Mathematical Toolkit **Autumn 2015**

Lecture 15: November 18, 2015

Lecturer: Madhur Tulsiani

1 Azuma's inequality

We now prove a concentration inequality for martingale sequences.

Proposition 1.1 (Azuma's inequality) *Let $Z_0, \ldots, Z_n$ be a martingale sequence with respect to the filter $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}_n$ such that for $Y_i = Z_i - Z_{i-1}$, we have that for all $i \in [n]$, $|Y_i| = |Z_i - Z_{i-1}| \leq c_i$. Then,*

$$\mathbb{P}[Z_n - Z_0 \geq t] \;\leq\; \exp\left(-\frac{t^2}{2 \cdot \sum_{i=1}^n c_i^2}\right) \quad \textit{and} \quad \mathbb{P}[Z_0 - Z_n \geq t] \;\leq\; \exp\left(-\frac{t^2}{2 \cdot \sum_{i=1}^n c_i^2}\right).$$

Proof: We first prove one side of the inequality. We get that for any $\lambda > 0$

$$\mathbb{P}[Z_n - Z_0 \geq t] \;=\; \mathbb{P}\left[e^{\lambda \cdot (Z_n - Z_0)} \geq e^{\lambda \cdot t}\right] \;\leq\; e^{-\lambda \cdot t} \cdot \mathbb{E}\left[e^{\lambda \cdot (Z_n - Z_0)}\right],$$

using Markov's inequality. Splitting the term in the exponent and conditioning on $\mathcal{F}_{n-1}$, we get

$$\begin{aligned} \mathbb{E}\left[e^{\lambda \cdot (Z_n - Z_0)}\right] \;&=\; \mathbb{E}\left[e^{\lambda \cdot (Y_n + Z_{n-1} - Z_0)}\right] \;=\; \mathbb{E}\left[\mathbb{E}\left[e^{\lambda \cdot (Y_n + Z_{n-1} - Z_0)} \mid \mathcal{F}_{n-1}\right]\right] \\ &=\; \mathbb{E}\left[e^{\lambda \cdot (Z_{n-1} - Z_0)} \cdot \mathbb{E}\left[e^{\lambda \cdot Y_n} \mid \mathcal{F}_{n-1}\right]\right], \end{aligned}$$

using the fact that Z_{n-1} and Z_0 are both measurable in the σ-algebra $\mathcal{F}_{n-1}$. We not bound the expectation $\mathbb{E}\left[e^{\lambda \cdot Y_n} \mid \mathcal{F}_{n-1}\right]$ using convexity of the function e^x. Let $\alpha \in [-1, 1]$ and $M \in \mathbb{R}$ be any real number. Then, we have

$$\alpha \cdot M \;=\; \left(\frac{1+\alpha}{2}\right) \cdot M + \left(\frac{1-\alpha}{2}\right) \cdot M.$$

Thus, using convexity of the function e^x, we get that

$$e^{\alpha \cdot M} \;\leq\; \left(\frac{1+\alpha}{2}\right) \cdot e^M + \left(\frac{1-\alpha}{2}\right) \cdot e^{-M}.$$

Taking $\alpha = Y_n / c_n$ and $M = \lambda \cdot c_n$, we get that

$$e^{\lambda \cdot Y_n} \;\leq\; \left(\frac{1 + (Y_n/c_n)}{2}\right) \cdot e^{\lambda \cdot c_n} + \left(\frac{1 - (Y_n/c_n)}{2}\right) \cdot e^{-\lambda \cdot c_n}.$$

Using the fact that $\mathbb{E}[Y_n \mid \mathcal{F}_{n-1}] = 0$, we get that

$$\begin{aligned}\mathbb{E}\left[e^{\lambda\cdot Y_n} \mid \mathcal{F}_{n-1}\right] &\leq \mathbb{E}\left[\left(\frac{1+(Y_n/c_n)}{2}\right)\cdot e^{\lambda\cdot c_n} + \left(\frac{1-(Y_n/c_n)}{2}\right)\cdot e^{-\lambda\cdot c_n} \mid \mathcal{F}_{n-1}\right] \\ &= \frac{e^{\lambda\cdot c_n}+e^{-\lambda\cdot c_n}}{2} \leq e^{(\lambda\cdot c_n)^2/2},\end{aligned}$$

where the last inequality uses the fact that $(e^x + e^{-x})/2 \leq e^{x^2/2}$ (which can be verified using the Taylor expansion). Thus, we get

$$\mathbb{P}[Z_n - Z_0 \geq t] \leq e^{-\lambda\cdot t}\cdot e^{(\lambda^2/2)\cdot c_n^2}\cdot \mathbb{E}\left[^{\lambda\cdot(Z_{n-1}-Z_0)}\right].$$

Continuing by induction, we can deduce

$$\mathbb{P}[Z_n - Z_0 \geq t] \leq \exp\left(-\lambda\cdot t + (\lambda^2/2)\cdot\sum_{i=1}^{n} c_i^2\right)$$

Since the above holds for any $\lambda > 0$, we can optimize over λ to minimize the above bound. Check that the above expression is minimized for $\lambda = \frac{t}{\sum_{i=1}^n c_i^2}$, which gives

$$\mathbb{P}[Z_n - Z_0 \geq t] \leq \exp\left(-\frac{t^2}{2\cdot\sum_{i=1}^n c_i^2}\right).$$

The bound for $\mathbb{P}[Z_0 - Z_n \geq t]$ follows similarly. ∎

2 Applications to large-deviation bounds

Many applications of Azuma's require bounding the deviation from mean of a function $f(x_1, \ldots, x_n)$ whose inputs are chosen at random. Taking X_i as a random variable, whose value is thought of at the input x_i, we are thus interested in understanding the random variable $f(X_1, \ldots, X_n)$. We define the Doob martingale sequence

$$Z_i = \mathbb{E}[f \mid X_1, \ldots, X_i].$$

A particularly useful case is when f does not change significantly by the change of any one input variable. A function f is said to be Lipschitz in the i^{th} coordinate with Lipschitz constant c_i if for all sets of inputs $x_1, \ldots, x_n$, and all $x_i \neq x_i'$, we have

$$\left|f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)\right| \leq c_i.$$

Prove that this implies that for the above martingale sequence, we have $|Z_i - Z_{i-1}| \leq c_i$. In fact, in most applications all c_is are equal but occasionally one needs different Lipschitz constants for each coordinate.

2.1 Balls and Bins

Consider throwing m balls independently at random in n bins. Let X_i denote the index of the bin in which the i^{th} ball lands. Let A be a random variable denoting the number of empty of empty bins. Then, we have $A = f(X_1, \ldots, X_m)$ since we can compute the number of empty given the information about all the balls.

We again define the Doob martingale

$$Z_i = \mathbb{E}\left[f(X_1, \ldots, X_i)\right] .$$

Since changing any X_i (changing the bin in which the i^{th} ball was thrown) can only change the number of empty bins by at most one, we get by the above discussion that for all $i \in [m]$,

$$|Z_i - Z_{i-1}| \;\leq\; 1 .$$

Thus, by Azuma's inequality we get that

$$\mathbb{P}\left[|Z_n - Z_0| \geq t\right] \;\leq\; 2 \cdot \exp\left(-t^2/(2m)\right) .$$

Note that $Z_0 = \mathbb{E}\left[f\right]$ and hence we get that

$$\mathbb{P}\left[|f - \mathbb{E}\left[f\right]| \geq t\right] \;\leq\; 2 \cdot \exp\left(-t^2/(2m)\right) .$$

Thus, with high probability, the number of empty bins is within $O(\sqrt{m})$ of its expectation.

Exercise 2.1 *Check that* $\mathbb{E}\left[f\right] = n \cdot \left(1 - \frac{1}{n}\right)^m$.

2.2 3-SAT

Consider a 3-SAT formula with n variables $x_1, \ldots, x_n$, and m clauses. Let each variable be contained in at most k clauses. Consider assigning each of the variables to be 0 or 1 independently with probability 1/2 each. Let

$$Z \;=\; f(x_1, \ldots, x_n) \;=\; \text{number of clauses satisfied} .$$

As before, we define the martingale sequence

$$Z_i \;=\; \mathbb{E}\left[Z \mid X_1, \ldots, X_i\right] ,$$

and note that changing the value of any variable can change the number of satisfied clauses by at most k. Thus, we get by Azuma's inequality that

$$\mathbb{P}\left[|Z - \mathbb{E}\left[Z\right]| \geq t\right] \;\leq\; 2 \cdot \exp\left(-t^2/(2k^2 n)\right) .$$

As we saw earlier, $\mathbb{E}\left[7m/8\right]$ and hence we get that with high probability Z is within $O(k\sqrt{n})$ of $7m/8$.

2.3 Chromatic number of a random graph

Definition 2.2 *Let $G = (V, E)$ be an undiredcted graph and let $k \in \mathbb{N}$. A* valid k-coloring *of G is map $g : V \to [k]$, which assigns one of k colors to each vertex, and satisfies*

$$g(i) \neq g(j) \quad \forall \{i, j\} \in E\,.$$

For a graph G, the chromatic number *of G, denoted by $\chi(G)$, is defined to be the least non-negative integer k such that G has a valid k-coloring.*

Recall that $\mathcal{G}_{n,p}$ denotes a distribution on graphs where we pick each of the $\binom{n}{2}$ pairs to be in E independently with probability p. Thus, the graph is defined by the random variables $\left\{X_{\{i,j\}}\right\}_{i \neq j}$, where each $X_{\{i,j\}} = 1$ independently with probability p and is 0 otherwise. Let X_i denote the collection $\left\{X_{\{i,j\}}\right\}_{j>i}$ which containts information about all the neighbors of the vertex i in the graph generate as above. Let $Z = \chi(G)$ be the random variable denoting the chromatic number of the random graph generated generated as above. Note that

$$Z = f(X_1, \ldots, X_n)\,,$$

and changing any X_i changes Z by at most 1. Thus, we have $|Z_i - Z_{i-1}| \leq 1$ for the martingale sequence $Z_i = \mathbb{E}\left[f \mid X_1, \ldots, X_i\right]$. Again, Azuma's inequality gives

$$\mathbb{P}\left[|Z - \mathbb{E}\left[Z\right]| \geq t\right] \;\leq\; 2 \cdot \exp\left(-t^2/(2n)\right)\,.$$

However, unlike the previous examples, computing $\mathbb{E}\left[Z\right]$ here is nontrivial. A more subtle use of martingales by Bollobás [Bol88] shows that $\chi(G) = (1 + o(1)) \cdot \frac{n}{2\log_{(1-1/p)} n}$ with high probability.

Note that the martingale used here revealed information about one vertex of the graph at a time. We could also have revealed just one $X_{\{i,j\}}$ at a time and the function $\chi(G)$ would still be Lipschitz, but the martingale sequence would have $\binom{n}{2}$ terms and the concentration bound would have been much weaker. The martingale sequence of the form used above is known as a vertex exposure martingale.

References

[Bol88] Béla Bollobás, *The chromatic number of random graphs*, Combinatorica **8** (1988), no. 1, 49–55.