Algorithms for Big Data Fall 2020

Lecture 02 — October 5, 2020

Prof. Cyrus Rashtchian *Topics: Probability Review*

Overview. In this lecture, we review basic probability inequalities. We will use these heavily throughout the course, so it is good to familiarize yourself with them and try out examples yourself to make sure that you understand how they work. The proofs of the inequalities will be less important than their applications.

1 Probability Review

We are mainly discussing discrete random variables. For this section we consider random variables as taking values in some set $S \subset \mathbb{R}$. Recall the expectation of X is defined

$$\mathbb{E}[X] = \sum_{j \in S} j \cdot \Pr(X = j).$$

Similarly, the variance of X is

$$\mathrm{Var}[X] = \mathbb{E}[(X - \mathbb{E}X)^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

We now state a few basic lemmas and facts that will be useful throughout the course. For more details (and proofs and examples), there are many excellent resources on Wikipedia and the course website (under "Supplemental Links").

Lemma 1 (Linearity of Expectation)**.** *For all reals a and b and random variables X and Y,*

$$\mathbb{E}[aX + bY] = a \cdot \mathbb{E}[X] + b \cdot \mathbb{E}[Y].$$

Lemma 2 (Union Bound)**.** *Let $\mathcal{E}_1, \ldots, \mathcal{E}_t$ be any collection of events in a shared probability space. Then,*

$$\Pr\left(\bigcup_{i=1}^{t} \mathcal{E}_i\right) \leq \sum_{i=1}^{t} \Pr(\mathcal{E}_i).$$

Lemma 3 (Markov's Inequality)**.** *If X is a nonnegative random variable, then for any $\lambda > 0$,*

$$\Pr(X \geq \lambda) \leq \frac{\mathbb{E}X}{\lambda}$$

Lemma 4 (Chebyshev's Inequality)**.** *For any random variable X and parameter $\lambda > 0$, we have*

$$\Pr(|X - \mathbb{E}X| \geq \lambda) \leq \frac{\mathbb{E}(X - \mathbb{E}X)^2}{\lambda^2} = \frac{\mathrm{Var}[X]}{\lambda^2}$$

Proof. Observe that

$$\Pr(|X - \mathbb{E}X| \geq \lambda) = \Pr((X - \mathbb{E}X)^2 > \lambda^2).$$

The claim follows by Markov's inequality. □

Also note that

$$\Pr(|X - \mathbb{E}X| \geq \lambda) = \Pr(|X - \mathbb{E}X|^p > \lambda^p)$$

for all $p \geq 1$. If we apply Markov's inequality to this statement, we get a more general version of Chebyshev's inequality:

$$\Pr(|X - \mathbb{E}X| \geq \lambda) \leq \frac{\mathbb{E}|X - \mathbb{E}X|^p}{\lambda^p}.$$

1.1 Exponential Tail Bounds

For a Normal random variable, we have very tight concentration.

Lemma 5 (Gaussian Tail Bound.)**.** *If $X \sim \mathcal{N}(\mu, \sigma^2)$, then for any $\lambda > 0$, we have*

$$\mathbf{Pr}[|X - \mu| > \lambda] \leq e^{-\frac{\lambda^2}{2\sigma^2}}.$$

Of course, we cannot expect the same thing to hold for all random variables (e.g., the uniform distribution is *not* concentrated). Fortunately, there is a common situation where we always get exponential tail bounds, that is, when a random variable X can be expressed as the sum of i.i.d. random variables $X = \sum_{i=1}^n x_i$. The resulting bound on the deviation probability is known as a Chernoff bound. There are many variants, such as Hoeffdings inequality (stated later), which deals with variables that have a larger range.

Lemma 6 (Chernoff Bound)**.** *Let $X_1, X_2, \ldots, X_n$ be independent random variables with $X_i \in [0, 1]$. Let $X = \sum_{i=1}^n X_i$ with $\mu = \mathbb{E}X$. If $0 < \varepsilon < 1$, then*

$$\Pr(|X - \mathbb{E}X| > \varepsilon\mu) \leq 2 \cdot e^{-\varepsilon^2\mu/3}.$$

Proof Idea. We will prove a weaker statement to gain intuition: we assume that the X_i are independent Bernoulli. Let $X_i = 1$ with probability p_i and $X_i = 0$ otherwise. Note then $\mu = \sum_{i=1}^n p_i$. Furthermore, we also do not attempt to achieve the constant $1/3$ in the exponent, but are happy to settle for any fixed constant. For the upper tail, we have for any $t > 0$ that

$$\begin{aligned}
\Pr(X > (1+\varepsilon)\mu) &= \Pr\left(e^{tX} > e^{t(1+\varepsilon)\mu}\right) \\
&\leq e^{-t(1+\varepsilon)\mu} \cdot \mathbb{E}e^{tX} \qquad \text{(Markov)} \\
&= e^{-t(1+\varepsilon)\mu} \cdot \mathbb{E}e^{\sum_i tX_i} \\
&= e^{-t(1+\varepsilon)\mu} \cdot \mathbb{E}\prod_i e^{tX_i} \\
&= e^{-t(1+\varepsilon)\mu} \cdot \prod_i \mathbb{E}e^{tX_i} \qquad \text{(independence)} \\
&= e^{-t(1+\varepsilon)\mu} \cdot \prod_i (1 - p_i + p_i e^t) \\
&= e^{-t(1+\varepsilon)\mu} \cdot \prod_i (1 + p_i(e^t - 1)) \\
&\leq e^{-t(1+\varepsilon)\mu} \cdot \prod_i e^{p_i(e^t - 1)} \qquad \text{(since } 1 + x \leq e^x\text{)} \\
&= e^{-t(1+\varepsilon)\mu + (e^t - 1)\mu} \\
&= \left(\frac{e^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}}\right)^\mu \qquad \text{(set } t = \ln(1+\varepsilon)\text{)} \qquad (1)
\end{aligned}$$

There are two regimes of interest for Eqn. (1). When $\varepsilon < 2$, by Taylor's theorem we have $\ln(1+\varepsilon) = \varepsilon - \varepsilon^2/2 + O(\varepsilon^3)$. Then replacing $(1+\varepsilon)^{1+\varepsilon}$ by $e^{(1+\varepsilon)\ln(1+\varepsilon)}$ and applying Taylor's theorem, this leads to (1) being $\approx e^{-\Theta(\varepsilon^2\mu)}$ in that regime. When $\varepsilon > 2$ we have $\ln(1+\varepsilon) = \Theta(\ln(\varepsilon))$, in which case the tail bounds becomes $\varepsilon^{-\Theta(\varepsilon\mu)}$. The lower tail analysis for $\Pr(X < (1-\varepsilon)\mu)$ is similar, noting that $X < (1-\varepsilon)\mu$ iff $e^{-tX} > e^{-t(1-\varepsilon)\mu}$. We then apply Markov then analyze $\mathbb{E}e^{-tX}$ and eventually set $t = -\ln(1-\varepsilon)$. The right hand side then becomes $(e^{-\varepsilon}/(1-\varepsilon)^{1-\varepsilon})^\mu$. $\square$

Lemma 7 (Hoeffding Bound)**.** *Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in [a_i, b_i]$, where $a_i \le b_i$ are parameters. Let $X = \sum_{i=1}^n X_i$. For any $\lambda > 0$, we have*

$$\Pr(|X - \mathbb{E}X| > \lambda) \le 2 \cdot e^{-\frac{\lambda^2}{\sum_{i=1}^n (b_i - a_i)^2}}.$$