Summary

Last Class:

• Course logistics/overview of planned content.
• Intro to randomized algorithms: Las Vegas vs. Monte Carlo
• Randomized complexity classes including RP, ZPP, BPP, PP.

This Class:

• Basic probability review with algorithmic applications.
  • Conditional probability, Baye's theorem, and independence.
  • Application to polynomial identity testing.
  • Linearity of expectation and variance. Application to randomized Quicksort analysis.
  • Maybe start on concentration inequalities (Markov's and Chebyshev's).
Summary

Last Class:

• Course logistics/overview of planned content.
• Intro to randomized algorithms: Las Vegas vs. Monte Carlo
• Randomized complexity classes including RP, ZPP, BPP, PP.

This Class: Basic probability review with algorithmic applications.

• Conditional probability, Baye’s theorem, and independence. Application to polynomial identity testing.
• Linearity of expectation and variance. Application to randomized Quicksort analysis.
• Maybe start on concentration inequalities (Markov’s and Chebyshev’s).
Basic Probability Review
Consider two random events $A$ and $B$.

- **Conditional Probability:**
  \[
  P(A|B) = \frac{P(A \cap B)}{P(B)}.
  \]

- **Bayes’s Theorem:**
  \[
  P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A)}.
  \]

- **Independence:** $A$ and $B$ are independent if:
  \[
  P(A|B) = P(A).
  \]

Using the definition of conditional probability, independence means:

\[
P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A) \implies P(A \cap B) = P(A) \cdot P(B).
\]
Sets of events: For a set of $n$ events, $A_1, \ldots, A_n$, the events are $k$-wise independent if for any subset $S$ of at most $k$ events,

\[ \Pr \left( \bigcap_{i \in S} A_i \right) = \prod_{i \in S} \Pr(A_i). \]

For $k = n$ we just say the events ‘are independent’.

\[ D_1 = 1 \]
\[ D_2 = 1 \]
\[ D_3 = 1 \]

\[ P(A) = \frac{1}{6} \]
\[ P(B) = \frac{1}{6} \]

\[ P(A \cap B) = \frac{1}{36} = P(A) \cdot P(B) \]

\[ P(A \cap B \cap C) = \frac{1}{36} \neq P(A) \cdot P(B) \cdot P(C) \]
Sets of events: For a set of \( n \) events, \( A_1, \ldots, A_n \), the events are \( k \)-wise independent if for any subset \( S \) of at most \( k \) events,

\[
\Pr \left( \bigcap_{i \in S} A_i \right) = \prod_{i \in S} \Pr(A_i).
\]

For \( k = n \) we just say the events ‘are independent’.

Random Variables: Two random variables \( X, Y \) are independent if for all \( s, t \), \( X = s \) and \( Y = t \) are independent events. In other words:

\[
\Pr(X = s \cap Y = t) = \Pr(X = s) \cdot \Pr(Y = t).
\]
Application 1: Polynomial Identity Testing
Polynomial Identity Testing

Given an $n$-variable degree-$d$ polynomial $p(x_1, x_2, \ldots, x_n)$, determine if the polynomial is identically zero. I.e., if $p(x_1, x_2, \ldots, x_n) = 0$ for all $x_1, \ldots, x_n$. 

E.g., you are given:

\[ p(x_1, x_2, \ldots, x_3) = x_3(x_1 - x_2)^3 + (x_1 + 2x_2 - x_3)^2 - x_1(x_2 + x_3)^2. \]

• Can expand out all the terms and check if they cancel. But the number of terms can be as large as $\binom{n+d}{d}$ – i.e., exponential in the number of variables $n$ and the degree $d$.

Extremely Simple Randomized Algorithm:
Just pick random values for $x_1, \ldots, x_n$ and evaluate the polynomial at these values. With high probability, if $p(x_1, x_2, \ldots, x_n) = 0$, the polynomial is identically 0!
Given an $n$-variable degree-$d$ polynomial $p(x_1, x_2, \ldots, x_n)$, determine if the polynomial is identically zero. I.e., if $p(x_1, x_2, \ldots, x_n) = 0$ for all $x_1, \ldots, x_n$. E.g., you are given:

$$p(x_1, x_2, \ldots, x_3) = x_3(x_1 - x_2)^3 + (x_1 + 2x_2 - x_3)^2 - x_1(x_2 + x_3)^2.$$ 

\[\begin{align*}
x_3x_1^2 & \cdot x_3 \cdot 2x_1x_2 + \ldots = 0
\end{align*}\]
Given an \( n \)-variable degree-\( d \) polynomial \( p(x_1, x_2, \ldots, x_n) \), determine if the polynomial is identically zero. I.e., if \( p(x_1, x_2, \ldots, x_n) = 0 \) for all \( x_1, \ldots, x_n \). E.g., you are given:

\[
p(x_1, x_2, \ldots, x_3) = x_3(x_1 - x_2)^3 + (x_1 + 2x_2 - x_3)^2 - x_1(x_2 + x_3)^2.
\]

- Can expand out all the terms and check if they cancel. But the number of terms can be as large as \( \binom{n+d}{d} \) – i.e., exponential in the number of variables \( n \) and the degree \( d \).
Polynomial Identity Testing

Given an $n$-variable degree-$d$ polynomial $p(x_1, x_2, \ldots, x_n)$, determine if the polynomial is identically zero. I.e., if $p(x_1, x_2, \ldots, x_n) = 0$ for all $x_1, \ldots, x_n$. E.g., you are given:

$$p(x_1, x_2, \ldots, x_3) = x_3(x_1 - x_2)^3 + (x_1 + 2x_2 - x_3)^2 - x_1(x_2 + x_3)^2.$$ 

• Can expand out all the terms and check if they cancel. But the number of terms can be as large as $\binom{n+d}{d}$ – i.e., exponential in the number of variables $n$ and the degree $d$.

Extremely Simple Randomized Algorithm: Just pick random values for $x_1, \ldots, x_n$ and evaluate the polynomial at these values. With high probability, if $p(x_1, \ldots, x_n) = 0$, the polynomial is identically 0!

$$p(5, 2, \ldots, -1) = -1(5 - 2)^3 + (5 + 2 \cdot 2 + 1)^2 - 5(2 - 1)^2 = 68.$$
Polynomial Identity Testing

Given an $n$-variable degree-$d$ polynomial $p(x_1, x_2, \ldots, x_n)$, determine if the polynomial is identically zero. I.e., if $p(x_1, x_2, \ldots, x_n) = 0$ for all $x_1, \ldots, x_n$. E.g., you are given:

$$p(x_1, x_2, \ldots, x_3) = x_3(x_1 - x_2)^3 + (x_1 + 2x_2 - x_3)^2 - x_1(x_2 + x_3)^2.$$

• Can expand out all the terms and check if they cancel. But the number of terms can be as large as ${n+d \choose d}$ - i.e., exponential in the number of variables $n$ and the degree $d$.

Extremely Simple Randomized Algorithm: Just pick random values for $x_1, \ldots, x_n$ and evaluate the polynomial at these values. With high probability, if $p(x_1, \ldots, x_n) = 0$, the polynomial is identically 0!

$$p(5, 2, \ldots, -1) = -1(5 - 2)^3 + (5 + 2 \cdot 2 + 1)^2 - 5(2 - 1)^2 = 68.$$
Polynomial Identity Testing Proof

**Schwartz-Zippel Lemma:** For any $n$-variable degree-$d$ polynomial $p(x_1, \ldots, x_n)$ and any set $S$, if $z_1, \ldots, z_n$ are selected independently and uniformly at random from $S$, then $\Pr[p(z_1, \ldots, z_n) \neq 0] \geq 1 - \frac{d}{|S|}$.

\[
\begin{align*}
\text{Let } k &= \text{max degree of } x_1 \\
q &= x_1^k \cdot \text{non-zero } (n-1) \text{-variable } d-k \text{-degree polynomial} \\
\Pr[q(z_2, \ldots, z_n) \neq 0] &\geq 1 - \frac{d-k}{|S|} \\
\text{If } q(z_2, \ldots, z_n) \neq 0, \text{ then } p(x_1, z_2, \ldots, z_n) \text{ is a degree } k \text{ non-zero univariate polynomial in } x_1.
\end{align*}
\]
Schwartz-Zippel Lemma: For any \( n \)-variable degree-\( d \) polynomial \( p(x_1, \ldots, x_n) \) and any set \( S \), if \( z_1, \ldots, z_n \) are selected independently and uniformly at random from \( S \), then \( \Pr[p(z_1, \ldots, z_n) \neq 0] \geq 1 - \frac{d}{|S|} \).

Proof: Via induction on the number of variables \( n \)

Base Case \( n = 1 \):

\[
p(x_1) = x_1^2 + 3x_3 + 4x_4.\]

\( p \) has at most \( d \) roots.

\( p(x) = 0 \) for at most \( d \) values of \( x \).

In worst case, all roots are in \( S \).

Probability that \( z_1 \) is a root so \( p(z_1) = 0 \) is \( \leq \frac{d}{|S|} \).

\[
\Pr(p(z_1) = 0) \leq 1 - \frac{d}{|S|}.
\]
Schwartz-Zippel Lemma: For any \( n \)-variable degree-\( d \) polynomial \( p(x_1, \ldots, x_n) \) and any set \( S \), if \( z_1, \ldots, z_n \) are selected independently and uniformly at random from \( S \), then \( \Pr[p(z_1, \ldots, z_n) \neq 0] \geq 1 - \frac{d}{|S|} \).

Proof: Via induction on the number of variables \( n \)

Induction Step \( n > 1 \):
Schwartz-Zippel Lemma: For any $n$-variable degree-$d$ polynomial $p(x_1, \ldots, x_n)$ and any set $S$, if $z_1, \ldots, z_n$ are selected independently and uniformly at random from $S$, then $\Pr[p(z_1, \ldots, z_n) \neq 0] \geq 1 - \frac{d}{|S|}$.

Proof: Via induction on the number of variables $n$.

Induction Step $n > 1$:

- Let $k$ be the max degree of $x_1$ in $p(\cdots)$. Assume w.l.o.g. that $k > 0$. Write $p(x_1, \ldots, x_n) = x_1^k \cdot q(x_2, \ldots, x_n) + r(x_1, \ldots, x_n)$. E.g.,

$$x_1^2 x_2 + x_1^2 x_3 + x_1 x_2 x_3 + x_2 x_3 = x_1^2 \cdot (x_2 + x_3) + x_1 x_2 x_3 + x_2 x_3.$$
Schwartz-Zippel Lemma: For any $n$-variable degree-$d$ polynomial $p(x_1, \ldots, x_n)$ and any set $S$, if $z_1, \ldots, z_n$ are selected independently and uniformly at random from $S$, then $\Pr[p(z_1, \ldots, z_n) \neq 0] \geq 1 - \frac{d}{|S|}$.

Proof: Via induction on the number of variables $n$

Induction Step $n > 1$:

• Let $k$ be the max degree of $x_1$ in $p(\cdots)$. Assume w.l.o.g. that $k > 0$. Write $p(x_1, \ldots, x_n) = x_1^k \cdot q(x_2, \ldots, x_n) + r(x_1, \ldots, x_n)$. E.g.,
  \[x_1^2x_2 + x_1^2x_3 + x_1x_2x_3 + x_2x_3 = x_1^2 \cdot (x_2 + x_3) + x_1x_2x_3 + x_2x_3.\]

• Observe: $q(\cdot)$ is non-zero, with $n - 1$ variables and degree $d - k$. 
Polynomial Identity Testing Proof

Schwartz-Zippel Lemma: For any \( n \)-variable degree-\( d \) polynomial \( p(x_1, \ldots, x_n) \) and any set \( S \), if \( z_1, \ldots, z_n \) are selected independently and uniformly at random from \( S \), then \( \Pr[p(z_1, \ldots, z_n) \neq 0] \geq 1 - \frac{d}{|S|} \).

Proof: Via induction on the number of variables \( n \)

Induction Step \( n > 1 \):

- Let \( k \) be the max degree of \( x_1 \) in \( p(\cdots) \). Assume w.l.o.g. that \( k > 0 \). Write \( p(x_1, \ldots, x_n) = x_1^k \cdot q(x_2, \ldots, x_n) + r(x_1, \ldots, x_n) \). E.g.,
  \[
  x^2_1x_2 + x_1^2x_3 + x_1x_2x_3 + x_2x_3 = x_1^2 \cdot (x_2 + x_3) + x_1x_2x_3 + x_2x_3.
  \]
- Observe: \( q(\cdot) \) is non-zero, with \( n - 1 \) variables and degree \( d - k \).
- So, by inductive assumption, \( \Pr[q(z_2, \ldots, z_n) \neq 0] \geq 1 - \frac{d-k}{|S|} \).
Schwartz-Zippel Lemma: For any $n$-variable degree-$d$ polynomial $p(x_1, \ldots, x_n)$ and any set $S$, if $z_1, \ldots, z_n$ are selected independently and uniformly at random from $S$, then $\Pr[p(z_1, \ldots, z_n) \neq 0] \geq 1 - \frac{d}{|S|}$.

Proof: Via induction on the number of variables $n$

Induction Step $n > 1$:

- Let $k$ be the max degree of $x_1$ in $p(\cdots)$. Assume w.l.o.g. that $k > 0$. Write $p(x_1, \ldots, x_n) = x_1^k \cdot q(x_2, \ldots, x_n) + r(x_1, \ldots, x_n)$. E.g.,
  
  \[
  x_1^2x_2 + x_1^2x_3 + x_1x_2x_3 + x_2x_3 = x_1^2 \cdot (x_2 + x_3) + x_1x_2x_3 + x_2x_3.
  \]

- Observe: $q(\cdot)$ is non-zero, with $n - 1$ variables and degree $d - k$.
- So, by inductive assumption, $\Pr[q(z_2, \ldots, z_n) \neq 0] \geq 1 - \frac{d-k}{|S|}$.
- Assuming $q(z_2, \ldots, z_n) \neq 0$, then $p(x_1, z_2, \ldots, z_n)$ is a degree $k$ non-zero univariate polynomial in $x_1$. 
Polynomial Identity Testing Proof

Assuming \( q(z_2, \ldots, z_n) \neq 0 \), then \( p(x_1, z_2, \ldots, z_n) \) is a degree \( k \) non-zero univariate polynomial in \( x_1 \).

Example:

\[
p(x_1, x_2, x_3) = x_1^2 x_2 + x_1^2 x_3 + x_1 x_2 x_3 + x_2 x_3 = x_1^2 \underbrace{(x_2 + x_3)}_{q(\cdots)} + x_1 x_2 x_3 + x_2 x_3.
\]

\[
p(x_1, z_2, z_3) = p(x_1, 2, 1) = x_1^2 \cdot 3 + 2x_1 + 2.
\]
Assuming $q(z_2, \ldots, z_n) \neq 0$, then $p(x_1, z_2, \ldots, z_n)$ is a degree $k$ non-zero univariate polynomial in $x_1$.

Example:

$$p(x_1, x_2, x_3) = x_1^2 x_2 + x_1^2 x_3 + x_1 x_2 x_3 + x_2 x_3 = x_1^2 \cdot (x_2 + x_3) + x_1 x_2 x_3 + x_2 x_3.$$  

$$p(x_1, z_2, z_3) = p(x_1, 2, 1) = x_1^2 \cdot 3 + 2x_1 + 2.$$

Next Step: Again applying the inductive hypothesis,

$$\Pr[p(z_1, \ldots z_n) \neq 0 | q(z_2, \ldots, z_n) \neq 0] \geq 1 - \frac{k}{|S|}.$$
Polynomial Identity Testing Proof

Assuming \( q(z_2, \ldots, z_n) \neq 0 \), then \( p(x_1, z_2, \ldots, z_n) \) is a degree \( k \) non-zero univariate polynomial in \( x_1 \).

Example:

\[
p(x_1, x_2, x_3) = x_1^2x_2 + x_1^2x_3 + x_1x_2x_3 + x_2x_3 = x_1^2 \cdot (x_2 + x_3) + \underbrace{x_1x_2x_3 + x_2x_3}_{q(\cdot)} + \underbrace{\cdot}_{r(\cdot)}.
\]

\[
p(x_1, z_2, z_3) = p(x_1, 2, 1) = x_1^2 \cdot 3 + 2x_1 + 2.
\]

Next Step: Again applying the inductive hypothesis,

\[
\Pr[p(z_1, \ldots z_n) \neq 0 | q(z_2, \ldots, z_n) \neq 0] \geq 1 - \frac{k}{|S|}.
\]

Overall:

\[
\Pr[p(z_1, \ldots z_n) \neq 0] \geq \Pr[p(z_1, \ldots z_n) \neq 0 \cap q(z_2, \ldots, z_n) \neq 0] = \Pr[p(\cdot) \neq 0 | q(\cdot) \neq 0] \cdot \Pr[q(\cdot) \neq 0] \geq \left(1 - \frac{k}{|S|}\right) \cdot \left(1 - \frac{d - k}{|S|}\right) \geq 1 - \frac{d}{|S|}.
\]

This completes the proof of Schwartz-Zippel.
Polynomial Identity Testing Proof

Assuming \( q(z_2, \ldots, z_n) \neq 0 \), then \( p(x_1, z_2, \ldots, z_n) \) is a degree \( k \) non-zero univariate polynomial in \( x_1 \).

Example:

\[
p(x_1, x_2, x_3) = x_1^2 x_2 + x_1^2 x_3 + x_1 x_2 x_3 + x_2 x_3 = x_1^2 \cdot (x_2 + x_3) + x_1 x_2 x_3 + x_2 x_3.
\]

\[
p(x_1, z_2, z_3) = p(x_1, 2, 1) = x_1^2 \cdot 3 + 2x_1 + 2.
\]

Next Step: Again applying the inductive hypothesis,

\[
\Pr[p(z_1, \ldots z_n) \neq 0 \mid q(z_2, \ldots, z_n) \neq 0] \geq 1 - \frac{k}{|S|}.
\]

Overall:

\[
\Pr[p(z_1, \ldots z_n) \neq 0] \geq \Pr[p(z_1, \ldots z_n) \neq 0 \cap q(z_2, \ldots, z_n) \neq 0]
\]

\[
\frac{d}{|S|} = 1 \quad \frac{d}{|S|} = 0
\]

\[
1 - \frac{d}{|S|} \geq \left(1 - \frac{k}{|S|}\right) \cdot \left(1 - \frac{d - k}{|S|}\right) \geq 1 - \frac{d}{|S|}.
\]

This completes the proof of Schwartz-Zippel.
Expectation and Variance Review
Expectation and Variance

Consider a random $X$ variable taking values in some finite set $S \subset \mathbb{R}$. E.g., for a random dice roll, $S = \{1, 2, 3, 4, 5, 6\}$. 

- **Expectation:** $\mathbb{E}[X] = \sum_{s \in S} \Pr(X = s) \cdot s$.  
- **Variance:** $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$. 

**Exercise:** Verify that for any scalar $\alpha$, $\mathbb{E}[\alpha \cdot X] = \alpha \cdot \mathbb{E}[X]$ and $\text{Var}[\alpha \cdot X] = \alpha^2 \cdot \text{Var}[X]$. 

![Graph showing the distributions of $\mathbb{E}[X]$, $\text{Var}[X]$, and $\mathbb{E}[\alpha \cdot X]$ and $\text{Var}[\alpha \cdot X]$ for different values of $\alpha$.](image)
Linearity of Expectation

\[ \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \] for any random variables \( X \) and \( Y \). No matter how correlated they may be!
Linearity of Expectation

\[ \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \]

for any random variables \( X \) and \( Y \). No matter how correlated they may be!

Proof:

\[ \mathbb{E}[X + Y] = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot (s + t) \]

\[ X + Y \text{ where } X \sim S + Y \sim T \]
Linearity of Expectation

\[ E[X + Y] = E[X] + E[Y] \] for any random variables \( X \) and \( Y \). No matter how correlated they may be!

Proof:

\[
E[X + Y] = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot (s + t) \\
= \sum_s \left( \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot s \right) + \sum_t \sum_s \Pr(X = s \cap Y = t) \cdot t \\
= \sum_s \Pr(X = s) \cdot s + \sum_t \Pr(Y = t) \cdot t \\
= E[X] + E[Y].
\]
Linearity of Expectation

\[ \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \] for any random variables \( X \) and \( Y \). No matter how correlated they may be!

Proof:

\[ \mathbb{E}[X + Y] = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot (s + t) \]

\[ = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot s + \sum_{t \in T} \sum_{s \in S} \Pr(X = s \cap Y = t) \cdot t \]

\[ = \sum_{s \in S} \Pr(X = s) \cdot s + \sum_{t \in T} \Pr(Y = t) \cdot t \]
Linearity of Expectation

\[ \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \] for any random variables \( X \) and \( Y \). No matter how correlated they may be!

Proof:

\[
\mathbb{E}[X + Y] = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot (s + t)
\]

\[
= \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot s + \sum_{t \in T} \sum_{s \in S} \Pr(X = s \cap Y = t) \cdot t
\]

\[
= \sum_{s \in S} \Pr(X = s) \cdot s + \sum_{t \in T} \Pr(Y = t) \cdot t
\]

\[
= \mathbb{E}[X] + \mathbb{E}[Y].
\]
Linearity of Expectation

\[ \mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \] for any random variables \( X \) and \( Y \). No matter how correlated they may be!

Proof:

\[ \mathbb{E}[X + Y] = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot (s + t) \]

\[ = \sum_{s \in S} \sum_{t \in T} \Pr(X = s \cap Y = t) \cdot s + \sum_{t \in T} \sum_{s \in S} \Pr(X = s \cap Y = t) \cdot t \]

\[ = \sum_{s \in S} \Pr(X = s) \cdot s + \sum_{t \in T} \Pr(Y = t) \cdot t \]

\[ = \mathbb{E}[X] + \mathbb{E}[Y]. \]

Maybe the single most powerful tool in the analysis of randomized algorithms.
Linearity of Variance

\[ \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \quad \text{when } X \text{ and } Y \text{ are independent.} \]
Linearity of Variance

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \text{ when } X \text{ and } Y \text{ are independent.}$$

Claim 1: (exercise) $$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \text{ (via linearity of expectation)}$$

Claim 2: (exercise) $$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \text{ (i.e., } X \text{ and } Y \text{ are uncorrelated) when } X, Y \text{ are independent.}$$
Linearity of Variance

\[ \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \text{ when } X \text{ and } Y \text{ are independent.} \]

**Claim 1: (exercise)** \( \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \) (via linearity of expectation)

**Claim 2: (exercise)** \( \mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \) (i.e., \( X \) and \( Y \) are uncorrelated) when \( X, Y \) are independent.

Together give:

\[ \text{Var}[X + Y] = \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \]
Linearity of Variance

\[ \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \] when \( X \) and \( Y \) are independent.

Claim 1: (exercise) \( \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \) (via linearity of expectation)

Claim 2: (exercise) \( \mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \) (i.e., \( X \) and \( Y \) are uncorrelated) when \( X, Y \) are independent.

Together give: \[
\mathbb{E}[X^2 + 2XY + Y^2]
\]

\[
\text{Var}[X + Y] = \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2
= \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2
\]
Linearity of Variance

\[ \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \] when \( X \) and \( Y \) are independent.

Claim 1: (exercise) \( \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \) (via linearity of expectation)

Claim 2: (exercise) \( \mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \) (i.e., \( X \) and \( Y \) are uncorrelated) when \( X, Y \) are independent.

Together give:

\[ \text{Var}[X + Y] = \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \]

\[ = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \]

\[ = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X] \cdot \mathbb{E}[Y] - \mathbb{E}[Y]^2 \]

\[ \approx \text{Var}(X) + \text{Var}(Y) + 2\mathbb{E}[XY] - 2\mathbb{E}[X] \cdot \mathbb{E}[Y] \]

0 by Claim 10.
Linearity of Variance

\[ \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \] when \( X \) and \( Y \) are independent.

Claim 1: (exercise) \( \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \) (via linearity of expectation)

Claim 2: (exercise) \( \mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \) (i.e., \( X \) and \( Y \) are uncorrelated) when \( X, Y \) are independent.

Together give:

\[ \text{Var}[X + Y] = \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \]
\[ = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \]
\[ = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X] \cdot \mathbb{E}[Y] - \mathbb{E}[Y]^2 \]
\[ = \mathbb{E}[X^2] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - \mathbb{E}[Y]^2 \]
Linearity of Variance

\[ \text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \text{ when } X \text{ and } Y \text{ are independent.} \]

Claim 1: (exercise) \( \text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 \) (via linearity of expectation)

Claim 2: (exercise) \( \mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y] \) (i.e., \( X \) and \( Y \) are uncorrelated) when \( X, Y \) are independent.

Together give:

\[ \text{Var}[X + Y] = \mathbb{E}[(X + Y)^2] - \mathbb{E}[X + Y]^2 \]
\[ = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - (\mathbb{E}[X] + \mathbb{E}[Y])^2 \]
\[ = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - 2\mathbb{E}[X] \cdot \mathbb{E}[Y] - \mathbb{E}[Y]^2 \]
\[ = \mathbb{E}[X^2] + \mathbb{E}[Y^2] - \mathbb{E}[X]^2 - \mathbb{E}[Y]^2 \]
\[ = \text{Var}[X] + \text{Var}[Y]. \]
Exercise: Verify that for random variables $X_1, \ldots, X_n,$

$$
\text{Var} \left( \sum_{i=1}^{n} X_i \right) = \sum_{i=1}^{n} \text{Var}(X_i),
$$

whenever the variables are 2-wise independent (also called pairwise independent).
Application 2: Quicksort with Random Pivots
Quicksort($X$): where $X = (x_1, \ldots, x_n)$ is a list of numbers.

1. If $X$ is empty: return $X$.
2. Else: select pivot $p$ uniformly at random from {$1, \ldots, n$}.
3. Let $X_{lo} = \{i \in X : x_i < x_p\}$ and $X_{hi} = \{i \in X : x_i \geq x_p\}$ (requires $n - 1$ comparisons with $x_p$ to determine).
4. Return the concatenation of the lists $[\text{Quicksort}(X_{lo}), (x_p), \text{Quicksort}(X_{hi})]$. 

What is the worst case running time of this algorithm?
Quicksort($X$): where $X = (x_1, \ldots, x_n)$ is a list of numbers.

1. If $X$ is empty: return $X$.
2. Else: select pivot $p$ uniformly at random from $\{1, \ldots, n\}$.
3. Let $X_{lo} = \{i \in X : x_i < x_p\}$ and $X_{hi} = \{i \in X : x_i \geq x_p\}$ (requires $n - 1$ comparisons with $x_p$ to determine).
4. Return the concatenation of the lists $[\text{Quicksort}(X_{lo}), (x_p), \text{Quicksort}(X_{hi})]$. 

| 4 | 5 | 2 | 8 | 1 | 3 | 6 | 9 | 7 | 0 |
Quicksort(\(X\)): where \(X = (x_1, \ldots, x_n)\) is a list of numbers.

1. If \(X\) is empty: return \(X\).

2. Else: select pivot \(p\) uniformly at random from \(\{1, \ldots, n\}\).

3. Let \(X_{lo} = \{i \in X : x_i < x_p\}\) and \(X_{hi} = \{i \in X : x_i \geq x_p\}\) (requires \(n - 1\) comparisons with \(x_p\) to determine).

4. Return the concatenation of the lists \([\text{Quicksort}(X_{lo}), (x_p), \text{Quicksort}(X_{hi})]\).
Quicksort($X$): where $X = (x_1, \ldots, x_n)$ is a list of numbers.

1. If $X$ is empty: return $X$.

2. Else: select pivot $p$ uniformly at random from $\{1, \ldots, n\}$.

3. Let $X_{lo} = \{i \in X : x_i < x_p\}$ and $X_{hi} = \{i \in X : x_i \geq x_p\}$ (requires $n - 1$ comparisons with $x_p$ to determine).

4. Return the concatenation of the lists
   $[\text{Quicksort}(X_{lo}), (x_p), \text{Quicksort}(X_{hi})]$. 

![Quicksort diagram]

What is the worst case running time of this algorithm?
Quicksort(X): where \( X = (x_1, \ldots, x_n) \) is a list of numbers.

1. If \( X \) is empty: return \( X \).
2. Else: select pivot \( p \) uniformly at random from \( \{1, \ldots, n\} \).
3. Let \( X_{lo} = \{i \in X : x_i < x_p\} \) and \( X_{hi} = \{i \in X : x_i \geq x_p\} \) (requires \( n - 1 \) comparisons with \( x_p \) to determine).
4. Return the concatenation of the lists 
   \([\text{Quicksort}(X_{lo}), (x_p), \text{Quicksort}(X_{hi})]\).

\[ \#E[T] = O(n \log n) \]
Quicksort(X): where X = (x₁, ..., xₙ) is a list of numbers.

1. If X is empty: return X.
2. Else: select pivot p uniformly at random from {1, ..., n}.
3. Let Xₜₒ = \{i ∈ X : xᵢ < xₚ\} and Xₜᵢ = \{i ∈ X : xᵢ ≥ xₚ\} (requires \(n - 1\) comparisons with \(xₚ\) to determine).
4. Return the concatenation of the lists
   \([\text{Quicksort}(Xₜₒ), (xₚ), \text{Quicksort}(Xₜᵢ)]\).

What is the worst case running time of this algorithm?

\(O(n²)\)
Randomized Quicksort Analysis

**Theorem:** Let $T$ be the number of comparisons performed by Quicksort($X$). Then $\mathbb{E}[T] = O(n \log n)$. 
Randomized Quicksort Analysis

**Theorem:** Let $T$ be the number of comparisons performed by Quicksort($X$). Then $\mathbb{E}[T] = O(n \log n)$.

- For any $i, j \in [n]$ with $i < j$, let $I_{ij} = 1$ if $x_i, x_j$ are compared at some point during the algorithm, and $I_{ij} = 0$ if they are not. An indicator random variable.
Theorem: Let $T$ be the number of comparisons performed by Quicksort($X$). Then $E[T] = O(n \log n)$.

- For any $i, j \in [n]$ with $i < j$, let $I_{ij} = 1$ if $x_i, x_j$ are compared at some point during the algorithm, and $I_{ij} = 0$ if they are not. An indicator random variable.
- We can write $T = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I_{ij}$. 

$\textbf{Randomized Quicksort Analysis}$
Randomized Quicksort Analysis

**Theorem:** Let $T$ be the number of comparisons performed by Quicksort($X$). Then $\mathbb{E}[T] = O(n \log n)$.

- For any $i, j \in [n]$ with $i < j$, let $I_{ij} = 1$ if $x_i, x_j$ are compared at some point during the algorithm, and $I_{ij} = 0$ if they are not. An indicator random variable.
- We can write $T = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I_{ij}$. Thus, via linearity of expectation

$$
\mathbb{E}[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}[I_{ij}]
$$

So we need to upper bound $\mathbb{P}[x_i, x_j$ are compared $]$. (Are these independent?)
Randomized Quicksort Analysis

**Theorem:** Let $T$ be the number of comparisons performed by Quicksort$(X)$. Then $\mathbb{E}[T] = O(n \log n)$.

- For any $i, j \in [n]$ with $i < j$, let $I_{ij} = 1$ if $x_i, x_j$ are compared at some point during the algorithm, and $I_{ij} = 0$ if they are not. An *indicator random variable*.

- We can write $T = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I_{ij}$. Thus, via linearity of expectation

\[
\mathbb{E}[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}[I_{ij}] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr[x_i, x_j \text{ are compared}]
\]
Theorem: Let $T$ be the number of comparisons performed by Quicksort($X$). Then $\mathbb{E}[T] = O(n \log n)$.

- For any $i, j \in [n]$ with $i < j$, let $I_{ij} = 1$ if $x_i, x_j$ are compared at some point during the algorithm, and $I_{ij} = 0$ if they are not. An indicator random variable.
- We can write $T = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I_{ij}$. Thus, via linearity of expectation

$$
\mathbb{E}[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{E}[I_{ij}] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr[x_i, x_j \text{ are compared}]
$$

So we need to upper bound $\Pr[x_i, x_j \text{ are compared}]$. 


Randomized Quicksort Analysis

Upper bounding $\Pr[x_i, x_j \text{ are compared}]:$

[Insert explanation or formula here]
Upper bounding $\Pr[x_i, x_j \text{ are compared}]$: $x_i < x_j$

- Assume without loss of generality that $x_1 \leq x_2 \leq \ldots \leq x_n$. This is just ‘renaming’ the elements of our list. Also recall that $i < j$. 
Randomized Quicksort Analysis

Upper bounding $\Pr[x_i, x_j \text{ are compared}]:$

- Assume without loss of generality that $x_1 \leq x_2 \leq \ldots \leq x_n$. This is just ‘renaming’ the elements of our list. Also recall that $i < j$.

- At exactly one step of the recursion, $x_i, x_j$ will be ‘split up’ with one landing in $X_{hi}$ and the other landing in $X_{lo}$, or one being chosen as the pivot. $x_i, x_j$ are only ever compared in this later case – if one is chosen as the pivot when they are split up.
Randomized Quicksort Analysis

Upper bounding $\Pr[x_i, x_j \text{ are compared}]:$

- Assume without loss of generality that $x_1 \leq x_2 \leq \ldots \leq x_n$. This is just ‘renaming’ the elements of our list. Also recall that $i < j$.
- At exactly one step of the recursion, $x_i, x_j$ will be ‘split up’ with one landing in $X_{hi}$ and the other landing in $X_{lo}$, or one being chosen as the pivot. $x_i, x_j$ are only ever compared in this later case – if one is chosen as the pivot when they are split up.
- The split occurs when some element between $x_i$ and $x_j$ is chosen as the pivot. The possible elements are $x_i, x_{i+1}, \ldots, x_j$.  

\[
\begin{array}{cccccccc}
4 & 5 & 2 & 1 & 3 & 0 & 6 & 8 & 9 & 7 \\
\end{array}
\]

\[
\Pr(x_i + x_j \text{ are compared}) = \frac{2}{j - i + 1}
\]
Randomized Quicksort Analysis

Upper bounding $\Pr[x_i, x_j$ are compared]:

- Assume without loss of generality that $x_1 \leq x_2 \leq \ldots \leq x_n$. This is just ‘renaming’ the elements of our list. Also recall that $i < j$.
- At exactly one step of the recursion, $x_i, x_j$ will be ‘split up’ with one landing in $X_{hi}$ and the other landing in $X_{lo}$, or one being chosen as the pivot. $x_i, x_j$ are only ever compared in this later case – if one is chosen as the pivot when they are split up.
- The split occurs when some element between $x_i$ and $x_j$ is chosen as the pivot. The possible elements are $x_i, x_{i+1}, \ldots, x_j$.

- $\Pr[x_i, x_j$ are compared] is equal to the probability that either $x_i$ or $x_j$ are chosen as the splitting pivot from this list. Thus, $\Pr[x_i, x_j$ are compared] $= \frac{2}{j-i+1}$.
So Far: Expected number of comparisons is given as:

\[ E[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr[x_i, x_j \text{ are compared}]. \]

And we computed \( \Pr[x_i, x_j \text{ are compared}] = \frac{2}{j-i+1} \).
Randomized Quicksort Analysis

**So Far:** Expected number of comparisons is given as:

\[
\mathbb{E}[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr[x_i, x_j \text{ are compared}].
\]

And we computed \( \Pr[x_i, x_j \text{ are compared}] = \frac{2}{j-i+1} \). Plugging in:

\[
\mathbb{E}[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}
\]
Randomized Quicksort Analysis

So Far: Expected number of comparisons is given as:

\[
E[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{Pr}[x_i, x_j \text{ are compared}].
\]

And we computed \( \text{Pr}[x_i, x_j \text{ are compared}] = \frac{2}{j-i+1} \). Plugging in:

\[
E[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} = \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k}.
\]
Randomized Quicksort Analysis

**So Far:** Expected number of comparisons is given as:

\[
E[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \text{Pr}[x_i, x_j \text{ are compared}].
\]

And we computed \( \text{Pr}[x_i, x_j \text{ are compared}] = \frac{2}{j-i+1} \). Plugging in:

\[
E[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1} = \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k}
\]

\[
\leq \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k} \leq 2 \cdot (n-1) \cdot \sum_{k=1}^{n} \frac{1}{k} \approx O(n \log n).
\]
So Far: Expected number of comparisons is given as:

\[ \mathbb{E}[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \Pr[x_i, x_j \text{ are compared}] \]

And we computed \( \Pr[x_i, x_j \text{ are compared}] = \frac{2}{j - i + 1} \). Plugging in:

\[ \mathbb{E}[T] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j - i + 1} = \sum_{i=1}^{n-1} \sum_{k=2}^{n-i+1} \frac{2}{k} \]

\[ \leq \sum_{i=1}^{n-1} \sum_{k=1}^{n} \frac{2}{k} \leq 2 \cdot (n - 1) \cdot \sum_{k=1}^{n} \frac{1}{k} = 2n \cdot H_n = O(n \log n). \]