Lecture 7: Proof of Azuma's inequality, applications CSE 525, Spring 2019
Instructor: James R. Lee

1 The Hoeffding-Azuma inequality

Say that a martingale $\{X_i\}$ has L-bounded increments if

$$|X_{i+1} - X_i| \leqslant L$$

for all $i \geqslant 0$. (The preceding inequality is meant to hold with probability 1.)

Theorem 1.1. *For every $L > 0$, if $\{X_i\}$ is a martingale with L-bounded increments, then for every $\lambda > 0$ and $n \geqslant 0$, we have*

$$\mathbb{P}[X_n \geqslant X_0 + \lambda] \leqslant e^{-\frac{\lambda^2}{2L^2 n}}$$
$$\mathbb{P}[X_n \leqslant X_0 - \lambda] \leqslant e^{-\frac{\lambda^2}{2L^2 n}}$$

We will prove this in the next lecture. It's useful to note the following special case of the theorem.

Corollary 1.2. *Suppose that $Z_1, Z_2, \ldots, Z_n$ are independent random variables taking values in the interval $[-L, L]$. Put $Z = Z_1 + \cdots + Z_n$ and $\mu = \mathbb{E}[Z]$. Then for every $\lambda > 0$, we have*

$$\mathbb{P}\left[Z \geqslant \mu + \lambda\right] \leqslant e^{-\lambda^2/(2L^2 n)}$$
$$\mathbb{P}\left[Z \leqslant \mu - \lambda\right] \leqslant e^{-\lambda^2/(2L^2 n)}$$

We will actually prove the following generalization of Theorem 1.1.

Theorem 1.3. *Suppose that $\{X_i\}$ is a sequence of random variables satisfying the property that for every subset of distinct indices $i_1 < i_2 < \cdots < i_k$, we have*

$$\mathbb{E}\left[X_{i_1} X_{i_2} \cdots X_{i_k}\right] = 0\,.$$

Then for every $\lambda > 0$ and $n \geqslant 1$, it holds that

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i \geqslant \lambda\right] \leqslant \exp\left(-\frac{\lambda^2}{2\sum_{i=1}^{n} \|X_i\|_\infty^2}\right).$$

Here, $\|X_i\|_\infty$ is the essential supremum of X_i, i.e. the least value L such that $|X_i| \leqslant L$ with probability one.

The reason Theorem 1.3 proves Theorem 1.1 is as follows: Suppose that $\{Z_i\}$ is a martingale with respect to the sequence of random variables $\{Y_i\}$, and let $X_i = Z_i - Z_{i-1}$. Consider distinct indices $i_1 < i_2 < \cdots < i_k$. Then:

$$\mathbb{E}[X_{i_1} \cdots X_{i_k}] = \mathbb{E}\left[X_{i_1} \cdots X_{i_{k-1}}\, \mathbb{E}\left[Z_{i_k} - Z_{i_k - 1} \mid Y_0, \ldots, Y_{i_k - 1}\right]\right] = 0\,,$$

where the final inequality follows from defining property of a martingale.

Proof of Theorem 1.3. Note that from our assumptions, we have that for any sequences of constants $\{a_i\}$ and $\{b_i\}$, we have

$$\mathbb{E}\left[\prod_{i=1}^{n}(a_i + b_i X_i)\right] = \prod_{i=1}^{n} a_i\,. \tag{1.1}$$

Also, observe that for any a, the functions $f(x) = e^{ax}$ is convex. Thus for $x \in [-1,1]$, it lies below the line connecting e^{-a} to e^{a}. In other words, for $x \in [-1,1]$,

$$e^{ax} \leqslant \frac{e^{a}+e^{-a}}{2} + x\frac{e^{a}-e^{-a}}{2} = \cosh(a) + x\sinh(a)\,.$$

Combining this with (1.1), we have for any t:

$$\mathbb{E}\left[e^{t\sum_{i=1}^{n} X_i}\right] \leqslant \mathbb{E}\left[\prod_{i=1}^{n}\cosh(t\|X_i\|_\infty) + \frac{X_i}{\|X_i\|_\infty}\sinh(t\|X_i\|_\infty)\right] = \prod_{i=1}^{n}\cosh(t\|X_i\|_\infty) \leqslant e^{t^2\|X_i\|_\infty^2/2}\,,$$

where the final inequality follows from $\cosh(x) = \sum \frac{x^{2k}}{(2k)!} \leqslant \sum \frac{x^{2k}}{2^k k!} = e^{x^2/2}$.

Now we are in position to apply the method of Laplace transforms:

$$\mathbb{P}\left[\sum_{i=1}^{n} X_i > \lambda\right] \leqslant \frac{\mathbb{E}[e^{t\sum_{i=1}^{n} X_i}]}{e^{t\lambda}} \leqslant e^{(t^2/2)\sum_{i=1}^{n}\|X_i\|_\infty^2 - t\lambda}\,.$$

Setting $t = \frac{\lambda}{\sum_{i=1}^{n}\|X_i\|_\infty^2}$ finishes the proof. □

2 Some applications

2.1 Concentration in product spaces

Define $\mathcal{U} = \{1,2,\ldots,6\}^n$. Define the *hamming distance* between $x, y \in \mathcal{U}$ by

$$H(x,y) := \#\{i \in [n] : x_i \neq y_i\},$$

and if $A \subseteq \mathcal{U}$, define $H(x,A) := \min\{H(x,y) : y \in A\}$. The following theorem shows that $\mathcal{U}$ exhibits "concentration of measure." Starting with any sufficiently large set $A \subseteq \mathcal{U}$, most of the points in $\mathcal{U}$ will be very close to A (the distance to A will be much smaller than the diameter of $\mathcal{U}$).

Theorem 2.1. *Consider any subset $A \subseteq \mathcal{U}$ with $|A| \geqslant 6^{n-1}$. Then for any $c > 0$,*

$$\frac{\left|\left\{x \in \mathcal{U} : H(x,A) \leqslant (c+2)\sqrt{n}\right\}\right|}{6^n} \geqslant 1 - e^{-c^2/2}. \tag{2.1}$$

Proof. Let $Z = (Z_1,\ldots,Z_n) \in \mathcal{U}$ be a uniformly random point. Define the Doob martingale $X_i = \mathbb{E}[H(Z,A) \mid Z_1,\ldots,Z_i]$. Since the map $x \mapsto H(x,A)$ is 1-Lipschitz, we know that $|X_i - X_{i-1}| \leqslant 1$ for every $i = 1,2,\ldots,n$. Thus if $\mu = \mathbb{E}[H(Z,A)]$, Azuma's inequality yields

$$\mathbb{P}[H(Z,A) \leqslant \mu - c\sqrt{n}] \leqslant e^{-c^2/2}$$
$$\mathbb{P}[H(Z,A) \geqslant \mu + c\sqrt{n}] \leqslant e^{-c^2/2}.$$

It is not immediately obvious how to calculate μ, but we can get a good bound using concentration. If $\mu < 2\sqrt{n}$, then the first inequality gives

$$\mathbb{P}[H(Z,A) = 0] \leqslant e^{-2^2/2} = e^{-2} < 1/6,$$

but we know that $\mathbb{P}[Z = 0] = |A|/6^n \geqslant 1/6$, thus $\mu \geqslant 2\sqrt{n}$. Now apply the second inequality, yielding

$$\mathbb{P}[H(Z, A) \geqslant 2\sqrt{n} + c\sqrt{n}] \leqslant e^{-c^2/2}.$$

This is precisely our goal (2.1). □

2.2 Tighter concentration of the chromatic number

Previously, using the vertex exposure martingale we were able to prove reasonable concentration for $\chi(G)$ when $G \sim \mathcal{G}_{n,p}$. In what follows, we will put $p = n^{-\alpha}$ for some $\alpha > 0$. We will show that, surprisingly, if $\alpha > 5/6$, then with probability tending to one, $\chi(G)$ is concentrated on one of four values. In what follows, we will say that an event $\mathcal{E}_n$ (explicitly or implicitly indexed by n) holds "with high probability" if $\mathbb{P}(\mathcal{E}_n) \to 1$ as $n \to \infty$.

Lemma 2.2. *For any $c > 0$ and $\alpha > 5/6$, the following holds for $G \sim \mathcal{G}_{n,p}$: With high probability, every induced subgraph of size at most $c\sqrt{n}$ is 3-colorable.*

Proof sketch. Let S be the smallest subset of $V(G)$ that is not 3-colorable (if no such set exists, we are done). Then every $x \in S$ must have at least three neighbors in S, otherwise since $S \setminus \{x\}$ is 3-colorable, it would be the case that S is also 3-colorable. Thus the number of edges in the induced subgraph $G[S]$ is at least $3|S|/2$.

But it is unlikely that any set S with $|S| \leqslant c\sqrt{n}$ has at least $3|S|/2$ edges inside it. To see this, let $t = |S|$, and we'll compute the probability for a fixed set S: It's at most

$$p^{3t/2}\binom{\binom{t}{2}}{3t/2} \leqslant p^{3t/2}O(t)^{3t/2}.$$

Now we take a union bound over all sets of size at most T:

$$\sum_{t \leqslant T} p^{3t/2}O(t)^{3t/2}\binom{n}{t} \leqslant O(pT)^{3T/2}O\left(\frac{n}{T}\right)^T.$$

The latter inequality holds as long as $T \ll n$. Now using $p = n^{-\alpha}$ and $T \leqslant c\sqrt{n}$, this is bounded by

$$O(n)^{(1/2-\alpha)3T/2}O(n)^{T/2},$$

and the latter quantity is $o(1)$ as long as $3/2(1/2 - \alpha) < 1/2$, i.e. $\alpha > 5/6$. □

Theorem 2.3. *With high probability, $\chi(G)$ takes one of four different values.*

Proof. Fix a number $\varepsilon > 0$ that we will send to 0. Let $u = u(n, p, \varepsilon)$ be the smallest integer so that $\mathbb{P}[\chi(G) \leqslant u] > \varepsilon$. Observe that, by the choice of u, we have $\mathbb{P}[\chi(G) > u - 1] \geqslant 1 - \varepsilon$.

Let $Y = Y(G)$ be the minimal size of a set of vertices S such that $\chi(G \setminus S) \leqslant u$. Consider the vertex exposure martingale for $G \sim \mathcal{G}_{n,p}$. Note that Y is 1-Lipschitz with respect to the exposure process because we could always add the modified vertex to S. Thus we can apply Azuma's inequality to the corresponding Doob martingale to conclude that

$$\mathbb{P}[Y \geqslant \mu + \lambda\sqrt{n}] \leqslant e^{-\lambda^2/2} \tag{2.2}$$

$$\mathbb{P}[Y \leqslant \mu - \lambda\sqrt{n}] \leqslant e^{-\lambda^2/2}, \tag{2.3}$$

where $\mu = \mathbb{E}[Y]$.

Let us choose λ so that $e^{-\lambda^2/2} = \varepsilon$. By the definition of u, we have $\mathbb{P}[Y = 0] > \varepsilon$. We conclude from (2.3) that $\mu \leqslant \lambda\sqrt{n}$. Now using (2.2), we see that $\mathbb{P}[Y \geqslant 2\lambda\sqrt{n}] \leqslant \varepsilon$.

By Lemma 2.2, we may assume that every subset of size at most $2\lambda\sqrt{n}$ is 3-colorable by throwing away an ε-fraction of graphs. Now observe that $Y < 2\lambda\sqrt{n}$ implies that G is $u + 3$ colorable since $G \setminus S$ is u-colorable and $|S| < 2\lambda\sqrt{n}$ so S can be colored with an additional 3 colors. We conclude that

$$\mathbb{P}[\chi(G) \in \{u, u+1, u+2, u+3\}] \geqslant 1 - 3\varepsilon \,.$$

Sending $\varepsilon \to 0$ completes the proof. □