**CS271 Randomness & Computation** **Spring 2018**

Lecture 20: March 22

*Instructor: Alistair Sinclair*

**Disclaimer**: *These notes have not been subjected to the usual scrutiny accorded to formal publications. They may be distributed outside this class only with the permission of the Instructor.*

In this lecture we continue our exploration of martingales through their application to random graph coloring and random geometric TSP. We shall see how Azuma's Inequality can be used to obtain useful tail bounds for these problems, even though the associated random variables are not independent.

# 20.1 Coloring Random Graphs (continued)

Recall that a (proper) coloring maps each vertex of a given graph $G$ to a color such that no two adjacent vertices share the same color. The *chromatic number* $\chi(G)$ is the minimum number of colors in a proper coloring of $G$, and is NP-complete to compute for general $G$. However, we shall now see that for a *random* graph $G$, $\chi(G)$ has a fairly well-determined value. In the last lecture we proved the following concentration result for $\chi(G)$ for a random graph $G \in \mathcal{G}_{n,p}$.

**Theorem 20.1 (Shamir-Spencer [SS87]).** *If $G \in \mathcal{G}_{n,p}$ then, for any $\lambda > 0$,*

$$\Pr\left[|\chi(G) - \mathrm{E}\left[\chi(\mathrm{G})\right]| \geq \lambda\right] \quad \leq \quad 2\exp\left(-\frac{\lambda^2}{2n}\right).$$

Recall that its proof was based on Azuma's inequality (Theorem 18.3) and did not tell us the value of $\mathrm{E}\left[\chi(\mathrm{G})\right]$. Today we will use a use a slightly more sophisticated martingale argument to calculate $\mathrm{E}\left[\chi(\mathrm{G})\right]$. The main result for the first half of the lecture is the following theorem:

**Theorem 20.2** *For $G \in \mathcal{G}_{n,p}$, we have $\mathrm{E}\left[\chi(\mathrm{G})\right] \sim \frac{\mathrm{n}}{2\log_{1/(1-\mathrm{p})}\mathrm{n}}$.*

Since $\mathrm{E}\left[\chi(\mathrm{G})\right] \gg \sqrt{\mathrm{n}}$, then combining theorems 20.1 and 20.2, we conclude that $\chi(G)$ is tightly concentrated around its expected value.

Note first that a lower bound on $\mathrm{E}\left[\chi(\mathrm{G})\right]$ is immediate from the observation that the vertices sharing the same color must form an independent set of $G$. Independent sets in $G$ correspond to cliques in the complement graph $\overline{G}$ and as we saw in Lecture 8, the size of a maximum clique in $G \in \mathcal{G}_{n,p}$ is $2\log_{1/p} n$ a.a.s. This implies that the size of a maximum independent set is a.a.s. $2\log_{1/(1-p)} n$, so the number of colors must be at least $\frac{n}{2\log_{1/(1-p)} n}$, giving this value as a lower bound on $\mathrm{E}\left[\chi(\mathrm{G})\right]$.

Establishing this as an upper bound is trickier, and was first proved by Bollobás [B88], using martingales in a rather subtle way as we shall see. This greatly improved on the previous result of Grimmett and McDiarmid [GM75] that $\mathrm{E}\left[\chi(\mathrm{G})\right] \lesssim \frac{\mathrm{n}}{\log_{1/(1-\mathrm{p})}\mathrm{n}}$, which is off by a factor of 2.

For simplicity we will prove the upper bound for $p = \frac{1}{2}$ only. In this special case we have the convenience of cliques and independent sets being equivalent under $\mathcal{G}_{n,1/2}$; however, the same arguments carry over to the case of any constant $p$.

We start by recalling some facts from Lecture 8. The expected number of $k$-cliques in $G \in \mathcal{G}_{n,\frac{1}{2}}$ is $g(k) = \binom{n}{k}2^{-\binom{k}{2}}$. Define $k_0(n) = \max_k\{g(k) \geq 1\}$; we know that $k_0(n) \sim 2\log_2 n$. If we now set $k_2(n) = k_0(n) - 3$, then it is not hard to check that $g(k_2(n)) \to \infty$, and indeed more precisely that $g(k_2(n)) = n^{3+o(1)}$. (We shall actually need such a lower bound on $g(k_2(n))$ later.)

The following lemma, which we will prove later, will be crucial in proving Theorem 20.2.

**Lemma 20.3** *Let $G \in \mathcal{G}_{n,\frac{1}{2}}$ then* $\Pr[G$ *contains no independent set of size* $\geq k_2(n)] \leq \exp\left(-n^{2-o(1)}\right)$.

Note that we already know from Lecture 8 that $G$ a.a.s. contains an independent set of size $k_2$. The point about this lemma is that it gives a very sharp upper bound on the probability that this fails to happen.

**Proof of Theorem 20.2:** We have seen that it suffices to prove only the upper bound, and we will restrict attention for simplicity to $p = \frac{1}{2}$. Our aim is to color $G \in \mathcal{G}_{n,1/2}$ with at most $\frac{n}{2\log_2 n}(1 + o(1))$ colors. Let $S$ be an arbitrary subset of vertices of $G$, with $m = |S| = \frac{n}{(\log_2 n)^2}$. Let $G|_S$ denote the restriction of $G$ to $S$. Then $G|_S \in \mathcal{G}_{m,\frac{1}{2}}$ is just a random graph on a slightly smaller vertex set. So by Lemma 20.3, $G|_S$ contains an independent set of size $k_2(m) \sim 2\log_2 m \sim 2\log_2 n$ with probability at least $1 - \exp\left(-m^{2-o(1)}\right) = 1 - \exp\left(-n^{2-o(1)}\right)$.

Applying the union bound over all $\binom{n}{m}$ choices of $S$ implies

$$\begin{aligned}
\Pr[\exists\, S \text{ s.t. } G|_S \text{ contains no independent set of size } k_2(m)] &\leq \binom{n}{m}\exp\left(-n^{2-o(1)}\right) \\
&\leq 2^n \exp\left(-n^{2-o(1)}\right) \\
&= o(1). \qquad (20.1)
\end{aligned}$$

Now consider the following method for coloring $G$:

> **while** $\exists$ more than $m$ uncolored vertices in $G$ **do**
>     pick an arbitrary uncolored subset $S \subseteq V(G)$ of size $m$
>     pick a new color and apply this to a largest independent set in $S$
> color each remaining vertex of $G$ with a different new color

By Equation (20.1), we know that a.a.s. every iteration of the while-loop colors at least $k_2(m)$ vertices. Hence the number of colors used by the above scheme is a.a.s. at most

$$\frac{n}{k_2} + m = \frac{n}{2\log_2 n}(1 + o(1)).$$

This proves the required upper bound on $\chi(G)$. ■

It remains only to go back and prove Lemma 20.3. This is where we will use martingales.

**Proof of Lemma 20.3:** Let $Y$ be the size of a maximal family of edge-disjoint $k_2(n)$-cliques in $G$. Then $Y = 0$ if $G$ contains no cliques of size $\geq k_2(n)$. Also, $Y = f(Z_1, \ldots, Z_{\binom{n}{2}})$ for some function $f$, where $Z_i$ indicates the presence or absence of the $i^{th}$ edge is in the "edge-exposure" process of $G$ (as defined in the previous lecture). Thus $X_i = \mathrm{E}[\mathrm{Y} \mid \mathrm{Z_1, \ldots, Z_i}]$ is a Doob martingale. Furthermore, since flipping one edge $i$ (an indicator $Z_i$) affects at most one element of a maximal family of edge-disjoint $k_2(n)$-cliques of $G$, it follows that $f$ is 1-Lipschitz. Note that it is crucial here that we are talking about *edge-disjoint* cliques; otherwise the function would not necessarily be 1-Lipschitz!

We first make the following claim.

**Claim 20.4**

$$\mathrm{E}\left[\mathrm{Y}\right] \geq \frac{n^2}{2k_2(n)^4}\left(1+o(1)\right)$$

We leave the proof of the claim to the end of this section. Combining the fact that $f$ is 1-Lipschitz with Claim 20.4, and using Azuma's inequality yields:

$$\begin{aligned}
\Pr\left[G \text{ contains no cliques of size} = k_2(n)\right] &\leq \Pr\left[Y=0\right] \\
&= \Pr\left[Y - \mathrm{E}\left[\mathrm{Y}\right] \leq -\mathrm{E}\left[\mathrm{Y}\right]\right] \\
&\leq \exp\left(-\frac{\left(\mathrm{E}\left[\mathrm{Y}\right]\right)^2}{2\binom{n}{2}}\right) \\
&\leq \exp\left(-\frac{n^2}{4k_2(n)^8}\left(1+o(1)\right)\right) \\
&= \exp\left(-n^{2-o(1)}\right).
\end{aligned}$$

■

We now prove the claim using a probabilistic method argument.

**Proof of Claim 20.4:** Let $K_2$ be the set of all $k_2(n)$-cliques in $G$ and let $\mu = \mathrm{E}\left[|\mathrm{K}_2|\right]$. Then as discussed earlier $\mu = g(k_2(n)) = n^{3+o(1)}$. Let $P$ be the set of all pairs of $k_2(n)$-cliques with non-trivial intersection – that is, the set of all pairs of distinct $k_2(n)$-cliques whose intersection contains betwen two and $k_2(n)-1$ vertices. Recall from the second moment calculations in Lecture 8 that

$$\frac{\mathrm{E}\left[|\mathrm{P}|\right]}{\mu^2} \sim \frac{1}{2}\frac{k_2(n)^4}{n^2}.$$

Now let $K'$ be a random subset of $K_2$ obtained by choosing each $k_2(n)$-clique with probability $q$, where $q$ is a parameter that we will optimize. Let $P'$ be the associated set of pairs of cliques from $P$. Then we see that

$$\begin{aligned}
\mathrm{E}\left[|\mathrm{K}'|\right] &= q\mathrm{E}\left[|\mathrm{K}|\right] = \mathrm{q}\mu \\
\mathrm{E}\left[|\mathrm{P}'|\right] &= q^2\mathrm{E}\left[|\mathrm{P}|\right] \sim \mathrm{q}^2\frac{1}{2}\frac{\mathrm{k_2(n)^4}}{\mathrm{n}^2}\mu^2.
\end{aligned}$$

Remove from $K'$ one element of each pair in $P'$. This now gives an edge disjoint family $Y$ of $k_2(n)$-cliques with

$$\begin{aligned}
\mathrm{E}\left[|\mathrm{Y}|\right] &\geq \mathrm{E}\left[|\mathrm{K}'|\right] - \mathrm{E}\left[|\mathrm{P}'|\right] \\
&\sim q\mu - q^2\frac{1}{2}\frac{k_2(n)^4}{n^2}\mu^2.
\end{aligned}$$

By choosing $q = \frac{n^2}{\mu k_2(n)^4} < 1$ to maximize this lower bound we get

$$\mathrm{E}\left[|\mathrm{Y}|\right] \geq \frac{n^2}{2k_2(n)^4}(1+o(1)).$$

■

**Remark 20.5** *Claim 20.4 says that the number of edges of the random $G \in \mathcal{G}_{n,1/2}$ that can be covered by edge-disjoint $k_2(n)$-cliques is on the order of $\frac{n^2}{k_2(n)^2}$. Hence of the $\Theta(n^2)$ edges in $G$, only on the order of one* *in $k_2(n)^2$, that is one in $(\log_2 n)^2$, of them need be covered by the cliques. There is in fact good reason to believe that a constant fraction of the edges of $G$ can be covered by edge-disjoint $k_2$-cliques. Note also that, since $g(k_2(n)) = n^{3+o(1)}$, $G$ contains about $n^3$ $k_2$-cliques (albeit not edge-disjoint). Thus Claim 20.4 should not be surprising.*

**Remark 20.6** *We actually only require the even weaker fact that* $\mathrm{E}[\mathrm{Y}] > \mathrm{n}^{3/2+\epsilon}$*, which gives us a slightly different bound in Lemma 20.3 that can still be used to prove Theorem 20.2. Exercise: Why?*

## 20.2 Random Geometric Traveling Salesman Problem

Given $n$ points, the objective of the geometric traveling salesman problem is to find a shortest tour which visits each point exactly once. Let the random variables $Z_1, Z_2, \ldots, Z_n$ be $n$ points chosen independently and uniformly at random from inside the $d$-dimensional unit cube (Figure 20.1 shows an example for the 2-dimensional case). Let $L_n$ denote the length of the shortest traveling salesman (TS) tour through $Z_1, Z_2, \ldots, Z_n$. Computing $L_n$, in general, is NP-complete. However, we will see that, for a random instance, $L_n$ is tightly concentrated around its mean $\mathbb{E}[L_n]$.

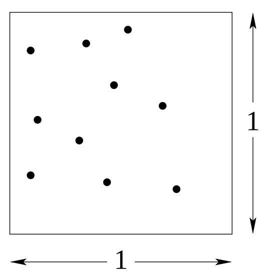


Figure 20.1: The unit square with $n$ randomly chosen points

### 20.2.1 Expected Length of TS Tour

We state, without proof, some results on the expected length of the shortest TS tour for a randomly chosen set $S = \{Z_1, \ldots, Z_n\}$ of $n$ points in a $d$-dimensional unit cube.

The following theorem about the expected length of the minimal tour is easy to show using subadditivity.

**Theorem 20.7** $\mathbb{E}[L_n] \sim \gamma_d n^{\frac{d-1}{d}}$*, where $\gamma_d$ is a constant depending on $d$.*

Furthermore, the following bounds for $\gamma_d$ $(d = 1, 2, 3)$ are known:

$$0.44 \leq \gamma_2/\sqrt{2} \leq 0.65$$
$$0.37 \leq \gamma_3/\sqrt{3} \leq 0.62$$
$$0.34 \leq \gamma_4/\sqrt{4} \leq 0.56$$

(It is natural to scale by $\sqrt{d}$, the diameter of the cube in $d$ dimensions.)

The following asymptotic result is due to Rhee [R92]:

**Theorem 20.8 (Rhee [R92])** $\lim_{d\to\infty} \frac{\gamma_d}{\sqrt{d}} = \frac{1}{\sqrt{2\pi e}} \approx 0.242.$

### 20.2.2 Sharp Concentration of Shortest TS Tour

We show, using Azuma's inequality, that with high probability, the length $L_n$ of the minimal tour does not deviate much from its mean $\mathbb{E}[L_n]$:

**Theorem 20.9** *For the TSP in $d = 2$ dimensions,* $\Pr[|L_n - \mathbb{E}[L_n]| \geq \lambda] \leq 2\exp\left(-\frac{A\lambda^2}{\log n}\right)$, *where $A$ is a universal constant.*

This theorem implies that deviations of size $\omega(\sqrt{\log n})$ are unlikely. Since $\mathbb{E}[L_n] \sim \gamma_2\sqrt{n}$, this constitutes a tight concentration about the mean.

A similar result holds for any dimension $d$:

$$\Pr[|L_n - \mathbb{E}[L_n]| \geq \lambda] \leq 2\exp\left(-\frac{A_d\lambda^2}{n^{(d-2)/d}}\right)$$

for some universal constant $A_d$. Thus deviations larger than $n^{(d-2)/2d}$ are unlikely. Since in this case $\mathbb{E}[L_n] \sim \gamma_d n^{(d-1)/d}$, this is again tight concentration.

**Proof of Theorem 20.9:** Let $f(Z_1, \ldots, Z_n)$ denote the length of the shortest TS tour in $\{Z_1, \ldots, Z_n\}$. Consider the Doob martingale $X_i = \mathbb{E}[f(Z_1, \ldots, Z_n) \mid Z_1 \ldots Z_i]$. It is easy to see (**exercise!**) that the differences $X_i - X_{i-1}$ are bounded by a constant, as in our previous examples, but this would give us only a tail bound of the form $\exp(-\frac{A\lambda^2}{n})$, which rules out deviations that are larger than $\sqrt{n}$, which for $d = 2$ is the same order as the mean. In order to get the much tighter concentration claimed in the theorem we will need to get better bounds on the differences. To this end, note first that we may trivially write

$$X_i - X_{i-1} = \mathbb{E}[f(Z_1, \ldots, Z_i, \ldots, Z_n) - f(Z_1, \ldots, \hat{Z}_i, \ldots, Z_n) \mid Z_1, Z_2, \ldots, Z_i]$$

where $\hat{Z}_i$ has the same distribution as $Z_i$ but is independent of it.

Define $\Delta_i = |f(Z_1, \ldots, Z_i, \ldots, Z_n) - f(Z_1, \ldots, \hat{Z}_i, \ldots, Z_n)|$. Observe that for any set of points $S$,

$$f(S) \leq f(S \cup \{z\}) \leq f(S) + 2\min_{y\in S} |y - z|.$$

To see this, consider Figure 20.2. The point $y \in S$ is a point in $S$ closest to $z$. Let the point $u$ be the next point after $y$ in the shortest tour in $S$. We get a tour in $S \cup \{z\}$ by taking the $y \to z \to u$ path instead of

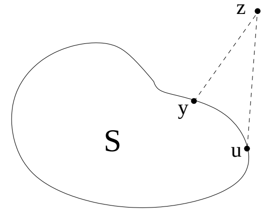


Figure 20.2: Adding a point $z$ to the set $S$

$y \to u$. By the triangle inequality, $|u - z| \leq |y - z| + |y - u|$. Therefore, the length of the shortest tour in $S \cup \{z\}$ is at most $|y - z| + |z - u| - |y - u| \leq 2|y - z|$ more than the one in $S$.

Setting $S = \{Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_n\}$, we get $\Delta_i \leq 2[q(Z_i) + q(\hat{Z}_i)]$, where $q(y)$ is the shortest distance from a point $y$ to the set $\{Z_{i+1}, \ldots, Z_n\}$. Taking conditional expectations with respect to $Z_1, \ldots, Z_i$, we get

$$X_i - X_{i-1} \leq 2\mathbb{E}[q(Z_i) + q(\hat{Z}_i)|Z_1, \ldots, Z_i] \leq 4\mathbb{E}[Q_i],$$

where the random variable $Q_i$ is the shortest distance from a fixed point to $n - i$ randomly selected points. Now consider the ball of radius $r$ centered at a fixed point (point) $z$ (the solid point in Figure 20.3).

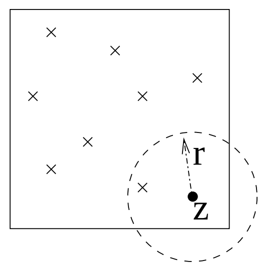


Figure 20.3: The solid dot is the point $z$; the crosses represent the randomly chosen $n - i$ points

Its area in the square is at most $Cr^2$ (for some constant $C$). Therefore, $\Pr[Q_i \geq r] \leq (1 - Cr^2)^{n-i}$. Hence, we may compute the expectation of $Q_i$ as follows:

$$\begin{aligned}
\mathbb{E}[Q_i] &= \int_0^\infty \Pr[Q_i > r]dr \\
&\leq \int_0^{\sqrt{2}} (1 - Cr^2)^{n-i} dr \\
&\leq \int_0^{\sqrt{2}} \exp[-Cr^2(n-i)]dr \\
&\leq \frac{C_1}{\sqrt{n-i}},
\end{aligned}$$

for some other constant $C_1$. Since $|X_i - X_{i-1}| \leq 4\mathbb{E}[Q_i]$, we have

$$|X_i - X_{i-1}| \leq \frac{C_2}{\sqrt{n-i}} = c_i.$$

This holds for $1 \leq i < n$. For $i = n$, we can just use the trivial bound $|X_n - X_{n-1}| \leq 4\sqrt{2} = c_n$.

Finally, by Azuma's inequality, we get

$$\Pr[|L_n - \mathbb{E}[L_n]| \geq \lambda^2] \leq 2\exp\left(\frac{-\lambda^2}{2\sum_{i=1}^n c_i{}^2}\right) = 2\exp\left(\frac{-\lambda^2}{2\left[(4\sqrt{2})^2 + \sum_{i=1}^{n-1} \frac{C_2{}^2}{n-i}\right]}\right) \leq 2\exp\left(\frac{-A\lambda^2}{\log n}\right).$$

■

**Exercise:** Extend the result of Theorem 20.9 to $d \in \mathbb{N}$ dimensions (see the paragraphs following the theorem for the correct form of the bound). [Hint: The proof is essentially the same, except for the calculations concerning the ball of radius $r$.]

## References

[B88] B. Bollobás, "The chromatic number of random graphs," *Combinatorica* **8** (1988), pp. 49–56.

[GM75] G. R. Grimmett and C. J. H. McDiarmid, "On coloring random graphs," *Math. Proc. Cambridge Philos. Soc.* **77** (1975), pp. 313–324.

[R92] W. Rhee, "On the Travelling Salesperson Problem in Many Dimensions," *Random Structures and Algorithms* **3** (1992), pp. 227–233.

[SS87] E. Shamir and J. Spencer, "Sharp concentration of the chromatic number on random graphs $G_{n,p}$," *Combinatorica* **7** (1987), pp. 121–129.