Probabilistic Combinatorics

These notes are to accompany the lectures in HT 2014 on Probabilistic Combinatorics for Part C Mathematics and Joint Schools, and for MFoCS. They are based on earlier notes from Oliver Riordan. Updates during the course: 13-02-2014 minor updates in chapters 1-3; 01-03-2014 minor updates in chapters 4-5 up to end L12 (Lemma 5.10).

Recommended books: For much of the course *The Probabilistic Method* (third edition, Wiley, 2008) by Alon and Spencer is the most accessible reference. Very good books containing a lot of material, especially about random graphs, are *Random Graphs* by Bollobás, and *Random Graphs* by Janson, Łuczak and Ruciński; but do not expect these books to be easy to read!

Contents

0 What is probabilistic combinatorics?

The first question is what is combinatorics! This is hard to define exactly, but should become clearer through examples, of which the main one are from graph theory.

Roughly speaking, combinatorics is the study of ‘discrete structures’. Here ‘discrete’ means either finite, or infinite but discrete in the sense that the integers are, as opposed to the reals. Usually in combinatorics, there are some underlying objects whose internal structure we ignore, and we study structures built on them: the most common example is graph theory, where we do not care what the vertices are, but study the abstract structure of graphs on a given set of vertices. Abstractly, a graph is just a set of unordered pairs of vertices, i.e., a symmetric irreflexive binary relation on its vertex set. More generally, we might study collections of general subsets of a given vertex set (not just pairs), for example.

Turning to probabilistic combinatorics, this is combinatorics with randomness involved. It can mean two things: (a) the use of randomness (e.g., random graphs) to solve deterministic combinatorial problems, or (b) the study of random combinatorial objects for their own sake. Historically, the subject started with (a), but after a while, the same objects (e.g., random graphs) come up again and again, and one realizes that it is not only important, but also interesting, to study these in themselves, as well as their applications. The subject has also led to new developments in probability theory, and interacts strongly with theoretical computer science.

The course will mainly be organized around proof techniques. However, each technique will be illustrated with examples, and one particular example (random graphs) will occur again and again, so by the end of the course we will have covered aim (b) in this special case as well as aim (a) above.

The first few examples will be mathematically very simple; nevertheless they will show the power of the method in general. Of course, modern applications are often not so simple.

1 First moment method

Perhaps the most basic inequality in probability is the *union bound*: if A_1 and A_2 are two events, then $\mathbb{P}(A_1 \vee A_2) \leqslant \mathbb{P}(A_1) + \mathbb{P}(A_2)$. (The notations $A_1 \vee A_2$ and $A_1 \cup A_2$ both mean A_1 or A_2.) More generally,

$$\mathbb{P}\big(A_1 \vee \cdots \vee A_n\big) \leqslant \sum_{i=1}^{n} \mathbb{P}(A_i).$$

This trivial fact is already useful.

Example (Ramsey numbers)**.** For positive integers k and ℓ, the *Ramsey number* $R(k, \ell)$ is the smallest n such that every red/blue colouring of the edges of the complete graph K_n contains either a red K_k or a blue K_ℓ. (These numbers are well-defined.)

Theorem 1.1 (Erdős, 1947)**.** *If $n, k \geqslant 1$ are integers such that $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$, then $R(k, k) > n$.*

Proof. Colour the edges of K_n red/blue at random so that each edge is red with probability 1/2 and blue with probability 1/2, and the colours of the edges are independent.

There are $\binom{n}{k}$ copies of K_k in K_n. Let A_i be the event that the ith copy is monochromatic. Then

$$\mathbb{P}(A_i) = 2\left(\frac{1}{2}\right)^{\binom{k}{2}} = 2^{1-\binom{k}{2}}.$$

Thus

$$\mathbb{P}(\exists \text{ monochromatic } K_k) \leqslant \sum_i \mathbb{P}(A_i) = \binom{n}{k} 2^{1-\binom{k}{2}} < 1.$$

Thus, in the random colouring, the probability that there is no monochromatic K_k is greater than 0. Hence it is *possible* that the random colouring is 'good' (contains no monochromatic K_k); that is, there must exist a 'good' colouring. □

To deduce an explicit bound on $R(k, k)$ involves a little calculation.

Corollary 1.2. *$R(k, k) \geqslant 2^{k/2}$ for $k \geqslant 3$.*

Proof. Set $n = \lfloor 2^{k/2} \rfloor$. Then

$$\binom{n}{k} 2^{1-\binom{k}{2}} \leqslant \frac{n^k}{k!} 2^{1-\binom{k}{2}} \leqslant \frac{2^{k^2/2}}{k!} 2^{1-k^2/2+k/2} = \frac{2^{1+k/2}}{k!},$$

which is smaller than 1 if $k \geqslant 3$. □

Remark. The result above is very simple, and may seem weak. But the best lower bound proved by non-random methods is roughly $2^{(\log k)^C}$ with C constant. This is *tiny* compared with the exponential bound above. Note that the known upper bounds are roughly 4^k, so exponential is the right order: the constant (if it exists) is unknown.

Often, the 'first-moment method' simply refers to using the union bound as above. But it is much more general than that. We recall another basic term from probability.

Definition. The *first moment* of a random variable X is simply its mean, or *expectation*, written $\mathbb{E}[X]$.

Recall that *expectation is linear.* If X and Y are (real-valued) random variables and λ is a (constant!) real number, then $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$, and $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X]$. Crucially, these ALWAYS hold, irrespective of any relationship (or not) between X and Y.

If A is an event, then its *indicator function* $\mathbb{I}_A$ is the random variable which takes the value 1 when A holds and 0 when A does not hold.

Let $A_1, \dots, A_n$ be events, let $\mathbb{I}_i$ denote the indicator function of A_i, and set $X = \sum_i \mathbb{I}_i$, so X is the (random) number of the events A_i that hold. Then

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[\mathbb{I}_i] = \sum_{i=1}^{n} \mathbb{P}(A_i).$$

We use the following observation about any random variable X with finite mean μ : it cannot be true that X is always smaller than μ, or always larger; that is $\mathbb{P}(X \geqslant \mu) > 0$ and $\mathbb{P}(X \leqslant \mu) > 0$.

Example (Ramsey numbers again)**.**

Theorem 1.3. *Let $n, k \geqslant 1$ be integers. Then*

$$R(k,k) > n - \binom{n}{k} 2^{1-\binom{k}{2}}.$$

Proof. Colour the edges of K_n as before. Let X denote the (random) number of monochromatic copies of K_k in the colouring. Then

$$\mu = \mathbb{E}[X] = \binom{n}{k} 2^{1-\binom{k}{2}}.$$

Since $\mathbb{P}(X \leqslant \mu) > 0$, there exists a colouring with at most μ monochromatic copies of K_k. Pick one vertex from each of these monochromatic K_ks – this may involve picking the same vertex more than once. Delete all the selected vertices. Then we have deleted at most μ vertices, and we are left with a 'good' colouring of K_m for some $m \geqslant n - \mu$. Thus $R(k,k) > m \geqslant n - \mu$. $\square$

The type of argument above is often called a 'deletion argument'. Instead of trying to avoid 'bad things' in our random structure, we first ensure that there are not too many, and then 'fix things' (here by deleting) to get rid of those few.

Corollary 1.4. $R(k,k) \geqslant (1-o(1))e^{-1}k2^{k/2}$.

Here we are using standard asymptotic notation. Explicitly, we mean that for any $\varepsilon > 0$ there is a k_0 such that $R(k,k) \geqslant (1-\varepsilon)e^{-1}k2^{k/2}$ for all $k \geqslant k_0$. (Theorem 1.1 does not quite yield this.)

Proof. Exercise: take $n = \lfloor e^{-1}k2^{k/2} \rfloor$. □

We now give a totally different example of the first-moment method.

Example (Sum-free sets)**.**

Definition. A set $S \subseteq \mathbb{R}$ is *sum-free* if there do not exist $a, b, c \in S$ such that $a + b = c$.

Note that $\{1, 2\}$ is *not* sum-free, since $1 + 1 = 2$. The set $\{2, 3, 7, 8, 12\}$ is sum-free, for example.

Theorem 1.5 (Erdős, 1965)**.** *Let $S = \{s_1, s_2, \ldots, s_n\}$ be a set of $n \geqslant 1$ (distinct) non-zero integers. There is some $A \subseteq S$ such that A is sum-free and $|A| > n/3$.*

Proof. We use a trick: we want a prime p not dividing any s_i, for example we may take $p > \max |s_i|$. There are infinitely many primes of the form $3k + 2$: we fix such a p not dividing any s_i. (A prime of the form $3k + 1$ works nearly as well.)

Let $I = \{k + 1, \ldots, 2k + 1\}$. Then I is *sum-free modulo* p: there do not exist $a, b, c \in I$ such that $a + b \equiv c \mod p$. (For if $a, b \in I$ then $2k + 2 \leq a + b \leq 4k + 2 = (3k + 2) + k$.) end L1 2014

Pick r uniformly at random from $1, 2, \ldots, p-1$, and set $t_i = rs_i \mod p$. Thus each t_i is a random element of $\{1, 2, \ldots, p-1\}$. For each fixed i, as r runs from 1 to $p - 1$, t_i takes each possible value $1, 2, \ldots, p-1$ exactly once: to see this note that no value can be repeated, since if $rs_i \equiv r's_i$ then $p|(r - r')s_i$ and so $p|(r - r')$. Hence

$$\mathbb{P}(t_i \in I) = \frac{|I|}{p-1} = \frac{k+1}{3k+1} > \frac{1}{3}.$$

We use the first moment method: we have

$$\mathbb{E}[\#i \text{ such that } t_i \in I] = \sum_{i=1}^{n} \mathbb{P}(t_i \in I) > n/3.$$

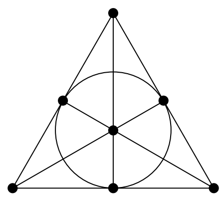
It follows that there is some r such that, for this particular r, the number of i with $t_i \in I$ is greater than $n/3$. For this r, let $A = \{s_i : t_i \in I\}$, so $A \subseteq S$ and $|A| > n/3$. If we had $s_i, s_j, s_k \in A$ with $s_i + s_j = s_k$ then we would have $rs_i + rs_j = rs_k$, and hence $t_i + t_j \equiv t_k \mod p$, which contradicts the fact that I is sum-free modulo p. □

The proof above is an example of an *averaging argument*. This particular example is not so easy to dream up but it is hopefully easy to follow.

Example (2-colouring hypergraphs)**.** A *hypergraph* H is simply an ordered pair (V, E) where V is a set of *vertices* and E is a set of *edges* (or *hyperedges*), i.e., a set of subsets of V.

Note that E is a *set*, so each possible edge (subset of V) is either present or not, just as each possible edge of a graph is either present or not. If we wanted to allow multiple copies of the same edge, we could define *multi-hypergraphs* in analogy with *multigraphs*.

H is r*-uniform* if $|e| = r$ for all $e \in E$, i.e., if every edge consists of exactly r vertices. In particular, a 2-uniform hypergraph is simply a graph.



An example of a 3-uniform hypergraph is the *Fano plane* shown in the figure. This has 7 vertices and 7 edges; in the drawing, the 6 straight lines and the circle each represent an edge. (As usual, how they are drawn is irrelevant, all that matters is which vertices each hyperedge contains.)

A *(proper) 2-colouring* of a hypergraph H is a red/blue colouring of the vertices such that every edge contains vertices of both colours. If H is 2-uniform, this is the same as a proper (vertex) 2-colouring of H as a graph. We say that H is *2-colourable* if it has a 2-colouring. (This was once called having *property B*.)

Let $m(r)$ be the minimum m such that there exists a non 2-colourable r-uniform hypergraph with m edges. The Fano plane is not 2-colourable (exercise), and so $m(3) \leq 7$. It is easy to check that $m(2) = 3$. It is harder to check that $m(3) = 7$ (there is no need to do this!).

Theorem 1.6. *For $r \geqslant 2$ we have $m(r) \geqslant 2^{r-1}$.*

Proof. Let $H = (V, E)$ be any r-uniform hypergraph with $m < 2^{r-1}$ edges. Colour the vertices red and blue randomly: each red with probability $1/2$ and blue with probability $1/2$, with different vertices coloured independently. For any $e \in E$, the probability that e is monochromatic is $2(1/2)^r = (1/2)^{r-1}$. By the union bound, it follows that the probability that there is at least one monochromatic edge is at most $m(1/2)^{r-1} < 1$. Thus there exists a 'good' colouring. □

We can also obtain a bound in the other direction; this is slightly harder.

Theorem 1.7 (Erdős, 1964)**.** *If r is large enough then $m(r) \leqslant 3r^2 2^r$.*

Proof. Fix $r \geqslant 3$. Let V be a set of n vertices, where n (which depends on r) will be chosen later. Let $m = 3r^2 2^r$.

Let $e_1, \ldots, e_m$ be chosen independently and uniformly at random from all $\binom{n}{r}$ possible hyperedges on V. Although repetitions are possible, the hypergraph

$$H = (V, \{e_1, \ldots, e_m\})$$

certainly has *at most* m hyperedges.

Let χ be any red/blue colouring of V (*not* a random one this time). Then χ has either at least $n/2$ red vertices, or at least $n/2$ blue ones. It follows that at least (crudely) $\binom{\lceil n/2 \rceil}{r}$ of the possible hyperedges are monochromatic with respect to χ.

Let $p = p_\chi$ denote the probability that e_1 (a hyperedge chosen uniformly at random from all possibilities) is monochromatic with respect to χ. Then

$$\begin{aligned} p \geqslant \frac{\binom{\lceil n/2 \rceil}{r}}{\binom{n}{r}} &\geq \frac{(n/2)(n/2-1)\cdots(n/2-r+1)}{n(n-1)\cdots(n-r+1)} \\ &\geqslant \left(\frac{n/2-r+1}{n-r+1}\right)^r \geqslant \left(\frac{n/2-r}{n-r}\right)^r = 2^{-r}\left(1 - \frac{r}{n-r}\right)^r. \end{aligned}$$

Set $n = r^2$. Then $p \geqslant 2^{-r}(1 - 1/(r-1))^r$. Since $(1 - 1/(r-1))^r \to e^{-1}$ as $r \to \infty$, we see that $p \geqslant p_0 := \frac{1}{3 \cdot 2^r}$ if r is large enough, which we assume from now on.

The probability that the given, fixed colouring χ is a proper 2-colouring of our random hypergraph H is simply the probability that none of $e_1, \ldots, e_m$ is monochromatic with respect to χ. Since $e_1, \ldots, e_m$ are independent, this probability is $(1-p)^m \leq (1-p_0)^m$.

By the union bound, the probability that H is 2-colourable is at most the sum over all possible χ of the probability that χ is a 2-colouring, which is at most $2^n(1-p_0)^m$. Using the standard inequality $1-x \leqslant e^{-x}$, we have

$$2^n(1-p_0)^m \leqslant 2^n e^{-p_0 m} \leqslant 2^{r^2} e^{-\frac{3r^2 2^r}{3 \cdot 2^r}} = 2^{r^2} e^{-r^2} < 1.$$

Therefore there exists an r-uniform hypergraph H with at most m edges and no 2-colouring. $\square$

Remark. Why does the first moment method work? Often, there is some complicated event A whose probability we want to know or at least bound. For example, A might be the event that the random colouring χ is a 2-colouring of a fixed (complicated) hypergraph H. Often, A is constructed by taking the union or intersection of simple events $A_1, \ldots, A_k$.

- If $A_1, \ldots, A_k$ are independent, then

$$\mathbb{P}(A_1 \wedge \cdots \wedge A_k) = \prod_i \mathbb{P}(A_i) \quad \text{and} \quad \mathbb{P}(A_1 \vee \cdots \vee A_k) = 1 - \prod_i (1 - \mathbb{P}(A_i)).$$

- If $A_1, \ldots, A_k$ are mutually exclusive, then

$$\mathbb{P}(A_1 \vee \cdots \vee A_k) = \sum_i \mathbb{P}(A_k).$$

(For example, these give us the probability $2(1/2)^{|e|}$ that a fixed hyperedge e is monochromatic in a random 2-colouring of the vertices.)

In general, the relationship between the A_i may be very complicated. However, if X is the number of A_i that hold, then we *always* have $\mathbb{E}[X] = \sum_i \mathbb{P}(A_i)$ and

$$\mathbb{P}(\bigvee_i A_i) = \mathbb{P}(X > 0) \leqslant \sum_i \mathbb{P}(A_i).$$

The key point is that while the left-hand side is complicated, the right-hand side is simple: we evaluate it by looking at one simple event at a time.

So far we have used the expectation only via the observations that $\mathbb{P}(X \leqslant \mathbb{E}[X]) > 0$ and $\mathbb{P}(X \geqslant \mathbb{E}[X]) > 0$, together with the union bound. A slightly more sophisticated (but still simple) way to use it is via Markov's inequality.

Lemma 1.8 (Markov's inequality)**.** *If X is a random variable taking only non-negative values and $t > 0$, then $\mathbb{P}(X \geqslant t) \leqslant \mathbb{E}[X]/t$.*

Proof. Always $X \geq tI_{X \geq t}$. Take expectations. $\square$

end L2 2014

We now start on one of our main themes, the study of the random graph $G(n,p)$.

Definition. Given an integer $n \geqslant 1$ and a real number $0 \leqslant p \leqslant 1$, the *random graph* $G(n,p)$ is the graph with vertex set $[n] = \{1, 2, \ldots, n\}$ in which each of the $\binom{n}{2}$ possible edges ij, $1 \leqslant i < j \leqslant n$, is present with probability p, independently of the others.

Thus, for any graph H on $[n]$,

$$\mathbb{P}\big(G(n,p) = H\big) = p^{e(H)}(1-p)^{\binom{n}{2}-e(H)}.$$

For example, if $p = 1/2$, then all $2^{\binom{n}{2}}$ graphs on $[n]$ are equally likely.

Remark. It is important to remember that we work with 'labelled' graphs. For example, the probability that $G(3,p)$ is a path with three vertices is $3p^2(1-p)$, since there are three (isomorphic) graphs on $\{1,2,3\}$ that are paths.

We use the notation $\mathcal{G}(n,p)$ for the probability space of graphs on $[n]$ with the probabilities above. All of $G \in \mathcal{G}(n,p)$, $G = G(n,p)$ and $G \sim G(n,p)$ mean exactly the same thing, namely that G is a random graph with this distribution.

This model of random graphs is often called the Erdős–Rényi model (or binomial model), although in fact it was first defined by Gilbert. Erdős and Rényi introduced an essentially equivalent model, and were the real founders of the theory of random graphs, so associating the model with their names is reasonable!

Example (High girth and chromatic number)**.** Let us recall some definitions. The *girth* $g(G)$ of a graph G is the minimum length of a cycle in G, or ∞ if G contains no cycles. The *chromatic number* $\chi(G)$ is the least k such that G has a *proper* k-colouring (that is, a colouring of the vertices with k colours in which adjacent vertices receive different colours); and the *independence number* $\alpha(G)$ is the maximum number of vertices in an independent set in G, i.e., a set of vertices of G no two of which are joined by an edge.

Since a proper k-colouring partitions the vertex set into k independent sets, $v(G) \le k\,\alpha(G)$, and so

$$\chi(G) \geqslant v(G)/\alpha(G).$$

Theorem 1.9 (Erdős, 1959)**.** *For any k and ℓ there exists a graph G with $\chi(G) \geqslant k$ and $g(G) \geqslant \ell$.*

There are non-random proofs of this, but it is not so easy.

The idea of the proof is to consider $G(n,p)$ for suitable n and p. We will show *separately* that (a) very likely there are few short cycles, and (b) very likely there is no large independent set. Then it is likely that the properties in (a) and (b) *both* hold, and after deleting a few vertices (to kill the short cycles), we obtain the graph we need.

Proof. Fix $k, \ell \geqslant 3$. For $r \ge 3$, there are

$$\frac{n(n-1)\cdots(n-r+1)}{2r}$$

possible cycles of length r in $G(n,p)$: the numerator counts sequences of r distinct vertices, and the denominator accounts for the fact that each cycle corresponds to $2r$ sequences, depending on the choice of starting point and direction.

Let X_r be the number of r-cycles in $G(n,p)$. Then

$$\mathbb{E}[X_r] = \frac{n(n-1)\cdots(n-r+1)}{2r} p^r \leqslant \frac{n^r p^r}{2r}.$$

Set $p = p(n) = n^{-1+1/\ell}$, and let X be the number of 'short' cycles, i.e., cycles with length less than ℓ. Then $X = X_3 + X_4 + \cdots + X_{\ell-1}$, so

$$\mathbb{E}[X] = \sum_{r=3}^{\ell-1} \mathbb{E}[X_r] \leqslant \sum_{r=3}^{\ell-1} \frac{(np)^r}{2r} = \sum_{r=3}^{\ell-1} \frac{n^{r/\ell}}{2r} = O(n^{\frac{\ell-1}{\ell}}) = o(n).$$

By Markov's inequality it follows that

$$\mathbb{P}(X \geqslant n/2) \leqslant \frac{\mathbb{E}[X]}{n/2} \to 0 \quad (\text{ as } \; n \to \infty).$$

Set $m = m(n) = \lfloor n^{1-1/(2\ell)} \rfloor$. Let Y be the number of independent sets in $G(n,p)$ of size (exactly) m. Then, using bounds from problem set 0,

$$\mathbb{E}[Y] = \binom{n}{m}(1-p)^{\binom{m}{2}} \leqslant \left(\frac{en}{m}\right)^m e^{-p\binom{m}{2}} = \left(\frac{en}{m} e^{-p\frac{m-1}{2}}\right)^m.$$

Now

$$p\frac{m-1}{2} \sim \frac{pm}{2} \sim \frac{n^{-1+\frac{1}{\ell}}n^{1-\frac{1}{2\ell}}}{2} = \frac{n^{\frac{1}{2\ell}}}{2}.$$

Thus $p(m-1)/2 \geqslant 2\log n$ if n is large enough, which we may assume. But then

$$\mathbb{E}[Y] \leqslant \Big(\frac{en}{m}n^{-2}\Big)^m \to 0,$$

and by Markov's inequality we have $\mathbb{P}(Y \geqslant 1) \leqslant \mathbb{E}[Y] \to 0$; that is, $\mathbb{P}(\alpha(G) \geqslant m) \to 0$.

Combining the two results above, we have $\mathbb{P}(X \geqslant n/2 \text{ or } \alpha(G) \geqslant m) \to 0$. Hence, if n is large enough, there exists some graph G with n vertices, with fewer than $n/2$ short cycles, and with $\alpha(G) < m$.

Construct G^* by deleting one vertex from each short cycle of G. Then $g(G^*) \geqslant \ell$, and $v(G^*) \geqslant n - n/2 = n/2$. Also, $\alpha(G^*) \leqslant \alpha(G) < m$. Thus

$$\chi(G^*) \geqslant \frac{v(G^*)}{\alpha(G^*)} \geqslant \frac{n/2}{m} \geqslant \frac{n/2}{n^{1-\frac{1}{2\ell}}} = \frac{1}{2}n^{\frac{1}{2\ell}},$$

which is larger than k if n is large enough. □

2 Second moment method

Suppose (X_n) is a sequence of random variables, each taking non-negative integer values. By Markov's inequality, if $\mathbb{E}[X_n] \to 0$ as $n \to \infty$, then we have $\mathbb{P}(X_n > 0) = \mathbb{P}(X_n \geqslant 1) \leqslant \mathbb{E}[X_n] \to 0$. Under what conditions can we show that $\mathbb{P}(X_n > 0) \to 1$? Simply $\mathbb{E}[X_n] \to \infty$ is *not* enough: it is easy to find examples where $\mathbb{E}[X_n] \to \infty$, but $\mathbb{P}(X_n = 0) \to 1$. We want some control on the difference between X_n and $\mathbb{E}[X_n]$.

Definition. The *variance* $\mathrm{Var}[X]$ of a random variable X is defined by

$$\mathrm{Var}[X] = \mathbb{E}\big[(X - \mathbb{E}X)^2\big] = \mathbb{E}\big[X^2\big] - \big(\mathbb{E}X\big)^2.$$

(We assume that $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ are finite.) We recall a basic fact from probability.

Lemma 2.1 (Chebyshev's Inequality)**.** *Let X be a random variable and let $t > 0$. Then*

$$\mathbb{P}\big(|X - \mathbb{E}X| \geqslant t\big) \leqslant \frac{\mathrm{Var}[X]}{t^2}.$$

Proof. By Markov's inequality applied to $Y = (X - \mathbb{E}X)^2$ we have

$$\mathbb{P}\big(|X - \mathbb{E}X| \geqslant t\big) = \mathbb{P}\big(Y \geqslant t^2\big) \leq \frac{\mathbb{E}[Y]}{t^2} = \frac{\mathrm{Var}[X]}{t^2}.$$

$\square$

In practice, we usually use this as follows.

Corollary 2.2. *Let (X_n) be a sequence of random variables with $\mathbb{E}[X_n] = \mu_n > 0$ and $\mathrm{Var}[X_n] = o(\mu_n^2)$. Then $\mathbb{P}(X_n = 0) \to 0$.*

Proof.

$$\mathbb{P}(X_n = 0) \leqslant \mathbb{P}\big(|X_n - \mu_n| \geqslant \mu_n\big) \leqslant \frac{\mathrm{Var}[X_n]}{\mu_n^2} \to 0.$$

$\square$

Remark. The mean $\mu = \mathbb{E}[X]$ is usually easy to calculate. Since $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mu^2$, this means that knowing the variance is like knowing the *second moment* $\mathbb{E}[X^2]$. In particular, with $\mu_n = \mathbb{E}[X_n]$, the condition $\mathrm{Var}[X_n] = o(\mu_n^2)$ is equivalent to $\mathbb{E}[X_n^2] = (1 + o(1))\mu_n^2$, i.e., $\mathbb{E}[X_n^2] \sim \mu_n^2$:

$$\mathrm{Var}[X_n] = o(\mu_n^2) \iff \mathbb{E}[X_n^2] \sim \mu_n^2.$$

Sometimes the second moment is more convenient to calculate than the variance.

Suppose that $X = I_1 + \cdots + I_k$, where each I_i is the indicator function of some event A_i. We have seen that $\mathbb{E}[X]$ is easy to calculate; $\mathbb{E}[X^2]$ is not too much harder:

$$\mathbb{E}[X^2] = \mathbb{E}\big[\sum_i I_i \sum_j I_j\big] = \mathbb{E}\big[\sum_i \sum_j I_i I_j\big] = \sum_i \sum_j \mathbb{E}[I_i I_j] = \sum_{i=1}^k \sum_{j=1}^k \mathbb{P}(A_i \cap A_j).$$

end L3 2014

Example (K_4s in $G(n,p)$)**.**

Theorem 2.3. *Let $p = p(n)$ be a function of n.*

1. *If $n^{2/3}p \to 0$ as $n \to \infty$, then $\mathbb{P}(G(n,p)$ contains a $K_4) \to 0$.*
2. *If $n^{2/3}p \to \infty$ as $n \to \infty$, then $\mathbb{P}(G(n,p)$ contains a $K_4) \to 1$.*

Proof. Let X (really X_n, as the distribution depends on n) denote the number of K_4s in $G(n,p)$. For each set S of 4 vertices from $[n]$, let A_S be the event that S induces a K_4 in $G(n,p)$. Then

$$\mu = \mathbb{E}[X] = \sum_S \mathbb{P}(A_S) = \binom{n}{4}p^6 = \frac{n(n-1)(n-2)(n-3)}{4!}p^6 \sim \frac{n^4p^6}{24}.$$

In case 1 it follows that $\mathbb{E}[X] \to 0$, so $\mathbb{P}(X > 0) \to 0$, as required.

For the second part of the result, we have $\mathbb{E}[X^2] = \sum_S \sum_T \mathbb{P}(A_S \cap A_T)$. The contributions from all terms where S and T meet in a given number of vertices are as follows:

$\lvert S \cap T \rvert$	contribution
0	$\binom{n}{4}\binom{n-4}{4}p^{12} \sim \frac{n^4}{24}\frac{n^4}{24}p^{12} \sim \mu^2$
1	$\binom{n}{4}4\binom{n-4}{3}p^{12} = \Theta(n^7p^{12})$
2	$\binom{n}{4}\binom{4}{2}\binom{n-4}{2}p^{11} = \Theta(n^6p^{11})$
3	$\binom{n}{4}\binom{4}{3}\binom{n-4}{1}p^{9} = \Theta(n^5p^{9})$
4	$\binom{n}{4}p^6 = \mu$

Recall that by assumption $n^4p^6 \to \infty$, so $\mu \to \infty$ and the last contribution μ is $o(\mu^2)$. How do the other contributions compare to μ^2? Firstly, since $\mu^2 = \Theta(n^8p^{12})$, we have $n^7p^{12} = o(\mu^2)$. For the others, we have

$$\frac{n^6p^{11}}{n^8p^{12}} = \frac{1}{n^2p} = o(1)$$

and

$$\frac{n^5p^9}{n^8p^{12}} = \frac{1}{(np)^3} = o(1).$$

Putting this all together, $\mathbb{E}[X^2] = \mu^2 + o(\mu^2)$, so $\mathrm{Var}[X] = o(\mu^2)$, and by Corollary 2.2 we have $\mathbb{P}(X = 0) \to 0$. □

Definition. Let $\mathcal{P}$ be a property of graphs (e.g., 'contains a K_4'). A function $p^*(n)$ is called a *threshold function* for $\mathcal{P}$ in the model $G(n,p)$ if

- $p(n) = o(p^*(n))$ implies that $\mathbb{P}(G(n,p(n))$ has $\mathcal{P}) \to 0$, and

- $p(n)/p^*(n) \to \infty$ implies that $\mathbb{P}(G(n,p(n)) \text{ has } \mathcal{P}) \to 1$.

Theorem 2.3 says that $n^{-2/3}$ is a threshold function for $G(n,p)$ to contain a K_4. Note that threshold functions are not quite uniquely defined (e.g., $2n^{-2/3}$ is also one). (Call a property *increasing* if whenever $G = (V,E)$ has the property then so does each graph $G' = (V,E')$ with $E \subseteq E'$. Every increasing property has a threshold function.)

Suppose as usual that $X = I_1 + \ldots + I_k$, with I_i the indicator function of A_i. When applying the second moment method, our aim is to estimate the variance, showing that it is small compared to the square of the mean, so Corollary 2.2 applies. So far we first calculated $\mathbb{E}[X^2]$, due to the simplicity of the formula $\sum_i \sum_j \mathbb{P}(A_i \cap A_j)$. However, this involves some 'unnecessary' work when many of the events are independent. We can avoid this by directly calculating the variance.

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
&= \sum_i \sum_j \mathbb{P}(A_i \cap A_j) - \Big(\sum_i \mathbb{P}(A_i)\Big)\Big(\sum_j \mathbb{P}(A_j)\Big) \\
&= \sum_i \sum_j \big(\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j)\big).
\end{aligned}$$

Write $i \sim j$ if $i \neq j$ and A_i and A_j are dependent. (More precisely, we ensure that if $i \neq j$ and $i \not\sim j$ then A_i and A_j must be independent.) The contribution from terms where A_i and A_j are independent is zero by definition, so

$$\begin{aligned}
\mathrm{Var}[X] &= \sum_i \big(\mathbb{P}(A_i) - \mathbb{P}(A_i)^2\big) + \sum_i \sum_{j \sim i} \big(\mathbb{P}(A_i \cap A_j) - \mathbb{P}(A_i)\mathbb{P}(A_j)\big) \\
&\leqslant \mathbb{E}[X] + \sum_i \sum_{j \sim i} \mathbb{P}(A_i \cap A_j).
\end{aligned}$$

Note that the second last line is an *exact* formula for the variance: the last line is just an upper bound, but this upper bound is often good enough.

The bound above gives another standard way of applying the 2nd moment method. We suppress the dependence on n in the notation here.

Corollary 2.4. *If $\mu := \mathbb{E}[X] \to \infty$ and $\Delta := \sum_i \sum_{j \sim i} \mathbb{P}(A_i \cap A_j) = o(\mu^2)$, then $\mathbb{P}(X > 0) \to 1$.*

Proof. We have
$$\frac{\mathrm{Var}[X]}{\mu^2} \leqslant \frac{\mu + \Delta}{\mu^2} = \frac{1}{\mu} + \frac{\Delta}{\mu^2} \to 0.$$
Now apply Chebyshev's inequality in the form of Corollary 2.2. $\square$

Definition. An *isomorphism* between graphs G and H is a bijection $\phi : V(G) \to V(H)$ such that $ij \in E(G)$ if and only if $\phi(i)\phi(j) \in E(H)$. An *automorphism* of H is an isomorphism between H and itself; and $\mathrm{aut}(H)$ denotes the number of automorphisms of H.

For example the path P_3 with 3 vertices has $\mathrm{aut}(P_3) = 2$.

Example (Appearance of H in $G(n,p)$)**.** Fix a graph H with v vertices and e edges. What is the threshold for copies of H to appear in $G = G(n,p)$?

Let X be the number of copies of H in G. (We count the number of pairs (W, F) where $W \subseteq V(G)$, $F \subseteq E(G)$, and the graph (W, F) is isomorphic to H.) For example, if H is P_3, then $\mathbb{E}[X] = n(n-1)(n-2)/2\ p^2$.

In general, there are $n(n-1)\cdots(n-v+1)$ injective maps $\phi : V(H) \to [n]$. Suppose that for $i = 1, 2$ we have a map $\phi_i : V(H) \to W$ that is an isomorphism between H and (W, F_i). Then $F_1 = F_2$ iff $\phi_1^{-1} \circ \phi_2$ is an automorphism γ of H; that is, if and only if $\phi_2 = \phi_1 \circ \gamma$. Thus if $\gamma_1, \ldots, \gamma_k$ are the automorphisms of H, then the maps that give the same copy of H as ϕ_1 are $\phi_1 \circ \gamma_1, \ldots, \phi_1 \circ \gamma_k$. Thus there are

$$\frac{n(n-1)\cdots(n-v+1)}{\mathrm{aut}(H)}$$

possible copies of H. It follows that

$$\mathbb{E}[X] = \frac{n(n-1)\cdots(n-v+1)}{\mathrm{aut}(H)} p^e \sim \frac{n^v p^e}{\mathrm{aut}(H)} = \Theta(n^v p^e).$$

This *suggests* that the threshold should be $p = n^{-v/e}$. end L4 2014

This worked for K_4 but can it be right in general? Consider for example H to be a K_4 with an extra edge hanging off, so $v = 5$ and $e = 7$. Our proposed threshold is $p = n^{-5/7}$, which is much smaller than $p = n^{-2/3}$. Consider the range in between, where $p/n^{-5/7} \to \infty$ but $p/n^{-2/3} \to 0$. Then $\mathbb{E}[X] \to \infty$, but the probability that $G(n,p)$ contains a K_4 tends to 0, so the probability that $G(n,p)$ contains a copy of H tends to 0. The problem is that H contains a subgraph K_4 which is hard to find, because its e/v ratio is larger than that of H.

Definition. The *edge density* $d(H)$ of a graph H is $e(H)/v(H)$, that is, 1/2 times the average degree of H.

Definition. H is *balanced* if each subgraph H' of H has $d(H') \leqslant d(H)$, and *strictly balanced* if each subgraph $H' \neq H$ has $d(H') < d(H)$.

Examples of strictly balanced graphs are complete graphs, trees, and connected regular graphs. Examples of balanced but not strictly balanced graphs are disconnected regular graphs, or a triangle with an extra edge attached (making four vertices and four edges).

For balanced graphs, $p = n^{-v/e}$ does turn out to be the threshold.

Theorem 2.5. *Let H be a balanced graph with $v(H) = v$ and $e(G) = e$. Then $p(n) = n^{-v/e}$ is a threshold function for the property of containing a copy of H.*

Proof. Let X denote the number of copies of H in $G(n,p)$, and set $\mu = \mathbb{E}X$, so $\mu = \Theta(n^v p^e)$. If $p/n^{-v/e} \to 0$ then $\mu \to 0$, so $\mathbb{P}(X \geqslant 1) \to 0$.

Suppose that $p/n^{-v/e} \to \infty$, i.e., that $n^v p^e \to \infty$. We must show that $\mathbb{P}(X \geqslant 1) \to 1$.

Let $H_1, \ldots, H_N$ list all possible copies of H with vertices in $[n]$, and let A_i denote the event that the ith copy H_i is present in $G = G(n,p)$. Let $H_i \cap H_j$ denote the graph with vertex set $V(H_i) \cap V(H_j)$ (when this is nonempty) and edge set $E(H_i) \cap E(H_j)$. Observe that A_i and A_j are dependent if and only if $e(H_i \cap H_j) > 0$.

$$\Delta := \sum_i \sum_{j \sim i} \mathbb{P}(A_i \cap A_j) = \sum_i \sum_{j \sim i} \mathbb{P}(E(H_i) \cup E(H_j) \subseteq E(G)).$$

We split the sum by the number r of vertices of $H_i \cap H_j$ ($2 \leq r \leq v$) and the number s of edges of $H_i \cap H_j$. Note that $H_i \cap H_j$ is a subgraph of H_i, which is isomorphic to the balanced graph H. We thus have

$$\frac{s}{r} = d(H_i \cap H_j) \leqslant d(H) = \frac{e}{v},$$

so $s \leqslant re/v$.

The contribution to Δ from terms with given r and s is

$$\Theta\big(n^{2v-r} p^{2e-s}\big) = \Theta\big(\mu^2/(n^r p^s)\big).$$

Now

$$n^r p^s \geqslant n^r p^{re/v} = (n^v p^e)^{r/v} = \Theta(\mu^{r/v}).$$

Since $\mu \to \infty$ and $r/v > 0$, it follows that $n^r p^s \to \infty$, so the contribution from this pair (r,s) is $o(\mu^2)$.

Since there are only a fixed number of pairs to consider, it follows that $\Delta = o(\mu^2)$. Hence by Corollary 2.4, $\mathbb{P}(X > 0) \to 1$. $\square$

Remark. In general the threshold is $n^{-1/d(H')}$, where H' is a densest subgraph of H. The proof is almost the same.

Remark. If H is strictly balanced and $p = cn^{-v/e}$, then μ tends to a constant and the rth *factorial moment* $\mathbb{E}_r[X] = \mathbb{E}[X(X-1)\cdots(X-r+1)]$ satisfies $\mathbb{E}_r[X] \sim \mu^r$, from which one can show that the number of copies of H has asymptotically a Poisson distribution. We shall not do this.

3 Lovász Local Lemma

Suppose that we have some 'bad' events $A_1, \dots, A_n$, and we want to show that it's *possible* that no A_i holds, no matter how unlikely. If $\sum_i \mathbb{P}(A_i) < 1$ then the union bound gives what we want. What about if the sum is large? In general of course it might be that $\bigvee_i A_i$ always holds. One trivial case where we can rule this out is when the A_i are independent. Then

$$\mathbb{P}\left(\bigwedge_i A_i^{\mathrm{c}}\right) = \prod_i \mathbb{P}(A_i^{\mathrm{c}}) = \prod_{i=1}^{n}(1 - \mathbb{P}(A_i)) > 0,$$

provided each A_i has probability less than 1.

What if each A_i depends only on *a few* others?

Recall that $A_1, \dots, A_n$ are *independent* if for all disjoint S, $T \subseteq [n]$ we have

$$\mathbb{P}\left(\bigwedge_{i\in S} A_i \wedge \bigwedge_{i\in T} A_i^{\mathrm{c}}\right) = \prod_{i\in S}\mathbb{P}(A_i)\prod_{i\in T}\mathbb{P}(A_i^{\mathrm{c}}).$$

(If $S = \emptyset$ then $\wedge_{i\in S} A_i$ is the whole probability space Ω, and $\mathbb{P}(\wedge_{i\in S} A_i) = 1$.) This is not the same as each pair of events being independent (see below).

Definition. An event A is *independent of a family* $(B_1, \dots, B_n)$ *of events* if, for all disjoint S, $T \subseteq [n]$, we have $\mathbb{P}(A \wedge E) = \mathbb{P}(A)\,\mathbb{P}(E)$ where

$$E = \bigwedge_{i\in S} B_i \wedge \bigwedge_{i\in T} B_i^{\mathrm{c}}.$$

This is equivalent to requiring that for all such events E with $\mathbb{P}(E) > 0$ we have $\mathbb{P}(A|E) = \mathbb{P}(A)$; that is, knowing that certain B_i hold and certain others do not does not affect the probability that A holds.

For example, suppose that each of the following four sequences of coin tosses happens with probability 1/4: TTT, THH, HTH and HHT. Let A_i be the event that the ith toss is H. Then one can check that any two events

A_i are independent, but (A_1, A_2, A_3) is not a family of independent events. Similarly, A_1 is *not* independent of $\{A_2, A_3\}$, since $\mathbb{P}(A_1 \mid A_2 \cap A_3) = 0$.

Recall that a *digraph* on a vertex set V is a set of ordered pairs of distinct elements of V, i.e., a 'graph' in which each edge has an orientation, there are no loops, and there is at most one edge from a given i to a given j, but we may have edges in both directions. We write $i \to j$ if there is an edge from i to j.

Definition. A digraph D on $[n]$ is called a *dependency digraph* for the events $A_1, \ldots, A_n$ if for each i the event A_i is independent of the family of events $(A_j : j \neq i, i \not\to j)$.

Roughly speaking, A_i is 'allowed to depend on A_j when $i \to j$'. More precisely, A_i must be independent of the remaining A_j *as a family*, not just individually. end L5 2014

Theorem 3.1 (Local Lemma, general form)**.** *Let D be a dependency digraph for the events $A_1, \ldots, A_n$. Suppose that there are real numbers $0 \leq x_i < 1$ such that*

$$\mathbb{P}(A_i) \leqslant x_i \prod_{j\,:\,i\to j} (1 - x_j)$$

for each i. Then

$$\mathbb{P}\left(\bigwedge_{i=1}^{n} A_i^{\mathrm{c}}\right) \geqslant \prod_{i=1}^{n} (1 - x_i).$$

In particular, the probability that no event A_i occurs is > 0.

Proof. We *claim* that for any proper subset S of $[n]$ and any $i \notin S$ we have $\mathbb{P}(\bigwedge_{j\in S} A_j^{\mathrm{c}}) > 0$ and

$$\mathbb{P}\left(A_i^{\mathrm{c}} \mid \bigwedge_{j\in S} A_j^{\mathrm{c}}\right) \geqslant 1 - x_i, \tag{1}$$

that is

$$\mathbb{P}\left(A_i \mid \bigwedge_{j\in S} A_j^{\mathrm{c}}\right) \leqslant x_i. \tag{2}$$

Assuming the claim, then

$$\begin{aligned}
\mathbb{P}\left(\bigwedge_{i=1}^{n} A_i^{\mathrm{c}}\right) &= \mathbb{P}(A_1^{\mathrm{c}})\mathbb{P}(A_2^{\mathrm{c}} \mid A_1^{\mathrm{c}})\mathbb{P}(A_3^{\mathrm{c}} \mid A_1^{\mathrm{c}} \cap A_2^{\mathrm{c}}) \cdots \mathbb{P}(A_n^{\mathrm{c}} \mid \bigwedge_{i=1}^{n-1} A_i^{\mathrm{c}}) \\
&\geqslant (1-x_1)(1-x_2)(1-x_3)\cdots(1-x_n) \\
&= \prod_{i=1}^{n}(1-x_i).
\end{aligned}$$

It remains to prove the claim. For this we use induction on $|S|$.

For the base case $|S| = 0$ we have $\mathbb{P}(\bigwedge_{j\in S} A_j^{\mathrm{c}}) = 1 > 0$ and

$$\mathbb{P}\left(A_i \mid \bigwedge_{j\in S} A_j^{\mathrm{c}}\right) = \mathbb{P}(A_i) \leqslant x_i \prod_{j\,:\,i\to j}(1-x_j) \leqslant x_i,$$

as required.

Now let $1 \leq r \leq n-1$ and suppose the claim holds whenever $|S| < r$; and consider S with $|S| = r$, and $i \notin S$. Observe first that $\mathbb{P}(\bigwedge_{j\in S} A_j^{\mathrm{c}}) > 0$; for if $j_0 \in S$ and $S' = S \setminus \{j_0\}$ then by the induction hypothesis $\mathbb{P}(\bigwedge_{j\in S'} A_j^{\mathrm{c}}) > 0$ and

$$\mathbb{P}(\bigwedge_{j\in S} A_j^{\mathrm{c}}) = \mathbb{P}(A_{j_0}^{\mathrm{c}} \mid \bigwedge_{j\in S'} A_j^{\mathrm{c}})\,\mathbb{P}(\bigwedge_{j\in S'} A_j^{\mathrm{c}}) \geq (1-x_{j_0})\,\mathbb{P}(\bigwedge_{j\in S'} A_j^{\mathrm{c}}) > 0.$$

To establish (2), let $S_1 = \{j \in S : i \to j\}$ and $S_0 = S \setminus S_1 = \{j \in S : i \not\to j\}$, and consider $B = \bigwedge_{j\in S_1} A_j^{\mathrm{c}}$ and $C = \bigwedge_{j\in S_0} A_j^{\mathrm{c}}$. The left-hand side of (2) is

$$\mathbb{P}(A_i \mid B \wedge C) = \frac{\mathbb{P}(A_i \wedge B \wedge C)}{\mathbb{P}(B \wedge C)} = \frac{\mathbb{P}(A_i \wedge B \wedge C)}{\mathbb{P}(C)}\frac{\mathbb{P}(C)}{\mathbb{P}(B \wedge C)} = \frac{\mathbb{P}(A_i \wedge B \mid C)}{\mathbb{P}(B \mid C)}. \tag{3}$$

To upper bound the numerator, note that $\mathbb{P}(A_i \wedge B \mid C) \leqslant \mathbb{P}(A_i \mid C) = \mathbb{P}(A_i)$, since A_i is independent of the family of events $(A_j : j \in S_0)$. Hence, by the assumption of the theorem,

$$\mathbb{P}(A_i \wedge B \mid C) \leqslant \mathbb{P}(A_i) \leqslant x_i \prod_{j\,:\,i\to j}(1-x_j). \tag{4}$$

Now we want to lower bound the denominator in (3). Observe first that if $S_1 = \emptyset$ then $\mathbb{P}(B \mid C) = 1$ and (2) follows using (3) and (4); so we may assume $S_1 \neq \emptyset$. Write S_1 as $\{j_1, \ldots, j_a\}$. Then

$$\begin{aligned}
\mathbb{P}(B \mid C) &= \mathbb{P}(A_{j_1}^{\mathrm{c}} \wedge \cdots \wedge A_{j_a}^{\mathrm{c}} \mid C) \\
&= \prod_{t=1}^{a}\mathbb{P}(A_{j_t}^{\mathrm{c}} \mid C \wedge A_{j_1}^{\mathrm{c}} \wedge \cdots \wedge A_{j_{t-1}}^{\mathrm{c}}).
\end{aligned}$$

In each conditional probability in the product, we condition on the intersection of at most $r-1$ events, and j_t is not one of their indices, so the induction hypothesis (1) applies, and thus

$$\mathbb{P}(B \mid C) \geqslant \prod_{t=1}^{a}(1-x_{j_t}) = \prod_{j\in S_1}(1-x_j) \geqslant \prod_{j\,:\,i\to j}(1-x_j)$$

since $S_1 \subseteq \{j : i \to j\}$. Together with (3) and (4) this gives $\mathbb{P}(A_i \mid B \wedge C) \leqslant x_i$, which is exactly (2). This completes the proof by induction. $\square$

Dependency digraphs are slightly slippery. First recall that given the events $A_1, \ldots, A_n$, we *cannot* construct D simply by taking $i \to j$ if A_i and A_j are dependent. Considering three events such that each pair is independent but (A_1, A_2, A_3) is not, a legal dependency digraph must have at least one edge from vertex 1 (since A_1 *does* depend on (A_2, A_3)), and similarly from each other vertex.

The same example shows that (even minimal) dependency digraphs are not unique: in this case there are 8 minimal dependency digraphs.

There is an important special case where dependency digraphs are easy to construct; we state it as an easy lemma.

Lemma 3.2. *Suppose that $(X_\alpha)_{\alpha\in F}$ is a set of independent random variables, and that $A_1, \ldots, A_n$ are events where A_i is determined by $(X_\alpha : \alpha \in F_i)$ for some $F_i \subseteq F$. Then the (di)graph in which, for distinct i and j, $i \to j$ and $j \to i$ if and only if $F_i \cap F_j \neq \emptyset$ is a dependency digraph for $A_1, \ldots, A_n$.*

Proof. For each i, the events $(A_j : j \neq i, i \not\to j)$ are (jointly) determined by the variables $(X_\alpha : \alpha \in F \setminus F_i)$, and A_i is independent of this family of variables. $\square$

We now turn to a more user-friendly version of the local lemma. The *out-degree* of a vertex i in a digraph D is simply the number of j such that $i \to j$.

Theorem 3.3 (Local Lemma, Symmetric version). *Let $A_1, \ldots, A_n$ be events having a dependency digraph D with maximum out-degree at most d. If $\mathbb{P}(A_i) \leqslant p$ for all i and $ep(d+1) \leqslant 1$, then $\mathbb{P}(\bigwedge_i A_i^{\mathrm{c}}) > 0$.*

Proof. Set $x_i = 1/(d+1)$ for all i and apply Theorem 3.1. We have $|\{j : i \to j\}| \leqslant d$, and $(1+1/d)^d \leq e$, so

$$x_i \prod_{j\,:\,i\to j}(1-x_j) \geqslant \frac{1}{d+1}\left(\frac{d}{d+1}\right)^d \geqslant \frac{1}{e(d+1)} \geqslant p \geqslant \mathbb{P}(A_i),$$

and Theorem 3.1 applies. $\square$

Remark. Considering $d+1$ disjoint events each with probability $1/(d+1)$ shows that the constant (here e) must be > 1. In fact, the constant e is best possible for large d.

end L6 2014

Example (Hypergraph colouring)**.**

Theorem 3.4. *Let H be an r-uniform hypergraph in which each edge meets at most d other edges. If $d+1 \leqslant 2^{r-1}/e$ then H has a 2-colouring.*

Proof. Colour the vertices randomly in the usual way, each red/blue with probability $1/2$, independently of the others. Let A_i be the event that the ith edge e_i is monochromatic, so $\mathbb{P}(A_i) = 2^{1-r} = p$.

By Lemma 3.2 we may form a dependency digraph for the A_i by joining i to j (both ways) if e_i and e_j share one or more vertices. The maximum out-degree is at most d by assumption, and

$$ep(d+1) \leqslant e2^{1-r}(2^{r-1}/e) = 1.$$

Now Theorem 3.3 gives $\mathbb{P}(\wedge_i A_i^{\mathrm{c}}) > 0$, so there exists a good colouring. $\square$

Example (Ramsey numbers again)**.**

Theorem 3.5. *If $k \geqslant 3$ and $e2^{1-\binom{k}{2}}\binom{k}{2}\binom{n}{k-2} \leqslant 1$ then $R(k,k) > n$.*

Proof. Colour the edges of K_n as usual, each red/blue with probability $1/2$, independently of the others. For each $S \subseteq [n]$ with $|S| = k$ let A_S be the event that the complete graph on S is monochromatic, so $\mathbb{P}(A_S) = 2^{1-\binom{k}{2}}$.

For the dependency digraph, by Lemma 3.2 we may join S and T if they share an edge, i.e., if $|S \cap T| \geqslant 2$. The maximum degree d is

$$d = |\{T : |S \cap T| \geqslant 2\}| < \binom{k}{2}\binom{n}{k-2}.$$

By assumption $ep(d+1) \leqslant 1$, so Theorem 3.3 applies, giving the result. $\square$

Corollary 3.6. $R(k,k) \geqslant (1+o(1))\frac{k\sqrt{2}}{e}2^{k/2}$.

Proof. Straightforward(ish) calculation; you won't be asked to do it! $\square$

Note: this improves the bound from the first moment method by a factor of $\sqrt{2}$. This is not much, but this is the best lower bound known!

Example ($R(3,k)$)**.** In the previous example, the local lemma didn't make so much difference, because each event depended on very many others. If we consider off-diagonal Ramsey numbers the situation changes, but we can't use the symmetric form. The point here is to understand how to apply the lemma when we have 'two types' of events; the details of the calculation are not important.

Colour the edges of K_n red with probability p and blue with probability $1-p$, independently of each other, where $p = p(n) \to 0$.

For each $S \subseteq [n]$ with $|S| = 3$ let A_S be the event that S spans a red triangle, and for each $T \subseteq [n]$ with $|T| = k$ let B_T be the event that T spans a blue K_k. Note that

$$\mathbb{P}(A_S) = p^3 \qquad \text{and} \qquad \mathbb{P}(B_T) = (1-p)^{\binom{k}{2}}.$$

As usual, we can form the dependency digraph by joining two events if they involve one or more common edges. Each A event is joined to

- at most $3n$ other A events, and
- at most $\binom{n}{k} \leqslant n^k$ B events (as there are only $\binom{n}{k}$ B events in total).

Also, each B event is joined to

- at most $\binom{k}{2}n$ A events, and
- at most n^k B events.

Our aim is to apply Theorem 3.1 with $x_i = x$ for all A events and $x_i = y$ for all B events, to conclude that the probability that none of the A_S or B_T holds is positive, which gives $R(3,k) > n$. The conditions are satisfied provided we have

$$p^3 \leqslant x(1-x)^{3n}(1-y)^{n^k} \tag{5}$$

and

$$(1-p)^{\binom{k}{2}} \leqslant y(1-x)^{\binom{k}{2}n}(1-y)^{n^k}. \tag{6}$$

It turns out that

$$p = \frac{1}{6\sqrt{n}} \qquad x = \frac{1}{12n^{3/2}} \qquad k \sim 30\sqrt{n}\log n \qquad y = n^{-k}$$

satisfies (5) and (6) if n is large enough. This gives the following result.

Theorem 3.7. *There exists a constant $c > 0$ such that $R(3,k) \geqslant ck^2/(\log k)^2$ if k is large enough.*

Proof. The argument above shows that, for sufficiently large n, we have $R(3,k) > n$ if $k \sim 30\sqrt{n}\log n$, that is if $n \sim \frac{k^2}{(60\log k)^2}$. $\square$

end L7 2014

Remark. This bound is best possible apart from one factor of $\log k$. Removing this factor was not easy, and was a major achievement of J.H. Kim. We now (2013) know that

$$(\frac{1}{4} + o(1))\frac{k^2}{\log k} \leq R(3,k) \leq (1 + o(1))\frac{k^2}{\log k}.$$

4 Chernoff bounds

Often we are interested in whether a random graph $G(n,p)$ has some property almost always (with probability tending to one as $n \to \infty$), or almost never. For example, this is enough to allow us to show the existence of graphs with various combinations of properties, using the fact that if two or three properties individually hold almost always, then their intersection holds almost always. Sometimes, however, we need to consider a number k of properties (events) that tends to infinity as $n \to \infty$. This means that we would like tighter bounds on the probability that individual events fail to hold.

For example, let $G = G(n,p)$ and consider its maximum degree $\Delta(G)$. For any d we have $\mathbb{P}(\Delta(G) \geqslant d) \leqslant n\mathbb{P}(d_v \geqslant d)$, where d_v is the degree of a given vertex v. In turn this is at most $n\mathbb{P}(X \geqslant d)$ where $X \sim \mathrm{Bin}(n,p)$. To show that $\mathbb{P}(\Delta(G) \geqslant d) \to 0$ for some $d = d(n)$ we would need a bound of the form

$$\mathbb{P}(X \geqslant d) = o(1/n). \tag{7}$$

Recall that if $X \sim \mathrm{Bin}(n,p)$ then $\mu = \mathbb{E}[X] = np$ and $\sigma^2 = \mathrm{Var}[X] = np(1-p)$. For example, if $p = 1/2$ then $\mu = n/2$ and $\sigma = \sqrt{n}/2$. Chebyshev's inequality gives $\mathbb{P}(X \geqslant \mu + \lambda\sigma) \leqslant \lambda^{-2}$; to use this for (7) we need $\lambda \gg \sqrt{n}$ (that is, $\lambda/\sqrt{n} \to \infty$ as $n \to \infty$). If $p = 1/2$ this gives $\lambda\sigma \gg n$, which is useless.

On the other hand, the central limit theorem *suggests* that as $n \to \infty$

$$\mathbb{P}(X \geqslant \mu + \lambda\sigma) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \geqslant \lambda\right) \to \mathbb{P}(N(0,1) \geqslant \lambda) \approx e^{-\lambda^2/2}$$

where $N(0,1)$ is the standard normal distribution. But the $\to$ here is valid only for λ constant, so again it is no use for (7) (and the final $\approx$ should really be $\approx \lambda^{-1}e^{-\lambda^2/2}$, valid for large λ).

Our next aim is to prove a bound similar to the above, but valid no matter how λ depends on n.

Theorem 4.1. *Suppose that $n \geqslant 1$ and $p, x \in (0,1)$. Let $X \sim \mathrm{Bin}(n,p)$. Then*

$$\mathbb{P}(X \geqslant nx) \leqslant \left[\left(\frac{p}{x}\right)^x \left(\frac{1-p}{1-x}\right)^{1-x}\right]^n \qquad \text{if } x \geqslant p,$$

and

$$\mathbb{P}(X \leqslant nx) \leqslant \left[\left(\frac{p}{x}\right)^x \left(\frac{1-p}{1-x}\right)^{1-x}\right]^n \qquad \text{if } x \leqslant p.$$

Note that the exact expression is in some sense not so important; what matters is (a) the proof technique, and (b) that it is exponential in n if x and p are fixed.

Proof. The idea is simply to apply Markov's inequality to the random variable e^{tX} for some number t that we will choose so as to optimize the bound.

Consider X as a sum $X_1 + \ldots + X_n$ where the X_i are independent with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = 0) = 1 - p$. Then

$$\begin{aligned}\mathbb{E}[e^{tX}] &= \mathbb{E}[e^{tX_1} e^{tX_2} \cdots e^{tX_n}] \\ &= \mathbb{E}[e^{tX_1}] \cdots \mathbb{E}[e^{tX_n}] \\ &= (pe^t + (1-p)e^0)^n,\end{aligned}$$

where we used independence for the second equality.

For any $t > 0$, using the fact that $y \mapsto e^{ty}$ is increasing and Markov's inequality, we have

$$\begin{aligned}\mathbb{P}(X \geqslant nx) &= \mathbb{P}(e^{tX} \geqslant e^{tnx}) \\ &\leqslant \mathbb{E}[e^{tX}]/e^{tnx} \\ &= [(pe^t + 1 - p)e^{-tx}]^n.\end{aligned} \tag{8}$$

To get the best bound we minimize over t (by differentiating and equating to zero).

For $x > p$, the minimum occurs when

$$e^t = \frac{x}{p}\frac{1-p}{1-x} > 1,$$

so $t > 0$ and we can use this value: we obtain

$$\mathbb{P}(X \geqslant nx) \leqslant \left[\left(x\frac{1-p}{1-x} + 1 - p\right)\left(\frac{p}{x}\right)^x \left(\frac{1-x}{1-p}\right)^x\right]^n = \left[\left(\frac{p}{x}\right)^x \left(\frac{1-p}{1-x}\right)^{1-x}\right]^n,$$

proving the first part of the theorem. (The case $x = p$ is trivial since the bound is 1.)

For the second part, let $Y = n - X$, so $Y \sim \mathrm{Bin}(n, 1-p)$. Then $\mathbb{P}(X \leqslant nx) = \mathbb{P}(Y \geqslant n(1-x))$, and apply the first part. $\square$

Remark. Theorem 4.1 gives the best possible bound among bounds of the form $\mathbb{P}(X \geqslant nx) \leqslant g(x,p)^n$ where $g(x,p)$ is some function of x and p.

In the form above, the bound is a little hard to use. Here are some more practical forms.

Corollary 4.2. *Let $X \sim \mathrm{Bin}(n,p)$. Then for $h, t > 0$*

$$\mathbb{P}\big(X \geqslant np + nh\big) \leqslant e^{-2h^2 n}$$

and

$$\mathbb{P}\big(X \geqslant np + t\big) \leqslant e^{-2t^2/n}.$$

Also, for $0 \leqslant \varepsilon \leqslant 1$ we have

$$\mathbb{P}\big(X \geqslant (1+\varepsilon)np\big) \leqslant e^{-\varepsilon^2 np/4}$$

and

$$\mathbb{P}\big(X \leqslant (1-\varepsilon)np\big) \leqslant e^{-\varepsilon^2 np/2}.$$

Proof. Fix p with $0 < p < 1$. For $x > p$ or $x < p$ Theorem 4.1 gives $\mathbb{P}(X \geqslant nx) \leqslant e^{-f(x)n}$ or $\mathbb{P}(X \leqslant nx) \leqslant e^{-f(x)n}$, where

$$f(x) = x \log\left(\frac{x}{p}\right) + (1-x)\log\left(\frac{1-x}{1-p}\right).$$

We aim to bound $f(x)$ from below by some simpler function. Note that $f(p) = 0$. Also,

$$f'(x) = \log x - \log p - \log(1-x) + \log(1-p),$$

so $f'(p) = 0$ and

$$f''(x) = \frac{1}{x} + \frac{1}{1-x}.$$

If $f''(x) \geqslant a$ for all a between p and $p+h$ then (e.g., by Taylor's Theorem) we get $f(p+h) \geqslant ah^2/2$.

Now for any x we have $f''(x) \geqslant \inf_{x>0}\{1/x + 1/(1-x)\} = 4$, so $f(p+h) \geqslant 2h^2$, giving the first bound; the second is the same bound in different notation, setting $t = nh$.

For the third bound, if $p \leqslant x \leqslant p(1+\varepsilon) \leqslant 2p$ then $f''(x) \geqslant 1/x \geqslant 1/(2p)$, giving $f(p+\varepsilon p) \geqslant \frac{\varepsilon^2 p^2}{2}\frac{1}{2p}$, which gives the result.

For the final bound, if $0 < x \leqslant p$ then $f''(x) \geqslant 1/x \geqslant 1/p$, giving $f(p-\varepsilon p) \geqslant \frac{\varepsilon^2 p^2}{2}\frac{1}{p}$. □

end L8 2014

Remark. Recall that $\sigma = \sqrt{np(1-p)}$, so when p is small then $\varepsilon np \sim \varepsilon\sqrt{np}\sigma$. The central limit theorem *suggests* that the probability of a deviation this large should be around $e^{-\varepsilon^2 np/2}$ as in the final bound above. The third bound is weaker (and can be improved by replacing the 4 by a 3, but not by a 2).

In general, think of the bounds as of the form $e^{-c\lambda^2}$ for the probability of being λ standard deviations away from the mean. Alternatively, deviations on the scale of the mean are exponentially unlikely.

The Chernoff bounds apply more generally than just to binomial distributions; they apply to other sums of independent variables where each variable has bounded range.

Example (The maximum degree of $G(n,p)$)**.**

Theorem 4.3. *Let $p = p(n)$ satisfy $np \geqslant 10\log n$, and let Δ be the maximum degree of $G(n,p)$. Then*

$$\mathbb{P}\big(\Delta \geqslant np + 3\sqrt{np\log n}\big) \to 0$$

as $n \to \infty$.

Proof. Let $d = np + 3\sqrt{np\log n}$. As noted at the start of the section,

$$\mathbb{P}(\Delta \geqslant d) \leqslant n\mathbb{P}(d_v \geqslant d) \leqslant n\mathbb{P}(X \geqslant d)$$

where $d_v \sim \mathrm{Bin}(n-1,p)$ is the degree of a given vertex, and $X \sim \mathrm{Bin}(n,p)$. Applying the third bound in Corollary 4.2 with $\varepsilon = 3\sqrt{\log n/(np)} \leqslant 1$, we have

$$n\mathbb{P}(X \geqslant d) \leqslant ne^{-\varepsilon^2 np/4} = ne^{-9(\log n)/4} = nn^{-9/4} = n^{-5/4} \to 0,$$

giving the result. □

Note that for large n there will be some vertices with degrees any given number of standard deviations above the average. The result says however that all degrees will be at most $C\sqrt{\log n}$ standard deviations above. This is best possible, apart from the constant.

5 Phase Transition in $G(n, p)$

[Overview of results.]

5.1 Branching processes

Let $\mathbf{p} = (p_0, p_1, p_2, \ldots)$ be a probability distribution on the non-negative integers (that is, each $p_k \geq 0$ and $\sum_k p_k = 1$). Let Z be a random variable with this distribution. The *Galton–Watson branching process with offspring distribution* $\mathbf{p}$ is defined as follows:

- Generation 0 consists of a single individual.
- Generation $t+1$ consists of the children of individuals in generation t (all distinct)
- The number of children of each individual has distribution $\mathbf{p}$, and is independent of everything else, i.e., of the history so far, and of other individuals in the same generation.

We write X_t for the number of individuals in generation t, and $\mathbf{X} = (X_0, X_1, \ldots)$ for the random sequence of generation sizes. Note that $X_0 = 1$, and given the values of $X_0, \ldots, X_t$ with $X_t = k$, the conditional distribution of X_{t+1} is the sum of k independent copies of Z.

Let $\lambda = \mathbb{E}[Z]$. Suppose $\lambda < \infty$. Then $\mathbb{E}[X_{t+1} \mid X_t = k] = k\lambda$; and so

$$\begin{aligned} \mathbb{E}[X_{t+1}] &= \sum_k \mathbb{P}(X_t = k)\mathbb{E}[X_{t+1} \mid X_t = k] \\ &= \sum_k \mathbb{P}(X_t = k)k\lambda = \lambda\mathbb{E}[X_t]. \end{aligned}$$

Hence $\mathbb{E}[X_t] = \lambda^t$ for all t.

The branching process *survives* if $X_t > 0$ for all t, and *dies out* or *goes extinct* if $X_t = 0$ for some t.

If $\lambda = \mathbb{E}[Z] < 1$, then for any t we have

$$\mathbb{P}(\mathbf{X} \text{ survives}) \leqslant \mathbb{P}(X_t > 0) \leqslant \mathbb{E}[X_t] = \lambda^t.$$

Letting $t \to \infty$ shows that $\mathbb{P}(\mathbf{X} \text{ survives}) = 0$.

What if $\lambda > 1$? Note that any branching process with $\mathbb{P}(Z = 0) > 0$ *may* die out – the question is, can it survive?

We recall some basic properties of probability generating functions.

Definition. The *probability generating function* of the distribution $\mathbf{p}$, or of the random variable Z, is the function $f_{\mathbf{p}}$ (or f_Z) : $[0,1] \to \mathbb{R}$ defined by

$$f_{\mathbf{p}}(x) = \mathbb{E}[x^Z] = \sum_{k=0}^{\infty} p_k x^k.$$

The following facts are easy to check

- $f_{\mathbf{p}}(0) = p_0$ and $f_{\mathbf{p}}(1) = 1$.
- $f_{\mathbf{p}}$ is continuous on $[0,1]$.
- $f_{\mathbf{p}}$ is strictly increasing and $f'_{\mathbf{p}}$ is non-decreasing ($f_{\mathbf{p}}$ is convex) with $f'_{\mathbf{p}}(1) = \mathbb{E}[Z]$.
- If $p_0 + p_1 < 1$, or equivalently if $\mathbb{P}(Z \geqslant 2) > 0$, then $f'_{\mathbf{p}}$ is strictly increasing ($f_{\mathbf{p}}$ is strictly convex).

For the last two observations, note that for $0 < x < 1$ we have

$$f'_{\mathbf{p}}(x) = \sum_{k=1}^{\infty} k p_k x^{k-1} \geqslant 0;$$

and

$$f''_{\mathbf{p}}(x) = \sum_{k \geqslant 2} k(k-1) p_k x^{k-2} \geqslant 0$$

with strict inequality if $\mathbb{P}(Z \geqslant 2) > 0$. Further $f'_{\mathbf{p}}(x)$ increases to $\mathbb{E}[Z]$ as x increases to 1. It follows using the Mean Value Theorem that $f'_{\mathbf{p}}(1)$ exists ('from the left') and equals $\mathbb{E}[Z]$. For, given $0 < x < 1$, $\frac{1-f_{\mathbf{p}}(x)}{1-x} = f'_{\mathbf{p}}(y)$ for some $x < y < 1$, and $y \to 1$ as $x \to 1$.

Let $\eta_t = \mathbb{P}(X_t = 0)$. Then $\eta_0 = 0$ and

$$\eta_{t+1} = \sum_k \mathbb{P}(X_{t+1} = 0 \mid X_1 = k)\mathbb{P}(X_1 = k) = \sum_k \eta_t^k \mathbb{P}(Z = k) = f_{\mathbf{p}}(\eta_t),$$

since, given the number of individuals in the first generation, the descendants of each of them form an independent copy of the branching process.

Let $\mathbf{X}_{\mathbf{p}}$ denote the Galton–Watson branching process with offspring distribution $\mathbf{p}$. Let $\eta = \eta(\mathbf{p})$ denote the *extinction probability* of $\mathbf{X}_{\mathbf{p}}$, i.e., the probability that the process dies out.

Theorem 5.1. *For any probability distribution* $\mathbf{p}$ *on the non-negative integers,* $\eta(\mathbf{p})$ *is the smallest non-negative solution to* $f_{\mathbf{p}}(x) = x$*.*

Proof. As above, let $\eta_t = \mathbb{P}(X_t = 0)$, so $0 = \eta_0 \leqslant \eta_1 \leqslant \eta_2 \cdots$. Since the events $\{X_t = 0\}$ are nested and their union is the event that the process dies out, we have $\eta_t \to \eta$ as $t \to \infty$. end L9 2014

As shown above, $\eta_{t+1} = f_{\mathbf{p}}(\eta_t)$. Since $f_{\mathbf{p}}$ is continuous, taking the limit of both sides gives $\eta = f_{\mathbf{p}}(\eta)$, so η is *a* non-negative solution to $f_{\mathbf{p}}(x) = x$.

Let x_0 be any non-negative solution to $f_{\mathbf{p}}(x) = x$. Then $\eta_0 = 0 \leqslant x_0$. Since $f_{\mathbf{p}}$ is increasing, this gives

$$\eta_1 = f_{\mathbf{p}}(\eta_0) \leqslant f_{\mathbf{p}}(x_0) = x_0.$$

Similarly, by induction we obtain $\eta_t \leqslant x_0$ for all t, so taking the limit, $\eta \leqslant x_0$. □

Corollary 5.2. *If* $\mathbb{E}[Z] > 1$ *then* $\eta = \eta(\mathbf{p}) < 1$, *i.e., the probability that* $\mathbf{X}_{\mathbf{p}}$ *survives is positive. If* $\mathbb{E}[Z] < 1$ *then* $\eta = 1$. *If* $\mathbb{E}[Z] = 1$ *and* $p_1 \neq 1$ *then* $\eta = 1$.

Proof. The question is simply whether the curves $f_{\mathbf{p}}(x)$ and x meet anywhere in $[0,1]$ other than at $x = 1$; sketch the graphs!

For the first statement, suppose that $\mathbb{E}[Z] > 1$. Then $f'_{\mathbf{p}}(1) > 1$, so there exists $\varepsilon > 0$ such that $f_{\mathbf{p}}(1-\varepsilon) < 1-\varepsilon$. Since $f_{\mathbf{p}}(0) \geqslant 0$, applying the Intermediate Value Theorem to $f_{\mathbf{p}}(x) - x$, there must be some $x \in [0, 1-\varepsilon]$ for which $f_{\mathbf{p}}(x) = x$. But then $\eta \leqslant x \leqslant 1 - \varepsilon < 1$.

We have already proved the second statement; so let us focus on the third, with $\mathbb{E}[Z] = 1$ and $p_1 \neq 1$. Now $f_{\mathbf{p}}(x)$ has strictly increasing derivative, and so $g(x) = f_{\mathbf{p}}(x) - x$ has strictly increasing derivative. But $g'(1) = 0$ so $g'(y) < 0$ for all $y \in (0,1)$, and thus $g(x) > 0$ for all $x \in [0,1)$, by the Mean Value Theorem. □

Note that when $\mathbb{E}[Z] > 1$, there is a *unique* solution to $f_{\mathbf{p}}(x) = x$ in $[0,1)$; this follows from the strict convexity of $f_{\mathbf{p}}$.

Definition. For $c > 0$, a random variable Z has the *Poisson distribution with mean c*, written $Z \sim \mathrm{Po}(c)$, if

$$\mathbb{P}(Z = k) = \frac{c^k}{k!} e^{-c}$$

for $k = 0, 1, 2, \ldots$.

Lemma 5.3. *Suppose* $n \to \infty$ *and* $p \to 0$ *with* $np \to c$ *where* $c > 0$ *is constant. Let* Z_n *have the binomial distribution* $\mathrm{Bin}(n,p)$, *and let* $Z \sim \mathrm{Po}(c)$. *Then* Z_n converges in distribution *to* Z*; that is, for each fixed* k, $\mathbb{P}(Z_n = k) \to \mathbb{P}(Z = k)$.

Proof. For k fixed,

$$\mathbb{P}(Z_n = k) = \binom{n}{k} p^k (1-p)^{n-k} \sim \frac{n^k}{k!} p^k (1-p)^n = \frac{(np)^k}{k!} e^{-np+O(np^2)} \to \frac{c^k}{k!} e^{-c},$$

since $np \to c$ and $np^2 \to 0$. □

As we shall see shortly, there is a very close connection between components in $G(n, c/n)$ and the Galton–Watson branching process $\mathbf{X}_{\mathrm{Po}(c)}$ where the offspring distribution $\mathbf{p}$ is Poisson with mean c. The extinction probability of this process will be especially important.

Theorem 5.4. *Let $c > 0$. Then the extinction probability $\eta = \eta(c)$ of the branching process $\mathbf{X}_{\mathrm{Po}(c)}$ satisfies the equation*

$$\eta = e^{-c(1-\eta)}.$$

Furthermore, $\eta < 1$ if and only if $c > 1$.

Proof. The probability generating function of the Poisson distribution with mean c is given by

$$f(x) = \sum_{k=0}^{\infty} \frac{c^k}{k!} e^{-c} x^k = e^{cx} e^{-c} = e^{c(x-1)} = e^{-c(1-x)}.$$

The result now follows from Theorem 5.1 and Corollary 5.2. □

Lemma 5.5. *Let $c > 0$ be fixed. Let $m = m(n) \to \infty$ and $p = p(n)$ with $0 \le p \le 1$ satisfy $mp \to c$ as $n \to \infty$. Let $\mathbf{p}_n \sim \mathrm{Bin}(m, p)$. Then*

$$\eta(\mathbf{p}_n) \to \eta(c) \quad \text{as } n \to \infty.$$

Proof. Fix $t \in (0, 1)$. Recall that $1 - x = e^{-x+O(x^2)}$ as $x \to 0$. Thus as $n \to \infty$

$$\begin{aligned} f_{\mathbf{p}_n}(t) &= (1 - p + pt)^m \ = \ (1 - p(1-t))^m \\ &= \exp(-mp(1-t) + O(mp^2)) \\ &\to e^{-c(1-t)} \ = \ f_{\mathrm{Po}(c)}(t). \end{aligned}$$

From our discussion of branching processes, if $t < \eta(c)$ then $f_{\mathrm{Po}(c)}(t) > t$ and if $t > \eta(c)$ then $f_{\mathrm{Po}(c)}(t) < t$. Hence, if $t < \eta(c)$ then $f_{\mathbf{p}_n}(t) > t$ for n sufficiently large, and so $t < \eta(\mathbf{p}_n)$; and similarly if $t > \eta(c)$ then $f_{\mathbf{p}_n}(t) < t$ for n sufficiently large, and so $t > \eta(\mathbf{p}_n)$. □

end L10 2014

5.2 Component exploration

In the light of Lemma 5.3, we may hope that the Poisson branching process gives a good 'local' approximation to the neighbourhood of a vertex of $G(n, c/n)$. To make this precise, we shall 'explore' the component of a vertex in a certain way. First we describe the (simpler) exploration for the branching process.

Exploration process for a branching process.

Start with $Y_0^{bp} = 1$, meaning one live individual (the root). At each step t, select a live individual; it has Z_t children and then dies. Let Y_t^{bp} be the number of individuals alive after t steps. Then

$$Y_t^{bp} = \begin{cases} Y_{t-1}^{bp} + Z_t - 1 & \text{if } Y_{t-1}^{bp} > 0 \\ 0 & \text{if } Y_{t-1}^{bp} = 0. \end{cases}$$

The process dies out if $Y_m^{bp} = 0$ for some m; in this case the total number T^{bp} of individuals is $\min\{m : Y_m^{bp} = 0\}$. (We set $T^{bp} = \infty$ if there is no such m.)

Until it hits zero, the sequence (Y_t^{bp}) is a random walk with iid steps $Z_1 - 1, Z_2 - 1, \ldots$ taking values in $\{-1, 0, 1, 2, \ldots\}$. Further, each step has expectation $\mathbb{E}[Z - 1] = \lambda - 1$. The idea is that $\lambda < 1$ implies negative drift and the walk will hit 0, i.e. the process will die, with probability 1. If $\lambda > 1$ then the drift is positive, so the probability of never hitting 0 is positive, i.e. the probability that the process survives is positive.

It will be convenient to consider the infinite random walk $(\tilde{Y}_t^{bp})$, which as above starts at 1 and has iid increments distributed like $Z - 1$. We may use exactly the same increments as for (Y_t^{bp}) as long as $t < T^{bp}$. Let $\tilde{T}^{bp}$ be the least m such that $\tilde{Y}_m^{bp} = 0$ (and $= \infty$ if there is no such m). Observe that $T^{bp} = \tilde{T}^{bp}$, and indeed $Y_t^{bp} = \tilde{Y}_t^{bp}$ for each $t \leq T^{bp} (= \tilde{T}^{bp})$.

Component exploration in $G(n, p)$.

Let v be a fixed vertex of a graph G. At each stage, each vertex u of G will be 'live', 'unreached', or 'processed'. Y_t^{gr} will be the number of live vertices after t steps, there will be exactly t processed vertices, and $U_t = n - t - Y_t$ unreached vertices.

At $t = 0$, mark v as live and all other vertices as unreached, so $Y_0^{gr} = 1$ and $U_0 = n - 1$.

At each step t, pick a live vertex w, if there is one. For each unreached w', check whether $ww' \in E(G)$; if so, make w' live. After completing these checks, set w to be processed.

Let R_t be the number of w' which become live during step t. (Think of this as the number of vertices Reached in step t.) Then

$$Y_t^{gr} = \begin{cases} Y_{t-1}^{gr} + R_t - 1 & \text{if } Y_{t-1}^{gr} > 0 \\ 0 & \text{if } Y_{t-1}^{gr} = 0. \end{cases}$$

Let $T^{gr} = \min\{m : Y^{gr} = 0\}$. The process stops at time T^{gr}, when we have reached all vertices in the component C_v of G containing v (since each vertex of C_v must become live at some step, and be processed). In particular, $|C_v| = T^{gr}$.

So far, G could be any graph. Now suppose that $G = G(n,p)$. Then each edge is present with probability p independently of the others. No edge is tested twice (we only check edges from live to unreached vertices, and then one end becomes processed). It follows that given $Y_0^{gr} = y_0, \ldots, Y_{t-1}^{gr} = y_{t-1}$, the number R_t of vertices reached in step t has the distribution

$$R_t \sim \mathrm{Bin}(u_{t-1}, p) \quad \text{where } u_{t-1} = n - (t-1) - y_{t-1}. \tag{9}$$

Extended process As with branching process exploration, it is convenient to introduce a simpler extended version of the exploration process. The extended process $(\tilde{Y}_t^{gr})$ always runs for $t = 0, 1, \ldots, n$. Let $\tilde{Y}_0^{gr} = 1$. For each $t = 1, \ldots, n$, conditional on $\tilde{Y}_1^{gr} = y_1, \ldots, \tilde{Y}_{t-1}^{gr} = y_{t-1}$ where $\tilde{u}_{t-1} := n - (t-1) - y_{t-1} \geq 0$, we generate $\tilde{R}_t \sim \mathrm{Bin}(\tilde{u}_{t-1}, p)$; and set $\tilde{Y}_t^{gr} = y_{t-1} - 1 + \tilde{R}_t$.

We use exactly the same increments as for (Y_t^{gr}) as long as $t < T^{gr}$. Let $\tilde{T}^{gr}$ be the least m such that $\tilde{Y}_m^{gr} = 0$ (and $= \infty$ if there is no such m). Then $T^{gr} = \tilde{T}^{gr}$, and indeed $Y_t^{gr} = \tilde{Y}_t^{gr}$ for each t up to this time.

5.3 Probability bounds for component size

Recall that C_v is the component containing v in $G(n,p)$. We shall establish the probability bounds (10), (11) and (12) for $|C_v|$.

Let k be an integer with $0 \leq k < n$. Let $(\mathbf{X}_n^1)$ be the branching process with offspring distribution $\mathrm{Bin}(n,p)$. Then

$$\mathbb{P}(|C_v| > k) \ \leq \ \mathbb{P}(|\mathbf{X}_n^1| > k) \ \leq \ \mathbb{P}(\mathrm{Bin}(kn, p) \geq k). \tag{10}$$

Now let $(\mathbf{X}_n^2)$ be the branching process with offspring distribution $\mathrm{Bin}(n-k, p)$. end L11 2014

Then

$$\mathbb{P}(|C_v| > k) \geq \mathbb{P}(|\mathbf{X}_n^2| > k), \tag{11}$$

and

$$\mathbb{P}(|C_v| = k) \leq \mathbb{P}(\mathrm{Bin}(k(n-k), p) \leq k-1). \tag{12}$$

Let us first prove (10). The distribution of R_t is dominated by a $\mathrm{Bin}(n, c/n)$ distribution. More precisely, we can define iid rvs $Z_t \sim \mathrm{Bin}(n, c/n)$ for $t = 1, 2, \ldots, n$ so that in the extended process always $\tilde{R}_t \leqslant Z_t$. [To be totally explicit, construct the random variables step by step. At step t, $\tilde{R}_t \sim \mathrm{Bin}(x, c/n)$ for some $0 \leq x \leqslant n$ that depends on what has happened so far. Toss x biased coins to determine $\tilde{R}_t$, and then $n - x$ further coins, taking the total number of heads to be Z_t; each coin has probability p of landing heads.]

Corresponding to the branching process $\mathbf{X}_n^1$, let $(\tilde{Y}_t^{bp})$ be the extended random walk with $\tilde{Y}_0^{bp} = 1$ and increments $Z_t - 1$, so $\tilde{Y}_t \leqslant \tilde{Y}_t^{bp}$ for all $t = 0, 1, \ldots, n$. Then for any k we have

$$\begin{aligned}
\mathbb{P}(|C_v| > k) &= \mathbb{P}\big(\tilde{Y}_1^{gr}, \ldots, \tilde{Y}_k^{gr} > 0\big) \\
&\leqslant \mathbb{P}\big(\tilde{Y}_1^{bp}, \ldots, \tilde{Y}_k^{bp} > 0\big) = \mathbb{P}(|\mathbf{X}_n^1| > k) \\
&\leqslant \mathbb{P}(\tilde{Y}_k^{bp} > 0).
\end{aligned}$$

Also, since $\tilde{Y}_k^{bp}$ has the simple distribution

$$\tilde{Y}_k^{bp} + k - 1 = \sum_{t=1}^{k} Z_t \sim \mathrm{Bin}(kn, p),$$

we have $\mathbb{P}(\tilde{Y}_k^{bp} > 0) = \mathbb{P}\big(\mathrm{Bin}(nk, p) \geqslant k\big)$; and we have completed the proof of (10).

Now let us prove (11) and (12). Corresponding to the branching process $(\mathbf{X}_n^2)$, for $t = 1, 2, \ldots$ let Z_t be iid random variables with distribution $\mathrm{Bin}(n - k, p)$, and let $(\tilde{Y}_t^{bp})$ be the random walk starting with $\tilde{Y}_0^{bp} = 1$ and with increments $Z_t - 1$. As long as we have used up at most k vertices, the increment $\tilde{R}_t$ dominates Z_t. It follows that

$$\begin{aligned}
\mathbb{P}(|C_v| > k) &= \mathbb{P}(\tilde{Y}_1^{gr}, \ldots, \tilde{Y}_k^{gr} > 0) \\
&\geq \mathbb{P}(\tilde{Y}_1^{bp}, \ldots, \tilde{Y}_k^{bp} > 0) = \mathbb{P}(|\mathbf{X}_n^2| > k).
\end{aligned}$$

Similarly

$$\mathbb{P}(|C_v| = k) = \mathbb{P}((\tilde{Y}_1^{gr}, \ldots, \tilde{Y}_{k-1}^{gr} > 0) \wedge \tilde{Y}_k^{gr} = 0) \leq \mathbb{P}(\tilde{Y}_k^{bp} \leq 0).$$

As before, $\tilde{Y}_k^{bp}$ has a simple distribution, namely $\mathrm{Bin}(k(n - k), p) - k + 1$; and (12) follows. amplify?

5.4 Vertices in small components

Let $\rho_k(c)$ denote the probability that $|\mathbf{X}_{\mathrm{Po}(c)}| = k$, where $|\mathbf{X}| = \sum_{t\geqslant 0} X_t$ denotes the total number of individuals in all generations of the branching process $\mathbf{X}$.

Lemma 5.6. *Suppose that $p = p(n)$ satisfies $np \to c$ where $c > 0$ is constant. Let v be a given vertex of $G(n,p)$. For each constant k we have*

$$\mathbb{P}(|C_v| = k) \to \rho_k(c) \quad \textit{as } n \to \infty.$$

Proof. The idea is simply to show that the random walks (Y_t^{gr}) and (Y_t^{bp}) have almost the same probability of first hitting zero at $t = k$. We do this by comparing the probabilities of individual trajectories.

Define (Y_t^{gr}) and (R_t) as in the graph exploration above. Then $|C_v| = k$ if and only if $Y_k^{gr} = 0$ and $Y_t^{gr} > 0$ for all $t < k$. Let $\mathcal{S}_k$ be the set of all possible corresponding sequences $\mathbf{y} = (y_0, \ldots, y_k)$ of values for $\mathbf{Y}^{gr} = (Y_0^{gr}, \ldots, Y_k^{gr})$: that is, $y_0 = 1$, $y_k = 0$, $y_t > 0$ for $t < k$, and each y_t is an integer with $y_t \geqslant y_{t-1} - 1$. Then

$$\mathbb{P}(|C_v| = k) = \sum_{\mathbf{y}\in\mathcal{S}_k} \mathbb{P}(\mathbf{Y}^{gr} = \mathbf{y}).$$

Similarly,

$$\rho_k(c) = \mathbb{P}\big(|\mathbf{X}_{\mathrm{Po}(c)}| = k\big) = \sum_{\mathbf{y}\in\mathcal{S}_k} \mathbb{P}(\mathbf{Y}^{bp} = \mathbf{y}).$$

Fix any sequence $\mathbf{y} \in \mathcal{S}_k$. For each t let $r_t = y_t - y_{t-1} + 1$, so (r_t) is the sequence of R_t values corresponding to $\mathbf{Y}^{gr} = \mathbf{y}$. From (9) we have

$$\mathbb{P}(\mathbf{Y}^{gr} = \mathbf{y}) = \prod_{t=1}^{k} \mathbb{P}\big(\mathrm{Bin}(n - (t-1) - y_{t-1}, p) = r_t\big).$$

In each factor, $t-1$, y_{t-1} and r_t are constants. As $n \to \infty$ we have $n - (t-1) - y_{t-1} \sim n$, so $(n - (t-1) - y_{t-1})p \sim c$. Applying Lemma 5.3 to each factor in the product, it follows that

$$\mathbb{P}(\mathbf{Y}^{gr} = \mathbf{y}) \to \prod_{t=1}^{k} \mathbb{P}\big(\mathrm{Po}(c) = r_t\big).$$

But this is just $\mathbb{P}(\mathbf{Y}^{bp} = \mathbf{y})$, from the exploration for the branching process. Summing over the finite number of possible sequences $\mathbf{y}$ gives the result. $\square$

We use this last result, together with (11), to prove:

Lemma 5.7. *Let $k = k(n) \to \infty$ with $k = o(n)$. Then*

$$\mathbb{P}(|C_v| \le k) \to \eta(c) \quad \textit{as } n \to \infty.$$

Proof. Let $\varepsilon > 0$. Since $\mathbb{P}(|\mathbf{X}_{\mathrm{Po}(c)}| \le a) \to \eta(c)$ as $a \to \infty$, we may choose a constant a sufficiently large that $\mathbb{P}(|\mathbf{X}_{\mathrm{Po}(c)}| \le a) \ge \eta(c) - \varepsilon/2$. Then by Lemma 5.6, $\mathbb{P}(|C_v| \le a) \ge \eta(c) - \varepsilon$ for n sufficiently large. So if $b = b(n) \to \infty$ as $n \to \infty$ then $\liminf_n \mathbb{P}(|C_v| \le b) \ge \eta(c)$.

Now let $b = b(n) = o(n)$. By (11) with $k = b$, $\mathbb{P}(|C_v| > b) \ge \mathbb{P}(|\mathbf{X}_n^2| = \infty)$. But $(n-b)p \to c$ as $n \to \infty$, so by Lemma 5.5 $\mathbb{P}(|\mathbf{X}_n^2| = \infty) \to 1 - \eta(c)$. Thus $\limsup_n \mathbb{P}(|C_v| \le b) \le \eta(c)$. □

Write $N_{\le k}(G)$ for the number of vertices in components with at most k vertices.

Corollary 5.8. *Suppose that $np \to c$ where $c > 0$ is constant. Let $k = k(n) \to \infty$ with $k = o(n)$. Then $\mathbb{E}[N_{\le k}(G(n,p))] \sim \eta(c)\, n$ as $n \to \infty$.*

Proof. The expectation is $\sum_v \mathbb{P}(|C_v| \le k) = n\, \mathbb{P}(|C_v| \le k) \sim \eta(c)\, n$. □

The last result shows that the branching process 'predicts' the expected number of vertices in 'small' components. We will use the second moment method to show that in fact this number is concentrated around its mean.

Definition. Let (X_n) be a sequence of real-valued random variables and a a (constant) real number. Then X_n *converges to a in probability*, written $X_n \xrightarrow{\mathrm{p}} a$, if for all (fixed) $\varepsilon > 0$ we have $\mathbb{P}(|X_n - a| > \varepsilon) \to 0$ as $n \to \infty$.

Lemma 5.9. *Suppose that $\mathbb{E}[X_n] \to a$ and $\mathbb{E}[X_n^2] \to a^2$. Then $X_n \xrightarrow{\mathrm{p}} a$.*

Proof. Note that $\mathrm{Var}[X_n] = \mathbb{E}[X_n^2] - (\mathbb{E}X_n)^2 \to a^2 - a^2 = 0$, and apply Chebyshev's inequality. □

In fact, whenever we showed that some quantity X_n was almost always positive by using the second moment method, we really showed more, that $X_n/\mathbb{E}[X_n] \xrightarrow{\mathrm{p}} 1$, i.e., that X_n is 'concentrated around its mean'.

We know from Corollary 5.8 that $\mathbb{E}[N_{\le k}(G(n,p))/n] \to \eta(c)$ (for suitable k): now we show more.

Lemma 5.10. *Suppose that $np \to c$ where $c > 0$ is constant. Let $k = k(n) \to \infty$ with $k = o(n)$, and let $N_{\le k} = N_{\le k}(G(n,p))$. Then $N_{\le k}/n \xrightarrow{\mathrm{p}} \eta(c)$ as $n \to \infty$.*

end L12 2014

details not examinable

Proof. For each vertex v and positive integer $i \leq k$ let I_{vi} be the indicator function of the event that $|C_v| = i$, so $N_{\leq k} = \sum_v \sum_i I_{vi}$ and

$$N_k^2 = \sum_v \sum_w \sum_i \sum_j I_{vi} I_{wj} = A + B,$$

where

$$A = \sum_v \sum_w \sum_i \sum_j I_{vi} I_{wj} I_{\{C_v = C_w\}} = \sum_v \sum_w \sum_i I_{vi} I_{\{C_v = C_w\}}$$

is the part of the sum from vertices in the same component, and

$$B = \sum_v \sum_w \sum_i \sum_j I_{vi} I_{wj} I_{\{C_v \neq C_w\}}$$

is the part from vertices in different components.

Consider A. If $I_{vi} = 1$ then $|C_v| = i$ and so $\sum_w I_{\{C_v = C_w\}} = i$. Thus, since $i \leq k$,

$$A = \sum_v \sum_i I_{vi} \sum_w I_{\{C_v = C_w\}} = \sum_v \sum_i I_{vi}\, i \leq k \sum_v \sum_i I_{vi} = k N_{\leq k}.$$

Hence $A \leq kn$, and $\mathbb{E}[A] = o(n^2)$.

Consider B. Since all vertices are equivalent, let us fix a vertex v. We may write I_{vi} as $\sum_{C \ni v, |C| = i} I_{\{C_v = C\}}$, where we are summing over all i-sets $C \subseteq V = V(G)$ with $v \in C$. Thus

$$\begin{aligned} \mathbb{E}[B] &= n\, \mathbb{E}\left[\sum_i \sum_{C \ni v, |C| = i} I_{\{C_v = C\}} \sum_w \sum_j I_{wj} I_{\{C_v \neq C_w\}} \right] \\ &= n \sum_i \sum_{C \ni v, |C| = i} \mathbb{E}\left[I_{\{C_v = C\}} \sum_w \sum_j I_{wj} I_{\{C_w \subseteq V \setminus C\}} \right]. \end{aligned}$$

But

$$I_{\{C_v = C\}} \sum_w \sum_j I_{wj} I_{\{C_w \subseteq V \setminus C\}} = I_{\{C_v = C\}} N_{\leq k}(G[V \setminus C])$$

since when $C_v = C$ there are no edges between C and $V \setminus C$. But $I_{\{C_v = C\}}$ and $N_{\leq k}(G[V \setminus C])$ are independent, since they involve disjoint sets of edges.

Also $N_{\leq k}(G[V \setminus C]) \sim N_{\leq k}(G(n-i, p(n))$ when $|C| = i$. Hence

$$\begin{aligned}
\mathbb{E}[B] &= n \sum_i \sum_{C \ni v, |C|=i} \mathbb{E}[I_{\{C_v = C\}}]\, \mathbb{E}[N_{\leq k}(G(n-i, p(n))] \\
&= n \sum_i \mathbb{P}(|C_v| = i)\, \mathbb{E}[N_{\leq k}(G(n-i, p(n))].
\end{aligned}$$

But by Corollary 5.8, $\mathbb{E}[N_{\leq k}(G(n-i, p(n))] \sim n\,\eta(c)$ for each $i \leq k$, and $n \sum_i \mathbb{P}(|C_v| = i) = \mathbb{E}[N_{\leq k}] \sim n\,\eta(c)$; and it follows that $\mathbb{E}[B] \sim (n\,\eta(c))^2$.

Hence, $\mathbb{E}[N_k^2] = \mathbb{E}[A] + \mathbb{E}[B] \sim (n\,\eta(c))^2$, and so $\mathbb{E}[(N_{\leq k}/n)^2] \to \eta(c)^2$. Lemma 5.9 now gives the result. □

For each fixed k, we know quite precisely how many vertices are in components of size k; and if $k = k(n) \to \infty$ with $k = o(n)$ (say $k = n/\log n$) then similarly we know how many vertices are in components of size at most k. Does this mean that we know the whole component structure? Not quite: if $c > 1$, so $\eta < 1$, then we know that there are around $(1-\eta)n$ vertices in components of size $> k$. But are these components really of around that size, or much larger?

5.5 The phase transition

We say that an event (formally a sequence of events) holds *with high probability* or *whp* if its probability tends to 1 as $n \to \infty$. Consider first the subcritical case $c < 1$.

Lemma 5.11. *Let $0 < c < 1$ be constant. There is a constant $A > 0$ (which depends on c) such that whp every component of $G(n, c/n)$ has size at most $A \log n$.*

Proof. By (10)

$$\mathbb{P}(|C_v| > k) \leq \mathbb{P}\big(\mathrm{Bin}(nk, c/n) \geqslant k\big) = \mathbb{P}\big(\mathrm{Bin}(nk, c/n) \geqslant ck + (1-c)k\big).$$

Since the mean of the binomial is ck, setting $\varepsilon = \min\{(1-c)/c, 1\}$, the Chernoff bound shows that the last probability is at most $e^{-\varepsilon^2 ck/4}$. If we set $k = \lceil A \log n \rceil$ with $A \geq 8/(\varepsilon^2 c)$, then we have $\mathbb{P}(|C_v| > k) \leqslant e^{-2\log n} = 1/n^2$. Then the expected number of vertices in components larger than $A \log n$ is at most $1/n = o(1)$. □

Now suppose that $c > 1$. Fix any $\delta > 0$, and let $k_0 = \lfloor (1 - c^{-1} - \delta)n \rfloor$. Then by (12), for each $1 \leq k \leq k_0$,

$$\mathbb{P}(|C_v| = k) \leq \mathbb{P}\big(\mathrm{Bin}(k(n-k), c/n) \leq k-1\big);$$

and

$$k(n-k) \geq k(n-k_0) \geq kn(c^{-1}+\delta)$$

so the binomial has mean $\mu \geq k + \delta ck$. Thus $k-1 \leq k \leq \mu(1-\varepsilon)$ for $\varepsilon = \delta c/(1+\delta c)$, where $0 < \varepsilon < 1$ is a constant. By the Chernoff bound, the probability above is thus at most $e^{-\varepsilon^2\mu/2} \leqslant e^{-\varepsilon^2 k/2}$.

Choose a constant $A \geq 6/\varepsilon^2$ and let $k^- = A\log n$. Then for $k^- \leqslant k \leqslant k_0$ we have

$$\mathbb{P}(|C_v| = k) \leqslant e^{-\varepsilon^2 k/2} \leqslant e^{-\varepsilon^2 k^-/2} \leqslant e^{-3\log n} = 1/n^3.$$

Applying the union bound over $k^- \leqslant k \leqslant k_0$ and over all n vertices v, it follows that whp there are *no vertices at all* in components of size between k^- and k_0. In other words, whp *all* components are either *small*, i.e. of size at most $k^- = O(\log n)$, or *large*, i.e. of size at least k_0.

Given a graph G, let $L_i(G)$ denote the number of vertices in the ith largest component. Note that which component is the ith largest may be ambiguous, if there are ties, but the value of $L_i(G)$ is unambiguous.

We know that there are close to ηn vertices in small components; hence there are close to $(1-\eta)n$ vertices in large components. We close this section by showing that almost all vertices not in 'small' components are in a *unique* 'giant' component.

Theorem 5.12. *Let $c > 0$ be constant, and let $G = G(n, c/n)$.*

If $c < 1$ there is a constant A such that $L_1(G) \leqslant A\log n$ holds whp.

If $c > 1$ then $L_1(G)/n \xrightarrow{\mathrm{p}} 1-\eta(c)$, and there is a constant A such that $L_2(G) \leqslant A\log n$ holds whp.

Proof. We have almost completed the proof. The only remaining ingredient is to show in the $c > 1$ case that there cannot be two large components.

The simplest way to show this is just to choose $\delta > 0$ so that $(1-1/c-\delta) > (1-\eta)/2$. Then we don't have enough vertices in large components to have two or more large components. But is this possible? Such a δ exists if and only if $(1-1/c) > (1-\eta)/2$, i.e., if and only if $\eta > 2/c - 1$.

Recall that $\eta = \eta(c)$ is the smallest solution to $\eta = e^{-c(1-\eta)}$. Furthermore (drawing the graphs), for $x < \eta$ we have $x < e^{-c(1-x)}$ and for $\eta < x < 1$ we have $x > e^{-c(1-x)}$. So what we have to show is that $x = 2/c - 1$ falls into the first case, i.e., that $2/c - 1 < e^{-c(1-(2/c-1))} = e^{2-2c}$.

Multiplying by c, let $f(c) = ce^{2-2c} + c - 2$, so our task is to show $f(c) > 0$ for $c > 1$. This is easy by calculus: we have $f(1) = 0$, $f'(1) = 0$ and $f''(c) > 0$ for $c > 1$. (In fact $f''(c) = 4(c-1)e^{2-2c}$.) $\square$

6 Correlation and concentration

6.1 Harris's Lemma

In this section we turn to the following simple question and its generalizations. Does conditioning on $G = G(n,p)$ containing a triangle make G more or less likely to be connected? Note that if we condition on a fixed set E of edges being present, then this is the same as simply adding E to $G(n,p)$, which does increase the chance of connectedness. But conditioning on *at least one* triangle being present is not so simple.

Let X be any non-empty finite set, the *ground set*. For $0 \leqslant p \leqslant 1$ let X_p be a random subset of X obtained by selecting each element independently with probability p. A *property of subsets of* X is just some collection $\mathcal{A} \subseteq \mathcal{P}(X)$ of subsets of X. For example, the property 'contains element 1 or element 3' may be identified with the set $\mathcal{A}$ of all subsets A of X with $1 \in A$ or $3 \in A$.

We write $\mathbb{P}_p^X(\mathcal{A})$ for

$$\mathbb{P}(X_p \in \mathcal{A}) = \sum_{A \in \mathcal{A}} p^{|A|}(1-p)^{|X|-|A|}.$$

Most of the time, we omit X from the notation, writing $\mathbb{P}_p(\mathcal{A})$ for $\mathbb{P}_p^X(\mathcal{A})$.

We say that $\mathcal{A} \subseteq \mathcal{P}(X)$ is an *up-set*, or *increasing property*, if $A \in \mathcal{A}$ and $A \subseteq B \subseteq X$ implies $B \in \mathcal{A}$. Similarly, $\mathcal{A}$ is a *down-set* or *decreasing property* if $A \in \mathcal{A}$ and $B \subseteq A$ implies $B \in \mathcal{A}$. Note that $\mathcal{A}$ is an up-set if and only if $\mathcal{A}^{\mathrm{c}} = \mathcal{P}(X) \setminus \mathcal{A}$ is a down-set.

To illustrate the definitions, consider the (for us) most common special case. Here X consists of all $\binom{n}{2}$ edges of K_n, and X_p is then simply the edge-set of $G(n,p)$. Then a property of subsets of X is just a set of graphs on $[n]$, e.g., all connected graphs on $[n]$. A property is increasing if it is preserved by adding edges, and decreasing if it is preserved by deleting edges.

Lemma 6.1 (Harris's Lemma)**.** *If $\mathcal{A}$, $\mathcal{B} \subseteq \mathcal{P}(X)$ are up-sets and $0 \leqslant p \leqslant 1$ then*

$$\mathbb{P}_p(\mathcal{A} \cap \mathcal{B}) \geqslant \mathbb{P}_p(\mathcal{A})\mathbb{P}_p(\mathcal{B}). \tag{13}$$

In other words, $\mathbb{P}\big(X_p \in \mathcal{A} \text{ and } X_p \in \mathcal{B}\big) \geqslant \mathbb{P}(X_p \in \mathcal{A})\mathbb{P}(X_p \in \mathcal{B})$, i.e., $\mathbb{P}\big(X_p \in \mathcal{A} \mid X_p \in \mathcal{B}\big) \geqslant \mathbb{P}(X_p \in \mathcal{A})$, i.e., 'increasing properties are positively correlated'.

Proof. We use induction on $n = |X|$. Consider the base case $n = 1$: we may take $X = \{1\} = [1]$. The three possible up-sets are $\emptyset$, $\{\{1\}\}$ and $\{\emptyset, \{1\}\}$. If

$\mathcal{A}$ or $\mathcal{B}$ is $\emptyset$ then the RHS of (13) is 0. If $\mathcal{A}$ or $\mathcal{B}$ is $\{\emptyset, \{1\}\}$ then (13) holds at equality. It remains only to consider the case $\mathcal{A} = \mathcal{B} = \{\{1\}\}$: but now

$$\mathbb{P}_p(\mathcal{A} \cap \mathcal{B}) = p \geq p^2 = \mathbb{P}_p(\mathcal{A})\mathbb{P}_p(\mathcal{B}).$$

Now suppose that $|X| = n \geqslant 2$ and that the result holds for smaller sets X. Without loss of generality, let $X = [n] = \{1, 2, \ldots, n\}$.

For any $\mathcal{C} \subseteq \mathcal{P}(X)$ let

$$\mathcal{C}_0 = \{C \in \mathcal{C} : n \notin C\} \subseteq \mathcal{P}([n-1])$$

and

$$\mathcal{C}_1 = \{C \setminus \{n\} : C \in \mathcal{C},\, n \in C\} \subseteq \mathcal{P}([n-1]).$$

Thus $\mathcal{C}_0$ and $\mathcal{C}_1$ correspond to the subsets of $\mathcal{C}$ not containing and containing n respectively, except that for $\mathcal{C}_1$ we delete n from every set to obtain a collection of subsets of $[n-1]$.

Note that

$$\mathbb{P}_p(\mathcal{C}) = (1-p)\mathbb{P}_p(\mathcal{C}_0) + p\mathbb{P}_p(\mathcal{C}_1). \tag{14}$$

More precisely,

$$\mathbb{P}_p^{[n]}(\mathcal{C}) = (1-p)\mathbb{P}_p^{[n-1]}(\mathcal{C}_0) + p\mathbb{P}_p^{[n-1]}(\mathcal{C}_1).$$

Suppose now that $\mathcal{A}$ and $\mathcal{B} \subseteq \mathcal{P}([n])$ are up-sets. Then $\mathcal{A}_0$, $\mathcal{A}_1$, $\mathcal{B}_0$ and $\mathcal{B}_1$ are all up-sets in $\mathcal{P}([n-1])$. Also, $\mathcal{A}_0 \subseteq \mathcal{A}_1$ and $\mathcal{B}_0 \subseteq \mathcal{B}_1$. Let $a_0 = \mathbb{P}_p(\mathcal{A}_0)$ etc, so certainly $a_0 \leqslant a_1$ and $b_0 \leqslant b_1$.

Since $(\mathcal{A} \cap \mathcal{B})_i = \mathcal{A}_i \cap \mathcal{B}_i$, by (14) and the induction hypothesis we have

$$\begin{aligned}
\mathbb{P}_p(\mathcal{A} \cap \mathcal{B}) &= (1-p)\mathbb{P}_p((\mathcal{A} \cap \mathcal{B})_0) + p\mathbb{P}_p((\mathcal{A} \cap \mathcal{B})_1) \\
&= (1-p)\mathbb{P}_p(\mathcal{A}_0 \cap \mathcal{B}_0) + p\mathbb{P}_p(\mathcal{A}_1 \cap \mathcal{B}_1) \\
&\geqslant (1-p)a_0 b_0 + p a_1 b_1 = x,
\end{aligned}$$

say. On the other hand

$$\mathbb{P}_p(\mathcal{A})\mathbb{P}_p(\mathcal{B}) = \big((1-p)a_0 + p a_1\big)\big((1-p)b_0 + p b_1\big) = y,$$

say. So it suffices to show that $x \geqslant y$. But

$$\begin{aligned}
x - y &= \big((1-p) - (1-p)^2\big)a_0 b_0 - p(1-p)a_0 b_1 - p(1-p)a_1 b_0 + (p - p^2)a_1 b_1 \\
&= p(1-p)(a_1 - a_0)(b_1 - b_0) \geqslant 0,
\end{aligned}$$

recalling that $a_0 \leqslant a_1$ and $b_0 \leqslant b_1$. $\square$

Harris's Lemma has an immediate corollary concerning two down-sets, or one up- and one down-set.

Corollary 6.2. *If $\mathcal{U}$ is an up-set and $\mathcal{D}_1$ and $\mathcal{D}_2$ are down-sets, then*

$$\mathbb{P}_p(\mathcal{U} \cap \mathcal{D}_1) \leqslant \mathbb{P}_p(\mathcal{U})\mathbb{P}_p(\mathcal{D}_1),$$

and

$$\mathbb{P}_p(\mathcal{D}_1 \cap \mathcal{D}_2) \geqslant \mathbb{P}_p(\mathcal{D}_1)\mathbb{P}_p(\mathcal{D}_2).$$

Proof. Exercise, using the fact that $\mathcal{D}_i^{\mathrm{c}}$ is an up-set. □

6.2 Janson's inequalities

We have shown (e.g., from the Chernoff bounds) that, roughly speaking, if we have many independent events and the expected number that hold is large, then the probability that none holds is very small. What if our events are not quite independent, but each 'depends on' only a few others?

Let $E_1, \ldots, E_k$ be sets of possible edges of $G(n,p)$, and let A_i be the event that all edges in E_i are present in $G(n,p)$. Note that each A_i is an up-set. Let X be the number of A_i that hold. (For example, if the E_i list all possible edge sets of triangles, then X is the number of triangles in $G(n,p)$.)

As usual, let $\mu = \mathbb{E}[X] = \sum_i \mathbb{P}(A_i)$.

As in Chapter 2, write $i \sim j$ if $i \neq j$ and A_i and A_j are dependent, i.e., if $i \neq j$ and $E_i \cap E_j \neq \emptyset$, and let

$$\Delta = \sum_i \sum_{j \sim i} \mathbb{P}(A_i \cap A_j).$$

Theorem 6.3. *In the setting above, we have $\mathbb{P}(X = 0) \leqslant e^{-\mu + \Delta/2}$.*

Before turning to the proof, note that

$$\begin{aligned}
\mathbb{P}(X=0) &= \mathbb{P}(A_1^{\mathrm{c}} \cap \cdots \cap A_k^{\mathrm{c}}) \\
&= \mathbb{P}(A_1^{\mathrm{c}})\mathbb{P}(A_2^{\mathrm{c}} \mid A_1^{\mathrm{c}}) \cdots \mathbb{P}(A_k^{\mathrm{c}} \mid A_1^{\mathrm{c}} \cap \cdots \cap A_{k-1}^{\mathrm{c}}) \\
&\geqslant \prod_{i=1}^{k} \mathbb{P}(A_i^{\mathrm{c}}) = \prod_{i=1}^{k} (1 - \mathbb{P}(A_i)),
\end{aligned}$$

where we used Harris's Lemma and the fact that the intersection of two or more down-sets is again a down-set. In the (typical) case that all $\mathbb{P}(A_i)$ are small, the final bound is roughly $e^{-\sum \mathbb{P}(A_i)} = e^{-\mu}$, so (if Δ is small), Theorem 6.3 is saying that the probability that $X = 0$ is not much larger than the minimum it could possibly be.

Proof. Let $r_i = \mathbb{P}(A_i \mid A_1^{\mathrm{c}} \cap \cdots \cap A_{i-1}^{\mathrm{c}})$. Note that

$$\mathbb{P}(X=0) = \mathbb{P}(A_1^{\mathrm{c}} \cap \cdots \cap A_k^{\mathrm{c}}) = \prod_{i=1}^{k} (1-r_i). \tag{15}$$

Our aim is to show that r_i is not much smaller than $\mathbb{P}(A_i)$.

Fix i, and let D_1 be the intersection of those A_j^{c} where $j < i$ and $j \sim i$. Let D_0 be the intersection of those A_j^{c} where $j < i$ and $j \not\sim i$. Then

$$\begin{aligned} r_i = \mathbb{P}(A_i \mid D_0 \cap D_1) = \frac{\mathbb{P}(A_i \cap D_0 \cap D_1)}{\mathbb{P}(D_0 \cap D_1)} &\geqslant \frac{\mathbb{P}(A_i \cap D_0 \cap D_1)}{\mathbb{P}(D_0)} \\ = \mathbb{P}(A_i \cap D_1 \mid D_0) = \mathbb{P}(A_i \mid D_0) &- \mathbb{P}(A_i \cap D_1^{\mathrm{c}} \mid D_0). \end{aligned} \tag{16}$$

Now D_0 depends only on the presence of edges in $\bigcup_{j \not\sim i} E_j$, which is disjoint from E_i. Hence

$$\mathbb{P}(A_i \mid D_0) = \mathbb{P}(A_i). \tag{17}$$

For the other term, D_1 is a down-set, so D_1^{c} and $A_i \cap D_1^{\mathrm{c}}$ are up-sets. Since D_0 is a down-set, Corollary 6.2 gives

$$\begin{aligned} \mathbb{P}(A_i \cap D_1^{\mathrm{c}} \mid D_0) &\leqslant \mathbb{P}(A_i \cap D_1^{\mathrm{c}}) \\ &= \mathbb{P}\left(A_i \cap \bigcup_{j<i,\, j\sim i} A_j \right) \\ &= \mathbb{P}\left(\bigcup_{j<i,\, j\sim i} (A_i \cap A_j) \right) \\ &\leqslant \sum_{j<i,\, j\sim i} \mathbb{P}(A_i \cap A_j). \end{aligned}$$

With (16) and (17) this gives

$$r_i \geqslant \mathbb{P}(A_i) - \sum_{j<i,\, j\sim i} \mathbb{P}(A_i \cap A_j).$$

By (15) we thus have

$$\begin{aligned} \mathbb{P}(X=0) &= \prod (1-r_i) \leqslant \prod e^{-r_i} = \exp\Big(-\sum r_i\Big) \\ &\leqslant \exp\left(-\sum_{i=1}^{k} \mathbb{P}(A_i) + \sum_i \sum_{j\sim i,\, j<i} \mathbb{P}(A_i \cap A_j) \right) \\ &= \exp(-\mu + \Delta/2). \end{aligned}$$

$\square$

When Δ is much larger than μ, Theorem 6.3 is not very useful. But there is a trick to deduce something from it in this case.

Theorem 6.4. *Under the assumptions of Theorem 6.3, if $\Delta \geqslant \mu$ then $\mathbb{P}(X=0) \leqslant e^{-\frac{\mu^2}{2\Delta}}$.*

Proof. For any $S \subseteq [k]$, by Theorem 6.3 we have

$$\mathbb{P}(X=0) = \mathbb{P}\Big(\bigcap_{i=1}^{k} A_i^{\mathrm{c}}\Big) \leqslant \mathbb{P}\Big(\bigcap_{i\in S} A_i^{\mathrm{c}}\Big) \leqslant e^{-\mu_S+\Delta_S/2}, \tag{18}$$

where

$$\mu_S = \sum_{i\in S} \mathbb{P}(A_i) = \sum_{i=1}^{k} I_{\{i\in S\}} \mathbb{P}(A_i)$$

and

$$\Delta_S = \sum_{i\in S} \sum_{j\in S,\, j\sim i} \mathbb{P}(A_i \cap A_j) = \sum_{i} \sum_{j\sim i} I_{\{i,j\in S\}} \mathbb{P}(A_i \cap A_j).$$

Suppose now that $0 \leqslant r \leqslant 1$, and let S be the random subset of $[k]$ obtained by selecting each element independently with probability r. Then μ_S and Δ_S become random variables. By linearity of expectation we have

$$\mathbb{E}[\mu_S] = \sum_i r\mathbb{P}(A_i) = r\mu$$

and

$$\mathbb{E}[\Delta_S] = \sum_i \sum_{j\sim i} \mathbb{P}(A_i \cap A_j)\mathbb{P}(i,j \in S) = r^2\Delta.$$

Thus $\mathbb{E}[\mu_S - \Delta_S/2] = r\mu - r^2\Delta/2$.

Since a random variable cannot always be smaller than its mean, there exists *some* set S such that $\mu_S - \Delta_S/2 \geqslant r\mu - r^2\Delta/2$. Applying (18) to *this particular* set S it follows that

$$\mathbb{P}(X=0) \leqslant e^{-r\mu + r^2\Delta/2}.$$

This bound is valid for any $0 \leqslant r \leqslant 1$; to get the best result we optimize, which simply involves setting $r = \mu/\Delta \leqslant 1$. Then we obtain

$$\mathbb{P}(X=0) \leqslant e^{-\frac{\mu^2}{\Delta} + \frac{\mu^2}{2\Delta}} = e^{-\frac{\mu^2}{2\Delta}}.$$

$\square$

Together Theorems 6.3 and 6.4 give the following.

Corollary 6.5. *Under the assumptions of Theorem 6.3*

$$\mathbb{P}(X = 0) \leqslant \exp\bigl(-\min\{\mu/2, \mu^2/(2\Delta)\}\bigr).$$

Proof. For $\Delta < \mu$ apply Theorem 6.3; for $\Delta \geqslant \mu$ apply Theorem 6.4. $\square$

How do the second moment method and Janson's inequalities compare? Suppose that X is the number of events A_i that hold, let $\mu = \mathbb{E}[X]$, and let $\Delta = \sum_i \sum_{j \sim i} \mathbb{P}(A_i \cap A_j)$, as in the context of Corollary 2.4. Then Corollary 2.4 says that if $\mu \to \infty$ and $\Delta = o(\mu^2)$ (i.e. $\mu^2/\Delta \to \infty$), then $\mathbb{P}(X = 0) \to 0$. More concretely, if $\mu \geqslant L$ and $\mu^2/\Delta \geqslant L$, then the proof of Corollary 2.4 gives

$$\mathbb{P}(X = 0) \leqslant 2/L.$$

Janson's inequality, in the form of Corollary 6.5, has more restrictive assumptions: the events A_i have to be events of a specific type. When this holds, the Δ there is the same Δ as before. When $\mu \geqslant L$ and $\mu^2/\Delta \geqslant L$, the conclusion is that

$$\mathbb{P}(X = 0) \leqslant e^{-L/2}.$$

Both bounds imply that $\mathbb{P}(X = 0) \to 0$ when μ and μ^2/Δ both tend to infinity, but when Janson's inequalities apply, the concrete bound they give is *exponentially* stronger than that from the second moment method.

7 Clique and chromatic number of $G(n, p)$

We shall illustrate the power of Janson's inequality by using it to study the chromatic number of $G(n, p)$. The ideas are more important than the details of the calculations. We start by looking at something much simpler: the clique number.

Throughout this section p is *constant* with $0 < p < 1$.

Recall that the *clique number* $\omega(G)$ of a graph G is the maximum k such that G contains a copy of K_k. For $k = k(n)$ let X_k be the number of copies of K_k in $G = G(n, p)$, and

$$\mu_k := \mathbb{E}[X_k] = \binom{n}{k} p^{\binom{k}{2}}.$$

Note that

$$\frac{\mu_{k+1}}{\mu_k} = \binom{n}{k+1}\binom{n}{k}^{-1} p^{\binom{k+1}{2} - \binom{k}{2}} = \frac{n-k}{k+1} p^k, \tag{19}$$

which is a decreasing function of k. It follows that the ratio is at least 1 up to some point and then at most 1, so μ_k first increases from $\mu_0 = 1$, $\mu_1 = n$, ..., and then decreases.

We define

$$k_0 = k_0(n,p) = \min\{k : \mu_k < 1\}.$$

Lemma 7.1. *With $0 < p < 1$ fixed we have $k_0 \sim 2\log_{1/p} n = 2\frac{\log n}{\log(1/p)}$ as $n \to \infty$.*

Proof. Using standard bounds on the binomial coefficent $\binom{n}{k}$,

$$\left(\frac{n}{k}\right)^k p^{k(k-1)/2} \leqslant \mu_k \leqslant \left(\frac{en}{k}\right)^k p^{k(k-1)/2}.$$

Taking the kth root it follows that

$$\mu_k^{1/k} = \Theta\left(\frac{n}{k}p^{(k-1)/2}\right) = \Theta\left(\frac{n}{k}p^{k/2}\right).$$

Let $\varepsilon > 0$ be given.

If $k \leqslant (1-\varepsilon)2\log_{1/p} n$ then $k/2 \leqslant (1-\varepsilon)\log_{1/p} n$, so $(1/p)^{k/2} \leqslant n^{1-\varepsilon}$, i.e., $p^{k/2} \geqslant n^{-1+\varepsilon}$. Thus $\mu_k^{1/k}$ is at least a constant times $nn^{-1+\varepsilon}/\log n = n^{\varepsilon}/\log n$, so $\mu_k^{1/k} > 1$ if n is large. Hence $\mu_k > 1$, so $k_0 \geqslant k$.

Similarly, if $k \geqslant (1+\varepsilon)2\log_{1/p} n$ then $p^{k/2} \leqslant n^{-1-\varepsilon}$ and if n is large enough it follows that $\mu_k < 1$, so $k_0 \leqslant k$. So for any fixed ε we have

$$(1-\varepsilon)2\log_{1/p} n \leqslant k_0 \leqslant (1+\varepsilon)2\log_{1/p} n$$

if n is large enough, i.e., $k_0 \sim 2\log_{1/p} n$. □

Note for later that if $k \sim k_0$ then

$$\left(\frac{1}{p}\right)^k = n^{2+o(1)} \tag{20}$$

so from (19) we have

$$\frac{\mu_{k+1}}{\mu_k} = \frac{n - O(\log n)}{\Theta(\log n)} n^{-2+o(1)} = n^{-1+o(1)}. \tag{21}$$

Lemma 7.2. *With $0 < p < 1$ fixed we have $\mathbb{P}\big(\omega(G(n,p)) > k_0\big) \to 0$ as $n \to \infty$.*

Proof. We have $\omega(G(n,p)) > k_0$ if and only if $X_{k_0+1} > 0$, which has probability at most $\mathbb{E}[X_{k_0+1}] = \mu_{k_0+1}$. Now $\mu_{k_0} < 1$ by definition, so by (21) we have $\mu_{k_0+1} \leqslant n^{-1+o(1)}$, so $\mu_{k_0+1} \to 0$. □

Let Δ_k be the expected number of ordered pairs of distinct k-cliques sharing at least one edge. This is exactly the quantity Δ appearing in Corollaries 2.4 and 6.5 when X is the number of k-cliques.

Lemma 7.3. *Suppose that $k \sim k_0$. Then*

$$\frac{\Delta_k}{\mu_k^2} \leqslant \max\left\{ n^{-2+o(1)}, \frac{n^{-1+o(1)}}{\mu_k} \right\}.$$

In particular, if $\mu_k \to \infty$ then $\Delta_k = o(\mu_k^2)$.

Proof. We have

$$\Delta_k = \binom{n}{k} \sum_{s=2}^{k-1} \binom{k}{s} \binom{n-k}{k-s} p^{2\binom{k}{2} - \binom{s}{2}},$$

so

$$\frac{\Delta_k}{\mu_k^2} = \sum_{s=2}^{k-1} \alpha_s,$$

where

$$\alpha_s = \frac{\binom{k}{s}\binom{n-k}{k-s}}{\binom{n}{k}} p^{-\binom{s}{2}}.$$

We will show that the α_s first decrease then increase as s goes from 2 to $k-1$. Let

$$\beta_s = \frac{\alpha_{s+1}}{\alpha_s} = \frac{k-s}{s+1} \frac{k-s}{n-2k+s+1} p^{-s},$$

so

$$\beta_s = n^{-1+o(1)} \left(\frac{1}{p} \right)^s. \tag{22}$$

In particular, using (20) we have $\beta_s < 1$ for $s \leqslant k/4$, say, and $\beta_s > 1$ for $s \geqslant 3k/4$. In between we have $\beta_{s+1}/\beta_s \sim 1/p$, so $\beta_{s+1}/\beta_s \geqslant 1$, and β_s is increasing when s runs from $k/4$ to $3k/4$.

It follows that there is some $s_0 \in [k/4, 3k/4]$ such that $\beta_s \leqslant 1$ for $s \leqslant s_0$ and $\beta_s > 1$ for $s > s_0$. In other words, the sequence α_s decreases and then increases.

Hence, $\max\{\alpha_s : 2 \leqslant s \leqslant k-1\} = \max\{\alpha_2, \alpha_{k-1}\}$, so

$$\frac{\Delta_k}{\mu_k^2} = \sum_{s=2}^{k-1} \alpha_s \leqslant k \max\{\alpha_2, \alpha_{k-1}\} = n^{o(1)} \max\{\alpha_2, \alpha_{k-1}\}.$$

Either calculating directly, or using $\alpha_0 \leqslant 1$, $\alpha_2 = \alpha_0\beta_0\beta_1$, and the approximate formula for β_s in (22), one can check that $\alpha_2 \leq n^{-2+o(1)}$. Similarly, $\alpha_k = 1/\mu_k$ and $\alpha_{k-1} = \alpha_k/\beta_{k-1} = n^{-1+o(1)}/\mu_k$, using (22) and (20). $\square$

Theorem 7.4. *Let $0 < p < 1$ be fixed. Define $k_0 = k_0(n,p)$ as above, and let $G = G(n,p)$. Then*

$$\mathbb{P}\big(k_0 - 2 \leqslant \omega(G) \leqslant k_0\big) \to 1$$

Proof. The upper bound is Lemma 7.2. For the lower bound, let $k = k_0 - 2$. Note that $\mu_{k_0-1} \geqslant 1$ by the definition of k_0, so by (21) we have $\mu_k \geqslant n^{1-o(1)}$, and in particular $\mu_k \to \infty$. Then by Lemma 7.3 we have $\Delta_k = o(\mu_k^2)$. Hence by the second moment method (Corollary 2.4) we have $\mathbb{P}(\omega(G) < k) = \mathbb{P}(X_k = 0) \to 0$. $\square$

Note that we have 'pinned down' the clique number to one of three values; with only a very little more care, we can pin it down to at most two values. Indeed typically we can specify a single value (when μ_{k_0-1} is much larger than one, μ_{k_0} much smaller than one).

Using Janson's inequality, we can get a very tight bound on the probability that the clique number is significantly smaller than expected.

Theorem 7.5. *Under the assumptions of Theorem 7.4 we have*

$$\mathbb{P}\big(\omega(G) < k_0 - 3\big) \leqslant e^{-n^{2-o(1)}}.$$

Note that this is a truly tiny probability: the probability that $G(n,p)$ contains *no edges at all* is $(1-p)^{\binom{n}{2}} = e^{-\Theta(n^2)}$.

Proof. Let $k = k_0 - 3$. Then arguing as above we have $\mu_k \geqslant n^{2-o(1)}$. Hence by Lemma 7.3 we have $\Delta_k/\mu_k^2 \leqslant n^{-2+o(1)}$, so $\mu_k^2/\Delta_k \geqslant n^{2-o(1)}$. Thus by Janson's inequality (Corollary 6.5) we have $\mathbb{P}(X_k = 0) \leqslant e^{-n^{2-o(1)}}$. $\square$

Why is such a good error bound useful? Because it allows us to study the chromatic number, by showing that with high probability *every* subgraph of a decent size contains a fairly large independent set.

Theorem 7.6 (Bollobás)**.** *Let $0 < p < 1$ be constant and let $G = G(n,p)$. Then for any fixed $\varepsilon > 0$, whp*

$$(1-\varepsilon)\frac{n}{2\log_b n} \leqslant \chi(G) \leqslant (1+\varepsilon)\frac{n}{2\log_b n}$$

where $b = 1/(1-p)$.

Proof. Apply Theorem 7.4 to the complement G^c of G, noting that $G^c \sim G(n, 1-p)$. Writing $\alpha(G)$ for the independence number of G, we find that whp $\alpha(G) = \omega(G^c) \leqslant k_0(n, 1-p) \sim 2\log_b n$. Since $\chi(G) \geqslant n/\alpha(G)$, this gives the lower bound.

For the upper bound, let $m = n/(\log n)^2$, say. For each subset W of $V(G)$ with $|W| = m$, let E_W be the event that $G[W]$ contains an independent set of size at least $k = k_0(m, 1-p) - 3$. Note that

$$k \sim 2\log_b m \sim 2\log_b n.$$

Applying Theorem 7.5 to the complement of $G[W]$, which has the distribution of $G(m, 1-p)$, we have

$$\mathbb{P}(E_W^c) \leqslant e^{-m^{2-o(1)}} = e^{-n^{2-o(1)}}.$$

Let $E = \bigwedge_{|W|=m} E_W$. Considering the $\binom{n}{m} \leqslant 2^n$ possible sets W separately, the union bound gives

$$\mathbb{P}(E^c) = \mathbb{P}(\bigvee_W E_W^c) \leqslant 2^n e^{-n^{2-o(1)}} \to 0.$$

It follows that E holds whp. But when E holds one can colour by greedily choosing independent sets of size at least k for the colour classes, until at most m vertices remain, and then simply using one colour for each vertex. Since we use at most n/k sets of size at least k, this shows that, when E holds,

$$\chi(G(n,p)) \leqslant \frac{n}{k} + m = (1+o(1))\frac{n}{2\log_b n} + \frac{n}{(\log n)^2} \sim \frac{n}{2\log_b n},$$

completing the proof. □

8 Postscript: other models

(These concluding remarks are non-examinable.) There are several standard models of random graphs on the vertex set $[n] = \{1, 2, \ldots, n\}$. We have focussed on $G(n,p)$, where each possible edge is included independently with probability p.

The model originally studied by the founders of the theory of random graphs, Erdős and Rényi, is slightly different. Fix $n \geqslant 1$ and $0 \leqslant m \leqslant N = \binom{n}{2}$. The random graph $G(n,m)$ is the graph with vertex set $[n]$ obtained by choosing exactly m edges randomly, with all $\binom{N}{m}$ possible sets of m edges equally likely.

For most natural questions (but not, for example, 'is the number of edges even?'), $G(n,p)$ and $G(n,m)$ behave very similarly, provided we choose the density parameters in a corresponding way, i.e. we take $p \sim m/N$.

Often, we consider random graphs of different densities *simultaneously*. In $G(n,m)$, there is a natural way to do this, called the *random graph process*. This is the random sequence $(G_m)_{m=0,1,\ldots,N}$ of graphs on $[n]$ obtained by starting with no edges, and adding edges one-by-one in a random order, with all $N!$ orders equally likely. Note that each individual G_m has the distribution of $G(n,m)$: we take the first m edges in a random order, so all possibilities are equally likely. The key point is that in the *sequence* (G_m), we define all the G_m together, in such a way that if $m_1 < m_2$, then $G_{m_1} \subset G_{m_2}$. (This is called a 'coupling' of the distributions $G(n,m)$ for different m.)

There is a similar coupling in the $G(n,p)$ setting, the *continuous time random graph process*. This is the random 'sequence' $(G_t)_{t\in[0,1]}$ defined as follows: for each possible edge, let U_e be a random variable with the uniform distribution on the interval $[0,1]$, with the different U_e independent. Let the edge set of G_t be $\{e : U_e \leqslant t\}$. (Formally this defines a random function $t \mapsto G_t$ from $[0,1]$ to the set of graphs on $[n]$.) One can think of U_e as giving the 'time' at which the edge e is born; G_t consists of all edges born by time t. For any p, G_p has the distribution of $G(n,p)$, but again these distributions are coupled in the natural way: if $p_1 < p_2$ then $G_{p_1} \subseteq G_{p_2}$.

Of course there are many other random graph models not touched on in this course (as well as many more results about $G(n,p)$). These include other classical models, such as the 'configuration model' for random regular graphs, random geometric graphs, and also new 'inhomogeneous' models introduced as more realistic models for networks in the real world.