MASSACHUSETTS INSTITUTE OF TECHNOLOGY

6.265/15.070J Lecture 8 Feb 15, SP17

Lecturer: Yury Polyanskiy Based on scribe notes by Dheeraj Nagaraj

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They are posted to serve class purposes.*

Concentration of Measure II

Content.

1. McDiarmid Theorem
2. Concentration of Lipschitz functions
3. Applications

Many absolute constants are not consistent in the results below. But order of magnitudes are accurate.

1 McDiarmid Theorem

Theorem 1. *Let $f(x_1, ..., x_n)$ be such that $|f(x_1, .., x_i, .., x_n) - f(x_1, .., x_i', .., x_n)| \leq c_i$, $\forall x, x_i', i$. Then, for independent random variables, $X_1, ..X_n$,*

$$\mathbb{P}(|f(X) - \mathbb{E}f(X)| > t) \leq \exp^{\frac{-t^2}{\sum_{k=1}^{n} c_k^2}}$$

Proof. Define $S_k = \mathbb{E}[f(X)|X_1, .., X_k]$. This is clearly a Doob martingale. We need to show that $|S_k - S_{k-1}| = |\mathbb{E}[f|X_1, .., X_k] - \mathbb{E}[f|X_1, ..., X_{k-1}]| \leq c_k$. We check for $k = 1$ and then generalize the argument.

$$\begin{aligned}
|S_1 - S_0| &= |\sum_{X_2,..X_n} \mathbb{P}(X_2, .., X_n|X_1) f(X_1, .., X_n) - \sum_{X_1', X_2, ..., X_n} P(X_1', X_2, ..., X_n) f(X_1', X_2, .., X_n)| \\
&\leq \sum_{X_1', X_2, ..., X_n} P(X_1', X_2, ..., X_n) \left[f(X_1, X_2, .., X_n) - f(X_1', X_2, .., X_n)\right] \\
&\leq c_1
\end{aligned}$$

Therefore, S_k is a bounded difference martingale. We apply Azuma-Hoeffding Theorem to conclude the result. ☐

2 Metric Measure Spaces

Definition 1. *(χ, μ, d) is called a metric measure space if χ is a metric space with distance d and μ is the borel measure with respect to metric topology of d.*

Definition 2. *A function $f : \chi \to \mathbb{R}$ is called c-Lipschitz if $\forall x, x' \in \chi$,*

$$|f(x) - f(x')| \leq cd(x, x')$$

Lemma 2. *f is c-lipschitz in d_H if and only if $|f(X_i, X_{\sim i}) - f(X_i', X_{\sim i})| \leq c$ for every i, X, X_i'. Similarly, a function is c-Lipschitz in ℓ_1 iff it is c-Lipschitz in each coordinate.*

Proof. Only if condition is trivial. For the sufficiency part : Consider X and X' to be arbitrary.

$$\begin{aligned}|f(X) - f(X')| &\leq |f(X) - f(X_1', X_{\sim 1})| + |f(X_1', X_{\sim 1}) - f(X')| \\ &\leq c\mathbb{1}_{X_1 \neq X_1'} + |f(X_1', X_{\sim 1}) - f(X')|\end{aligned}$$

Continuing recursively, we conclude the result. □

Example :

1. χ^n, $d_H(X^n, Y^n) \triangleq \sum_{i=1}^n \mathbb{1}_{X_i \neq Y_i}$, called the Hamming distance.
2. $\mathbb{R}^n$, $l_1(X, Y) = \sum_{i=1}^n |X_i - Y_i|$
3. $\mathbb{R}^n$ with euclidean distance (also called l_2 metric)

Exercise 1. *Plot the curves $l_1(X, 0) = 1$, $d_H(X, 0) = 1$ and $l_2(X, 0) = 1$ over $\mathbb{R}^2$*

Example : Consider $f(X^n) = \sum_{i=1}^n X_i$. This is 1-Lipschitz with l_1 metric and $\sqrt{n}$-Lipschitz with l_2 metric.

Corollary 3. *Any $1-$Lipschitz function with Hamming distance over any product measure $\mu = \prod_{i=1}^n \mu_i$ on $\mathcal{X}^n$ satisfies the following concentration inequality.*

$$\mu(|f - \mathbb{E}f| > t) \leq 2e^{-\frac{t^2}{2n}} \quad (1)$$

Proof. Use McDiarmid's theorem. □

Corollary 4. *Any 1-Lip function (ℓ_1) on $[0, 1]^n$ with product measure over $\mathcal{X}^n$ satisfies equation 1*

Definition 3. *(χ, μ, d) is said to have (b, ν) conc. if every 1-Lipschitz function (b, ν) - sub-gaussian.*

We prove the following useful lemma to check for c Lipschizness with respect to Hamming distance.

3 Applications

3.1 Application 1

[Learning Theorem] Let $X_i \stackrel{\text{iid}}{=} \mathbb{P}_X$ on $\mathbb{R}$. Define $\hat{F_n}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{X_i \leq t}$ (empirical CDF). Let $Z = \sup_t |\hat{F}_n(t) - F_X(t)| := f(X_1, .., X_n)$ where F_X is the CDF of X.

$$|f(X_{\sim i}, X_i) - f(X_i, X_i')| \leq 2sup_t \frac{1}{n} \sum_{i=1}^{n} |\mathbb{1}_{X_i \leq t} - \mathbb{1}_{X_i' \leq t}| \leq \frac{2}{n}$$

Therefore f is a $\frac{2}{n}$ - Lipschitz. Using McDiarmid's theorem, we conclude $Z = EZ \pm O_p(\frac{1}{\sqrt{n}})$ Therefore, we conclude that the empirical distribution function converges to the distribution function uniformly.

This is a stronger version of Glivenko Cantelli Theorem.

3.2 Application 2

We consider throwing n balls into m bins at random. Consider i.i.d random variables $X_i \in [m]$ be uniformly distributed, $i \in \{1, 2, ..., n\}$. i-th ball goes into the X_ith bin. Define

$M_j = f_j(X) =$ number of bins with j -balls.

M_0 = number of empty bins.

Clearly, f_j is 2-Lipschitz with respect to d_H. Applying McDiarmid's theorem, we show that

$$\mathbb{P}[|M_i - \mathbb{E}(M_i)| > t] \leq e^{t^2} 8n$$

Therefore, $M_0 = \mathbb{E}M_0 \pm O_p(\sqrt{n})$.

3.3 Application 3

Consider the problem of longest common subsequence of independent random binary sequences - LCS(X^n, Y^n). Clearly, by our co-ordinate wise characterization of Lipschitz functions, this is a 1-Lipschitz function with respect to the Hamming distance. Therefore, by McDiarmid's theorem, LCS $= \mathbb{E}$LCS $\pm$ $O_p(\sqrt{n})$. This is a very useful result since $\mathbb{E}\text{LCS} = \theta(n)$

3.4 Application 4: Chromatic number concentrates within $O(\sqrt{n})$

Note: This method is what introduced bounded-differences method to analysis of Lipschitz functions. In particular, it predateds McDiarmid's theorem.

Consider $G = G(n,p)$. Define $\chi(G)$ to be chromatic number of G. Let $e \in \binom{[n]}{2}$ and $X_e = 1$ if e is present in G and $X = 0$ otherwise. Clearly, $\chi(G) = f(X_e, e \in \binom{[n]}{2})$. Clearly, f is a 1- Lipschitz function since addition or deletion of one edge can change the chromatic number by at most 1. But since there are $\frac{n(n-1)}{2} = \theta(n^2)$ edges. Therefore, McDiarmid's theorem implies $\mathbb{P}(|\chi - \mathbb{E}\chi| > t) \leq 2e^{-\frac{t^2}{n^2}}$. $\Rightarrow \chi = \mathbb{E}\chi \pm O_p(n)$. This is a useless result since $\chi = O(n)$ a.s.

We get a better result using Shamir-Spencer method. $Y_1 = \emptyset$, $Y_2 \in \{0,1\}$, ... , $Y_i \in \{0,1\}^i$. The random i-tuple Y_i describe the existence/ non-existence of edges from vertex i to vertices $\{1, ..., i-1\}$ with $Y_i(k) = 0$ if there is no edge between i and k and $Y_i(k) = 1$ otherwise. This is the same description as X_e, but we have bunched together edges emanating from a given vertex to vertices with lower index. Therefore, $\chi = f(Y_1, ..., Y_n)$. We use "vertex exposure" process to show that χ is 1-Lipschitz.

Consider any arbitrary change in Y_i i.e, addition or removal of edges from vertex i to a vertex of index lesser than i. The neighborhood of every other vertex remains the same except for addition or subtraction of vertex i. Therefore, for the next minimal coloring, we might need to add a completely new color to i or change the color of i to one existing before because it was removed from a neighborhood with that pre existing color. Therefore, χ can change by utmost 1. By our characterization of $c-$ Lipschitz function with respect to Hamming distance, we show that χ is $1-$ Lipschitz with respect to Y. Clearly, Y_i are all independent. We can use McDiarmid's theorem to show that $\chi = \mathbb{E}\chi \pm O_p(\sqrt{n})$.

Every vertex which is part of a clique of G must have a distinct color. Therefore, to study the chromatic number, we study the appearance of cliques of different sizes in G.

3.5 Application 5: all $\log n$-subsets of $G(n,p)$ are not triangle-free

Let $k_3(G)$ denote the number of triangles in G. Clearly, $\mathbb{E}k_3(G) = \binom{n}{3}p^3 \sim \frac{n^3p^3}{6}$. $k_3(G) = f(X_e)$ where X_e are edge indicator functions as defined before. Any edge can be part of at most n distinct triangles. Therefore we conclude that $k_3(G)$ is n-Lipschitz with respect to d_H. We apply McDiarmid's theorem to conclude

$$\mathbb{P}(k_3(G) = 0) \leq \mathbb{P}(|k_3 - \mathbb{E}k_3| \geq \mathbb{E}k_3) \leq 2e^{-\frac{(\mathbb{E}k_3)^2}{2n^4}} \sim e^{-n^2p^6}. \qquad (2)$$

(Note: Janson's inequality (see homework) gives better estimate $e^{-n^2 p}$.)

Proposition 5. *Any m-vertex induced subgraph of $G(n,p)$ with $m \geq \lceil c_p \log n \rceil$ contains a triangle with high probability for some constant c_p dependent only on p.*

Proof. Let S be any set of vertices and $G[S]$ be the subgraph induced by the vertices in S. Clearly, $\mathbb{P}(\exists S : |S| = m, G[S] \text{ is triangle free}) \leq \binom{n}{m}\mathbb{P}(G(m,p) \text{ is triangle free}) \approx e^{m \log n - m^2 p^6} = e^{(\log n)^2 c_p (p^6 - c_p)}$. The RHS $\to 0$ as $n \to \infty$ if $c_p \geq p^6$, concluding our proof. □

3.6 Application 6: Chromatic number of dense random graphs

Our main theorem is one due to Bollobas. Let $G = G(n,p)$ with $p \in (0,1)$ fixed.

Theorem 6 (Bollobas)**.** *Fix $p \in (0,1)$ and let $n \to \infty$. Then $\chi(G(n,p)) = (1 + o(1))\frac{n}{2\log n} \log \frac{1}{1-p}$ with high probability.*

Let $k_r(G)$ denote the number of r-cliques in G. Then, $\mathbb{E}(k_r(G)) = \binom{n}{r} p^{\binom{r}{2}} \triangleq E(n,r)$. The function $r \mapsto E(n,r)$ is unimodal. It increases until abour $r \approx \log \frac{1}{p} n$ and then monotonically decreases. Let r_0 be such that $E(n, r_0) < 1 < E(n, r_0 - 1)$, then from tedious Stirling's approximations we can get:

- $r_0 = 2 \log_{\frac{1}{p}} n - 2 \log \log n + O(1)$.
- Around r_0 the decay of $E(n,r)$ is polynomial: $E(n, r_0 + \delta) \sim n^{-\delta}$ and $E(n, r_0 - \delta) \sim n^{\delta}$
- For all $1.01 \log_{\frac{1}{p}} n < r < 2.99 \log_{\frac{1}{p}} n$ we have:

$$\mathrm{var}(k_r(G)) \leq \Delta \sim E(n,r)^2 (\frac{1}{n^2} + \frac{1}{E(n,r)})$$

Let $\omega(G)$ denote the size of the largest clique in G (the clique number). For $r = r_0 + \delta$,

$$\mathbb{P}(\omega(G) \geq r) = \mathbb{P}(k_r(G) \geq 1) \leq E(n,r) \leq \frac{1}{n^{\delta}}$$

For $r = r_0 - \delta$,

$$\mathbb{P}(\omega(G) < r) = \mathbb{P}(k_r = 0) \leq \frac{\mathrm{var}(k_r)}{E(n,r)^2} \sim \frac{1}{n^{\min(\delta,2)}}$$

Therefore, we conclude that w.h.p, $\omega(G) = r_0 \sim 2 \log_{\frac{1}{p}} n$.

Lemma 7. *$\chi(G)\omega(\bar{G}) \geq n$, where $\bar{G}$ is the complement of graph G.*

Proof. For any valid coloring of G, vertices of the same color must form a clique in $\bar{G}$. Thus $\chi(G)\omega(\bar{G}) \geq n$. ☐

Lemma 8. *If r_1 is such that $E(n, r_1) \sim n^\alpha$ for $\alpha \in (\frac{1}{2}, 2)$ then*

$$\mathbb{P}(k_r(G) = 0) \leq e^{-(1+o(1))\frac{E(n,r_1)^2}{2n^2}} = e^{-O(n^{2\alpha-2})}.$$

Note: this a bound one would get had $k_r(G)$ been 1-Lipschitz with respect to $\frac{n^2}{2}$ edge variables.

Proof. The function $k_r(G)$ has a very large Lipschitz constant in edge variables, so McDiarmid method does not apply directly. A clever workaround of Bollobas was to consider a different random variable U_r equal to the number of r-cliques that do not share edge with any other r-clique. $U_r(G)$ is evidently 1-Lipschitz in edge-variables and also $k_r(G) \geq U_r(G)$. Furthermore, we have

$$\begin{aligned} E(n, r_1) \geq \mathbb{E}(U_{r_1}) &\geq \binom{n}{r_1} p^{\binom{n}{r_1}} - \sum_{l=2}^{r_1} \binom{r_1}{l} \binom{n}{r_1} \binom{n-r_1}{r_1-l} p^{2\binom{r_1}{2} - \binom{l_1}{2}} \\ &\gtrsim E(n, r_1) - n^{2\alpha-2} r_1^4 \\ &= E(n, r_1)(1 - n^{\alpha-2} r_1^4) \\ &= E(n, r_1)(1 + o(1)) \end{aligned}$$

where we lower-bounded $\mathbb{E}[U_{r_1}]$ by choosing all cliques and removing all possible overlapping r-cliques. We omitted some tedious Stirling-based computations. Similarly to (2) the Lemma follows after noticing $\mathbb{P}(k_{r_1}(G) = 0) = \mathbb{P}(U_{r_1}(G) = 0)$. ☐

We proceed with the proof of theorem 6

Proof.

1. Lower bound : When $G \sim G(n, p)$ we have $\bar{G} \sim G(n, 1-p)$ and hence
$$\chi(G) \geq \frac{n}{\omega(\bar{G})} = \frac{n}{\omega(\bar{G})} \overset{\text{w.h.p}}{=} \frac{n}{2 \log_{\frac{1}{1-p}} n}$$

2. Upper bound: Fix n_1 and r_1 such that $n_1 = \frac{n}{\log n^2}$ and $E(n_1, r_1) = (1 + o(1))n^{\frac{5}{3}}$. Then we have $r_1 = 2\frac{\log n}{\log \frac{1}{1-p}} + O(\log\log n)$. By the Lemma any n_1-subset of vertices contains an independent set of size r_1 with probability greater than $1 - e^{-\tilde{O}(n^{\frac{4}{3}})}$ and since there are at most 2^n

subsets of size n_1, by the union bound w.h.p. every n_1-subset has r_1-independent set.

Then proceed as follows. Pick an r_1-sized independent set and color it with a new color. Subtract it from the set of vertices. Proceed until no r_1-sized independent set is found. We must have left with fewer than $n_1 \ll \frac{n}{\log n}$ vertices and so can color those each with a unique color. Thus, we proved that w.h.p. $\chi(G) \leq \frac{n}{r_1}$. This concludes our proof.

$\square$

Remark: Bollobas notes that after these tight concentration results and interesting question is to prove anti-concentration results. E.g. at present there is no proof that $\chi(G(n,p))$ is not w.h.p. in some $O(1)$-interval (the natural high-probability range should be $O(\sqrt{n})$).