CS 591: Data Analytics – Spring 2017 January 30, 2017

Lecture 3

Lecturer: Charalampos E. Tsourakakis *Scribe: Charalampos E. Tsourakakis*

1 Basic tools

We will use the following inequalities to bound the binomial coefficient $\binom{n}{k}$.

$$\boxed{\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{en}{k}\right)^k.}$$

We will also need to be able to upper- and lower-bound certain expressions. Here are some useful inequalities.

$$\boxed{(1-x)^n \geq 1-nx, \;\; \forall 0 \leq x \leq 1.}$$

$$\boxed{e^x \geq x+1, \;\; \forall x.}$$

$$\boxed{e^x \leq x^2+x+1, \quad \forall 0 < |x| < 1.}$$

$$\boxed{\log{(1+x)} = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots, \quad \forall 0 < |x| < 1.}$$

$$\boxed{\big(\sum_{i=1}^n a_i^2\big)\big(\sum_{i=1}^n b_i^2\big) \geq \big(\sum_{i=1}^n a_i b_i\big)^2.}$$

The above inequality is the Cauchy-Schwartz inequality, which is a special case of Hölder's inequality for $p = q = 2$.

Theorem 1 (Hölder's inequality) *For any positive real numbers p, q such that $\frac{1}{p} + \frac{1}{q} = 1$*

$$\boxed{\big(\sum_{i=1}^n |a_i b_i|\big) \leq \big(\sum_{i=1}^n |x_i|^p\big)^{1/p}\big(\sum_{i=1}^n |x_i|^p\big)^{1/p}.}$$

The following inequalities are basic probabilistic tools.

Theorem 2 (Markov's Inequality) *Let X a be non-negative integer valued random variable. Then,*

$$\mathbf{Pr}\left[X \geq t\right] \leq \frac{\mathbb{E}\left[X\right]}{t}.$$

Proof:

$$\mathbb{E}\left[X\right] = \sum_{k \geq 1} k\mathbf{Pr}\left[X = k\right] \geq \sum_{k=t} k\mathbf{Pr}\left[X = k\right] \geq t \sum_{k=t} \mathbf{Pr}\left[X = k\right] = t\mathbf{Pr}\left[X \geq t\right].$$

■

We will use this inequality in two ways in our class. First, it is the basis of the first moment method. In many cases we will need to show that $\mathbf{Pr}\left[X > 0\right] = o(1)$, where X is a non-negative random variable of interest. It turns out that computing $\mathbb{E}\left[X\right]$ can be much easier than directly computing $\mathbf{Pr}\left[X > 0\right]$ in numerous cases. If $\mathbb{E}\left[X\right] = o(1)$ then by Markov's inequality

$$\boxed{\mathbf{Pr}\left[X > 0\right] \leq \mathbb{E}\left[X\right]}$$

we obtain that $X = 0$ **whp** . Furthermore, we will use Markov's inequality to obtain probabilistic inequalities for higher order moments. This is a special case of the following observation. If ϕ is a strictly monotonically increasing function, then

$$\mathbf{Pr}\left[X \geq t\right] = \mathbf{Pr}\left[\phi(X) \geq \phi(t)\right] \leq \frac{\mathbb{E}\left[\phi(X)\right]}{\phi(t)}.$$

For instance, if $\phi(x) = x^2$, then we obtain Chebyshev's inequality.

Theorem 3 (Chebyshev's Inequality) *Let X be any random variable. Then,*

$$\mathbf{Pr}\left[|X - \mathbb{E}\left[X\right]| \geq t\right] \leq \frac{\mathbb{V}ar\left[X\right]}{t^2}.$$

A simple corollary of Chebyshev's inequality is the following:

Corollary 4 (Second moment method) *Let X be a non-negative integer valued random variable. Then,*

$$\boxed{\mathbf{Pr}\left[X = 0\right] \leq \frac{\mathbb{V}ar\left[X\right]}{(\mathbb{E}\left[X\right])^2}.}$$

For completeness, here is the proof.

Proof:

$$\mathbf{Pr}\left[X = 0\right] \leq \mathbf{Pr}\left[|X - \mathbb{E}\left[X\right]| \geq \mathbb{E}\left[X\right]\right] \leq \frac{\mathbb{V}ar\left[X\right]}{(\mathbb{E}\left[X\right])^2}.$$

■

The use of the above corollary is known as the second moment method. Here is how we will typically use it in our class. Let the random variable X of interest be the sum of m indicator random variables $X_1, \ldots, X_m$, where $\mathbf{Pr}\left[X_i = 1\right] = p_i$, i.e.,

$$X = X_1 + \ldots + X_m.$$

We will be interested in showing that $X > 0$ **whp** . Even if $\mathbb{E}\left[X\right]$ will tend to $+\infty$ this does not suggest that $X > 0$ **whp** . In order to prove this kind of statement, we will use the second moment method. Since $\mathbf{Pr}\left[X = 0\right] \leq \frac{\mathbb{V}ar[X]}{(\mathbb{E}[X])^2}$ it will suffice to prove that $\frac{\mathbb{V}ar[X]}{(\mathbb{E}[X])^2} = o(1)$. The problem therefore is reduced to computing or actually *upper-bounding* the variance.

In our typical setting,

$$\mathbb{V}ar\left[X\right] = \sum_{i=1}^{m} \mathbb{V}ar\left[X_i\right] + \sum_{i \neq j} \mathbb{C}ov\left[X_i, X_j\right] \leq \mathbb{E}\left[X\right] + \sum_{i \neq j} \mathbb{C}ov\left[X_i, X_j\right].$$

To see how we obtained the inequality, notice that $\mathbb{V}ar\left[X_i\right] = p_i(1-p_i) \leq p_i = \mathbb{E}\left[X_i\right]$. Hence by the linearity of expectation $\sum_i \mathbb{V}ar\left[X_i\right] \leq \sum_i \mathbb{E}\left[X_i\right] = \mathbb{E}\left[X\right]$. The covariance of two random variables A, B is defined as

$$\mathbb{C}ov\left[A, B\right] = \mathbb{E}\left[AB\right] - \mathbb{E}\left[A\right]\mathbb{E}\left[B\right].$$

In the case of indicator random variables we obtain the following expression:

$$\mathbb{C}ov\left[X_i, X_j\right] = \mathbf{Pr}\left[X_i = X_j = 1\right] - \mathbf{Pr}\left[X_i = 1\right]\mathbf{Pr}\left[X_j = 1\right].$$

So, when we apply the second moment, typically the hard part it to upper bound the sum of covariances.

2 Exponential Moment Method

2.1 Outline

In the previous section, we saw that we can use Markov's Inequality to obtain probabilistic inequalities for higher order moments. Specifically, we saw that if ϕ is a strictly monotonically increasing function, then

$$\mathbf{Pr}\left[X \geq t\right] = \mathbf{Pr}\left[\phi(X) \geq \phi(t)\right] \leq \frac{\mathbb{E}\left[\phi(X)\right]}{\phi(t)}.$$

For instance, for $\phi(x) = x^2$ we obtained Chebyshev's inequality, that tells us that the random variable X takes value $\mathbb{E}\left[X\right] + O(\lambda \mathbb{V}ar\left[X\right])$ with probability $1 - O(\lambda^{-2})$. This means that the tail of the probability distribution decays as $O(\lambda^{-2})$. In numerous cases, we are able to get control of higher moments of the variable X. So we may ask, whether we can use this to get better tail estimates. Today, we are going to discuss the exponential moment method which results in deriving the famous Chernoff bounds. The typical setting we are going to deal with is when the random variable is the sum of random variables which are either jointly independent or negatively associated or "almost" independent, in the sense that there may be dependencies but they will be weak. We are going to focus typically on integer-valued, non-negative variables in the context of our class, but keep in mind that these results apply to other settings as well. For instance, there exist Chernoff bounds for complex-valued random variables. Furthermore, in this and the next lectures we are going to see applications of probabilistic tools on random graphs. These tools come up very often in the analysis of randomized algorithms, see [Motwani and Raghavan, 2010].

2.2 Details

In the place of ϕ above, we will use the exponential function. Specifically, let $t > 0, \lambda \in \mathbb{R}$. Then, we obtain

$$\mathbf{Pr}\left[X \geq \lambda\right] = \mathbf{Pr}\left[e^{tX} \geq e^{t\lambda}\right] \leq \frac{\mathbb{E}\left[e^{tX}\right]}{e^{t\lambda}},$$

and

$$\mathbf{Pr}\left[X \leq -\lambda\right] = \mathbf{Pr}\left[e^{-tX} \geq e^{t\lambda}\right] \leq \frac{\mathbb{E}\left[e^{-tX}\right]}{e^{t\lambda}}.$$

The core idea of Chernoff bounds is to set t to a value that minimizes the right-hand side probabilities. Sometimes, we may choose sub-optimal values of t in order to get simpler bounds that will still be good enough for our purposes. The function $M_X(t) = \mathbb{E}\left[e^{tX}\right]$ has a special name since it is an important function. So let's define it so that we can frequently refer to it.

Definition 5 (Moment Generating Function) *The function $t \mapsto \mathbb{E}\left[e^{tX}\right]$ is known as the moment generating function (***mgf***) of X.*

It is called like this since by the Taylor series for the exponential

$$\mathbb{E}\left[e^{tX}\right] = 1 + t\mathbb{E}\left[X\right] + \frac{t^2}{2!}\mathbb{E}\left[X^2\right] + \ldots + \frac{t^n}{n!}\mathbb{E}\left[X^n\right] + \ldots$$

we see that all moments of X appear. Specifically, if we take the n-th derivative of $M_X(t)$ and set $t = 0$, then we see that $\mathbb{E}\left[X^n\right] = M_X^{(n)}(0)$. There is a technicality here: we assumed that we can exchange the operands of expectation and differentiation. In general, this is valid when the moment generating function exists in a neighborhood of zero. A well-known distribution whose moment generating function takes infinite values for all $t \neq 0$ is the Cauchy distribution with density $f(x) = \frac{1}{\pi}\frac{1}{1+x^2}$. In the cases we are going to see in this class, the following assumption is going to hold and therefore we can safely exchange the operands of expectation and differentiation.

Assumption: Throughout this class, we will assume that the **mgf** exists in the sense that there exists a positive number $b > 0$ such that $M_X(t)$ is finite for all $|t| < b$.

Two more facts which are useful to keep in mind about **mgf**s follow.

Fact 1: The moment generating function uniquely defined the distribution. Speficially, let X, Y be two random variables. If $M_X(t) = M_Y(t)$ for all $t \in (-\delta, \delta)$ for some $\delta > 0$ then X, Y have the same distribution.

Fact 2: Let's assume that X, Y are two independent random variables. Then the **mgf** $M_{X+Y}(t)$ of the random variables X+Y is $M_X(t)M_Y(t)$. Since the proof is one line, let's see why this is true. The only things we need to use are definitions and the fact that e^{tX}, e^{tY} are independent.

$$M_{X+Y}(t) = \mathbb{E}\left[e^{t(X+Y)}\right] = \mathbb{E}\left[e^{tX}e^{tY}\right] = \mathbb{E}\left[e^{tX}\right]\mathbb{E}\left[e^{tY}\right] = M_X(t)M_Y(t).$$

Example: Let's compute the **mgf** of the binomial $Bin(n,p)$. The binomial is the sum of n independent Bernoulli random variables with parameter p, i.e., if $X \sim Bin(n,p)$, then $X = X_1 + \ldots + X_n$ where $X_i \sim Bernoulli(p)$ for all i. The **mgf** of such a variable is $\mathbb{E}\left[e^{tX_1}\right] = pe^t + (1-p)$. By fact 2, the **mgf** of X is $M_X(t) = (pe^t + (1-p))^n$.

Exercise: Work out the **mgf**s of the following two discrete probability distributions: the Poisson distribution with parameter λ and the geometric distribution with parameter p.

The following theorem shows the way we apply the exponential moment method.

Theorem 6 *Let X_i for $1 \le i \le n$ be jointly independent random variables with $\mathbf{Pr}\left[X_i = 1\right] = \mathbf{Pr}\left[X_i = -1\right] = \frac{1}{2}$. Let $S_n = \sum_{i=1}^n X_i$ and $\alpha > 0$. Then*

$$\mathbf{Pr}\left[|S_n| > \alpha\right] < 2e^{-\frac{\alpha^2}{2n}}.$$

Proof: By symmetry it suffices to prove that $\mathbf{Pr}\left[S_n > \alpha\right] < e^{-\frac{\alpha^2}{2n}}$. Let $t > 0$ be arbitrary. For $1 \le i \le n$

$$\mathbb{E}\left[e^{tX_i}\right] = \frac{e^t + e^{-t}}{2} = \cosh(t).$$

Since $\cosh(t) \le e^{t^2/2}$, see Lemma 7, for all $t > 0$ we obtain the following valid inequality:

$$\mathbb{E}\left[e^{tS_n}\right] = \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right] = (\cosh(t))^n \le e^{nt^2/2}.$$

Therefore, by the exponential moment method we obtain

$$\mathbf{Pr}\left[S_n > \alpha\right] \le e^{nt^2/2 - t\alpha}.$$

Setting $t = \frac{\alpha}{n}$ to minimize the right-hand side we get

$$\mathbf{Pr}\left[S_n > \alpha\right] < e^{-\frac{\alpha^2}{2n}}.$$

■

Lemma 7 *For all $t > 0$*

$$\cosh(t) \le e^{t^2/2}.$$

Proof: We will compare the Taylor expansions of the two hand-sides. By the definition of $\cosh(t)$ and the Taylor expansion for e^t we get

$$\cosh(t) = \frac{e^t + e^{-t}}{2} = \sum_{i=0}^{+\infty} \frac{t^{2i}}{(2i)!},$$

since the odd terms of the Taylor expansions for the exponentials cancel out, and the even terms get multiplied by 2 and then divided by 2. Check it. The Taylor expansion for the right-hand side is

$$e^{t^2/2} = \sum_{i=0}^{+\infty} \frac{(t^2/2)^i}{i!} = \sum_{i=0}^{+\infty} \frac{t^{2i}}{2^i i!}.$$

It is easy now to check that the inequality is true. Notice that for all $i \geq 1$

$$(2i)! = (2i)(2i-1)\dots(i+1) \times i \dots 1 \geq 2^i i!.$$

■

One can use the exponential moment method to derive the following useful Chernoff bounds, stated as facts. In the next homework you will have the chance to practice the exponential moment method.

Theorem 8 (Chernoff bound for $Bin(n,p)$) *For any $0 \leq t \leq np$*

$$\boxed{\mathbf{Pr}\left[|Bin(n,p) - np| > t\right] < 2e^{-\frac{t^2}{3np}}.}$$

For $t > np$

$$\boxed{\mathbf{Pr}\left[|Bin(n,p) - np| > t\right] < \mathbf{Pr}\left[|Bin(n,p) - np| > np\right] < 2e^{-\frac{np}{3}}.}$$

For all t

$$\boxed{\mathbf{Pr}\left[|Bin(n,p) - np| > t\right] < 2\exp\Big(-np\Big((1+\frac{t}{np})\ln(1+\frac{t}{np}) - \frac{t}{np}\Big)\Big).}$$

Let's prove another Chernoff-type bound for a random variable S_n that is the sum of the n jointly independent random variables $X_1, \dots, X_n$. We will assume that X_is are bounded. Despite the strong assumptions we make, what we will derive is a very useful bound. We will start by proving the following lemma.

Lemma 9 *Let X be a random variable with $|X| \leq 1$, $\mathbb{E}[X] = 0$. Then for any $|t| \leq 1$ the following holds:*

$$M_X(t) \leq e^{t^2 \mathbb{V}ar[X]}.$$

Proof: Given that $|tX| \leq 1$ the inequality $e^{tX} \leq 1 + tX + (tX)^2$ holds. By the linearity of expectation and the fact that $\mathbb{E}[tX] = t\mathbb{E}[X] = 0$ we obtain

$$\mathbb{E}\left[e^{tX}\right] \leq 1 + t^2\mathbb{E}\left[X^2\right] = 1 + t^2\mathbb{V}ar[X] \leq e^{t^2\mathbb{V}ar[X]}.$$

■

Theorem 10 (Chernoff for bounded variables) *Assume that $X_1, \ldots, X_n$ are jointly independent random variables where $|X_i - \mathbb{E}[X_i]| \le 1$ for all i. Let $S_n = \sum_{i=1}^n X_i$ and $\sigma = \sqrt{\mathbb{V}ar[S_n]}$ be the standard deviation of S_n. Then for any $\lambda > 0$*

$$\boxed{\mathbf{Pr}[|S_n - \mathbb{E}[S_n]| \ge \lambda\sigma] \le 2\max\left(e^{-\lambda^2/4}, e^{-\lambda\sigma/2}\right).}$$

Proof: Without loss of generality we may assume that $\mathbb{E}[X_i] = 0$ since if not we can subtract a constant from each of the X_is and normalize. Observe that by symmetry it suffices to prove $\mathbf{Pr}[S_n \ge \lambda\sigma] \le e^{-\frac{t\lambda\sigma}{2}}$ where $t = \min\left(\frac{\lambda}{2\sigma}, 1\right)$. Applying the exponential moment method and by taking into account the joint independence of X_is, $\sum_{i=1}^n \mathbb{V}ar[X_i] = \sigma^2$ and the previous lemma we obtain

$$\mathbf{Pr}[X \ge \lambda\sigma] \le e^{-t\lambda\sigma} \prod_{i=1}^n \mathbb{E}\left[e^{tX_i}\right] \le e^{-t\lambda\sigma} \prod_{i=1}^n e^{t^2 \mathbb{V}ar[X_i]} = e^{-t\lambda\sigma + t^2\sigma^2}.$$

Since $t \le \lambda/(2\sigma)$, the proof is complete. ■

Finally, a bound due to Hoeffding which is also known as Chernoff bound or Chernoff-Hoeffding bound is the following. Again, we make the same assumptions, namely $S_n = X_1 + \ldots + X_n$ where X_is are jointly independent and bounded.

Theorem 11 (Chernoff-Hoeffding bound) *Suppose $a_i \le X_i \le b_i$ for $i = 1, \ldots, n$. Then for all $t > 0$*

$$\boxed{\mathbf{Pr}[S_n \ge \mathbb{E}[S_n] + t] \le e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}},}$$

and

$$\boxed{\mathbf{Pr}[S_n \le \mathbb{E}[S_n] - t] \le e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}}.}$$

Combining them we get that

$$\boxed{\mathbf{Pr}[|S_n - \mathbb{E}[S_n]| \ge t] \le 2e^{-\frac{2t^2}{\sum_{i=1}^n (b_i - a_i)^2}}.}$$

3 Balls and Bins

Theorem 12 *When n balls are thrown independently and uniformly at random into n bins, the probability that the maximum load is more than $3\frac{\ln(n)}{\ln\ln(n)}$ is at most $\frac{1}{n}$ for n sufficiently large.*

Proof:[Sketch, details on whiteboard] Let Q_1 be the event that the first bin receives at least k balls. Then,

$$\mathbf{Pr}[Q_1] \le \binom{n}{k}\left(\frac{1}{n}\right)^k. \le \left(\frac{e}{k}\right)^k$$

Therefore, by setting $k = 3\frac{\ln(n)}{\ln\ln(n)}$

$$\mathbf{Pr}\left[\cup Q_i\right] \leq n\Big(\frac{e\ln\ln(n)}{3\ln(n)}\Big)^{3\frac{\ln(n)}{\ln\ln(n)}} \leq .. \leq \frac{1}{n}$$

■

References

[Motwani and Raghavan, 2010] Motwani, R. and Raghavan, P. (2010). Randomized algorithms. In *Algorithms and theory of computation handbook*, pages 12–12. Chapman & Hall/CRC.