CHAPTER 2

First and second moment methods

In the previous lecture, we have seen two important principles that play a role in the probabilisic method: the averaging principle, and the counting principle. In today's lecture, we will see two tools that have many applications when combined with the counting principle.

1. Markov's inequality and Chebychev's inequality

THEOREM 2.1 (Markov's inequality). *Suppose that* $\mathbf{X} \geq 0$ *has finite expectation, then* $\mathbb{P}\left(\mathbf{X} \geq k\right) \leq \frac{\mathbb{E}[\mathbf{X}]}{k}$ *for all* $k > 0$.

PROOF. Let $k > 0$ be given. Note that

$$\mathbf{X} \geq \mathbb{1}_{\{\mathbf{X} \geq k\}} \mathbf{X} \geq \mathbb{1}_{\{\mathbf{X} \geq k\}} k. \tag{2.1}$$

The claim follows by taking expectations of the expressions on the extreme sides, and rearranging. □

A very useful consequence of Markov's inequality is the following.

COROLLARY 2.2. *Let* $\mathbf{X} \geq 0$ *be an integer-valued random variable with finite mean, then* $\mathbb{P}\left(\mathbf{X} = 0\right) \geq 1 - \mathbb{E}\left[\mathbf{X}\right]$.

Recall that the variance of a random variable is a measure of variability of a random variable. The following result, known as Chebychev's inequality, quantifies this intuition. It shows that if a random variable has finite variance, typical realisations will be close to its expected value. THis phenomenon is called **concentration of measure**.

THEOREM 2.3 (Chebychev's inequality). *Let* $\mathbf{X}$ *be a random variable with finite expectation* $\mu = \mathbb{E}\left[\mathbf{X}\right]$ *and finite second moment, then* $\mathbb{P}\left(|\mathbf{X} - \mu| \geq k\right) \leq \frac{Var(\mathbf{X})}{k^2}$.

PROOF. Chebychev's inequality follows from Markov's inequality applied to the random variable $(\mathbf{X} - \mu)^2$. □

Note that Chebychev's inequality works for very general random variables. Later, we will see other concentration of measure results that make stronger assumptions, and give stronger results as well.

For integer-valued random variables, Chebychev's inequality has the following consequence.

COROLLARY 2.4. *Suppose that* $\mathbf{X}$ *is integer-valued with expectation* μ *and finite second moment, then* $\mathbb{P}\left(\mathbf{X} = 0\right) \leq \frac{Var(\mathbf{X})}{\mu^2}$.

2. Graphs with large girth and chromatic number

The **chromatic number** $\chi(G)$ of a graph is the smallest number of colours necessary for a proper vertex-colouring. In general, the problem is computing the chromatic number of a graph is NP-complete, but for certain classes of graphs it is trivial.

For example, when G is a forest, $\chi(G) = 2$ (except for the degenerate case in which G does not contain any edges, in which case $\chi(G) = 1$). A natural generalisation of this result is that if G does not contain any short cycles, then it has small chromatic number as well. Indeed, if G does not contain any cycles of length less than ℓ, the $(\ell/2 - 1)$-neighbourhood of any vertex is a tree, and hence may be two-coloured.

The following result, first proved by Erdős [Erd59], may therefore be very surprising. The **girth** $g(G)$ of a graph is the length of its shortest cycle (or infinity if the graph is a forest).

THEOREM 2.5. *For every k and ℓ, there exists a graph G with $\chi(G) \geq k$ and $g(G) > \ell$.*

In the proof of this theorem, we will work with the independence number of a graph, rather than with its chromatic number. The two quantities are related via the following result:

LEMMA 2.6. *For every graph $G = (V, E)$, $\alpha(G)\chi(G) \geq |V|$.*

PROOF. Consider a proper vertex-colouring of G using $\chi(G)$ colours. The colour classes form a partition of the vertices in G into independent sets, each of which has size at most $\alpha(G)$. $\square$

PROOF OF THEOREM 2.5. Let us first give a brief outline of the proof. We will show that for sufficiently large n and suitably chosen p, with positive probability the random graph $\mathbf{G}(n,p)$ contains few small cycles, and has small independence number. Hence, by the counting principle, there is a graph with few small cycles, and small independence number. We then pass to a subgraph of this graph by removing from each small cycle a single vertex. This breaks all the small cycles that may be present, and does not increase the independence number; hence, the resulting graph has large girth and, by Lemma 2.6, large chromatic number. We will now fill in the details.

Fix some parameter $\theta < 1/\ell$, and set $p = n^{\theta-1}$. Let $\mathbf{X}_{\leq\ell}$ be the number of cycles of length at most ℓ in $\mathbf{G}(n,p)$.

The number of cycles of length j in K_n is $\frac{n!}{2j(n-j)!}$. This follows, since there are $\binom{n}{j}$ ways of choosing j vertices, which can then be ordered in $j!$ ways. Given any such set of ordered vertices, we can "start" the cycle at any of the j vertices, and reverse the order, accounting for the factor $2j$. It follows that

$$\mathbb{E}\left[\mathbf{X}_{\leq\ell}\right] = \sum_{j=3}^{\ell} \frac{n!}{2j(n-j)!} \leq \sum_{j=3}^{\ell} (np)^j \leq n^{\theta\ell}. \tag{2.2}$$

Note that this is not sufficient to prove that $\mathbf{X}_{\leq\ell} = 0$ with positive probability using Markov's inequality, but we are able to show that

$$\mathbb{P}\left(\mathbf{X}_{\leq\ell} \geq \frac{n}{2}\right) \leq \frac{\mathbb{E}\left[\mathbf{X}_{\leq\ell}\right]}{n/2} \leq 2n^{\theta\ell-1} = o(1). \tag{2.3}$$

In particular, for n sufficiently large, $\mathbf{G}(n,p)$ contains more than $n/2$ short cyles with probability at most $1/3$.

We will now turn to the independence number of $\mathbf{G}(n,p)$. Let $a = \frac{n}{2k}$. For a fixed set of a vertices, the probability that it is an independent set is $(1-p)^{\binom{a}{2}}$. Hence, by the union bound,

$$\begin{aligned}
\mathbb{P}\left(\alpha(\mathbf{G}(n,p)) \geq a\right) &\leq \binom{n}{a}(1-p)^{\binom{a}{2}} \\
&\leq n^a \mathrm{e}^{-p\binom{a}{2}} \\
&\leq \exp\left(\frac{n}{2k}\ln n - n^{\lambda-1}\frac{n^2}{8k^2}\right) = o(1),
\end{aligned} \tag{2.4}$$

so in particular, for sufficiently large n, it follows that $\alpha(\mathbf{G}(n,p)) \geq a$ with probability at most $1/3$.

By another application of the union bound, we obtain that, for sufficiently large n,

$$\mathbb{P}\left(\mathbf{X}_{\leq\ell} \geq \frac{n}{2} \text{ or } \alpha(\mathbf{G}(n,p)) \geq a\right) \leq \frac{1}{3} + \frac{1}{3} < 1. \tag{2.5}$$

By the counting principle, there exists a graph G on n vertices with fewer than $n/2$ cycles of length at most ℓ, and $\alpha(G) < a$. If we remove one vertex from each cycle in G, the resulting graph G' contains at least $n/2$ vertices, has girth $g(G') > \ell$, and $\alpha(G') < a$. By Lemma 2.6,

$$\chi(G') \geq \frac{|V(G')|}{\alpha(G')} > \frac{n/2}{n/(2k)} = k, \tag{2.6}$$

which concludes the proof. □

3. Clique number of the Erdős-Rényi random graph

3.1. The Erdős-Rényi random graph. Suppose that we start with the complete graph on n vertices, K_n. For each edge in this graph we flip a coin to decide whether we want to keep this edge: if it comes up heads (with probability p) we keep it, and if it comes up tails (with probability $1-p$) we discard the edge.

The resulting random graph is called the Erdős-Rényi random graph, and we denote it by $\mathbf{G}(n,p)$.[1] Note that in this model, the number of vertices is always equal to n, while the number of edges follows a binomial distribution with parameters n and p. If H is any fixed graph on n vertices and m edges, we have

$$\mathbb{P}\left(\mathbf{G}(n,p) = H\right) = p^m(1-p)^{\binom{n}{2}-m}. \tag{2.7}$$

In particular, when $p = 1/2$, the random graph $G(n,p)$ is uniformly distributed over all n-vertex graphs.

3.2. Clique number. If G is a graph, then its clique number $\omega(G)$ is the size of the largest clique contained in G. In this section, we analyse the clique number of the Erdős-Rényi random graph $\mathbf{G}(n,1/2)$. In particular, we will show that almost all random graphs have clique number close to $2\log n$.

Theorem 2.7. *For every $\varepsilon > 0$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left((2-\varepsilon)\log n \leq \omega(\mathbf{G}(n,1/2)) \leq (2+\varepsilon)\log n\right) = 0.$$

We will prove the theorem in two parts, corresponding to the upper bound (proved in Theorem 2.8) and the lower bond (proved in Theorem 2.9).

Theorem 2.8. *For every $\varepsilon > 0$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\omega(\mathbf{G}(n,1/2)) \geq (2+\varepsilon)\log n\right) = 0.$$

Proof. Define $k = (2+\varepsilon)\log n$, and let us introduce the random variable $\mathbf{X}_k$, which counts the number of k-cliques in $\mathbf{G}(n,1/2)$. Note that

$$\mathbf{X}_k = 0 \qquad \Longleftrightarrow \qquad \omega(\mathbf{G}(n,1/2)) < k. \tag{2.8}$$

Our goal is to show that $\mathbf{X}_k = 0$ with probability tending to 1, as this would immediately imply the result that we are after. We first compute the expectation of $\mathbf{X}_k$:

$$\mathbb{E}\left[\mathbf{X}_k\right] = \binom{n}{k} 2^{-\binom{k}{2}} \leq 2^{k\log n - \frac{k(k-1)}{2}}. \tag{2.9}$$

Substituting the value for k, we find

$$\begin{aligned} \mathbb{E}\left[\mathbf{X}_k\right] &\leq \exp_2\left((2+\varepsilon)\log^2 n - \frac{(2+\varepsilon)^2\log^2 n}{2} + \frac{(2+\varepsilon)\log n}{2}\right) \\ &= \exp_2\left(-(2\varepsilon+\varepsilon^2/2)\log^2 n + (1+\varepsilon/2)\log n\right) = o(1). \end{aligned} \tag{2.10}$$

Using Markov's inequality, we bound

$$\mathbb{P}\left(\mathbf{X}_k = 0\right) \geq 1 - \mathbb{E}\left[\mathbf{X}_k\right] = 1 - o(1) \tag{2.11}$$

Hence, with probability tending to 1, $\omega(\mathbf{G}(n,1/2)) < k$, which proves the theorem. □

[1] Actually, Erdős and Rényi [ER59] introduced the different but related model $\mathbf{G}(n,M)$, in which of all possible edges exactly M are chosen at random. The model $\mathbf{G}(n,p)$, that is prevalent in modern literature, is due to Gilbert [Gil59].

THEOREM 2.9. *For any $\varepsilon > 0$, we have*

$$\lim_{n\to\infty} \mathbb{P}\left(\omega(\mathbf{G}(n,1/2)) \le (2-\varepsilon)\log n\right) = 0.$$

It is not hard to show that if $k = (2-\varepsilon)\log n$, the expected number of k-cliques in $G(n, 1/2)$ tends to infinity. Therefore, Markov's inequality is not the right tool to prove the corresponding lower bound on $\omega(\mathbf{G}(n, 1/2))$. Fortunately, Chebychev's inequality gives the right type of result, but it requires a little more work, as now we have to compute the second moment of the number of k-cliques.

PROOF. Let $k = (2-\varepsilon)\log n$. For a k-subset of vertices S, let $\mathbf{X}_S$ be the indicator random variable of the event that S forms a clique in $\mathbf{G}(n, 1/2)$. The number of k-cliques in $\mathbf{G}(n, 1/2)$ is given by $\mathbf{X}_k = \sum_S \mathbf{X}_S$, where the sum is over k-subsets of vertices. As in the proof of Theorem 2.8,

$$\mathbb{E}\left[\mathbf{X}_k\right] = \binom{n}{k} 2^{-\binom{k}{2}}. \tag{2.12}$$

For an application of Chebychev's inequality, we need to bound the variance of $\mathbf{X}_k$. Note that

$$\begin{aligned}
\operatorname{Var}\left(\mathbf{X}_k\right) &= \mathbb{E}\left[\mathbf{X}_k^2\right] - \left(\mathbb{E}\left[\mathbf{X}_k\right]\right)^2 \\
&= \mathbb{E}\left[\sum_S \sum_T \mathbf{X}_S \mathbf{X}_T\right] - \left(\mathbb{E}\left[\sum_S \mathbf{X}_S\right]\right)^2 \\
&= \sum_S \sum_T \left(\mathbb{E}\left[\mathbf{X}_S \mathbf{X}_T\right] - \mathbb{E}\left[\mathbf{X}_S\right]\mathbb{E}\left[\mathbf{X}_T\right]\right),
\end{aligned} \tag{2.13}$$

where in the final step we used linearity of expectation.

It will be useful to further split up the sum by the size of the intersection $S \cap T$, so let us call $\ell = |S \cap T|$. Note that if $\ell \le 2$, then $\mathbf{X}_S$ and $\mathbf{X}_T$ are independent, which implies that $\mathbb{E}\left[\mathbf{X}_S \mathbf{X}_T\right] - \mathbb{E}\left[\mathbf{X}_S\right]\mathbb{E}\left[\mathbf{X}_T\right]$. Therefore, it suffices to sum over $\ell \ge 2$ only. This leads to

$$\begin{aligned}
\operatorname{Var}\left(\mathbf{X}_k\right) &= \sum_{\ell=2}^{k} \sum_S \sum_{T:|S\cap T|=\ell} \left(\mathbb{E}\left[\mathbf{X}_S \mathbf{X}_T\right] - \mathbb{E}\left[\mathbf{X}_S\right]\mathbb{E}\left[\mathbf{X}_T\right]\right) \\
&\le \sum_{\ell=2}^{k} \sum_S \sum_{T:|S\cap T|=\ell} \mathbb{E}\left[\mathbf{X}_S \mathbf{X}_T\right].
\end{aligned} \tag{2.14}$$

Given the size of the intersection ℓ, the summand does not depend on the particular choice of S and T, which means that we can simplify this expression even further. For fixed ℓ, there are $\binom{n}{k}$ choices for S. As $S \cap T$ is a subset of S of size ℓ, it follows that after choosing S, there are $\binom{k}{\ell}$ ways of choosing $S \cap T$; the intersection can be extended in $\binom{n-k}{k-\ell}$ ways to the full set T. Hence, it follows that

$$\operatorname{Var}\left(\mathbf{X}_k\right) \le \sum_{\ell=2}^{k} \binom{n}{k}\binom{k}{\ell}\binom{n-k}{k-\ell} \mathbb{E}\left[\mathbf{X}_S \mathbf{X}_T\right], \tag{2.15}$$

where now S and T are generic k-subsets that intersect in ℓ elements.

Note that the product $\mathbf{X}_S \mathbf{X}_T$ takes only the values 0 and 1, and that it is equal to 1 precisely when both S and T are cliques, which happens precisely when all the edges spanned by vertices in S and those spanned by vertices in T are present. The number of such edges is $2\binom{k}{2} - \binom{\ell}{2}$. It follows that

$$\mathbb{E}\left[\mathbf{X}_S \mathbf{X}_T\right] = \mathbb{P}\left(S \text{ and } T \text{ are both cliques}\right) = 2^{-2\binom{k}{2}+\binom{\ell}{2}}, \tag{2.16}$$

which we can plug into (2.15)

$$\operatorname{Var}\left(\mathbf{X}_k\right) \le \sum_{\ell=2}^{k} \binom{n}{k}\binom{k}{\ell}\binom{n-k}{k-\ell} 2^{-2\binom{k}{2}+\binom{\ell}{2}}. \tag{2.17}$$

We now have all the ingredients necessary for application of Corollary 2.4.

$$\mathbb{P}\left(\mathbf{X}_k = 0\right) \leq \frac{\mathrm{Var}\left(\mathbf{X}_k\right)}{\left(\mathbb{E}\left[\mathbf{X}_k\right]\right)^2} \leq \frac{\sum_{\ell=2}^{k}\binom{n}{k}\binom{k}{\ell}\binom{n-k}{k-\ell}2^{-2\binom{k}{2}+\binom{k}{\ell}}}{\left(\binom{n}{k}2^{-\binom{k}{2}}\right)^2} = \sum_{\ell=2}^{k}\frac{\binom{k}{\ell}\binom{n-k}{k-\ell}2^{\binom{\ell}{2}}}{\binom{n}{k}}. \tag{2.18}$$

The remainder of the proof is dedicated to showing that this upper bound tends to zero as n tends to infinity. We apply a number of standard bounds on binomial coefficients, in particular, $\binom{k}{\ell} \leq k^\ell$, $\binom{n-k}{k-\ell} \leq \frac{(n-k)^{k-\ell}}{(k-\ell)!}$, $\binom{\ell}{2} \leq \ell^2/2$, and $\binom{n}{k} \geq \frac{(n-k)^k}{k!}$. It follows that

$$\sum_{\ell=2}^{k}\frac{\binom{k}{\ell}\binom{n-k}{k-\ell}2^{\binom{\ell}{2}}}{\binom{n}{k}} \leq \sum_{\ell=2}^{k}\frac{k^{2\ell}}{(n-k)^\ell}2^{\ell^2/2}. \tag{2.19}$$

For n sufficiently large, we have $n-k \geq n/2$, so the right-hand side of the previous display is at most

$$\begin{aligned}\sum_{\ell=2}^{k}k^{2\ell}\left(\frac{2}{n}\right)^\ell 2^{\ell^2/2} &= \sum_{\ell=2}^{k}\exp_2\left(\ell(2\log k - (\log n - 1) + \ell/2\right)\\ &\leq \sum_{\ell=2}^{k}\exp_2\left(\ell(1 + 2\log k - (\varepsilon/2)\log n)\right),\end{aligned} \tag{2.20}$$

where the final step follows since $\ell/2 \leq k/2 = (1-\varepsilon/2)\log n$. Finally, for n sufficiently large, $1 + 2\log k \leq (\varepsilon/4)\log n$. It follows that the sum in the right-hand side of (2.20) is at most

$$k2^{-(\varepsilon/2)\log n} = \frac{(2-\varepsilon)\log n}{n^{\varepsilon/2}} = o(1), \tag{2.21}$$

which proves that with probability tending to 1, the random graph $\mathbf{G}(n, 1/2)$ contains a k-clique as n tends to infinity. □

Bibliography

[ER59] P. Erdős and A. Rényi. "On random graphs I". In: *Publ. Math. Debrecen* 6 (1959), pp. 290–297.

[Erd59] P. Erdős. "Graph theory and probability". In: *Canad. J. Math* 11 (1959), pp. 34–38.

[Gil59] Edgar N. Gilbert. "Random graphs". In: *The Annals of Mathematical Statistics* (1959), pp. 1141–1144.