LEcTUrE 9

Last time
• Variance, covariance
• Chebyshev’s inequality
• Variance of Binomial and Geometric RVs

In Discussion
• Median of a RV

Today
• Computing the median of an array
Recall: Chebyshev’s Inequality

- **Theorem.** For a random variable $X$ and $a > 0$,
  \[
  \Pr[|X - \mathbb{E}[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}.
  \]

- **Alternatively:** Then, for all $t > 1$,
  \[
  \Pr[|X - \mathbb{E}[X]| \geq t \cdot \sigma[X]] \leq \frac{1}{t^2}.
  \]
Chebyshev’s Inequality: Example

$X \sim \text{Bin}(n, 1/2)$. Bound $\Pr\left[X > \frac{3n}{4}\right]$ using Markov and Chebyshev.

Solution: $\mathbb{E}[X] = \quad ; \ \text{Var}[X] = \quad$

Markov: $\Pr\left[X > \frac{3n}{4}\right] \leq \quad$

Chebyshev: $\Pr[|X - \mathbb{E}[X]| \geq \quad ] \leq \quad$

$\Pr\left[X > \frac{3n}{4}\right] = \quad$

For $n > 3$, Chebyshev is much better!
Median of a random variable

• A value \( m \) is the median of a random variable \( X \) if \( \Pr[X \leq m] \geq 1/2 \) and \( \Pr[X \geq m] \geq 1/2 \).

• **Example 1:** \( X \) is uniform over \( x_1, \ldots, x_{2k+1} \), where \( x_1 < \cdots < x_{2k+1} \). What is the median?

• **Example 2:** \( X \) is uniform over \( x_1, \ldots, x_{2k} \), where \( x_1 < \cdots < x_{2k} \). Find all medians.
Theorem. For a random variable $X$ with a finite expectation $\mu$ and a finite median $m$,

1. the expectation $\mu$ is the value of $c$ that minimizes the expression 
   $$E[(X - c)^2];$$

2. the median $m$ is a value of $c$ that minimizes the expression 
   $$E[|X - c|].$$
**Theorem.** For a random variable $X$ with expectation $\mu$, median $m$, and standard deviation $\sigma$, 
$$|\mu - m| \leq \sigma.$$
Randomized Algorithm for the Median of an Array

- Given elements $a_1 \leq a_2 \leq \cdots \leq a_n$, their median is $a_{\lfloor n/2 \rfloor}$.

Task: Find the median of an array if the elements are not sorted.

- Deterministic algorithm (median of medians): $O(n)$ time.
- Today: simple randomized algorithm: $O(n)$ time.

- Simplifying assumptions:
  - all elements are distinct;
  - $n$ is odd;
  - we can sample from the array in constant time.

- Idea: Sample to find elements $\ell$ and $u$ such that
  1. $\ell \leq m \leq u$, where $m$ denotes the median.
  2. The number of input elements that lie in the interval $(\ell, u)$ is small.
Randomized Median Algorithm

Input: array $A$ of elements $a_1, \ldots, a_n$

Output: median of $A$

1. Let $R$ be an array $r_1, \ldots, r_t$, where each $r_i$ is chosen from $A$ u.i.r. with replacement, where $t = \lceil n^{3/4} \rceil$.

2. Sort $R$.

3. Let $\ell$ be the $\left\lfloor \frac{n^{3/4}}{2} - \sqrt{n} \right\rfloor$-th smallest element in $R$.

4. Let $u$ be the $\left\lceil \frac{n^{3/4}}{2} + \sqrt{n} \right\rceil$-th smallest element in $R$.

5. Use PARTITION from Quicksort to compute $C = \{ a \in A \mid \ell \leq a \leq u \}$,

   $n_\ell = |\{ a \in A \mid a < \ell \}|$ and $n_u = |\{ a \in A \mid a > u \}|$

6. If $n_\ell > \left\lfloor \frac{n}{2} \right\rfloor$ or $n_u > \left\lceil \frac{n}{2} \right\rceil$ then fail.

7. If $|C| \leq 4n^{3/4}$ then sort $C$; otherwise fail.

8. Output the $\left(\left\lfloor \frac{n}{2} \right\rfloor - n_\ell + 1 \right)$-th smallest element in $C$.

View of sorted $A$: 

$R$

$C$
Randomized Median Algorithm (RMA) terminates in $O(n)$ time. It outputs either `fail` or the median.

RMA outputs `fail` with probability at most $n^{-1/4}$. 
Monte Carlo vs. Las Vegas

- **Monte Carlo**: a randomized algorithm that may fail or produce an incorrect answer.
- **Las Vegas**: a randomized algorithm that always returns the right answer.

- We can get a Las Vegas algorithm from a Monte Carlo algorithm that may fail by repeating until it succeeds.