7 Correlation and Janson's inequalities

7.1 The Harris-FKG inequality

We're often interested in understanding probabilities related to a graph $G(n, p)$: for example, we might want the probability of some event occurring, such as the probability of having a triangle (which we don't know how to compute yet).

But we can also ask questions like "does this probability increase or decrease if we condition on some other event?" For example, if we have a Hamiltonian cycle, does this increase or decrease our chance of having a triangle?

These don't seem all that relevant to each other, but let's look more closely. Having a Hamiltonian cycle is indicative of having more edges in the graph, and thus the probability of having a triangle should go up. This isn't a proof, but the basic idea here is correct. Similarly, if we know that our graph is planar, we should expect fewer edges, so we should expect a smaller probability of having a triangle. We can rigorize this!

Let's say we have n independent Bernoulli variables $x_1, \cdots, x_n$ (for most applications, the probabilities are the same). An event A is **increasing** if changing some 0s to 1s never destroys the events. In other words, if $x \le x'$ pointwise, and $x \in A$, then $x' \in A$ as well. Notice that we can view A as a subset of $\{0, 1\}^n$: this is an **up-set**, because it's closed upwards. Similarly, decreasing events are **down-sets**.

Example 7.1

Let's say we have a graph $G(N, p)$, and we have $n = \binom{N}{2}$ random variables for the edges. Having a Hamiltonian cycle and being connected are increasing, while having average degree at least 5, planarity, and being 4-colorable are decreasing.

Proposition 7.2 (Harris inequality)

If A and B are increasing events of independent boolean variables, then

$$\Pr(A \cap B) \ge \Pr(A)\Pr(B).$$

We also have that $\Pr(A|B) \ge \Pr(A)$.

(Both of these imply that A and B are **positively correlated**.) More generally, we can let each Ω_i be a probability space that is linearly ordered (for example, $\{0, 1\}$ in the case above).

Definition 7.3

A function $f(x_1, \cdots, x_n)$ is **monotone increasing** if given two vectors $x \le x'$ (pointwise in every coordinate), $f(x) \le f(x')$.

Theorem 7.4 (More general version of the Harris inequality)

Let f, g be increasing functions of independent random variables $x_1, \cdots, x_n$. Then $\mathbb{E}[fg] \ge \mathbb{E}[f]\mathbb{E}[g]$.

(This implies the first version by picking f and g to be indicator functions.)

Later generalizations were made by "FKG" (Fortuin, Kastelyn, Ginibre), but we won't discuss FKG inequalities in their full generality. The idea is that we can relax the independence condition, use a distributive lattice, and use a few other conditions.

Proof of the Harris inequality. We will use induction on the number of random variables. If $n = 1$, then we're saying that given two functions f, g that are monotone increasing on one variable,

$$\mathbb{E}[fg] \geq \mathbb{E}[f]\mathbb{E}[g].$$

This is due to Chebyshev (the same guy as earlier in the course): the proof is that picking x, y independently,

$$0 \leq \mathbb{E}\left[(f(x) - f(y))(g(x) - g(y))\right]$$

since $f(x) - f(y)$ and $g(x) - g(y)$ have the same sign, and expanding this out,

$$= 2\mathbb{E}[fg] - 2\mathbb{E}[f]\mathbb{E}[g],$$

implying the result for one variable. Now for the inductive step, let $h = fg$. We'll fix x_1, defining a new function

$$f_1(x) = \mathbb{E}[f \mid x_1 = x] :$$

basically, fix one of the variables in our function f. Likewise, let $g_1(x) = \mathbb{E}[g \mid x_1 = x]$ and similar for h_1. We know that f_1, g_1 are monotone increasing functions on the remaining variables. Note that

$$h_1(x_1) \geq f_1(x_1)g_1(h_1)$$

by the induction hypothesis, so

$$\mathbb{E}[fg] = \mathbb{E}[h] = \mathbb{E}[h_1]$$

(letting x_1 be random as well now), and pointwise, this is

$$\geq \mathbb{E}[f_1 g_1] \geq \mathbb{E}[f_1]\mathbb{E}[g_1] = \mathbb{E}[f]\mathbb{E}[g]$$

by the base case, since f_1 and g_1 are one-variable functions. □

Corollary 7.5

Decreasing events are also positively correlated:

$$\Pr(A \cap B) \geq \Pr(A) \cap \Pr(B).$$

(Take the complement of a decreasing event to get an increasing event.) Similarly, if one event is increasing and another is decreasing, they are negatively correlated. Finally, if all A_is are increasing or all decreasing, we can say that

$$\Pr(A_1 \cdots A_k) \geq \Pr(A_1) \cdots \Pr(A_k).$$

7.2 Applications of correlation

Example 7.6

Let's find the probability that $G(n, p)$ is triangle-free.

There are lots of possible appearances of triangles, and lots of dependent probabilities. But we know that these events are all correlated! In particular, let A_{ijk} be the event that (i, j, k) is not a triangle. A_{ijk} is a decreasing event (on the edges), and this means the probability of having no triangles at all is (by Harris' inequality)

$$= \Pr\left(\bigwedge_{ijk} \overline{A_{ijk}}\right) \geq \prod_{ijk} \Pr(\overline{A_{ijk}}) = (1-p^3)^{\binom{n}{3}}.$$

How close is this to the truth? Taking $p = o(1)$, we can approximate this as

$$\geq e^{-(1+o(1))p^3n^3/6}.$$

(By the way, the probability $G(n,p)$ is triangle free is monotone for p by coupling - having a higher chance of including each edge just makes our chances worse.) One way to obtain an upper bound is by Janson's inequality, which is kind of dual to the Lovász Local Lemma, but we'll see that in the next few sections.

Example 7.7

What is the probability that $G\left(n, \frac{1}{2}\right)$ has maximum degree at least $\frac{n}{2}$?

Let A_v be the event that the degree of v is at most $\frac{n}{2}$: each of these has probability at least $\frac{1}{2}$, so by the Harris inequality, the probability is at least the product of the individual vertex probabilities, which is at least 2^{-n}.

Is this close to the truth - what's the actual value? It turns out the probability is indeed of the form $(c+o(1))^n$. Is $c = \frac{1}{2}$, meaning that our correlation inequality is essentially tight, or is $c > \frac{1}{2}$, which means our lower bound is not very good?

Theorem 7.8 (Riordan-Selby, 2000)

The probability that $G\left(n, \frac{1}{2}\right)$ has maximum degree at least $\frac{n}{2}$ is $(c+o(1))^n$, where $c \approx 0.6102$.

This is very technical, but let's do a "physicist proof" to see where the number comes from.

Solution motivation. Use a continuous model instead. Instead of making each variable Bernoulli, put a standard normal distribution on each (undirected) edge of K_n instead. Now the degree is just the sum of the standard normals of the edges connected to each vertex.

Let $W_v = \sum_{u \neq v} Z_{uv}$ be this sum: We know each event $W_v \leq 0$ has a $\frac{1}{2}$ chance of being at most 0. What's the probability all W_vs are less than 0? We know the W_vs are a joint normal distribution, entirely dependent on their covariance matrix. Then the variance of W_v is $n-1$, and the covariance between W_u and W_v is 1 because of the shared edge. So now we can directly compute by using a different model with the same covariance matrix: this is identically distributed as

$$\sqrt{n-2}\,(Z_1', \cdots, Z_n') + Z_0'(1, 1, \cdots, 1)$$

(where each Z_i' is a standard normal distribution independent of the others).

Then what is the probability that this vector has all entries less than or equal to 0? That's an explicit calculation: let Φ be the cumulative distribution of the standard normal distribution: conditioning on Z_0', we find that the desired probability is

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} \Phi\left(\frac{-z}{\sqrt{n-2}}\right)^n dz$$

where dz refers to us picking Z_0' first. To evaluate this, we can substitute $z = y\sqrt{n}$ for scaling, this integral becomes

$$\sqrt{\frac{n}{2\pi}} \int_{-\infty}^{\infty} e^{-nf(y)} dy,$$

where $f(y) = \frac{y^2}{2} - \log \Phi\left(y\sqrt{\frac{n}{n-2}}\right)$. We can pretend $f(y) = \frac{y^2}{2} - \log \Phi(y)$ and bound the error, and then look at the asymptotic property of this integral as $n \to \infty$. Well, there's a general principle:

Fact 7.9

If f is a "sufficiently nice" function with a unique minimum at y_0, then

$$\int_{-\infty}^{\infty} e^{-nf(y)} dy = \left(e^{-f(y_0)} + o(1)\right)^n.$$

Basically, as $n \to \infty$, we only get contributions from the smallest $f(y)$. So the rest is just finding the right value of y_0: we can just do this by taking the derivative, and that yields $c \approx 0.6102$ as desired. □

The actual proof is very technical, but this is just a general idea to explain where the constant c potentially comes from. A lot of probability theory now is rigorizing physical intuitions!

7.3 The first Janson inequality: probability of non-existence

The Harris-FKG inequality gives us lower bounds on the probabilities of certain events, but those are not necessarily tight bounds. We'll now start to explore some methods of obtaining upper bounds that are hopefully close to the Harris lower bounds.

Setup 7.10

Pick a random subset R of $[N]$, where each element is chosen independently (usually with probability $\frac{1}{2}$). Refer to $[N]$ as the "ground set." Suppose we have some subsets $S_1, S_2, \cdots, S_k \subseteq [N]$, and A_i be the "bad event" that $S_i \subseteq R$ for all i.

Denote $X = \sum_i 1_{A_i}$ to be the number of A_is that occur. Note that

$$\mu = \mathbb{E}[X] = \sum_i \Pr(A_i),$$

and we have a dependency graph $i \sim j$ if $i \neq j$ and $S_i \cap S_j \neq \varnothing$ (the two underlying subsets overlap). Much like in our earlier covariance calculations, let

$$\Delta = \sum_{(i,j):i\sim j} \Pr(A_i \wedge A_j).$$

(Δ is an upper bound on the variance.) For example, in the current problem we're considering, the S_is are the triangles, and $[N]$ is the set of edges.

Back with the second moment method, we found that if the standard deviation was small relative to the mean, then we have concentration, so we want Δ to generally be small. Janson's inequalities are going to give us better control over our concentration!

Theorem 7.11 (First Janson inequality)

With the definitions in Setup 7.10, the probability that no bad events occur is

$$\Pr(X = 0) \leq e^{-\mu + \frac{\Delta}{2}}.$$

So if Δ is small relative to the mean, then we are essentially upper bounded by $e^{-\mu}$. By the way, this is pretty close to the truth: if all bad events occur with some probability $p = o(1)$, and $\Delta = o(\mu)$, then our lower bound from Harris is

$$\Pr(X = 0) \geq \prod_i \Pr(\overline{A_i}) = e^{-(1+o(1))\mu}$$

by using $(1+x)^n \approx 1 + nx$. The original proof interpolated the derivative of the exponential generating function, but we'll look at a different one.

Proof by Boppana and Spencer, with a modification by Warnke. Let

$$r_i = \Pr(A_i \mid \overline{A}_1\overline{A}_2 \cdots \overline{A}_{i-1})$$

(so this is conditioned on the probability that none of the previous bad events occur). Then the probability that no bad events occur is a chain of conditional probabilities:

$$\Pr(X = 0) = \Pr(\overline{A_1}) \Pr(\overline{A_2} \mid \overline{A_1}) \cdots \Pr(\overline{A_k} \mid \overline{A_1} \cdots \overline{A_{k-1}}) = (1 - r_1)(1 - r_2) \cdots (1 - r_k) \leq e^{-r_1 - r_2 - \cdots - r_k}.$$

It now suffices to show that for all $i \in [k]$, we have

$$r_i \geq \Pr(A_i) - \sum_{j<i, j\sim i} \Pr(A_i \wedge A_j).$$

where the sum only accounts for those A_j with $j < i$ that are dependent on A_i. Then we'd be done, since we have

$$\Pr(X = 0) \leq e^{-\sum \Pr(a_i) + \sum_{i\sim j} \Pr(A_i \wedge A_j)} = e^{-\mu + \frac{\Delta}{2}},$$

as desired. Well, the proof of that statement is somewhat reminiscent of the Lovász Local Lemma proof.

Fix i, and split up the events into those that depend and don't depend on i: let D_0 be $\bigwedge_{j<i:j\not\sim i} \overline{A_j}$ and D_1 be $\bigwedge_{j<i:j\sim i} \overline{A_j}$. This partitions all events A_j with $j < i$, and now

$$r_i = \Pr(A_i | D_0 D_1) = \frac{\Pr(A_i D_0 D_1)}{\Pr(D_0 D_1)} \geq \frac{\Pr(A_i D_0 D_1)}{\Pr(D_0)}$$

by Bayes' formula. We're trying to use the fact that D_0 is independent of A_i here, so by Bayes' rule again, this is equal to

$$= \Pr(A_i D_1 | D_0).$$

We can write this as the probability

$$\Pr(A_i | D_0) - \Pr(A_i \overline{D_1} | D_0).$$

By independence of A_i and D_0, this first term is now $\Pr(A_i)$. Now because A_i is increasing, $\overline{D_1}$ is increasing, and D_0 is decreasing, we can use the Harris-FKG inequality: conditioning on a decreasing event must make the probability of an increasing event go down, so

$$\Pr(A_i \overline{D_1} | D_0) \leq \Pr(A_i \overline{D_i}) = \Pr(A_i \cap \vee_{j<i, j\sim i} A_j)$$

and by a union bound, this is at most

$$\leq \sum_{j<i, j\sim i} \Pr(A_i \cap A_j),$$

as desired.

$\square$

Fact 7.12
Janson's inequality is originally about random subsets including a certain S_i, but we actually only need that the A_is are increasing events. So they don't necessarily have to be of the form $S_i \subseteq R$ for our random subset R.

Previously, our dependency graph had edges wherever the S_is overlapped. In general, there's a difference between pairwise independence and mutual independence, so it seems that we have to be careful. However, we get lucky:

Lemma 7.13
If A, B, C are increasing events, and A is independent of B and C, then A is independent of $B \wedge C$.

So the pairwise dependence graph is the same as the dependency graph. This may not be that intuitive: remember that one counterexample was $R \subset [3]$, where A_1 is the event "$|R \cap \{1,2\}|$ is even", A_2 is the event"$|R \cap \{1,3\}|$ is even," and A_3 is the event "$|R \cap \{2,3\}|$ is even:" we have pairwise independence but not mutual independence.

Proof. Note that $\Pr(A \wedge (B \wedge C)) + \Pr(A \wedge (B \vee C)) = \Pr(A \wedge B) + \Pr(A \wedge C) = \Pr(A)(\Pr(B) + \Pr(C))$. But on the other hand, by Harris-FKG, since the events are increasing,

$$\Pr(A \wedge (B \wedge C)) \geq \Pr(A)\Pr(B \wedge C), \Pr(A \wedge (B \vee C)) \geq \Pr(A)(\Pr(B \vee C)).$$

But these last two inequalities actually add to the first equality! So equality must occur, and A is independent of $B \wedge C$ and $B \vee C$. $\square$

Let's use the first Janson inequality to get an upper bound on the probability that $G(n, p)$ is triangle-free. Recall the Second Moment Method calculations that we made earlier: the expected number of triangles is

$$\mu = \binom{n}{3} p^3 \asymp n^3 p^3.$$

Meanwhile, Δ is counting the number of pairs of triangles with a shared edge: these look like 4-cycles with a diagonal, and that evaluates out to

$$\Delta \asymp n^4 p^5.$$

So $\Delta \ll \mu \iff p \ll n^{-1/2}$, and therefore the probability that $G(n, p)$ is triangle free is $e^{-(1+o(1))\mu}$ if $p \ll n^{-1/2}$. Note that this is exactly the right asymptotic behavior (see the logic after Theorem 7.11).

Well, what if p is larger - does the formula still hold when $p \gtrsim n^{-1/2}$, and is Harris a good approximation? Bipartite graphs are triangle-free, so the probability of a triangle-free $G(n, p)$ is at least the probability that it is bipartite. This is at least the probability that $G(n, p)$ has no edges, which is $(1-p)^{\binom{n}{2}} = e^{-(1+o(1)pn^2/2}$ for $p \ll 1$. This is actually much larger than $e^{-c\mu}$ if $p \gg n^{-1/2}$! So the lower bound of Harris is true, but it's inferior to very stupid bounds.

7.4 The second Janson inequality

Let's try to strengthen our inequality when $\Delta > \mu$

Theorem 7.14 (Second Janson Inequality)
Again using the assumptions of Setup 7.10, if $\Delta \geq \mu$, then

$$\Pr(X = 0) \leq e^{-\frac{\mu^2}{2\Delta}}.$$

Proof. For each subset of the bad events $T \subset [k]$, let

$$X_T = \sum_{i \in T} 1_{A_i}$$

be the number of bad events in T only, and $\mu_T = \sum_{i \in T} \Pr(A_i), \Delta_T = \sum_{(i,j) \in T^2, i \sim j} \Pr(A_i \wedge A_j)$ be defined similarly. Then the probability that none of the bad events occur is always

$$\Pr(X = 0) \le \Pr(X_T = 0) \le e^{-\mu_T + \frac{\Delta_T}{2}}$$

Choose T randomly: include each element independently with some probability q (to be determined). Then μ_T has expectation $q\mu$, and Δ_T has expectation $q^2\mu$ (since both A_i and A_j need to be kept for any term to count). Thus,

$$\mathbb{E}(-\mu_T + \frac{\Delta_T}{2}) = -q\mu + \frac{q^2\Delta}{2}.$$

Minimizing this, pick $q = \frac{\mu}{\Delta}$, which is at most 1 by the theorem statement. This yields $\frac{\mu^2}{2\Delta}$, so there exists some choice of T so that $-\mu_T + \frac{\Delta_T}{2} \le \frac{-\mu^2}{2\Delta}$, as desired. □

These two Janson inequalities work in different regimes, and it's interesting that the proof of the second uses the proof of the first!

Remark. *This "bootstrapping argument," where we start with a weak inequality and make it stronger, is reminiscent of the crossing number inequality. We had*

$$cr(G) \ge |E| - 3|V|,$$

and this was only quadratic in n. To get a stronger result, we sampled our graph G, which gave us a much stronger inequality of the form $cr(G) \gtrsim \frac{|E|^3}{|V|^2}$.

How do these compare to the second moment method calculations? There, we said that if $\Delta \ll \mu^2$, and $\mu^2 \to \infty$, then X is concentrated around its mean, meaning $\Pr(X = 0) = o(1)$. But here, we have an explicit exponential decay, rather than just knowing that the probability goes to zero.

Does this give a better bound than the first Janson inequality for $G(n, p)$ being triangle free (when p is large)? Say that $p \gg n^{-1/2}$, so that we have $\Delta \ge \mu$. By Janson's second inequality, the probability that $G(n, p)$ is triangle free is now

$$\le e^{-\frac{\mu^2}{2\Delta}} = e^{-\Theta(n^2 p)}.$$

The exponent matches the order of "probability $G(n, p)$ has no edges" from above, which means this is essentially tight! So

$$\Pr(G(n, p) \text{ is triangle free}) = \begin{cases} e^{-(1+o(1))n^3p^3/6} & p \ll n^{-1/2} \\ e^{-\Theta(n^2 p)} & p \gtrsim n^{-1/2}. \end{cases}$$

What is the constant here in the Θ? We can do a bit better than "$G(n, p)$ has no edges," since $G(n, p)$ has probability of being bipartite at least

$$(1 - p)^{2\binom{n/2}{2}} = e^{-(1+o(1))n^2 p/4}.$$

Is this the dominating way of generating graphs with no triangles? The answer turns out to be yes, but we don't yet have the tools to show that. The modern way to think about this is through something called the container method.

Fact 7.15

A lot of probability distributions that are nonnegative-integer-valued are Poisson or Gaussian. In particular, for the $p \ll n^{-1/2}$ case, if the exponent converges to a constant, we get Poisson behavior for the number of triangles in $G(n, p)$.

7.5 Lower tails: the third Janson inequality

One more time, let X be the number of triangles in $G(n, p)$. This time, we want to estimate the probability that X is at most $0.9\mathbb{E}[X]$ (or some other constant times the mean): this is a generalization of estimating the probability that $X = 0$.

This is on a larger order than the standard deviation of X, so Chebyshev-like tools don't help us. It turns out that in these cases, we have exponential decay:

Theorem 7.16 (Third Janson inequality)

Use the assumptions in Setup 7.10 again. For any $0 \le t \le \mu$,

$$\Pr(X \le \mu - t) \le \exp\left(\frac{-t^2}{2(\mu + \Delta)}\right).$$

We'll come back to the proof of this later - interestingly, it also bootstraps the first Janson inequality. First, let's look at a consequence of this by looking some more at triangle counts (here, triangle can also be replaced with any subgraph H). If we still let X be the number of triangles in $G(n, p)$, and we let $t = c\mu \asymp n^3p^3$ for some constant c, then

$$\Pr(X \le (1 - c)\mathbb{E}[X]) \le \exp\left(-\Theta\left(\frac{n^6p^6}{n^3p^3 + n^4p^5}\right)\right).$$

We can clean this up a bit by splitting by dominant term:

$$\Pr(X \le (1 - c)\mathbb{E}[X]) = \begin{cases} \exp(-\Theta(n^3p^3)) & p \lesssim n^{-1/2} \\ \exp(-\Theta(n^2p)) & p \gtrsim n^{-1/2} \end{cases}$$

Are these inequalities tight – that is, do we have a corresponding lower bound? Turns out the answer is yes! The probability of having at most $(1 - c)\mathbb{E}[X]$ includes the probability that there at exactly 0 triangles, and the upper bounds that we just found have exponents on the same order as what we found for the 0-triangle case. This means our bounds are tight up to constant factors in the exponent.

Unfortunately, we don't actually know the value of the constants except for some values of c. We are essentially asking "what is the best way of getting few triangles?", and one good way to do this is to uniformly decrease probability of edges everywhere, which helps for small c. On the other hand, when c is close to 1, we expect that bipartite graphs dominate the few-triangle space. However, it's still a research problem to look at values of c in between and find the dominating graphs.

By the way, there's a reason we're only mentioning the lower tail: the upper tail is completely different. If we use $X \ge (1 + c)\mathbb{E}[X]$, our inequalities are false!

Proof of Theorem 7.16 by Warnke. Define the parameter $q \in [0, 1]$ (value to be determined later). Let $T \subseteq [k]$, where each element is included with probability q independently; let's consider

$$X_T = \sum_{i \in T} 1_{A_i}.$$

We can alternatively write this as a sum over the original bad events:

$$= \sum_{i \in [k]} 1_{A_i} W_i,$$

where each W_i is distributed according to a Bernoulli distribution: 1 with probability q if $i \in T$ and 0 otherwise.

Notice that X, our actual number of bad events, tells us

$$\Pr(X_T = 0 | X) = (1-q)^X,$$

because this is the probability that none of the bad events that occurred are included in T. Taking expectations on both sides,

$$\mathbb{E}[(1-q)^X] = \Pr(X_T = 0) \le e^{-\mu' + \Delta'/2}$$

by the first Janson inequality, where $\mu' = \mathbb{E}[X_T] = q\mu$ and Δ' (similarly) is $= q^2\Delta$. We can think of the left side as a moment generating function if we want to get exponential bounds!

So now by Markov's inequality,

$$\Pr(X \le \mu - t) = \Pr((1-q)^X \ge (1-q)^{\mu - t}) \le (1-q)^{-\mu+t}\mathbb{E}[(1-q)^X]$$

and plugging in our result,

$$\le (1-q)^{-\mu+t} e^{-q\mu' + q^2\frac{\Delta}{2}}.$$

Now we just optimize for q. One thing we can do is take the derivative and set things equal to 0, but instead, we can give ourselves some slack: let's let $1 - q = e^{-\lambda}$ for some $\lambda \ge 0$. By the Taylor expansion, $\lambda - \frac{\lambda^2}{2} \le q \le \lambda$. Plugging this in,

$$\Pr(X \ge \mu - t) \le \exp\left[\lambda(\mu - t) - \left(\lambda - \frac{\lambda^2}{2}\right) + \frac{\lambda^2\Delta}{2}\right] = \exp\left[-\lambda t + \frac{\lambda^2}{2}(\mu + \Delta)\right],$$

and now set $\lambda = \frac{1}{\mu + \Delta}$ to get the result. □

Notice that the proof of Janson's third inequality only works for lower tails: in particular, the proof starts by using the probability that $X = 0$ and builds up from there. In fact, upper tails are completely different! Let's show that for $p \gtrsim n^{-1/2}$,

$$\Pr(X \ge (1+c)\mathbb{E}[X]) \ge e^{-Cn^2p}.$$

To do this, we ask: what's the "cheapest" possible way to generate lots of triangles? If we plant a clique, we get something pretty good:

$$\Pr(X \ge 2\mathbb{E}[X]) \ge \Pr(G(n,p) \text{ has a clique on the first } 10np \text{ vertices}),$$

because this already gives $\binom{10np}{3}$ triangles, which is more than $2\binom{n}{3}p^2$. Then the probability $G(n,p)$ has a clique in the first $10np$ vertices is

$$p^{\binom{10np}{2}} \ge e^{-cn^2p^2\log(1/p)},$$

and this is exponentially larger than the previous bound for lower tails.

So what's the truth here? There was a paper written about the "infamous upper tail:" it just demonstrated that a wide range of techniques did not work for showing bounds on the upper tail. About 10 years ago, though, the following result (on the same order of the planted clique) was proved:

$$\Pr(X \geq 2\mathbb{E}[X]) = e^{-\Theta(n^2p^2\log(1/p))} \text{ if } p \gtrsim \frac{\log n}{n}.$$

The constant in the Θ here is a bit tricky, too. Basically, there are two main constructions for generating lots of triangles: have a hub of size cnp^2 and connect them to everything else, or have a clique of some select size. But this isn't obvious, and there's lots of open research here!

7.6 Revisiting clique numbers

Recall that in our Second Moment Method discussions, we found that

$$\omega\left(G\left(n, \frac{1}{2}\right)\right) \sim 2\log_2(n)$$

with high probability (this is Theorem 4.21). We actually found two-point concentration: let's review the ideas of the proof so that we can look at them more closely.

Proof. Let X_k be the number of k-cliques in $G\left(n, \frac{1}{2}\right)$: the expected value of X_k is then

$$\mu(k) = \binom{n}{k}2^{-\binom{k}{2}}.$$

If k is a function of n, and our mean $\mu_k \to 0$, then $X = 0$ with high probability by Markov's inequality. Meanwhile, if $\mu_k \to \infty$, then $\Delta \ll \mu^2$: variance is much smaller than the squared mean, so by the second moment method, there is always a k-clique with high probability due to concentration. $\square$

Where does the quantity μ_k cross a threshold? We defined k_0 to be the largest k such that $\mu_k \geq 1$. Another routine calculation shows us that around that value of $k \sim 2\log_2 n$,

$$\frac{\mu_{k+1}}{\mu_k} = n^{-1+o(1)}$$

and this implies that the clique number of $G\left(n, \frac{1}{2}\right)$ is concentrated around $2\log_2 n$ with high probability: in fact, it's either k_0 or $k_0 - 1$ with high probability.

Problem 7.17

Is the clique number of $G\left(n, \frac{1}{2}\right)$ really concentrated around two different points, or do we just have a weakness in our proof?

Solution. Let's say that we have k as a function of n so that $\mu = \binom{n}{k}2^{-\binom{k}{2}} \to c$ converges as $n, k \to \infty$. Doing the relevant calculations, we find that $\Delta = o(1)$, so $\Delta \ll \mu$ here. For all S that are k-element subsets of $[n]$, let A_S be the event that S forms a clique: then the probability for any given A_S is $2^{-\binom{k}{2}}$. Then $X = 0$, which is the probability that there are no k-cliques, is

$$\Pr\left(\omega\left(G\left(n, \frac{1}{2}\right)\right) < k\right) \leq e^{-\mu+\Delta/2}$$

by the first Janson inequality, and we also have $e^{-\mu(1+o(1))}$ as a lower bound by Harris. So because $\Delta \ll \mu$, the lower and upper bounds grow closer:

$$\Pr\left(\omega\left(G\left(n, \frac{1}{2}\right)\right) < k\right) = e^{-(1+o(1))\mu} \to e^{-c},$$

which is some specific value between 0 and 1. To be more rigorous about this, we're saying that for $\lambda \in (-\infty, \infty)$, if $n_0(k)$ is the minimum n such that $\binom{n}{k}2^{-\binom{k}{2}} \geq 1$, then $n = n_0(k)\left(1 + \frac{\lambda + o(1)}{k}\right)$, and thus the mean

$$\mu = \binom{n}{k}2^{-\binom{k}{2}} = e^{\lambda} + o(1).$$

So that means that

$$\Omega\left(G\left(n, \frac{1}{2}\right)\right) = \begin{cases} k-1 & \text{with probability } 1 - e^{-e^{\lambda}} + o(1) \\ k & \text{with probability } e^{-e^{\lambda}} + o(1) \end{cases}.$$

and we can get sequences with two-point concentration of any probability we want. □

Fact 7.18

On the other hand, most n do have one-point concentration, so both cases (one- and two-point concentration) do occur. One way to view this problem in general is that these situations look sort of Poisson in their distribution, and that's the right kind of case to use Janson's inequalities.

7.7 Revisiting chromatic numbers

We're ready to go back to thinking about this result from earlier in the class now. Remember that Corollary 4.23, we found that the independence number

$$\alpha\left(G\left(n, \frac{1}{2}\right)\right) \sim 2\log_2 n$$

with high probability, and because each color class is an independent set,

$$\chi(G) \geq \frac{n}{\alpha(G)} \geq (1 + o(1))\frac{n}{2\log_2(n)}$$

with high probability. This is a lower bound: the methods we had before didn't allow us to get an upper bound, but now we have the Janson inequalities.

Theorem 7.19

We have with high probability that

$$\chi\left(G\left(n, \frac{1}{2}\right)\right) \sim \frac{n}{2\log_2(n)}.$$

Proof. The idea is that we will use a "greedy coloring," since our goal is to show that we can color all of G using $(1 + o(1))\frac{n}{2\log_2 n}$ colors with high probability.

At each step of our strategy, we take out an independent set of size $(1 + o(1))2\log_2(n)$. Each time we do that, color all of those with one of the colors, and remove the independent set from our graph. We'll stop when we have $o\left(\frac{n}{\log n}\right)$ vertices left: at that point, we finish by just coloring with a different color for every vertex.

Why can we perform that step of removing an independent set of size $(1 + o(1))2\log_2(n)$? It is sufficient to show that with "very" high probability, every "not-too-small" subset of G has an independent set of size $(1 + o(1))2\log_2 n$: here "very high" means exponentially small.

Lemma 7.20

There exists $k \sim 2\log_2 n$ such that

$$\Pr\left(\omega\left(G\left(n, \frac{1}{2}\right)\right) < k\right) < e^{-n^{2-o(1)}}.$$

By extension, this is also true for the independence number of G.

Proof of lemma. Use the notation from the clique number calculations (specifically k_0). Let $k = k_0 - 3$ (so that we have a significantly smaller number of cliques): what's the probability that we don't have any k-cliques? This should be very small if there's high probability of k_0-clique! In particular, the mean of the number of k-cliques is (because we change by a factor of n each time we change k by one)

$$\mu_k > n^{3-o(1)}.$$

We can also calculate

$$\Delta \sim \mu^2 \frac{k^2}{n^2} > n^{4+o(1)} > \mu,$$

so by the second Janson inequality,

$$\Pr\left(\omega\left(G\left(n, \frac{1}{2}\right)\right) < k\right) \le e^{-\frac{\mu^2}{2\Delta}} = e^{-n^{2+o(1)}}$$

as desired. $\square$

This decays very quickly. So now given a graph $G \sim G\left(n, \frac{1}{2}\right)$, let's take $m = \left\lfloor \frac{n}{\log^2 n} \right\rfloor$: for every set of m vertices, let $G[S]$ be the graph induced by the vertices S. The probability the independence number of $G[S]$ is less than k is

$$e^{-m^{2-o(1)}} = e^{-n^{2-o(1)}},$$

because the $G[s]$ looks like $G\left(m, \frac{1}{2}\right)$ for some $k \sim 2\log_2 m \sim 2\log_2 n$. Summing over all S and doing a union bound, the probability that

$$\Pr(\alpha(G[S]) < k \text{ for some } S) < 2^n e^{-n^{2-o(1)}} = o(1),$$

so with high probability, every m-element subset of G contains a k-element independent set. Thus we can carry out our greedy coloring, and the total number of colors we use (including the last part where we use one color for each vertex) is

$$\frac{n-m}{k} + m = (1 + o(1))\frac{n}{2\log_2(n)},$$

as desired. $\square$

Note that this proof only works because we can get an exponential bound from the Janson inequalities! Bollobás' theorem guarantees some kind of concentration, but the window of deviation is still basically an open problem.