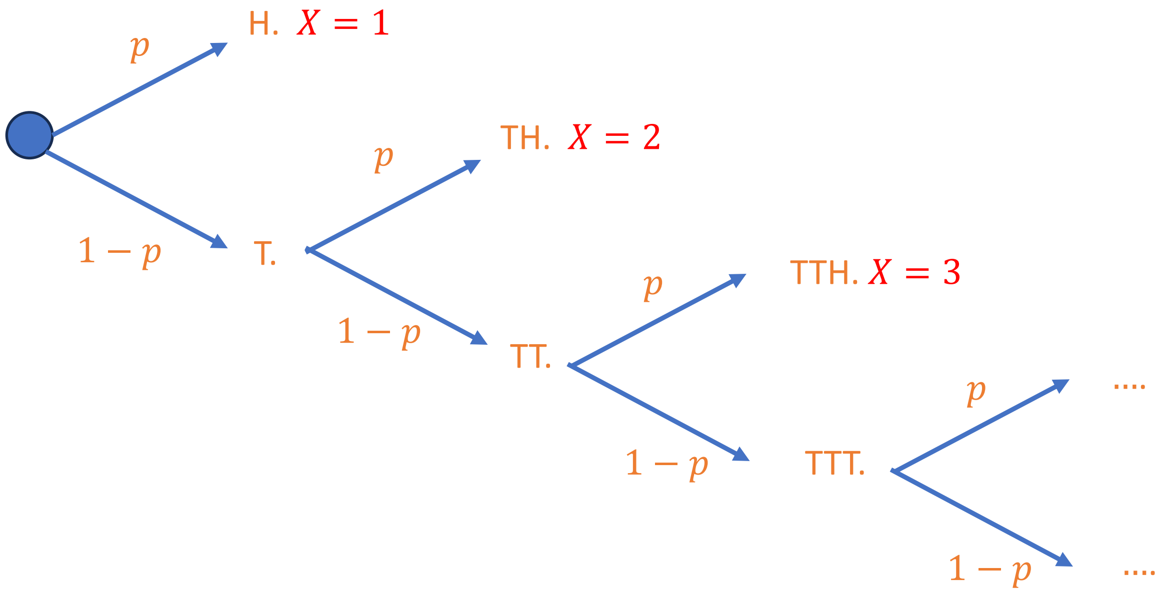
Lecture 17: Random Counts & Famous distributions



Personal takes on how to learn math

- Math is a language.

- Sorted by importance:
  - Definitions
  - Examples
  - Proofs.

- The point of a proof is not only to prove the theorem is true, but more importantly to convey the intuition of why it is true.

Law of total expectation

Theorem (Law of total expectation).

For any event E and variable X,

$$\mathbb{E}[X] = \mathbb{E}[X \mid E] \cdot \mathbb{P}[E] + \mathbb{E}[X \mid \neg E] \cdot \mathbb{P}[\neg E].$$

Law of total expectation

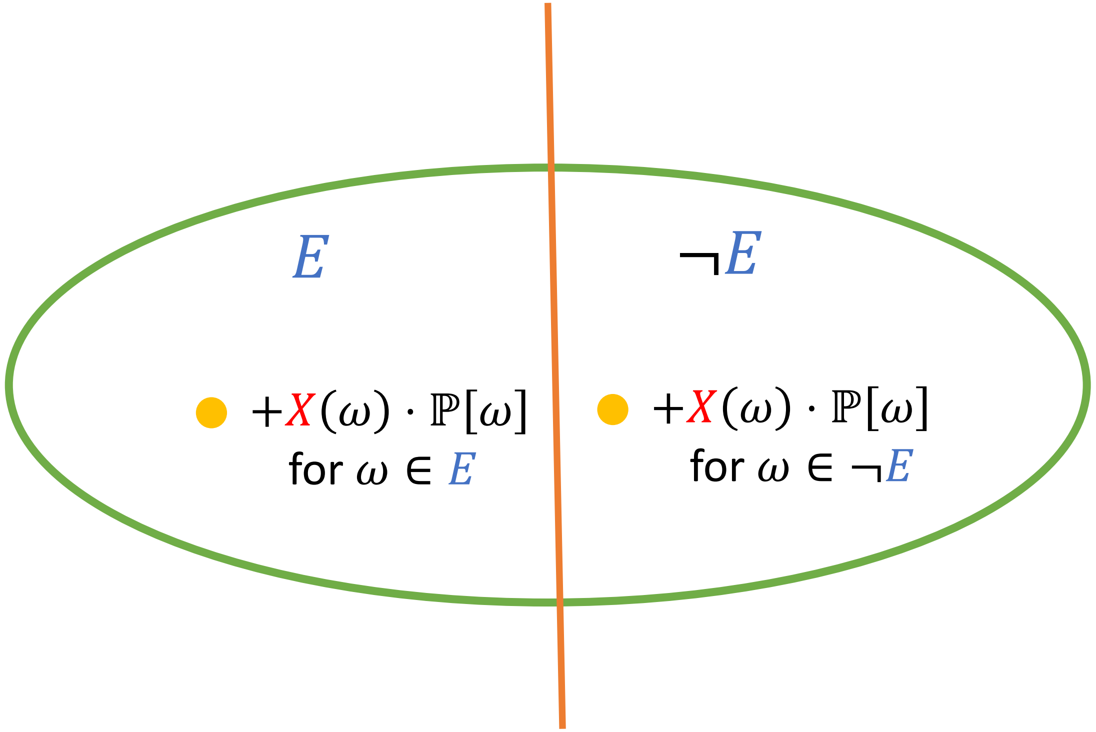
Theorem (Law of total expectation).

For any event E and variable X,

$$\mathbb{E}[X] = \mathbb{E}[X \mid E] \cdot \mathbb{P}[E] + \mathbb{E}[X \mid \neg E] \cdot \mathbb{P}[\neg E].$$

Proof.

$$\begin{aligned}\mathbb{E}[X] &= \textstyle\sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}[\omega] \\ &= \textstyle\sum_{\omega \in \mathrm{E}} X(\omega) \cdot \mathbb{P}[\omega] + \sum_{\omega \in \neg \mathrm{E}} X(\omega) \cdot \mathbb{P}[\omega]\end{aligned}$$



Law of total expectation

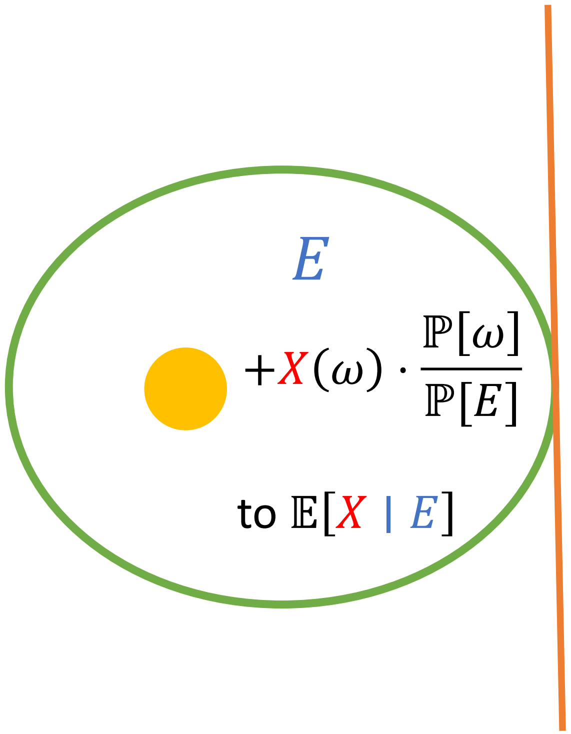
Theorem (Law of total expectation).

For any event E and variable X,

$$\mathbb{E}[X] = \mathbb{E}[X \mid E] \cdot \mathbb{P}[E] + \mathbb{E}[X \mid \neg E] \cdot \mathbb{P}[\neg E].$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \textstyle\sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}[\omega] \\
&= \textstyle\sum_{\omega \in E} X(\omega) \cdot \mathbb{P}[\omega] + \sum_{\omega \in \neg E} X(\omega) \cdot \mathbb{P}[\omega] \\
&= \left(\textstyle\sum_{\omega \in E} X(\omega) \cdot \frac{\mathbb{P}[\omega]}{\mathbb{P}[E]}\right) \cdot \mathbb{P}[E] \\
&\quad + \left(\textstyle\sum_{\omega \in \neg E} X(\omega) \cdot \frac{\mathbb{P}[\omega]}{\mathbb{P}[\neg E]}\right) \cdot \mathbb{P}[\neg E] \\
&= \mathbb{E}[X \mid E] \cdot \mathbb{P}[E] + \mathbb{E}[X \mid \neg E] \cdot \mathbb{P}[\neg E]
\end{aligned}$$



Law of total expectation

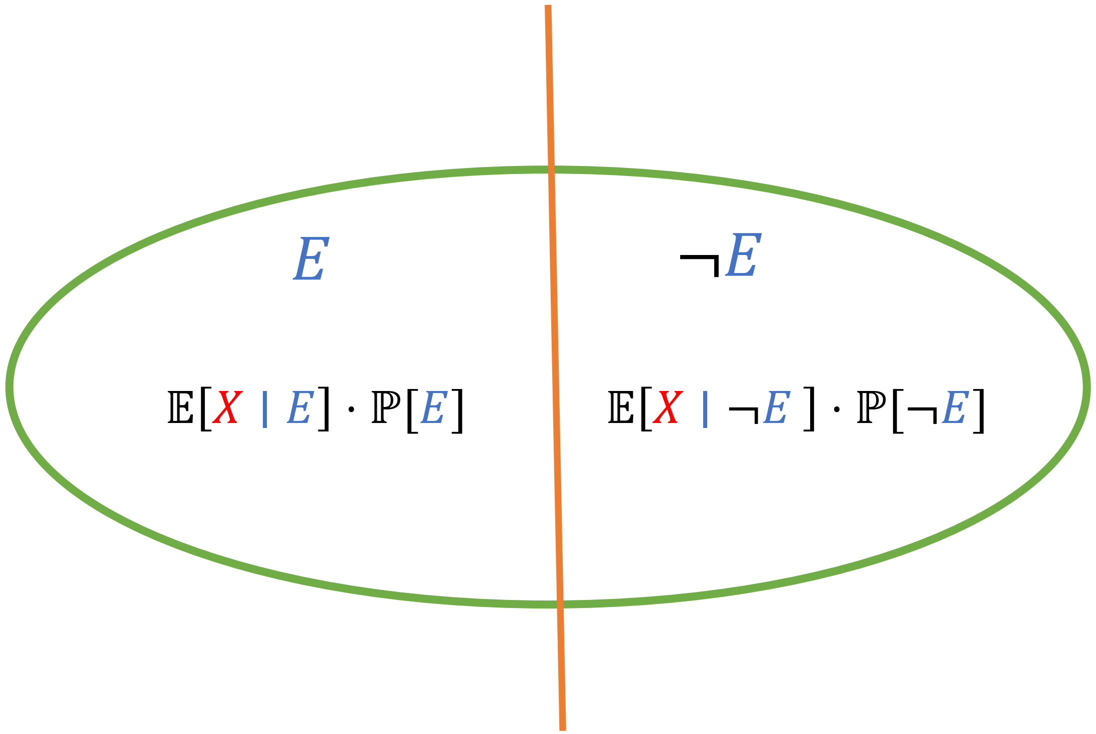
Theorem (Law of total expectation).

For any event E and variable X,

$$\mathbb{E}[X] = \mathbb{E}[X \mid E] \cdot \mathbb{P}[E] + \mathbb{E}[X \mid \neg E] \cdot \mathbb{P}[\neg E].$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{\omega \in \Omega} X(\omega) \cdot \mathbb{P}[\omega] \\
&= \sum_{\omega \in \mathrm{E}} X(\omega) \cdot \mathbb{P}[\omega] + \sum_{\omega \in \neg \mathrm{E}} X(\omega) \cdot \mathbb{P}[\omega] \\
&= \left(\sum_{\omega \in \mathrm{E}} X(\omega) \cdot \frac{\mathbb{P}[\omega]}{\mathbb{P}[E]}\right) \cdot \mathbb{P}[E] \\
&\quad + \left(\sum_{\omega \in \neg \mathrm{E}} X(\omega) \cdot \frac{\mathbb{P}[\omega]}{\mathbb{P}[\neg E]}\right) \cdot \mathbb{P}[\neg E] \\
&= \mathbb{E}[X \mid E] \cdot \mathbb{P}[E] + \mathbb{E}[X \mid \neg E] \cdot \mathbb{P}[\neg E]
\end{aligned}$$

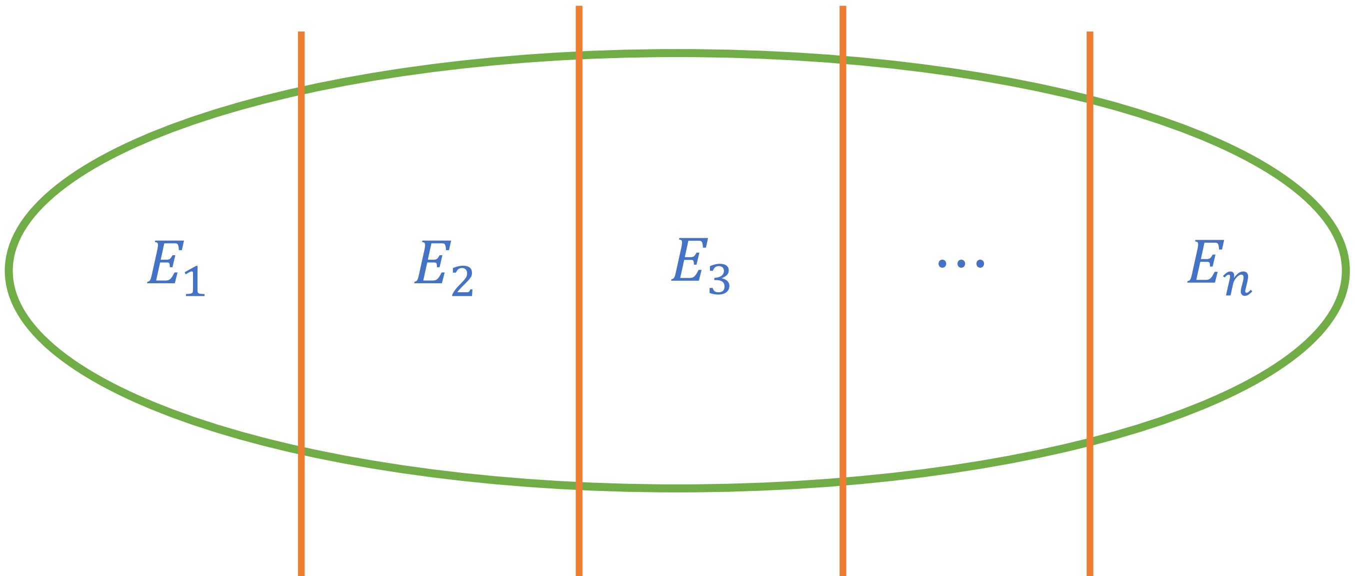


Law of total expectation

Theorem (Law of total expectation).

For any disjoint event $E_1, E_2, \ldots E_n$ that covers all possibilities (i.e., $E_1 \cup E_2, \cup \cdots \cup E_n = \Omega$) and variable X,

$$\mathbb{E}[X] = \mathbb{E}[X \mid E_1] \cdot \mathbb{P}[E_1] + \mathbb{E}[X \mid E_2] \cdot \mathbb{P}[E_2] + \cdots \mathbb{E}[X \mid E_n] \cdot \mathbb{P}[E_n].$$



Function of random variable

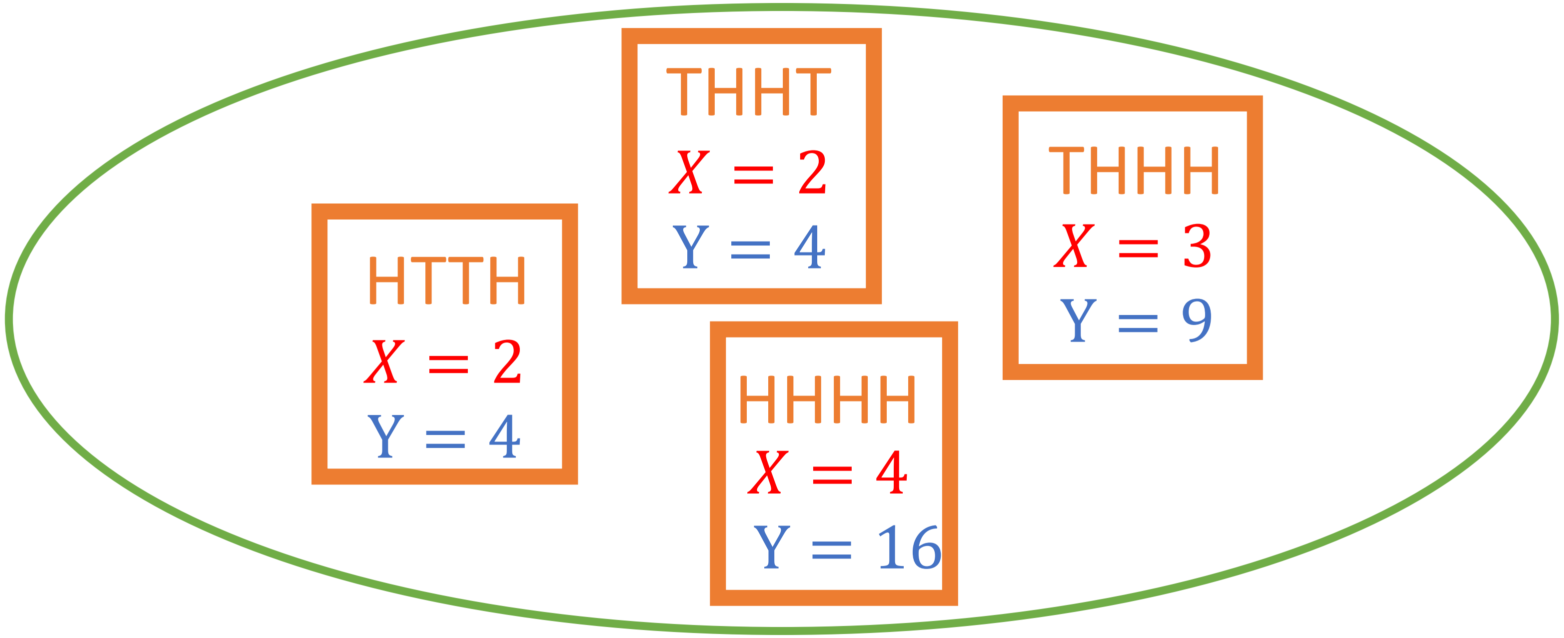
Definition.

If $X: \Omega \to \mathbb{R}$ is a random variable, then we can define another random variable $Y = f(X)$ for function f.

For every outcome ω, it has a value. $f(X(\omega))$.

Example.

$Y = X^2$



Function of random variable

LOTUS (Law of the unconscious Statistician).

If $X: \Omega \to \mathbb{R}$ is a random variable, then we can define another random variable $Y = f(X)$ for function f.

We can also talk about the expectation of that random variable.

$$\mathbb{E}[f(X)] = \sum_{a} f(a) \cdot \mathbb{P}[X = a]$$

Law of iterated expectation

Lemma.

For any two random variables X and Y,

$$\mathbb{E}_Y[\mathbb{E}_X[X|Y]] = \mathbb{E}[X]$$

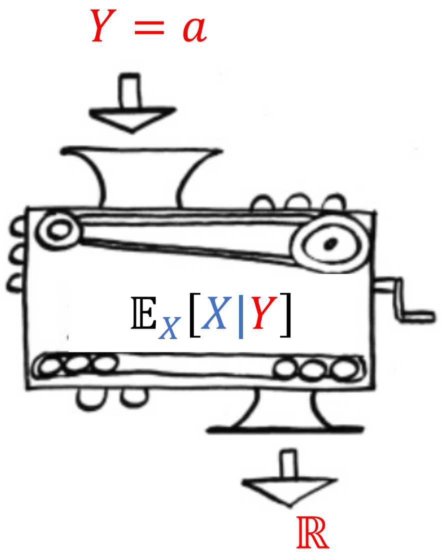
Or,

$$\sum_b \mathbb{E}_X[X|Y=b] \cdot \Pr[Y=b] = \mathbb{E}[X]$$

Break it down:

For $\mathbb{E}_X[X|Y]$, Y is a free variable.

This is a function about Y.



Law of iterated expectation

Lemma.

For any two random variables X and Y,

$$\mathbb{E}_Y[\mathbb{E}_X[X|Y]] = \mathbb{E}[X]$$

Or,

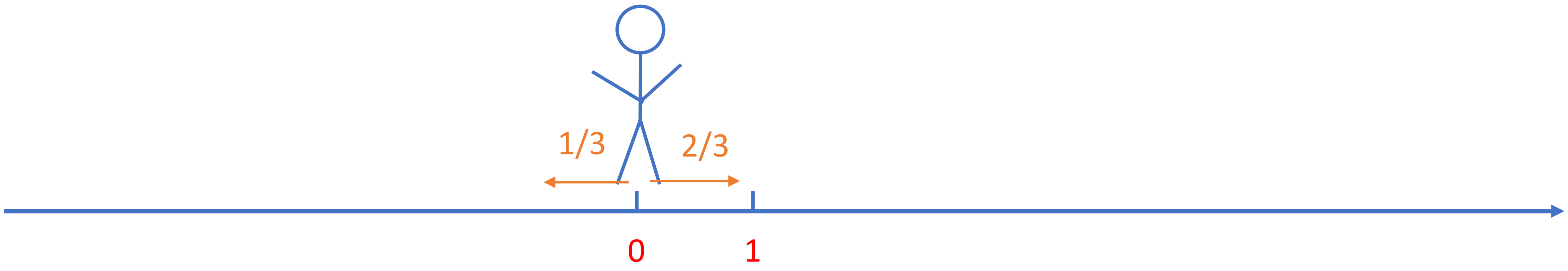
$$\sum_b \mathbb{E}_X[X|Y = b] \cdot \Pr[Y = b] = \mathbb{E}[X]$$

Break it down:

Let $f(b) = \mathbb{E}_X[X|Y = b]$.

$\mathbb{E}_Y[\mathbb{E}_X[X|Y]]$ simply means $\mathbb{E}[f(Y)]$.

Recall the random walk



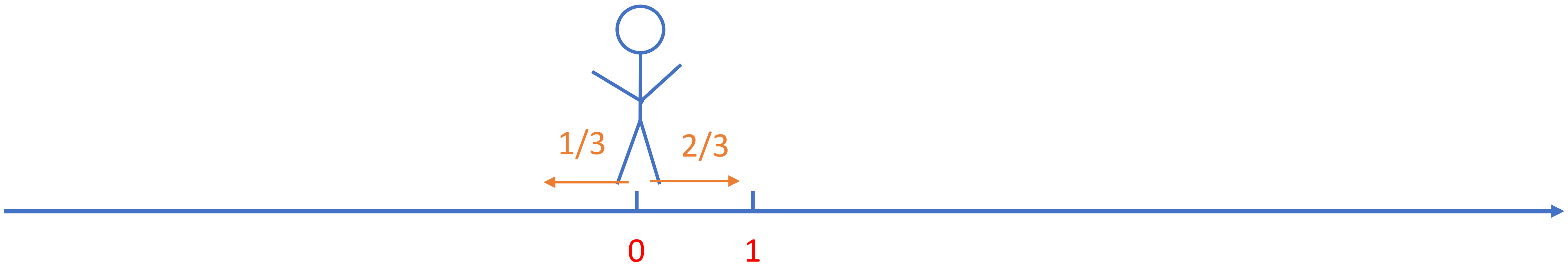
On a number axis that is infinitely long on both ends, you start from 0.

Each step:

With probability 2/3, you walk length one right.

With probability 1/3, you walk length one left.

Recall the random walk



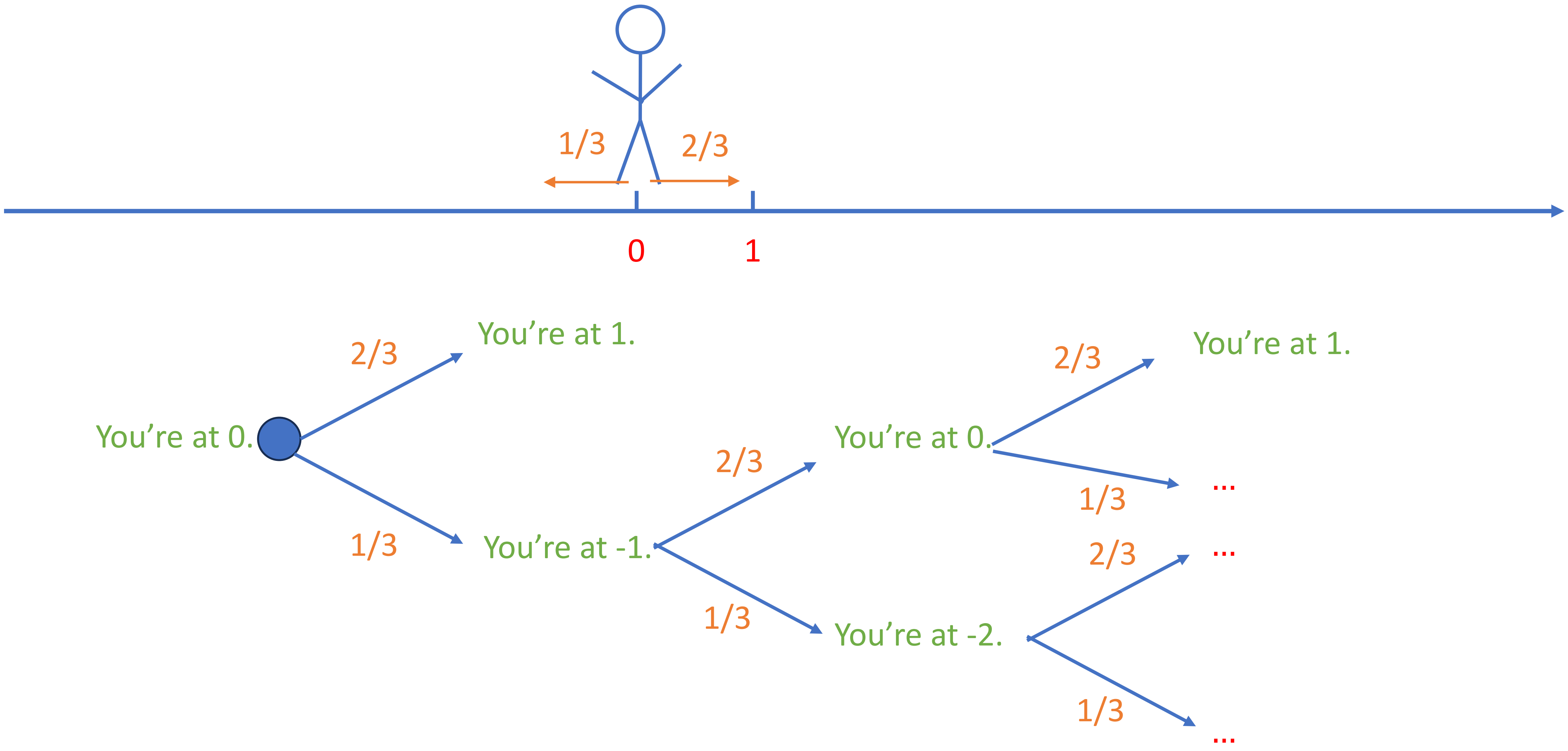
Initially, you are at $x_0 = 0$.

Each step:

With probability 2/3, $x_t = x_{t-1} - 1$.

With probability 1/3, $x_t = x_{t-1} + 1$.

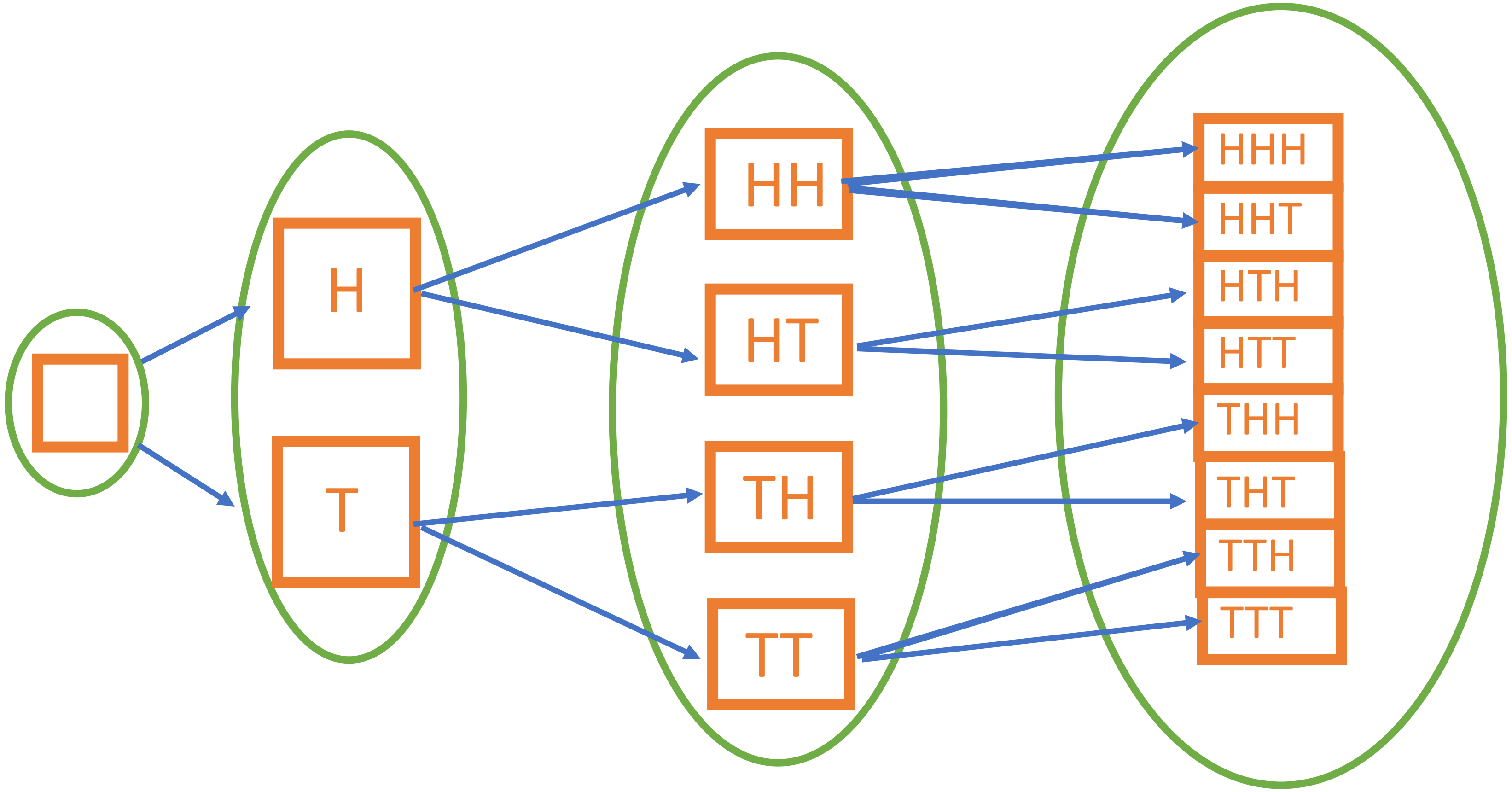
Recall the random walk



Random process (Intuition)

Probability Space

We can intuitively think of random process as evolving possibilities.



Bernoulli Process

Definition

A Bernoulli process with parameter p is the process of tossing coins, $c_1, c_2, \cdots, c_i, \cdots \in \{H, T\}$ where $c_i = H$ independently with probability p.

… … …

Let’s count things in this process!

Bernoulli Distribution

Random Variable

In a Bernoulli process with parameter p, let X be result of a single coin.
(X=1 for H, X=0 for T)

Distribution

$X = 0$ with probability 1-p.
$X = 1$ with probability p.

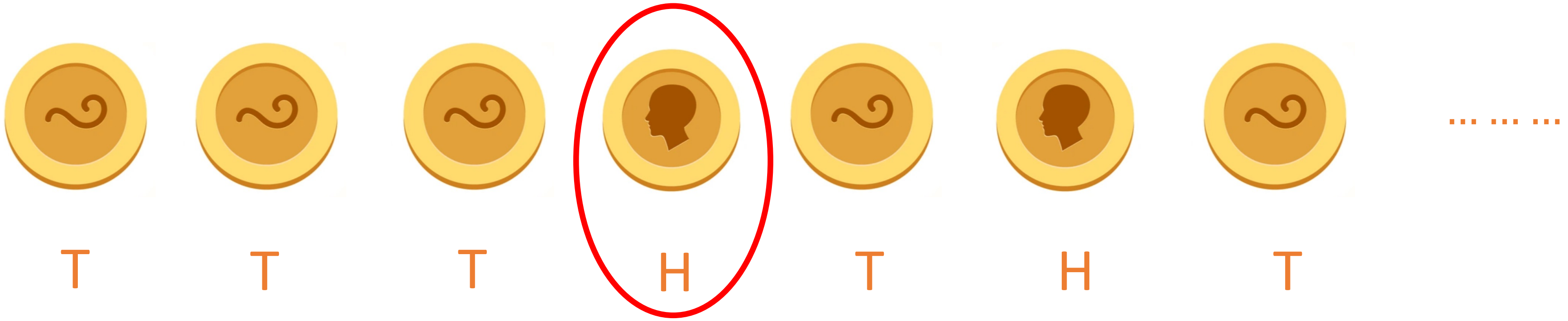
Expectation

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

Geometric Distribution

Random Variable

In a Bernoulli process with parameter p, let X be the position of the first head.



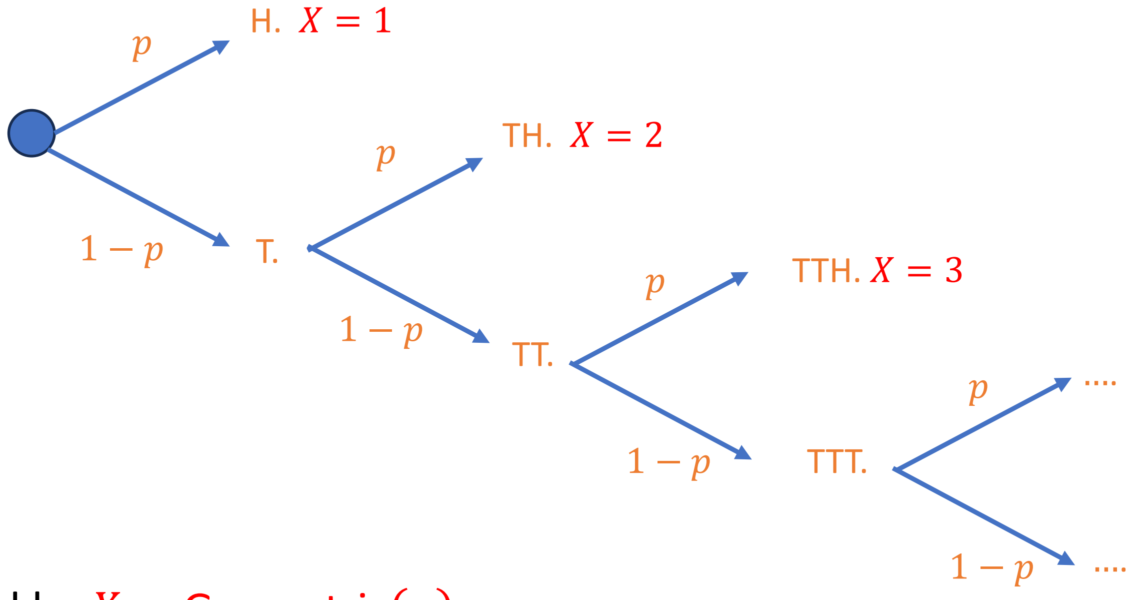
In this example, $X = 4$.

Geometric Distribution

Random Variable

In a Bernoulli process with parameter p, let X be the position of the first head.

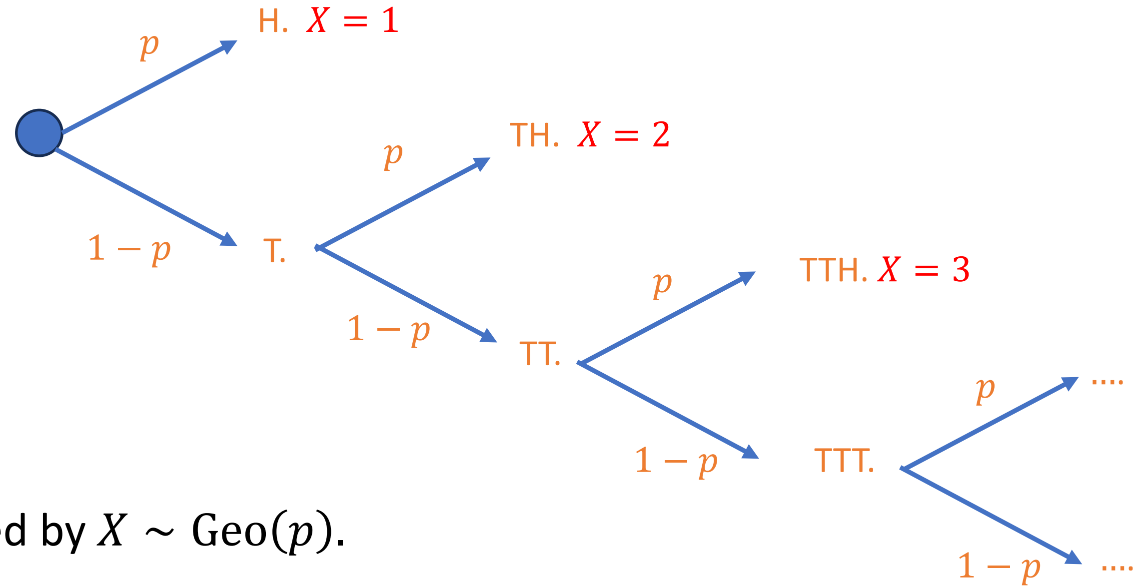
Distribution



Probability that $X = 1$: p
Probability that $X = 2$: $(1 - p) \cdot p$.
Probability that $X = 3$: $(1 - p)^2 \cdot p$.
....
Probability that $X =$ i : $(1 - p)^{i-1} \cdot p$.

This is called the Geometric distribution. Denoted by $X \sim \text{Geometric}(p)$.

Geometric Distribution



Distribution

Probability that $X = 1$: p
Probability that $X = 2$: $(1-p) \cdot p$.
Probability that $X = 3$: $(1-p)^2 \cdot p$.
....
Probability that $X = \mathrm{i}$: $(1-p)^{i-1} \cdot p$.

This is called the Geometric distribution. Denoted by $X \sim \mathrm{Geo}(p)$.

Check sum = 1

$$\sum_{i=1}^{\infty}(1-p)^{i-1} \cdot p = \frac{1}{1-(1-p)} \cdot p = 1$$

In the first step, we used
$1 + x + x^2 + \cdots = \frac{1}{1-x}$. This is true for all $-1 < x < 1$.

Geometric Distribution

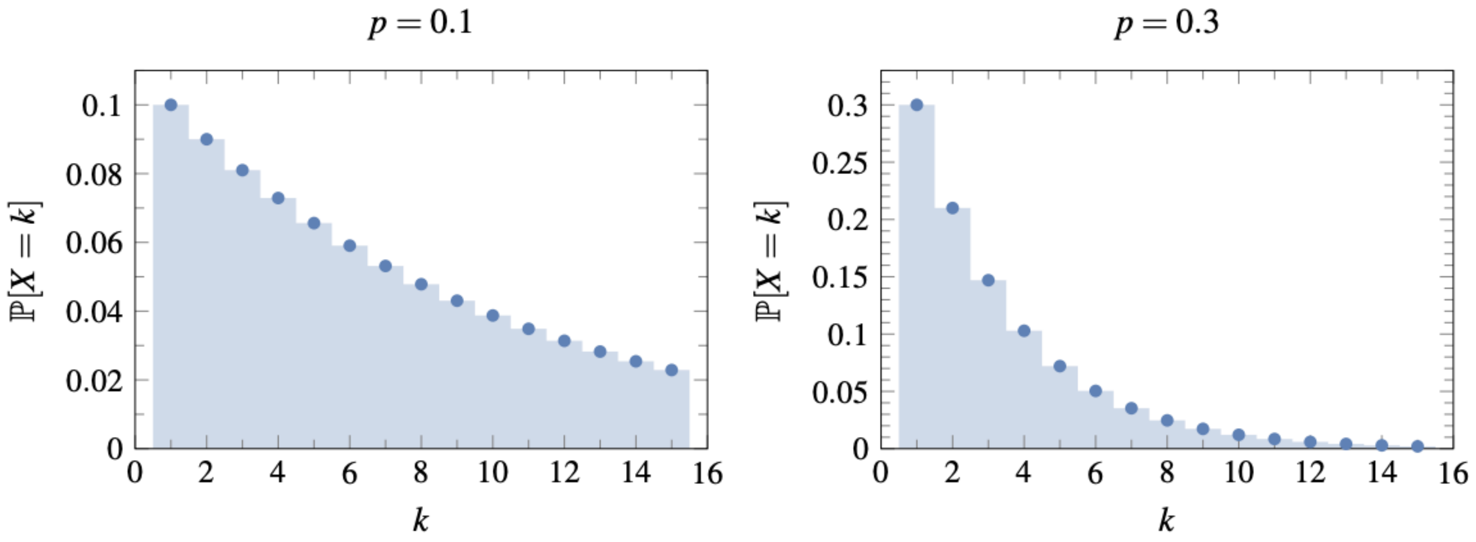
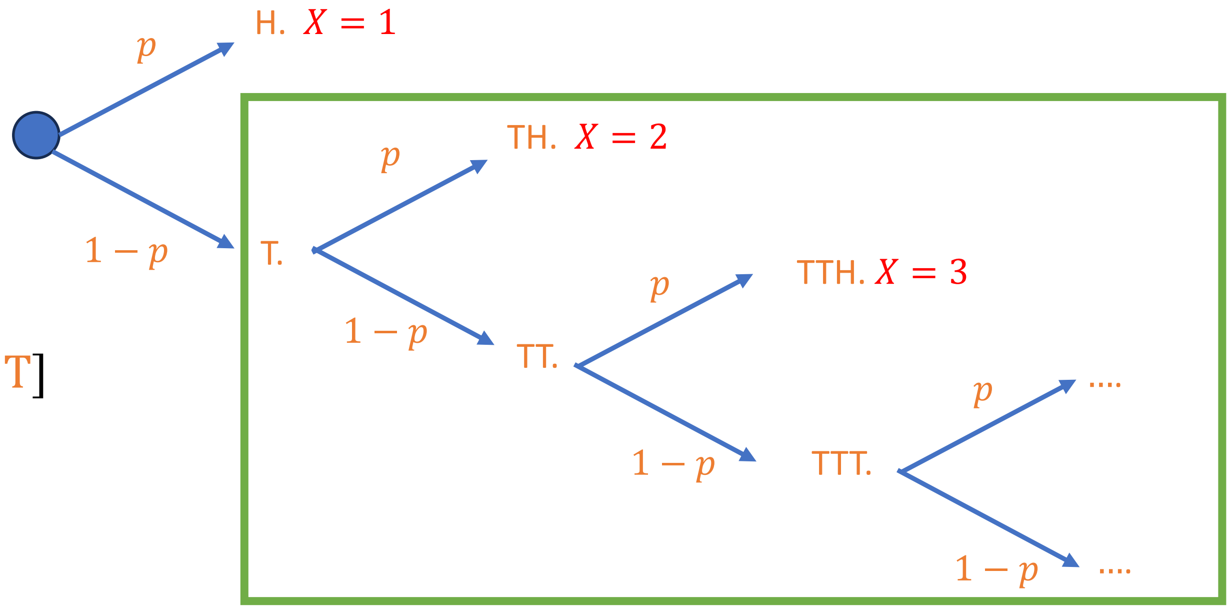


Figure 1: Illustration of the Geometric(p) distribution for $p = 0.1$ and $p = 0.3$.

Geometric Distribution



Expectation

Solution 1: The self-referencing trick.

$$\mathbb{E}[X] = p \cdot 1 + (1-p) \cdot \mathbb{E}[X \mid \text{first coin is T}]$$
$$= p + (1-p)\,(1 + \mathbb{E}[X])$$

$$\mathbb{E}[X] = \frac{1}{p}.$$

Geometric Distribution

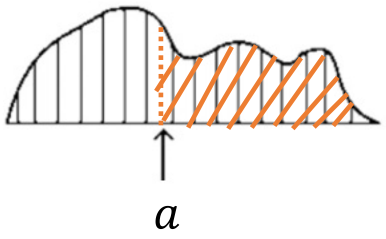
Expectation

Solution 2: The alternative formula for expectation.

Suppose $X \in \mathbb{Z}_+$.

Formula 1 (def): $\mathbb{E}[X] = \sum_{a \in \mathbb{Z}_+} a \cdot \mathbb{P}[X = a]$.

Formula 2 (new): $\mathbb{E}[X] = \sum_{a \in \mathbb{Z}_+} \mathbb{P}[X \geq a]$.



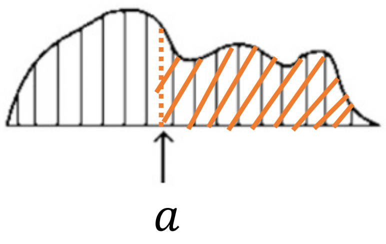
$$\mathbb{P}[X \geq 1] + \mathbb{P}[X \geq 2] + \mathbb{P}[X \geq 3] + \cdots$$
$$= (\mathbb{P}[X = 1]) + (\mathbb{P}[X = 1] + \mathbb{P}[X = 2]) + (\mathbb{P}[X = 1] + \mathbb{P}[X = 2] + \mathbb{P}[X = 3]) + \cdots$$
$$= 1 \cdot \mathbb{P}[X = 1] + 2 \cdot \mathbb{P}[X = 2] + 3 \cdot \mathbb{P}[X = 2] + \cdots$$

Geometric Distribution

Expectation

Solution 2: The alternative formula for expectation.

Suppose $X \in \mathbb{Z}_+$.



Formula 2 (new): $\mathbb{E}[X] = \sum_{a \in \mathbb{Z}_+} \mathbb{P}[X \geq a]$

$= \sum_{a \in \mathbb{Z}_+} (1-p)^{a-1}$ (first a-1 coins being T)

$= \frac{1}{1-(1-p)}$

$= \frac{1}{p}$

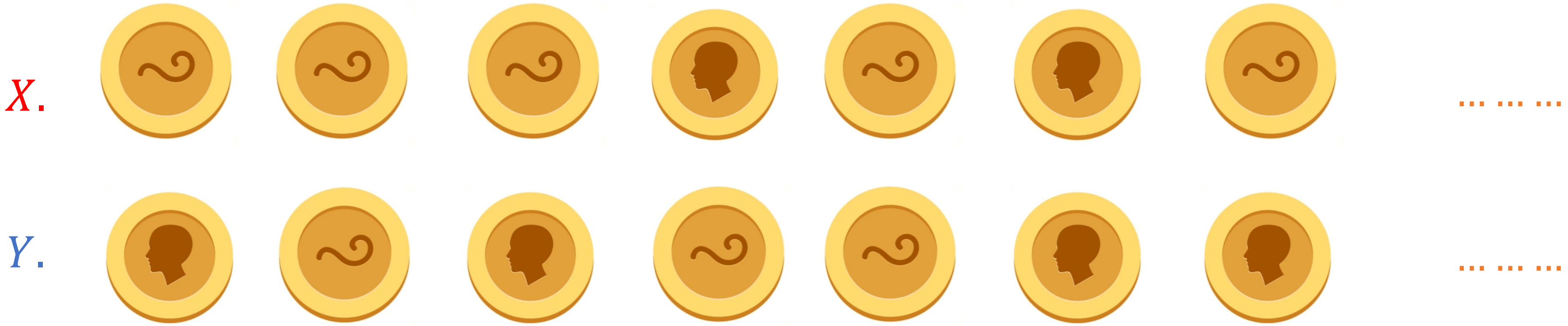
Geometric Distribution

Exercise 1

Let $X, Y \sim \text{Geometric}(p)$. What is $\mathbb{E}[\min(X, Y)]$?

A:

Consider the Bernoulli process:



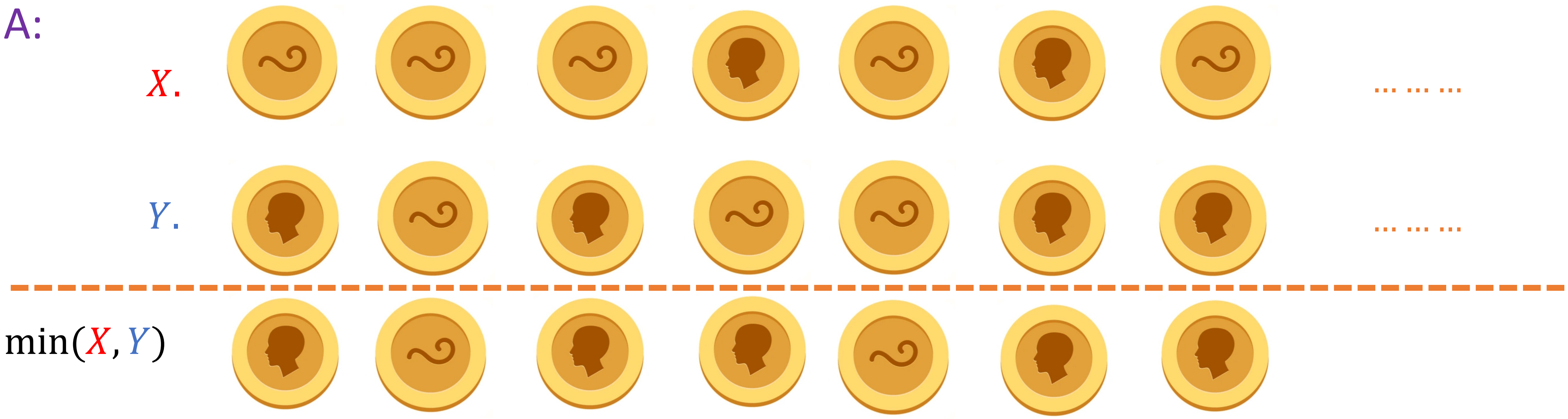
$\min(X, Y)$ is tossing two coins in each step & looking for first head.

Geometric Distribution

Exercise 1

Let $X, Y \sim \text{Geometric}(p)$. What is $\mathbb{E}[\min(X, Y)]$?

A:



This is the same as tossing a single coin with probability $1 - (1 - p)^2$

Geometric Distribution

Exercise 1

Let $X, Y \sim \text{Geometric}(p)$. What is $\mathbb{E}[\min(X, Y)]$?

A:

This is the same as tossing a single coin with probability $1 - (1-p)^2$

$$\mathbb{E}[\min(X, Y)] = \frac{1}{1 - (1-p)^2}$$

Geometric Distribution

Exercise 2

Let $X, Y \sim \mathrm{Geometric}(p)$. What is $\mathbb{E}[\max(X, Y)]$?

Geometric Distribution

Exercise 2

Let $X, Y \sim \text{Geometric}(p)$. What is $\mathbb{E}[\max(X, Y)]$?

Inclusion Exclusion for expectation

Let X, Y be random variables. We have

$$\mathbb{E}[\max(X, Y)] = \mathbb{E}[X] + \mathbb{E}[Y] - \mathbb{E}[\min(X, Y)]$$

Why?

$\max(X, Y) + \min(X, Y) = X + Y$ holds for all outcome.

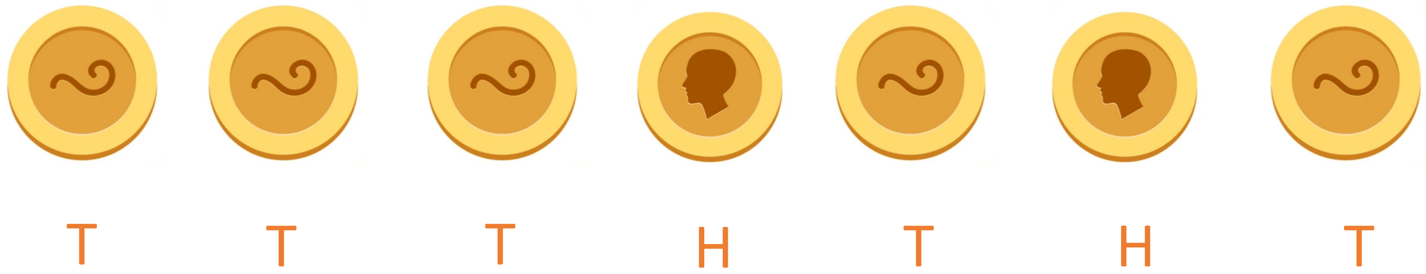
Then use linearity of expectation.

Binomial Distribution

Random Variable

In a Bernoulli process with parameter p with n coins,

let X be the total number of heads.



In this example, $n = 7$ and $X = 2$.

Binomial Distribution

Random Variable

In a Bernoulli process with parameter p with n coins,

let X be the total number of heads.

Distribution

$$\mathbb{P}[X = k] = \#\text{outcomes with } k \text{ heads} \cdot \mathbb{P}[\text{one such outcome}]$$
$$= \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

Denoted by $X \sim \text{Binomial}(n, p)$.

Binomial Distribution

Distribution

$$\mathbb{P}[X = k] = \#\text{outcomes with } k \text{ heads} \cdot \mathbb{P}[\text{one such outcome}]$$
$$= \binom{n}{k} \cdot p^k (1-p)^{n-k}$$

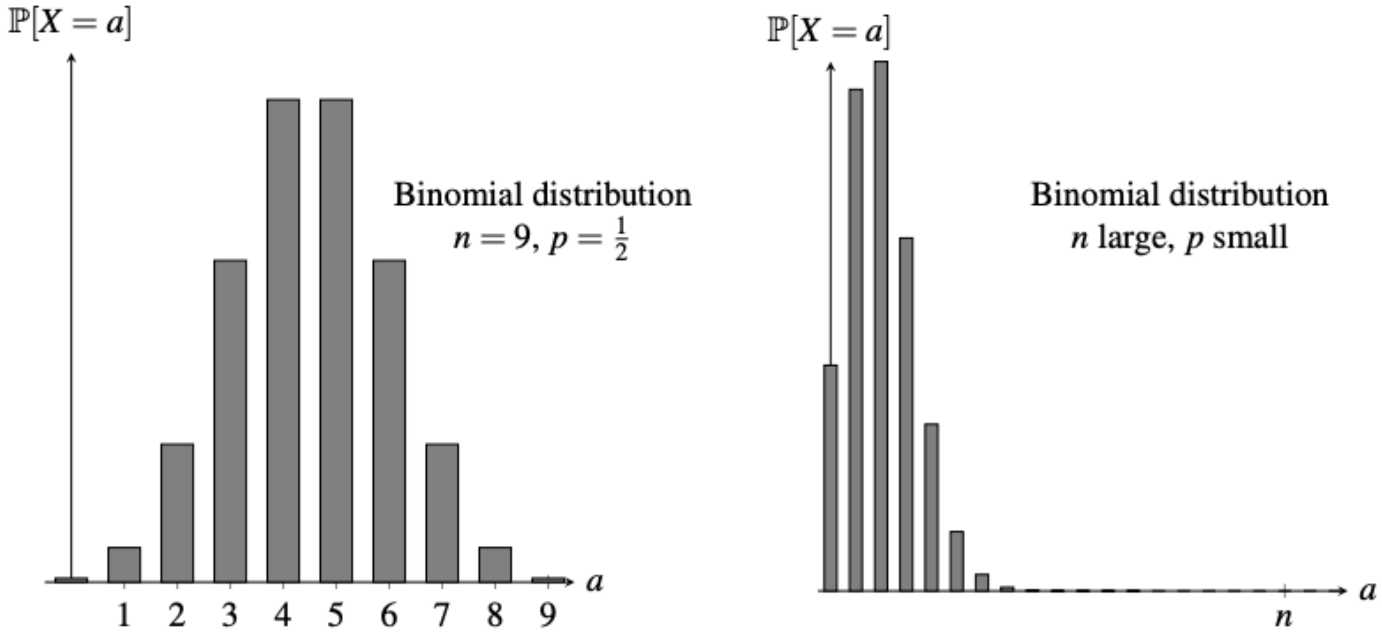


Figure 3: The binomial distributions for two choices of (n, p).

Binomial Distribution

Expectation

Let $X_i = \mathbf{1}[\text{the } i - \text{th coin is head}]$ be the indicator variable.

$X = X_1 + X_2 + X_3 + \cdots + X_n$ in any possible world.

By linearity of expectation,

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X_1] + \mathbb{E}[X_2] + \mathbb{E}[X_3] + \cdots + \mathbb{E}[X_n] \\ &= p + p + p + \cdots + p \\ &= np\end{aligned}$$

Poisson Process

Intuition

In reality, not everything is discrete.

E.g. When you walk in Tenderloin, SF around 10 pm, you could get robbed anytime.

It could happen at t = 10 : 00 : 3.141516……..

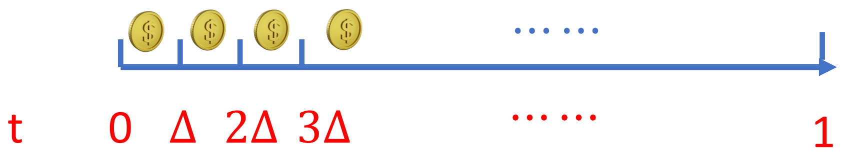
Or in a McDonald, people could walk in anytime.

How do we model this?

Poisson Process

Intuition (Discretization)

Let's take a unit time interval and divide it into $n \to \infty$ segments .



For each segment with length $\Delta = 1/n \to 0$,

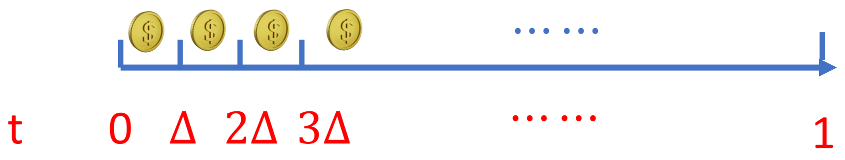
we view it as a coin with probability λ/n of being head.

This process is the limit of $\text{Binomial}\left(n, \frac{\lambda}{n}\right)$ when $n \to \infty$.

Poisson Process

Intuition (Discretization)

Let's take a unit time interval and divide it into $n \to \infty$ segments .



This process is the limit of Binomial $\left(n, \frac{\lambda}{n}\right)$ when $n \to \infty$.

$$\mathbb{P}[X = i] = \lim_{n \to \infty} \binom{n}{k} \cdot \left(\frac{\lambda}{n}\right)^{i} \cdot \left(1 - \frac{\lambda}{n}\right)^{n-i}$$

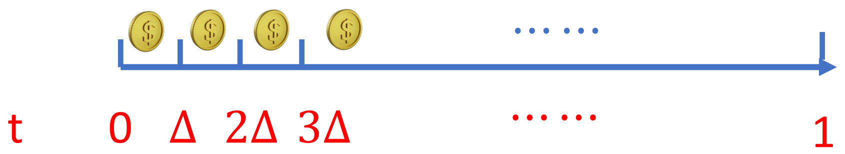
$$= \frac{\lambda^{i}}{i!} \cdot e^{-\lambda}$$

(If you are interested, see notes for the calculation.)

Poisson Distribution

Intuition (Discretization)

Let's take a unit time interval and divide it into $n \to \infty$ segments .



Let X be the random variable denoting the number of heads in this interval.

Its distribution is the limit of Binomial $\left(n, \frac{\lambda}{n}\right)$ when $n \to \infty$.

We call it Poisson distribution.

Poisson Distribution

Definition

A variable X is said to obey Poisson distribution with rate λ ($X \sim \text{Poisson}(\lambda)$) if

$$\mathbb{P}[X = i] = \frac{\lambda^i}{i!} \cdot e^{-\lambda}$$

Expectation

$$\mathbb{E}[X] = \lim_{n \to \infty} n \cdot \frac{\lambda}{n} = \lambda.$$

Sum of Poisson variables.

Lemma

If $X \sim \text{Poisson}(\lambda)$ and $\text{Y} \sim \text{Poisson}(\lambda)$, then their sum

$$\text{X} + \text{Y} \sim \text{Poisson}(2\lambda)$$

Proof.

What are X and Y? They are the number of heads in a unit interval.

What is $\text{X} + \text{Y}$? Number of heads in time two.

We can speed up the clock by 2x. The rate at which head occurs x2.

Sum of Poisson variables.

Lemma

If $X \sim \text{Poisson}(\lambda_1)$ and $\text{Y} \sim \text{Poisson}(\lambda_2)$, then their sum

$$\text{X} + \text{Y} \sim \text{Poisson}(\lambda_1 + \lambda_2)$$

Proof.

Try to convince yourself. Or look at the notes.

The first head?

Q:

In a Poisson process with rate λ, what is the probability that the first head occurs after time t?

A:

For a fixed time t, let X be the number of heads during this time.

$X \sim \text{Poisson}(\lambda \cdot t)$

Thus $\mathbb{P}[X = 0] = \frac{(\lambda t)^{-0}}{0!} \cdot e^{-\lambda t} = e^{-\lambda t}$.