18.405J/6.841J Advanced Complexity Theory March 1, 2001

Probability Handout

Daniel Spielman *Daniel Spielman*

1 Introduction

At the request of some students who took this class last year, I've written this short tutorial on some of the probability that we will use in this class. While I would like to believe that you've all seen all of this before, I've come to accept that this is not true. Really, the best reference for the probabilistic techniques that we use is "The Probabilistic Method" by Alon and Spencer. If you don't own it, you should. Another good source is "Randomized Algorithms" by Motwani and Raghavan.

2 Expectation

Throughout the following, X will be a discrete real random variable taking value x_i with probability p_i.

Definition 1 *The expectation of X, written $\mathbf{E}[X]$ is*

$$\mathbf{E}[X] = \sum_i p_i x_i.$$

The most useful inequality derived from the expectation of a random variable is:

Theorem 2 (Markov's Inequality) *For any $c > 1$,*

$$\mathbf{Pr}[X \geq c\mathbf{E}[X]] \leq 1/c.$$

Proof If the probability that X is greater than k is at least $1/c$, then the expectation of X is at least k/c. The theorem follows by setting $k = c\mathbf{E}[X]$. To derive this from the definition, let S be the set of i for which $x_i \geq k$. If the probability that $X \geq k$ is at least $1/c$, then

$$\sum_{i \in S} p_i \geq 1/c.$$

So,

$$\begin{aligned} \sum_i p_i x_i &\geq \sum_{i \in S} p_i x_i \\ &\geq k \sum_{i \in S} p_i \\ &\geq k/c. \end{aligned}$$

■

The most important equality involving expectation is

Lemma 3 (Linearity of Expectation) *Let X and Y be random variables. Then*

$$\mathbf{E}\left[X+Y\right]=\mathbf{E}\left[X\right]+\mathbf{E}\left[Y\right].$$

Proof Let Y take value y_i with probability q_i. Then

$$\begin{aligned}
\mathbf{E}\left[X+Y\right] &= \sum_{i,j} p_i q_j (x_i + y_j) \\
&= \sum_i p_i \sum_j q_j (x_i + y_j) \\
&= \sum_i p_i x_i \sum_j q_j + \sum_i p_i \sum_j q_j y_j \\
&= \mathbf{E}\left[X\right] + \mathbf{E}\left[Y\right],
\end{aligned}$$

where the last equality follows from the fact that $\sum_i p_i = 1$ and $\sum_j q_j = 1$. ■

The reason that linearity of expectation is so useful is that it does not depend on *any* property of X and Y. In particular, they don't have to be independent. Not even a little bit.

3 Variance

Definition 4 *The variance of X, written* $\mathbf{var}\left[X\right]$, *is*

$$\mathbf{var}\left[X\right]=\mathbf{E}\left[(X-\mathbf{E}\left[X\right])^2\right].$$

The standard deviation of X, written $\mathbf{std}\left[X\right]$, *is the square root of* $\mathbf{var}\left[X\right]$.

Using the variance, we can bound the probability that X is far from its expectation using:

Theorem 5 (Chebyshev's Inequality) *For any $\lambda > 1$,*

$$\mathbf{Pr}\left[\left|X-\mathbf{E}\left[X\right]\right| > \lambda \mathbf{std}\left[X\right]\right] \leq 1/\lambda^2.$$

Proof Follows from Markov's inequality. To see why, let

$$Y=(X-\mathbf{E}\left[X\right])^2.$$

Now,

$$\begin{aligned} Y &\geq \lambda^2 \mathbf{E}[Y] \Leftrightarrow \\ (X - \mathbf{E}[X])^2 &\geq \lambda^2 \mathbf{var}[X] \Leftrightarrow \\ |X - \mathbf{E}[X]| &\geq \lambda \mathbf{std}[X]. \end{aligned}$$

By applying Markov's inequality to Y, we find that the probability that this happens is at most $1/\lambda^2$. ■

A key property of the variance is

Lemma 6 (Additivity of Variance) *If X and Y are independent random variables, then*

$$\mathbf{var}[X + Y] = \mathbf{var}[X] + \mathbf{var}[Y].$$

Proof Left as an exercicse for the reader. ■

Definition 7 *For an event A, X is an indicator random variable for A if $X = 1$ when A is true, and $X = 0$ otherwise.*

To see if you've understood all this, consider events $A_1, \ldots, A_n$ that each occur with probability $1/2$. Let X_i be the indicator random variable for A_i. Now, prove that

$$\mathbf{Pr}\left[\left|\sum_i X_i - n/2\right| > \sqrt{n}\right] \leq 1/2.$$

Finally, here's an exercise that I gave out as a homework problem last year. I'll give out the solution in a few weeks. I'm just providing this for you to test yourself.

1. A family of functions $\mathcal{H} : A \to B$ is *pairwise-independent* if for all distinct a_1 and a_2 in A, and all b_1 and b_2 in B,

   $$\mathbf{Pr}_{h \in \mathcal{H}}[h(a_1) = b_1 \text{ and } h(a_2) = b_2] = 1/|B|^2.$$

   a. Fix m and n. Let $\mathcal{F} = GF(2)$, the field of two elements. Let $\mathcal{H}$ be the set of $h_{M,s}$ such that M is an $m \times n$ matrix over $\mathcal{F}$ and $s \in \mathcal{F}^n$, where $h_{M,s} \colon \mathcal{F}^m \to \mathcal{F}^n$ is given by $h_{M,s}(x) = xM + s$. Show that $\mathcal{H}$ is pairwise independent.

   b. Let $\mathcal{F}$ be a finite field (if you are unfamiliar general fields, just work with the integers modulo a prime). Let

   $$\mathcal{H} = \{h_{a,b} : a, b \in \mathcal{F}\}$$

   where $h_{a,b} : \mathcal{F} \mapsto \mathcal{F}$ is given by $h_{a,b} : x \mapsto ax + b$. Show that $\mathcal{H}$ is pairwise independent.

4 Chernoff bounds

When we need to show that it is exceedingly unlikely that a random variable is very far from its expectation, we use a Chernoff bound. There are many Chernoff bounds that apply in different situations. Here's one that should suffice for most purposes in class:

Theorem 8 (Chernoff Bound) *Let $p_1, \ldots, p_n$ be numbers between 0 and 1. Let*

$$p = (1/n) \sum_{i=1}^{n} p_i.$$

Let $X_1, \ldots, X_n$ be mutually independent random variables such that

$$\begin{aligned} \mathbf{Pr}\left[X_i = 1\right] &= p_i, \text{ and} \\ \mathbf{Pr}\left[X_i = 0\right] &= 1 - p_i. \end{aligned}$$

Then, for $c > 0$

$$\mathbf{Pr}\left[\sum_{i=1}^{n} X_i > (c+1)pn\right] < e^{-2c^2 n}.$$

Proof See Theorem A.4 in Alon-Spencer. ■