Variance Refresher

- **Expectation:** \( E[X] = \sum_r r P[X = r] \)
- **Linearity of expectation:** \( E[X + Y] = E[X] + E[Y] \)
- **Variance random variable:** \( \sigma^2_X = E[(X - E[X])^2] \)
- **Linearity of variance if \( X \) and \( Y \) are independent:**
  \[
  \sigma^2_{X + Y} = \sigma^2_X + \sigma^2_Y 
  \]
Examples of Random Variables

Example
Let $X$ have the binomial distribution $Bin(n, p)$:

$$P[X = i] = \binom{n}{i} p^i (1 - p)^{n-i}$$

“How many heads do we see when we toss a coin with probability $p$ of heads $n$ times?” $E[X] = np$ and $V[X] = np(1 - p)$.

Example
Let $X$ have the geometric distribution $Geom(p)$:

$$P[X = i] = (1 - p)^{i-1} p$$

“How many times do we toss a coin with probability $p$ of heads until we see a heads.” $E[X] = 1/p$, $V[X] = (1 - p)/p^2$. 
Outline

Markov and Chebyshev

Lazy Select
Markov Inequality

**Theorem (Markov)**

Let $Y$ be a positive random variable and let $\mu = \mathbb{E}[Y]$. Then, for $t > 0$,

$$
\mathbb{P}[Y \geq t\mu] \leq \frac{1}{t}.
$$

**Proof.**

1. $\mathbb{E}[Y] = \sum_r r \cdot \mathbb{P}[Y = r] \geq \sum_{r \geq t\mu} r \cdot \mathbb{P}[Y = r] \geq \mathbb{P}[Y \geq t\mu] \cdot t \cdot \mu$

2. Therefore, $\mathbb{P}[Y \geq t\mu] \leq \frac{1}{t}$. 

\[\square\]
Chebyshev Inequality

Theorem (Chebyshev)

Let $X$ be a random variable with expectation $\mu$ and variance $\sigma^2$. Then for $t > 0$,

$$\mathbb{P}[|X - \mu| \geq t\sigma] \leq \frac{1}{t^2}.$$ 

Proof.

- Note that $\mathbb{P}[|X - \mu| \geq t\sigma] = \mathbb{P}[(X - \mu)^2 \geq t^2\sigma^2]$ 
- Let $Y = (X - \mu)^2$ and note $\mathbb{E}[Y] = \sigma^2$ 
- Use Markov’s inequality to show $\mathbb{P}[Y \geq t^2\mathbb{E}[Y]] \leq 1/t^2$
Theorem
Let $X_1, \ldots, X_n$ be independent boolean random variables such that $P[X_i = 1] = p_i$. Then, for $X = \sum_i X_i$, $\mu = E[X]$, and $\delta > 0$,

$$P[X > (1 + \delta)\mu] < \left[\frac{e^{\delta}}{(1 + \delta)^{1+\delta}}\right]^\mu$$

Other versions: For $0 < \delta \leq 1$

$$P[X \geq (1 + \delta)\mu] \leq e^{-\delta^2 \mu/3}$$

$$P[X \leq (1 - \delta)\mu] \leq e^{-\delta^2 \mu/2}$$

Will prove Chernoff Bound next time...
Outline

Markov and Chebyshev

Lazy Select
Lazy Select

Let $S$ be set of $n = 2k$ distinct values. Want to find $k$-th smallest value.

Algorithm

1. Add each element in $S$ to a set $R$ with probability $p = 1/n^{1/4}$.
2. Call this set $R$, Sort $R$ and let

   $$a = \left(\frac{n^{3/4}}{2} - 5\sqrt{n}\right) \text{ smallest element in } R.$$ 

   $$b = \left(\frac{n^{3/4}}{2} + 5\sqrt{n}\right) \text{ smallest element in } R.$$ 

3. Construct $S' = \{i \in S : a < y < b\}$ and let $t$ be the number of values less or equal to $a$ amongst $S$.
4. Sort $S'$ and return $(k - t)$th smallest value in $S'$. 