CSIS0351/8601: Randomized Algorithms
Lecture 4: Chernoff Bound: Measure Concentration
Lecturer: Hubert Chan
Date: 19 Sept 2011

These lecture notes are supplementary materials for the lectures. They are by no means substitutes for attending lectures or replacement for your own notes!

1 Measure Concentration

As we have seen in the previous lectures, the objective function of a problem can be expressed as some random variable Y, and we analyze the performance of a randomized algorithm in terms of the expectation (or mean) $E[Y]$. We often wish to show that with a large probability, the random variable Y is near its mean $E[Y]$. We see that if Y is a sum of *independent* random variables, then this is indeed the case. This phenomenon is known as *measure concentration.*

1.1 Example: Chebyshev's Inequality

Suppose $X_0, X_1, \ldots, X_{n-1}$ are independent $\{0, 1\}$-random variables such that for each i, $Pr(X_i = 1) = p$ and $Pr(X_i = 0) = 1 - p$. Let $Y := \sum_i X_i$. We have $E[Y] = np$.

Remark 1.1 By using pairwise independence of the X_i's, we have $var[Y] = np(1 - p)$.

Using the Chebyshev's Inequality, we have for $0 < \epsilon < 1$,

$$Pr(|Y - E[Y]| \geq \epsilon E[Y]) \leq \frac{var[Y]}{(\epsilon E[Y])^2} = \frac{1-p}{\epsilon^2 p} \cdot \frac{1}{n}$$

We have only used the fact that any two different random variables X_i and X_j are independent. The goal is to show that if we fully exploit the fact that all the random variables $X_0, X_1, \ldots, X_{n-1}$ are independent of one another, we can obtain a much better result.

Theorem 1.2 (Basic Chernoff Bound) *Suppose Y is the sum of n independent $\{0, 1\}$-random variables X_i's such that for each i, $Pr(X_i = 1) = p$. Let $\mu := E[Y] = np$. Then, for $0 < \epsilon < 1$,*

$$Pr(|Y - E[Y]| \geq \epsilon E[Y]) \leq 2 \exp\{-\frac{1}{3}\epsilon^2 np\}.$$

2 Using Moment Generating Function

The bound in Theorem 1.2 measures, in terms of $E[Y]$, how far the random variable Y is away from its mean $E[Y]$. One can instead measure this in terms of the total number of random variables n, i.e., one wants to analyze the probability $Pr(|Y - E[Y]| \geq \epsilon n)$. Of course, a different bound would be obtained. There are a number of variations of this inequality: Hoeffding's Inequality, Azuma's Inequality, McDiarmid's Inequalities. Each one of them has slightly different assumptions, and it would be confusing to learn them separately. Fortunately, there is a generic method to obtain all of them: the method of moment generating function.

We describe in general terms. Suppose $X_0, X_1, \ldots, X_{n-1}$ are independent random variables. They can take any value (not necessarily in $\{0,1\}$), and need not even be identically distributed. Let $Y := \sum_i X_i$ and $\mu := E[Y]$. The goal is to give an upper bound on the probability $Pr[|Y - \mu| \geq \alpha]$, for some value $\alpha > 0$. We outline the steps in the following.

2.1 Transform the Inequality into a Convenient Form

We first use the equation:

$$Pr[|Y - \mu| \geq \alpha] = Pr[Y - \mu \geq \alpha] + Pr[Y - \mu \leq -\alpha]. \tag{2.1}$$

We bound each of the term on the right hand side separately. Recall that $Y := \sum_i X_i$. Sometimes it would be convenient to first rescale each random variable X_i. For example,

1. $Z_i := X_i$. The simplest case. We can just work with X_i.
2. $Z_i := X_i - E[X_i]$. We have $E[Z_i] = 0$.
3. $Z_i := \frac{X_i}{R}$. If X_i is in the range $[0, R]$, then we now have $Z_i \in [0, 1]$.

Since the X_i's are independent, the Z_i's are also independent. After the transformation, the two terms in (2.1) have the form

(i) $Pr[\sum_i Z_i \geq \beta]$, or

(ii) $Pr[\sum_i Z_i \leq \beta]$.

Note that the β in each case is different. The direction of the inequality is also different. We use a trick to turn both inequalities into the same form. In case (i), let $t > 0$; in case (ii), let $t < 0$. Now, both inequalities have the same form

$$Pr[t \sum_i Z_i \geq t\beta] \tag{2.2}$$

The value t would be chosen later to get the best possible bound. Note that we have to remember whether t is positive or negative.

Example.

As part of the Basic Chernoff Bound, suppose we wish to consider the part $Pr[Y - \mu \leq -\epsilon\mu]$. In this case, we just let $Z_i := X_i$ and let $t < 0$ to obtain

$Pr[Y - \mu \leq -\epsilon\mu] = Pr[t \sum_i X_i \geq t(1 - \epsilon)\mu]$.

2.2 Using Moment Generating Function and Independence

Notation: we write $\exp(x) = e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!}$.

Observe that the exponentiation function is strictly increasing, i.e. $x < y$ *iff* $\exp(x) < \exp(y)$. Hence,

$Pr[t \sum_i Z_i \geq t\beta] = Pr[\exp(t \sum_i Z_i) \geq \exp(t\beta)]$.

Notice that now both sides of the inequality are positive. Hence, by Markov's Inequality, we have

$Pr[\exp(t \sum_i Z_i) \geq \exp(t\beta)] \leq \exp(-t\beta) E[\exp(t \sum_i Z_i)]$.

The next step is where we use the fact that the Z_i's are independent:

$E[\exp(t \sum_i Z_i)] = \prod_i E[\exp(tZ_i)]$.

Definition 2.1 *Given a random variable Z, the moment generating function is given by the mapping $t \mapsto E[e^{tZ}]$.*

Hence, it suffices to find an upper bound for $E[e^{tZ_i}]$, for each i.

Remark 2.2 We wish to find an upper bound of the form $E[e^{tZ_i}] \leq \exp(g_i(t))$ for some appropriate function $g_i(t)$. Note that this is often the most technical part of the proof, and requires tools from calculus.

Hence, we obtain the bound

$Pr[t \sum_i Z_i \geq t\beta] \leq \exp(-t\beta) \prod_i E[e^{tZ_i}] \leq \exp(-t\beta) \prod_i \exp(g_i(t)) = \exp(-t\beta + \sum_i g_i(t)) = \exp(g(t))$,

where $g(t) := -t\beta + \sum_i g_i(t)$.

Example.

Continuing with our example, if $Z_i = X_i$ is a $\{0, 1\}$-random variable such that $Pr(X_i = 1) = p$, then we have

$E[e^{tZ_i}] = (1 - p) \cdot e^0 + p \cdot e^t = 1 + p(e^t - 1) \leq \exp(p(e^t - 1))$,

where we have used the inequality $1 + x \leq e^x$, for all real numbers x.

Hence,

$Pr[t \sum_i X_i \geq t(1 - \epsilon)\mu] \leq \exp\{-t(1 - \epsilon)\mu + np(e^t - 1)\} = \exp(g(t))$,

where $g(t) := \mu(e^t - t(1 - \epsilon) - 1)$.

2.3 Find the Best Value for t to Minimize $g(t)$

We find the value t that minimizes the function $g(t) := -t\beta + \sum_i g_i(t)$. Be careful to remember whether t is positive or negative!

Example.

In our example, we have $g(t) := \mu(e^t - t(1 - \epsilon) - 1)$.

Note that $g'(t) = \mu(e^t - (1-\epsilon))$ and $g''(t) = \mu e^t > 0$. It follows that g attains its minimum when $g'(t) = 0$, i.e., when $t = \ln(1-\epsilon) < 0$.

We check that in our example, $t < 0$. So, we can set the value $t := \ln(1-\epsilon)$. Using the expansion for $0 < \epsilon < 1$, $-\ln(1-\epsilon) = \sum_{i \geq 1} \frac{\epsilon^i}{i}$, we have $g(\ln(1-\epsilon)) \leq -\frac{\epsilon^2 \mu}{2} = -\frac{\epsilon^2 np}{2}$.

So, we have one part of the Basic Chernoff Bound,

$Pr[Y - \mu \leq -\epsilon\mu] \leq \exp(-\frac{\epsilon^2 np}{2}) \leq \exp(-\frac{\epsilon^2 np}{3})$.

Theorem 2.3 *Suppose $X_0, X_1, \ldots, X_{n-1}$ are independent $\{0,1\}$-random variables, each having expectation p. Let $Y := \sum_i X_i$ and $\mu := E[Y]$.*

Then, for $0 < \epsilon < 1$, $Pr[Y \leq (1-\epsilon)\mu] \leq \exp(-\frac{\epsilon^2 \mu}{2})$.

3 The Other Half of Chernoff

To complete the proof of the Chernoff Bound, one also needs to obtain an upper bound for $[Y - \mu \geq \epsilon\mu]$. The same technique of moment generating function can be applied. The calculations might be different though. We would leave the details as a homework problem.

Lemma 3.1 *Suppose $X_0, X_1, \ldots, X_{n-1}$ are independent $\{0,1\}$-random variables, each having expectation p. Let $Y := \sum_i X_i$ and $\mu := E[Y]$.*

Then, for all $\epsilon > 0$, $Pr[Y \geq (1+\epsilon)\mu] \leq (\frac{e^\epsilon}{(1+\epsilon)^{1+\epsilon}})^\mu$.

Corollary 3.2 *For $0 < \epsilon < 1$, using the inequality $(1+\epsilon)\ln(1+\epsilon) \geq \epsilon + \frac{\epsilon^2}{3}$, we have:*

$Pr[Y \geq (1+\epsilon)\mu] \leq \exp(-\frac{\epsilon^2 \mu}{3})$.

Corollary 3.3 (Chernoff Bound with Large ϵ) *For all $\epsilon > 0$, using the inequality $\ln(1+\epsilon) > \frac{2\epsilon}{2+\epsilon}$, we have:*

$Pr[Y \geq (1+\epsilon)\mu] \leq \exp(-\frac{\epsilon^2 \mu}{2+\epsilon})$.

4 2-Coloring Subsets: Revisited

Consider a finite set U and subsets $S_1, S_2, \ldots, S_m$ of U such that each S_i has size $|S_i| = l$, where $l > 12 \ln m$. Is it possible to color each element of U *red* or *blue* such that each set S_i contains roughly the same number of red and blue elements?

Proposition 4.1 *Fix a subset S_i, let X_i be the number of red elements in S_i.*

Then, $Pr[|X_i - \frac{l}{2}| \geq \sqrt{3l \ln m}] \leq \frac{2}{m^2}$.

Proof: Note that $E[X_i] = \frac{l}{2}$. By Chernoff Bound, for $0 < \epsilon < 1$,

$Pr[|X_i - \frac{l}{2}| \geq \epsilon E[X_i]] \leq 2\exp(-\frac{1}{3}\epsilon^2 E[X_i])$.

Substituting $\epsilon := \sqrt{\frac{12 \ln m}{l}} < 1$, we have the result.

■

Corollary 4.2 *By the union bound,* $Pr[\exists i, |X_i - \frac{l}{2}| \geq \sqrt{3l \ln m}] \leq \frac{2}{m}$.

5 n Balls into n Bins: Load Balancing

Suppose one throws n balls into n bins, independently and uniformly at random. We wish to analyze the maximum number of balls in any single bin. A similar situation arises when there are n jobs independently and randomly assigned to n machines, and we wish to analyze the number of jobs assigned to the busiest machine.

Consider the first bin, and let Y_1 be the number of balls in it. Note that Y_1 is a sum of n independent $\{0, 1\}$-random variables, each having expectation $\frac{1}{n}$.

Proposition 5.1 $Pr[Y_1 \geq 4 \ln n + 1] \leq \frac{1}{n^2}]$.

Proof: Observe that $E[Y_1] = 1$, we use Chernoff Bound with large $\epsilon > 0$ (Corollary 3.3). We have:

$Pr[Y_1 \geq 1 + \epsilon] \leq \exp(-\frac{\epsilon^2}{2+\epsilon})$.

We wish to find a value for ϵ so that the last quantity is at most $\frac{1}{n^2}$.

For $\epsilon \geq 2$, we have $\frac{\epsilon^2}{2+\epsilon} \geq \frac{\epsilon^2}{2\epsilon} = \frac{\epsilon}{2}$. Hence, the last quantity is at most $\exp(-\frac{\epsilon}{2})$, which equals $\frac{1}{n^2}$, if we set $\epsilon := 4 \ln n \geq 2$. ■

Corollary 5.2 *Using union bound, the probability that there exists a bin with more than* $1 + 4 \ln n$ *balls is at most* $\frac{1}{n}$.

6 Homework Preview

1. **The Other Half of Chernoff.** Suppose $X_0, X_1, \ldots, X_{n-1}$ are independent $\{0, 1\}$-random variables, each having expectation p. Let $Y := \sum_i X_i$ and $\mu := E[Y]$. Using the method of moment generating function, prove the following.

   For all $\epsilon > 0$, $Pr[Y - \mu \geq \epsilon\mu] \leq \left(\frac{e^\epsilon}{(1+\epsilon)^{1+\epsilon}}\right)^\mu$.

2. n **Balls into** n **Bins (Revisited).** Using the Chernoff Bound from the previous question, we can obtain a better bound for the balls and bins problem. Suppose n balls are thrown independently and uniformly at random into n bins. Let Y_1 be the number of balls in the first bin.

   (a) Find a number N in terms of n such that $Pr[Y_1 \geq N] \leq \frac{1}{n^2}$. Please give the exact form and do not use big O notation for this part of the question.
   (Hint: if you need to find a number W such that $W \ln W \geq \ln n$, try setting $W := \frac{\lambda \ln n}{\ln \ln n}$, for some constant $\lambda > 0$. You can also assume that n is large enough, say $n \geq 100$.)

   (b) Show that with probability at least $1 - \frac{1}{n}$, no bin contains more than $\Theta(\frac{\log n}{\log \log n})$ balls.