3.2 Concentration of Measure on $I_n$

Define Lipschitz functions as before, then $\exists m_f$ such that,

\[ \mu_n(\{x \in I_n : f(x) \geq m_f\}) \geq \frac{1}{2} \]

and

\[ \mu_n(\{x \in I_n : f(x) \leq m_f\}) \geq \frac{1}{2} \]

3.2.1 Theorem. If $f : I_n \to \mathbb{R}$ is 1-Lipschitz and $m_f$ is its median then for $\epsilon > 0$,

\[ \mu_n(\{x \in I_n : |f(x) - m_f| \geq \epsilon \sqrt{n}\}) \leq 4e^{-\epsilon^2}. \]

Proof. Let,

\[ A_+ = \{x \in I_n : f(x) \geq m_f\} \]
\[ A_- = \{x \in I_n : f(x) \leq m_f\} \]

and\[ A_+(\epsilon) = \{x \in I_n : d(x, A_+) \leq \epsilon \sqrt{n}\} \]
\[ A_-(\epsilon) = \{x \in I_n : d(x, A_-) \leq \epsilon \sqrt{n}\} \]

By Corollary to Talagrand’s Theorem,

\[ \mu_n(A_+(\epsilon)) \geq 1 - \frac{e^{-\epsilon^2}}{\mu_n(A_+)} \geq 1 - 2e^{-\epsilon^2} \]

and

\[ \mu_n(A_-(\epsilon)) \geq 1 - 2e^{-\epsilon^2} \]

Now we have for $A(\epsilon) = A_+(\epsilon) \cap A_-(\epsilon)$ on which $f$ attains values between $m_f - \epsilon \sqrt{n}$ and $m_f + \epsilon \sqrt{n}$ that,

\[ \mu_n(A(\epsilon)) \geq 1 - 4e^\epsilon^2 \]

by the union bound. \qed
4 The Martingale Method for the Boolean Cube

4.1.1 Question. Concentration theorem is conveniently formulated with the median. What about the mean?

4.1.2 Definition. If $(X, \mathcal{A}, \mu)$ is a probability space, with measure $\mu$ and $\mathcal{F}$ a sub-$\sigma$-algebra of $\mathcal{A}$, then for $f : X \rightarrow \mathbb{R}$, $\mathcal{A}$-measurable, $h : X \rightarrow \mathbb{R}$ is a conditional expectation with respect to $\mathcal{F}$ if it is $\mathcal{F}$-measurable and for all $A \in \mathcal{F}$,

$$\int_A hd\mu = \int_A fd\mu.$$ 

4.1.3 Remark. Conditional Expectation exists even for $\sigma$-finite measure spaces by Radon-Nikodym Theorem. For finite $X$, and $\mu(\{x\}) > \epsilon$ for all $x \in X$, $\mathcal{F}$ is generated by some partition of $X$. Then $h$ is a constant on sets of partition generating $\mathcal{F}$, so for $A$ in this partition, $y \in A$,

$$\int_A h(x)d\mu(x) = \int_A f(x)d\mu(x)$$

$$\Rightarrow \ h(y)\mu(A) = \int_A f(x)d\mu(x)$$

$$\Rightarrow \ h(y) = \frac{1}{\mu(A)} \int_A f(x)d\mu(x)$$

We abbreviate these local averages by,

$$h = E[f|\mathcal{F}]$$

We also denote $E[f] = \int_X f d\mu$.

4.1.4 Note. 1. $E[f] = E[E[f|\mathcal{F}]]$.

2. If $\mathcal{F}_1 \subseteq \mathcal{F}_2$ then,

$$E[E[f|\mathcal{F}_2]|\mathcal{F}_1] = E[f|\mathcal{F}_1].$$

3. If $f(x) \leq g(x)$ for all $x \in X$ then,

$$E[f|\mathcal{F}] \leq E[g|\mathcal{F}]$$

pointwise on $X$.

4. If $g$ is $\mathcal{F}$-measurable then,

$$E[fg|\mathcal{F}] = gE[f|\mathcal{F}].$$

4.1.5 Definition. Given a sequence of $\sigma$-algebras, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_n$, with $\mathcal{F}_0 = \{\emptyset, X\}$ and $\{x\} \in \mathcal{F}_n$ for all $x \in X$, then for $f : X \rightarrow \mathbb{R}$, $f_i = E[f|\mathcal{F}_i]$, the sequence $f_0, f_1, \ldots f_n$ is called a martingale.

We use this to prove concentration results.
4.1.6 Theorem. Let $X$ be a finite probability space and let $\sigma$-algebras $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \ldots \subseteq \mathcal{F}_n$ together with $f : X \to \mathbb{R}$ define a martingale. If there are $d_1, d_2, \ldots, d_n$ such that,

$$\|f_i - f_{i-1}\|_\infty \leq d_i$$

for each $i$ and $a = E[f]$, $D = \sum_{i=1}^n d_i^2$, then for $t \geq 0$,

$$\mu(\{x \in X : f(x) \geq a + t\}) \leq e^{\frac{-t^2}{2D}}$$

and

$$\mu(\{x \in X : f(x) \leq a - t\}) \leq e^{\frac{-t^2}{2D}}$$

As a first step we bound the Laplace Transform.

4.1.7 Lemma. Let $f : X \to \mathbb{R}$ be of zero mean and bounded by $d$, that is,

$$\int_X f \, d\mu = 0 \text{ and } |f(x)| \leq d \quad \forall \ x \in X$$

Then for $\lambda \geq 0$,

$$\int_X e^{\lambda f} \, d\mu \leq e^{\lambda d} + e^{-\lambda d} \leq e^{\frac{\lambda d^2}{2}}$$

Proof. By scaling assume $d = 1$. Note $t \to e^{\lambda t}$ is convex. So by interpolating linearly between $-1$ and $1$, we get the inequality,

$$e^{\lambda t} \leq e^{-\lambda} + \frac{e^{\lambda} - e^{-\lambda}}{2} (t + 1)$$

$$= \frac{e^{\lambda} + e^{-\lambda}}{2} + \frac{e^{\lambda} - e^{-\lambda}}{2} t$$

when $-1 \leq t \leq 1$.

Now inserting $f(x)$ instead of $t$ and integrating gives,

$$\int_X e^{\lambda f} \, d\mu \leq \frac{e^{\lambda} + e^{-\lambda}}{2} + 0 \leq e^{\frac{\lambda^2}{2}}$$

The last step holds by Taylor expansion.

Proof. (of the theorem)

We only need to show the first inequality, $\mu(\{x \in X : f(x) \geq a + t\}) \leq e^{\frac{-t^2}{2D}}$ for all $f$ because the second one follows by replacing $f \mapsto -f$, $a \mapsto -a$.

By Laplace Transform method,

$$\mu(\{x \in X : f(x) - a \geq t\}) \leq e^{-\lambda t} \int_X e^{\lambda(f-a)} \, d\mu$$
Now we can insert,
\[ f - a = f_n - f_0 = \sum_{i=0}^{n} (f_i - f_{i-1}). \]

Denoting \( g_i = f_i - f_{i-1} \),
\[
\int_X e^{\lambda(f-a)} d\mu = \sum_{i=1}^{\lambda} e^{\lambda g_i} d\mu = E_0 [E_1[...E_{n-1}[e^{\lambda g_1} e^{\lambda g_2} ... e^{\lambda g_n}]]] \\
with\ E_k[h] \equiv E[h|F_k]
\]

We note for \( 1 \leq k \leq n \),
\[
E_0[E_1[...E_{k-1}[e^{\lambda g_1} e^{\lambda g_2} ... e^{\lambda g_k}]]] = E_0[E_1[...E_{k-2}[e^{\lambda g_1} e^{\lambda g_2} ... e^{\lambda g_{k-1}} E[e^{\lambda g_k}]]]]
\]
and recall, \(|g_k(x)| \leq d_k\) for all \( x \), as well as,
\[
E_{k-1}[g_k] = E[f_k|F_{k-1}] - E[f_{k-1}|F_{k-1}] = f_k - f_{k-1} = 0.
\]

So by applying the lemma to each block in partition \( E_{k-1}[e^{\lambda g_k}] \leq e^{\frac{\lambda^2 d_k^2}{2}} \). This implies,
\[
E_0[E_1[...E_{k-1}[e^{\lambda g_1} e^{\lambda g_2} ... e^{\lambda g_k}]]] \leq e^{\frac{\lambda^2 d_k^2}{2}} E_0[E_1[...E_{k-2}[e^{\lambda g_1} e^{\lambda g_2} ... e^{\lambda g_{k-1}}]]]
\]
\[
\vdots
\]
\[
\leq e^{\frac{\lambda^2}{2} \sum_{i=1}^{k} d_i^2}
\]
This gives,
\[
E[e^{\lambda g_1} e^{\lambda g_2} ... e^{\lambda g_n}] \leq e^{\frac{\lambda^2}{2} \sum_{i=1}^{k} d_i^2} = e^{\frac{\lambda^2 D^2}{2}}
\]

Now choosing \( \lambda = \frac{t}{D} \), we have,
\[
\mu(\{x \in X : f(x) - a \geq t\}) \leq e^{-\lambda t} \int_X e^{\lambda(f-a)} d\mu \leq e^{\frac{-t^2}{2D^2}} e^{\frac{t^2}{2D^2}} = e^{\frac{-t^2}{2D^2}}.
\]