Martingale concentration and the chromatic number of $\mathcal{G}(n, n^{-\alpha})$

Mriganka Basu Roy Chowdhury

Berkeley Student Probability Seminar

January 31, 2024

Contents

1 Doob's improvement on Markov

Recall that a martingale is a sequence of random variables X_n along with a *filtration* $\mathcal{F}_n$*, which remembers its past via the (conditional) mean*:

$$\mathbb{E}\left[X_{n+1} \middle| \mathcal{F}_n\right] = X_n, \quad \forall n \geq 1.$$

A submartingale is a variant of this which is biased upwards:

$$\mathbb{E}\left[X_{n+1} \middle| \mathcal{F}_n\right] \geq X_n, \quad \forall n \geq 1.$$

For submartingales, Doob proved the following beautiful inequality, which says that essentially the burden of deviation falls on the last element in the sequence. Formally, if $\{X_n\}$ is a nonegative submartingale

$$\mathbb{P}\left(X_n^* \geq a\right) \leq \frac{\mathbb{E}X_n}{a}, \quad \forall a > 0,$$

where $X_n^* = \sup_{j \leq n} X_j$ is the *record* at time n. Contrast this to vanilla Markov, which will show that

$$\mathbb{P}\left(X_k \geq a\right) \leq \frac{\mathbb{E}\left[X_k\right]}{a} \leq \frac{\mathbb{E}\left[X_n\right]}{a}, \quad \forall k \leq n,$$

so that

$$\sup_{k\le n} \mathbb{P}\left(X_k \ge a\right) \le \frac{\mathbb{E}\left[X_n\right]}{a}.$$

The proof of this fact uses the observation that, not only at fixed times t, but at arbitrary stopping times σ, the "projected growth" in the future is only positive, i.e.,

$$\mathbb{E}\left[X_\sigma\right] \le \mathbb{E}\left[X_n\right], \quad \sigma \text{ is a stopping time always} \le n.$$

Of course, finding the maximum is not a stopping time! But, we don't need to. To check if the supremum is bigger than a, we can just stop as soon as some element is bigger than a, which will produce a stopping time:

$$\sigma \stackrel{\text{def}}{=} \inf\{k \le n : X_k \ge a\},$$

with edge cases handled appropriately.

2 Exponentials of martingales

If X_n is a martingale, then $Y_n \stackrel{\text{def}}{=} \exp\left(X_n\right)$ is a submartingale. To see define the *martingale difference sequence* $D_n \stackrel{\text{def}}{=} X_n - X_{n-1}$ for $n \ge 1$, and observe

$$\begin{aligned}
\mathbb{E}\left[Y_{n+1}|\mathcal{F}_n\right] &= \mathbb{E}\left[\exp\left(X_n + D_{n+1}\right)|\mathcal{F}_n\right] \\
&= \exp\left(X_n\right) \cdot \mathbb{E}\left[\exp\left(D_{n+1}\right)|\mathcal{F}_n\right] \\
&\stackrel{\text{Jensen}}{\ge} Y_n \cdot \exp\left(\mathbb{E}\left[D_{n+1}|\mathcal{F}_n\right]\right) = Y_n,
\end{aligned}$$

since $\mathbb{E}\left[D_{n+1}|\mathcal{F}_n\right] = \mathbb{E}\left[X_{n+1} - X_n|\mathcal{F}_n\right] = 0$. In fact, exponential is not special. This holds for any convex ϕ.

We have that exponentials of martingales are submartingales, and have an improvement over Markov for submartingales. This should motivate us to check if Chernoff can be improved. Indeed, it can, and the result is called *Azuma-Hoeffding:*

Theorem 1 (Azuma-Hoeffding)**.** *Let X_n be a martingale w.r.t the filtration $\mathcal{F}_n$, and let D_n be the difference sequence. Assume that D_n is sub-Gaussian with constant c_n^2, conditioned on $\mathcal{F}_{n-1}$, i.e.,*

$$\mathbb{E}\left[\exp(\lambda(D_n - \mathbb{E}D_n))\middle|\mathcal{F}_{n-1}\right] \le \exp\left(\lambda^2 c_n^2/2\right).$$

Then, the supremum of the martingale $X_n - X_0$ is also sub-Gaussian with constant $\sum_{k=1}^n c_k^2$, in particular, has the tail

$$\mathbb{P}\left((X_\cdot - X_0)^* \ge a\right) \le \exp\left(-\frac{a^2}{2\sum_{k=1}^n c_k^2}\right).$$

In the most common version of this statement: $D_n \in [A_n, B_n]$ with $B_n - A_n \le \ell_n$ where B_n, A_n are predictable processes. The maximal sub-Gaussian constant of such a random variable is $\ell_n^2/4$, yielding

$$\mathbb{P}\left((X_\cdot - X_0)^* \ge a\right) \le \exp\left(-2 \cdot \frac{a^2}{\sum_{k=1}^n \ell_k^2}\right).$$

3 What if I don't care about martingales

But almost surely, you care about (complicated) functions F of many independent random variables $\mathbf{X} = (X_1, \ldots, X_n)$. In such cases, you can define the *Doob martingale* (which is a "free"-martingale):

$$Z_n \stackrel{\text{def}}{=} \mathbb{E}\left[F \;\middle|\; X_1, \ldots, X_n\right], \quad n \geq 1.$$

This is a martingale, and we can try to apply Azuma-Hoeffding. A condition on F is needed to satisfy A-H though. We impose

$$\Delta_i \stackrel{\text{def}}{=} \sup_{\mathbf{X}} \sup_{y} \left[F(\mathbf{X}) - F(\mathbf{X}')\right] < \infty, \quad \mathbf{X}' = \mathbf{X} \text{ except } \mathbf{X}'_i = y,$$

that is, the maximum amount $F(\mathbf{X})$ can change by changing the ith coordinate.

This assumption allows us to bound the difference sequence of the Doob martingale, since if we replace $\mathbf{X}_k$ by an independent sample from the distribution of $\mathbf{X}_k$ and call this new entity $\mathbf{X}'$,

$$\begin{aligned} Z_k &= \mathbb{E}\left[\mathbf{X} | \mathcal{F}_k\right], \\ Z_{k-1} &= \mathbb{E}\left[\mathbf{X}' | \mathcal{F}_k\right]. \end{aligned}$$

The rest is not difficult, yielding the McDiarmid inequality:

Theorem 2 (McDiarmid inequality)**.** *Let* $\mathbf{X} = (X_1, \ldots, X_n)$ *be independent, and let* Δ_i *defined as above are finite. Then*

$$\mathbb{P}\left(F - \mathbb{E}F \geq a\right) \leq \exp\left(-2 \cdot \frac{a^2}{\sum_{i \leq n} \Delta_i^2}\right).$$

Replacing F *by* $-F$ *and union-bounding we get*

$$\mathbb{P}\left(|F - \mathbb{E}F| \geq a\right) \leq 2\exp\left(-2 \cdot \frac{a^2}{\sum_{i \leq n} \Delta_i^2}\right).$$

Remark: There is no restriction on the spaces in which X_i take values. They don't need to be i.i.d. either.

Side note: For the "resampling" Markov chain, McDiarmid shows concentration of the stationary measure. The related Poincare-type bounds for such chains are also available.

4 Chromatic numbers via vertex revealing

Given an undirected graph G, its chromatic number is the *smallest* number of colors required to color its vertices with no two adjacent vertices having the same color. For example, the complete graph K_n has chromatic number n, and the path P_n has chromatic number 2 unless $n = 1$. Our goal is to understand the chromatic number of Erdős–Rényi graphs $\mathcal{G}(n, p)$ for some particular choices of p.

Firstly, think of the chromatic number N as $N(\{X_e\}_e)$ as a function of the boolean variables given by the presence/absence of edges e. Changing the status of some edge cannot change the chromatic number by more than 1. Clearly, if we add an edge, it can increase N by at most one since we can just use a new color. Conversely, if we drop an edge and chromatic number drops by more than 1, we can reverse the change to obtain a contradiction.

Thus

$$\mathbb{P}\left(|N - \mathbb{E}N| \geq a\right) \leq 2\exp\left(\approx -4a^2/n^2\right),$$

a horrible bound since this says that the fluctuation order is at most n, which is trivial, since the chromatic number of bounded between 0 and n.

We can do much better if we realize that the logic above holds even if we change the status of *any vertex*! That is, even if we change the status of all edges adjacent to a given vertex v, the chromatic number changes by at most 1. Observe that our variables X_v are now $X_v = (X_e : e = (v, w), w \leq v)$ (assuming some order on the vertices), which lie in different spaces – but that does not harm us, yielding the bound

$$\mathbb{P}\left(|N - \mathbb{E}N| \geq a\right) \leq 2\exp\left(-2a^2/(n-1)\right)$$

since $\Delta_1 = 0$, showing a fluctuation order of $n^{1/2}$, much better!

5 Chromatic numbers exhibit constant order fluctuation

Much more is known about chromatic numbers. When $\alpha > 1/2$, N is in fact concentrated on two values! We however will only show

Theorem 3. *Let $\alpha > 5/6$, and let $G \sim \mathcal{G}(n, p_n \stackrel{def}{=} n^{-\alpha})$. Then for any $\varepsilon > 0$, there is a number $\phi_n = \phi_n(\alpha, \varepsilon)$ such that*

$$\mathbb{P}\left(N \in [\phi_n, \phi_n + 3]\right) \geq 1 - \varepsilon.$$

Proof. The proof proceeds via showing that there is a choice of ϕ_n such that with high probability,

- there is a $O_\varepsilon(\sqrt{n})$ size subset of vertices, except which the whole graph is ϕ_n colorable,
- and, every subset of this size is 3-colorable,

combining which we get our result.

We choose ϕ_n to be the smallest integer such that

$$\mathbb{P}\left(G \sim \mathcal{G}(n, p_n) \text{ is } \phi_n \text{ colorable}\right) > \varepsilon/3.$$

Step 1: Let $F = F(\{X_v\}_v)$ be the function of the vertex revealing variables X_v defined above, counting the smallest size of a subset whose removal ensures that G is ϕ_n colorable. For example, $\mathbb{P}(F = 0) > \varepsilon/3$, by assumption above. This fact will be used later. We now claim: *F satisfies bounded differences with $\Delta_v = 1$.*

To see why, fix v and observe that (1) F is monotone in the number of edges adjacent to v, and (2) between the cases where there are no edges adjacent to v versus every edge adjacent to v, the gap in F can be at most 1 since we can drop v itself from the graph.

This enables us to apply McDiarmid's inequality to assert:

$$\mathbb{P}\left(F - \mathbb{E}F \leq -b\sqrt{n-1}\right) \leq \exp\left(-2b^2\right) = \varepsilon/3,$$

where $b = b(\varepsilon)$ is chosen so that $\exp\left(-2b^2\right) = \varepsilon/3$. We will now use these facts to bound $\mathbb{E}F$, which will then allow us to derive a tail on F without the $\mathbb{E}F$ additive term.

Observe that if $\mathbb{E}F > b\sqrt{n-1}$, then

$$\mathbb{P}(F=0) \leq \mathbb{P}\left(F - \mathbb{E}F < -b\sqrt{n-1}\right) \leq \varepsilon/3,$$

but we saw earlier that $\mathbb{P}(F=0) > \varepsilon/3$. So $\mathbb{E}F \leq b\sqrt{n-1}$.

Using McDiarmid in the opposite direction then yields

$$\mathbb{P}\left(F \geq 2b\sqrt{n-1}\right) \leq \mathbb{P}\left(F - \mathbb{E}F \geq b\sqrt{n-1}\right) \leq \exp\left(-2b^2\right) = \varepsilon/3,$$

finishing the proof of Step 1.

Step 2: This will be a proof by first moment, where we count the number Y of the number of **minimal** vertex subsets of size at most $c\sqrt{n}$ such that the induced subgraph in G is not 3-colorable. The claim is that w.h.p. $Y = 0$.

The requirement is of course not lossy since we want to show that there are *no non-3-colorable "small subsets"*. However, imposing minimality allows us to control the subsets via the following observation: *Let W be a subset of vertices constituting a minimal subset. Then, each vertex in W has degree at least 3 in W.* If not, let v be such a vertex. Dropping v cannot make the graph 3-colorable, since if we add v back in, it will still be 3-colorable as we can just pick the color not adjacent to v (of which there is at least one, since v has degree two at most). Thus W couldn't possibly be minimal non-3-colorable.

Suppose $|W| = \ell$. The above observation then means that the number of edges induced by W satisfies $|E(W)| \geq 3\ell/2$. Thus, the probability that W is a *minimal* subset is at most the probability that it has $3\ell/2$ edges induced in it:

$$\binom{\binom{\ell}{2}}{3\ell/2} p_n^{3\ell/2}.$$

Therefore, a first moment bound yields

$$\begin{aligned}
\mathbb{P}(Y>0) &\leq \mathbb{E}Y \\
&\leq \sum_{\ell=4}^{c\sqrt{n}} \binom{n}{\ell}\binom{\binom{\ell}{2}}{3\ell/2} n^{-3\ell\alpha/2} \\
&\leq \sum_{\ell=4}^{c\sqrt{n}} \left(\frac{en}{\ell}\right)^{\ell}\left(\frac{e\ell^2}{3\ell/2}\right)^{3\ell/2} n^{-3\ell\alpha/2} \\
&= \sum_{\ell=4}^{c\sqrt{n}} \exp\left(\frac{5\ell}{2} + \left(1 - \frac{3\alpha}{2}\right)\ell\log n + 2\ell\log\ell - \frac{3\ell}{2}\log\left(\frac{3\ell}{2}\right)\right) \\
&\leq \sum_{\ell=4}^{c\sqrt{n}} \exp\left(\frac{5\ell}{2} + \left(1 - \frac{3\alpha}{2}\right)\ell\log n + \frac{\ell}{2}\log\ell\right).
\end{aligned}$$

It is tempting to substitute $\ell = c\sqrt{n}$ but we have to be careful since $1 - 3\alpha/2 < 0$. We can substitute $\ell = c\sqrt{n}$ in the $\log\ell$ and not elsewhere:

$$\leq \sum_{\ell=4}^{c\sqrt{n}} \exp\left(c'\ell + \left(1 - \frac{3\alpha}{2} + \frac{1}{4}\right)\ell\log n\right), \quad c' = 5/2 + (1/2)\log c.$$

If $\alpha > 5/6$, $1 - 3\alpha/2 + 1/4 = \delta < 0$, so that the first term is dominant, showing that the sum goes to 0. Let $c = b$ from above, and choose n large enough so that this is smaller than $\varepsilon/3$.

Combining all the results above concludes the proof. $\square$