CPSC 536N: Randomized Algorithms 2014-15 Term 2

Lecture 2

Prof. Nick Harvey *University of British Columbia*

In this lecture we continue our introduction to randomized algorithms by discussing the Max Cut problem. We will achieve a non-trivial result for this problem using only some elementary tools of discrete probability. An important technical point that arises is the difference between algorithms that are good *in expectation* and those that are good *with high probability.*

1 Example: Max Cut

The **Max Cut problem** is a foundational problem in combinatorial optimization and approximation algorithms; many important techniques have been developed during the study of this problem. It is defined as follows.

Let $G = (V, E)$ be an undirected graph. For $U \subseteq V$, let

$$\delta(U) = \{ \, uv \in E \, : \, u \in U \text{ and } v \notin U \, \} .$$

The set $\delta(U)$ is called the **cut** determined by vertex set U. The Max Cut problem is to solve

$$\max \{ \, |\delta(U)| \, : \, U \subseteq V \, \} .$$

This is NP-hard (and in fact was one of the original problems shown to be NP-hard by Karp in his famous 1972 paper), so we cannot hope to solve it exactly. Instead, we will be content to find a cut that is sufficiently large.

More precisely, let OPT denote the size of the maximum cut. We want an algorithm for which there exists a factor $\alpha > 0$ (independent of G) such that the set U output by the algorithm is guaranteed to have $|\delta(U)| \geq \alpha \mathrm{OPT}$. If the algorithm is randomized, we want this guarantee to hold with some probability close to 1.

Here is a brief summary of what is known about this problem.

- Folklore: there is an algorithm with $\alpha = 1/2$. In fact, there are several such algorithms.
- Goemans and Williamson 1995: there is an algorithm with $\alpha = 0.878...$.
- Håstad 2001: no efficient algorithm has $\alpha > 16/17$, unless $P = NP$.
- Khot, Kindler, Mossel, O'Donnel and Oleszkiewicz 2004-2005: no efficient algorithm has $\alpha > 0.878...$, assuming the Unique Games Conjecture. Khot won the Nevanlinna Prize in 2014 partially for this result.

We will give a randomized algorithm achieving $\alpha = 1/2$. In fact, this algorithm appears in an old paper of Erdos [1]. The algorithm couldn't possibly be any simpler: it simply lets U be a uniformly random subset of V. One can check that this is equivalent to independently adding each vertex to U with probability $1/2$. (This relates to the first question in assignment 0.) Note that the algorithm does not even look at the edges of G! Philosophically, this algorithm as trying to "find hay in a haystack", where the cuts are the haystack and the near-maximum cuts are the hay.

The following claim analyzes this algorithm.

Claim 1 *Let U be the set chosen by the algorithm. Then* $\mathrm{E}[\,|\delta(U)|\,] \geq \mathrm{OPT}/2$.

PROOF: For every edge $uv \in E$, let X_{uv} be the indicator random variable which is 1 if $uv \in \delta(U)$. Then

$$\begin{aligned}
\mathrm{E}[\,|\delta(U)|\,] &= \mathrm{E}\left[\sum_{uv \in E} X_{uv}\right] \\
&= \sum_{uv \in E} \mathrm{E}[\,X_{uv}\,] \quad \text{(linearity of expectation)} \\
&= \sum_{uv \in E} \Pr[\,X_{uv} = 1\,]
\end{aligned}$$

Now we note that

$$\begin{aligned}
\Pr[\,X_{uv} = 1\,] &= \Pr[\ (u \in U \wedge v \notin U) \ \vee \ (u \notin U \wedge v \in U)\ ] \\
&= \Pr[\,u \in U \wedge v \notin U\,] \ + \ \Pr[\,u \notin U \wedge v \in U\,] \quad \text{(these are disjoint events)} \\
&= \Pr[\,u \in U\,] \cdot \Pr[\,v \notin U\,] \ + \ \Pr[\,u \notin U\,] \cdot \Pr[\,v \in U\,] \quad \text{(independence)} \\
&= 1/2.
\end{aligned}$$

Thus $\mathrm{E}[\,|\delta(U)|\,] = |E|/2 \geq \mathrm{OPT}/2$, since clearly $\mathrm{OPT} \leq |E|$. □

So we have shown that the algorithm outputs a cut whose *expected size* is large. We might instead prefer a different sort of guarantee: perhaps we'd like to know that, *with high probability*, the algorithm outputs a cut that is large. Mathematically, we have shown that $\mathrm{E}[\,|\delta(U)|\,] \geq \mathrm{OPT}/2$, but perhaps we want to know that $\Pr[\,|\delta(U)| \geq \mathrm{OPT}/2\,]$ is large. To connect these two statements, we need to show that $|\delta(U)|$ is likely to be close to its expectation.

2 Concentration Inequalities

One of the most important tasks in analyzing randomized algorithms is to understand what random variables arise and how well they are **concentrated**. A variable with good concentration is one that is close to its mean with good probability. A "concentration inequality" is a theorem proving that a random variable has good concentration. Such theorems are also known as "tail bounds".

2.1 Markov's inequality

This is the simplest concentration inequality. The downside is that it only gives very weak bounds, but the upside is that needs almost no assumptions about the random variable. It is often useful in scenarios where not much concentration is needed, or where the random variable is too complicated to be analyzed by more powerful inequalities.

Theorem 2 *Let Y be a real-valued random variable that assumes only nonnegative values*[1]*. Then, for all $a > 0$,*

$$\Pr[\,Y \geq a\,] \ \leq \ \frac{\mathrm{E}[\,Y\,]}{a}.$$

[1] Pedantic detail: we should also assume that $\mathrm{E}[\,Y\,]$ is finite.

References: Mitzenmacher-Upfal Theorem 3.1, Motwani-Raghavan Theorem 3.2, Wikipedia, Grimmett-Stirzaker Lemma 7.2.7, Durrett Theorem 1.6.4.

Note that if $a \leq \mathrm{E}[\,Y\,]$ then the right-hand side of the inequality is at least 1, and so the statement is trivial. So Markov's inequality is only useful if $a > \mathrm{E}[\,Y\,]$. Typically we use Markov's inequality to prove that Y has only a constant probability of exceeding its mean by a constant factor, e.g., $\Pr[\,Y \geq 2 \cdot \mathrm{E}[\,Y\,]\,] \leq 1/2$.

Proof: Let X be the indicator random variable that is 1 if $Y \geq a$. Since Y is non-negative, we have

$$X \;\leq\; Y/a. \tag{1}$$

Then

$$\Pr[\,Y \geq a\,] \;=\; \Pr[\,X \geq 1\,] \;=\; \mathrm{E}[\,X\,] \;\leq\; \mathrm{E}[\,Y/a\,] \;=\; \frac{\mathrm{E}[\,Y\,]}{a},$$

where the inequality comes from taking the expectation of both sides of (1). □

Note that Markov's inequality only bounds the **right tail** of Y, i.e., the probability that Y is much greater than its mean.

2.2 The Reverse Markov inequality

In some scenarios, we would also like to bound the probability that Y is much smaller than its mean. Markov's inequality can be used for this purpose if we know an upper-bound on Y. The following result is an immediate corollary of Theorem 2.

Corollary 3 *Let Y be a random variable that is never larger than B. Then, for all $a < B$,*

$$\Pr[\,Y \leq a\,] \;\leq\; \frac{\mathrm{E}[\,B - Y\,]}{B - a}.$$

References: Grimmett-Stirzaker Theorem 7.3.5.

2.3 Application to Max Cut

Let's now analyze the probability that our randomized algorithm for Max Cut gives a large cut. Let $Y = |\delta(U)|$ and $B = |E|$, and note that Y is never larger than B. Fix any $\epsilon \in [0, 1/2]$ and set $a = (\frac{1}{2} - \epsilon)|E|$. By the Reverse Markov inequality,

$$\begin{aligned}
\Pr[\,|\delta(U)| \leq (1/2-\epsilon)|E|\,] \;&\leq\; \frac{\mathrm{E}[\,|E| - |\delta(U)|\,]}{|E| - (1/2-\epsilon)|E|} \\
&=\; \frac{|E| - \mathrm{E}[\,|\delta(U)|\,]}{(1/2+\epsilon)|E|} \qquad \text{(linearity of expectation)} \\
&=\; \frac{1}{1+2\epsilon} \qquad \text{since } \mathrm{E}[\,|\delta(U)|\,] = |E|/2 \\
&\leq\; 1-\epsilon,
\end{aligned}$$

where we have used Inequality 2 from the Notes on Convexity Inequalities.

This shows that, with probability at least ϵ, the algorithm outputs a set U satisfying

$$|\delta(U)| \;>\; (1/2 - \epsilon)\,\mathrm{OPT}.$$

This statement is quite unsatisfying. If we want a 0.499 approximation, then we have only shown that the algorithm has probability 0.001 of succeeding. Next we will show how to increase the probability of success.

3 Amplification by Independent Trials

In many cases where the probability of success is positive but small, we can "amplify" that probability by perfoming several independent trials and taking the best outcome. We already used this technique in the first lecture to improve the success probability of our equality test.

For Max Cut, consider the following algorithm. First it picks several sets $U_1, \ldots, U_k \subseteq V$, independently and uniformly at random. Let j be the index for which $|\delta(U_j)|$ is largest. The algorithm simply outputs the set U_j. (Our initial Max Cut algorithm above did not even look at the edges of the graph, but this new algorithm must look at the edges to compute $|\delta(U_j)|$.)

To analyze this algorithm, we wish to argue that $|\delta(U_j)|$ is large. Well, $|\delta(U_j)|$ is small only if *all* $|\delta(U_i)|$ are small, and this is rather unlikely.

$$\begin{aligned}
\Pr\left[\, |\delta(U_i)| \leq (1/2-\epsilon)|E| \;\forall i = 1,\ldots,k \,\right] \;&=\; \prod_{i=1}^{k} \Pr\left[\, |\delta(U_i)| \leq (1/2-\epsilon)|E| \,\right] \qquad \text{(by independence)}\\
&\leq\; (1-\epsilon)^k \qquad \text{(by our analysis above)}\\
&\leq\; e^{-\epsilon k}.
\end{aligned}$$

Here we have used the standard trick $1 + x \leq e^x$, which is Inequality 1 from the Notes on Convexity Inequalities.

Thus, setting $k = \log(1/\delta)/\epsilon$, we obtain that

$$\Pr\left[\,\text{every } |\delta(U_i)| \leq (1/2-\epsilon)|E|\,\right] \;\leq\; \delta.$$

In summary, our new Max Cut algorithm picks $k = \log(1/\delta)/\epsilon$ sets $U_1, \ldots, U_k$, finds the j which maximizes $|\delta(U_j)|$, then outputs U_j. With probability at least $1 - \delta$ we have

$$\Pr\left[\, |\delta(U_j)| \geq (1/2-\epsilon)|E| \,\right].$$

In this analysis we used the following general fact, which is elementary but often very useful.

Claim 4 *Consider a random trial in which the probability of success is p. If we perform k mutually independent trials, then*

$$\Pr\left[\,\textit{all failures}\,\right] \;=\; (1-p)^k \;\leq\; e^{-pk}$$

and therefore

$$\Pr\left[\,\textit{at least one success}\,\right] \;=\; 1-(1-p)^k \;\geq\; 1-e^{-pk}.$$

3.1 Example: Flipping Coins

Consider flipping a fair coin. If the coin turns up heads, call this a "success". So the probability of success is $p = 1/2$. By Claim 4,

$$\Pr\left[\,\text{at least one heads after } k \text{ trials}\,\right] \;=\; 1 - 1/2^k.$$

This is a true statement, but not all that interesting: it is quite obvious that after k coin flips, the probability of seeing at least one head is very close to 1. Note that the expected number of heads is $k/2$. Can't we say that there are probably close to $k/2$ heads? This takes us beyond the subject of amplification and into a discussion of the binomial distribution.

4 The binomial distribution

Again, let us flip a fair coin k times and let X be the number of heads seen. Then X has the **binomial distribution**. So

$$\Pr[\,\text{exactly } i \text{ heads}\,] \;=\; \binom{k}{i} 2^{-k}.$$

How can we show that X is probably close to its expected value (which is $k/2$). Well, the probability of X being "small" is:

$$\Pr[\,\text{at most } i \text{ heads}\,] \;=\; \sum_{0 \leq j \leq i} \binom{k}{j} 2^{-k}. \tag{2}$$

Here the meaning of "small" depends on the choice of i. For what values of i is this sum small? Unfortunately the sum is a bit too complicated to get a feel for its magnitude. Can we simplify the expression so it's easier to see what's going on?

As far as I know, this sum does *not* have a simple closed-form expression. Instead, can we usefully approximate that sum? It would be great to have a provable upper bound for the sum that is simple, useful and asymptotically nearly tight. The **Chernoff bound**, which we discuss in Section 5, is a very general and powerful tool for analyzing tails of probability distributions, and it gives a fantastic bound on (2). But first, to motivate the bound that we will obtain, let us describe some interesting behavior of binomial distributions.

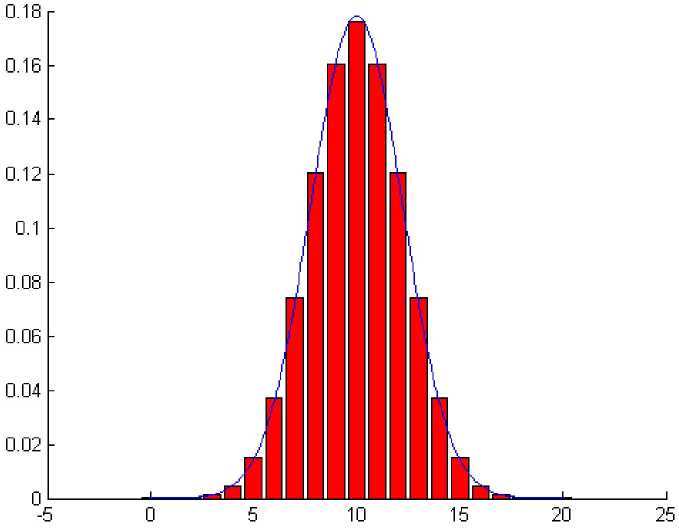
4.1 Different behavior in different regimes

In most introductory courses on probability theory, one often encounters statements that describe the limiting behavior of the binomial distribution. For example:

Fact 5 *Let X be a binomially distributed random variable with parameters n and p. (That is, X gives the number of successes in n independent trials where each trial succeeds with probability p.) As n gets large while p remains fixed, the distribution of X is "well approximated" by the normal distribution with mean np and variance $np(1-p)$.*

References: Durrett Theorem 3.4.1 and Example 3.4.3.

The following plot shows the binomial distribution for $n = 20$ and $p = 1/2$ in red and the corresponding normal approximation in blue.

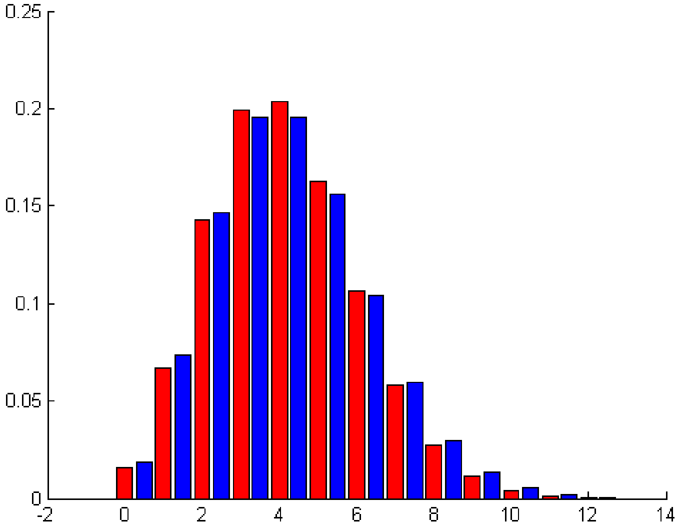


However the binomial distribution has qualitatively different behavior when p is very small.

Fact 6 *Let X be a binomially distributed random variable with parameters n and p. As n gets large while np remains fixed, the distribution of X is "well approximated" by the Poisson distribution with parameter $\lambda = np$.*

References: Mitzenmacher-Upfal Theorem 5.5, Durrett Theorem 3.6.1.

The following plot shows the binomial distribution for $n = 50$ and $p = 4/n$ in red and the corresponding Poisson approximation in blue. Note that the red plot is quite asymmetric, so we would not expect a normal distribution (which is symmetric) to approximate it well.



Ideally would like an upper bound on (2) which works well for all ranges of parameters, and captures the phenomena described in Fact 5 and Fact 6. Remarkably, the Chernoff bound is able to capture both of these phenomena.

5 The Chernoff Bound

The Chernoff bound is used to bound the tails of the distribution for a sum of independent random variables, under a few mild assumptions. Since binomial random variables are sums of independent Bernoulli random variables, it can be used to bound (2). Not only is the Chernoff bound itself very useful, but its proof techniques are very general and can be applied in a wide variety of settings.

The Chernoff bound is by far the most useful tool in randomized algorithms. Numerous papers on randomized algorithms have only three ingredients: Chernoff bounds, union bounds and cleverness. Of course, there is an art in being clever and finding the right way to assemble these ingredients!

5.1 Formal Statement

Let $X_1, \ldots, X_n$ be independent random variables such that X_i always lies in the interval $[0,1]$. Define $X = \sum_{i=1}^n X_i$ and $p_i = \mathrm{E}[\,X_i\,]$. Let $\mu_{\min} \leq \sum_i \mathrm{E}[\,X_i\,] \leq \mu_{\max}$.

Theorem 7 *For all $\delta > 0$,*

$$\begin{aligned}
\Pr[\,X \geq (1+\delta)\mu_{\max}\,] &\overset{(a)}{\leq} \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu_{\max}} \overset{(b)}{\leq} \begin{cases} e^{-\delta^2\mu_{\max}/3} & (\text{if } \delta \leq 1) \\ e^{-(1+\delta)\ln(1+\delta)\mu_{\max}/4} & (\text{if } \delta \geq 1) \\ e^{-\delta\mu_{\max}/3} & (\text{if } \delta \geq 1) \end{cases} \\
\Pr[\,X \leq (1-\delta)\mu_{\min}\,] &\overset{(c)}{\leq} \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_{\min}} \overset{(d)}{\leq} e^{-\delta^2\mu_{\min}/2}.
\end{aligned}$$

Inequalities (c) and (d) are only valid for $\delta < 1$, but $\Pr[\,X \leq (1-\delta)\mu_{\min}\,] = 0$ *if* $\delta > 1$.

Remarks. All of these bounds are useful in different scenarios. In inequality (b), the cases $\delta \leq 1$ and $\delta > 1$ are different due to the different phenomena illustrated in Facts 5 and 6.

Historical remark. Chernoff actually only considered the case in which the X_i random variables are identically distributed. The more general form stated here is due to Hoeffding.

References

[1] P. Erdős. Gráfok páros körüljárású részgráfjairól (On bipartite subgraphs of graphs, in Hungarian). *Mat. Lapok*, 18:283–288, 1967.

1