Lecture 16

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

Lemma we proved last time

Suppose that

$$\boxed{\lambda H'(\lambda) - H(\lambda) \le \rho(\lambda)\left(-a H'(\lambda) + v\right)}$$

for $\lambda \in [0, \theta)$, where

(i) $-a H'(\lambda) + v > 0 \quad \forall\ 0 \le \lambda < \theta$;

(ii) $H'(0) = H(0) = 0$

Also, suppose

$$\boxed{\lambda G_0'(\lambda) - G_0(\lambda) \ge \rho_0(\lambda)\left(-a G_0'(\lambda) + 1\right)}$$

where

(iii) $-a G_0'(\lambda) + 1 > 0, \quad \forall\ 0 \le \lambda < \theta$,

(iv) $G_0'(0) = G_0(0) = 0$.

(v) $G_0''(0) = 1$.

If $\rho_0(\lambda) \ge \rho(\lambda)$, then $\boxed{H(\lambda) \le v\, G_0(\lambda)}$, $0 \le \lambda < \theta$.

Proof simplified:

$$\left(\frac{\lambda H'(\lambda) - H(\lambda)}{-a H'(\lambda) + v}\right) \ge \left(\frac{\lambda G_0'(\lambda) - G_0(\lambda)}{-a G_0'(\lambda) + 1}\right)$$

$$\Leftrightarrow \frac{d}{d\lambda}\left(\frac{H(\lambda)}{(\lambda - aG_0(\lambda))} - \frac{v\, G_0(\lambda)}{(\lambda - aG_0(\lambda))} \right) \le 0$$ ②

■

Lemma we are yet to prove

For $a \geq 1/3$ and $v > 0$, suppose that

$$\lambda H'(\lambda) - H(\lambda) \le \phi(\lambda)\left(-a H'(\lambda) + v\right),$$

$$\forall \quad 0 \le \lambda < 1/a.$$

Then, $H(\lambda) \le \dfrac{v\lambda^2}{2}$.

Proof. In view of the previous lemma, our goal is to identify an upper bound $\phi_0(\lambda) \ge \phi(\lambda)$ in $0 \le \lambda < 1/a$ and show that $G_0(\lambda) = \lambda^2/2$ solves

$$(\#) \quad \boxed{\lambda G_0'(\lambda) - G_0(\lambda) = \phi_0(\lambda)\left(-a G_0'(\lambda) + 1\right)}.$$

(a) $\phi(\lambda) \le \dfrac{\lambda^2}{2(1-\lambda/3)} \le \dfrac{\lambda^2}{2(1-a\lambda)} =: \phi_0(\lambda), \quad 0 \le \lambda < \frac{1}{a} \le 3.$

(b) For $G_0(\lambda) = \lambda^2/2$,

$$\lambda G_0'(\lambda) - G_0(\lambda) = \lambda . \lambda - \frac{\lambda^2}{2} = \frac{\lambda^2}{2}$$

$$\phi_0(\lambda)\left(-a G_0'(\lambda) + 1\right) = \frac{\lambda^2}{2(1-a\lambda)} \cdot (-a\lambda + 1)$$

Also, for $\lambda < 1/a$, $1 - a G_0'(\lambda) > 0$. Thus, $G_0(\lambda) = \lambda^2/2$ solves (#), and so $H(\lambda) \le \dfrac{v\lambda^2}{2}$, $0 \le \lambda < \frac{1}{a}$.

■

Review:-

(a) Hoeffding, Azuma, Mc Diarmid

→ need bound for fluctuation along each coordinate (BDP)

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

(b) Efron-Stein gives a handle over variance.

Can be extended to conc. (around mean & median) using a differential inequality.

But we only get $\leq \exp(-t/\sqrt{v})$

(c) Entropy method: (subadditivity → log-Sobolev ↓ differential inequality for $\psi_{Z-\mathbb{E}Z}(\lambda)$)

Could get: $\sum_{i=1}^{n}(Z-Z_i)^2 \leq v$, $Z_i = \inf_{x_i'} f(X^{i-1}, x_i', X_{i+1}^n)$

$$\Rightarrow \boxed{\mathbb{P}(Z-\mathbb{E}Z > t) \leq \exp\left(-\frac{t^2}{v}\right)}$$

Additionally, if $Z-Z_i \leq 1$,

$$\boxed{\mathbb{P}(Z-\mathbb{E}Z < -t) \leq \exp\left(-v\,h\left(\frac{t}{v}\right)\right)}.$$

(d) What is still missing?

→ Extensions of Bennett, Bernstein (require only $\mathbb{E}[\cdot]$ bound)

Could do it for weakly (a,b)-self bounding.

For $a \geq 1/3$, 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 $\mathbb{P}(X_i = 1) = p$.

Fix a set $A \subseteq \{-1,+1\}^n$. Let $\mathbb{P}_p(A)$ denote $\mathbb{P}(X \in A)$.

* For reasonable sets (?) A, $\mathbb{P}_p(A)$ increases from 0 to 1 as p increases from 0 to 1.

* What is amazing is that the shift from 0 to 1 happends over a very narrow interval around a "threshold" p_0.

$\boxed{A}$ Influence of Boolean functions

Recall the binary log-Sobolev inequality:

$f : \{-1,1\}^n \to \mathbb{R}$

$$\mathrm{Ent}(f^2) \le 2\mathcal{E}(f) = \frac{1}{2}\sum_{i=1}^{n}\left(f(X) - f(\bar{X}^{(i)})\right)^2$$

For the special case of Boolean f, namely $f : \{-1,1\}^n \to \{0,1\}$, we can represent f as $\mathbb{1}_A$.

Then, denoting $\mathbb{P}_{1/2}(A)$ by $\mathbb{P}(A)$

$\to$ $\mathrm{Ent}(f^2) = \mathrm{Ent}(f) = -\ \mathbb{P}(A)\log \mathbb{P}(A)$

$\to$ $\mathbb{E}\left[(f(X) - f(\bar{X}^{(i)}))^2\right] = \mathbb{P}\left(f(X) \ne f(\bar{X}_i)\right)$

* 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 or A is defined as

$$I(A) \equiv I(f) = \sum_{i=1}^{n} I^i(f).$$

* If $f(X) \ne f(\bar{X}^{(i)})$, we say that the i^{th} variable is pivotal or X_i is pivotal (at X).
* Similarly, we can define $I_p^i(A)$ and $I_p(A)$.

→ In this new langauge, BLSI is the same as

$$\boxed{\mathbb{P}(A)\log\frac{1}{\mathbb{P}(A)} \le \frac{1}{2}\, I(A)}$$

B Monotone Sets and Russo's Lemma

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

Simple Obsevation: For a monotone set $A \neq \phi$ or $\{-1,1\}^n$

$$\mathbb{P}_0(A) = 0 \text{ and } \mathbb{P}_1(A) = 1.$$

Theorem (Russo's Lemma)

For every monotone set $A \neq \phi$ on $\{-1,1\}^n$,

$$\boxed{\frac{d}{dp}\,\mathbb{P}_p(A) = I_p(A).}$$

Here $I_p^i(A) = \mathbb{P}_p\left(X \in A \text{ and } \bar{X}^{(i)} \notin A \text{ or } X \notin A \text{ and } X^{(i)} \in A\right)$

$$I_p(A) = \sum_{i=1}^{n} I_p^i(A)$$