CSE 525 Randomized Algorithms & Probabilistic Analysis **Spring 2013**

Lecture 3: April 9

Lecturer: Anna Karlin *Scribe: Tyler Rigsby & John MacKinnon*

3.1 Kinds of randomization in algorithms

So far in our discussion of randomized and probabilistic algorithms, we have yet to make much of a distinction regarding the quality of randomization in our methods, and how this randomization affects the execution and results of our algorithms. The following are two major classes of randomized and probabilistic algorithms:

3.1.1 Monte Carlo Algorithms

Monte Carlo algorithms have a fixed, deterministic running time, but they have a small probability of producing an incorrect result. Thus, they have the property that running a Monte Carlo algorithm multiple times will decrease the probability of outputting an incorrect result.

An example of a Monte Carlo algorithm is the randomized min-cut algorithm from the previous lecture; it always runs in time $\tilde{O}(n^2)$, but may produce an incorrect result with probability $O(\frac{1}{n^k})$, where k is some (constant) number of times the algorithm is executed.

3.1.2 Las Vegas Algorithms

Unlike Monte Carlo algorithms, Las Vegas Algorithms always produce the correct answer. However, this comes at the price of the running time: rather than correctness being a random variable (as is the case in Monte Carlo algorithms), the running time becomes the random variable. This fact draws a unique parallel beteween Last Vegas and Monte Carlo algorithms - that is, we may convert any Las Vegas algorithm to a Monte Carlo algorithm by simply terminating early after a given number of iterations and outputting the current answer. The converse, however, is not true; there is no known way to universally convert Monte Carlo algorithms to Las Vegas algorithms.

An example of a Las Vegas Algorithm algorithm is Quicksort, where the pivot is chosen randomly. Depending upon the pivots chosen, the runtime may range from $O(n^2)$ in the worst case (poor pivot choices), to $O(n * log(n))$ in the best (pivot is always the median). However, no matter how long the algorithm takes, the output of Quicksort is always a sorted list, and is thus correct.

3.2 Inequalities and Bounds

In exploring randomized algorithms, it is useful to consider the tail bounds and concentration inequalities of the random variables we use. While these bounds are generally somewhat loose, they at least provide us with a metric for bounding the 'badness' of any arbitrary random variable, under very mild assumptions. The following subsections explore some of these bounds that will be useful in analyzing our randomized algorithms.

3.2.1 Markov's Inequality

Markov's inequality is the weakest bound we explore, but also requires the fewest assumptions. It's implications are not very tight, but they give us a handle to relate expected values and probabilities for almost any arbitrary random variable.

Assumptions:

1. X is a random variable, that is **non-negative**
2. We know the expected value of X

then, under these minimal assumptions:

Theorem 3.1

$$\forall \alpha > 0 : Pr(X \geq \alpha) \leq \frac{E(X)}{\alpha} \tag{3.1}$$

Another useful way to view the inequality is as such:

Corollary 3.2

$$\forall \alpha > 0 : Pr(X \geq \alpha * E(X)) \leq \frac{1}{\alpha} \tag{3.2}$$

Proof: The proof of Markov's inequality follows directly from the defintion of expectation of a random variable:

$$\begin{aligned}
E(X) &= \sum_x x * Pr(x) \\
&= \sum_{x<\alpha} x * Pr(x) + \sum_{x \geq \alpha} x * Pr(x) && [split\ up\ sum] \\
&\geq 0 + \sum_{x \geq \alpha} x * Pr(x) && [given\ x\ nonnegative] \\
&\geq 0 + \sum_{x \geq \alpha} \alpha * Pr(x) && [x \geq \alpha] \\
&\geq a * Pr(X \geq \alpha) && [defintion\ of\ sum] \\
\frac{E(X)}{\alpha} &\geq Pr(X \geq \alpha) && [arithmetic]
\end{aligned}$$

■

We can see that the Markov inequality is a useful tool for bounding the probability of a random variable versus its expected value, because there is virtually no extra information or assumptions about the variable itself. For example, it can be proven that Quicksort's expected runtime is $\simeq 1.4n * log(n)$. Without knowing anything else about Quicksort, we may plug 2 in for the α value in our inequality, and we very simply get that there is less than a $\frac{1}{2}$ probability that Quicksort takes 2 times longer than expected. However, the bound can also be exceptionally loose. Take for example a fair six-sided die: we know that $Pr(rollingX \geq 3) = \frac{2}{3}$; however, by Markov's inequality, noting that $E(X) = \frac{7}{2}$, we get that this same probility $Pr(X \geq 3) \leq \frac{7}{6}$, or ≤ 1. This is trivially true of *all* probabilities, so the Markov analysis is useless here. For cases such as this, stronger bounds are necessary.

3.2.2 Chebychev's Inequality

If a little more is known about our random variable, then we may make use of Chebychev's inequality, in order to produce a tighter bound. The inequality is based upon knowing the variance of our random variable, and at a high level, asserts that 'nearly all' values from any probability distribution will lie near to the mean.

Assumptions:

1. X is a random variable
2. We know the expected value and variance of X

Then, with these conditions met, Chebychev's inequality is as follows:

Theorem 3.3

$$\forall \alpha > 0 : Pr(|X - E(X)| \geq \alpha) \leq \frac{Var(X)}{\alpha^2} \tag{3.3}$$

Written with standard deviation in mind:

Corollary 3.4

$$let\ \sigma = \sqrt{Var(X)}$$
$$\forall \alpha > 0 : Pr(X \geq \alpha * \sigma) \leq \frac{1}{\sigma^2} \tag{3.4}$$

Proof: The proof of Chebychev's inequality follows as a direct application of Markov's Inequality:

$$\begin{aligned}
let\ X &= (Y - E(Y))^2 \\
Pr(|Y - E(Y)| \geq \alpha) &= Pr(X \geq \alpha^2) && [square\ both\ sides] \\
Pr(|Y - E(Y)| \geq \alpha) &\leq \frac{E(X)}{\alpha^2} && [by\ (3.1)] \\
Pr(|Y - E(Y)| \geq \alpha) &\leq \frac{Var(Y)}{\alpha^2} && [Var(X) = E[(X - E(X))^2]]
\end{aligned}$$

■

Translated into english, the essential takeaway from Chebychev's Inequality is that no more than $\frac{1}{\alpha^2}$ of any distribution's values may lie more than α standard deviations away from the mean. While this is still fairly loose, Chebychev's inequality is somewhat more useful than Markov's because it takes into account the spread of the given random variable. Take for example two normal distributions with equal means and standard deviations that differ by a factor of 2. Markov bounds give us the same probability values for any point in both distributions, whereas Chebychev is more closely tailored to each, giving us probabilities that differ by a factor of 4 for any point. However, bounds derived from Chebychev's inequality aren't always incredibly tight either. For example: if we plug in the value 2 for α, then we get that no more than $\frac{1}{4}$ of a distribution's values may lie more than 2 standard deviations from the mean; in other words, at least 75% of all values must lie within 2 standard deviations of the mean. In the case of the normal distribution, we know by the empirical rule that 95% of all values lie within 2 standard deviations of the mean, so stronger bounds than Chebychev's may be needed in some situations.

3.2.3 Chernoff-Hoeffding Bounds

The Chernoff-Hoeffding Bounds require our strongest assumptions, but they're also our tightest, strongest bounds. The bounds are particularly useful for bounding the tails of sums of independent random variables.

Assumptions:

1. $X_1, X_2, \ldots, X_n$ be independent random variables such that $X_i \in [0, 1]$
2. $X = \sum_{i=1}^{n} X_i$ and $\mu = E(X)$ (Therefore, if $E(X_i) = p_i$, then $\mu = \sum_{i=1}^{n} p_i$)

3.2.3.1 Chernoff-Hoeffding Upper Trail

The strongest version of the upper bound inequality is as follows:

Theorem 3.5

$$\forall \delta > 0 : Pr(X \geq (1+\delta)\mu) \leq e^{-\mu[(1+\delta)\ln(1+\delta)-\delta]}$$

However, there are also some weaker, but still useful versions:

$$\forall \delta \in [0, 1] : Pr(X \geq (1+\delta)\mu) \leq e^{-\frac{\mu\delta^2}{3}} \tag{3.5}$$

$$\forall \delta \geq 1 : Pr(X \geq (1+\delta)\mu) \leq e^{-\frac{\mu\delta}{3}} \tag{3.6}$$

Note that, though the statement is weaker, it suffices to have $\mu \geq E(X)$.

Proof: Note that a key property of independent random variables is:

$$E(YZ) = E(Y)E(Z)$$

However, in our case, we have sums. To solve this, we will convert our sums to products using the exponential function.

Let $Y_i = e^{tX_i}$, where t will be determined later.

Then,

$$Y = e^{tX} = e^{t\sum X_i} = \prod_{i=1}^{n} e^{tX_i} = \prod_{i=1}^{n} Y_i$$

Because the X's were independent random variables and the Y's are derived from the X's, the Y's are also indepndent variables and we may apply the equation above to conclude:

$$E(Y) = \prod_{i=1}^{n} E(Y_i)$$

Now, we may begin to bound the probability that the sum of multiple i.r.v.'s is not above a certain point.

$$Pr(X \geq a) = Pr(e^{tX} \geq e^{ta})$$

for all $t > 0$. By the Markov inequality,

$$Pr(e^{tX} \geq e^t a) \leq \frac{E(e^{tX})}{e^{ta}}$$

Thus,

$$Pr(X \geq a) \leq \frac{E(e^{tX})}{e^{ta}} \tag{3.7}$$

By our definitions above,

$$\frac{E(e^{tX})}{e^{ta}} = \frac{E(Y)}{e^{ta}} = \frac{\prod e^{tX_i}}{e^{ta}} \tag{3.8}$$

Because we want to choose the t which gives us the best possible upper bound, we want to find:

$$Pr(X \geq a) \leq \min_{t>0} \frac{E(e^{tX})}{e^{ta}} \tag{3.9}$$

Lemma 3.6 $E(e^{tX_i}) \leq 1 + p_i(e^t - 1) \leq e^{(e^t-1)p_i}$

Proof: Suppose X_i is a Bernoulli random variable; that is,

$$X_i = \begin{cases} 1 & \text{with probability p} \\ 0 & \text{otherwise} \end{cases}$$

Then,

$$E(e^{tX_i}) = e^t p_i + (1 - p_i)$$

For all $X_i \in [0, 1]$, observe that e^{tx} is a convex function of x (i.e. the 2nd derivative of the function is always positive). This means that for all $x \in [0, 1]$:

$$e^{tx} \leq 1 + (e^t - 1)x$$

Thus,

$$\begin{aligned} E(e^{tX_i}) &\leq E(1 + (e^t - 1)X_i) \\ &\leq 1 + (e^t - 1)E(X_i) \\ &\leq 1 + (e^t - 1)p_i \end{aligned}$$

Because $1 + x \leq e^x \ \forall x$, it follows that:

$$E(e^{tX_i}) \leq e^{(e^t-1)p_i}$$

■

Now, to complete the proof, consider, by (3.7) and (3.8):

$$Pr(X \geq (1+\delta)\mu) \leq \frac{E(e^{tX})}{e^{t(1+\delta)\mu}} \leq \frac{\prod_{i=1}^{n} e^{(e^t-1)p_i}}{e^{t(1+\delta)\mu}} = e^{(e^t-1)\sum p_i - t(1+\delta)\mu} \tag{3.10}$$

Plugging this back into (3.9), we should now choose some $t > 0$ to minimize:

$$f(t) = e^{[(e^t-1)-t(1+\delta)]\mu}$$

Taking the first derivative, we get:

$$f'(t) = [e^t - (1+\delta)]e^{[(e^t-1)-t(1+\delta)]\mu}$$

$f'(t) = 0$ when $e^t = 1+\delta$; thus, when $t = \ln(1+\delta)$.

Taking the second derivative, we get:

$$f''(t) = [e^t - (1+\delta)]^2 e^{[(e^t-1)-t(1+\delta)]\mu} + e^t e^{[(e^t-1)-t(1+\delta)]\mu}$$

The exponentials enforce that this is always positive; thus the critical point at $t = \ln(1+\delta)$ is, in fact, a minimum.

Plugging this value of t back into (3.10), we get:

$$\begin{aligned} Pr(X \geq (1+\delta)\mu) &\leq e^{(e^{\ln(1+\delta)}-1)-\ln(1+\delta)(1+\delta)\mu} \\ &\leq e^{[(1+\delta)-1)-(1+\delta)\ln(1+\delta)]\mu} \\ &\leq e^{-\mu[(1+\delta)\ln(1+\delta)-\delta]} \end{aligned}$$

This proves the strongest version of the inequality. For the weaker version (3.5), we may start with the stronger version and simply relax it:

$$\begin{aligned} e^{-\mu[(1+\delta)\ln(1+\delta)-\delta]} &\leq e^{-\frac{\mu\delta^2}{3}} && \forall\delta \in [0,1] \\ -\mu[(1+\delta)\ln(1+\delta)-\delta] &\leq -\mu\frac{\delta^2}{3} && \forall\delta \in [0,1] \\ (1+\delta)\ln(1+\delta)-\delta &\geq \frac{\delta^2}{3} && \forall\delta \in [0,1] \\ (1+\delta)\ln(1+\delta)-\delta-\frac{\delta^2}{3} &\geq 0 && \forall\delta \in [0,1] \end{aligned}$$

The function is 0 at $\delta = 0$. Taking the first derivative, we get:

$$\ln(1+\delta) - \frac{2\delta}{3}$$

The derivative is positive along $x \in [0, 1]$; thus, the function itself must be increasing along that interval, satisfying the inequality.

Finally, for the other weaker version (3.6), we take a similar approach.

$$\begin{aligned} e^{-\mu[(1+\delta)\ln(1+\delta)-\delta]} &\leq e^{-\frac{\mu\delta}{3}} && \forall\delta \geq 1 \\ -\mu[(1+\delta)\ln(1+\delta)-\delta] &\leq -\mu\frac{\delta}{3} && \forall\delta \geq 1 \\ (1+\delta)\ln(1+\delta)-\delta &\geq \frac{\delta}{3} && \forall\delta \geq 1 \\ (1+\delta)\ln(1+\delta)-\frac{4\delta}{3} &\geq 0 && \forall\delta \geq 1 \end{aligned}$$

The function is positive at $\delta = 1$. Taking the first derivative, we get:

$$\ln(1+\delta) - \frac{1}{3}$$

The derivative is positive and monotonically increasing for all $\delta >= 1$; thus, the original inequality will always be satisfied.

■

3.2.3.2 Chernoff-Hoeffding Lower Trail

Theorem 3.7

$$\forall\delta \in [0,1] : Pr(X \leq (1-\delta)\mu) \leq e^{-\mu[\delta+(1-\delta)\ln(1-\delta)]}$$

A weaker, but also useful version, is:

$$\forall\delta \in [0,1] : Pr(X \leq (1-\delta)\mu) \leq e^{-\frac{\mu\delta^2}{2}}$$

Note that, though the statement is weaker, it suffices to have $\mu \leq E(X)$.

Proof: Observe that $Pr(X \leq (1-\delta)\mu) = Pr(-X \geq (-(1-\delta)\mu))$. Define $Y = e^{-tX}$ and proceed as in the proof for the upper bound. ■

3.2.4 Union Bound

The Union Bound is a seemingly trivial bound which proves to be very useful in analyzing many randomized algorithms.

Intuitively, the Union Bound shows that the probability of any events in the probability space occurring is bounded above by the sum of the probabilities of each event in the probability space.

Logically:

Let $E_1, E_2, \ldots, E_n$ be any collection of events in the probability space.

Then, $Pr(E_1 \vee E_2 \vee \ldots \vee E_n) \leq \sum_{i=1}^{n} Pr(E_i)$

3.2.4.1 Application: Balls in Bins

Consider the application of Chernoff and Union bounds to the following scenario: there are n bins, and n balls to be thrown. Balls are thrown one at a time, independently, and with equal probability of ending up in each bin (i.e. $Pr(ballilandsinballm) = \frac{1}{n}$). let B_i = the number of balls in bin i after all n balls have been thrown. Then the following is true:

Theorem 3.8

$$Pr(MAX(B_i) \geq \frac{e \ln n}{\ln \ln n}) \leq \frac{1}{n^c}$$
where c is some constant value > 0

The proof of this theorem is based upon the idea that any particular bin is so unlikely to exceed the given value, that the use of a Union bound will suffice to give us our desired probability bounds.

Proof: Consider a particular bin, say, bin 1.

$$let\ X_i = \begin{cases} 1 & \text{if } i^{th} \text{ ball landed in bin 1} \\ 0 & \text{otherwise} \end{cases}$$

as such,

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

There are n bins, each with equal probability of having the i^{th} ball land in the. We may thus calculate the expected value of each arbitrary X_i as follows:

$$E(X_i) = (\frac{1}{n} * 1) + (\frac{n-1}{n} * 0)$$
$$E(X_i) = \frac{1}{n}$$

And hence we may find the expected number of balls that land in bin 1:

$$E(B_1) = \sum_{i=1}^{n} \frac{1}{n}$$
$$E(X_i) = 1$$

With these values in place, we may apply our strongest Chernoff bounds:

$$Pr(B_1 \geq (1+\delta)E(B_1)) \leq e^{-E(B_1)[(1+\delta)\ln(1+\delta)-\delta]}$$

In this case, we define:

$$c = \frac{e \ln n}{\ln \ln n}$$

and we let

$$\delta = \frac{e \ln n}{\ln \ln n} - 1$$
$$\delta = c - 1$$

And thus, by plugging our variable definitions back into our probability:

$$Pr(B_1 \geq c) \leq e^{-c \ln c + c - 1}$$

which we may further weaken and simplify by noting that: $Pr(B_i \geq c) \leq e^{-c \ln c + c - 1} \leq e^{-c \ln c + c} = e^{-c(\ln c - 1)}$. Thus, by transitivity:

$$Pr(B_1 \geq c) \leq e^{-c(\ln c - 1)}$$

In order to simplify this probability further, we must bound our $\ln c$ term with something that simplifies nicely:

Lemma 3.9 $\ln c \geq 1 + \frac{\ln \ln n}{2}$

Proof: We start with the power series expansion of e:

$$e^v = 1 + v + \frac{v^2}{2} + ...$$

We can see from this expansion that it is trivially true that, when $v \geq 1$:

$$e^v \geq 2v$$

Now, letting $v = \ln \ln \ln n$:

$$\begin{aligned} e^{\ln \ln \ln n} &\geq 2 \ln \ln \ln n \\ \ln \ln n &\geq 2 \ln \ln \ln n \\ \frac{\ln \ln n}{2} &\geq \ln \ln \ln n \end{aligned}$$

hence:

$$\begin{aligned} \ln \ln n - \ln \ln \ln n &\geq \frac{\ln \ln n}{2} \\ 1 + \ln \ln n - \ln \ln \ln n &\geq 1 + \frac{\ln \ln n}{2} \end{aligned}$$

Now, recalling our definition of c:

$$\ln c = 1 + \ln \ln n - \ln \ln \ln n$$

And thus,

$$\ln c \geq 1 + \frac{\ln \ln n}{2}$$

thereby completing the proof of our lemma. ■

Now, with our lemma in place, we may substitute this bound for $\ln c$ back into our Chernoff bound from (3.2.4.1):

$$\begin{aligned}
Pr(B_1 \geq c) &\leq e^{-c(\ln c - 1)} \\
Pr(B_1 \geq c) &\leq e^{-c(1+\frac{\ln\ln n}{2}-1)} && [by\ (3.9)] \\
Pr(B_1 \geq c) &\leq e^{-c\frac{\ln\ln n}{2}} \\
Pr(B_1 \geq c) &\leq e^{-e(\frac{\ln n}{2})} && [c = \frac{e\ln n}{lnlnn}] \\
Pr(B_1 \geq c) &\leq n^{\frac{-e}{2}} && [arithmetic]
\end{aligned}$$

And we may simply bound our $\frac{-e}{2}$ term by noting that it is ≤ -1.35, giving us:

$$Pr(B_1 \geq c) \leq n^{-1.35} \tag{3.11}$$

So we have seen that the probability that the number of balls in a particular bin exceeds our value c is quite low. We may use this probability, with our Union Bound, in order to bound the probability that *any* bin might more than c balls. Now, let:

$$E_j \equiv \textit{the event that } B_j \geq c$$

So, with this notation in hand:

$$\begin{aligned}
Pr(max\ B_i \geq c) &= Pr(E_1 \vee ... \vee E_n) \\
Pr(max\ B_i \geq c) &\leq \sum_{i=1}^{n} Pr(E_i) && [by\ Union\ bound] \\
Pr(max\ B_i \geq c) &\leq \sum_{i=1}^{n} n^{-1.35} && [by\ (3.11)] \\
Pr(max\ B_i \geq c) &\leq n * n^{-1.35} \\
Pr(max\ B_i \geq c) &\leq n^{-0.35}
\end{aligned}$$

We have thus found our bound for the probability that any bin ends up with at least $\frac{e\ln n}{\ln\ln n}$ balls, and our proof is complete. ■