Applications of Martingales

1 Recovery of the Hoeffding Bound

Let the X_i be independent random variables with $a_i \leq X_i \leq b_i$ always. We are interested in $X = f(X_1, \ldots, X_n) = \sum_{i=1}^n X_i$. Let $Y_i = \mathbb{E}[X \mid X_1, \ldots, X_i]$. Since Y_i is a Doob martingale we know $Y_0 = \mathbb{E}[X]$ and $Y_n = X$.

We'd like to study $\Pr[|X - \mathbb{E}[X]| > \Delta]$. We'd like to use the corollary form of Azuma's Inequality from last time, so we need to show a Lipschitz condition. Notice that if we change x_i, then $f(x_1, \ldots, x_n)$ changes by exactly the amount of change in x_i. This is upper-bounded by $c_i = b_i - a_i$. So, f satisfies a Lipschitz condition with c_i in the i^{th} argument.

Thus,

$$\Pr[|X - \mathbb{E}[X]| > \Delta] \leq 2 \exp\big(\frac{-\Delta^2}{2\sum_{i=1}^n c_i^2}\big).$$

This is exactly a Hoeffding bound.

2 Largest Clique Size in a Random Graph

Recall that a random graph G has each edge (u, v) is present independently with probability $p_{u,v}$. Define indicator random variables $X_{u,v}$ for whether $(u, v) \in G$. Then, let $f((X_{u,v})_{u,v})$ be the size of the largest clique in the graph defined by the $X_{u,v}$.

Notice that f satisfies a Lipschitz condition in the sense that adding or removing an edge can change the size of the largest clique by at most one.

Recall the edge exposure martingale which reveals edges one at a time. Then we have

$$\begin{aligned}\Pr[|\text{clique size} - \mathbb{E}[\text{clique size}]| > \Delta] &\leq 2 \exp\big(\frac{-\Delta^2}{2\sum_{u,v} 1}\big)\\ &= 2 \exp\big(\frac{-\Delta^2}{n(n-1)}\big).\end{aligned}$$

How small can we make Δ so this is still small? We need $\Delta \geq n$ for any nontrivial bound. $\Delta \approx n$ is trivial because the clique size is always in $[n]$. But that is the best we can do with this bound.

We can try the vertex exposure martingale instead, which reveals nodes one at a time. Revealing node i reveals the presence or absence of all edges between v_i and the v_j for $j < i$. Let Y_i be a 0-1 vector of length $i - 1$ capturing all edges between v_i and v_j for $j < i$. Then, let $g(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)$ be the size of the largest clique in the graph defined by the $\boldsymbol{y}_i$.

Notice that g still satisfies a Lipschitz condition in the same sense as f, because adding or removing a node can change the size of the largest clique by at most one. Thus

$$\begin{aligned}\Pr[|\text{clique size} - \mathbb{E}[\text{clique size}]| > \Delta] &\leq 2 \exp\big(\frac{-\Delta^2}{\sum_{i=1}^n 1}\big)\\ &= 2 \exp\big(\frac{-\Delta^2}{2n}\big).\end{aligned}$$

Thus with probability at least $1 - n^{-c}$, the largest clique size is concentrated to within $\mathcal{O}(\sqrt{n \log n})$.

Notice that something similar happens when studying the chromatic number of a random graph. We have a Lipschitz condition with 1 in each coordinate for both martingales, since we at worst have to add a new color. This results in exactly the same concentration bounds as largest clique.

3 Connected Components in a Random Graph

In the vertex exposure martingale, if $v_j, j < i$ have no edges and v_i connects them all, then we could decrease the number of components by $i - 1$. So the best Lipschitz condition we can get is $c_i = i - 1$. Thus

$$\begin{aligned}
\Pr[|\#\text{components} - \mathbb{E}[\#\text{components}]| > \Delta] &\leq 2 \exp\big(\frac{-\Delta^2}{2\sum_{i=1}^{n}(i-1)^2}\big) \\
&= 2 \exp\big(\frac{-\Delta^2}{\Theta(n^3)}\big).
\end{aligned}$$

So the best we could get is $\Delta \approx \Theta(n^{3/2})$. Since there can be at most n connected components, this is useless!

Let's consider the edge exposure martingale instead. Since any edge can at most merge two components, and otherwise creates a new component, we have a Lipschitz condition with $c_i = 1$. Thus

$$\begin{aligned}
\Pr[|\#\text{components} - \mathbb{E}[\#\text{components}]| > \Delta] &\leq 2 \exp\big(\frac{-\Delta^2}{2\binom{n}{2}}\big) \\
&= 2 \exp\big(\frac{-\Delta^2}{n(n-1)}\big).
\end{aligned}$$

So the best bound would be $\Delta \approx \Theta(n)$, which is still useless! Connected components seem to fluctuate too much to get any use out of this technique.

4 Azuma's Inequality for Degree Distributions

The following is from N. Wormald: Models of Random Regular Graphs. We are interested in random graph models that produce a particular degree distribution. Typically, a degree distribution p_i says what fraction of nodes have degree i. Equivalently, we can prescribe the degree d_v of each node. There are generative models based on particular random processes that, with high probability, produce graphs with certain distributions (not the focus here, see CSCI 673). Instead, we want to draw uniformly from the set of all graphs in which each node v has a prescribed degree d_v.

To make our lives easier, we allow self-loops and parallel edges (and count them towards the degree). Consider the following generative model, called the configuration model: assume $\sum d_v$ is even (otherwise it's impossible), then for each node v, generate d_v "stubs". Find a uniformly random perfect matching of the $\sum d_v = 2m$ stubs, and "compress" the stubs back into the nodes, keeping the edges of the matching.

How do we find a uniformly random perfect matching of the stubs? One way is to generate a uniformly random order, then match $2i - 1$ with $2i$ for all i. Another is to put the stubs in an arbitrary order, go through the stubs in this order, and always look at the first unmatched stub i. Match it with a uniformly random unmatched stub (which necessarily has $j > i$).

We'd like similar concentration bounds to those for independent edge generation. The problem is, the presence or absence of edges is not independent, and it is not clear whether there is some other representation of randomness as $X_1, \ldots, X_k$ such that we can consider the graph determined by $X_1, \ldots, X_k$ independent and obtain useful Lipschitz conditions.

Instead, define a different approach. We say that G, G' differ in a "simple switching" if there exists u, v, u', v' such that: (i) $(u, v) \in G, (u', v') \in G$, (ii) $(u, v') \in G', (u', v) \in G'$, (iii) $G' = G \setminus \{(u, v), (u', v') \cup$

$\{(u, v'), (u', v)\}$. Basically, everything else in the graph is the same except two edges swap destinations. When there is a simple switching between G and G', we write $G \sim G'$. In some sense this is the minimal change you can make to move between two graphs with the same degree distribution.

Theorem 1. *[Wormald 2.19]. Let X_n be a random variable defined over random multigraphs with a given degree sequence (d_v) such that*

$$|X(G) - X(G')| < c$$

whenever $G \sim G'$. Then

$$\begin{aligned}\Pr[|X - \mathbb{E}[X]| \geq \Delta] &\leq 2\exp\big(\frac{-\Delta^2}{c^2 \sum_v d_v}\big) \\ &= 2\exp\big(\frac{-\Delta^2}{2c^2 m}\big).\end{aligned}$$

This is exactly what we would get out of an edge exposure martingale with m steps and a bounded difference of c for each step.

Proof. Define a Doob martingale with edges exposed in the order defined above: stubs are numbered arbitrarily $1, \ldots, 2m$ and we always match the lowest-numbered free stub with a uniformly random free stub. Define the random graph $G(k)$ to be the set of edges revealed after k iterations. Thus $Y_k := \mathbb{E}[X \mid G(k)]$ is a Doob martingale for X – that is, $Y_0 = \mathbb{E}[X]$ and $Y_m = X$. Note that $m = \frac{1}{2}\sum_v d_v$.

Our goal is to apply Azuma's Inequality, so we need to establish bounded differences: $|Y_{k+1} - Y_k| \leq c$ for all k. Then, the result follows immediately. We will compare

$$\mathbb{E}[X \mid G(k+1) = g(k+1)] \text{ vs. } \mathbb{E}[X \mid G(k) = g(k)].$$

We will consider a graph $g(k)$ and an edge $(i, j) = g(k+1) \setminus g(k)$. (Note that i and j are stubs, not nodes. We have

$$\begin{aligned}\mathbb{E}[X \mid G(k+1) = g(k+1)] &= \sum_x x \cdot \Pr[X = x \mid G(k+1) \mid g(k+1)] \\ &= \sum_{h \sim \text{all graphs w/correct degrees}} X(h) \cdot \Pr[\text{draw } h \mid G(k+1) = g(k+1)] \\ &= \sum_{h \supseteq g(k+1)} X(h) \cdot \Pr[\text{draw } h \mid G(k+1) = g(k+1)].\end{aligned}$$

Similarly, we know

$$\mathbb{E}[X \mid G(k) = g(k)] = \sum_{g \supseteq g(k)} X(j) \cdot \Pr[\text{draw } h \mid G(k) = g(k)].$$

In the end, we want

$$\begin{aligned}|Y_{k+1} - Y_k| = |&\sum_{h \subseteq g(k+1)} X(h) \cdot \Pr[h \mid G(k+1) = g(k+1)] \\ &- \sum_{g \supseteq g(k)} X(g) \cdot \Pr[g \mid G(k) = g(k)]|.\end{aligned}$$

So we want to somehow relate these two sums with a charging argument. For now, fix any $g \supseteq g(k)$ (where g is a completed graph with m edges) and recall $g(k+1) = g(k) \cup \{i, j\}$. Let j' be the match of i in g and ℓ be the match of j in g.

Define $\sigma(g) = g \setminus \{(i, j'), (j, \ell)\} \cup \{(i, j), (j', \ell)\}$. If i is matched with j in g, then $\sigma(g) = g$. Then, $\sigma(g) \supseteq g(k+1)$ because we "fixed" the one edge they disagreed on. Since matchings are drawn uniformly at random,

$$\Pr[\text{draw } \sigma(g) \mid G(k) = g(k) = \Pr[\text{draw } g \mid G(k) = g(k)].$$

How many different g have the same $\sigma(g)$? There are $2m-2k-1$ possible choices for j'. But i, j, j' determines ℓ, so j' fully determines $\sigma(g)$. So there are $2m-2k-1$ choices of g with $\sigma(g) = h$ total for every $h \supseteq g(k+1)$. We can now rewrite $|Y_{k+1} - Y_k|$ as

$$\begin{aligned} | \sum_{h \supseteq g(k+1)} & X(h) \cdot \Pr[h \mid G(k) = g(k)] \cdot (2m - 2k - 1) \\ & - \sum_{g \supseteq g(k)} X(g) \cdot \Pr[\sigma(g) \mid G(k) = g(k)]|. \end{aligned}$$

By the triangle inequality this is at most

$$\sum_{h \supseteq g(k+1)} \Pr[h \mid G(k) = g(k)] \cdot |(2m - 2k - 1) \cdot X(h) - \sum_{g : \sigma(g) = h} X(g)|.$$

Because the second sum contains exactly $2m - 2k - 1$ terms, we use the triangle inequality once more to obtain

$$\sum_{h \supseteq g(k+1)} \Pr[h \mid G(k) = g(k)] \cdot \sum_{g : \sigma(g) = h} |X(h) - X(g)|.$$

Since $\sigma(g)$ and g differ by a simple swap, this is at most

$$c \sum_{h \supseteq g(k+1)} \Pr[h \mid G(k) = g(k)] \cdot (2m - 2k - 1),$$

which is exactly equal to c because the inner sum is

$$\Pr[h \mid G(k+1) = g(k+1)].$$

$\square$