CS70
Random Variables

Kelvin Lee

UC Berkeley

April 6, 2021

Overview

Definitions

Definitions

(Random Variable).

Definitions

(Random Variable).

A **random variable** X on a sample space Ω is a function $X : \Omega \to \mathbb{R}$ that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

Definitions

(Random Variable).

A **random variable** X on a sample space Ω is a function $X : \Omega \to \mathbb{R}$ that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

(Distribution).

Definitions

(Random Variable).

A **random variable** X on a sample space Ω is a function $X : \Omega \to \mathbb{R}$ that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

(Distribution).

The **distribution** of a discrete random variable X is the collection of values $\{(x, \mathbb{P}(X = x)) : x \in \mathcal{X}\}$, where $\mathcal{X}$ is the set of all possible values taken by X.

Definitions

(Random Variable).

A **random variable** X on a sample space Ω is a function $X : \Omega \to \mathbb{R}$ that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

(Distribution).

The **distribution** of a discrete random variable X is the collection of values $\{(x, \mathbb{P}(X = x)) : x \in \mathcal{X}\}$, where $\mathcal{X}$ is the set of all possible values taken by X.

(Probability Mass Function).

Definitions

(Random Variable).

A **random variable** X on a sample space Ω is a function $X : \Omega \to \mathbb{R}$ that assigns to each sample point $\omega \in \Omega$ a real number $X(\omega)$.

(Distribution).

The **distribution** of a discrete random variable X is the collection of values $\{(x, \mathbb{P}(X = x)) : x \in \mathcal{X}\}$, where $\mathcal{X}$ is the set of all possible values taken by X.

(Probability Mass Function).

The **probability mass function**, or **PMF**, of a discrete random variable X is a function mapping X's values to their associated probabilities. It is the function $p :\to [0, 1]$ defined by

$$p_X(x) := \mathbb{P}(X = x).$$

Definitions

Definitions

(Joint Distribution).

Definitions

(Joint Distribution).

The **joint distribution** for two discrete random variables X and Y is the collection of values $\{((x, y), \mathbb{P}(X = x, Y = y)) : x \in \mathcal{X}, y \in \mathcal{Y}\}$, where $\mathcal{X}$ is the set of all possible values taken by X and $\mathcal{Y}$ is the set of all possible values taken Y.

Definitions

(Joint Distribution).

The **joint distribution** for two discrete random variables X and Y is the collection of values $\{((x, y), \mathbb{P}(X = x, Y = y)) : x \in \mathcal{X}, y \in \mathcal{Y}\}$, where $\mathcal{X}$ is the set of all possible values taken by X and $\mathcal{Y}$ is the set of all possible values taken Y.

(Marginal Distribution).

Definitions

(Joint Distribution).

The **joint distribution** for two discrete random variables X and Y is the collection of values $\{((x,y), \mathbb{P}(X = x, Y = y)) : x \in \mathcal{X}, y \in \mathcal{Y}\}$, where $\mathcal{X}$ is the set of all possible values taken by X and $\mathcal{Y}$ is the set of all possible values taken Y.

(Marginal Distribution).

Given the joint distribution for X and Y, the **marginal distribution** for X is as follows:

Definitions

(Joint Distribution).

The **joint distribution** for two discrete random variables X and Y is the collection of values $\{((x, y), \mathbb{P}(X = x, Y = y)) : x \in \mathcal{X}, y \in \mathcal{Y}\}$, where $\mathcal{X}$ is the set of all possible values taken by X and $\mathcal{Y}$ is the set of all possible values taken Y.

(Marginal Distribution).

Given the joint distribution for X and Y, the **marginal distribution** for X is as follows:

$$\mathbb{P}(X = x) = \sum_{y \in \mathcal{Y}} \mathbb{P}(X = x, Y = y).$$

Independence

Independence

(Independence).

Independence

(Independence).

Random variables X and Y are said to be **independent** if the events $X = x$ and $Y = y$ are independent for all values x, y. Equivalently, the joint distribution of independent R.V's decomposes as

Independence

(Independence).

Random variables X and Y are said to be **independent** if the events $X = x$ and $Y = y$ are independent for all values x, y. Equivalently, the joint distribution of independent R.V's decomposes as

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y), \quad \forall x, y.$$

Expectation

Expectation

(Expectation).

Expectation

(Expectation).

The **expectation** of a discrete random variable X is defined as

Expectation

(Expectation).

The **expectation** of a discrete random variable X is defined as

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}(X = x).$$

Expectation

(Expectation).

The **expectation** of a discrete random variable X is defined as

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}(X = x).$$

Alternatively, we also have

Expectation

(Expectation).

The **expectation** of a discrete random variable X is defined as

$$\mathbb{E}[X] = \sum_{x \in \mathcal{X}} x \cdot \mathbb{P}(X = x).$$

Alternatively, we also have

$$\mathbb{E}[X] = \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}(\omega).$$

Variance

Variance

(Variance).

Variance

(Variance).

The **variance** of a random variable X is

Variance

(Variance).

The **variance** of a random variable X is

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

Variance

(Variance).

The **variance** of a random variable X is

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

(Standard Deviation).

Variance

(Variance).

The **variance** of a random variable X is

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

(Standard Deviation).

The **standard deviation** of a random variable X

Variance

(Variance).

The **variance** of a random variable X is

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2].$$

(Standard Deviation).

The **standard deviation** of a random variable X

$$\sigma := \sqrt{\mathrm{Var}(X)}.$$

Variance

Variance

Theorem

For a random variable X,

Variance

Theorem

For a random variable X,

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Variance

Theorem

For a random variable X,

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Proof:

Variance

Theorem

For a random variable X,

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Proof:

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

Variance

Theorem

For a random variable X,

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Proof:

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] \end{aligned}$$

Variance

Theorem

For a random variable X,

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Proof:

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[2X\mathbb{E}[X]] + \mathbb{E}\left[\mathbb{E}[X]^2\right]
\end{aligned}$$

Variance

Theorem

For a random variable X,

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Proof:

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\ &= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] \\ &= \mathbb{E}\left[X^2\right] - \mathbb{E}[2X\mathbb{E}[X]] + \mathbb{E}\left[\mathbb{E}[X]^2\right] \\ &= \mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \end{aligned}$$

Variance

Theorem

For a random variable X,

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2.$$

Proof:

$$\begin{aligned}
\text{Var}(X) &= \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \\
&= \mathbb{E}\left[X^2 - 2X\mathbb{E}[X] + \mathbb{E}[X]^2\right] \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[2X\mathbb{E}[X]] + \mathbb{E}\left[\mathbb{E}[X]^2\right] \\
&= \mathbb{E}\left[X^2\right] - 2\mathbb{E}[X]^2 + \mathbb{E}[X]^2 \\
&= \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2.
\end{aligned}$$

Variance

Variance

Theorem

For independent random variables X, Y, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Variance

Theorem

For independent random variables X, Y, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof:

Variance

Theorem

For independent random variables X, Y, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof:

$$\mathbb{E}[XY] = \sum_x \sum_y xy \cdot \mathbb{P}(X = x, Y = y)$$

Variance

Theorem

For independent random variables X, Y, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof:

$$\begin{aligned}\mathbb{E}[XY] &= \sum_x \sum_y xy \cdot \mathbb{P}(X = x, Y = y) \\ &= \sum_x \sum_y xy \cdot \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y)\end{aligned}$$

Variance

Theorem

For independent random variables X, Y, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof:

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_x \sum_y xy \cdot \mathbb{P}(X = x, Y = y) \\
&= \sum_x \sum_y xy \cdot \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y) \\
&= \left(\sum_x x \cdot \mathbb{P}(X = x) \right) \cdot \left(\sum_y y \cdot \mathbb{P}(Y = y) \right)
\end{aligned}$$

Variance

Theorem

For independent random variables X, Y, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof:

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_x \sum_y xy \cdot \mathbb{P}(X = x, Y = y) \\
&= \sum_x \sum_y xy \cdot \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y) \\
&= \left(\sum_x x \cdot \mathbb{P}(X = x)\right) \cdot \left(\sum_y y \cdot \mathbb{P}(Y = y)\right) \\
&= \mathbb{E}[X] \cdot \mathbb{E}[Y]
\end{aligned}$$

Variance

Theorem

For independent random variables X, Y, $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Proof:

$$
\begin{aligned}
\mathbb{E}[XY] &= \sum_x \sum_y xy \cdot \mathbb{P}(X = x, Y = y) \\
&= \sum_x \sum_y xy \cdot \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y) \\
&= \left(\sum_x x \cdot \mathbb{P}(X = x) \right) \cdot \left(\sum_y y \cdot \mathbb{P}(Y = y) \right) \\
&= \mathbb{E}[X] \cdot \mathbb{E}[Y]
\end{aligned}
$$

where the second line follows from independence. □

Variance

Variance

Theorem

For independent random variables X, Y,

Variance

Theorem

For independent random variables X, Y,

$$Var(X + Y) = Var(X) + Var(Y).$$

Variance

Theorem

For independent random variables X, Y,

$$Var(X + Y) = Var(X) + Var(Y).$$

Proof:

Variance

Theorem

For independent random variables X, Y,

$$Var(X + Y) = Var(X) + Var(Y).$$

Proof:

Try it on your own.

Variance

Theorem

For independent random variables X, Y,

$$Var(X+Y) = Var(X) + Var(Y).$$

Proof:

Try it on your own. □

Bernoulli Distribution

Bernoulli Distribution

- $X \sim \text{Bernoulli}(p)$.

Bernoulli Distribution

- $X \sim \text{Bernoulli}(p)$.
- One coin flip with probability of p getting a head.

Bernoulli Distribution

- $X \sim \text{Bernoulli}(p)$.
- One coin flip with probability of p getting a head.
- **PMF:**

Bernoulli Distribution

- $X \sim \text{Bernoulli}(p)$.
- One coin flip with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = i) = \begin{cases} p, & \text{if } i = 1 \\ 1 - p, & \text{if } i = 0 \end{cases}$$

Bernoulli Distribution

- $X \sim \text{Bernoulli}(p)$.
- One coin flip with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = i) = \begin{cases} p, & \text{if } i = 1 \\ 1 - p, & \text{if } i = 0 \end{cases}$$

where $0 \leq p \leq 1$.

Bernoulli Distribution

- $X \sim \text{Bernoulli}(p)$.
- One coin flip with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = i) = \begin{cases} p, & \text{if } i = 1 \\ 1 - p, & \text{if } i = 0 \end{cases}$$

  where $0 \leq p \leq 1$.
- **Expectation:**

Bernoulli Distribution

- $X \sim \text{Bernoulli}(p)$.
- One coin flip with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = i) = \begin{cases} p, & \text{if } i = 1 \\ 1 - p, & \text{if } i = 0 \end{cases}$$

  where $0 \leq p \leq 1$.
- **Expectation:**

$$\mathbb{E}[X] = p.$$

Bernoulli Distribution

- $X \sim \text{Bernoulli}(p)$.
- One coin flip with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = i) = \begin{cases} p, & \text{if } i = 1 \\ 1 - p, & \text{if } i = 0 \end{cases}$$

  where $0 \leq p \leq 1$.

- **Expectation:**

$$\mathbb{E}[X] = p.$$

- **Variance:**

Bernoulli Distribution

- $X \sim \text{Bernoulli}(p)$.
- One coin flip with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = i) = \begin{cases} p, & \text{if } i = 1 \\ 1 - p, & \text{if } i = 0 \end{cases}$$

  where $0 \leq p \leq 1$.
- **Expectation:**

$$\mathbb{E}[X] = p.$$

- **Variance:**

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

Binomial Distribution

Binomial Distribution

- $X \sim \text{Bin}(n, p)$.

Binomial Distribution

- $X \sim \text{Bin}(n, p)$.
- n coin flips with probability of p getting a head.

Binomial Distribution

- $X \sim \text{Bin}(n, p)$.
- n coin flips with probability of p getting a head.
- **PMF:**

Binomial Distribution

- $X \sim \text{Bin}(n, p)$.
- n coin flips with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \ldots, n.$$

Binomial Distribution

- $X \sim \text{Bin}(n, p)$.
- n coin flips with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \ldots, n.$$

- **Expectation:**

Binomial Distribution

- $X \sim \text{Bin}(n, p)$.
- n coin flips with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \ldots, n.$$

- **Expectation:**

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = \sum_{i=1}^{n} p = np, \quad Y_i \sim \text{Bernoulli}(p).$$

Binomial Distribution

- $X \sim \text{Bin}(n, p)$.
- n coin flips with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \ldots, n.$$

- **Expectation:**

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = \sum_{i=1}^{n} p = np, \quad Y_i \sim \text{Bernoulli}(p).$$

- **Variance:**

Binomial Distribution

- $X \sim \text{Bin}(n, p)$.
- n coin flips with probability of p getting a head.
- **PMF:**

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{for } k = 0, 1, \ldots, n.$$

- **Expectation:**

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i=1}^{n} Y_i\right] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = \sum_{i=1}^{n} p = np, \quad Y_i \sim \text{Bernoulli}(p).$$

- **Variance:**

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^{n} Y_i\right) = np(1-p).$$

Indicator Random Variable

Indicator Random Variable

(Indicator Random variable).

Indicator Random Variable

(Indicator Random variable).

We use $\mathbb{I}_i$, or X_i to denotes the **indicator random variable** that takes on values $\{0, 1\}$ according to whether a specified event occurs or not.

Indicator Random Variable

(Indicator Random variable).

We use $\mathbb{I}_i$, or X_i to denotes the **indicator random variable** that takes on values $\{0, 1\}$ according to whether a specified event occurs or not.

- Usually $\{\mathbb{I}_i\}_{i=1}^n$ are mutually independent and they are said to be ***independent and identically distributed (i.i.d)***.

Indicator Random Variable

(Indicator Random variable).

We use $\mathbb{I}_i$, or X_i to denotes the **indicator random variable** that takes on values $\{0, 1\}$ according to whether a specified event occurs or not.

- Usually $\{\mathbb{I}_i\}_{i=1}^{n}$ are mutually independent and they are said to be ***independent and identically distributed (i.i.d)***.
- Will be very useful soon for computing expectations.

Geometric Distribution

Geometric Distribution

- $X \sim \text{Geo}(p)$.

Geometric Distribution

- $X \sim \text{Geo}(p)$.
- Number of coin flips until first head appears.

Geometric Distribution

- $X \sim \text{Geo}(p)$.
- Number of coin flips until first head appears.
- **PMF:**

Geometric Distribution

- $X \sim \text{Geo}(p)$.
- Number of coin flips until first head appears.
- **PMF:**

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad \text{for } i = 1, 2, 3, \ldots.$$

Geometric Distribution

- $X \sim \text{Geo}(p)$.
- Number of coin flips until first head appears.
- **PMF:**

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad \text{for } i = 1, 2, 3, \ldots .$$

- **Expectation:**

Geometric Distribution

- $X \sim \text{Geo}\,(p)$.
- Number of coin flips until first head appears.
- **PMF:**

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad \text{for } i = 1, 2, 3, \ldots .$$

- **Expectation:**

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i) = \sum_{x=1}^{\infty} (1-p)^{i-1} = \frac{1}{1-(1-p)} = \frac{1}{p},$$

Geometric Distribution

- $X \sim \text{Geo}(p)$.
- Number of coin flips until first head appears.
- **PMF:**

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad \text{for } i = 1, 2, 3, \ldots.$$

- **Expectation:**

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i) = \sum_{x=1}^{\infty} (1-p)^{i-1} = \frac{1}{1-(1-p)} = \frac{1}{p},$$

where the first equality uses the **tail sum formula** (discussed in next slide).

Geometric Distribution

- $X \sim \text{Geo}(p)$.
- Number of coin flips until first head appears.
- **PMF:**

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad \text{for } i = 1, 2, 3, \ldots.$$

- **Expectation:**

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i) = \sum_{x=1}^{\infty} (1-p)^{i-1} = \frac{1}{1-(1-p)} = \frac{1}{p},$$

where the first equality uses the **tail sum formula** (discussed in next slide).

- **Variance:**

Geometric Distribution

- $X \sim \text{Geo}\,(p)$.
- Number of coin flips until first head appears.
- **PMF:**

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad \text{for } i = 1, 2, 3, \ldots.$$

- **Expectation:**

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i) = \sum_{x=1}^{\infty} (1-p)^{i-1} = \frac{1}{1-(1-p)} = \frac{1}{p},$$

  where the first equality uses the **tail sum formula** (discussed in next slide).

- **Variance:**

$$\text{Var}(X) = \frac{1-p}{p^2}.$$

Tail Sum Formula

Tail Sum Formula

Let X be a random variable that takes values in $\{0, 1, 2, \ldots\}$. Then

Tail Sum Formula

Let X be a random variable that takes values in $\{0, 1, 2, \ldots\}$. Then

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i).$$

Tail Sum Formula

Let X be a random variable that takes values in $\{0, 1, 2, \dots\}$. Then

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i).$$

Proof:

Tail Sum Formula

Let X be a random variable that takes values in $\{0, 1, 2, \ldots\}$. Then

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i).$$

Proof:

$$\mathbb{E}[X] = \sum_{x=1}^{\infty} x\mathbb{P}(X = x)$$

Tail Sum Formula

Let X be a random variable that takes values in $\{0, 1, 2, \ldots\}$. Then

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i).$$

Proof:

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x=1}^{\infty} x\mathbb{P}(X = x) \\ &= \sum_{x=1}^{\infty} \sum_{i=1}^{x} \mathbb{P}(X = x) \end{aligned}$$

Tail Sum Formula

Let X be a random variable that takes values in $\{0, 1, 2, \dots\}$. Then

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i).$$

Proof:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} x\mathbb{P}(X = x) \\
&= \sum_{x=1}^{\infty} \sum_{i=1}^{x} \mathbb{P}(X = x) \\
&= \sum_{i=1}^{\infty} \sum_{x=i}^{\infty} \mathbb{P}(X = x)
\end{aligned}$$

Tail Sum Formula

Let X be a random variable that takes values in $\{0, 1, 2, \ldots\}$. Then

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i).$$

Proof:

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{\infty} x\mathbb{P}(X = x) \\
&= \sum_{x=1}^{\infty} \sum_{i=1}^{x} \mathbb{P}(X = x) \\
&= \sum_{i=1}^{\infty} \sum_{x=i}^{\infty} \mathbb{P}(X = x) \\
&= \sum_{i=1}^{\infty} \mathbb{P}(X \geq i).
\end{aligned}$$

□

A Clever Way of Computing the Geometric Expectation

A Clever Way of Computing the Geometric Expectation

The **Tail Sum Formula** certainly makes the computation a lot easier, but there's an even more simpler way.

A Clever Way of Computing the Geometric Expectation

The **Tail Sum Formula** certainly makes the computation a lot easier, but there's an even more simpler way.

- Suppose we toss our first coin. There are two possibilities: (1) we get a head with probability p and call it a day, (2) we get a tail with probability $1 - p$ and we are right back where we just started.

A Clever Way of Computing the Geometric Expectation

The **Tail Sum Formula** certainly makes the computation a lot easier, but there's an even more simpler way.

- Suppose we toss our first coin. There are two possibilities: (1) we get a head with probability p and call it a day, (2) we get a tail with probability $1 - p$ and we are right back where we just started.
- In the latter case, we expect $1 + \mathbb{E}[X]$ trials until our first success because we already wasted one trial. Hence,

A Clever Way of Computing the Geometric Expectation

The **Tail Sum Formula** certainly makes the computation a lot easier, but there's an even more simpler way.

- Suppose we toss our first coin. There are two possibilities: (1) we get a head with probability p and call it a day, (2) we get a tail with probability $1 - p$ and we are right back where we just started.
- In the latter case, we expect $1 + \mathbb{E}[X]$ trials until our first success because we already wasted one trial. Hence,

$$\mathbb{E}[X] = p \cdot 1 + (1 - p)(1 + \mathbb{E}[X]) \implies \mathbb{E}[X] = \frac{1}{p}.$$

A Clever Way of Computing the Geometric Expectation

The **Tail Sum Formula** certainly makes the computation a lot easier, but there's an even more simpler way.

- Suppose we toss our first coin. There are two possibilities: (1) we get a head with probability p and call it a day, (2) we get a tail with probability $1 - p$ and we are right back where we just started.
- In the latter case, we expect $1 + \mathbb{E}[X]$ trials until our first success because we already wasted one trial. Hence,

$$\mathbb{E}[X] = p \cdot 1 + (1 - p)(1 + \mathbb{E}[X]) \implies \mathbb{E}[X] = \frac{1}{p}.$$

This makes use of an important property called the **memoryless property**, which will be covered later in the class.

Poisson Distribution

Poisson Distribution

- $X \sim \text{Poisson}(\lambda)$.

Poisson Distribution

- $X \sim \text{Poisson}(\lambda)$.
- Models rare events, such as number of arrivals of a bus in an hour.

Poisson Distribution

- $X \sim \text{Poisson}(\lambda)$.
- Models rare events, such as number of arrivals of a bus in an hour.
- Defined in terms of a rate λ, which specifies the average number of times an event occurs in a time interval.

Poisson Distribution

- $X \sim \text{Poisson}(\lambda)$.
- Models rare events, such as number of arrivals of a bus in an hour.
- Defined in terms of a rate λ, which specifies the average number of times an event occurs in a time interval.
- **PMF:**

Poisson Distribution

- $X \sim \text{Poisson}(\lambda)$.
- Models rare events, such as number of arrivals of a bus in an hour.
- Defined in terms of a rate λ, which specifies the average number of times an event occurs in a time interval.
- **PMF:**

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{for } i = 0, 1, 2, \ldots.$$

Poisson Distribution

- $X \sim \text{Poisson}(\lambda)$.
- Models rare events, such as number of arrivals of a bus in an hour.
- Defined in terms of a rate λ, which specifies the average number of times an event occurs in a time interval.
- **PMF:**
$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{for } i = 0, 1, 2, \ldots.$$
- $$\sum_{i=0}^{\infty} \mathbb{P}(X = i) = \sum_{i=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

Poisson Distribution

- $X \sim \text{Poisson}(\lambda)$.
- Models rare events, such as number of arrivals of a bus in an hour.
- Defined in terms of a rate λ, which specifies the average number of times an event occurs in a time interval.
- **PMF:**
$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{for } i = 0, 1, 2, \ldots.$$
- $$\sum_{i=0}^{\infty} \mathbb{P}(X = i) = \sum_{i=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

  Recall the **Taylor series expansion** from calculus:

Poisson Distribution

- $X \sim \text{Poisson}(\lambda)$.
- Models rare events, such as number of arrivals of a bus in an hour.
- Defined in terms of a rate λ, which specifies the average number of times an event occurs in a time interval.
- **PMF:**
$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{for } i = 0, 1, 2, \ldots.$$
- $$\sum_{i=0}^{\infty} \mathbb{P}(X = i) = \sum_{i=0}^{\infty} \frac{\lambda^k e^{-\lambda}}{k!} = e^{-\lambda} \sum_{i=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

  Recall the **Taylor series expansion** from calculus:
$$e^x = \sum_{i=1}^{\infty} \frac{x^i}{i!}.$$

Poisson Distribution

Poisson Distribution

- **Expectation:**

Poisson Distribution

- **Expectation:**

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} i \cdot \mathbb{P}(X = i)$$

Poisson Distribution

- **Expectation:**

$$\mathbb{E}[X] = \sum_{i=0}^{\infty} i \cdot \mathbb{P}(X = i)$$
$$= \sum_{i=1}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!}$$

Poisson Distribution

- **Expectation:**

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=0}^{\infty} i \cdot \mathbb{P}(X = i) \\ &= \sum_{i=1}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \\ &= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}\end{aligned}$$

Poisson Distribution

- **Expectation:**

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{\infty} i \cdot \mathbb{P}(X = i) \\
&= \sum_{i=1}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \\
&= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} e^{\lambda} \qquad \left(e^{\lambda} = \sum_{j=1}^{\infty} \frac{\lambda^j}{j!} \text{ with } j = i-1\right)
\end{aligned}$$

Poisson Distribution

- **Expectation:**

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{i=0}^{\infty} i \cdot \mathbb{P}(X = i) \\
&= \sum_{i=1}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \\
&= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\
&= \lambda e^{-\lambda} e^{\lambda} && \left(e^{\lambda} = \sum_{j=1}^{\infty} \frac{\lambda^j}{j!} \text{ with } j = i - 1\right) \\
&= \lambda.
\end{aligned}$$

Poisson Distribution

- **Expectation:**

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=0}^{\infty} i \cdot \mathbb{P}(X = i) \\ &= \sum_{i=1}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \\ &= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} e^{\lambda} & \left(e^{\lambda} = \sum_{j=1}^{\infty} \frac{\lambda^j}{j!} \text{ with } j = i - 1\right) \\ &= \lambda.\end{aligned}$$

- **Variance:**

Poisson Distribution

- **Expectation:**

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i=0}^{\infty} i \cdot \mathbb{P}(X = i) \\ &= \sum_{i=1}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} \\ &= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda e^{-\lambda} e^{\lambda} \qquad \left(e^{\lambda} = \sum_{j=1}^{\infty} \frac{\lambda^j}{j!} \text{ with } j = i-1\right) \\ &= \lambda.\end{aligned}$$

- **Variance:**

$$\mathrm{Var}(X) = \lambda.$$

Problem Time!