Expectation/Linearity of expectation (review)
Markov’s Inequality
Variance
Chebyshev's Inequality
The **expectation** or **expected value** of a discrete random variable is:

\[ E[X] = \sum_{x \in \mathcal{X}} xp_X(x) \]

- **(Reminder) Linearity of expectation:**
  Consider a **non-random (deterministic) function** of a random variable:

  \[ Y = g(X) \quad \Rightarrow \quad E[Y] = E[g(X)] = \sum_x g(x)p_X(x) \]

  \[ g(E[X]) \neq E[g(X)] \]

**Linearity of expectation:**

\[ g(X) = aX + b \quad \Rightarrow \quad E[g(X)] = g(E[X]) \]
The expectation or expected value of a function of two discrete variables:

\[ E[(X, Y)] = \sum (x, y)p_{XY}(x, y) \]

Not what we want!

\[ E[g(X, Y)] = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} g(x, y)p_{XY}(x, y) \]
Expectation of Multiple Variables

➢ The *expectation* or *expected value* of a function of two discrete variables:

\[ E[g(X, Y)] = \sum_{x \in X} \sum_{y \in Y} g(x, y)p_{XY}(x, y) \]

**Linearity of expectation for two random variables:**

\[ g(X, Y) \text{ linear} \quad \rightarrow \quad E[g(X, Y)] = g(E[X], E[Y]) \]

Example 1. Sum is a linear function!

\[ g(X, Y) = X + Y \quad E[X + Y] = E[X] + E[Y] \]

Example 2. \[ g(X, Y) = aX + bY + c \]

\[ E[aX + bY + c] = aE[X] + bE[Y] + c \]

Linearity of expectation holds whether or not X and Y are independent
The *expectation* or *expected value* of a function of $n$ discrete variables:

Let $X_1, X_2, X_3, ..., X_n$ be $n$ random variables, having ranges $\chi_1, \chi_2, ..., \chi_n$ respectively.

Let $g$ be a function defined on $X_1, X_2, X_3, ..., X_n$. We have:

$$E[g(X_1, X_2, X_3, ..., X_n)] =$$

$$\sum_{x_1 \in \chi_1} \sum_{x_2 \in \chi_2} ... \sum_{x_n \in \chi_n} g(x_1, x_2, x_3, ..., x_n) P_{X_1, X_2, X_3, ..., X_n}(x_1, x_2, x_3, ..., x_n)$$

**Linearity of expectation.** If $g$ is a linear function

$$E[g(X_1, X_2, X_3, ..., X_n)] = g(E[X_1], E[X_2], ..., E[X_n])$$

We do not need independence!
Binomial Distribution

Let $Y$ be a binomial random variable with parameters $n$ and $p$

**Reminder:** range of $P_Y$ is $\{0,1,...,n\}$, $E[Y] = np$

Let $X_1X_2X_3,...,X_n$ be $n$ Bernoulli random variable with parameters $p$ *(not necessarily independent)*

**Reminder:** range of $P_{X_1X_2X_3...X_n}$ is an $n$ dimensional hypercube.

If we assume $X_1X_2X_3,...,X_n$ are independent, the probability distribution of $Y=\sum_{i=1}^{n}X_i$ is binomial with parameters $n$ and $p$
Going back to the Binomial Distribution

Let \( Y \) be a binomial random variable with parameters \( n \) and \( p \).

**Reminder:** \( E[Y] = np \)

Let \( X_1, X_2, X_3, \ldots, X_n \) be \( n \) Bernoulli random variable with parameters \( p \) (not necessarily independent). Let \( Y = \sum_{i=1}^{n} X_i \)

What is \( E[Y] = ? \)

**First approach.** Take \( g(x_1, x_2, \ldots, x_n) = \sum x_i \),

Calculate \( E[g] = \sum g(x_1, x_2, \ldots, x_n) P_{X_1X_2X_3\ldots X_n} (x_1, x_2, \ldots, x_n) \)

What is \( P_{X_1X_2X_3\ldots X_n} \)?
Going back to the Binomial Distribution

Let \( X_1X_2X_3, \ldots, X_n \) be \( n \) Bernoulli random variables all with parameters \( p \) (not necessarily independent).

What is \( E[\sum X_i] =? \)

Second approach. Take \( g(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} x_i \), sum is linear! Use linearity of expectation.

\[
E[g(X_1, X_2, X_3, \ldots, X_n)] = g(E[X_1], [X_2], \ldots, E[X_n])) = \sum_{i=1}^{n} E[X_i]
\]

No need to know \( P_{X_1X_2X_3\ldots X_n} \)!

\[
E[\sum X_i] = np \text{ whether or not } X_i \text{'s are independent!}
\]
Calculating Expectation

- We know $P_Y$
  - We can calculate the weighted average
  - It is a distribution that we know: Bernoulli, Binomial, Geometric

- We don’t know $P_Y$ but we have $Y = g(X_1, X_2, \ldots, X_n)$, and we know $P_{X_i}$’s and $g$ is a linear function
  - We calculate $E[X_i]$s, and use linearity of expectation
    \[
    E[g(X_1, X_2, X_3, \ldots, X_n)] = g(E[X_1], E[X_2], \ldots, E[X_n])
    \]

Examples

- Coupon collector
- Number of successes in non-independent trials
- The hat problem (Example 2.11 of your book)
Let $X_1 X_2 X_3, \ldots, X_n$ be $n$ random variables and $g(X_1, X_2, X_3, \ldots, X_n)$.

Consider r.v. $Y$ having unknown distribution $P_Y$. If we know $E[Y] = \mu$, can we conclude anything about $P_Y$?
Consider r.v. $Y$ having unknown distribution $P_Y$. If we know $E[Y] = \mu$, can we conclude anything about $P_Y$?

What can we say about $P(E[Y] = Y)$?

Let $Y = \{+1, -1\}$ with equal probability thus $E[Y] = 0, P(E[Y] = Y) = 0$.

What if we repeat $n$ times? $Y_1, Y_2, Y_3, \ldots, Y_n$ thus $E[Y] = E[Y_i]$. Can we claim $P(E[Y] = Y_i) > 0$? **No!** the same example.

What if instead of looking at $P(E[Y] = Y_i)$ we look at how close/far $E[Y]$ and $Y$ can be? Basically can we say $P(|E[Y] - Y| > k)$ gets small as $k$ gets bigger?
What if instead of looking at $P(E[Y] = Y_i)$ we look at how close/far $E[Y]$ and $Y$ can be? Basically can we say $P(|E[Y] - Y| > k)$ gets small as $k$ gets bigger?

For arbitrary $k$ let $Y = \{-2k, +2k\}$, thus, $E[Y] = 0$ and $P(|E[Y] - Y| > k) = 1$.

**Theorem**

[Markov Inequality] For any non-negative random variable, and for all $a > 0$,

$$Pr(X \geq a) \leq \frac{E[X]}{a}.$$
What if instead of looking at $P(E[Y] = Y_i)$ we look at how close/far $E[Y]$ and $Y$ can be? Basically can we say $P(|E[Y] - Y| > k)$ gets small as $k$ gets bigger?

For arbitrary $k$ let $Y = \{-2k, +2k\}$, thus, $E[Y] = 0$ and $P(|E[Y] - Y| > k) = 1$

**Theorem**

**[Markov Inequality]** For any non-negative random variable, and for all $a > 0$,

$$Pr(X \geq a) \leq \frac{E[X]}{a}.$$
Markov inequality

Theorem

[Markov Inequality] For any non-negative random variable, and for all \( a > 0 \),

\[
Pr(X \geq a) \leq \frac{E[X]}{a}.
\]

Fix some constant \( a > 0 \), and define

\[
Y_a = \begin{cases} 
0, & \text{if } X < a, \\
\frac{a}{X}, & \text{if } X \geq a.
\end{cases}
\]

\[
a P(X \geq a) = E[Y_a] \leq E[X]
\]
Markov’s Inequality

[Markov Inequality] For any non-negative random variable, and for all \( a > 0 \),

\[
Pr(X \geq a) \leq \frac{E[X]}{a}.
\]

Fix some constant \( a > 0 \), and define

\[
Y_a = \begin{cases} 
0, & \text{if } X < a, \\
a, & \text{if } X \geq a.
\end{cases}
\]

- No such inequality would hold if \( X \) could take negative values. Why?
- If \( a < E[X] \), Markov’s inequality is vacuous, but no better bound is possible. Why?
Reminder. The *expectation* or *expected value* of a random variable

\[ E[X] = \sum_{x \in \mathcal{X}} x p_X(x) \]

The *variance* is the *expected squared deviation* of a random variable from its mean (the following definitions are equivalent):

\[ \text{Var}[X] = E[(X - E[X])^2] = \sum_{x \in \mathcal{X}} (x - E[X])^2 p_X(x) \]

\[ \text{Var}[X] = E[X^2] - E[X]^2 = \left[ \sum_{x \in \mathcal{X}} x^2 p_X(x) \right] - \left[ \sum_{x \in \mathcal{X}} x p_X(x) \right]^2 \]

The *standard deviation* is the square root of the variance:

\[ \sigma_X = \text{Std}[X] = \sqrt{\text{Var}[X]} \]
A Bernoulli or indicator random variable $X$ has one parameter $p$:

$$p_X(1) = p, \quad p_X(0) = 1 - p, \quad \mathcal{X} = \{0, 1\}$$

For an indicator variable, expected values are probabilities:

$$E[X] = p$$

Variance of Bernoulli distribution:

$$\text{Var}[X] = E\left[(X - p)^2\right] = p(1 - p)$$

Fair coin ($p=0.5$) has largest variance.

Coins that always come up heads ($p=1.0$), or always come up tails ($p=0.0$), have variance 0.
Sums of Independent Variables

- If $Z = X + Y$ and random variables $X$ and $Y$ are independent, we have

$$E[Z] = E[X] + E[Y] \quad \text{Var}[Z] = \text{Var}[X] + \text{Var}[Y]$$

For any variables $X$, $Y$.  

Only for independent $X$, $Y$.

- Interpretation: Adding independent variables increases variance

$$\text{Var}[Z] \geq \text{Var}[X] \quad \text{and} \quad \text{Var}[Z] \geq \text{Var}[Y]$$

- Example when we do not have independence:

Let $X$ and $Y$ be two Bernoulli r.v. such that $P(X=Y)=1$

$$P_{XY} = \begin{bmatrix}
0 & 1 \\
0 & 1-p & 0 \\
1 & 0 & p
\end{bmatrix}$$

Sums of Independent Variables

➢ If $Z = X + Y$ and random variables $X$ and $Y$ are independent, we have

$$E[Z] = E[X] + E[Y] \quad \text{Var}[Z] = \text{Var}[X] + \text{Var}[Y]$$

For any variables $X, Y$. Only for independent $X, Y$.

➢ Interpretation: Adding independent variables increases variance

$$\text{Var}[Z] \geq \text{Var}[X] \quad \text{and} \quad \text{Var}[Z] \geq \text{Var}[Y]$$

➢ The standard deviation of a sum of independent variables is then

$$\sigma_Z = \sqrt{\sigma_X^2 + \sigma_Y^2} \quad \sigma_X = \sqrt{\text{Var}[X]}, \sigma_Y = \sqrt{\text{Var}[Y]}, \sigma_Z = \sqrt{\text{Var}[Z]}$$

➢ Identity used in proof: If $X$ and $Y$ are independent random variables,

$$E[XY] = E[X]E[Y] \quad \text{if} \quad p_{XY}(x, y) = p_X(x)p_Y(y)$$

This equality does not hold for general, dependent random variables.
Linear Combination of Independent Variables

\[ \text{Var}[X + a] = \text{Var}[X] + \text{Var}[a] = \text{Var}[X] \]

\[ \text{Var}[X] = \quad X: \Omega \to \mathcal{X}, \exists a \in \Omega; p_X(a) = 1 \]

Equivalently \[ \text{Var}[a] = \]

\[ \text{Var}[aX] = \]

\[ \text{Var}[aX + bY + c] = \]

\[ E[aX + bY + c] = \]
Suppose you flip \( n \) coins with bias \( p \), count number of heads

A **binomial** random variable \( X \) has parameters \( n, p \):

\[
p_{X}(k) = \binom{n}{k} p^{k} (1 - p)^{n-k}
\]

\( X_i \) is a Bernoulli variable indicating whether toss \( i \) comes up heads, because tosses are **independent**:

\[
X = \sum_{i=1}^{n} X_i
\]

\[
E[X] = np \quad Var[X] = np(1 - p)
\]

\( n = 20, p = 0.5 \)
Suppose you flip $n$ coins with bias $p$, count number of heads

A binomial random variable $X$ has parameters $n, p$:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$X_i$ is a Bernoulli variable indicating whether toss $i$ comes up heads, because tosses are independent:

$$X = \sum_{i=1}^{n} X_i \quad E[X] = np \quad \text{Var}[X] = np(1-p)$$

$n = 20, p = 0.2$
Suppose you flip \( n \) coins with bias \( p \), count number of heads

A binomial random variable \( X \) has parameters \( n, p \):

\[
p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}
\]

\( X_i \) is a Bernoulli variable indicating whether toss \( i \) comes up heads, because tosses are independent:

\[
X = \sum_{i=1}^{n} X_i \quad E[X] = np \quad Var[X] = np(1 - p)
\]

\( n = 100, p = 0.5 \)
Variance: What Comes Next?

- Expectation and Variance
- Markov’s Inequality
- Variance
- Chebyshev's Inequality

Can we say $P(|E[Y] - Y| > k)$ gets small as $k$ gets bigger?
Chebyshev’s Inequality

**Theorem**

For any random variable $X$, and any $a > 0$,

$$\Pr(|X - E[X]| \geq a) \leq \frac{\text{Var}[X]}{a^2}.$$ 

**Proof.**

$$\Pr(|X - E[X]| \geq a) = \Pr((X - E[X])^2 \geq a^2)$$

By Markov inequality

$$\Pr((X - E[X])^2 \geq a^2) \leq \frac{E[(X - E[X])^2]}{a^2} = \frac{\text{Var}[X]}{a^2}$$
Chebyshev’s Inequality

**Theorem**

*For any* random variable $X$, and any $a > 0$,

$$Pr(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}.$$ 

➢ **Another way of parameterizing** Chebyshev’s inequality:

$$\mu = E[X], \quad \sigma = \sqrt{Var[X]}$$

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

➢ **Chebyshev bound is vacuous** (above one) for events less than one standard deviation from the mean. But this could be likely!
Another way of parameterizing Chebyshev’s inequality:

$$\mu = E[X], \quad \sigma = \sqrt{\text{Var}[X]}$$

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

Chebyshev bound is vacuous (above one) for events less than one standard deviation from the mean. But this could be likely!
We flip a fair coin $n$ times. What is the probability of getting more than $3n/4$ heads?


Markov’s Inequality:

$$Pr\{X \geq \frac{3n}{4}\} \leq \frac{E[X]}{3n/4} = \frac{n/2}{3n/4} = \frac{2}{3}$$

Chebyshev’s Inequality:

$$Pr\{X \geq \frac{3n}{4}\} \leq Pr\{|X - \frac{n}{2}| \geq \frac{n}{4}\} \leq \frac{Var[X]}{(n/4)^2} = \frac{n/4}{n^2/16} = \frac{4}{n}$$
Chebyshev’s Inequality

• Let $X$ be a r.v. with:
  • $E[X] = \mu$
  • $Var[X] = \nu$

• Let $X_1, X_2, \ldots, X_n$ be $n$ copies of $X$.
  • They have all have the same expectation and variance

• Let $Y = \frac{X_1 + X_2 + \cdots + X_n}{n}$.
  • What is $E[Y]$?
Chebyshev’s Inequality

Let $X$ be a r.v. with $E[X] = \mu$ and $Var[X] = \nu$.

Let $X_1, X_2, \ldots, X_n$ be $n$ copies of $X$. (which means they all have the same expectation and variance)

Let $Y = \frac{X_1 + X_2 + \cdots + X_n}{n}$, thus $E[Y] = \mu$. What is $Var[Y]$?

**Theorem**

For any random variable $X$, and any $a > 0$,

$$Pr(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}.$$
Let $X$ be a r.v. with $E[X] = \mu$ and $Var[X] = \nu$.

Let $X_1, X_2, \ldots, X_n$ be $n$ independent copies of $X$. (which means they all have the same expectation and variance)

Let $Y = \frac{X_1 + X_2 + \cdots + X_n}{n}$, thus $E[Y] = \mu$. What is $Var[Y]$?

**Theorem**

*For any random variable $X$, and any $a > 0$,*

$$Pr(|X - E[X]| \geq a) \leq \frac{Var[X]}{a^2}.$$
Chebyshev’s Inequality

Let $X$ be a r.v. with $E[X] = \mu$ and $Var[X] = \nu$.

Let $X_1, X_2, \ldots, X_n$ be $n$ independent copies of $X$. (which means they all have the same expectation and variance)

Let $Y = \frac{X_1 + X_2 + \cdots + X_n}{n}$, thus $E[Y] = \mu$. $Var[Y] = \nu/n$.

What is $P(|X - \mu| \geq k)$? What is $P(|Y - \mu| \geq k)$

$P(|X - \mu| \geq k) \leq \frac{\nu}{k^2}$

$$P \left( \left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| \geq k \right) \leq \frac{\nu}{nk^2}$$
What are expectation and variance?

Let $X$ be a r.v. with $E[X] = \mu$; by the Law of Large Numbers

\[ X_1, X_2, \ldots, X_n, \ldots \]

\[ n \to \infty \quad P\left( \left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| \geq k \right) \to 0 \]

Let $X$ be a r.v. with $E[X] = \mu$ and $Var[X] = E[(X - E[X])^2] = \nu$. 

\[ X_1, X_2, \ldots, X_n \]

\[ P\left( \left| \frac{X_1 + X_2 + \cdots + X_n}{n} - \mu \right| \geq k \right) \leq \frac{\nu}{nk^2} \]