

## Lecture 3

Prof. Nick Harvey

University of British Columbia

## 1 Example: Balls and Bins

Last time we proved the Chernoff bound, the most useful analysis tool we will encounter in this course. Today we illustrate the power of the Chernoff bound by using it to analyze a fundamental balls-and-bins problem. But first of all we introduce the other (almost trivial) tool which is often used in conjunction with the Chernoff bound.

**Lemma 1 (The Union Bound)** *Let  $\mathcal{E}_1, \dots, \mathcal{E}_k$  be any collection of events. Then*

$$\Pr[\mathcal{E}_1 \vee \dots \vee \mathcal{E}_k] \leq \sum_{i=1}^k \Pr[\mathcal{E}_i].$$

PROOF: Mitzenmacher and Upfal, Lemma 1.2.  $\square$

Many interesting problems can be modeled using simple problems involving balls and bins. Today we are interested in the bin which has the most balls. Suppose there are  $n$  bins. We repeat the following experiment  $n$  times: throw a ball into a uniformly chosen bin. (The experiments are mutually independent.) Let  $B_i$  be the number of balls in bin  $i$ . What is  $\max_i B_i$ ?

**Theorem 2** *Let  $\alpha = e \ln n / \ln \ln n$ . Then  $\max_i B_i \leq \alpha$  with probability tending to 1 as  $n \rightarrow \infty$ .*

This theorem is optimal up to constant factors. It is known that  $\max_i B_i \geq \ln n / \ln \ln n$  with probability at least  $1 - 1/n$ . (See, e.g., Lemma 5.12 in Mitzenmacher-Upfal.)

PROOF: Let us focus on the first bin. Let  $X_1, \dots, X_n$  be indicator random variables where  $X_j$  is 1 if the  $j$ th ball lands in the first bin. Obviously  $E[B_1] = \sum_j E[X_j] = 1$ . What is the probability that this bin has more than  $\alpha$  balls?

We will analyze this using the Chernoff bound. This is possible since  $B_1$  is a sum of independent random variables, each of which take values in  $[0, 1]$ . Recall that the Chernoff bound says

$$\Pr[X \geq (1 + \delta)E[X]] \leq \exp\left(-E[X] \cdot ((1 + \delta) \ln(1 + \delta) - \delta)\right) \quad \forall \delta > 0.$$

We apply this to  $B_1$ , setting  $\delta = \alpha - 1$  and using  $E[B_1] = 1$ . We obtain

$$\Pr[B_1 \geq \alpha] \leq \exp(-\alpha \ln \alpha + \alpha - 1) \leq \exp(-\alpha(\ln \alpha - 1)).$$

Using our usual techniques from [Notes on Convexity Inequalities](#), one can show that

$$\ln \alpha = 1 + \ln \ln n - \ln \ln \ln n \geq 1 + (\ln \ln n)/2 \quad \forall n \geq 1.$$

Plugging that in,

$$\begin{aligned}\Pr[B_1 \geq \alpha] &\leq \exp(-\alpha(\ln \alpha - 1)) \\ &\leq \exp(-\alpha(\ln \ln n)/2) \\ &\leq \exp(-(e/2) \ln n) \\ &\leq n^{-1.35}.\end{aligned}$$

So, let  $\mathcal{E}_i$  be the event that  $B_i \geq \alpha$ . By Lemma 1,

$$\begin{aligned}\Pr[\text{for any bin } i \text{ we have } B_i \geq \alpha] &\leq \sum_i \Pr[B_i \geq \alpha] \\ &\leq \sum_i n^{-1.35} \\ &= n^{-0.35}\end{aligned}$$

Thus, with probability at least  $1 - n^{-0.35}$ , all bins have less than  $\alpha$  balls.  $\square$

## 2 Congestion Minimization

One of the classically important areas in algorithm design and combinatorial optimization is **network flows**. A central problem in that area is the maximum flow problem. We now look at a generalization of this problem.

An instance of the problem consists of a directed graph  $G = (V, A)$  and a sequence  $(s_1, t_1), \dots, (s_k, t_k)$  of pairs of vertices. Let  $n = |V|$ . (It is not crucial that the graph be directed; the problem is equally interesting in undirected graphs. However in network flow problems it is often more convenient to look at directed graphs. Feel free to think about whatever variant you find easier.)

A natural question to ask is: do there exist paths  $P_i$  from  $s_i$  to  $t_i$  for every  $i$  such that these paths share no arcs? This is called the **edge-disjoint paths** problem. Quite remarkably, it is NP-hard even in the case  $k = 2$ , assuming the graph is directed. For undirected graphs, it is polynomial time solvable if  $k$  is a fixed constant, but NP-hard if  $k$  is a function of  $n$ .

We will look at a variant of this problem called the **congestion minimization** problem. The idea is to allow each arc to be used in multiple paths, but not too many. The number of paths using a given arc is the “congestion” of that arc. We say that a solution has congestion  $C$  if it is a collection of paths  $P_i$  from  $s_i$  to  $t_i$ , where each arc is contained in at most  $C$  of the paths. The problem is to find the minimum value of  $C$  such that there is a solution of congestion  $C$ . This problem is still NP-hard, since determining if  $C = 1$  is the edge-disjoint paths problem.

We will look at the congestion minimization problem from the point of view of approximation algorithms. Let  $OPT$  be the minimum congestion of any solution. We would like to give an algorithm which can produce a solution with congestion at most  $\alpha \cdot OPT$  for some  $\alpha \geq 1$ . This factor  $\alpha$  is called the approximation factor of the algorithm.

**Theorem 3** *There is an algorithm for the congestion minimization problem with approximation factor  $O(\log n / \log \log n)$ .*

To design such an algorithm we will use linear programming. We write down an integer program (IP) which captures the problem exactly, relax that to a linear program (LP), then design a method for “rounding” solutions of the LP into solutions for the IP.

**The Integer Program.** Writing an IP formulation of an optimization problem is usually quite simple. That is indeed true for the congestion minimization problem. However, we will use an IP which you might find rather odd: our IP will have exponentially many variables. This will simplify our explanation of the rounding.

Let  $\mathcal{P}_i$  be the set of *all* paths in  $G$  from  $s_i$  to  $t_i$ . (Note that  $|\mathcal{P}_i|$  may be exponential in  $n$ .) For every path  $P \in \mathcal{P}_i$ , we create a variable  $x_P^i$ . This variable will take values only in  $\{0, 1\}$ , and setting it to 1 corresponds to including the path  $P$  in our solution.

The integer program is as follows

$$\begin{aligned} \min \quad & C \\ \text{s.t.} \quad & \sum_{P \in \mathcal{P}_i} x_P^i = 1 \quad \forall i = 1, \dots, k \\ & \sum_i \sum_{P \in \mathcal{P}_i \text{ with } a \in P} x_P^i \leq C \quad \forall a \in A \\ & x_P^i \in \{0, 1\} \quad \forall i = 1, \dots, k \text{ and } P \in \mathcal{P}_i \end{aligned}$$

The last constraint says that we must decide for every path whether or not to include it in the solution. The first constraint says that the solution must choose exactly one path between each pair  $s_i$  and  $t_i$ . The second constraint ensures that the number of paths using each arc is at most  $C$ . The optimization objective is to find the minimum value of  $C$  such that a solution exists.

Every solution to the IP corresponds to a solution for the congestion minimization problem with congestion  $C$ , and vice-versa. Thus the optimum value of the IP is  $OPT$ , which we previously defined to be the minimum congestion of any solution to the original problem.

**The Linear Program.** The integer program is NP-hard to solve, so we “relax” it into a linear program. This amounts to replacing the integrality constraints with non-negativity constraints. The resulting linear program is:

$$\begin{aligned} \min \quad & C \\ \text{s.t.} \quad & \sum_{P \in \mathcal{P}_i} x_P^i = 1 \quad \forall i = 1, \dots, k \\ & \sum_i \sum_{P \in \mathcal{P}_i \text{ with } a \in P} x_P^i \leq C \quad \forall a \in A \\ & x_P^i \geq 0 \quad \forall i = 1, \dots, k \text{ and } P \in \mathcal{P}_i \end{aligned}$$

This LP can be solved in time polynomial in the size of  $G$ , even though its number of variables is exponential in the size of  $G$ . This can be done either by the ellipsoid method or by finding a “compact formulation” of the LP which uses fewer variables (much like the usual LP that you have probably seen for the ordinary maximum flow problem).

So, without going into details, our algorithm will solve this LP and obtain a solution where the number of non-zero  $x_P^i$  variables is only polynomial in the size of  $G$ . Let  $C^*$  be the optimum value of the LP.

**Claim 4**  $C^* \leq OPT$ .

PROOF: The LP was obtained from the IP by *removing* constraints. Therefore any feasible solution for the IP is also feasible for the LP. In particular, the optimal solution for the IP is feasible for the LP. So the LP has a solution with objective value equal to  $OPT$ .  $\square$

**The Rounding.** Our algorithm will solve the LP and most likely obtain a “fractional” solution — a solution with some non-integral variables, which is therefore not feasible for the IP. The next step of

the algorithm is to “round” that fractional solution into a solution which is feasible for the IP. In doing so, the congestion might increase, but we will ensure that it does not increase too much.

The technique we will use is called **randomized rounding**. For each  $i = 1, \dots, k$ , we randomly choose *exactly one path*  $P_i$  by setting  $P_i = P$  with probability  $x_P^i$ . (The LP ensures that these are indeed probabilities: they are non-negative and sum up to 1.) The algorithm outputs the chosen paths  $P_1, \dots, P_k$ .

**Analysis.** All that remains is to analyze the congestion of these paths. Let  $Y_i^a$  be the indicator random variable that is 1 if  $a \in P_i$  and 0 otherwise. Let  $Y^a = \sum_i Y_i^a$  be the congestion on arc  $a$ . The expected value of  $Y^a$  is easy to analyze:

$$\mathbb{E}[Y^a] = \sum_i \mathbb{E}[Y_i^a] = \sum_i \sum_{P \in \mathcal{P}_i \text{ with } a \in P} x_P^i \leq C^*,$$

where the inequality comes from the LP’s second constraint. (Recall we assume that the fractional solution is optimal for the LP, and therefore  $C = C^*$ .)

The Chernoff bound says, if  $X$  is a sum of independent random variables each of which take values in  $[0, 1]$ , and  $\mu$  is an *upper bound* on  $\mathbb{E}[X]$ , then

$$\Pr[X \geq (1 + \delta)\mu] \leq \exp\left(-\mu \cdot ((1 + \delta) \ln(1 + \delta) - \delta)\right) \quad \forall \delta > 0.$$

We apply this to  $Y^a$ , taking  $\mu = C^*$  and  $\alpha = 1 + \delta = 6 \log n / \log \log n$ . Following the argument in Section 1,

$$\begin{aligned} \Pr[Y^a \geq \alpha C^*] &\leq \exp\left(-C^*(\alpha \ln \alpha + \alpha - 1)\right) \\ &\leq \exp\left(-\alpha \ln \alpha + \alpha - 1\right) \\ &\leq \exp\left(-\frac{6}{2} \ln n\right) = 1/n^3. \end{aligned}$$

We now use a union bound to analyze the probability of *any* arc having congestion greater than  $\alpha C^*$ .

$$\Pr[\text{any } a \text{ has } Y^a \geq \alpha C^*] \leq \sum_{a \in A} \Pr[Y^a \geq \alpha C^*] \leq \sum_{a \in A} 1/n^3 \leq 1/n,$$

since the graph has at most  $n^2$  arcs. So, with probability at least  $1 - 1/n$ , the algorithm produces a solution for which every arc has congestion at most  $\alpha C^*$ , which is at most  $\alpha \cdot OPT$  by Claim 4. So our algorithm has approximation factor  $\alpha = O(\log n / \log \log n)$ .

**Further Remarks.** The *rounding* algorithm that we presented is actually optimal: there are graphs for which  $OPT/C^* = \Omega(\log n / \log \log n)$ . Consequently, every rounding algorithm which converts a fractional solution of LP to an integral solution of IP must necessarily incur an increase of  $\Omega(\log n / \log \log n)$  in the congestion.

That statement does not rule out the possibility that there is a better algorithm which behaves completely differently (i.e., one which does not use IP or LP at all). But sadly it turns out that there is no better algorithm (for the case of directed graphs). It is known that every efficient algorithm must have approximation factor  $\alpha = \Omega(\log n / \log \log n)$ , assuming a reasonable complexity theoretic conjecture ( $NP \not\subseteq BPTIME(n^{O(\log \log n)})$ ). So the algorithm that we presented is optimal, up to constant factors.

### 3 The Negative Binomial Distribution

The **negative binomial distribution** is perhaps the probability distribution with the worst public relations. It shows up in many different randomized algorithms, but it is not taught or covered in textbooks as much as it should be.

There are a few ways to define this distribution. We adopt the following definition. There are two parameters,  $p \in [0, 1]$  and  $k \in \mathbb{N}$ . Suppose we perform a sequence of independent Bernoulli trials, each succeeding with probability  $p$ . Let  $Y$  be the number of trials performed until we see the  $k$ th success. Then  $Y$  is said to have the negative binomial distribution.

Note that this is quite different from the usual binomial distribution. For example, if  $X$  is a binomial random variable with parameters  $n$  and  $p$  then the value of  $X$  is always at most  $n$ . In contrast,  $Y$  has positive probability of taking any integer value larger than or equal to  $k$ . Nevertheless, there is a relationship between the tails of  $X$  and  $Y$ . The following claim is quite useful, although often not stated explicitly in the literature.

**Claim 5** *Let  $Y$  be a random variable distributed according to the negative binomial distribution with parameters  $k$  and  $p$ . Let  $X$  be a random variable distributed according to the binomial distribution with parameters  $n$  and  $p$ . Then  $\Pr[Y > n] = \Pr[X < k]$ .*

Informally, this is quite easy to see. The event  $\{Y > n\}$  is the event that, after performing  $n$  trials, we still have not seen  $k$  successes. And that is also the event  $\{X < k\}$ . That argument is not completely formal because the sample spaces of  $X$  and  $Y$  are not the same. The following argument explains the connection in more detail.

PROOF: Let  $q = 1 - p$ . We need the following facts about the sample space of the negative binomial distribution.

- **Probability of an elementary event.** The sample space underlying the negative binomial distribution consists of all finite sequences of successes and failures with exactly  $k$  successes and any number  $f \geq 0$  of failures, where the last outcome must be a success. For any such sequence, its probability in the sample space is  $p^k q^f$ . One can check that these probabilities sum up to 1 using properties of binomial coefficients generalized to negative numbers.
- **Probability of seeing a prefix.** For any sequence  $\sigma$  with  $i < k$  successes and any number  $f \geq 0$  of failures, the probability that  $\sigma$  gives the outcomes of the *first  $i + f$  trials* is  $p^i q^f$ . The proof of this is very similar to the proof of the previous property.

The event  $\{Y > n\}$  consists of all sequences with  $k$  successes and  $f \geq n - k + 1$  failures, where the last outcome is a success. To compute the total probability of this event, we can partition these sequences into groups where the members of each group all have the same prefix of length  $n$  (i.e., the same outcomes in the first  $n$  trials). For any group with  $i < k$  successes in the first  $n$  trials, the total probability of that group is  $p^i q^{n-i}$ , by the second property given above. Since there are  $\binom{n}{i}$  ways to choose the locations of the  $i$  successes in the  $n$  trials, we have

$$\Pr[Y > n] = \sum_{i=0}^{k-1} \binom{n}{i} p^i q^{n-i} = \Pr[X < k].$$

□

An important consequence of the previous claim is that Chernoff bounds give tail bounds on  $Y$ .

**Claim 6** *Let  $Y$  have the negative binomial distribution with parameters  $k$  and  $p$ . Pick  $\delta \in [0, 1]$  and set  $n = \frac{k}{(1-\delta)p}$ . Then  $\Pr[Y > n] \leq \exp(-\delta^2 k / 3(1-\delta))$ .*

PROOF: Let  $X$  have the binomial distribution with parameters  $n$  and  $p$ . Note that  $E[X] = np = k/(1 - \delta)$ . By Claim 5,

$$\begin{aligned}\Pr[Y > n] &= \Pr[X < k] \\ &\leq \exp(-\delta^2 E[X]/3) \\ &= \exp(-\delta^2 k/3(1 - \delta)),\end{aligned}$$

where the inequality comes from the Chernoff bound.  $\square$