Lecture 2

Lecturer: Elena Grigorescu *Scribe: Jacques Pienaar*

1 Overview

In this lecture:

- Basic facts about random variables and expectations;
- Markov's inequality;
- Chebyshev's inequality;
- Chernoff/Hoeffding's inequality;
- Applications.

2 Review of probability theory needed

Let D be a finite domain, then we call a function $p : D \to [0,1]$ a probability distribution if $\sum_{x \in D} p(x) = 1$. Examples of common probability distributions are

- Uniform distribution $u(x) = 1/|D|$;
- Bernoulli distribution (corresponding to flipping a biased coin which gives heads with probability q and tails with probability $1-q$)

$$p(x) = \begin{cases} q & \text{if } x = 1 \\ 1-q & \text{if } x = 0 \end{cases}$$

- Binomial distribution (corresponding to flipping a biased coin n times)

$$\Pr[\text{get } k \text{ heads in } n \text{ flips}] = \binom{n}{k} q^k (1-q)^{n-k}.$$

Definition 1 *A* random variable *is a function* $V : D \to S \subset \mathbb{R}$.

Definition 2 (Expectation) *For a random variable V defined over domain D and distributed according to a probability distribution p, we define the expectation of V as*

$$E[V] \equiv \sum_{x \in D} p(x)V(x).$$

Definition 3 (Indicator random variable) *Given an event A, we define the indicator random variable of A as*

$$I_A := \begin{cases} 1 & \textit{if } A \textit{ is true} \\ 0 & \textit{else.} \end{cases}$$

Proposition 4 *If A is an event then $E[I_A] = \Pr[A]$,*

Proof $E[I_A] = 1 \cdot \Pr[A \text{ is true}] + 0 \cdot \Pr[A \text{ is false}]$. ■

Definition 5 (Pairwise independence) *We call two random variables, A and B, over D pairwise independent if for all $a, b \in D$, $\Pr[A = a \wedge B = b] = \Pr[A = a] \cdot \Pr[B = b]$.*

Fact 6 *(Linearity of expectation) For any two random variables A, B (not necessarily independent) over D we have $E[A + B] = E[A] + E[B]$.*

Examples:

- $E[\text{sum of 3 dice}] = 3 \cdot 7/2$
- Expected value of a binomial distribution. Let us flip n biased coins (with each coin having q probability of head, $1 - q$ the probability of landing tails). Denote with X_i the indicator variable that the i^{th} coin landed heads. Then $X = \sum_{i=1\ldots n} X_i$ is the random variable for the number of heads, and its expectation is

$$E[X] = E\left[\sum_{1\ldots n} X_i\right] = \sum_{i=1\ldots n} E[X_i]$$
$$= \sum_{i=1\ldots n} q = nq.$$

Proposition 7 *If A and B are pairwise independent then $E[AB] = E[A]E[B]$.*

Definition 8 (Conditional probability) *For two random variables A and B over the domain D the conditional probability of event A occurring given that B occurs, denoted $\Pr[A \mid B]$, is defined as follows*

$$\Pr[A \mid B] = \frac{\Pr[A \wedge B]}{\Pr[B]}.$$

Definition 9 (Conditional expectation) *For a random variable X over a domain D,*

$$E[X \mid A] = \sum_{x \in D} \Pr[X = x \mid A] \cdot x.$$

For example, the expected value of a roll of a die, given the event A that we rolled something less than 3 is

$$E[X \mid A] = \sum_{x=1,2,3} \Pr[X = x \mid A] \cdot x = 1 \cdot 1/3 + 2 \cdot 1/3 + 3 \cdot 1/3 + 0 = 2.$$

Proposition 10 (Union bound) *Given events E_1 and E_2,the probability that at least one happens is bounded by $\Pr[E_1 \cup E_2] \leq \Pr[E_1] + \Pr[E_2]$.*

For example, if we roll 3 dice, then $\Pr[\text{at least one } = 6] \leq 3 \cdot 1/6$.

Theorem 11 (Markov's inequality) *Let X be a non-zero, random variable, and $a > 0$ then*

$$\Pr[|X| \geq a] \leq \frac{E[|X|]}{a}.$$

Or equivalently $\Pr[|X| \geq aE[|X|]] \leq 1/a$.

As an example, for the toss of n fair coins let X_i denote the event that the i^{th} coin lands heads. Then $E[\sum X_i] = n/2$ so the probability that more than 2/3's of the coins come up heads is

$$\Pr[2n/3 \text{ come up heads}] \leq \frac{n/2}{2n/3} = \frac{3}{4}.$$

Theorem 12 (Chebyshev's inequality)

$$\Pr[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2},$$

where $\text{Var}[X] := E[(X - E[X])^2]$.

Proof Let random variable $Y = |X - E[X]|$. Using Markov's inequality we have that

$$\Pr[Y \geq a] = \Pr[Y^2 \geq a^2] \leq \frac{E[Y^2]}{a^2} = \frac{\text{Var}[X]}{a^2}.$$

■

Theorem 13 (Chernoff/Hoeffding inequality) *Let $X_1, \ldots, X_n$ be independent random variables in the interval $[0, 1]$ and let $X = \sum X_i$. Then*

$$\Pr[|X - E[X]| \geq t] \leq 2\exp\left(-2t^2/n\right).$$

Equivalently

$$\Pr[|X - E[X]| \geq \epsilon \cdot E[X]] \leq 2\exp\left(-2\epsilon^2 E[X]^2/n\right),$$

and

$$\Pr[|X - E[X]| \geq \epsilon n] \leq 2\exp\left(-2\epsilon^2 n\right).$$

3 Applications

In this section we shall discuss two applications of the Chernoff bound.

3.1 Approximating the fraction of 1's in a binary string

Suppose we want to estimate the fraction of 1's in a given string $S \subset \{0, 1\}^n$. That is, we wish to find a fast randomized algorithm that, given ϵ and string S outputs a value V such that $|V - \text{fraction of 1's}| < \epsilon$ with probability 2/3.

Algorithm Pick $k = 1/\epsilon^2$ uniformly random indices in the string S and output the fraction of 1's in the sample.

Analysis Let $X_1, \ldots, X_k$ be random variables indicating if a 1 was found in the string position for the i^{th} index selected ($1 \leq i \leq k$). It follows that $E[X_i]$ is equal to the fraction of 1's in S. Let X be the random variable for the number of 1's in the sample, so $X = \sum X_i$ and the indicator variable for the value output by the algorithm (i.e. for the fraction of 1's in the sample) is $\frac{X}{k}$. Then by Chernoff's bound

$$\Pr\left[\left|\frac{X}{k} - \frac{E[X]}{k}\right| > \epsilon\right] \leq 2e^{-2\epsilon^2 k} = 2e^{-2} < 1/3$$

as $\epsilon^2 k = 1$.

Therefore

$$\Pr\left[\left|\frac{X}{k} - E[X_i]\right| > \epsilon\right] < 1/3,$$

meaning that we output a good estimate (i.e. within ϵ from the true fraction of 1's in the string) with probability $> 2/3$.

3.2 Improving a random algorithm's correctness

Suppose we are given a randomized algorithm A which on each input x from some domain D outputs a 0 or 1 answer and it is correct with probability $p = 2/3$. (In other words there is some function $f : D \to \{0,1\}$ such that, for any $x \in D$ we have that $\Pr[A(x) = f(x)] \geq 2/3$, where the probability is computed over the randomness of the algorithm A.) Let algorithm B run A for t times and output the majority answer. We next show that algorithm B is correct (on each input) with probability greater than $1 - 2^{-ct}$ for some constant c (that is, $\forall x \in D$, $Pr[B(x) = f(x)] \geq 1 - 2^{-ct}$.)

Analysis Fix some input x. Let $X_1, \ldots, X_t$ be indicator variables such that $X_i = 1$ if A outputs the correct answer in the i^{th} step. Therefore, $E[X_i] = p = 2/3$. Set $X = \sum X_i$, that is X is the random variable counting the number of correct answers, and notice that $E[X] = 2t/3$.

$$\begin{aligned}
\Pr[B \text{ outputs incorrect answer}] &= \Pr[A \text{ outputs incorrect answer more than } t/2 \text{ times}] \\
&= \Pr[X < t/2] \leq \Pr[X - 2t/3 < t/2 - 2t/3] \\
&= \Pr[X - 2t/3 < -t/6] \leq \Pr[|X - 2t/3| > t/6] \\
&\leq 2e^{-2t^2(1/6)^2/t} = 2^{-ct}.
\end{aligned}$$