LECTURE 41

Concentration of Measure

- All concentration inequalities we studied so far
  (Krein-Feller, Chernoff, Bernstein, matrix Bernstein)
  are for linear functions of r.v’s (their sums).
- Nonlinear: i.e. \[ |f(x) - f(y)| \leq L \|x - y\|_2 \text{ for all } x, y \in \mathbb{R}^d \]

**Theorem (Gaussian concentration)**

Let \( f: \mathbb{R}^d \to \mathbb{R} \) be an \( L \)-Lipschitz function, and \( X \sim N(0, \text{Id}) \). Then

\[
P\left\{ |f(X) - Ef(X)| \geq t \right\} \leq 2 \exp\left( -\frac{t^2}{2L^2} \right) \quad \forall t > 0
\]

The proof is based on an isoperimetric inequality:

"Among all subsets of \( \mathbb{R}^d \) with given volume, the Euclidean balls have minimal surface area."

More generally, consider the \( \varepsilon \)-neighborhood of \( A \subset \mathbb{R}^d \):

\[
A_\varepsilon := \{ x \in \mathbb{R}^d : \exists y \in A \text{ s.t. } \|x - y\|_2 \leq \varepsilon \}
\]

Surface area \( (A) = \lim_{\varepsilon \to 0} \frac{\text{Vol}(A_\varepsilon) - \text{Vol}(A)}{\varepsilon} \)

Isoperimetric inequality: Hence, among all subsets \( A \subset \mathbb{R}^d \) with given volume, the Euclidean balls minimize the volume of \( A_\varepsilon \).
Gaussian isoperimetric ineq \( \gamma(A) = \frac{1}{(2\pi)^{d/2}} \int_A e^{-\|x\|^2/2} \, dx \)

**Thm (Gauss Ineq)**

Let \( \epsilon > 0 \). Among all sets \( A \subset \mathbb{R}^d \) with given Gauss measure, the half-spaces minimize the Gauss measure of \( A \).

Hence: \( \gamma(A) \geq \gamma(\mathbb{H}) \) implies \( \gamma(A_\epsilon) \geq \gamma(\mathbb{H}_\epsilon) \) \( \forall \) half-space \( \mathbb{H} \).

\(*\)

Proof of Gauss concentration Thm (p1): WLOG \( \ell = 1 \).

Let \( M = \text{median of } f(x) \), i.e.

\[
\Pr\{f(x) \leq M\} \geq 1/2, \quad \Pr\{f(x) \geq M\} \geq 1/2.
\]

Let \( A := \{x \in \mathbb{R}^d : f(x) \leq M\} \), \( H := \{x \in \mathbb{R}^d : x_1 \leq 0\} \).

\[
\gamma(A) \geq \frac{1}{2} = \gamma(H) \quad \xrightarrow{(*)} \quad \gamma(A_t) \geq \gamma(H_t) = \Pr\{x_1 \leq t\} = 1 - \Pr\{x_1 > t\} = 1 - \exp(-t^2/2)
\]

Claim: \( A_t \subseteq \{x : f(x) \leq M + t\} \)

Indeed, if \( x \in A_t \), then \( \exists y \in A : \|x - y\| \leq t \)

\[
|f(x) - f(y)| \leq t \quad \Rightarrow \quad f(x) \leq f(y) + t \leq M + t
\]

\[
\Rightarrow \Pr\{f(x) \leq M + t\} \geq \Pr\{x \in A_t\} = \gamma(A_t) \geq 1 - \exp(-t^2/2)
\]

\[
\Rightarrow \Pr\{|f(x) - M - t| \leq t\} \leq \exp(-t^2/2)
\]

Repeat for \( -f \)

Replacing median by mean: DIY. QED
Application to random matrices:

**Theorem (Concentration of eigenvalues)** Let \( A \) be a \( n \times n \) symmetric random matrix with iid \( \mathcal{N}(0,1) \) entries on \& above diagonal. Then, for \( i \in \{1, \ldots, n\} \):

\[
P\left\{ |\lambda_i(A) - E\lambda_i(A)| \geq t \right\} \leq 2\exp\left(-\frac{t^2}{4} \right),
\]

\( \forall t > 0 \).

**Proof**

1. **Vectorize** \( A \) with the vector \( \bar{A} \in \mathbb{R}^{n(n+1)/2} \) that contains the entries of \( A \) on \& above diagonal.

2. **Apply Gauss. concentration** for \( f: \mathbb{R}^{n(n+1)/2} \to \mathbb{R} \):

\[
f(\bar{A}) = \lambda_i(A).
\]

3. **Weyl's inequality**

\[
|f(A) - f(B)| = |\lambda_i(A) - \lambda_i(B)| \leq \|A-B\|_F
\]

\[
\leq \|A-B\|_F \leq \sqrt{\sum (A-B)^2}_F = \sqrt{\text{square of entries}}\]

\[
\Rightarrow f \text{ is } \sqrt{2} - \text{Lipschitz}.
\]

4. Gauss. concentration \( \therefore \) \( \square \).

**Q:** \( E\lambda_i(A) = ? \)

Use Wigner's semicircle law (lec. 25, Oct 31):

- **Set area** \( \frac{1}{n} \#	ext{(eigenvalues } \geq x) \)
- **Solve** for \( x = E\lambda_i(A) \)

\[
\begin{align*}
-2\sqrt{n} & \quad 0 & \quad 2\sqrt{n} \\
\uparrow & & \uparrow \\
\text{area} & = \frac{1}{n} \#	ext{(eigenvalues } \geq x) & \text{Set area} = \frac{1}{n}, \text{ solve for } x = E\lambda_i(A)
\end{align*}
\]
CONCLUDING REMARKS

Further literature:

1. My book (e.g., beyond Gauss concentration)
2. Joel Tropp "Probability in high dimensions"
3. Van Handel "Probability in high dimensions"—more advanced
4. Wainwright "High-dimensional statistics"

2ND edition