CS 6501: Advanced Topics in Cryptography **Spring 2019**

Probability and Statistics

Instructor: David Wu

Below is a summary of some basic facts from probability and statistics that we will use throughout this course. The presentation here is adapted from Appendix A.2 of Arora and Barak, and we refer there for additional details as well as proofs of the different claims.

Probability theory. A finite probability space[1] consists of a finite set S with a probability function $\Pr\colon S \to [0,1]$ such that $\sum_{s \in S} \Pr[s] = 1$. The probability function defines a distribution $\mathcal{D}$ over S, and we write $x \leftarrow \mathcal{D}$ to denote a draw from $\mathcal{D}$ where each element $s \in S$ is sampled with probability $\Pr[s]$. We write $\mathsf{Uniform}(S)$ to denote the *uniform distribution* over S—namely, the distribution where $\Pr[s] = 1/|S|$ for all $s \in S$. We write $x \xleftarrow{\text{R}} S$ to denote sampling an element from $\mathsf{Uniform}(S)$.

Events. An event over a probability space S is defined to be a subset $E \subseteq S$. The probability that an event E occurs is defined to be $\Pr[E] = \sum_{x \in E} \Pr[x]$. Throughout this course, we will use the following simple bound on the probability that at least one event out of a collection of events occur:

Fact 1 (Union Bound)**.** Let $E_1, \ldots, E_n \subseteq S$ be a finite collection of events over a probability space S. Then,

$$\Pr\left[\bigcup_{i \in [n]} E_i\right] \le \sum_{i \in [n]} \Pr[E_i].$$

k-Wise Independence. We say that two events E_1 and E_2 are independent if $\Pr[E_1 \cap E_2] = \Pr[E_1]\Pr[E_2]$. More generally, we say that a collection of events $E_1, \ldots, E_n$ is k-wise independent if for every subset $T \subseteq [n]$ where $|T| \le k$,

$$\Pr\left[\bigcap_{i \in T} E_i\right] = \prod_{i \in T} \Pr[E_i].$$

We say that $E_1, \ldots, E_n$ is mutually independent if it is n-wise independent.

Conditional probabilities. Given two events E_1 and E_2, we define the conditional probability of E_1 given E_2 as

$$\Pr[E_1 \mid E_2] = \frac{\Pr[E_1 \cap E_2]}{\Pr[E_2]}.$$

Fact 2 (Law of Total Probability)**.** Let $F_1, \ldots, F_n$ be a collection of pairwise disjoint events over a probability space S where $\bigcup_{i \in [n]} F_i = S$. Then, for any event E over S,

$$\Pr[E] = \sum_{i \in [n]} \Pr[E \cap F_i] = \sum_{i \in [n]} \Pr[E \mid F_i]\Pr[F_i].$$

[1] While we can also define infinite probability spaces, in this course, we will only work with finite probability spaces. Thus, in the following, we will always assume a finite probability space.

Random variables. A random variable over a probability space S is a mapping $X\colon S \to \mathbb{R}$. Given a random variable $X\colon S \to T$ that maps onto a finite set T, we can associate a probability distribution over T where $\Pr[t] = \sum_{s\in S: X(s)=t} \Pr[s]$. We refer to this as the distribution of T.

Expectation. The expected value (or expectation) of a random variable $\mathbb{E}[X]$ is defined as $\mathbb{E}[X] = \sum_{s\in S} X(s)\cdot \Pr[s]$.

Fact 3 (Linearity of Expectation)**.** Let S be a probability space and $X, Y\colon S \to \mathbb{R}$ be random variables. We write $X+Y$ to denote the random variable that implements the mapping $s \mapsto X(s)+Y(s)$. Then, $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

Fact 4 (Markov's Inequality)**.** Let $X\colon S \to \mathbb{R}$ be a non-negative random variable. Then,

$$\Pr[X \geq k\cdot \mathbb{E}[X]] \leq 1/k.$$

Fact 5 (Chernoff Bounds)**.** Let $X_1,\ldots,X_n\colon S \to \{0,1\}$ be a collection of mutually independent random variables. Let $X = \sum_{i\in[n]} X_i$ and $\mu = \mathbb{E}[X] = \sum_{i\in[n]} \mathbb{E}[X_i]$. Then for every $\delta > 0$,

$$\Pr\left[X \geq (1+\delta)\mu\right] \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu} \quad \text{and} \quad \Pr\left[X \leq (1-\delta)\mu\right] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu}.$$

In many scenarios, it will be easier to use the following special case:

Corollary 6 (Chernoff Bound)**.** *Under the same conditions as in Fact 5, for every constant $c > 0$,*

$$\Pr\left[\left|X-\mu\right| \geq c\mu\right] \leq 2^{-\Omega(\mu)}.$$

Statistical distance. Throughout this course, we will use the following notion of the statistical distance between two distributions:

Definition 7 (Statistical Distance)**.** Let $\mathcal{D}_1, \mathcal{D}_2$ be two probability distributions over a finite set S. Then, the statistical distance between $\mathcal{D}_1, \mathcal{D}_2$ is defined to be

$$\begin{aligned}\Delta(\mathcal{D}_1,\mathcal{D}_2) &= \max_{T\subseteq S} |\Pr[x \leftarrow \mathcal{D}_1 : x \in T] - \Pr[x \leftarrow \mathcal{D}_2 : x \in T]| \\ &= \frac{1}{2}\sum_{s\in S} |\Pr[x \leftarrow \mathcal{D}_1 : x = s] - \Pr[x \leftarrow \mathcal{D}_2 : x = s]|\end{aligned}$$

We say that two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ are *identical* if $\Delta(\mathcal{D}_1,\mathcal{D}_2) = 0$. We denote this by writing $\mathcal{D}_1 \equiv \mathcal{D}_2$.