

5 The Chernoff bound

5.1 Setup and proof

The second moment essentially compares the values of $\mathbb{E}[X]$ and $\mathbb{E}[X^2]$ to each other. Why do we not take higher moments? In general, if we have independent random variables

$$X = X_1 + \cdots + X_n,$$

we can look at $p(X)$, which is some polynomial in X , and apply Markov's inequality in the same way that we did for Chebyshev's inequality. It turns out that if we're allowed to look at arbitrarily high-degree polynomials, it's usually better to just look at the following object:

Definition 5.1

The **moment generating function** of a random variable X is a function of t

$$\mathbb{E}[e^{tX}] = 1 + t\mathbb{E}[X] + \frac{t^2\mathbb{E}[X^2]}{2} + \cdots$$

What are its applications?

Theorem 5.2 (Chernoff bound)

Let $S_n = X_1 + \cdots + X_n$, where $X_i = \pm 1$ uniformly and independently. Then for all $\lambda > 0$,

$$\Pr(S_n \geq \lambda\sqrt{n}) \leq e^{-\lambda^2/2}.$$

This gives better tail decay! While the second moment method only gave us polynomial decay (right hand side of the form $\frac{1}{\lambda^2}$), this is exponential decay instead.

Proof. Let $t \geq 0$ be a real number, and consider the moment generating function

$$\mathbb{E}[e^{tS_n}].$$

Since S_n is a sum of random independent variables, this is

$$\mathbb{E}[e^{tX_1 + \cdots + tX_n}] = \mathbb{E}[e^{tX_1}]\mathbb{E}[e^{tX_2}] \cdots \mathbb{E}[e^{tX_n}] = \mathbb{E}[e^{tX_1}]^n = \left(\frac{e^{-t} + e^t}{2}\right)^n.$$

Since $\frac{e^{-t} + e^t}{2} \leq e^{t^2/2}$ by comparing coefficients of the Taylor expansions:

$$\frac{1}{(2n)!} \leq \frac{1}{n!2^n},$$

our moment generating function is $\leq e^{nt^2/2}$, and by Markov's inequality,

$$\Pr(S_n \geq \lambda\sqrt{n}) \leq \frac{\mathbb{E}[e^{tS_n}]}{e^{t\lambda\sqrt{n}}} \leq e^{-t\lambda\sqrt{n} + t^2n/2}$$

and setting $t = \frac{\lambda}{\sqrt{n}}$ gives the desired result. □

By symmetry, we have a bound for $S_n \leq -\lambda\sqrt{n}$ as well, so combining these, we obtain the following:

Corollary 5.3

Using the definition of S_n above,

$$\Pr(|S_n| \geq \lambda\sqrt{n}) \leq 2e^{-\lambda^2/2}$$

for all $\lambda > 0$.

But notice that S_n converges to a Gaussian distribution for large n , so something similar should be true for Gaussians as well. This is indeed true:

Fact 5.4

For the standard normal distribution $Z \sim N(0, 1)$, for all $\lambda \geq 0$,

$$\Pr(Z \geq \lambda) = \Pr(e^{tZ} \geq e^{t\lambda}) \leq e^{-t\lambda} \mathbb{E}[e^{tZ}] = e^{-t\lambda + t^2/2} \leq e^{-\lambda^2/2}$$

by taking $t = \lambda$.

This is pretty tight: it turns out that in general we're only losing a $c\sqrt{\lambda}$, and in reality we actually have

$$\Pr(Z \geq \lambda) \sim \frac{e^{-\lambda^2/2}}{\sqrt{2\pi\lambda}}.$$

See Appendix A of the textbook for different instantiations of the Chernoff bound. Similarly, we can find exponential decay for Bernoulli variables where $p \neq \frac{1}{2}$:

Fact 5.5

If Y is a sum of independent Bernoulli variables (with not necessarily the same probability), then for all $\varepsilon > 0$,

$$\Pr(|Y - \mathbb{E}[Y]| \geq \varepsilon \mathbb{E}[Y]) \leq 2e^{-C_\varepsilon \mathbb{E}[Y]}$$

for some constant $C_\varepsilon > 0$.

5.2 An application: discrepancy

Theorem 5.6

Let H be a k -uniform hypergraph with m edges. Then we can color the vertices red and blue so that every edge has an $O(\sqrt{k \log m})$ difference in the number of red and blue vertices.

Proof. Color each vertex uniformly at random: put ± 1 on every vertex. Then every edge is of the form $S_m = X_1 + \dots + X_m$ where all $X_m = \pm 1$, so by the Chernoff bound, the probability $|S_m|$ exceeds $\lambda\sqrt{k}$ is at most $2e^{-\lambda^2/2}$. Note that the absolute value of S_n is exactly the difference between the number of red and blue vertices.

In particular, we can now do a union bound: if $2me^{-\lambda^2/2} \leq 1$, then there exists a graph where none of the bad events happen. Inverting this gives the desired result. \square

This kind of log term usually comes from the Chernoff bound. If we only used the second moment method, we'd have a much worse result - polynomial instead of exponential?

Well, what's the truth? Suppose $m = k$: this theorem gives us a difference of $\sqrt{k \log k}$ between the red and blue vertices. But we can do much better:

Fact 5.7

Spencer's paper "Six standard deviations suffice" says that when $m = k$, we can get at most $6\sqrt{k}$ difference between the number of red and blue vertices on every edge.

5.3 Chromatic number and graph minors

Let's start with some graph theory results for motivation:

Proposition 5.8 (Kuratowski's theorem)

If G is not planar, then it contains a $K_{3,3}$ or K_5 subdivision.

Here, a **subdivision** of a graph H is H with some of the edges chopped into smaller pieces. Basically, K_{5s} and $K_{3,3s}$ are not allowed, nor are those graphs with extra vertices along the edges. There's another similar theorem that is actually equivalent to Kuratowski's theorem:

Proposition 5.9 (Wagner's theorem)

If G is not planar, then G contains a $K_{3,3}$ - or K_5 -minor.

Here, H is a **minor** of G if it can be obtained from deleting edges/vertices or **contracting** an edge. (Basically, take the two vertices of an edge and squish them together.) In particular, K_5 is a minor of a K_5 -subdivision.

Theorem 5.10 (Four-color theorem)

If $\chi(G) \geq 5$, then G is not planar.

In particular, if $\chi(G) \geq 5$, it must contain a $K_{3,3}$ -minor or a K_5 -minor. Having a K_5 -minor seems pretty relevant, since we need 5 colors to color a K_5 . But $K_{3,3}$ doesn't seem like as much of an obstruction, and that's quantified in the statement below:

Fact 5.11

If $\chi(G) \geq 5$, then G contains a K_5 -minor.

Well, does this hold if we replace 5 with other numbers?

Conjecture 5.12 (Hadwiger's conjecture)

If $\chi(G) \geq t$, then G contains a K_t -minor.

Many people consider this to be the biggest open problem in graph theory! We do have some small cases resolved: $t = 1, 2$ are trivial. $t = 3$ is not too hard: If G has no K_3 -minor, it is a tree, which is 2-colorable. $t = 4$ requires more work but is elementary, and $t = 5$ is equivalent to the four-color theorem (for which we only have a computer-assisted proof). But Robertson, Seymour, and Thomas showed that the four-color theorem actually implies $t = 6$, and all $t \geq 7$ are open.

Are there variations on this conjecture?

Proposition 5.13 (Hajos conjecture: even stronger)

If $\chi(G) \geq t$, then G has a K_t -subdivision.

Unfortunately, this is false. In fact, by the probabilistic method, Erdős and Fajtlowicz showed that $G(n, \frac{1}{2})$ fails this condition with high probability:

Theorem 5.14

With high probability, $G(n, \frac{1}{2})$ has chromatic number $\chi(G) \geq (1 + o(1)) \frac{n}{2 \log_2 n}$ and no $K_{\lceil 10\sqrt{n} \rceil}$ -subdivision.

So the theorem is very false in the relation between the two parameters, as well as in its likelihood! Note that the Hajos conjecture is still true for small t : it just fails for larger t due to the arguments below.

Proof. We already lower bounded by upper bounding color classes as independent sets:

$$\chi(G) \geq \frac{n}{\alpha(G)} \sim \frac{n}{2 \log_2 n}$$

with high probability. Let's work on the second part.

Suppose we have a K_t -subdivision, where $t = \lceil 10\sqrt{n} \rceil$. Out of the $\binom{t}{2}$ edges in K_t , about half of them are not contained in G , so they must use up other vertices to form paths, and we don't have enough of those.

Let's do this more rigorously. Let G have a K_t subdivision $S \subset V$, where $|S| = t = \lceil 10\sqrt{n} \rceil$. At most n edges in the subdivision can be paths of at least 2 edges (rather than just straight lines between vertices), since each path takes up an external vertex, and all paths use distinct vertices by definition. So the number of edges E involved in the subdivision satisfies

$$E \geq \binom{t}{2} - n \geq \frac{3}{4} \binom{t}{2},$$

where the \geq comes from picking a large enough constant in $t = c\sqrt{n}$ (we chose $c = 10$). But this inequality fails with high probability, since we're supposed to have $\frac{1}{2} \binom{t}{2}$ edges only. Indeed, for every fixed t -vertex S , each edge appears with probability $\frac{1}{2}$, so the number of edges in the subgraph induced by S satisfies

$$\Pr\left(E \geq \frac{3}{4} \binom{t}{2}\right) \leq e^{-t^2/10}$$

by the Chernoff bound. Now by a union bound, ranging over all t -element subsets of vertices, the probability of any subdivision being possible is bounded above by

$$\binom{n}{t} e^{-t^2/10} < n^t e^{-t^2/10} = o(1),$$

so there must not be a K_t -subdivision with high probability. □

MIT OpenCourseWare
<https://ocw.mit.edu>

18.218 Probabilistic Method in Combinatorics
Spring 2019

For information about citing these materials or our Terms of Use, visit: <https://ocw.mit.edu/terms>.