Day - 57, Jan - 26, 2026 (Magh 13, 2082 B.S.).

An Example

What is the probability that if I flip 5 coins, 2 of them land in heads?

$$0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \quad = \frac{1}{32}$$

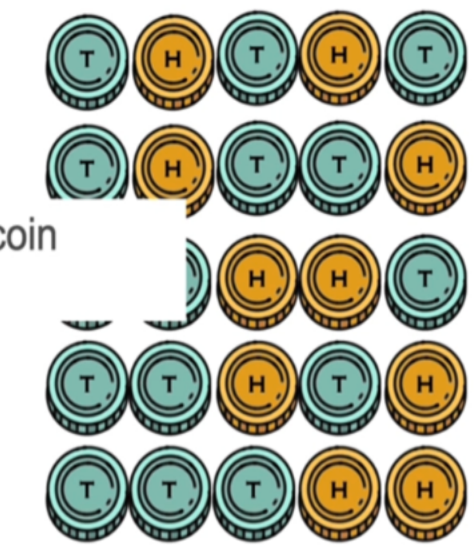
$$0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \cdot 0.5 \quad = \frac{1}{32}$$

DeepLearning.AI

An Example

What is the probability that if I flip 5 coins, 2 of them land in heads?

10 ways to have 2 heads in 5 coin tosses



DeepLearning.AI

An Example

What is the probability that if I flip 5 coins, 2 of them land in heads?

$$10 = \frac{5!}{2!(5-2)!} = \binom{5}{2}$$

Binomial coefficient

Number of ways you can get 2 heads in 5 coin tosses

Binomial Coefficient

In general:

$\binom{n}{k}$ counts all the combinations for landing k heads in n coin tosses

Property:

$$\binom{n}{k} = \binom{n}{n-k}$$

k heads is the same thing as obtaining n-k tails.

Binomial Coefficient

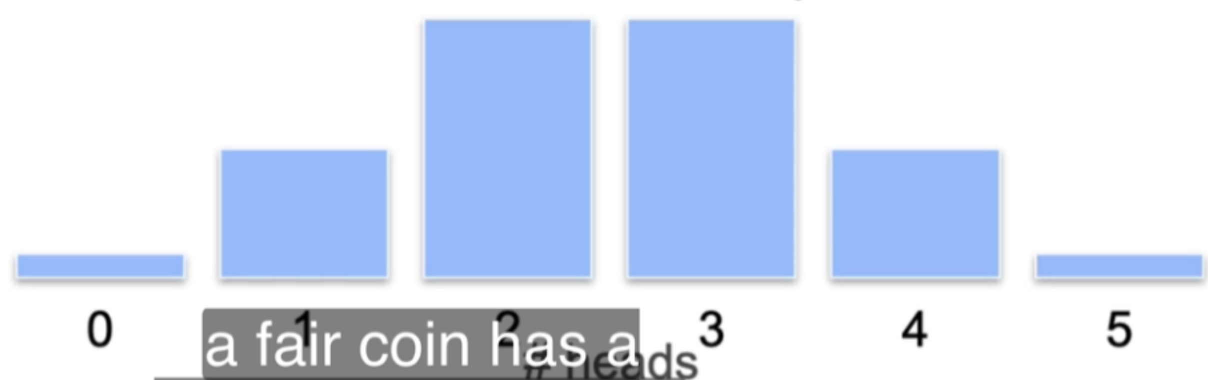
In general:

$\binom{n}{k}$ counts all the combinations for landing k heads in n coin tosses

Property:

$$\binom{n}{k} = \binom{n}{n-k}$$

The PMF with a fair coin is symmetrical



a fair coin has a symmetrical shape.

Binomial Distribution

General PMF for X : number of heads in 5 coin tosses?

Your coin has $\mathbf{P}(H) = p$

Event: $X = x$: x heads in 5 tosses

Probability of seeing $5 - x$ tails

All the possible orders

$$\boxed{\binom{5}{x}}\boxed{p^x}\boxed{(1-p)^{5-x}}$$

Probability of seeing x heads

so you need to take into account
all the possible orders,

Binomial Distribution

General PMF for X : number of heads in 5 coin tosses?

Your coin has $\mathbf{P}(H) = p$

Event: $X = x$: x heads in 5 tosses

$$p_X(x) = \binom{5}{x} p^x (1-p)^{5-x}, \quad x = 0,1,2,3,4,5$$

X follows a binomial distribution

$$X \sim Binomial(\boxed{5},\boxed{p})$$

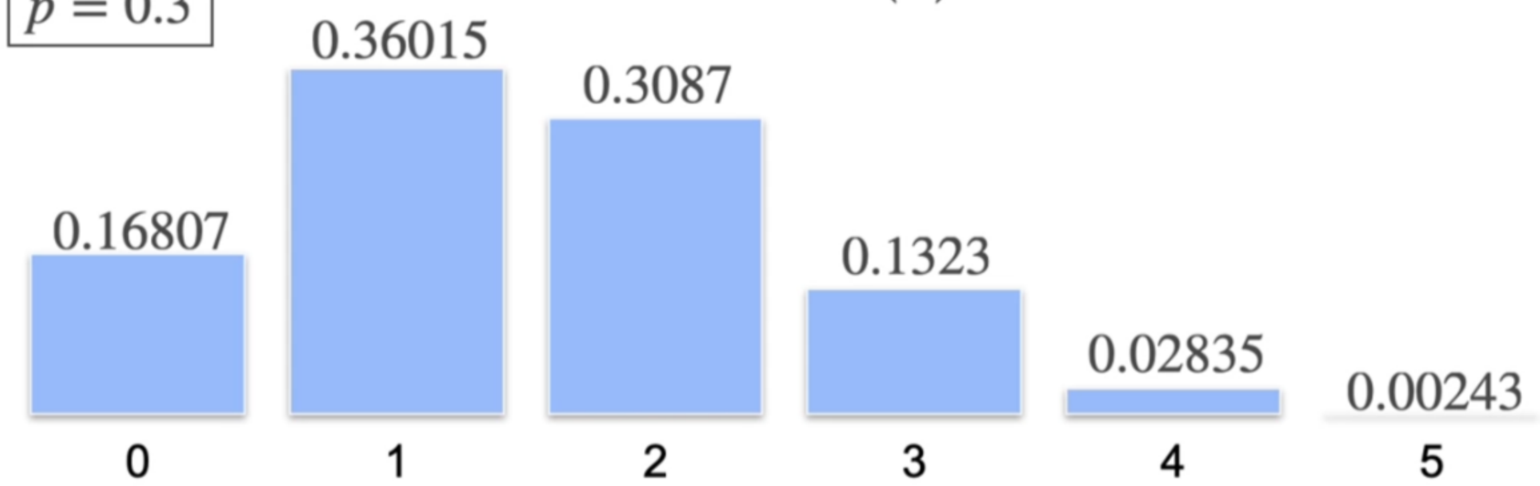
Number of flips $\qquad \mathbf{P}(H)$

We denote it as
binomial of 5 and p,

Binomial Distribution

$$\boxed{\begin{array}{c} n = 5 \\ p = 0.3 \end{array}}$$

$$p_X(x) = \mathbf{P}(X = x) = \binom{5}{k} 0.3^k 0.7^{5-k}$$



greater chances of seeing
a small number of heads,

DeepLearning.AI

Binomial Distribution

General PMF for X : number of heads in n coin tosses?

Your coin has $\mathbf{P}(H) = p$

Event: $X = x$: x heads in n tosses

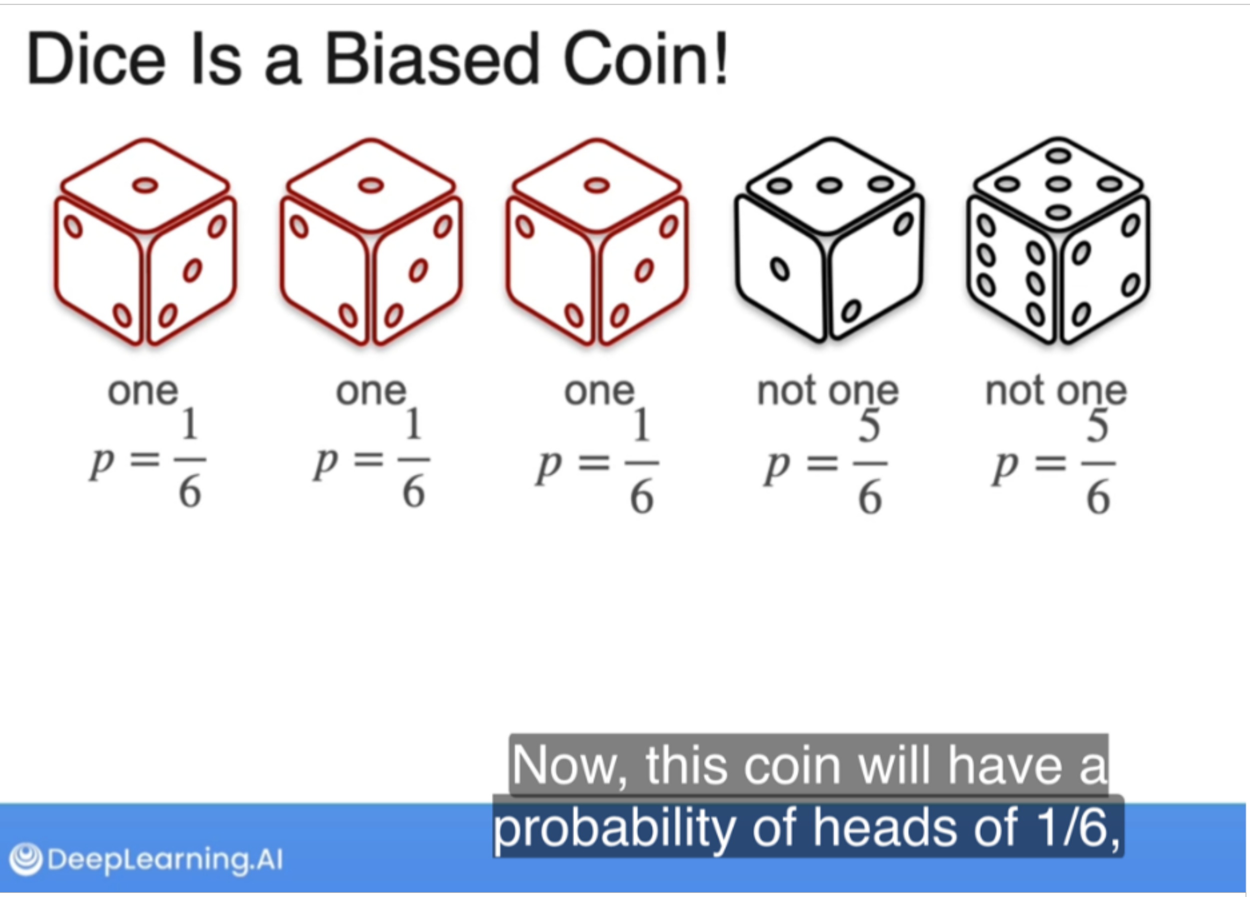
$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,1,2,\ldots,n$$

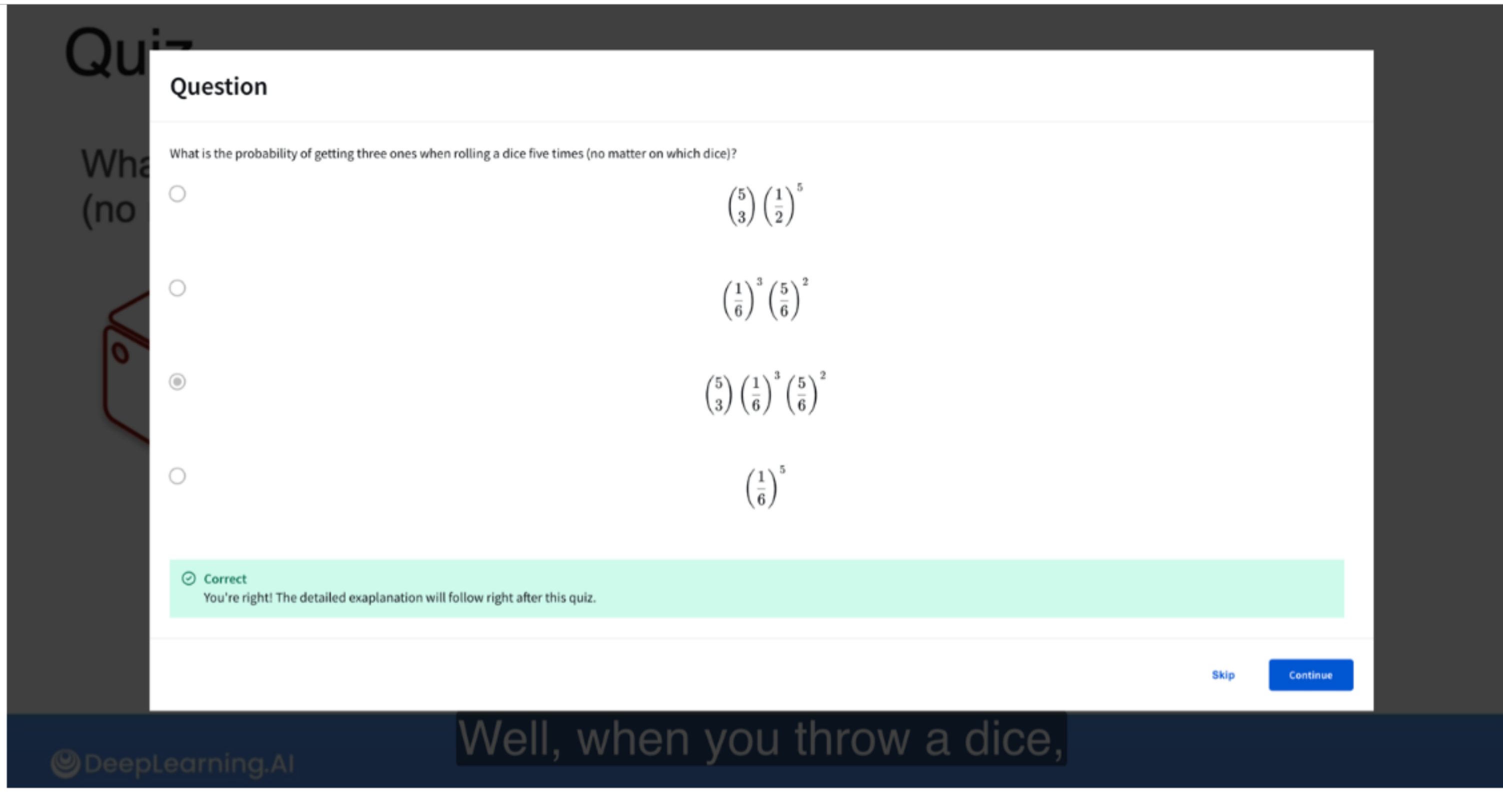
$X \sim Binomial(n, p)$

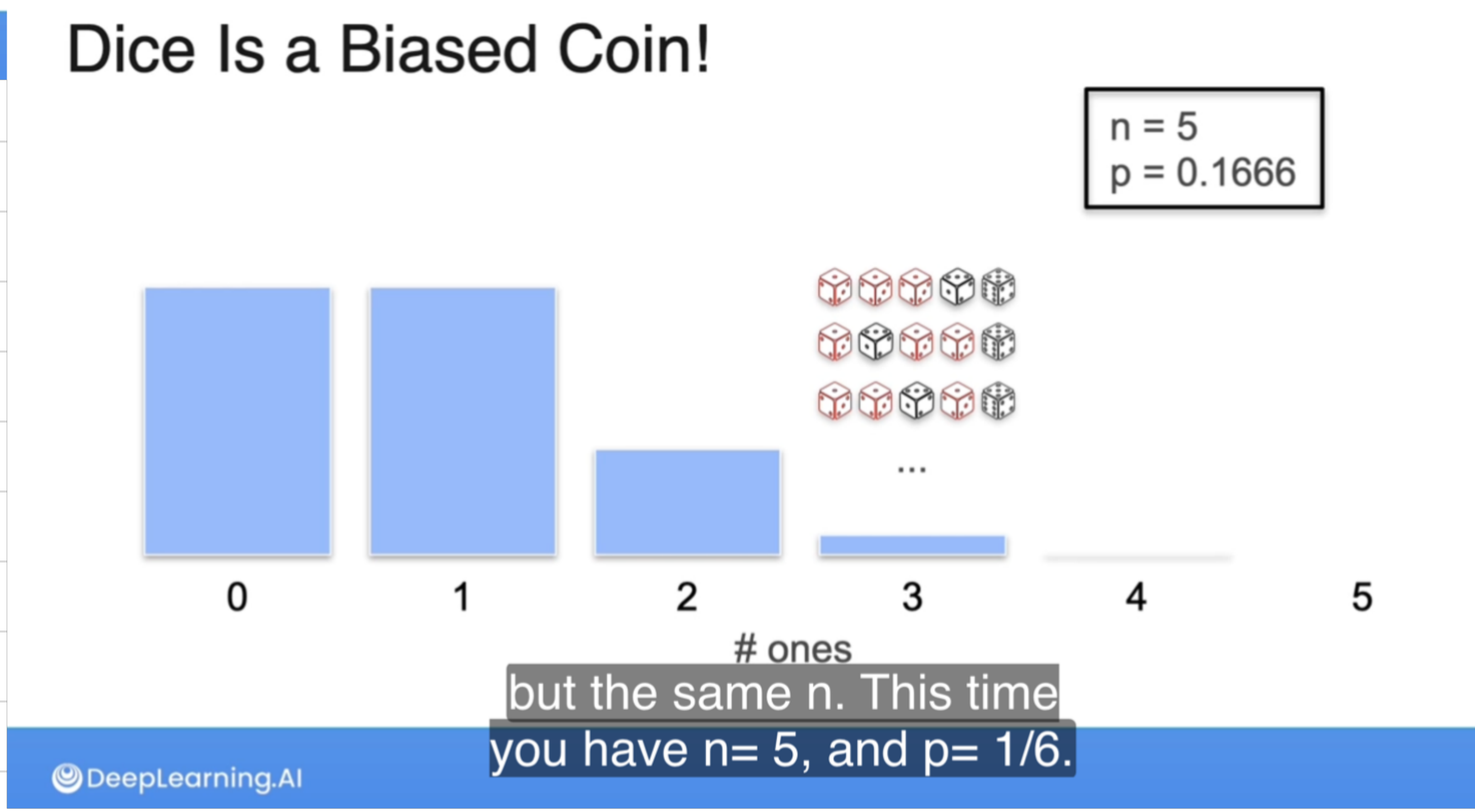
n and p are called the **parameters** of the binomial distribution

n is the number of tosses,

DeepLearning.AI


Dice Is a Biased Coin!
one
$p = \frac{1}{6}$
one
$p = \frac{1}{6}$
one
$p = \frac{1}{6}$
not one
$p = \frac{5}{6}$
not one
$p = \frac{5}{6}$
Now, this coin will have a
probability of heads of 1/6,
DeepLearning.AI


Question
What is the probability of getting three ones when rolling a dice five times (no matter on which dice)?
$\binom{5}{3}\left(\frac{1}{2}\right)^5$
$\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
$\binom{5}{3}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2$
$\left(\frac{1}{6}\right)^5$
Correct
You're right! The detailed exaplanation will follow right after this quiz.
Skip
Continue
DeepLearning.AI
Well, when you throw a dice,


Dice Is a Biased Coin!
n = 5
p = 0.1666
...
0
1
2
3
4
5
ones
but the same n. This time
you have n= 5, and p= 1/6.
DeepLearning.AI

Binomial Coefficient

Pick 1st number | Pick 2nd number | Pick 3rd number | Pick k-th number

n options → $n-1$ options → $n-2$ options → … → $n-(k-1)$options

Ordered sets of length k

$1,2,3,4,\ldots,k$

$2,1,3,4,\ldots,k$

$5,1,3,k,\ldots,2$

$n,1,3,4,\ldots,2$

…

Some of them repeat

$$n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-(k-1))$$

DeepLearning.AI

Question

If X is the number of times we get a 1 when rolling a dice ten times, then

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

Where n, p is equal to:

- (selected) n = 10, p = 1/6
- n = 6, p = 1/6
- n = 10, p = 1/2
- n = 10, p = 6

Correct
Correct. The probability of getting a 1 when throwing one dice is 1/6. And n is the total of experiments (10 times).

Skip | Continue

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

Where:
n: number of trials
k: number of successes
p: probability of success
$(1-p)$ Probability of failure

Binomial Coefficient

	Pick 1st	Pick 2nd	Pick 3rd	Pick 4th
	4 options	3 options	2 options	1 option

1,2,3,4

1,2,4,3

1,3,2,4

1,3,4,2

…

4,3,2,1

$$4\cdot3\cdot2\cdot1 = 4!$$

For five numbers:

$$5\cdot4\cdot3\cdot2\cdot1 = 5!$$

General solution: $k!$

DeepLearning.AI

k different objects.

Binomial Coefficient

Pick 1st number — Pick 2nd number — Pick 3rd number — Pick k-th number

n options $\to$ $n-1$ options $\to$ $n-2$ options $\to \dots \to$ $n-(k-1)$options

Ordered sets of length k

$1,2,3,4,\dots,k$
$2,1,3,4,\dots,k$
$5,1,3,k,\dots,2$ $\quad n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-(k-1))$
$n,1,3,4,\dots,2$
$\dots$

was n times n-1 all
the way to times n-k-1.

DeepLearning.AI

Binomial Coefficient

Pick 1st number — Pick 2nd number — Pick 3rd number — Pick k-th number

n options $\to$ $n-1$ options $\to$ $n-2$ options $\to \dots \to$ $n-(k-1)$options

Ordered sets of length k $\longrightarrow$ Unordered sets of length k

$1,2,3,4,\dots,k$
$2,1,3,4,\dots,k$
$5,1,3,k,\dots,2$
$n,1,3,4,\dots,2$
$\dots$

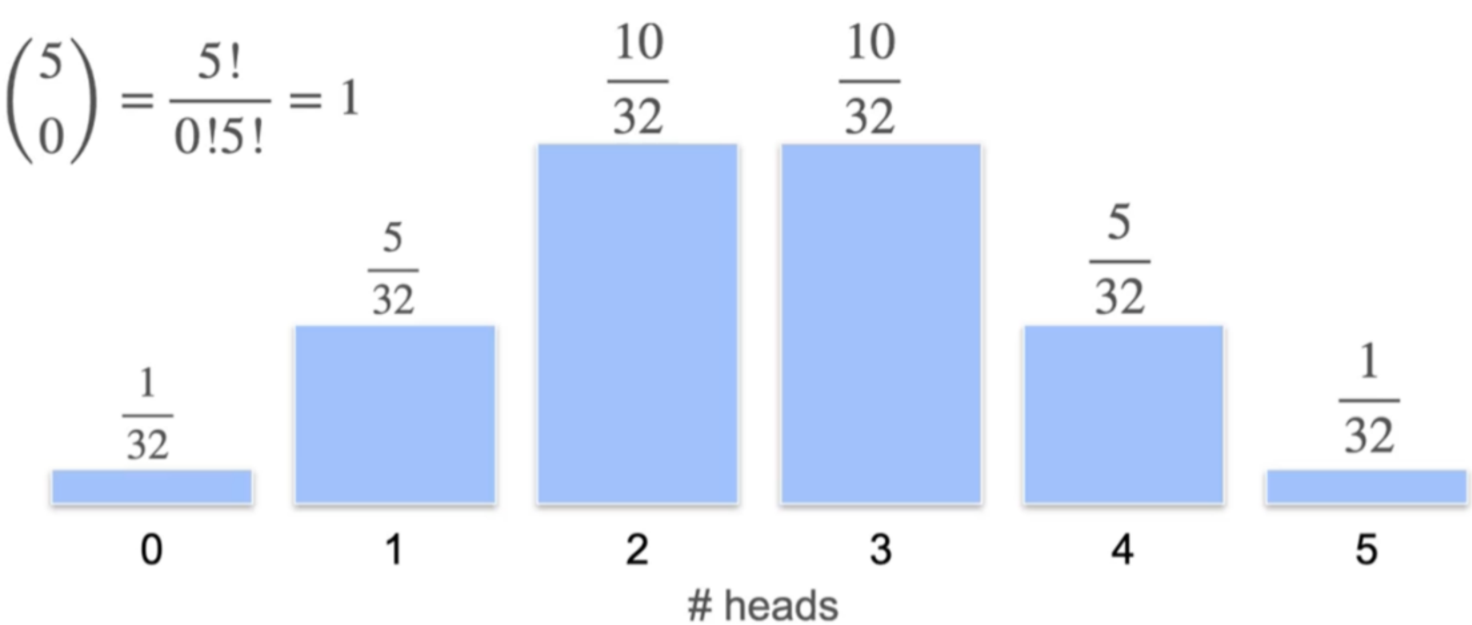
$$\frac{n\cdot(n-1)\cdot(n-2)\cdot\ldots\cdot(n-(k-1))}{k!} = \frac{n!}{(n-k)!\,k!} = \binom{n}{k}$$

DeepLearning.AI

Binomial Distribution models the number of successes in 'n' independent Bernoulli trials, where each trial has two outcomes: (with probability p) and without 1-p.

Binomial Distribution: Fair Coins

$$\binom{5}{0} = \frac{5!}{0!5!} = 1$$



DeepLearning.AI

Biased Coins

30 %

70 %

$0.3 \cdot 0.3 \cdot 0.3 \cdot 0.3 \cdot 0.3 = 0.00243$

$0.3 \cdot 0.3 \cdot 0.7 \cdot 0.7 \cdot 0.7 = 0.01323$

DeepLearning.AI

→ A biased coin is a coin where the probabilities of landing on heads or tails are not equal, unlike a fair coin. p of head cannot be 0.5.

Binomial Coefficient often denoted as $\binom{n}{k}$ represents the number of ways to choose k items from n items without considering the order.

Biased Coins

$0.3 \cdot 0.3 \cdot 0.3 \cdot 0.3 \cdot 0.3 = 0.3^5 \cdot 0.7^0$

$0.3 \cdot 0.3 \cdot 0.3 \cdot 0.3 \cdot 0.7 = 0.3^4 \cdot 0.7^1$

$0.3 \cdot 0.3 \cdot 0.3 \cdot 0.7 \cdot 0.7 = 0.3^3 \cdot 0.7^2$

$0.3 \cdot 0.3 \cdot 0.7 \cdot 0.7 \cdot 0.7 = 0.3^2 \cdot 0.7^3$

$0.3 \cdot 0.7 \cdot 0.7 \cdot 0.7 \cdot 0.7 = 0.3^1 \cdot 0.7^4$

$0.7 \cdot 0.7 \cdot 0.7 \cdot 0.7 \cdot 0.7 = 0.3^0 \cdot 0.7^5$

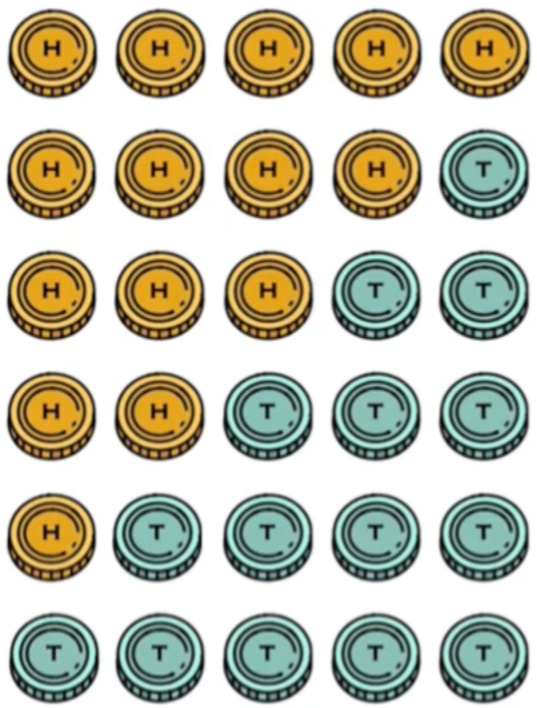
$= 0.3^k \cdot 0.7^{n-k}$

And here we have all the other ones

0.3 to the 4 times 0.7 to the 1,

DeepLearning.AI

Biased Coins



$0.3^5 \cdot 0.7^0$

$0.3^4 \cdot 0.7^1$

$0.3^3 \cdot 0.7^2$

$0.3^2 \cdot 0.7^3$

$0.3^1 \cdot 0.7^4$

$0.3^0 \cdot 0.7^5$

$$= 0.3^k \cdot 0.7^{n-k} \Rightarrow \binom{n}{k} 0.3^k \cdot 0.7^{n-k}$$

Account for all possible orders of heads and tails

DeepLearning.AI

And that's where n choose k comes

$n \Rightarrow$ no. of trials

$k \Rightarrow$ no. of favourable outcomes.

Question

Throwing a 4-sided fair dice and observing if it lands in 2 or not might be modeled as a Bernoulli distribution with p equals to:

- 1/4
- 1/6
- 1/2
- 3/4

Correct

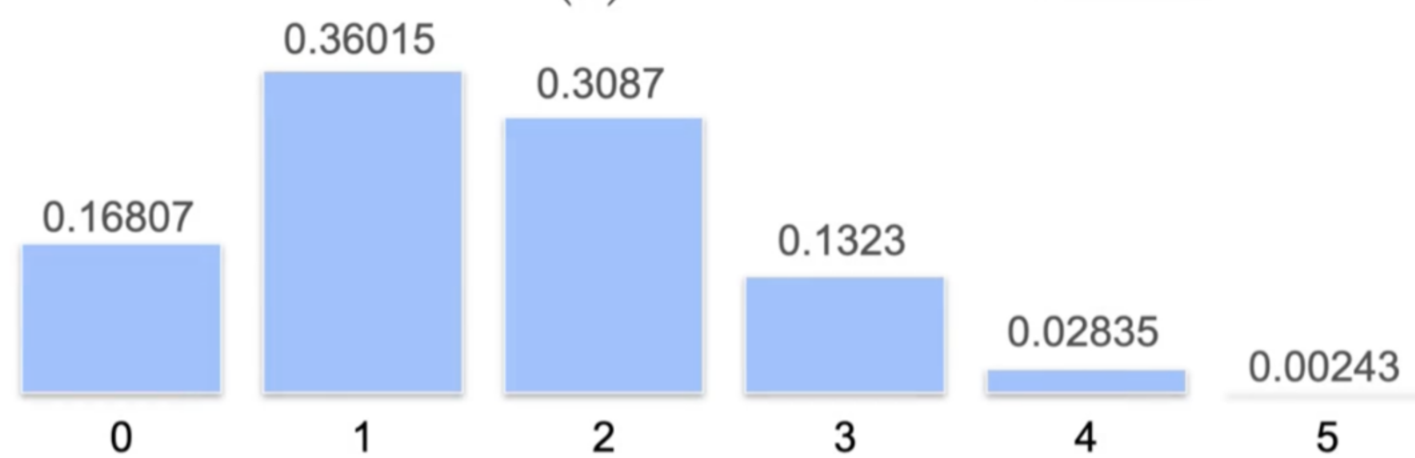
Correct! In a 4-sided fair dice, there is a 1/4 probability of landing in each face.

Skip Continue

Binomial Distribution

$$= \binom{n}{k} 0.3^k \cdot 0.7^{n-k}$$

n = 5
p = 0.3



Except that now you can calculate these probabilities using the formula that you

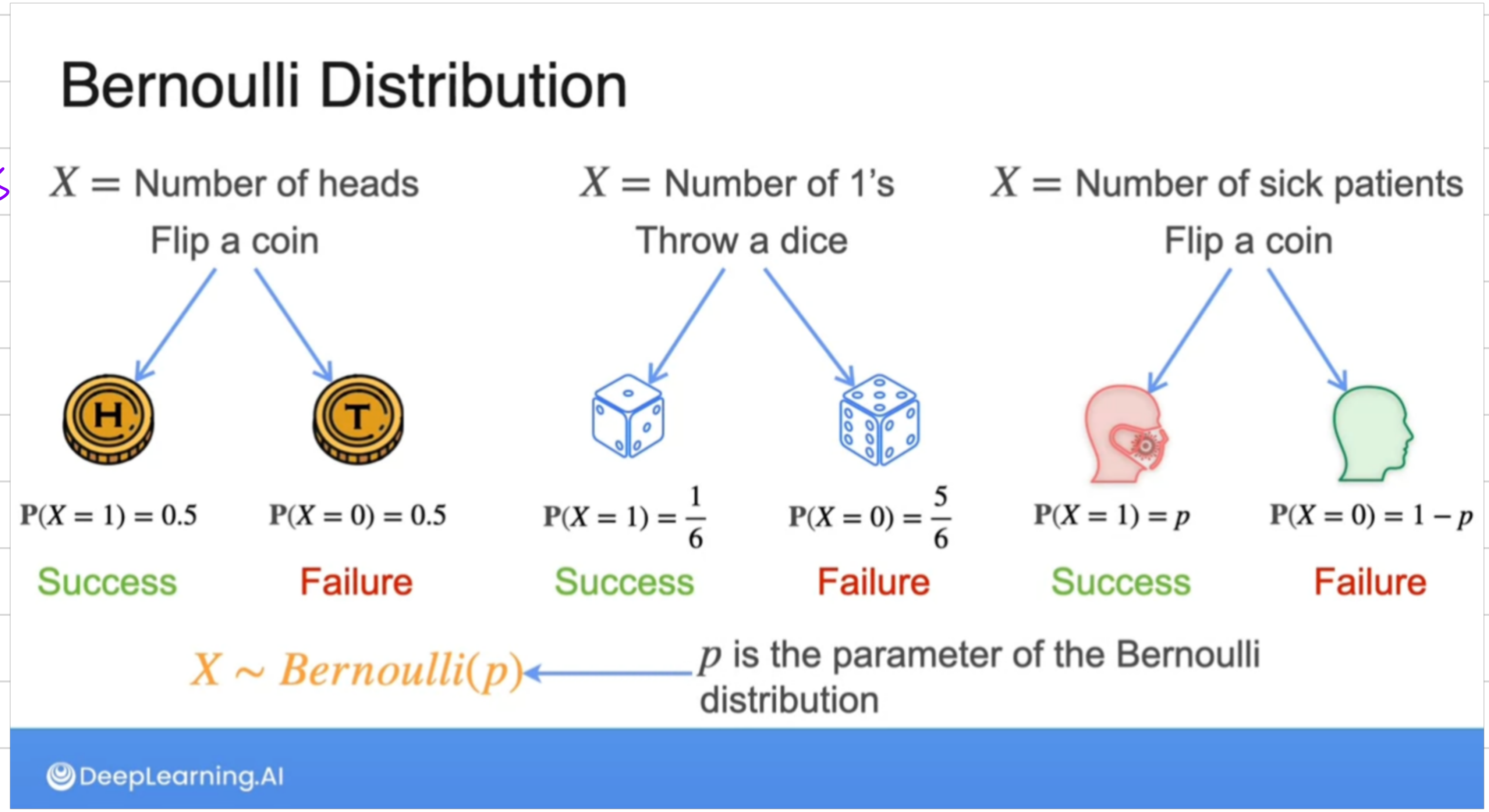
DeepLearning.AI

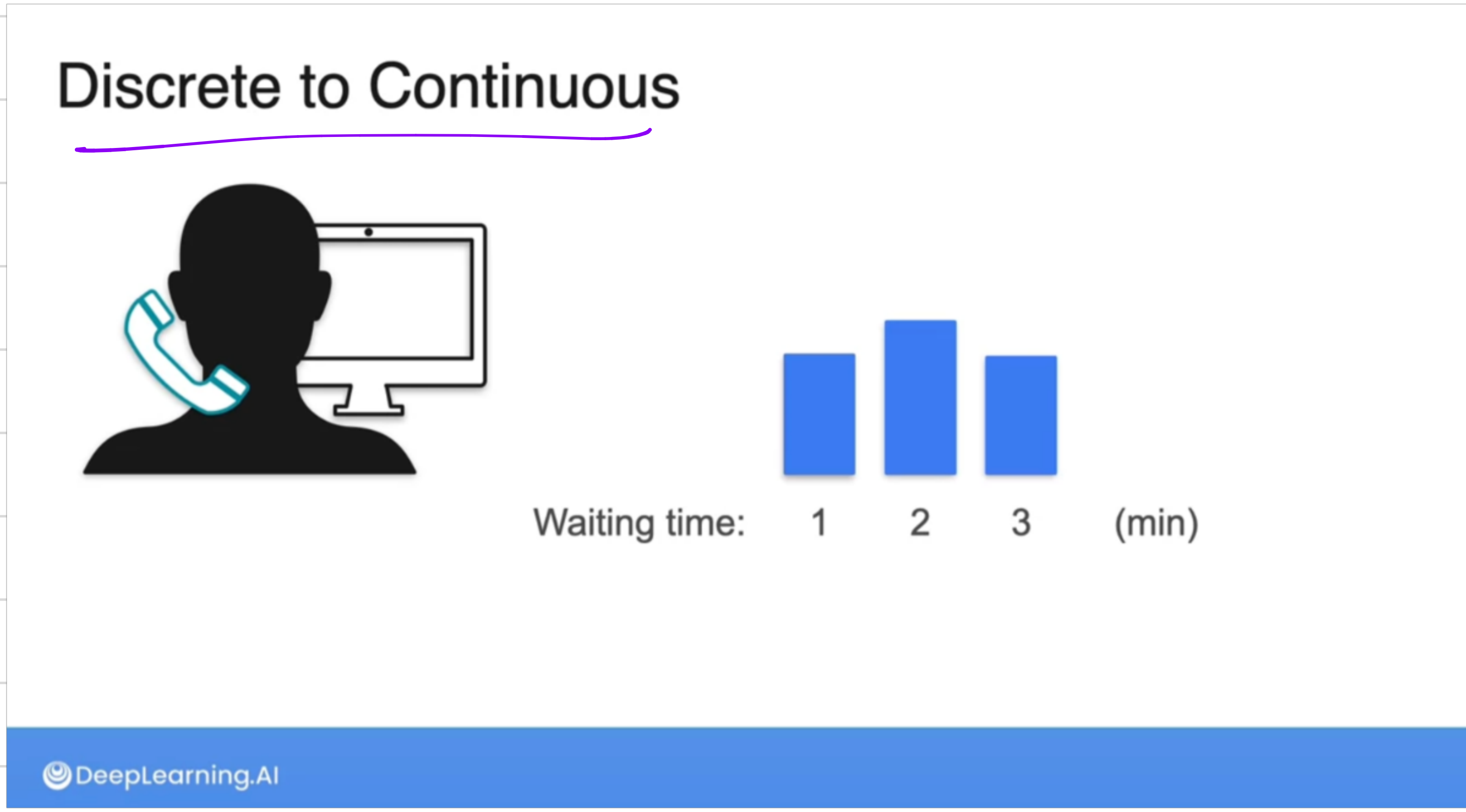
Bernoulli Distribution represents the outcome of a single trial of an experiment where there are only two possible outcomes:

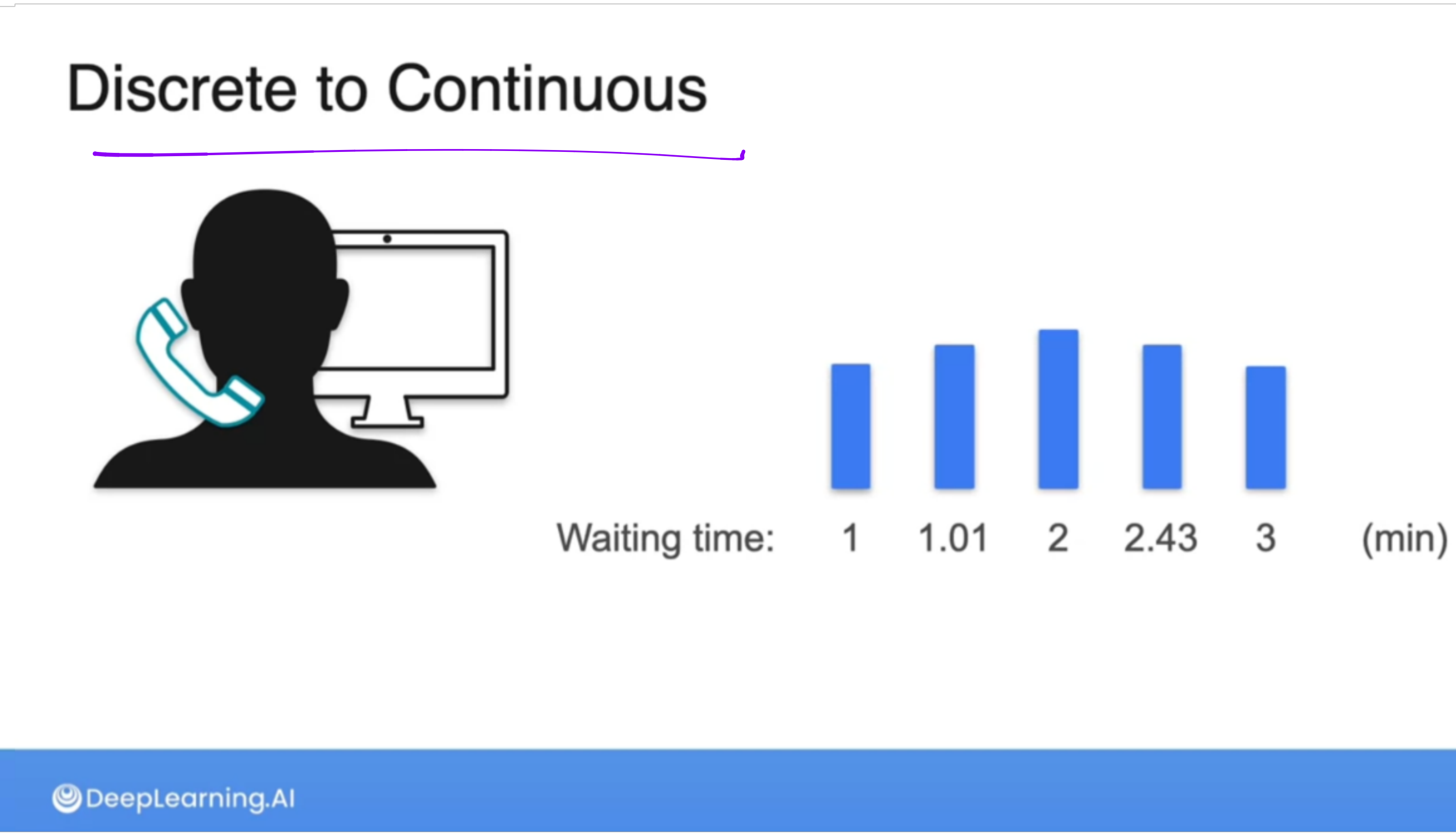
Success (often denoted as 1) and failure (often denoted as 0):

→ Success (1) with p

→ failure (0) with $1-p$.



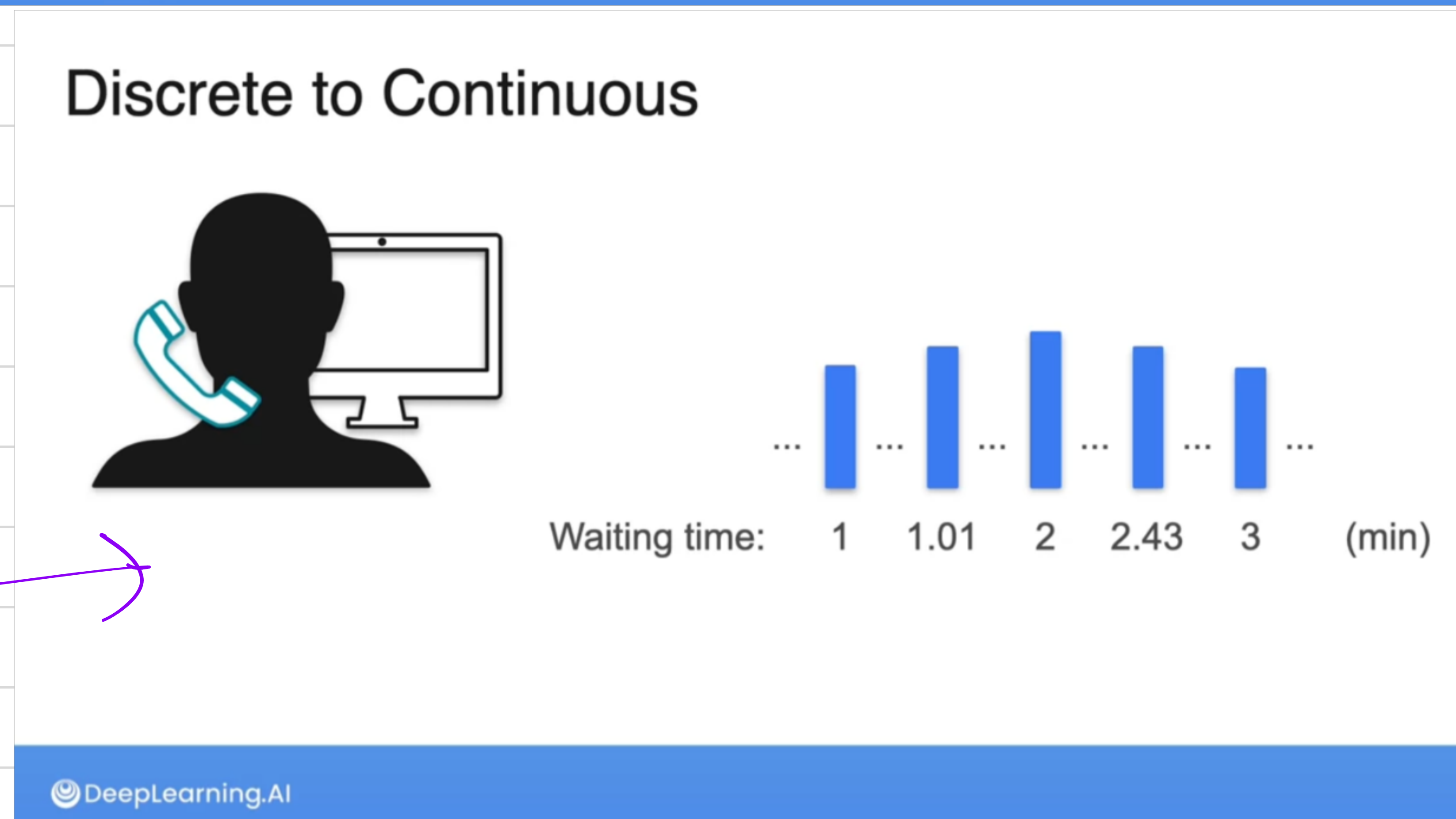




[1 – 1.01]

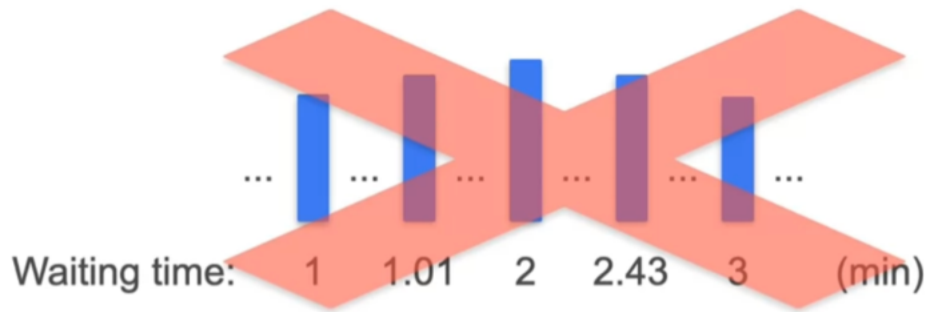
[2 – 2.43]

there are infinitely large numbers between 1 and 2.



Discrete to Continuous

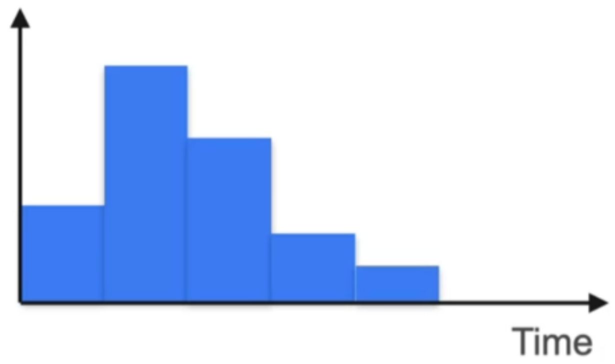
This is a continuous distribution!



What is the probability that you will wait EXACTLY one minute for the call?

Mixed both Discrete & Continuous But here ∞.

Discrete to Continuous



P(between 0 and 1 mins)
P(between 1 and 2 mins)
P(between 2 and 3 mins)
P(between 3 and 4 mins)
P(between 4 and 5 mins)

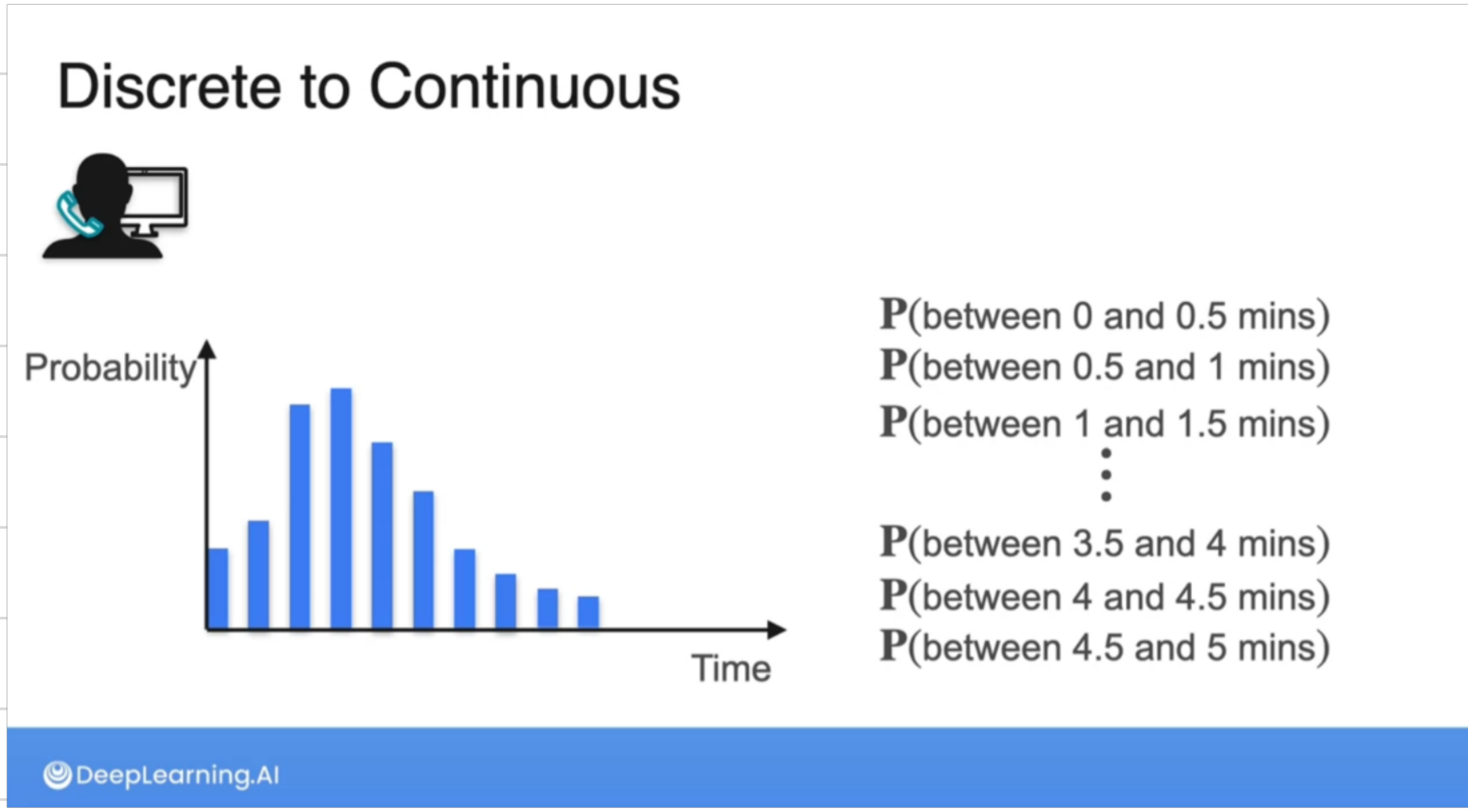
DeepLearning.AI

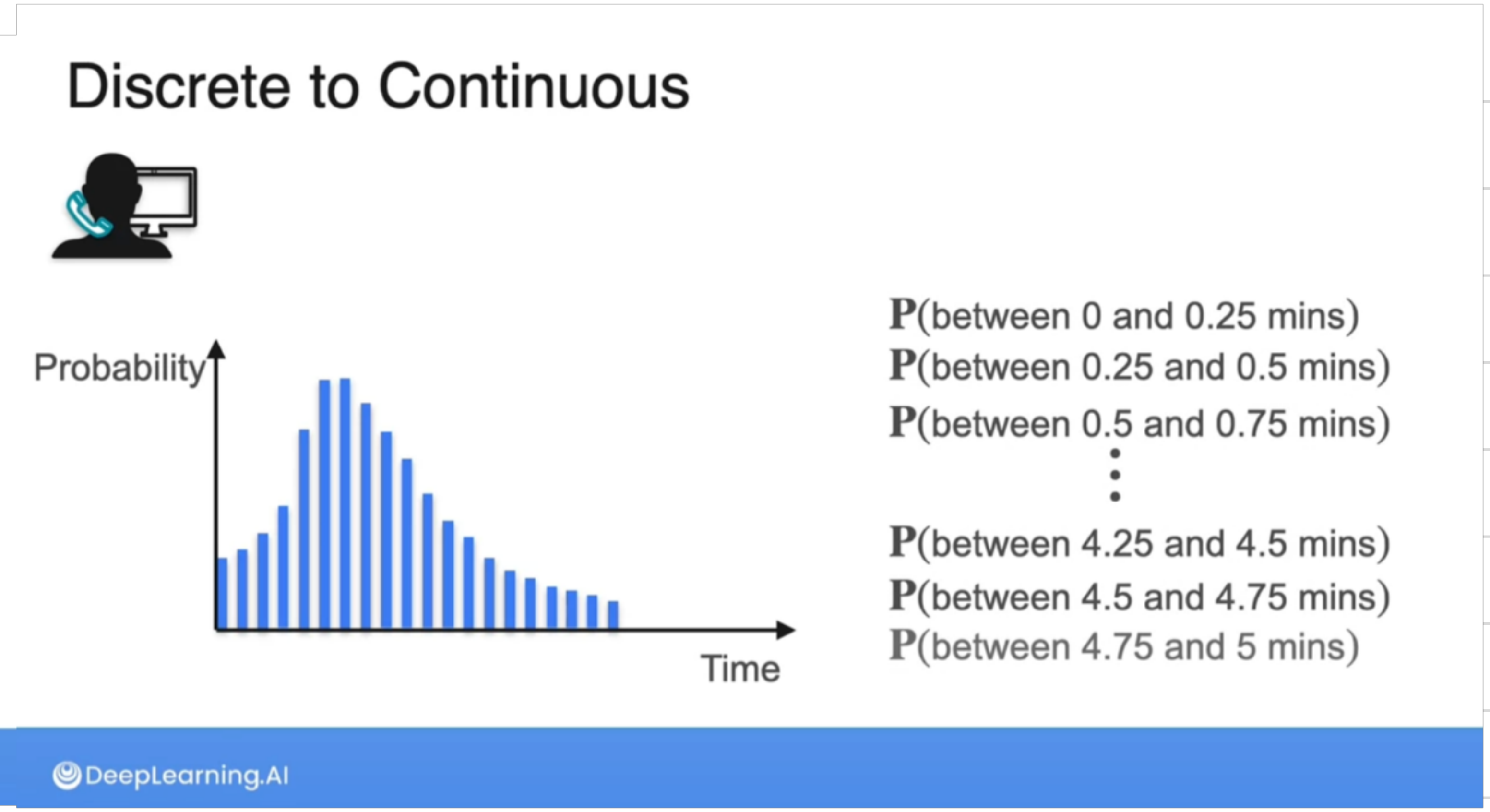
Transformation flow
Discrete Distributions
becomes
Continuous Distributions

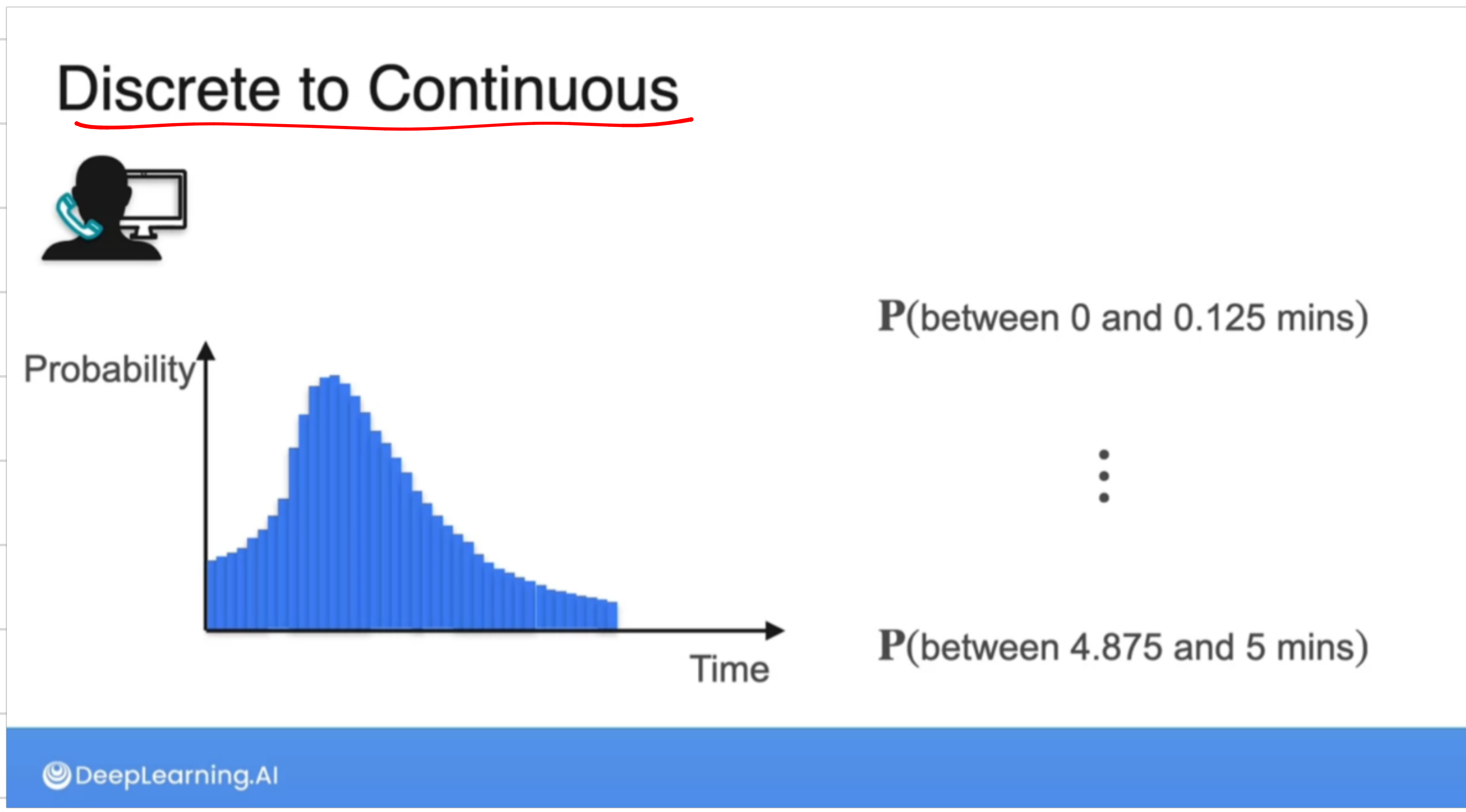
Discrete to Continuous

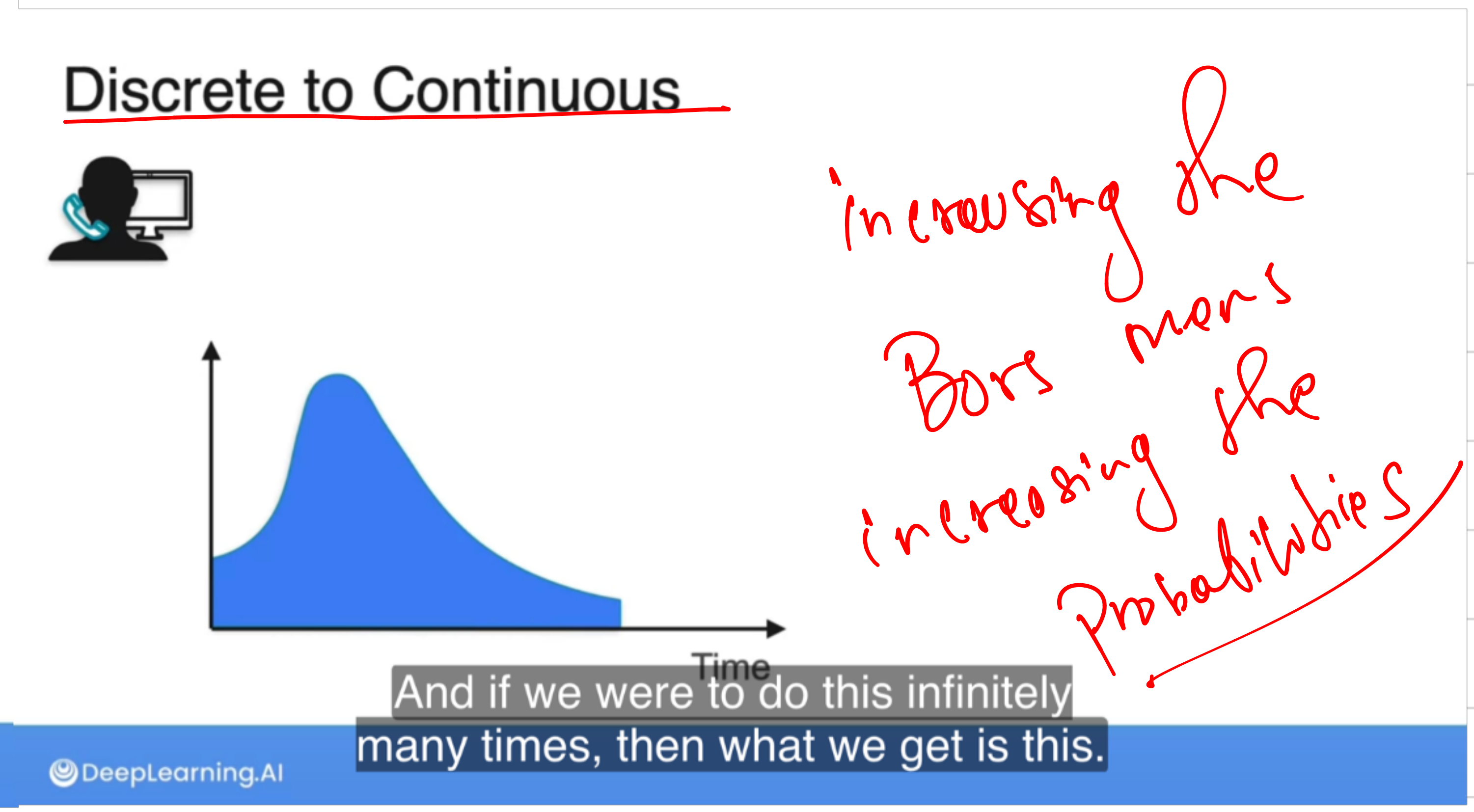
P(between 0 and 0.5 mins)
P(between 0.5 and 1 mins)
P(between 1 and 1.5 mins)
⋮
P(between 3.5 and 4 mins)
P(between 4 and 4.5 mins)
P(between 4.5 and 5 mins)

DeepLearning.AI

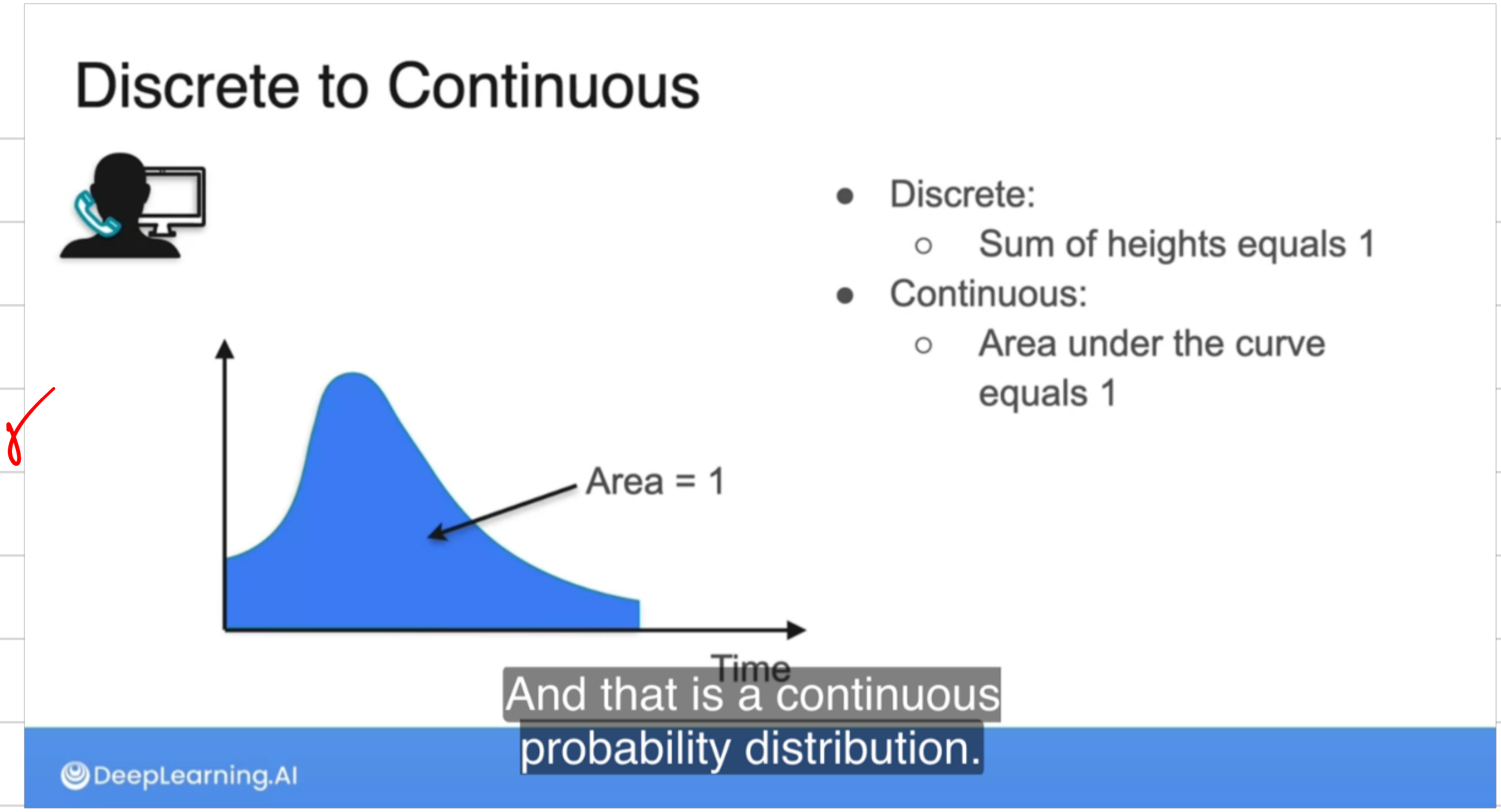

Discrete to Continuous
Probability
Time
P(between 0 and 0.5 mins)
P(between 0.5 and 1 mins)
P(between 1 and 1.5 mins)
⋮
P(between 3.5 and 4 mins)
P(between 4 and 4.5 mins)
P(between 4.5 and 5 mins)
DeepLearning.AI

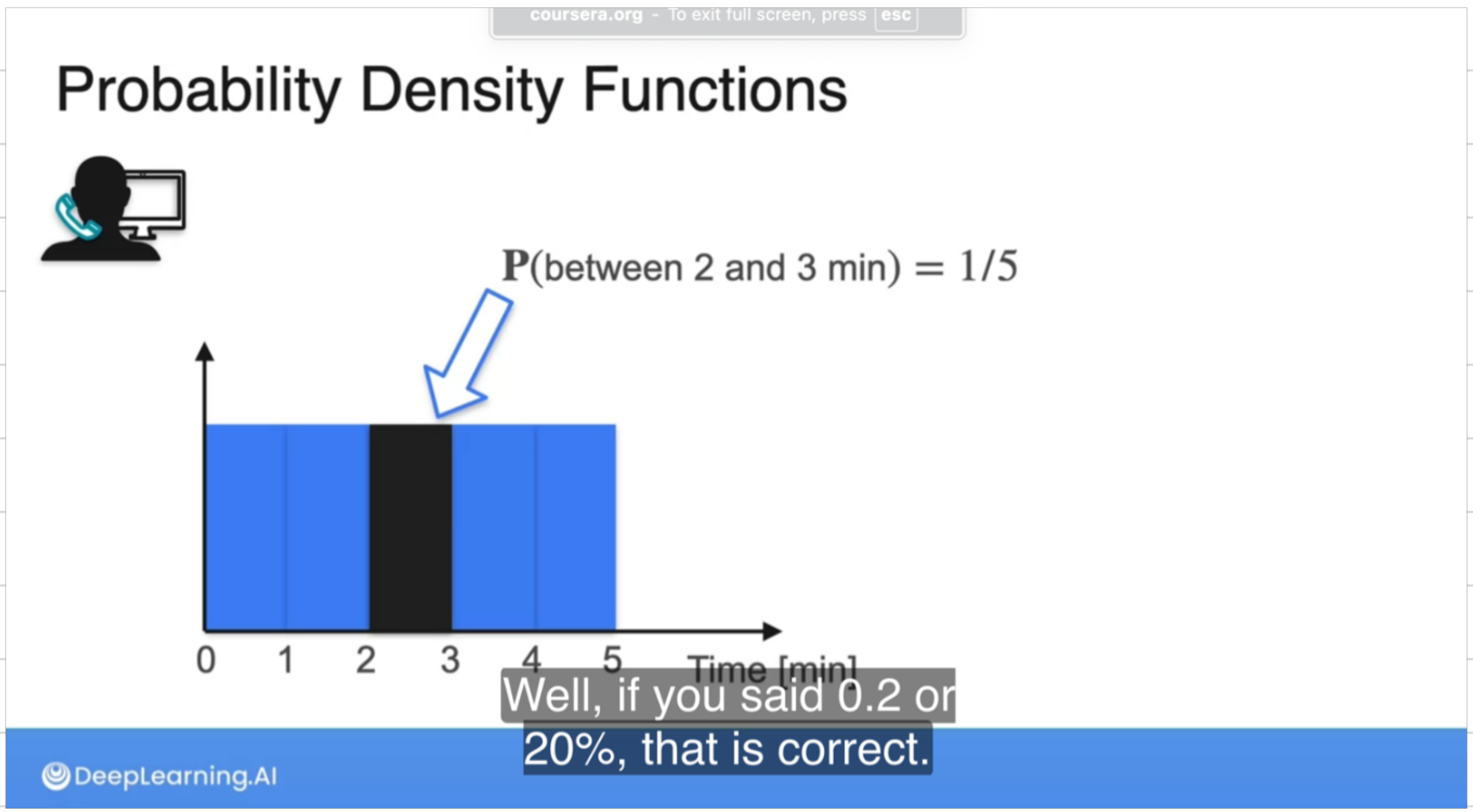

Discrete to Continuous
Probability
Time
P(between 0 and 0.25 mins)
P(between 0.25 and 0.5 mins)
P(between 0.5 and 0.75 mins)
⋮
P(between 4.25 and 4.5 mins)
P(between 4.5 and 4.75 mins)
P(between 4.75 and 5 mins)
DeepLearning.AI


Discrete to Continuous
Probability
Time
P(between 0 and 0.125 mins)
⋮
P(between 4.875 and 5 mins)
DeepLearning.AI


Discrete to Continuous
Time
increasing the Bars means increasing the Probabilities
And if we were to do this infinitely
many times, then what we get is this.
DeepLearning.AI

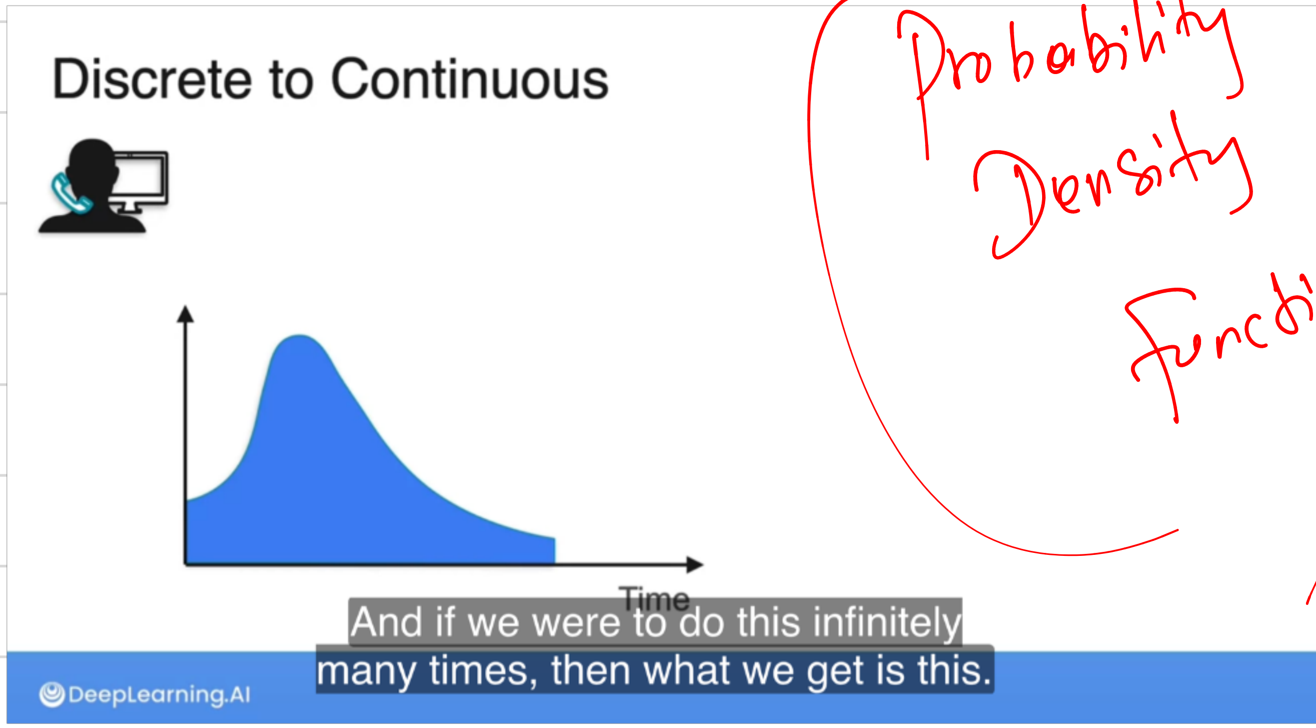
Both Discrete and Continuous have value under f. (Area = 1)



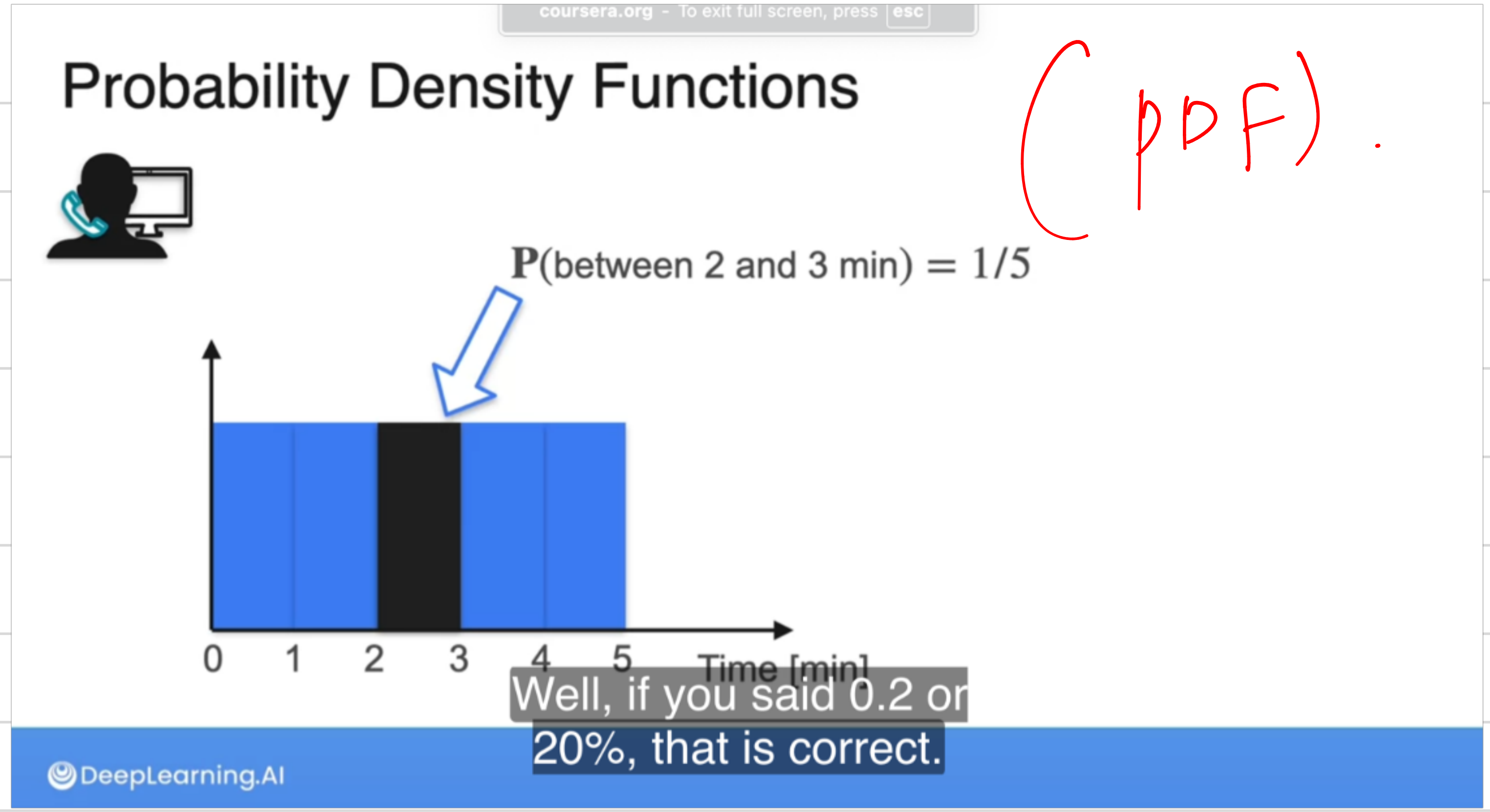


Probability (2 and 3 min) = 1/5

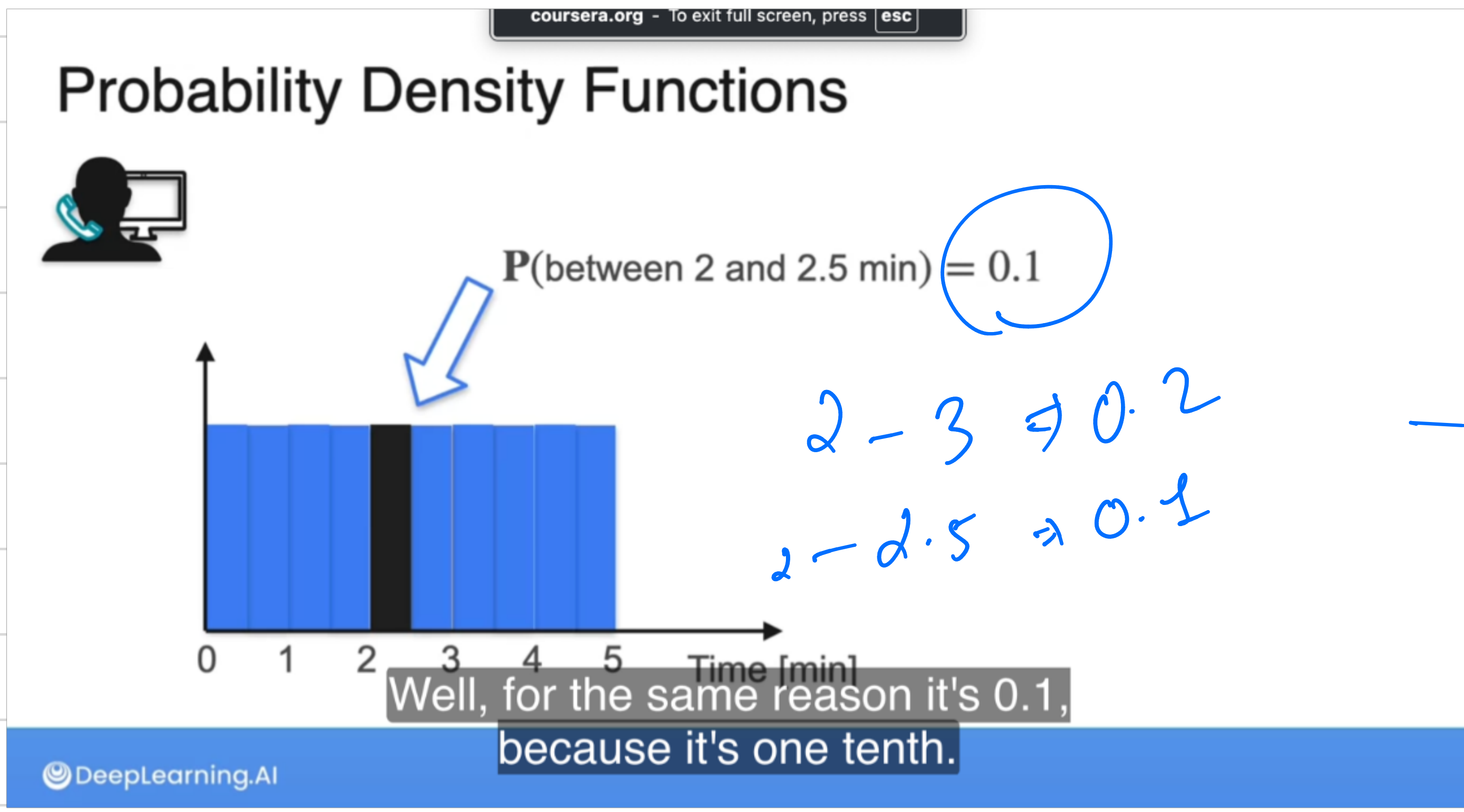
Since there are 5 events.



Probability Density Function



(PDF).

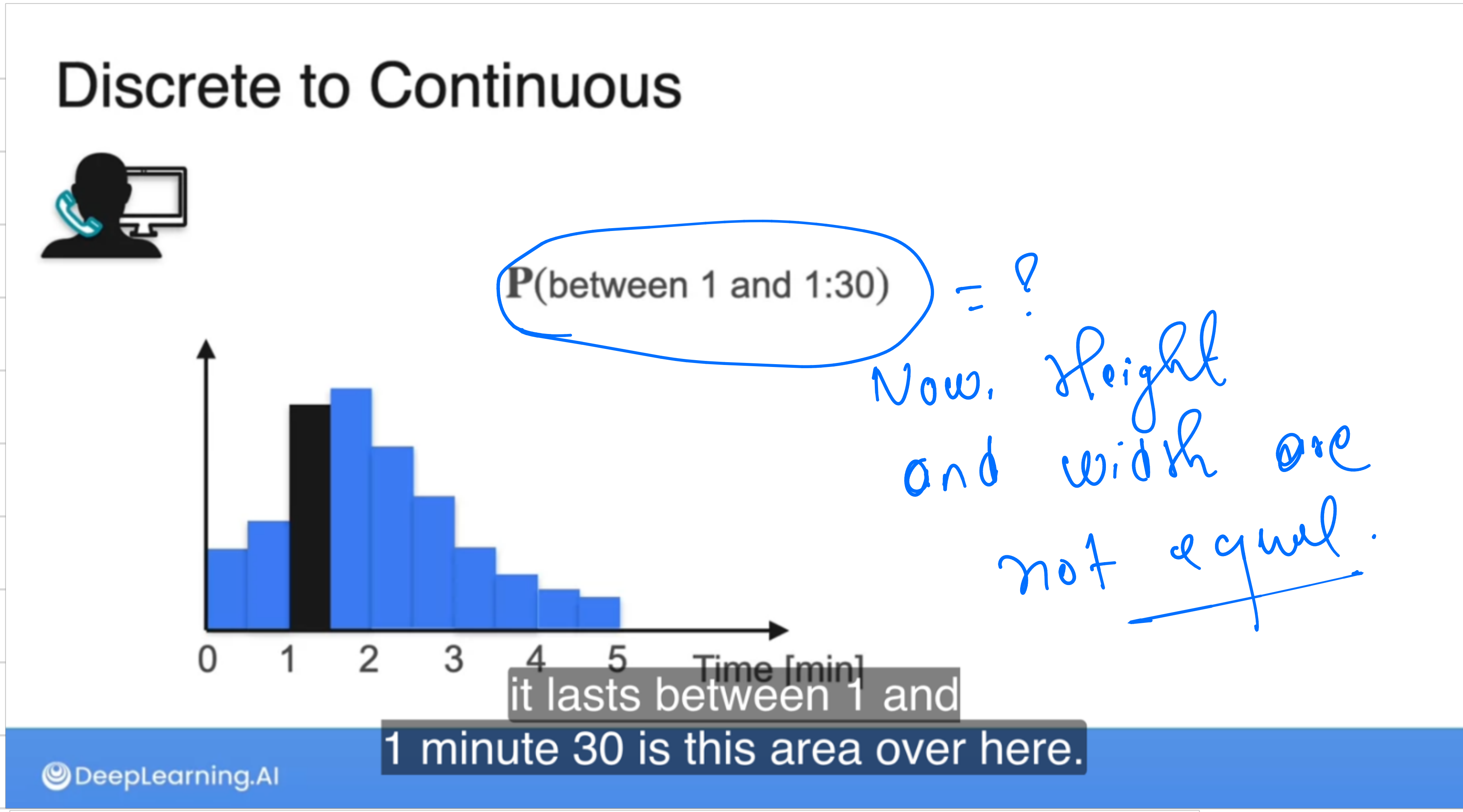


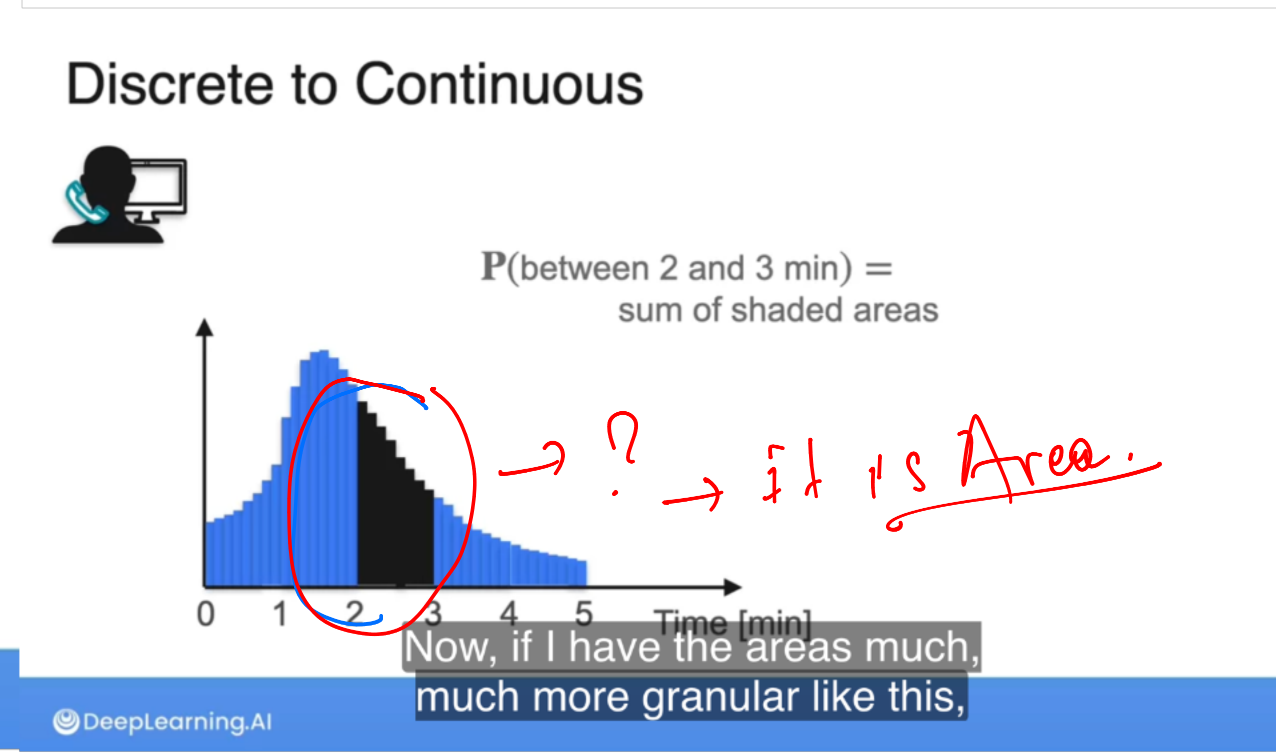
2 - 3 ⇒ 0.2

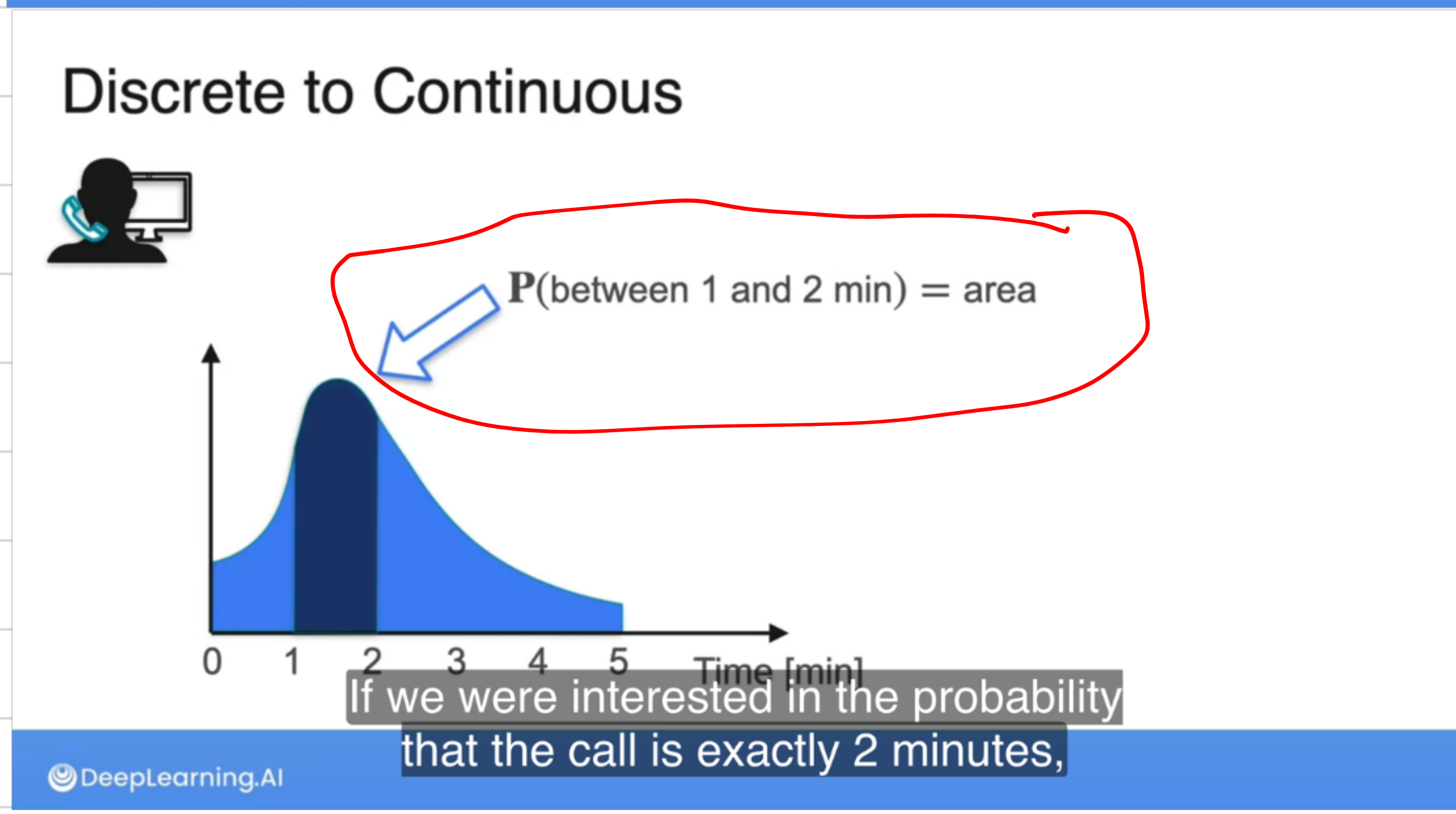
2 - 2.5 ⇒ 0.1

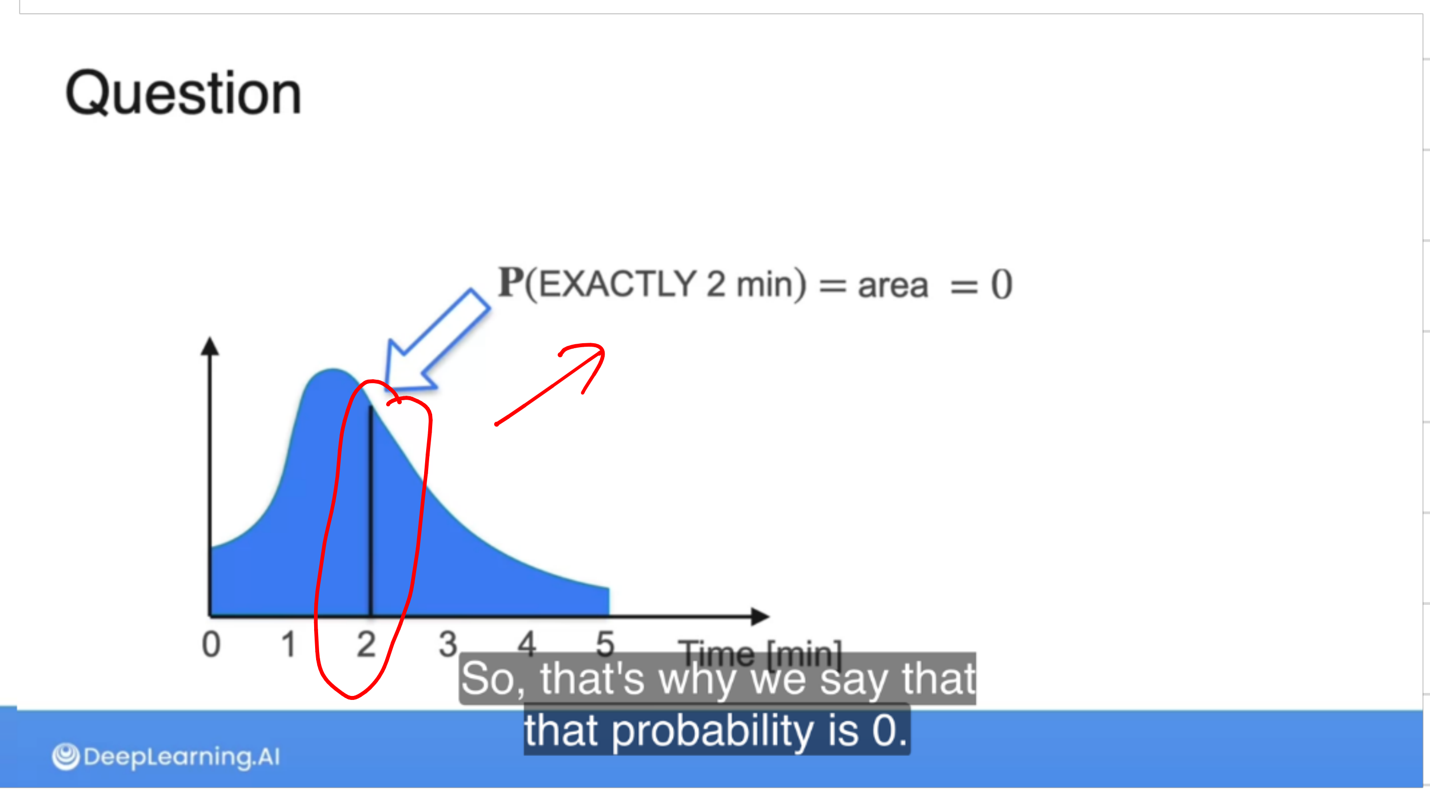
→ PDF ⇒ Since there are 5 and it is discrete event.

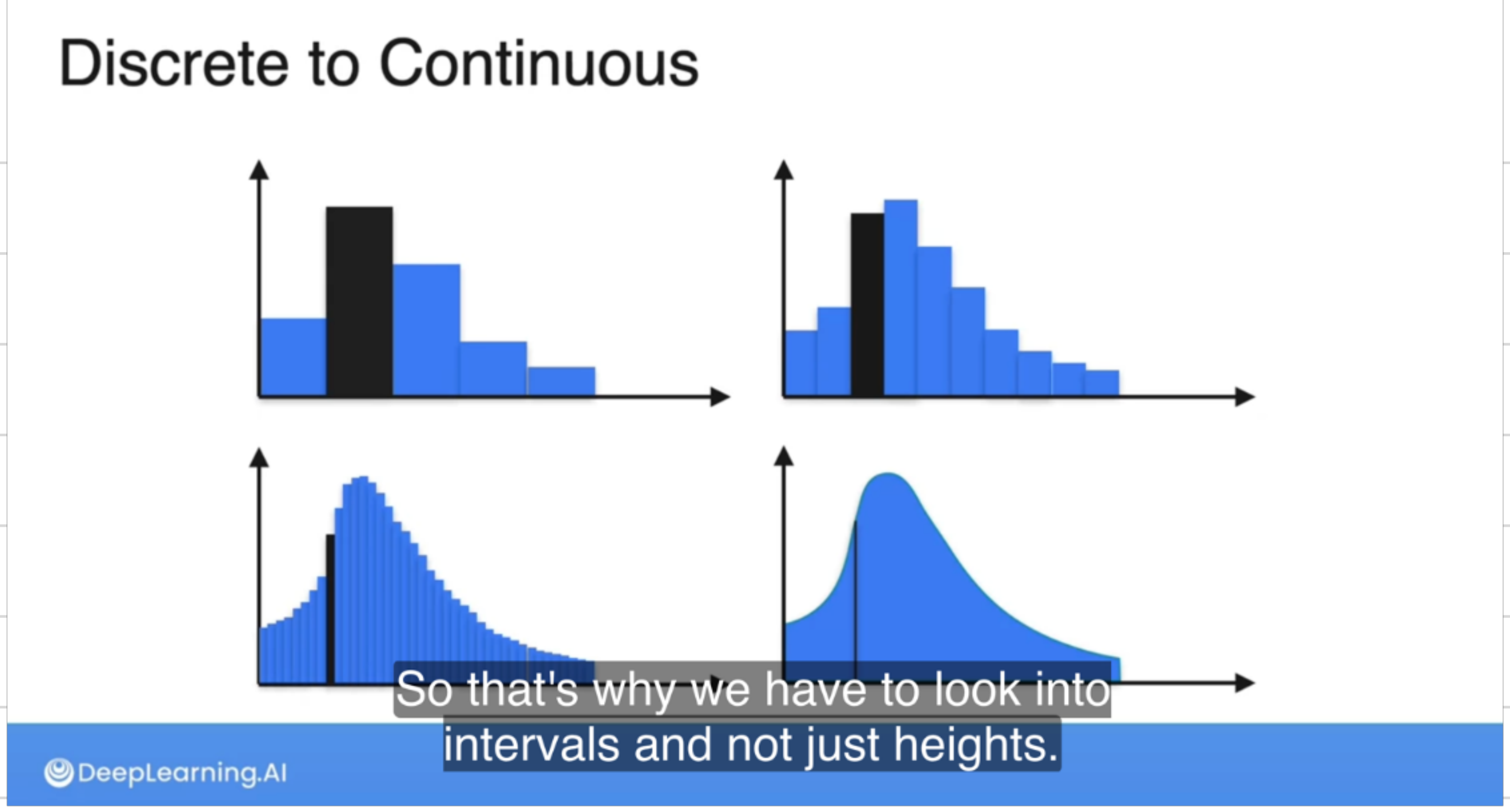
→, We're looking height, width we can say area [all equal].

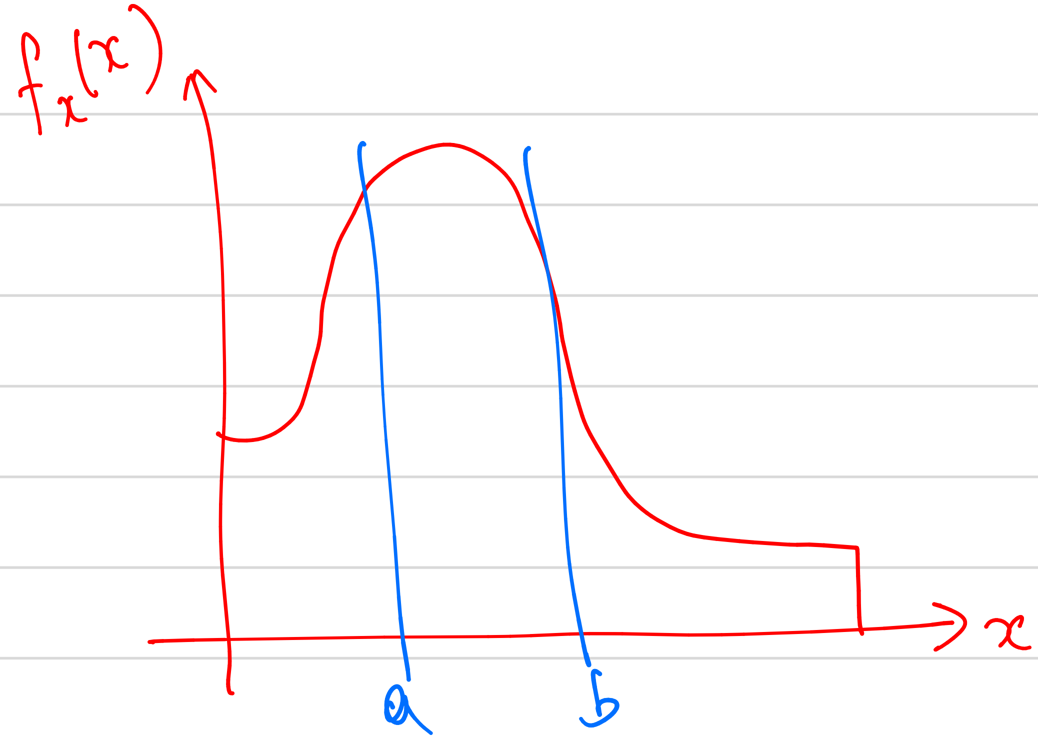

Discrete to Continuous
P(between 1 and 1:30)
= ?
Now, Height and width are not equal.
0 1 2 3 4 5 Time [min]
it lasts between 1 and
1 minute 30 is this area over here.
DeepLearning.AI


Discrete to Continuous
P(between 2 and 3 min) =
sum of shaded areas
→ ? → it's Area.
0 1 2 3 4 5 Time [min]
Now, if I have the areas much,
much more granular like this,
DeepLearning.AI


Discrete to Continuous
P(between 1 and 2 min) = area
0 1 2 3 4 5 Time [min]
If we were interested in the probability
that the call is exactly 2 minutes,
DeepLearning.AI


Question
P(EXACTLY 2 min) = area = 0
0 1 2 3 4 5 Time [min]
So, that's why we say that
that probability is 0.
DeepLearning.AI





Probability Density Function (PDF)

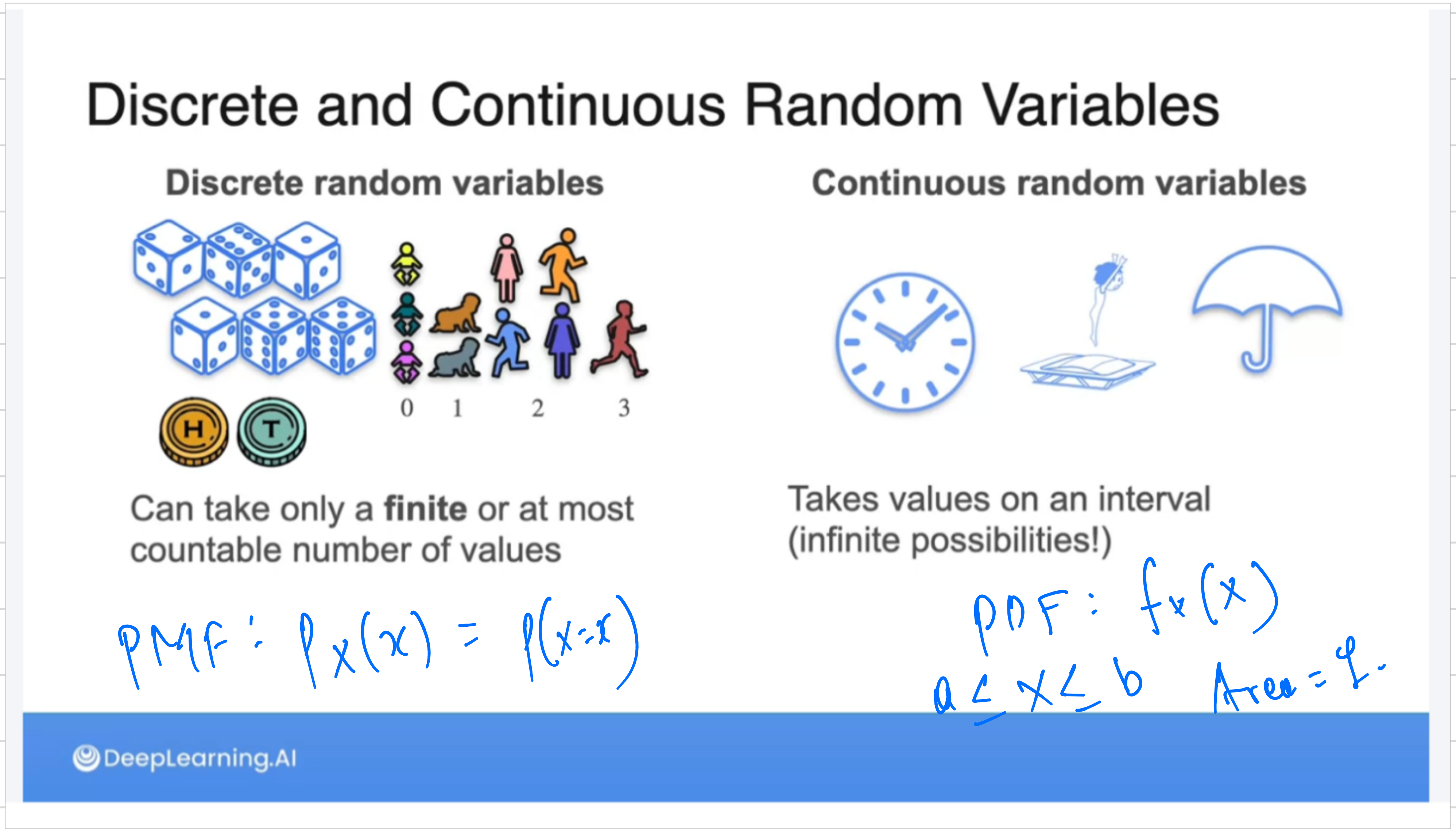
$f_X(x)$

Tells you the rate you accumulate probability around each point. Only defined for continuous variables!.

$P(a < X < b)$ = area under $f_X(x)$

$f_X(x)$ needs to satisfy:

- It is defined for all numbers
- $f_X(x) \geq 0$
- $f_X(x) = 1$



PMF: $p_X(x) = P(X=x)$

PDF: $f_X(x)$

$a \le X \le b$ Area = 1.

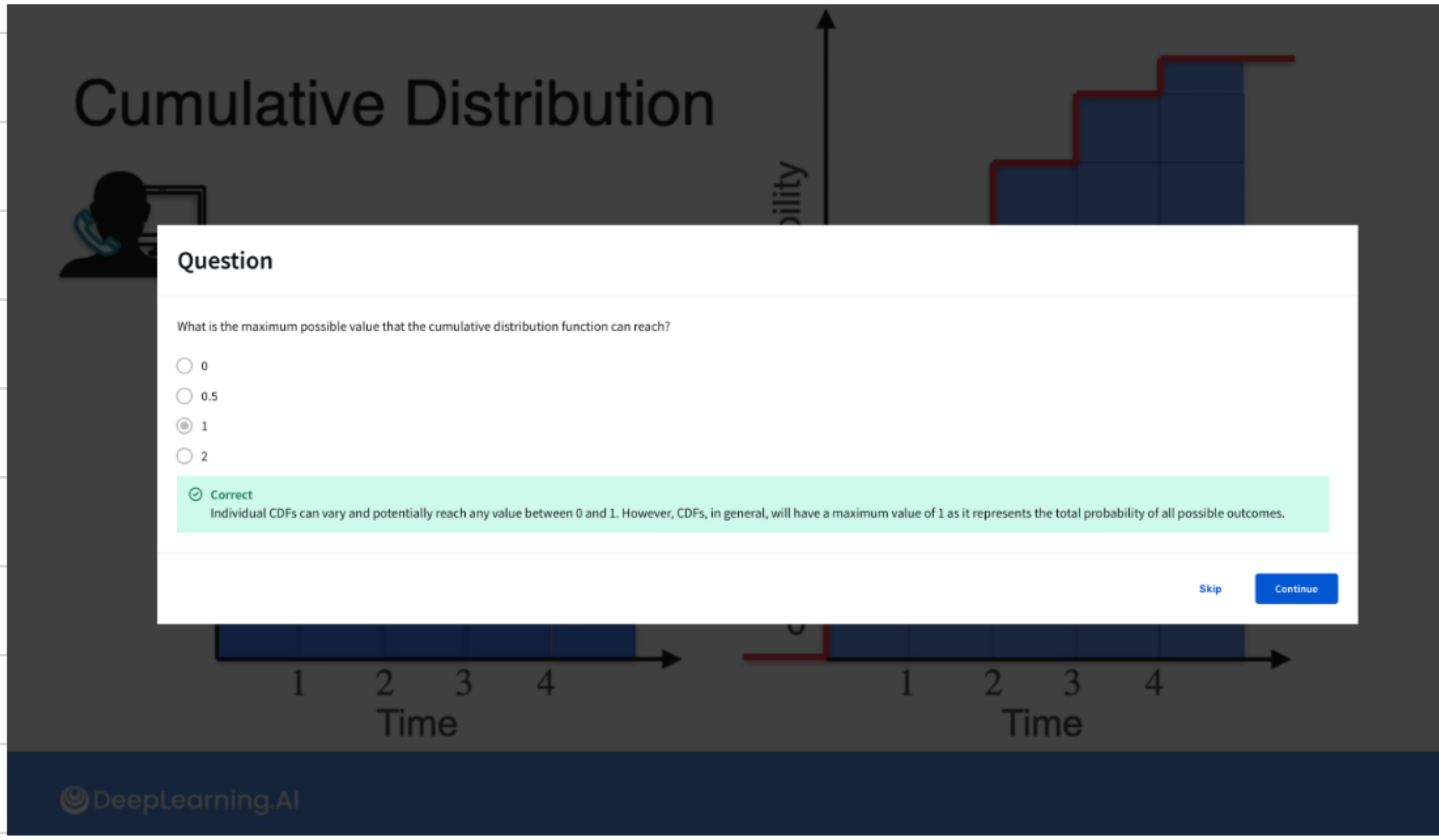
RV ⇒ Random Variables.

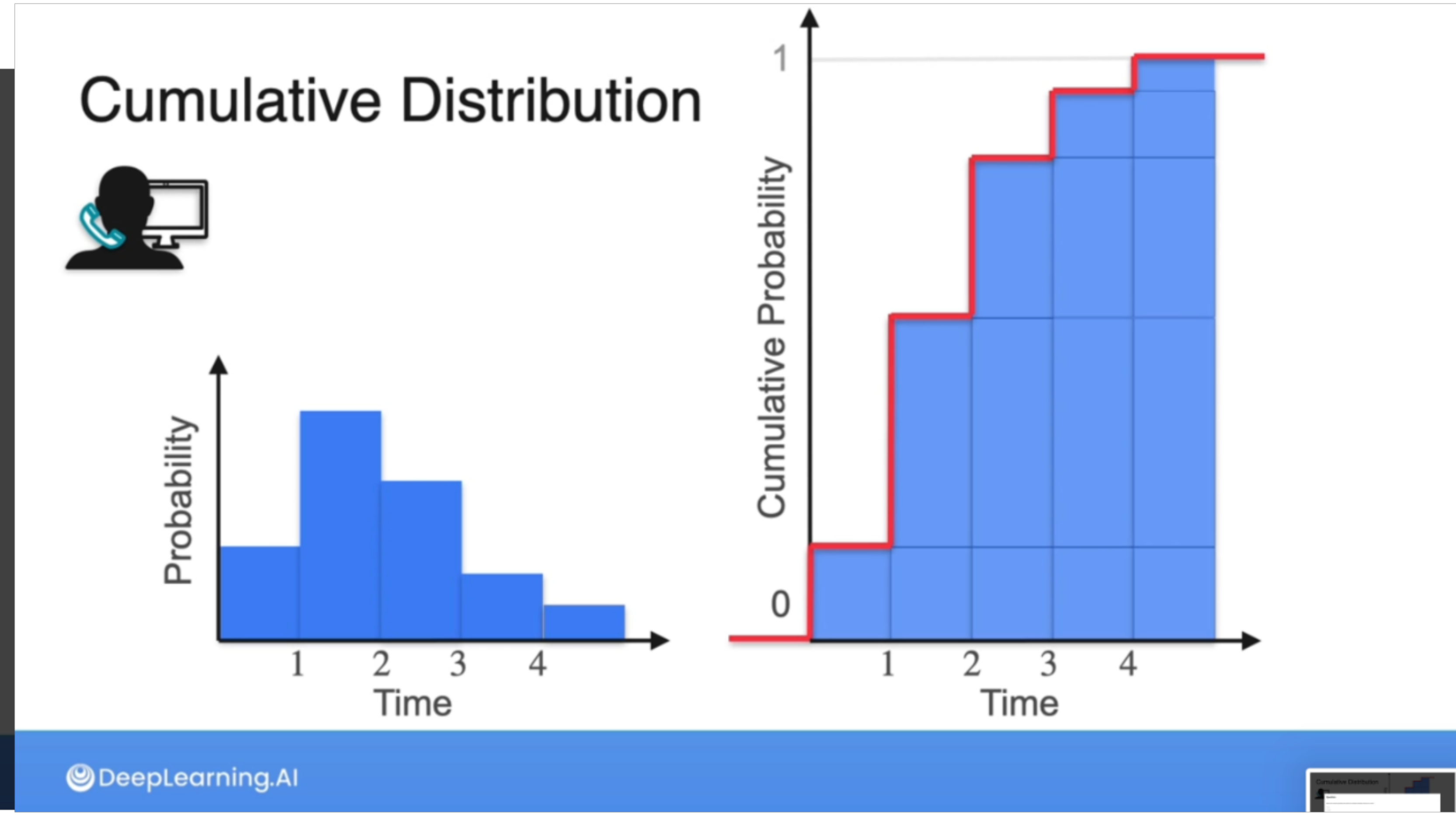
PMF ⇒ Discrete RV.

PDF ⇒ Continuous Random Variables.

Cummulative Distribution function:

CDF is the one that shows the actual probabilities that the event is between 0 and certain number of event.





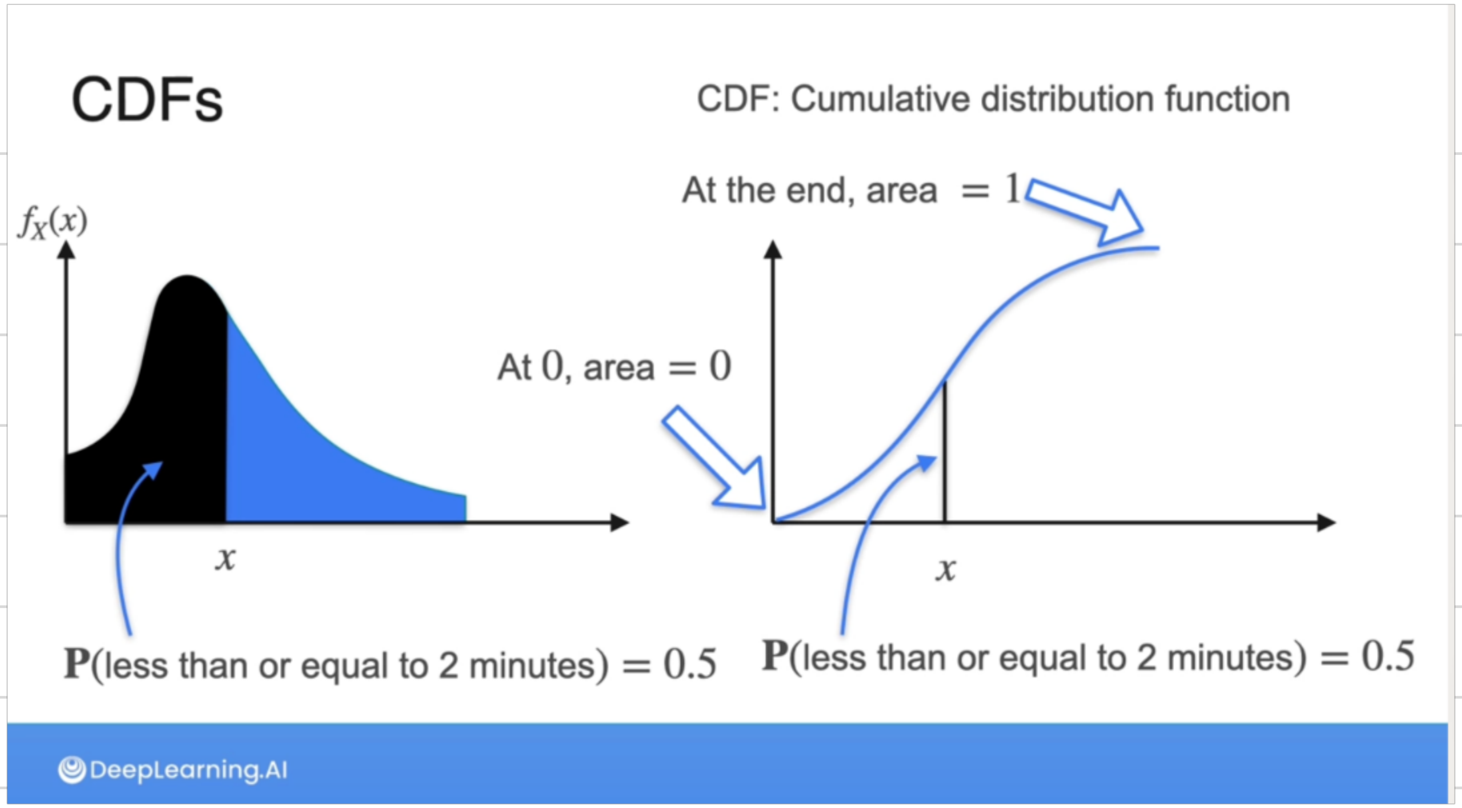
So, the p between 0 - 1 or the 1-2, and 2-3 and so on untill the Probability is 1. which means we start from 0 to 1.

← At the end, area = 1

P(less than or equal to 2 min) = 0.5

At 0, area = 0

P(less than or equal to 2 min) = 0.5



The CDF shows how much probability the variable has accumulated until a certain value

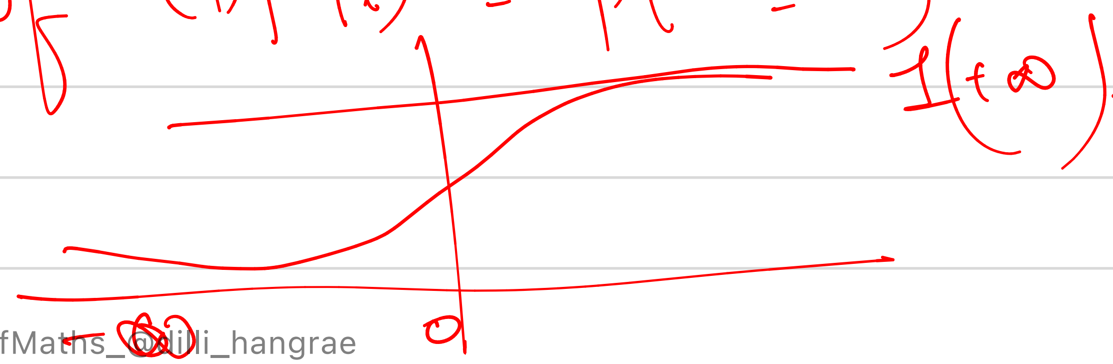
$CDF(x) = P(X \le x) \leftarrow$ It is defined for every real number

$[-\infty \text{ to } +\infty]$

$F_X(x) = P(X \le x)$

F sub (capital Y). $F_Y(x)$

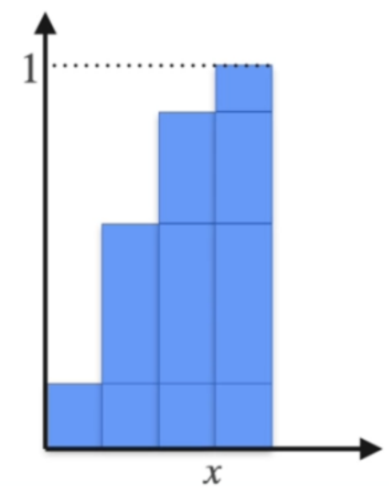
We get continuous function on curve of $CDF(x) = P(X \le x)$

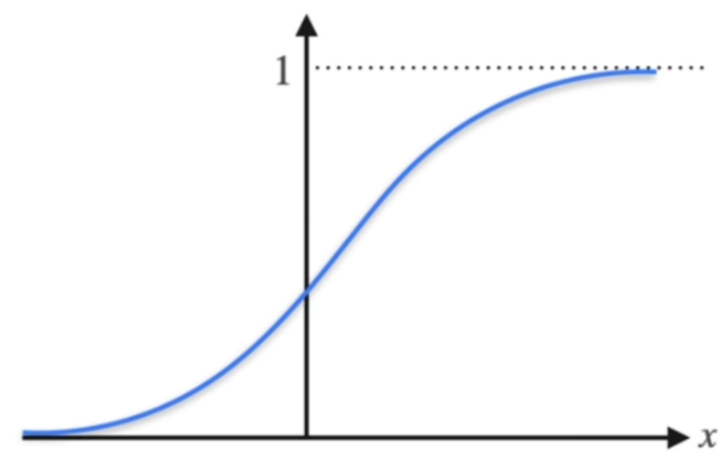


Cumulative Distribution Function: Formal Definition

The CDF shows how much probability the variable has accumulated until a certain value

That means that $F_X(x)$ $\text{CDF}(x) = \mathbf{P}(X \le x) \longleftarrow$ It is defined for every real number





DeepLearning.AI

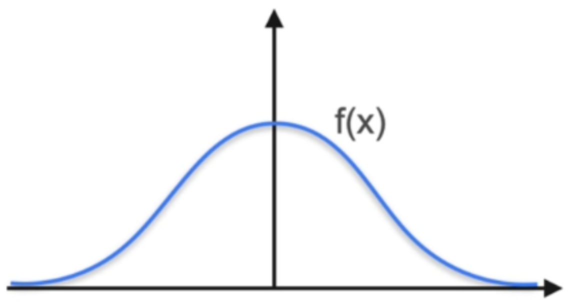
$0 \leq F_X(x) \leq 1$ —

[Left endpoint is 0
Right Endpoint is 1]

Never Decreases

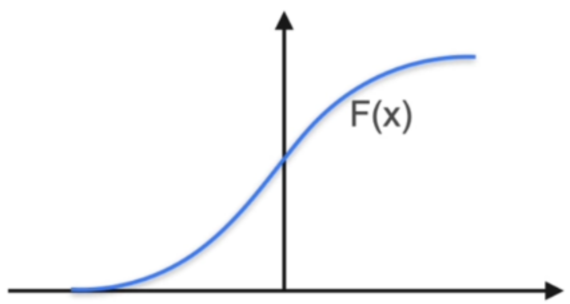
PDF and CDF Summary

PDF



- area = 1
- Always positive

CDF



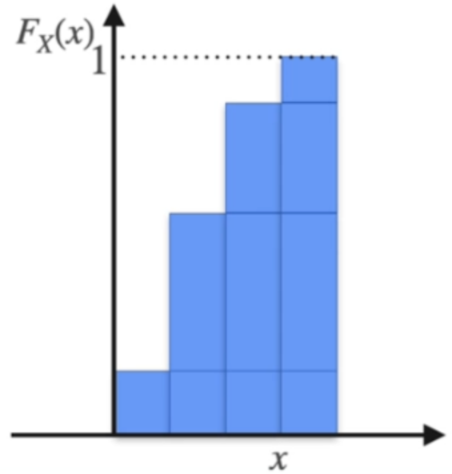
- left "endpoint" is 0
- right "endpoint" is 1
- (endpoints can be at infinity)
- Always positive and increasing

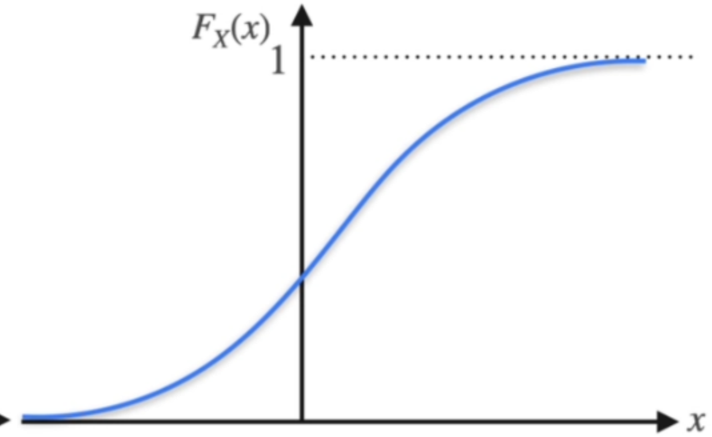
DeepLearning.AI

Cumulative Distribution Function: Formal Definition

The CDF shows how much probability the variable has accumulated until a certain value

That means that $F_X(x) = \mathbf{P}(X \leq x)$ ← It is defined for every real number





Properties

- $0 \leq F_X(x) \leq 1$
- Left "endpoint" is 0
- Right "endpoint" is 1
- **Never decreases**

DeepLearning.AI

Seems PDF follows the gausian Distribution Whereas CDF is like Sigmoid function S Curve.

PDF => Area under the Curve or the Density for Continuous Random Variable Probabilities.

Uniform Distribution

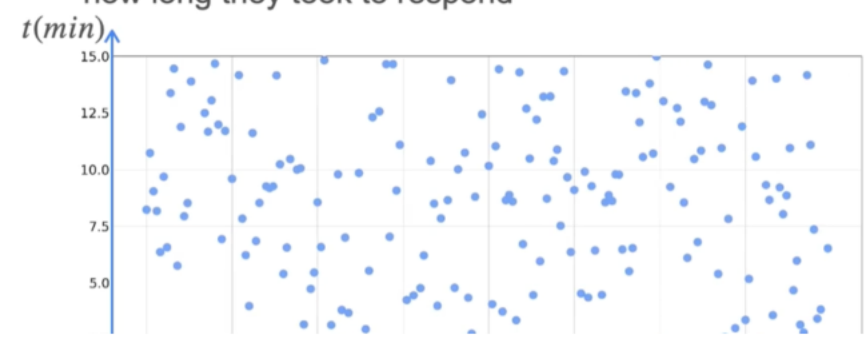
Graph looks like Homogenous in the Interval from 0-10 minutes. Sometimes I take 1 min, 5, 3, 2, 7, 9, 10 min as waiting time for bus arrival.

Uniform Distribution: Motivation

You're calling a tech support line.
They can answer any time between zero and 15 minutes and if they don't answer in this time, the line is disconnected.

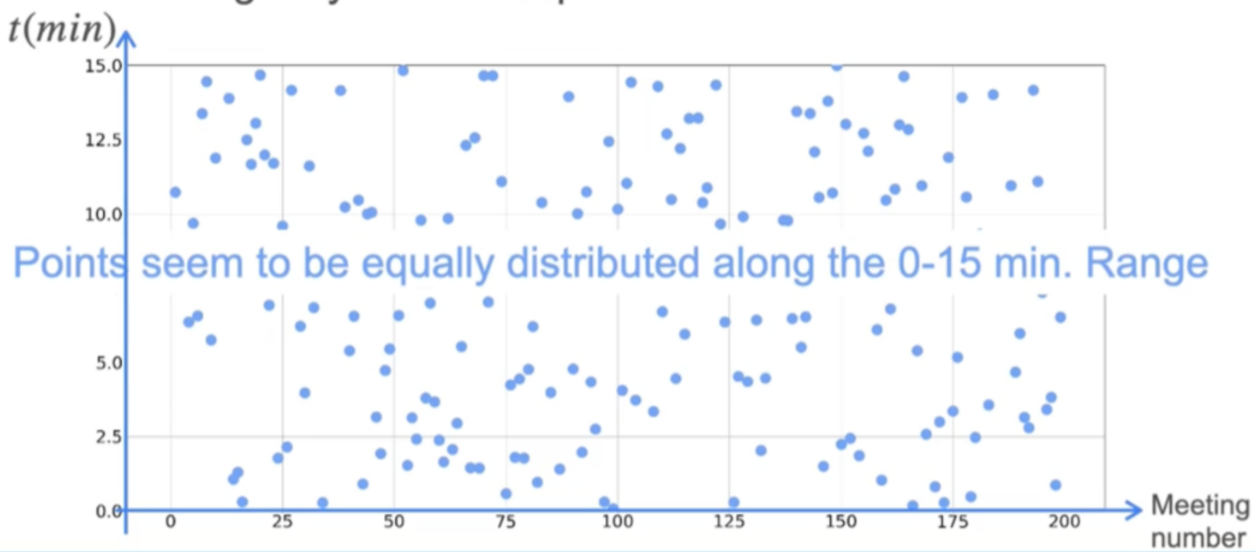
Uniform Distribution: Motivation

Last 200 times you called them, you took down notes of how long they took to respond

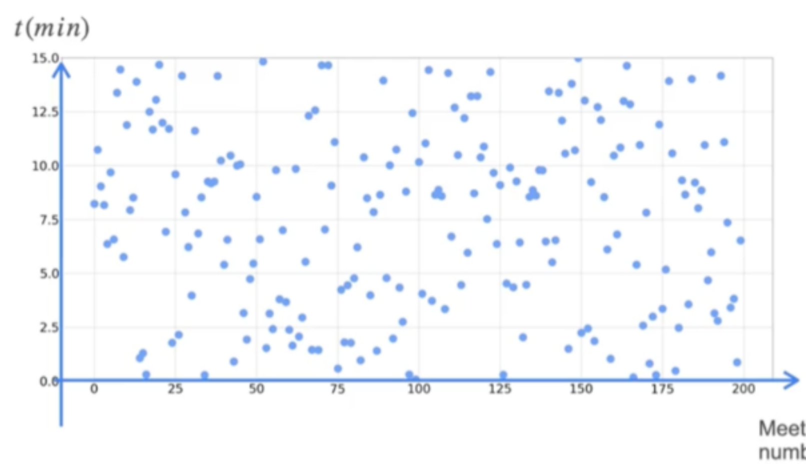


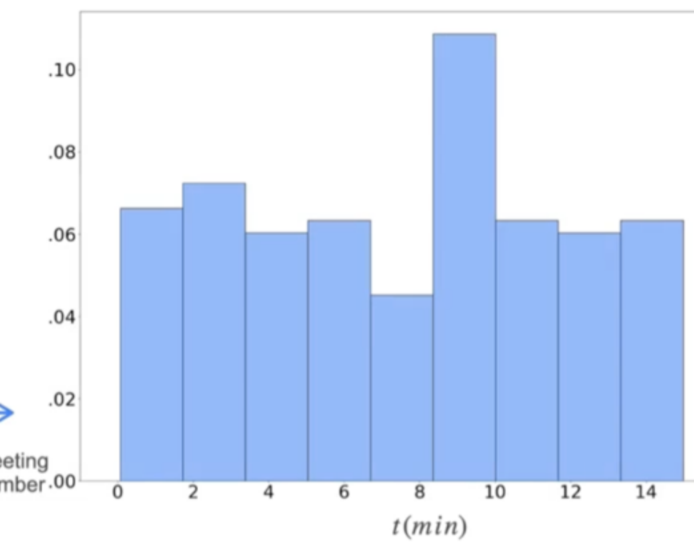
Uniform Distribution: Motivation

Last 200 times you called them, you took down notes of how long they took to respond



Uniform Distribution: Motivation





Uniform Distribution: Motivation

T: time (in minutes) you have to wait

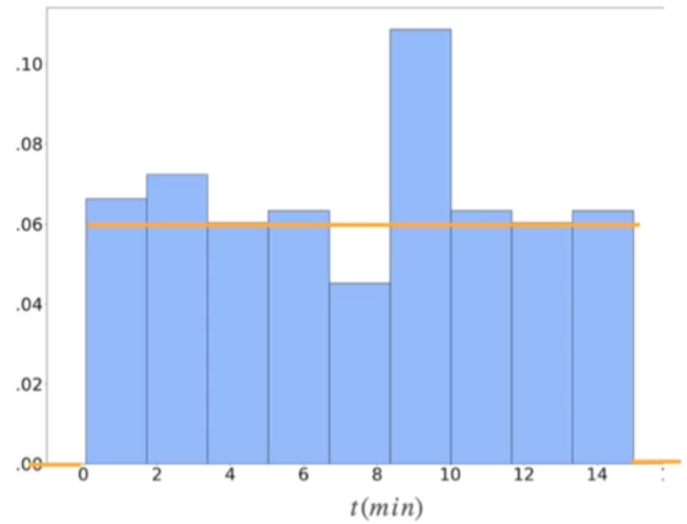
Any value between 0 and 15 minutes must have the same frequency of occurrence.

↓

The pdf must be constant for all values in the interval (0,15)

↓

Which constant? $\longrightarrow 15 \times h = 1 \longrightarrow h = \frac{1}{15} = 0.06$



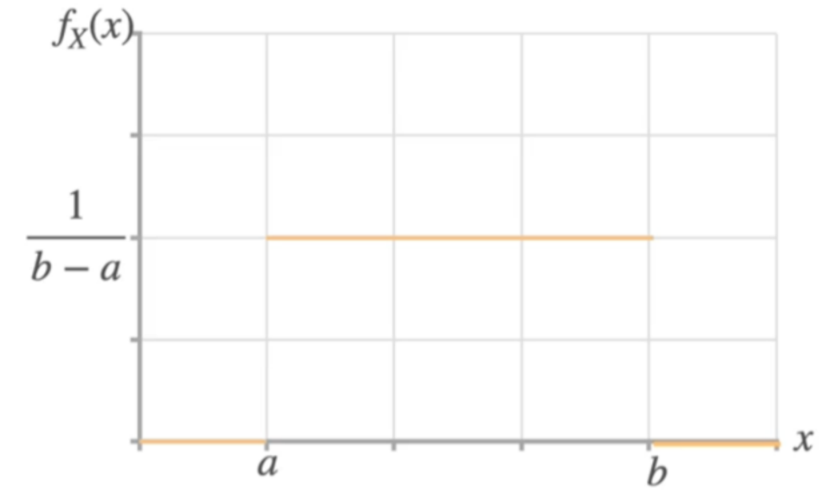
Uniform Distribution: Model

A continuous random variable can be modeled with a **uniform** distribution if all possible values lie in an interval and have the **same frequency** of occurrence

Parameters:

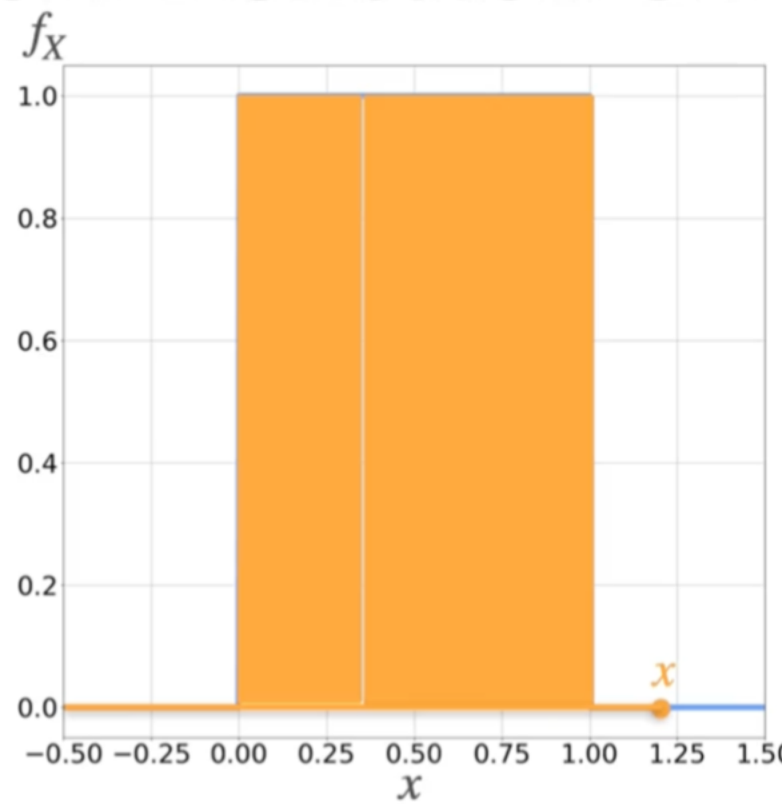
- a: beginning of the interval
- b: end of the interval

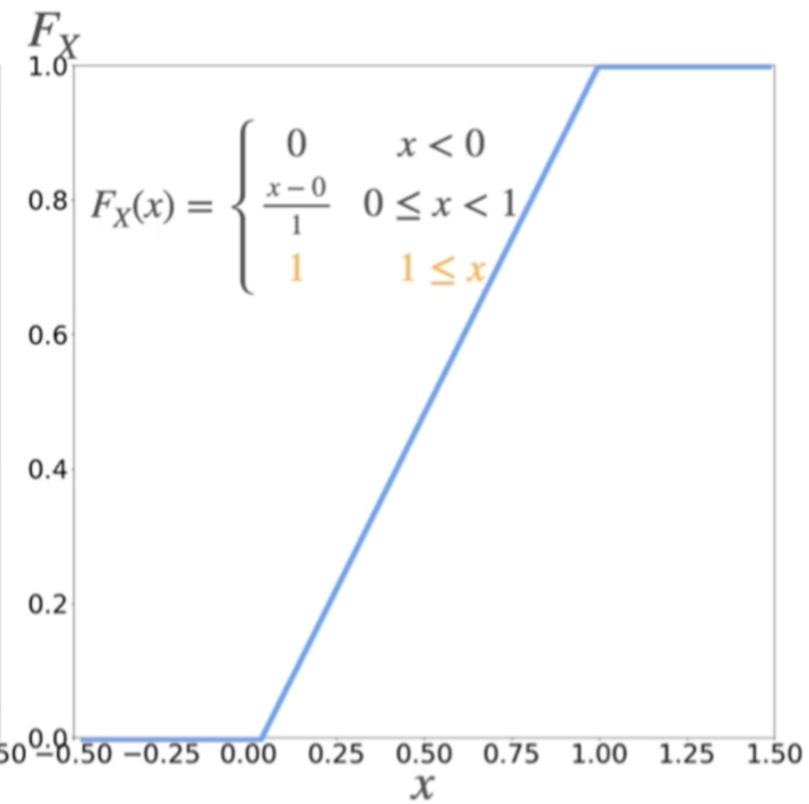
$$f_X(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & x \notin (a,b) \end{cases}$$



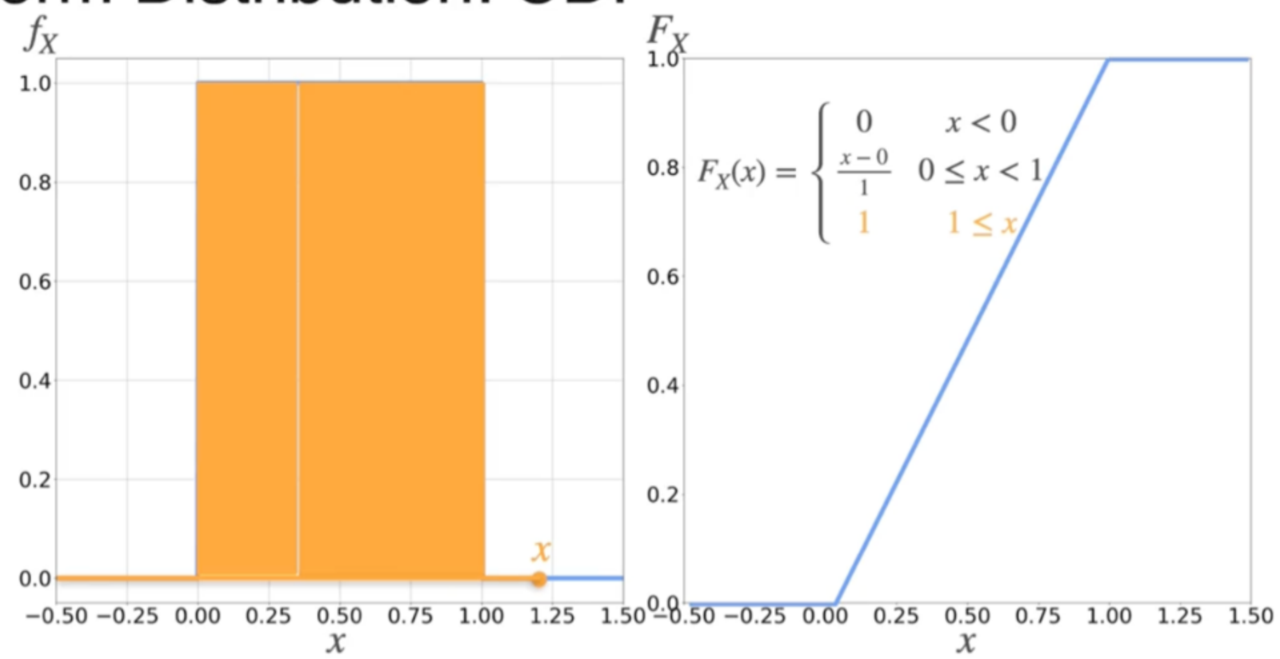
$$F_X(x) = \begin{cases} 0 \; ; & x < 0 \\ \frac{x-0}{1} & 0 \leq x < 1 \\ 1 & 1 \leq x \end{cases}$$

Uniform Distribution: CDF



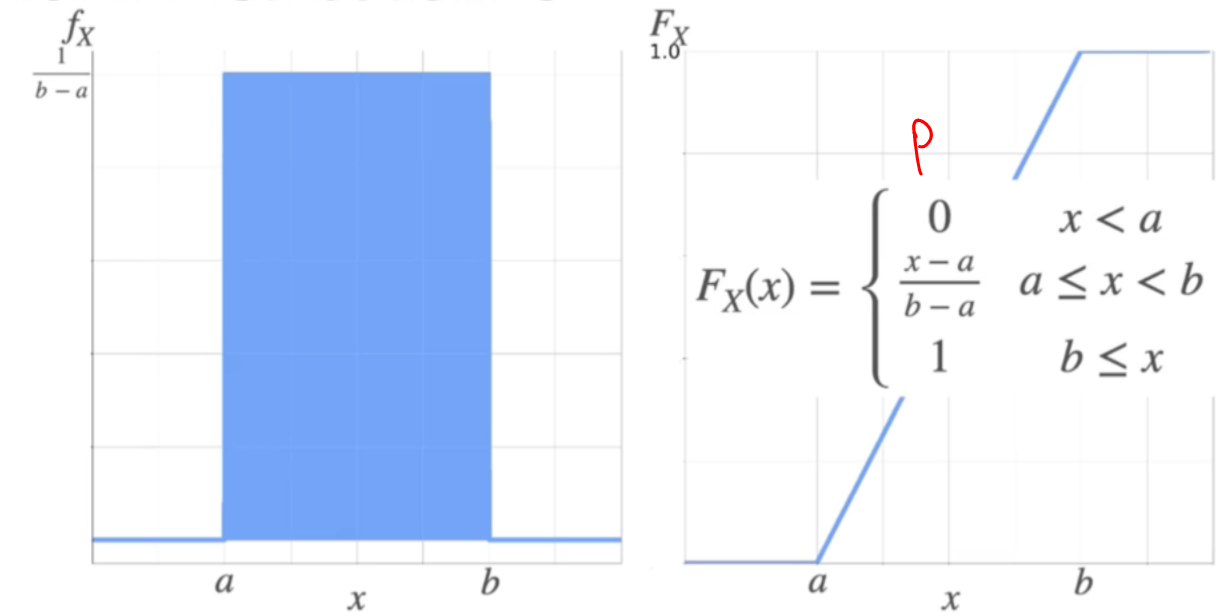


Uniform Distribution: CDF



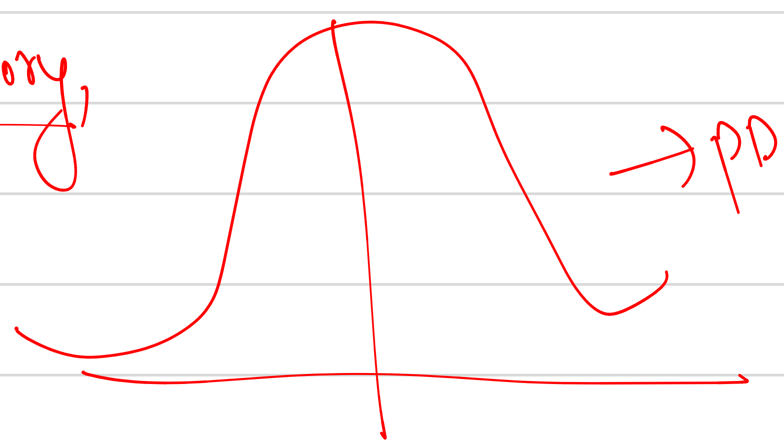
DeepLearning.AI

Uniform Distribution: CDF

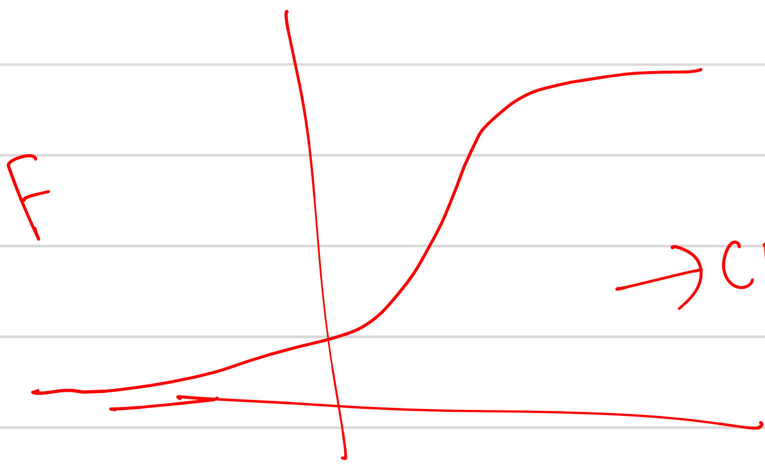


DeepLearning.AI

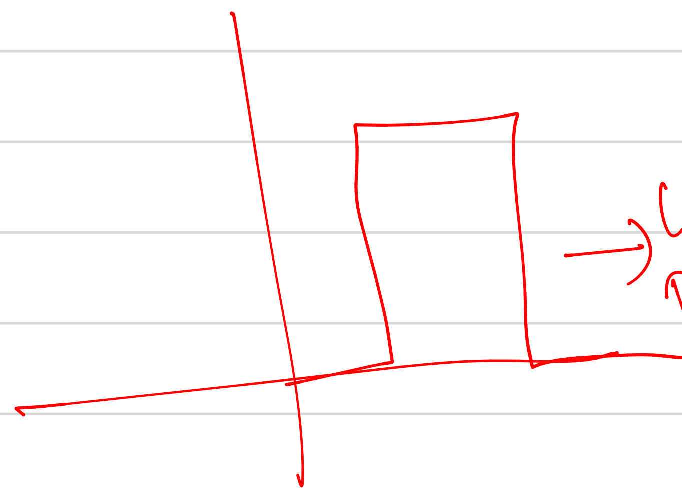
So In Summary,



→ PDF



→ CDF



→ Uniform Distribution

→ Area represents P Under the Continuous.

P starts from 0 to 1.

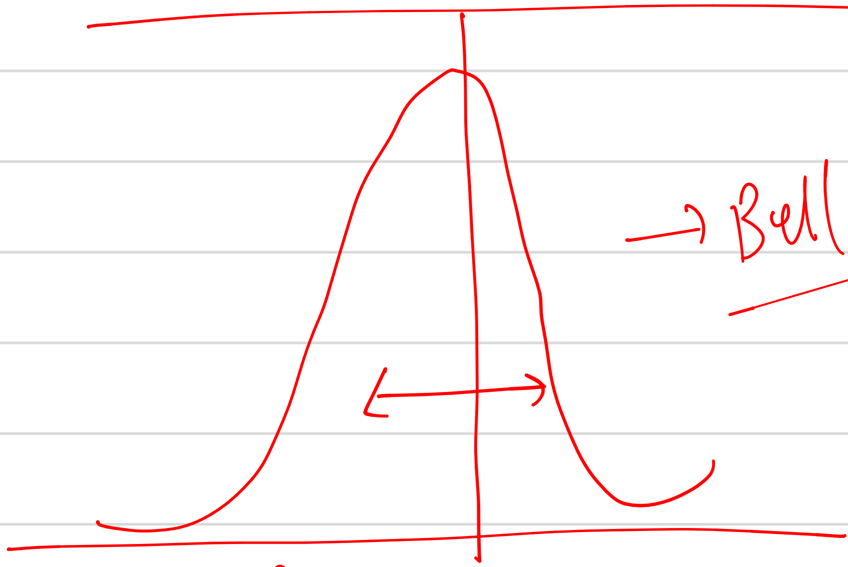
Eg: bus waiting time

→ PMF for Discrete.

Normal or Gaussian Distribution:

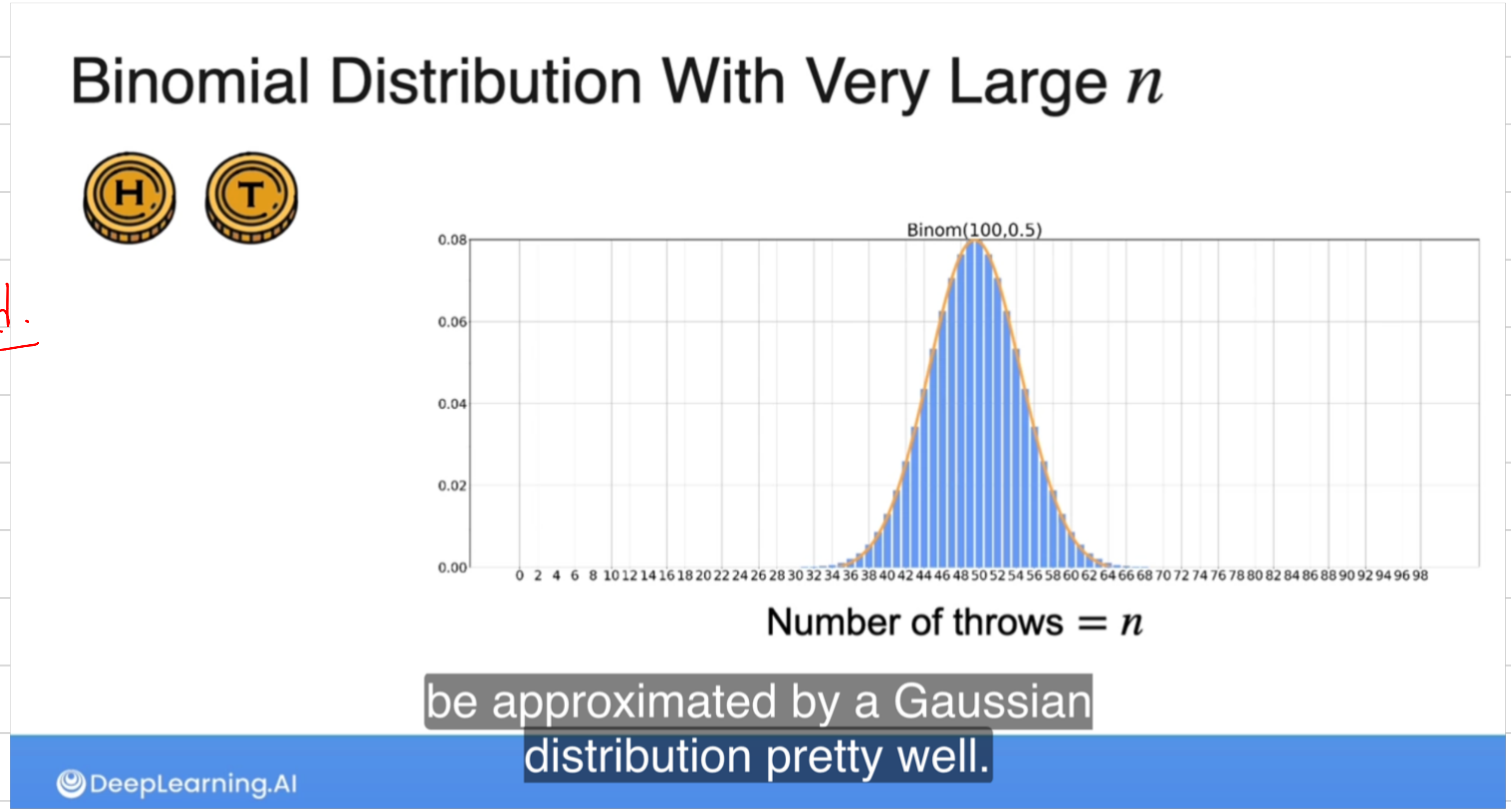
Let's take, $e^{-\frac{x^2}{2}}$

which can be bell-shaped.



→ Bell Curve.

$\sigma = 1$



$e^{-\frac{1}{2}\left(\frac{x-2}{3}\right)^2}$

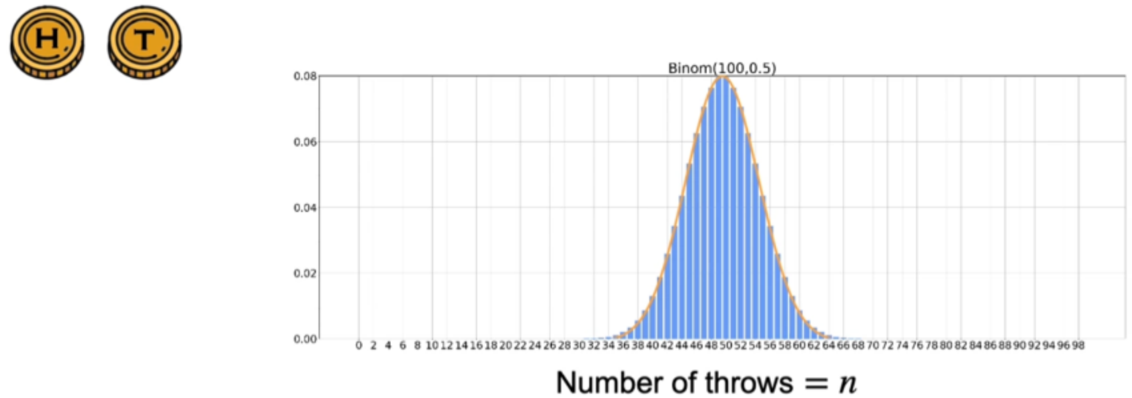
(σ tell how curve is thick.)

We have two parameters Mean = μ and Std. deviation = σ

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$\sigma \Rightarrow$ Flatness or wideness
whereas μ is the place where dist is centered.

Binomial Distribution With Very Large n

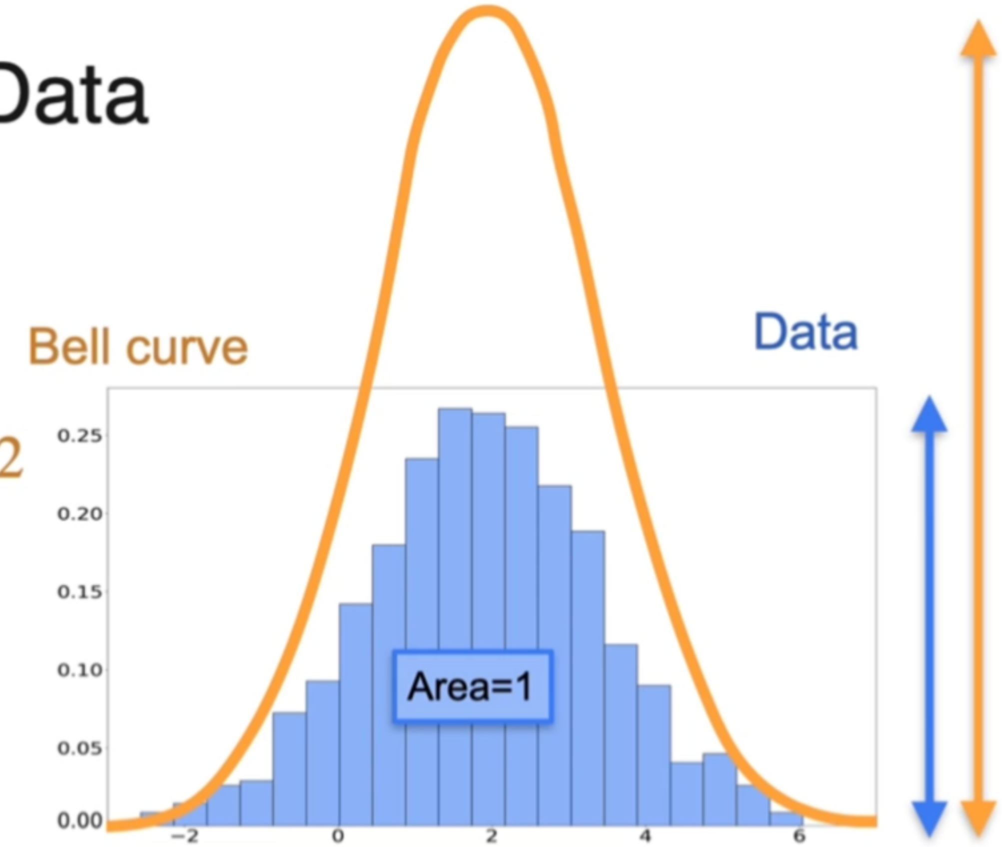


be approximated by a Gaussian distribution pretty well.

DeepLearning.AI

Bell Shaped Data

$$e^{-\frac{1}{2}\left(\frac{x-2}{3}\right)^2}$$

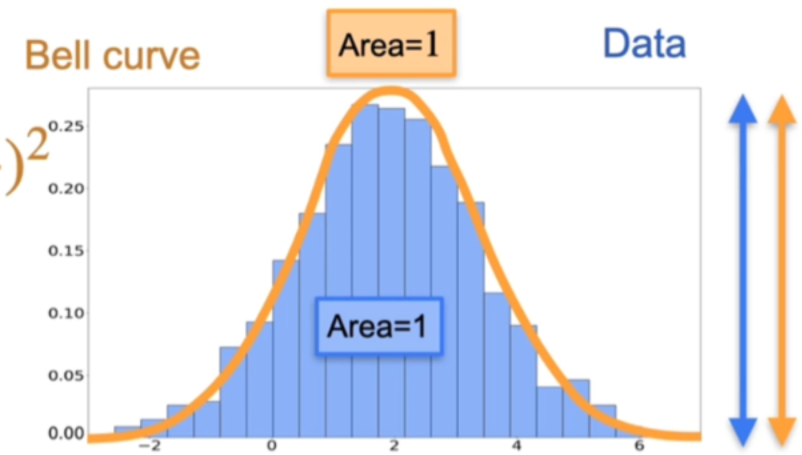


The area under the blue curve is 1 because it's a probability distribution, and

Making Adjustments we get,

Bell Shaped Data

$$\frac{1}{3\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{x-2}{3}\right)^2}$$



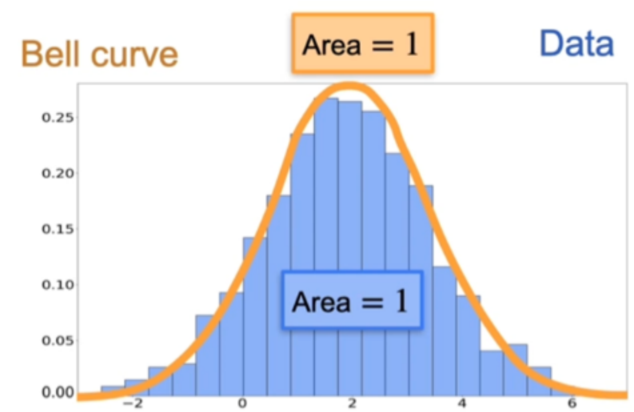
So that's the formula for the Gaussian or normal distribution.

DeepLearning.AI

Bell Shaped Data

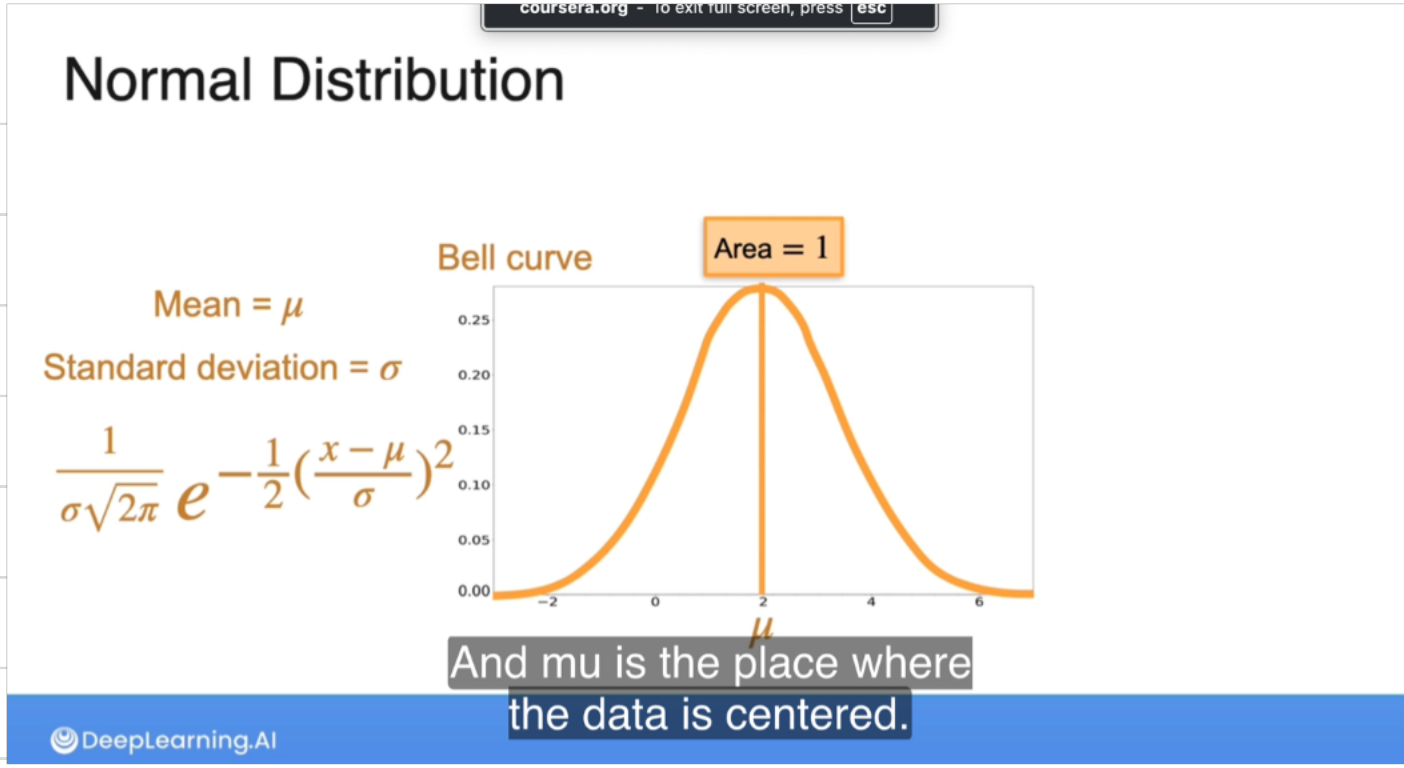
Mean = μ

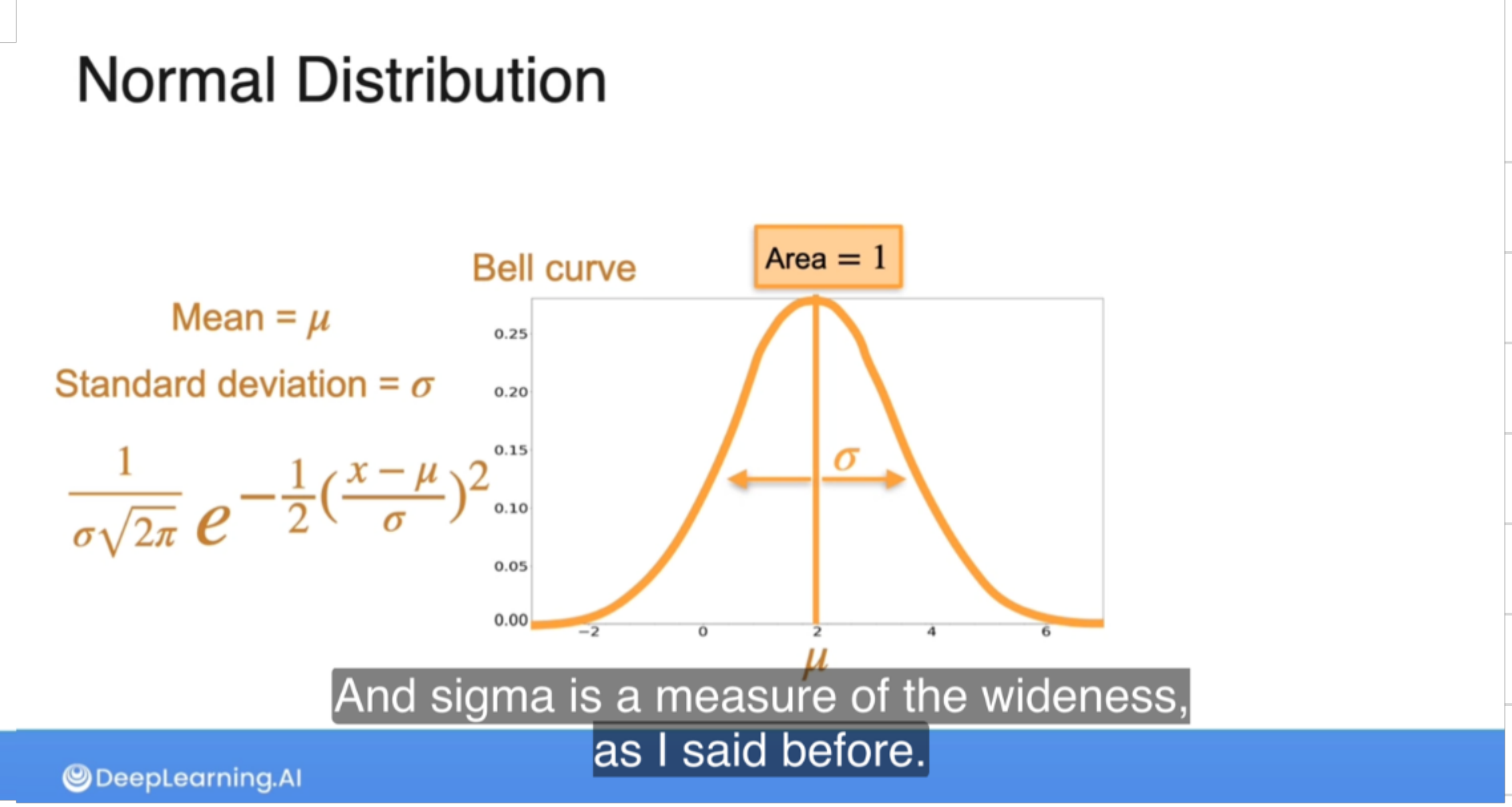
Standard deviation = σ



And if you want the mean to be mu and the standard deviation to be sigma,

DeepLearning.AI





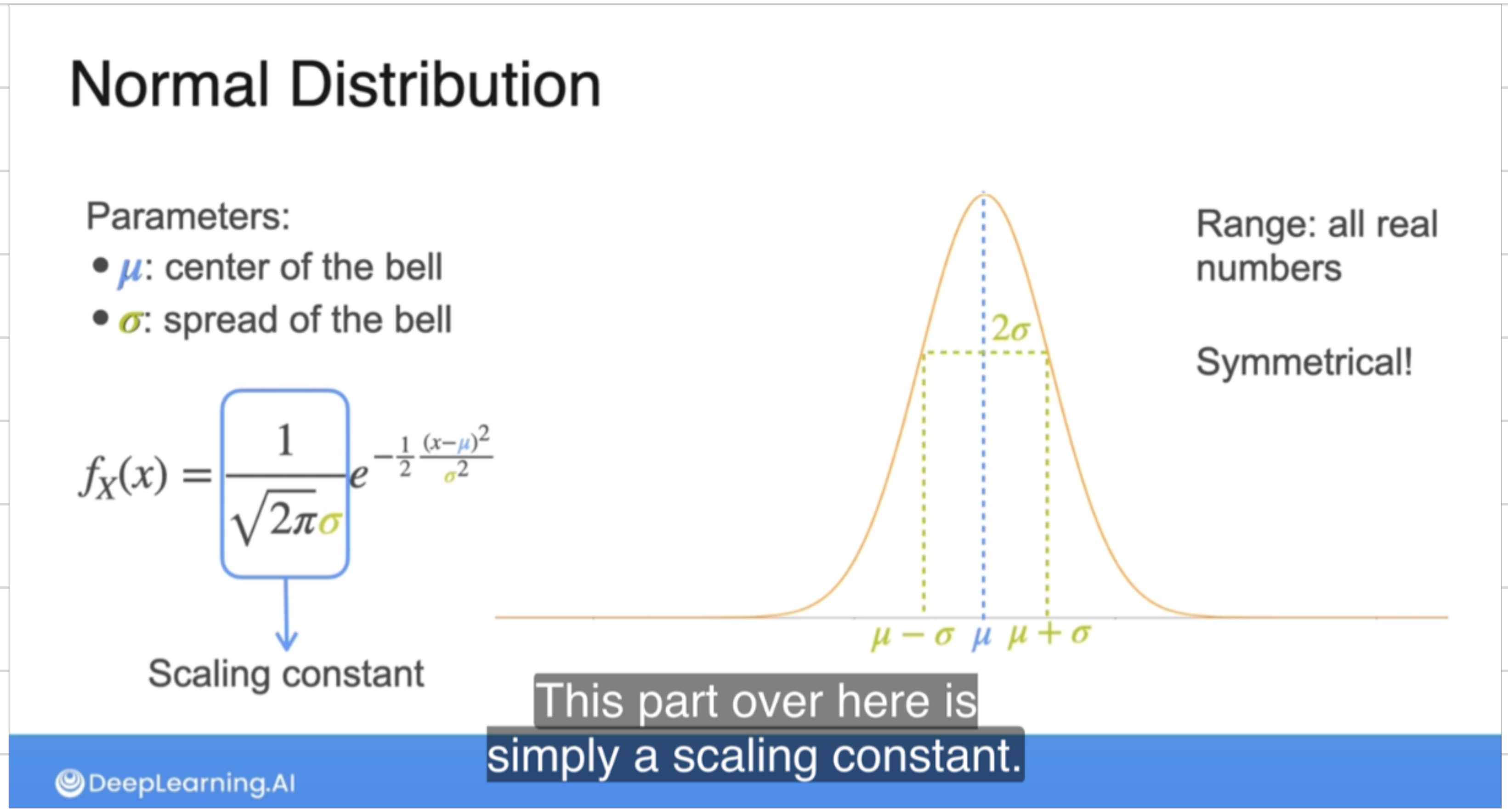
Parameters

μ : center of the bell

σ : spread of the bell

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} \cdot e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

Scaling Constant



So, depending upon μ and σ we have,

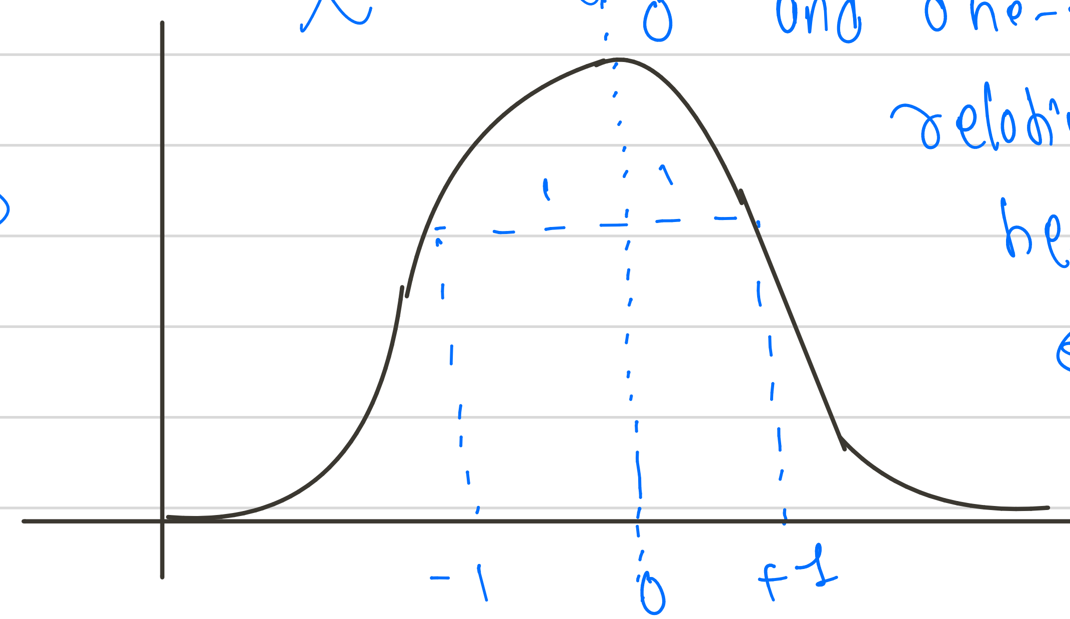
So if we have Random Variable X with this PDF we can write

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{2\sigma^2}} \Rightarrow X \sim N(\mu, \sigma^2)$$

— σ have to be always +ve.

$X \sim N(0,1)^2$ and one-to-one relationship between σ and μ.

Here, $\mu = 0$ and $\sigma = 2$



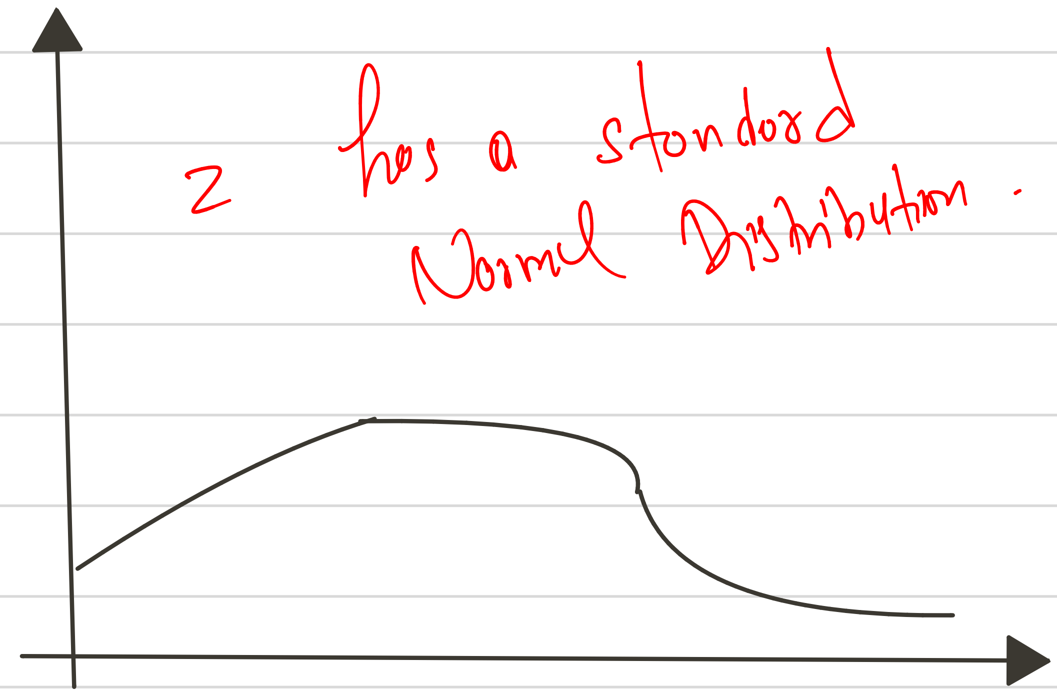
Standardization.

Convert any Normal Distribution to the standard one.

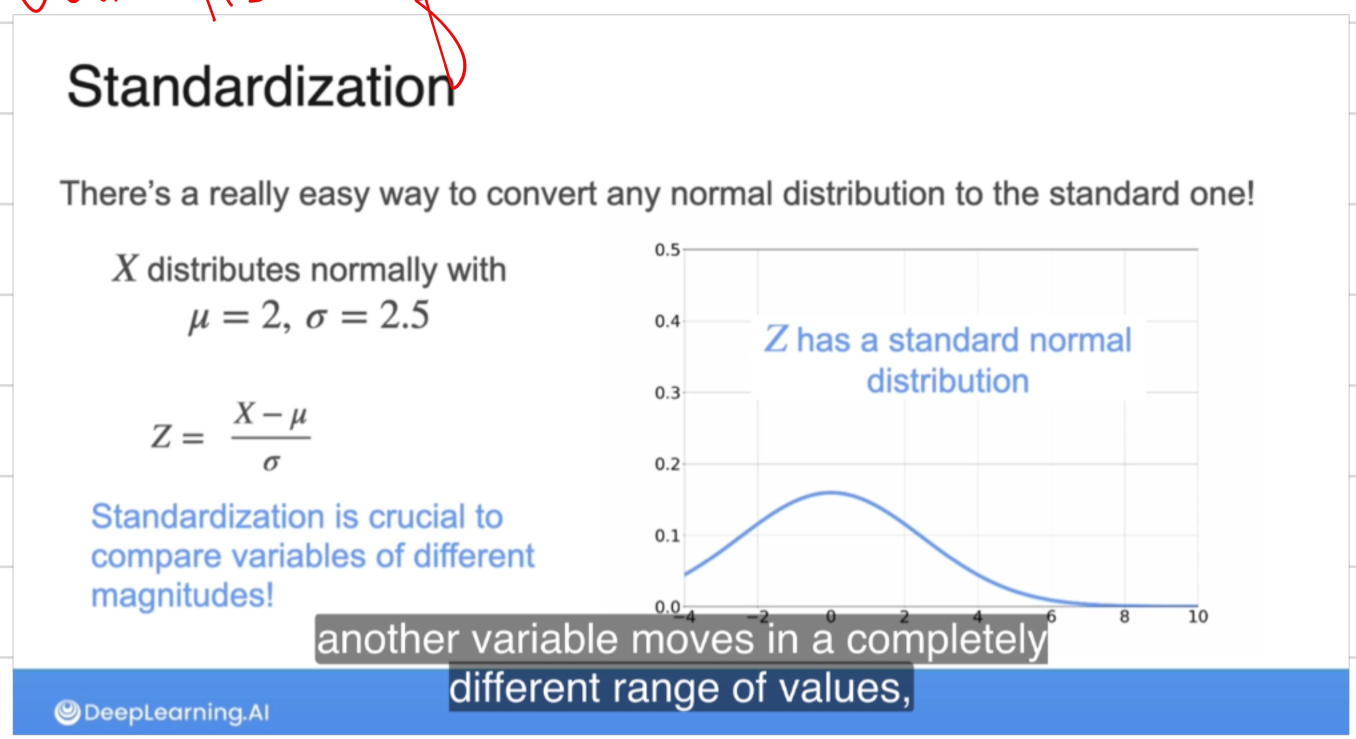
X distributes normally with

$\mu = 2, \ \sigma = 2.5$

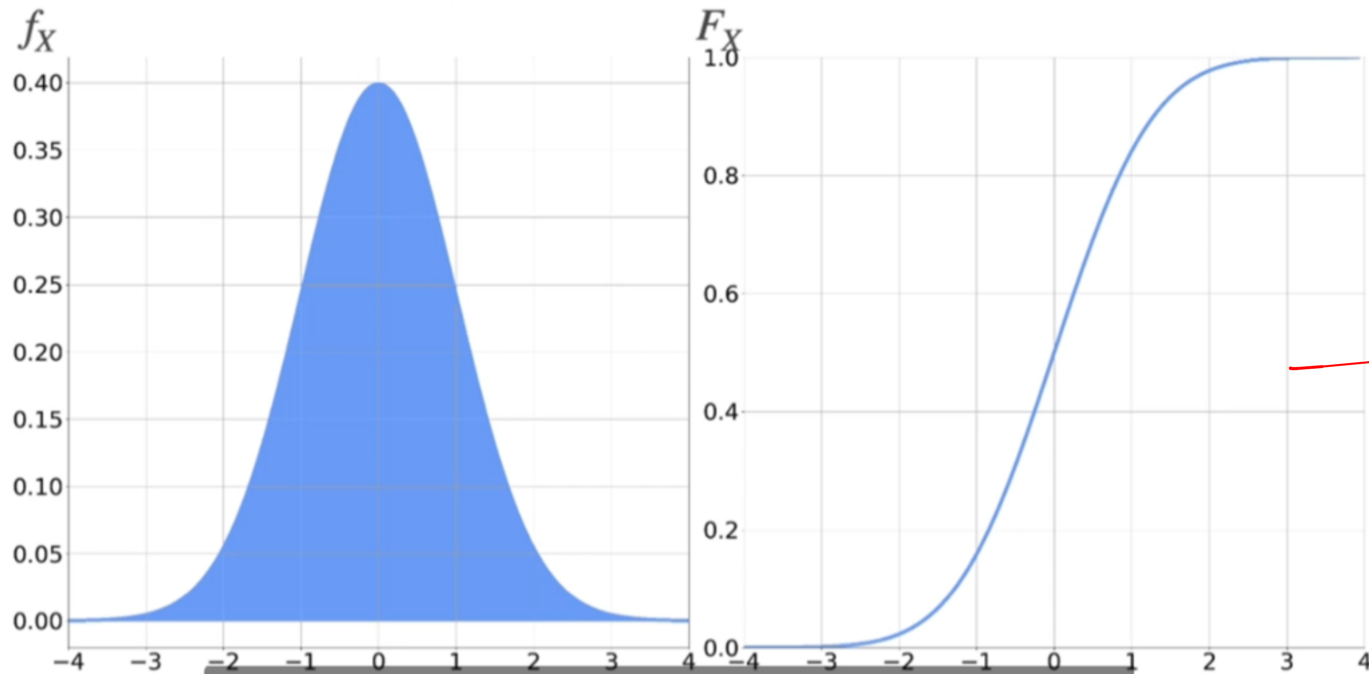
$$Z = \frac{X - \mu}{\sigma}$$



Standardization is crucial to compare variables of different magnitudes!



What Does the CDF Look Like?



As usual, it starts at 0 on the very left,
which is now the point at infinity,

DeepLearning.AI

Math can't be done by hand we use computer

→ Area computing is hard.

Source:

Coursera Probability & Statistics for Machine Learning.

Normal Distribution: Applications

Height

Weight

IQ

Noise in a communication channel

In general, characteristics that are the sum of many independent processes

Many models in ML are designed under the assumption that the variables follow a normal distribution

DeepLearning.AI