The Probabilistic Method

Techniques

- Union bound
- Argument from expectation
- Alterations
- **The second moment method**
- The (Lovasz) Local Lemma

And much more

- Alon and Spencer, "The Probabilistic Method"
- Bolobas, "Random Graphs"

Second Moment Method: Main Idea

Use Chebyshev's Inequality.

Example 1: Distinct Subset Sums

- A set $A = \{a_1, \cdots, a_k\}$ of positive integers has distinct subset sums if the sums of all subsets of A are distinct
- $f(n) =$ maximum k for which there's a k-subset of $[n]$ having distinct subset sums
- Example: $A = \{2^i \mid 0 \leq i \leq \lg n\}$

$$f(n) \geq \lfloor \lg n \rfloor + 1$$

- Open Problem: (Erdős offered 500usd)

$$f(n) \leq \log_2 n + c?$$

- Simple information-theoretic bound:

$$2^k \leq nk \;\Rightarrow\; k < \lg n + \lg \lg n + O(1).$$

A Bound for $f(n)$ Using Second Moment Method

Line of thought

- Fix n and k-subset $A = \{a_1, \cdots, a_k\}$ with distinct subset sums
- $X =$ sum of random subset of A, $\mu = \mathsf{E}[X], \sigma^2 = \mathsf{Var}[X]$
- For any integer i,
$$\mathsf{Prob}[X = i] \in \left\{0, \frac{1}{2^k}\right\}$$
- By Chebyshev, for any $\alpha > 1$
$$\mathsf{Prob}[|X - \mu| \geq \alpha\sigma] \leq \frac{1}{\alpha^2} \Rightarrow \mathsf{Prob}[|X - \mu| < \alpha\sigma] \geq 1 - \frac{1}{\alpha^2}$$
- There are at most $2\alpha\sigma + 1$ integers within $\alpha\sigma$ of μ; hence,
$$1 - \frac{1}{\alpha^2} \leq \frac{1}{2^k}(2\alpha\sigma + 1)$$
- σ is a function of n and k

More Specific Analysis

$$\sigma^2 = \frac{a_1^2 + \cdots + a_k^2}{4} \leq \frac{n^2 k}{4} \;\Rightarrow\; \sigma \leq n\sqrt{k}/2$$

There are at most $(\alpha n\sqrt{k} + 1)$ within $\alpha\sigma$ of μ

$$1 - \frac{1}{\alpha^2} \leq \frac{1}{2^k}(\alpha n\sqrt{k} + 1)$$

Equivalently,

$$n \geq \frac{2^k\left(1 - \frac{1}{\alpha^2}\right) - 1}{\alpha\sqrt{k}}$$

Recall $\alpha > 1$, we get

$$k \leq \lg n + \frac{1}{2}\lg\lg n + O(1).$$

Example 2: $\mathcal{G}(n,p)$ Model and $\omega(G) \geq 4$ Property

$\mathcal{G}(n,p)$

Space of random graphs with n vertices, each edge (u,v) is included with probability p
Also called the Erdős-Rényi Model.

Question

Does a "typical" $G \in \mathcal{G}(n,p)$ satisfy a given property?

- Is G connected?
- Does G have a 4-clique?
- Does G have a Hamiltonian cycle?

Threshold Function

- As p goes from 0 to 1, $G \in \mathcal{G}(n, p)$ goes from "typically empty" to "typically full"
- Some property may become more likely or less likely
- The property *having a* 4*-clique* will be come more likely

Threshold Function

$f(n)$ is a threshold function for property P if

- When $p \ll f(n)$ almost all $G \in \mathcal{G}(n, p)$ do not have P
- When $p \gg f(n)$ almost all $G \in \mathcal{G}(n, p)$ do have P

- It is not clear if any property has threshold function

The $\omega(G) \geq 4$ Property

- Pick $G \in \mathcal{G}(n,p)$ at random
- $S \in \binom{V}{4}$, X_S indicates if S is a clique
- $X = \sum_S X_S$ is the number of 4-clique
- $\omega(G) \geq 4$ iff $X > 0$

Natural line of thought:

$$\mathsf{E}[X] = \sum_S \mathsf{E}[X_S] = \binom{n}{4} p^6 \approx \frac{n^4 p^6}{24}$$

- When $p = o\left(n^{-2/3}\right)$, we have $\mathsf{E}[X] = o(1)$; thus,

$$\text{Prob}[X > 0] \leq \mathsf{E}[X] = o(1)$$

The $\omega(G) \geq 4$ Property

More precisely

$$p = o\left(n^{-2/3}\right) \Longrightarrow \lim_{n\to\infty} \text{Prob}[X > 0] = 0$$

In English

When $p = o\left(n^{-2/3}\right)$ and n sufficiently large, almost all graphs from $\mathcal{G}(n,p)$ do not have $\omega(G) \geq 4$

- What about when $p = \omega\left(n^{-2/3}\right)$?
- We know $\lim_{n\to\infty} \text{E}[X] = \infty$
- But it's not necessarily the case that $\text{Prob}[X > 0] \to 1$
- Equivalently, it's not necessarily the case that $\text{Prob}[X = 0] \to 0$
- Need more information about X

Here Comes Chebyshev

Let $\mu = \mathsf{E}[X]$, $\sigma^2 = \mathsf{Var}[X]$

$$\begin{aligned}
\mathsf{Prob}[X=0] &= \mathsf{Prob}[X-\mu = -\mu] \\
&\leq \mathsf{Prob}\left[\{X-\mu \leq -\mu\} \cup \{X-\mu \geq \mu\}\right] \\
&= \mathsf{Prob}\left[|X-\mu| \geq \mu\right] \\
&\leq \frac{\sigma^2}{\mu^2}
\end{aligned}$$

Thus, if $\sigma^2 = o\left(\mu^2\right)$ then $\mathsf{Prob}[X=0] \to 0$ as desired!

Lemma

For any random variable X

$$\mathsf{Prob}[X=0] \leq \frac{\mathsf{Var}[X]}{(\mathsf{E}[X])^2}$$

PTCF: Bounding the Variance

Suppose $X = \sum_{i=1}^n X_i$

$$\text{Var}\left[X\right] = \sum_{i=1}^n \text{Var}\left[X_i\right] + \sum_{i \neq j} \text{Cov}\left[X_i, X_j\right]$$

If X_i is an indicator for event A_i and $\text{Prob}[X_i = 1] = p_i$, then

$$\text{Var}\left[X_i\right] = p_i(1 - p_i) \leq p_i = \text{E}[X_i]$$

If A_i and A_j are independent, then

$$\text{Cov}\left[X_i, X_j\right] = \text{E}[X_i X_j] - \text{E}[X_i]\text{E}[X_j] = 0$$

If A_i and A_j are not independent (denoted by $i \sim j$)

$$\text{Cov}\left[X_i, X_j\right] \leq \text{E}[X_i X_j] = \text{Prob}[A_i \cap A_j]$$

PTCF: Bounding the Variance

Theorem

Suppose

$$X = \sum_{i=1}^{n} X_i$$

where X_i is an indicator for event A_i. Then,

$$\mathsf{Var}[X] \leq \mathsf{E}[X] + \sum_i \mathsf{Prob}[A_i] \underbrace{\sum_{j:j\sim i} \mathsf{Prob}[A_j \mid A_i]}_{\Delta_i}$$

Corollary

If $\Delta_i \leq \Delta$ for all i, then

$$\mathsf{Var}[X] \leq \mathsf{E}[X](1+\Delta)$$

Back to the $\omega(G) \geq 4$ Property

$$\begin{aligned}
\Delta_S &= \sum_{T \sim S} \text{Prob}[A_T \mid A_S] \\
&= \sum_{|T \cap S| = 2} \text{Prob}[A_T \mid A_S] + \sum_{|T \cap S| = 3} \text{Prob}[A_T \mid A_S] \\
&= \binom{n-4}{2}\binom{4}{2}p^5 + \binom{n-4}{1}\binom{4}{3}p^3 \;\; = \Delta
\end{aligned}$$

So,

$$\sigma^2 \leq \mu(1+\Delta)$$

- Recall: we wanted $\sigma^2/\mu^2 = o(1)$ – OK as long as $\Delta = o(\mu)$
- Yes! When $p = \omega\left(n^{-2/3}\right)$, certainly

$$\Delta = \binom{n-4}{2}\binom{4}{2}p^5 + \binom{n-4}{1}\binom{4}{3}p^3 = o\left(n^4 p^6\right)$$

The $\omega(G) \geq 4$ Property: Conclusion

Theorem

$f(n) = n^{-2/3}$ is a threshold function for the $\omega(G) \geq 4$ property

With essentially the same proof, we can show the following.

Let H be a graph with v vertices and e edges. Define the *density* $\rho(H) = e/v$. Call H *balanced* if every subgraph H' has $\rho(H') \leq \rho(H)$

Theorem

The property "$G \in \mathcal{G}(n,p)$ contains a copy of H" has threshold function $f(n) = n^{-v/e}$.

What Happens when $p \approx$ Threshold?

Theorem

Suppose $p = cp^{-2/3}$, *then* X *is approximately Poisson*$(c^6/24)$
In particular, $\text{Prob}[X = 0] \to 1 - e^{-c^6/24}$

Brief Summary

Let X be a non-negative integral random variable, $\mu = \mathsf{E}[X]$

- Since
$$\mathsf{Prob}[X > 0] \leq \mu,$$
if $\mu = o(1)$ then $X = 0$ almost always!
- If $\mu \to \infty$, then it does not not necessarily follow that $X > 0$ almost always.
- Chebyshev gives
$$\mathsf{Prob}[X = 0] \leq \frac{\sigma^2}{\mu^2}$$
So, if $\sigma^2 = o(\mu^2)$ then $X > 0$ almost always.
- Thus, need to bound the variance.