Basic Definitions

- A finite probability space \((\Omega, p)\) consists of a finite set \(\Omega = \{\omega_1, \ldots, \omega_n\}\) and a probability mass function \(p: \Omega \rightarrow [0, 1]\) such that \(\sum_{\omega \in \Omega} p(\omega) = 1\). We refer to \(\Omega\) as the sample space and \(\omega_i\) as a possible outcome of a probabilistic event. Throughout this handout, we will only consider finite probability spaces.

- An event \(E\) over a probability space \((\Omega, p)\) is a set \(A \subseteq \Omega\). The probability of event \(E\), denoted \(\Pr[E]\) is defined to be \(\Pr[E] := \sum_{\omega \in E} p(\omega)\). For an outcome \(\omega \in \Omega\), we will write \(\Pr[\omega]\) to denote \(p(\omega)\).

- A random variable \(X\) over a probability space \((\Omega, p)\) is a real-valued function \(X: \Omega \rightarrow \mathbb{R}\). For the remainder of this handout, we will assume all random variables are defined over a probability space \((\Omega, p)\).

Expected Value and Variance

- The expected value \(\mathbb{E}[X]\) of a random variable \(X\) is defined to be 
  \[
  \mathbb{E}[X] := \sum_{\omega \in \Omega} X(\omega) \Pr[\omega].
  \]

- Linearity of expectation: For all random variables \(X, Y\) and all \(\alpha, \beta \in \mathbb{R}\),
  \[
  \mathbb{E}[\alpha X + \beta Y] = \alpha \mathbb{E}[X] + \beta \mathbb{E}[Y].
  \]

- The variance \(\text{Var}(X)\) of a random variable \(X\) is defined to be
  \[
  \text{Var}(X) := \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2
  \]

Useful Bounds

- Union bound: For every collection of events \(E_1, \ldots, E_n\),
  \[
  \Pr\left[ \bigcup_{i \in [n]} E_i \right] \leq \sum_{i \in [n]} \Pr[E_i].
  \]

- Markov's inequality: Let \(X\) be a non-negative random variable. For all \(t > 0\),
  \[
  \Pr[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.
  \]

- Chebyshev's inequality: Let \(X\) be a random variable. For all \(t > 0\),
  \[
  \Pr[|X - \mathbb{E}[X]| \geq t] \leq \frac{\text{Var}(X)}{t^2}.
  \]
\textbf{Chernoff bounds:} Let $X_1, \ldots, X_n$ be independent binary-valued random variables (i.e., the value of $X_i$ is either 0 or 1). Let $X = \sum_{i \in [n]} X_i$ and $\mu = \mathbb{E}[X]$. Then, for every $t > 0$,

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

Often, the following simpler (and looser) bounds suffice:

$$\forall 0 \leq t \leq 1, \quad \Pr[X \leq (1 - t)\mu] \leq e^{-\frac{t^2}{2\mu}}$$

$$\forall 0 \leq t, \quad \Pr[X \geq (1 + t)\mu] \leq e^{\frac{t^2}{2\mu}}.$$

Another useful variant (by Hoeffding) gives a bound on the sum of any sequence of bounded random variables. Specifically, let $X_1, \ldots, X_n$ be independent random variables where each $X_i \in [a_i, b_i]$ for $a_i, b_i \in \mathbb{R}$. As before let $X = \sum_{i \in [n]} X_i$ and let $\mu = \mathbb{E}[X]$. Then, for all $t > 0$,

$$\Pr\left[\left|X - \mu\right| \geq t\right] \leq 2\exp\left(-\frac{2t^2}{\sum_{i \in [n]}(b_i - a_i)^2}\right).$$

For the special case where $X_i \in [0, 1]$ for all $i \in [n]$, the bound becomes

$$\Pr\left[\left|X - \mu\right| \geq t\right] \leq 2e^{-2t^2/n}.$$

\textbf{Example 1.} Suppose $X_1, \ldots, X_N$ are independent binary-valued random variables where $\Pr[X_i = 1] = \frac{1}{2} + \epsilon$. Let $\bar{X} = \frac{1}{N} \sum_{i \in [N]} X_i$. If $N = \lambda/e^2$, then

$$\Pr[\bar{X} \geq 1/2 + \epsilon/2] \geq 1 - \text{negl}(\lambda).$$

This follows by a direct application of the Chernoff/Hoeffding bound:

$$\Pr\left[\bar{X} < \frac{1}{2} + \frac{\epsilon}{2}\right] = \Pr\left[\sum_{i \in [N]} X_i - N\left(\frac{1}{2} + \epsilon\right) < -\frac{\epsilon}{2}N\right] \leq 2e^{-\epsilon^2N^2/2N} = 2e^{-\lambda/2} = \text{negl}(\lambda).$$

\textbf{Averaging Argument}

The basic averaging argument states that if $X_1, \ldots, X_n \in \mathbb{R}$ are values with mean $\mu = \frac{1}{n} \sum_{i \in [n]} X_i$, then there exists at least one $i \in [n]$ where $X_i \geq \mu$. There are several variants of this fact that often come in handy:

\textbf{Lemma 1.} If $X_1, \ldots, X_n \in [0, 1]$ whose average is $\mu$, then at least a $\mu/2$ fraction of the $X_i$'s are at least $\mu/2$.

\textit{Proof.} Let $t$ be the fraction of $X_i$'s where $X_i \geq \mu/2$. Then, $\mu < (1 - t)\mu/2 + t < \mu/2 + t$, so $t > \mu/2$. \hfill \Box

\textbf{Lemma 2.} Let $X_1, \ldots, X_n \in [0, 1]$ whose average is $\mu = p + \epsilon$. Then, at least an $\frac{\epsilon}{2(1 - p - \epsilon/2)} > \frac{\epsilon}{2(1 - p)}$ fraction of the $X_i$'s are at least $p + \epsilon/2$.

\textit{Proof.} Let $t$ be the fraction of $X_i$'s where $X_i \geq p + \epsilon/2$. Then,

$$\mu < \left(p + \frac{\epsilon}{2}\right)(1 - t) + t = \left(p + \frac{\epsilon}{2}\right) + \left(1 - p - \frac{\epsilon}{2}\right)t.$$

Since $\mu = p + \epsilon$, this means that $\epsilon/2 < (1 - p - \epsilon/2)t$ so $t > \frac{\epsilon}{2(1 - p - \epsilon/2)} > \frac{\epsilon}{2(1 - p)}$. \hfill \Box
Example 2. Let $f$ be a function. Suppose we have an algorithm $\mathcal{A}$ where

$$\Pr[x \leftarrow \{0, 1\}^n, y \leftarrow \{0, 1\}^n : \mathcal{A}(x, y) = f(x)] = \frac{1}{2} + \epsilon.$$ 

We say that a string $y^* \in \{0, 1\}^n$ is “good” if

$$\Pr[x \leftarrow \{0, 1\}^n : \mathcal{A}(x, y^*) = f(x)] = \frac{1}{2} + \frac{\epsilon}{2}. \quad (1)$$

By an averaging argument (Lemma 2), at least an $\epsilon$-fraction of $y$’s are good (i.e., sampling a random $y$ will satisfy Eq. (1) with probability $\epsilon$).