eigen values} \]
\[ \hat{C} = \frac{1}{n} X^T X \] where \( X \) is \( m \times N \) with i.i.d. \( \mathcal{N}(0,1) \) entries.

\textbf{Wishart Ensemble} \quad \text{Wish}(m, N)

Entries are random \Rightarrow Eigenvectors are random!

Highly non-linear functions of the entries
\Rightarrow not independent
\Rightarrow difficult to compute analytically

We want to understand what the (random) histogram of the eigenvalues looks like, for large \( N, m \).
20 bins
Same data \( \chi = \{ \lambda_1, \ldots, \lambda_n \} \) \( N = 1000 \)
eigenvalues of \( \text{Wish}(4000, 1000) \)

50 bins

How can we say anything exact about the histogram of (random) data that we cannot compute???
Method of Indicators

Really just need to compute all the (random) numbers

\[ h_{[a,b]}(x) = \frac{1}{N} \# \{ j : x_j \in [a,b] \} \]

Usually we introduce indicators to make it easy to compute expectations. Here?
Linear Statistics

Given a point set \( \mathbf{x} = \{ x_1, \ldots, x_n \} \) in \( \mathbb{R} \), a linear statistic is any quantity of the form

\[
L(\mathbf{x}) = \sum a_i x_i
\]

- Sample mean: \( \bar{x} \)
- Sample variance: \( S_N(x) \)
- Sample moments: \( M_k(x) \)
- Histograms: \( h_{[a,b]}(x) \)
Approximating with Polynomials

It will turn out to be possible to really compute linear statistics of eigenvalues, with a polynomial weight function $f$. We'd really like to do it with $f = 1_{[a,b]}$. Fortunately, we can approximate $1_{[a,b]}$ with polynomials.

**Theorem (Bernstein, 1920s)**

Let $[\alpha, \beta]$ be a closed interval, and let $f : [\alpha, \beta] \to \mathbb{R}$. There is a sequence $B_n$ of polynomials ($\deg P_n = n$) such that

$$
\lim_{n \to \infty} B_n(t) = f(t) \quad \text{for all } t \in [\alpha, \beta] 
$$

such that $f$ is continuous at $t$. 
Proof of Bernstein’s Theorem

Step 1: reduce to \([a, \beta] = [0, 1]\).

Step 2: the “Probabilistic Method”