Lecture 16

Continuing with the proof of lower tail bound for weakly \((a, b)\)-self bounding functions

Lemma we proved last time

Suppose that

\[ 2H'(\lambda) - H(\lambda) \leq p(\lambda) ( -aH'(\lambda) + v ) \]

for \( \lambda \in (0, \theta) \), where

(i) \(-aH'(\lambda) + v > 0 \) for \( 0 \leq \lambda < \theta \),

(ii) \( H'(0) = H(0) = 0 \)

Also, suppose

\[ 2G'_0(\lambda) - G_0(\lambda) \geq p_0(\lambda) ( -aG'_0(\lambda) + 1 ) \]

where

(iii) \(-aG'_0(\lambda) + 1 > 0 \), for \( 0 \leq \lambda < \theta \),

(iv) \( G'_0(0) = G_0(0) = 0 \).

(v) \( G''_0(0) = 1 \).

If \( p_0(\lambda) \geq p(\lambda) \), then \( H(\lambda) \leq vG_0(\lambda) \) for \( 0 \leq \lambda < \theta \).

Proof simplified:

\[
\left( \frac{2H'(\lambda) - H(\lambda)}{-aH'(\lambda) + v} \right) \geq \left( \frac{2G'_0(\lambda) - G_0(\lambda)}{-aG'_0(\lambda) + 1} \right)
\]
Lemma we are yet to prove

For \( a \geq \frac{1}{3} \) and \( \nu > 0 \), suppose that
\[
2 H'(\lambda) - H(\lambda) \leq \phi(\lambda) \left( -a H'(\lambda) + \nu \right),
\]
\[ 0 \leq \lambda < \frac{1}{a}. \]

Then, \( H(\lambda) \leq \nu \lambda^2 \).

Proof. In view of the previous lemma, our goal is to identify an upper bound \( \phi_0(\lambda) \) of \( \phi(\lambda) \) in \( 0 \leq \lambda < \frac{1}{a} \) and show that \( G_0(\lambda) = \lambda^2/2 \) solves
\[
(\#) \quad 2 G'_0(\lambda) - G_0(\lambda) = \phi_0(\lambda) \left( -a G'_0(\lambda) + 1 \right).
\]
(a) \( \phi(\lambda) \leq \frac{\lambda^2}{2(1 - \lambda/3)} \leq \frac{\lambda^2}{2(1 - a \lambda)} =: \phi_0(\lambda), \quad 0 \leq \lambda < \frac{1}{a} \)

(b) For \( G_0(\lambda) = \lambda^2/2 \),
\[
2 G'_0(\lambda) - G_0(\lambda) = \lambda - \frac{\lambda^2}{2} = \frac{\lambda^2}{2} \phi_0(\lambda) \left( -a G'_0(\lambda) + 1 \right) = \frac{\lambda^2}{2(1 - a \lambda)} \cdot ( -a + 1 )
\]
Also, for \( \lambda < \frac{1}{a} \), \( 1 - a G'_0(\lambda) > 0 \). Thus,
\( G_0(\lambda) = \lambda^2/2 \) solves \( (\#) \) and so, \( H(\lambda) \leq \nu \lambda^2, \quad 0 \leq \lambda < \frac{1}{a}. \)
Review:

(a) Hoeffding, Azuma, McDiarmid
→ need bound for fluctuation along each coordinate (BDP)

Q: How can we relax BDP to a bound on total fluctuation?

(b) Efron-Stein gives handle over variance.
Can be extended to conc (around mean & median)
using a differential inequality.
But we only get \( \leq \exp(-t/\Delta^2) \)

(c) Entropy method: (Subadditivity → Log-Sobolev →
      differential inequality)

    Could get: \( \sum_{i=1}^{n} (Z - Z_0)^2 \leq \Delta^2 \),
    \( Z_0 = \inf_{x \in \mathbb{R}^n} \{ f(x_i, x') \} \)

    \( \Rightarrow P(Z - \mathbb{E}Z > t) \leq \exp(-\frac{t^2}{\Delta^2}) \)

    Additionally, if \( Z - Z_0 \leq 1 \),

    \( P(Z - \mathbb{E}Z < -t) \leq \exp(-t h(\frac{t}{\Delta})) \)

(d) What is still missing?
→ Extensions of Bennett, Bernstein (require only)

Could do it for weakly \((a, b)\)-self-bounding.
For \( a \geq 1/2 \), we get sub-Gaussian style bound.
We can use the Entropy method to get the "Bennett" extension. But we won't.

Coming up: (1) Concentration and Isoperimetric Inequality

* Talagrand's convex distance ineq.
* Concentration around median

(2) Transportation Inequality Method

* We will get the "Bennett" extension, that too with a simple proof.

But before we continue with concentration bounds, we take a detour to cover:

* The threshold phenomenon (application of log-Sobolev ineq.)
* Hypercontractivity (connection with log-Sobolev)
* Analysis of LASSO (Gaussian concentration)

The Threshold Phenomenon

Consider the binary hypercube \([-1, +1^n]\).
Let \(X_1, \ldots, X_n\) be iid over \([-1, +1]\) with \(P(X_i = 1) = p\).
Fix a set \(A \subseteq [-1, +1^n]\). Let \(P_p(A)\) denote \(P(X \in A)\).
* For reasonable sets \(A\), \(P_p(A)\) increases from 0 to 1 as \(p\) increases from 0 to 1.
What is amazing is that the shift from 0 to 1 happens over a very narrow interval around a "threshold" \( p_0 \).

### Influence of Boolean functions

Recall the binary log-Sobolev inequality:

\[
\text{Ent}(f^2) \leq 2 \mathcal{E}(f) = \frac{1}{2} \sum_{i=1}^{n} (f(x) - f(\overline{x}^{(i)}))^2
\]

For the special case of Boolean \( f \), namely \( f: \{-1,1\}^n \to \{0,1\} \), we can represent \( f \) on \( \mathbb{B}^n \). Then, denoting \( P_x(A) \) by \( P(A) \)

\[
\text{Ent}(f^2) = \text{Ent}(f) = -P(A) \log P(A)
\]

\[
\mathbb{E} \left[ (f(x) - f(\overline{x}^{(i)}))^2 \right] = P \left( f(x) \neq f(\overline{x}^{(i)}) \right)
\]

* This quantity on the right is called the \( i \)-th influence of \( f \), denote \( I^i(f) \) or \( I^i(A) \).

* The influence of \( f \) on \( A \) is defined as

\[
I(A) = I(f) = \sum_{i=1}^{n} I^i(f)
\]

* If \( f(x) \neq f(\overline{x}^{(i)}) \), we say that the \( i \)-th variable is pivotal on \( x \), or \( x \) is pivotal (at \( x \)).

* Similarly, we can define \( I_p^i(A) \) and \( I_p(A) \).
In this new language, BLSI is the same as

\[ P(A) \log \frac{1}{P(A)} \leq \frac{1}{2} I(A) \]

**B Monotone Sets and Russo’s Lemma**

**Definition** A set \( A \) is monotone if when \( x \in A \), then \( x_i^+ = (x^+, 1, x_{i+1}^-) \in A \).

**Simple Observation** For a monotone set \( A \neq \emptyset \) or \( x \in L_1^\infty \)

\[ P_0(A) = 0 \text{ and } P_1(A) = 1 \]

**Theorem (Russo’s Lemma)**

For every monotone set \( A \neq \emptyset \) on \( L_1^\infty \),

\[ \frac{d}{d\rho} P_\rho(A) = I_\rho(A) \]

Here \( I_\rho(A) = P_\rho(\{ x \in A \text{ and } x_{i+1}^+ \neq A \text{ or } x_{i+1}^- \in A \}) \)

\[ I_\rho(A) = \sum_{i=1}^n I_\rho(A) \]

**Proof** Let \( U_1, ..., U_n \) be iid uniform \( (0, 1) \). Given \( \rho = (\rho_1, ..., \rho_n) \) and \( \hat{\rho} = (\hat{\rho}_1, ..., \hat{\rho}_1, \rho_{i+1}, ..., \rho_n) \), let

\[ X_j = 2 \mathbb{I}_{\{ U_j < \rho_j \}} - 1 \text{ and } \hat{X}_j = 2 \mathbb{I}_{\{ U_j < \hat{\rho}_j \}} - 1, 1 \leq j \leq n. \]

Consider the case \( \hat{\rho}_i > \rho_i \). Then, since \( A \) is monotone,
Thus, \( P(\hat{X} \in A) - P(X \in A) \)
\[ = P(\hat{X} \in A, X \notin A) \]
\[ = P(\bigcup_{i \in (p, \hat{p})_i}, X^+ \in A, X^- \notin A) \]
\[ = (\hat{p}_i - p_i)P(X^+ \in A, X^- \notin A). \]

For \( p_j = p \) for every \( j \neq i \), we get
\[ P(\hat{X} \in A) - P(X \in A) = (\hat{p}_i - p_i)I^i_p(A) \]
and similarly, when \( \hat{p}_i < p_i \),
\[ P(\hat{X} \in A) - P(X \in A) = -(\hat{p}_i - p_i)I^i_p(A). \]

Thus, denoting by \( Q_p(A) \) the prob. \( X \in A \) for \( p = (p, \ldots, p) \),
\[ \frac{d}{dp_i} Q_p(A) = I^i_p(A) \]

which gives
\[ \frac{dP_p(A)}{dP} = \sum_{i=1}^{n} \frac{dQ_p(A)}{dp_i} = \sum_{i=1}^{n} I^i_p(A) = 1(A). \]