Tools from Probability Theory

1. Let $X$ be a real-valued random variable with finite support $\{x_1, \ldots, x_n\}$.

   \[ E[X] = \sum_{i=1}^{n} x_i \cdot Pr[X=x_i] \]  
   (sum is replaced with integral if $X$ is continuous)

2. **Fact 1**: Expectation is linear.
   - For all $c \in \mathbb{R}$, $E[cX] = cE[X]$.

Variance:

3. $\sigma^2(X) := E[(X-E[X])^2]$.

Fact 2: Let $\mu := E[X]$. Then $\sigma^2(X) = E[X^2] - \mu^2$.

Proof:

\[
\begin{align*}
\sigma^2(X) & = E[(X-\mu)^2] = E[(X^2 - 2X\mu + \mu^2)] \\
& = E[X^2] - 2\mu E[X] + \mu^2 \\
& = E[X^2] - 2\mu^2 + \mu^2 \\
& = E[X^2] - \mu^2.
\end{align*}
\]

Fact 3: $\sigma^2(cX) = c^2 \sigma^2(X)$.

Two random variables $X, Y$ are independent if for every $x, y \in \mathbb{R}$,

\[ Pr[X=x, Y=y] = Pr[X=x] \cdot Pr[Y=y]. \]

Variables $X_1, X_2, \ldots, X_n$ are mutually independent if for every $x_1, x_2, \ldots, x_n \in \mathbb{R}$,

\[ Pr[X_1=x_1, \ldots, X_n=x_n] = \prod_{i=1}^{n} Pr[X_i=x_i]. \]
Variables $X_1, \ldots, X_n$ are pairwise independent if for all $i \neq j$ and all $x_i, x_j \in \mathbb{R}$, $\Pr[X_i = x_i, X_j = x_j] = \Pr[X_i = x_i] \cdot \Pr[X_j = x_j]$

Independence and linearity of variance

**Fact 4**: If $X_1, \ldots, X_n$ are pairwise independent, then

$$\sigma^2(\sum_{i} X_i) = \sum_{i} \sigma^2(X_i),$$

**Proof sketch**: I will just cover the case that $\mathbb{E}[X_i] = 0$ for all $i$.

Note that the assumption means $\mathbb{E}[(\sum_{i} X_i)^2] = \left(\sum_{i} \mathbb{E}[X_i]\right)^2 = 0^2 = 0$.

So $\sigma^2(\sum_{i} X_i) \overset{\text{def}}{=} \mathbb{E}[(\sum_{i} X_i)^2] - (\mathbb{E}[\sum_{i} X_i])^2 = \mathbb{E}[(\sum_{i} X_i)^2]$

$$= \mathbb{E}[(\sum_{1 \leq i \leq n} X_i^2)] \overset{\text{linearity of expectation}}{=} \sum_{1 \leq i \leq n} \mathbb{E}[X_i^2]$$

$$= \sum_{1 \leq i \leq n} \mathbb{E}[X_i] \cdot \mathbb{E}[X_i] = \sum_{i} \sigma^2(X_i) \quad \square$$

**Variance**

- Averaging pairwise independent random variables reduces variance.

**Proof**: If $X_1, \ldots, X_n$ are pairwise independent and all have the same mean $\mu$ and variance $\sigma^2$, then $\frac{1}{n} \sum_{i} X_i$ also has mean $\mu$ (by linearity of expectation) and variance $\frac{1}{n^2} \sigma^2(\sum_{i} X_i) \overset{\text{fact 4}}{=} (\frac{1}{n})(n \cdot \sigma^2) = \frac{\sigma^2}{n}$.

- **by Fact 3**

we'll see why variance reduction beneficial (Chebyshev's Inequality)
Def: An event is a set of possible outcomes of an experiment. E.g., in the experiment of drawing a single card from a 52-card deck, example events are "draw the jack of clubs", "draw a heart", "draw a red card", etc.

Def: For two events $E_1$ and $E_2$,

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

(undefined if $\Pr[E_1] = 0$)

Fact 5: If two random variables $X_1, X_2$ are independent, then for all $x_1, x_2 \in \mathbb{R}$,

$$\Pr[X_2 = x_2 \mid X_1 = x_1] = \Pr[X_2 = x_2].$$

Techniques for upper bounding the probability of a "bad event" occurring (typically our randomized algorithms will return the right answer in expectation, and we want to bound the probability the answer is far from its expectation).

Union-Bound: For any events $E_1, E_2, \ldots, E_k$,

$$\Pr[E_1 \cup E_2 \cup \cdots \cup E_k] \leq \Pr[E_1] + \cdots + \Pr[E_k].$$

Inclusion-Exclusion principle:

$$\sum_{i=1}^{k} \Pr[E_i] - \sum_{1 \leq i < j \leq k} \Pr[E_i \cap E_j] + \cdots + (-1)^{k+1} \Pr[E_1 \cap E_2 \cap \cdots \cap E_k]$$
Markov's inequality. Let $\mu = \mathbb{E}[X]$. If $X$ is a non-negative random variable, then for every $t > 0$, $\Pr[X \geq t\mu] \leq \frac{1}{t}$.

Proof: $\mathbb{E}[X] = \sum_i \Pr[X = i] \cdot i$

$X$ non-negative

$\geq \sum_{i \geq t\mu} \Pr[X = i] \cdot i$

$\geq t\mu \sum_{i \geq t\mu} \Pr[X = i]$

$= t\mu \cdot \Pr[X \geq t\mu]$

So cancel the $i$'s and write but $\mu$ sides by $t$. 

Bound can be weak and doesn't work for bounding probability $X$ is far below its expectation.
Chebyshev's Inequality: Recall \( \mu = \text{E}(X) \) and \( \sigma^2 \) is variance. \( \sigma = \sqrt{\sigma^2} \) is the standard deviation of \( X \).

For every \( \epsilon > 0 \), \( \Pr \left[ |X - \mu| \geq \epsilon \sigma \right] \leq \frac{1}{\epsilon^2} \)

Proof: Apply Markov to \( |X - \mu| \). That is,

\[
\Pr \left[ |X - \mu| \geq \epsilon \sigma \right] \leq \frac{\Pr \left[ |X - \mu|^2 \geq \epsilon^2 \sigma^2 \right]}{\epsilon^2} \quad \text{(Square both sides of the inequality)}
\]

Definition of variance:

\[
\Pr \left[ |X - \mu|^2 \geq \epsilon^2 \sigma^2 \right] = \Pr \left[ \sum \left( X_i - \mu \right)^2 \geq \epsilon^2 \sigma^2 \right] = \frac{1}{\epsilon^2}
\]

Chernoff Bound: Let \( X_1, \ldots, X_n \) be i.i.d. random variables with range \([0, 1]\) and \( \text{E}(X) = \mu \).

Then \( \sum \frac{X_i}{n} \leq \mu \) and \( 0 < \delta \leq \mu \), and \( n \mu = \sum X_i = np \),

\[
\Pr \left[ |\sum X_i - n \mu| \geq \delta \mu \right] \leq 2 \exp \left( -\frac{n \delta^2}{3} \right)
\]

Example: Say you toss a fair coin \( n \) times. What is the probability more than \( \frac{n}{2} + \sqrt{n \log n} \) of them are heads?

Let \( X_i = \left\{ \begin{array}{ll} 1 & \text{if \( \text{heads} \)} \\ 0 & \text{if \( \text{tails} \)} \end{array} \right. \). Then \( \mu = \frac{n}{2} \) and \( \sum_{i=1}^{n} X_i = \# \text{ of heads} \)

\[
\Pr \left[ |X - \mu| \geq \sqrt{n \log n} \right] \leq \Pr \left[ \left| X - \mu \right| \geq \sqrt{n \log n} \right] \leq \frac{Q \delta^2}{4} \quad \text{for} \quad \delta = \frac{\sqrt{n \log n}}{n}
\]

\[
\Pr \left[ X = \frac{n}{2} + \frac{\sqrt{n \log n}}{n} \right] \leq 2 \exp \left( -\frac{n}{2} \cdot \left( \frac{4 \log n}{n} \right) \right)
\]

\[
2 \cdot \sqrt{n \log n} = 2 \exp \left( -\frac{n}{2} \cdot \frac{3}{5} \log n \right) = \frac{2}{n^{3/5}}.
\]