An overview of Distributions

A distribution shows the possible values a random variable can take and how frequently they occur.

Important Notation for Distributions:

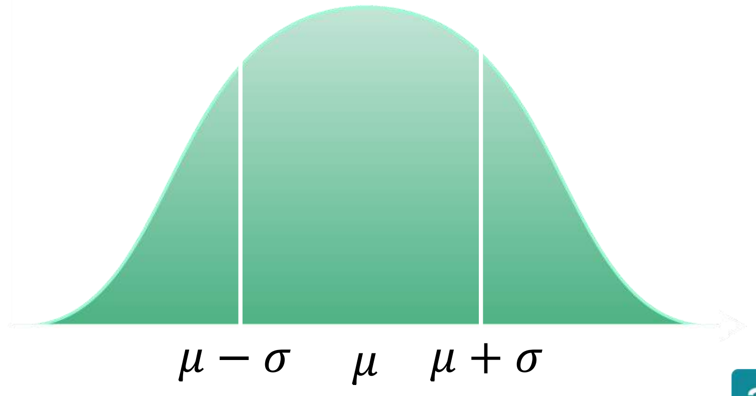
Y actual outcome

y one of the possible outcomes

$P(Y = y)$ is equivalent to $p(y)$.

We call a function that assigns a probability to each distinct outcome in the sample space, a **probability function.**

	Population	Sample
Mean	μ	$\bar{x}$
Variance	σ^2	s^2
Standard Deviation	σ	s



Notation

$Y \rightarrow$ The actual outcome of an event

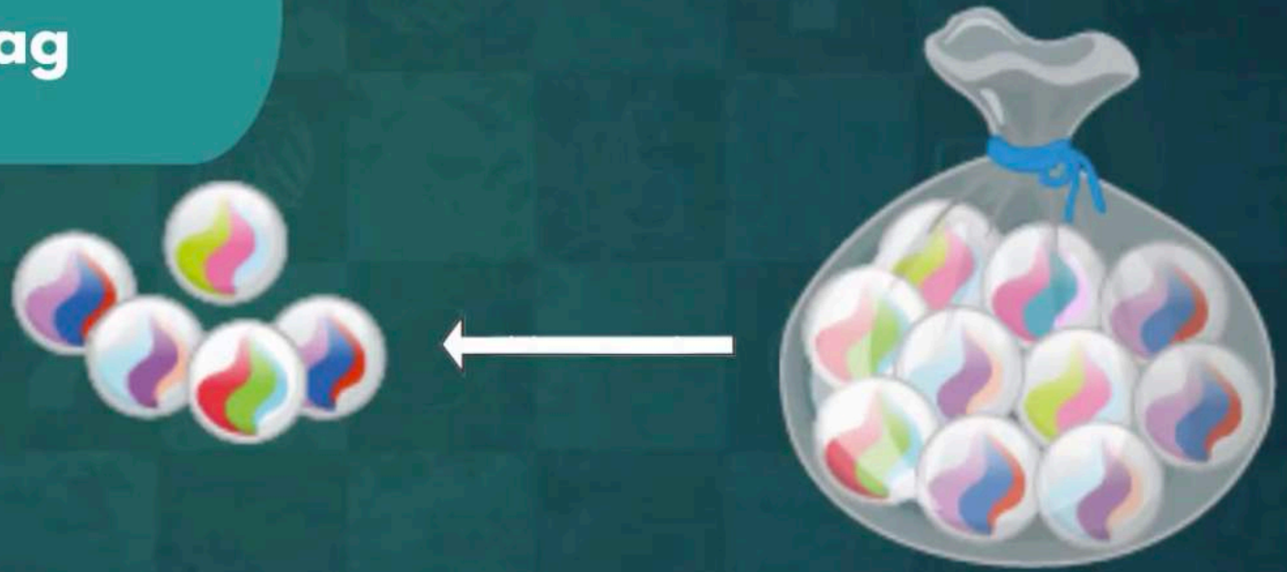
$y \rightarrow$ One of the possible outcomes

$P(Y=y)$ or $p(y)$

Example

Y → The number of red marbles we draw out of a bag

y → 5



P (Y = 5) or p (5)

$P(Y=5)$ or $p(5)$

1.40

Probability Frequency Distribution

Probabilities measure the likelihood of an outcome.

Fullscreen

Recall:

Probability Frequency Distribution

Recorded the frequency for each unique value and divide it by the total number of elements.

Sum	Frequency	Probability
2	1	1/36
3	2	1/18
4	3	1/12
5	4	1/9
6	5	5/36
7	6	1/6
8	5	5/36
9	4	1/9
10	3	1/12
11	2	1/18
12	1	1/36

Definitions

Two characteristics:

Fullscreen

mean → average value → μ

variance → how spread out the data is → σ^2

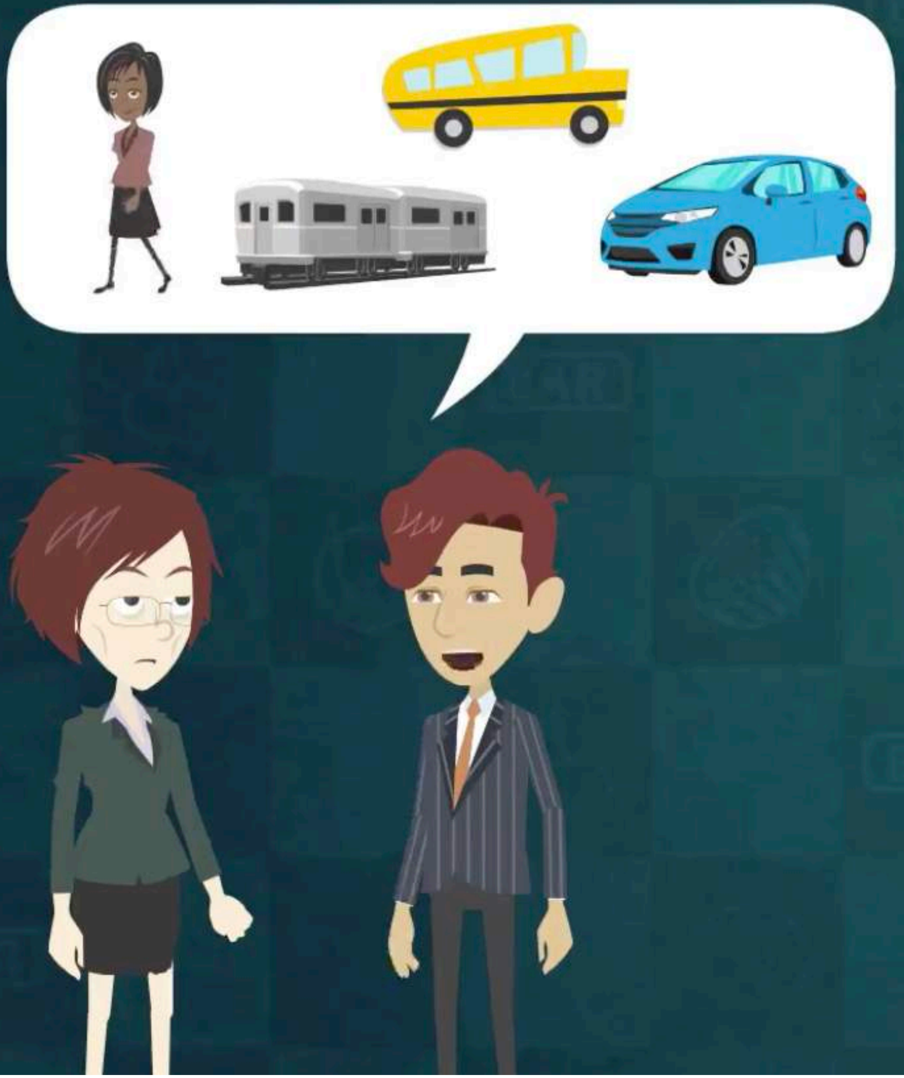
Population vs Sample

Population data — "all" the data

vs

Sample data — just a part of it

Population vs Sample



Entire department

Population of the department

Sample of the whole company

Population vs Sample

! Different notation:

sample mean

$\bar{x}$

sample variance

s^2

Variance

measured in squared units

X seconds

σ^2 seconds^2

← no direct interpretation

Standard deviation

Standard Deviation

Standard deviation → square root of variance $\sqrt{\sigma^2}$

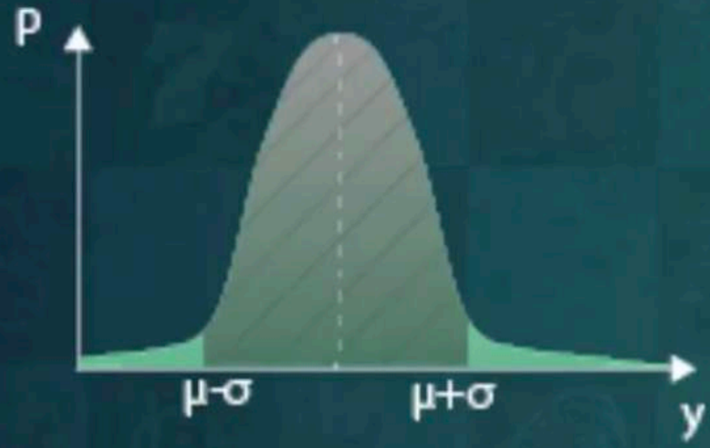
σ	population
s	sample

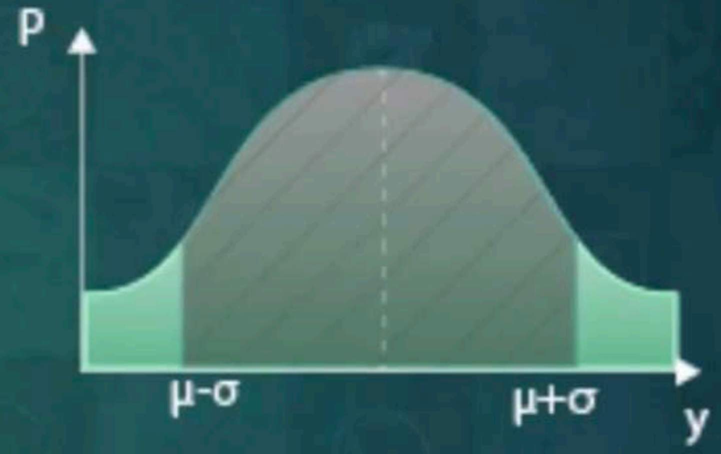
- same units as the mean
- we can directly interpret it

1.40

Standard Deviation

Fullscreen





the more congested the middle of the distribution

the more data falls within that interval

the less data falls within the interval

the more dispersed the data is

Mean and Variance

Fullscreen

a constant relationship between mean and variance

$$\sigma^2 = E((Y - \mu)^2) = E(Y^2) - \mu^2$$

Types of Distributions

Certain distributions share characteristics, so we separate them into **types**. The well-defined types of distributions we often deal with have elegant statistics. We distinguish between two big types of distributions based on the type of the possible values for the variable – discrete and continuous.

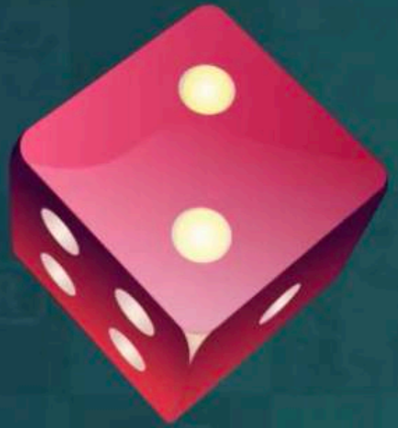
Discrete

- Have a finite number of outcomes.
- Use formulas we already talked about.
- Can add up individual values to determine probability of an interval.
- Can be expressed with a table, graph or a piece-wise function.
- Expected Values might be unattainable.
- Graph consists of bars lined up one after the other.

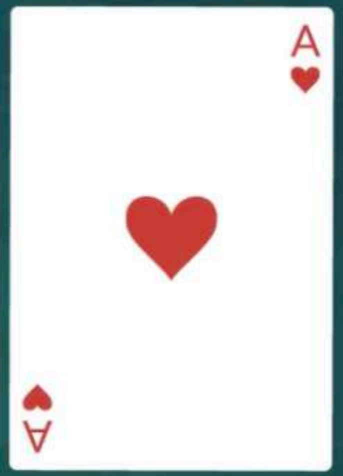
Continuous

- Have infinitely many consecutive possible values.
- Use new formulas for attaining the probability of specific values and intervals.
- Cannot add up the individual values that make up an interval because there are **infinitely many** of them.
- Can be expressed with a graph or a continuous function.
- Graph consists of a smooth curve.

Types of Probability Distributions



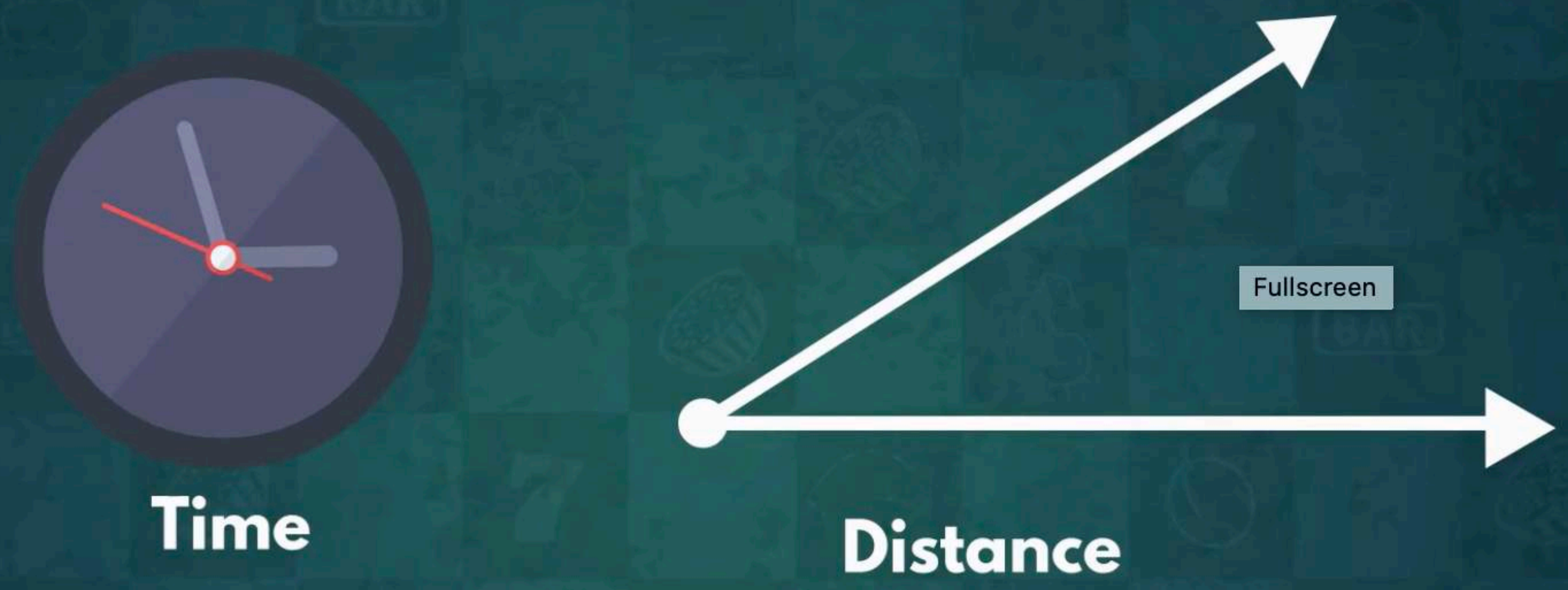
Die



Picking a card

Finite number of outcomes → Discrete distributions

Types of Probability Distributions



Infinitely many outcomes ➡ **Continuous distributions**

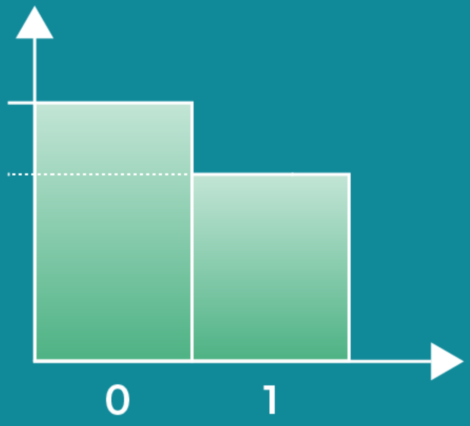
Notation

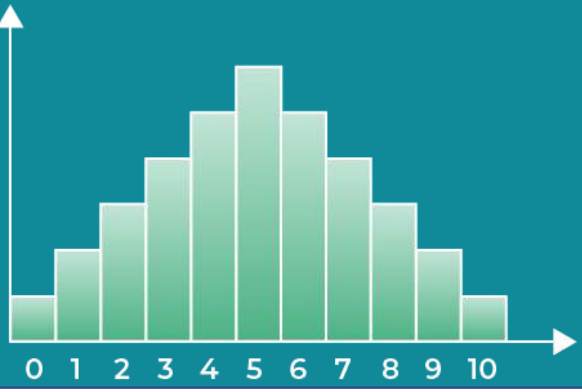
$$X \sim N(\mu, \sigma^2)$$ ← May vary

Variable Tilde Type Characteristics

Discrete Distributions

Discrete Distributions have finitely many different possible outcomes. They possess several key characteristics which separate them from continuous ones.





Key characteristics of discrete distribution

- Have a finite number of outcomes.
- Use formulas we already talked about.
- Can add up individual values to determine probability of an interval.
- Can be expressed with a table, graph or a piece-wise function.
- Expected Values might be unattainable.
- Graph consists of bars lined up one after the other.
- $\boldsymbol{P(Y \leq y) = P(Y < y + 1)}$

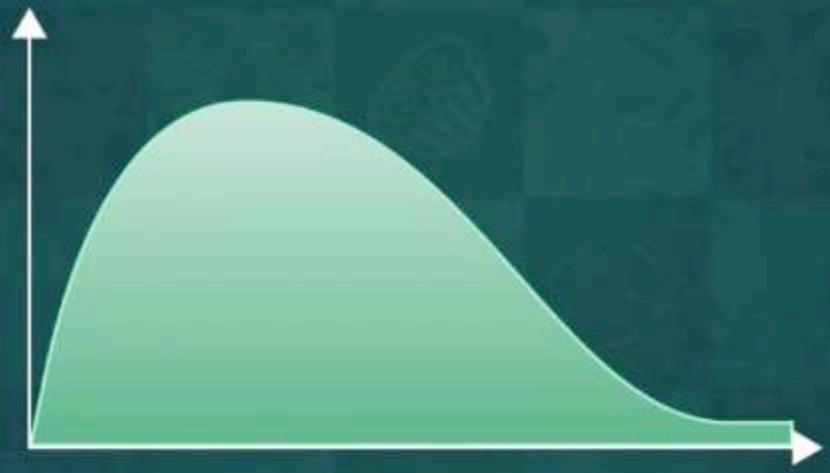
Examples of Discrete Distributions:

- Discrete Uniform Distribution
- Bernoulli Distribution
- Binomial Distribution
- Poisson Distribution

Discrete Distributions

Finitely many distinct outcomes

Y	P(y)

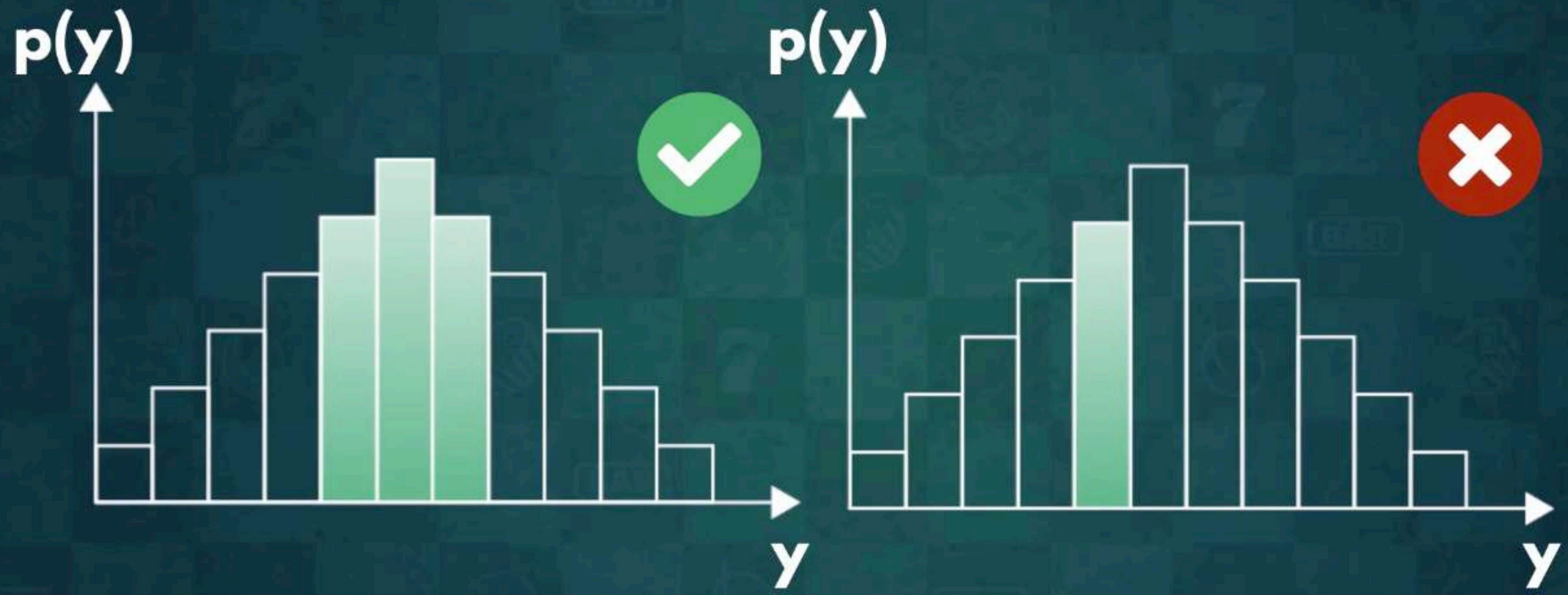


$$P(y) = \frac{1}{y!e}$$

Every unique outcome has a probability

1.50

Intervals



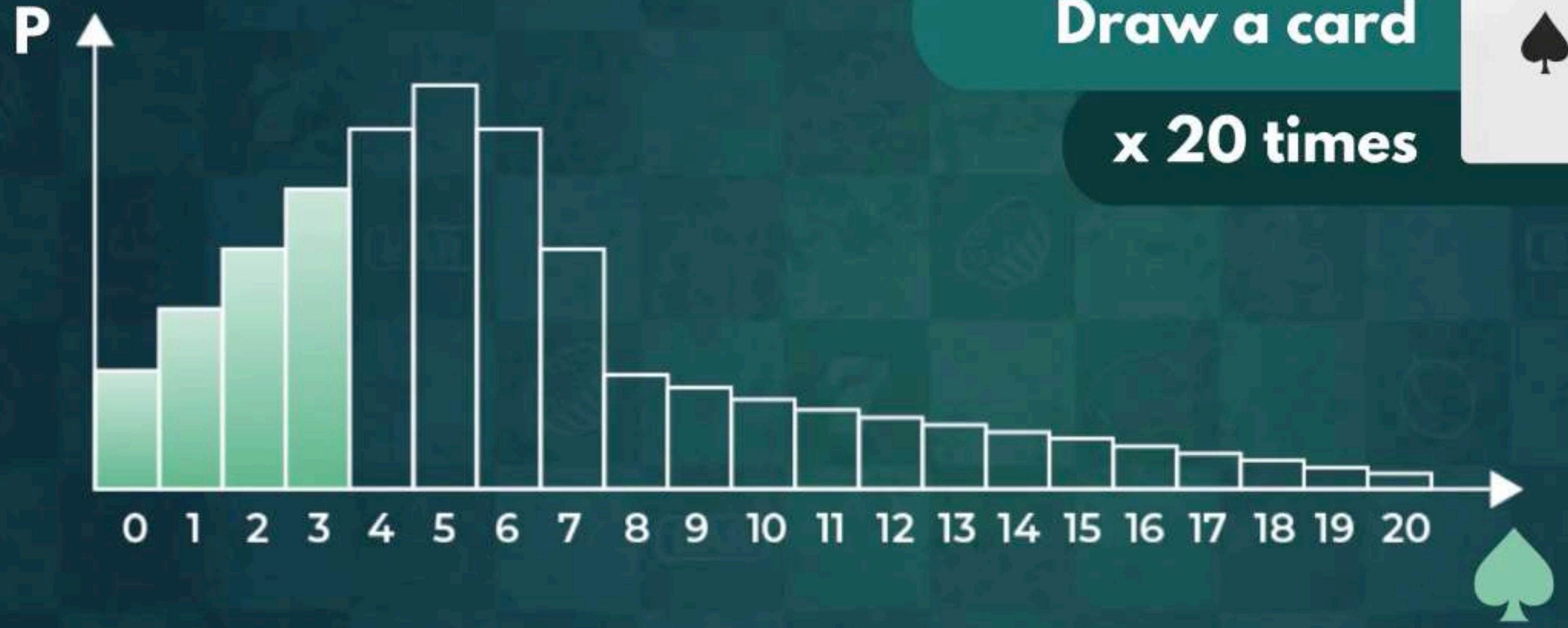
Add up the probabilities for all the values v

365√DataScience

Card Example

Draw a card

x 20 times



$$P(0) + P(1) + P(2) + P(3) = P(y \leq 3)$$

Peculiarity

$$P(Y \leq y) = P[Y < (y + 1)]$$

$$P(\spadesuit \leq 3) = P(\spadesuit < 4)$$

Uniform Distribution

A distribution where all the outcomes are equally likely is called a **Uniform Distribution**.

Notation:

- $Y \sim U(a, b)$
- * alternatively, if the values are categorical, we simply indicate the number of categories, like so: $Y \sim U(a)$

Key characteristics

- All outcomes are equally likely.
- All the bars on the graph are equally tall.
- The expected value and variance have no predictive power.

Example and uses:

- Outcomes of rolling a single die.
- Often used in shuffling algorithms due to its fairness.

Discrete Distributions

All outcomes are equally likely → Equiprobable

<u>Uniform</u> Distribution

365√DataScience

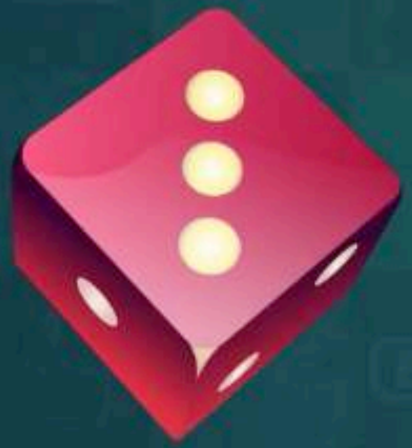
Uniform Distribution

$$U(a, b)$$

$$X \sim U(3, 7)$$

Uniform Distribution

All outcomes have equal probability



$$p(a) = p(b)$$

Die Example

$$P(1) = P(2) = P(3) = P(4) = P(5) = P(6)$$

Each individual outcome is equally likely

365 DataScience

Everyday Situation

X ➡ Choosing a specific chocolate

$X \sim U(3)$

365 DataScience

Expected Value

Provides us no relevant information

E () = 3.5

No predictive power

Mean and Value

$\mu = 3.5$

$\sigma^2 = \frac{105}{36}$

- Completely uninterpretable
- No real intuition behind what they mean

Main Takeaway

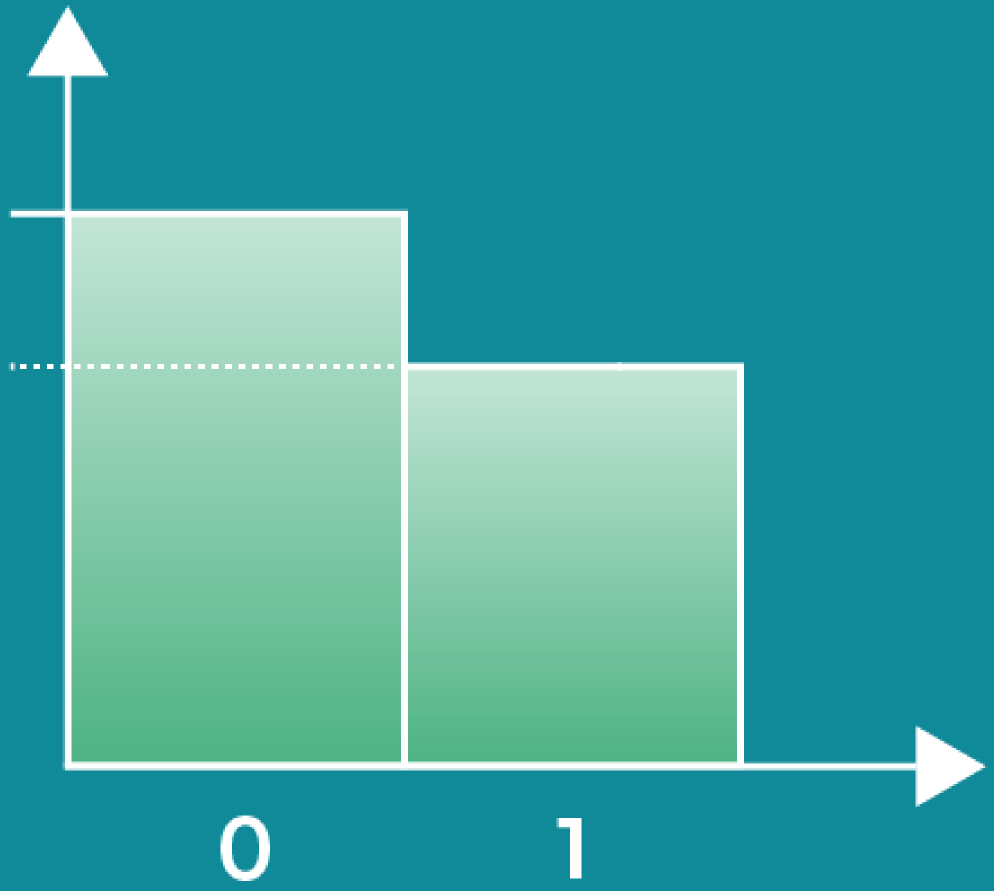
$$X \sim U(a, b)$$

- Each outcome is equally likely
- Both the mean and the variance are uninterpretable
- No predictive power

365 DataScience

Bernoulli Distribution

A distribution consisting of a single trial and only two possible outcomes – success or failure is called a **Bernoulli Distribution.**



Notation:

- $Y \sim Bern(p)$

Key characteristics

- One trial.
- Two possible outcomes.
- $E(Y) = p$
- $Var(Y) = p \times (1 - p)$

Example and uses:

- Guessing a single True/False question.
- Often used in when trying to determine what we expect to get out a single trial of an experiment.

Discrete Distributions

Events with only two possible outcomes

True False

Bernoulli Distribution

Any event with two outcomes can be transformed into a Bernoulli event

Bernoulli Distribution

7 Native → ✓ 3 International → ✗

1.40

Bernoulli Distribution

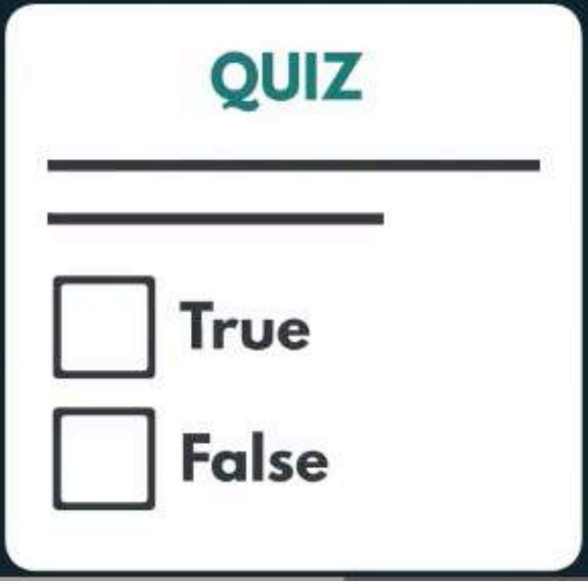
$$\mathrm{Bern}(p)$$

$$X \sim \mathrm{Bern}(p)$$

Bernoulli Distribution

Events with:

- 1 Trial
- 2 Possible outcomes



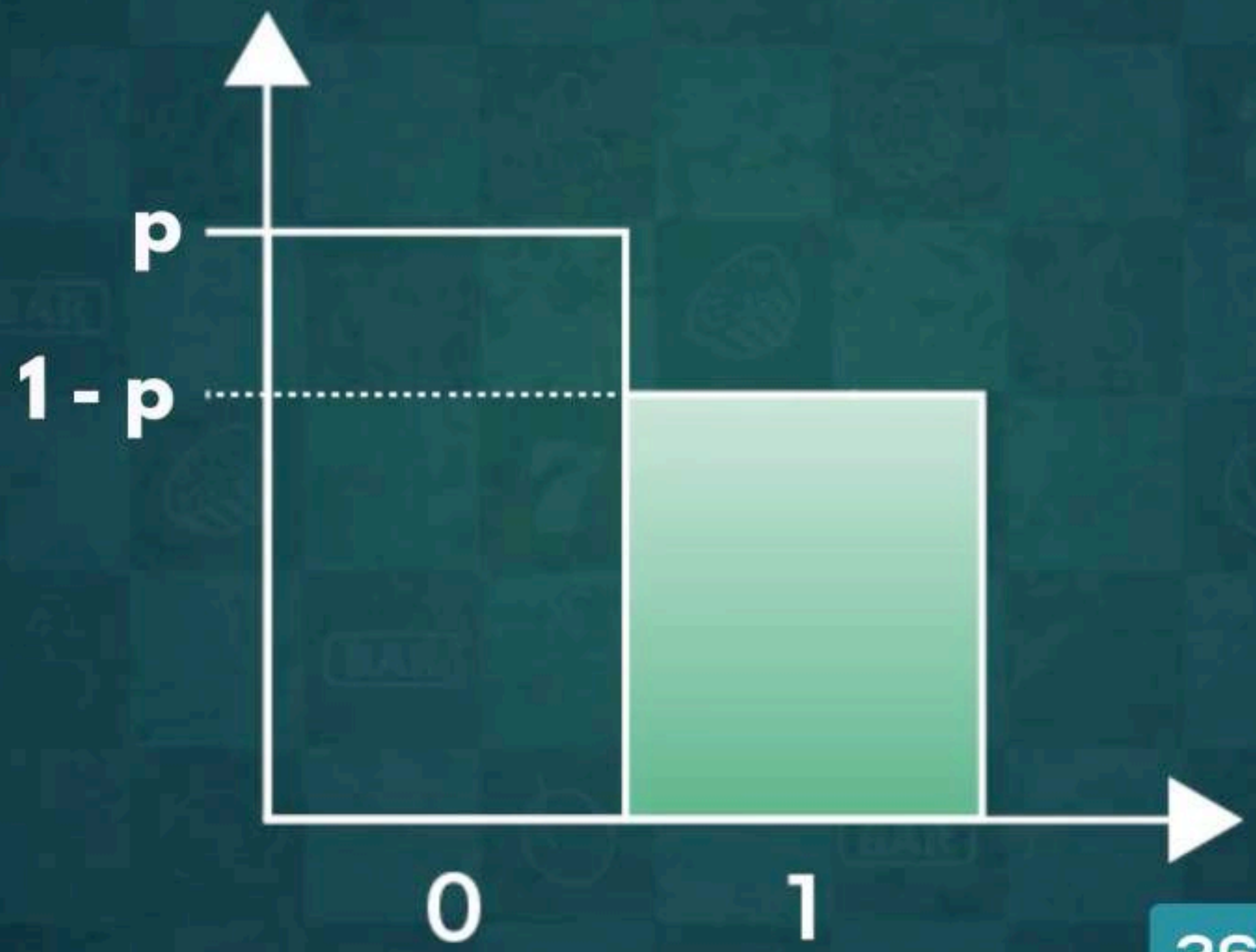
Bernoulli Distribution

p is known

OR

past data indicating some experimental probability

Graph of Bernoulli Distribution



Assigning 0 and 1

Assigning which outcome is 0, and which outcome is 1

 → 0

 → 1

$$E(X) = p$$

Assigning 0 and 1

Conventionally:

$p \leftarrow 1$

$1 - p \leftarrow 0$

$p > 1 - p$

Expected Values

$$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p$$

The likelihood of the favoured event

1 Trial ➡ We expect that outcome to occur

Variance

$$\sigma^2 = (x_0 - \mu)^2 \cdot P(x_0) + (x_1 - \mu)^2 \cdot P(x_1) =$$

$$= (0 - p)^2 \cdot (1 - p) + (1 - p)^2 \cdot p =$$

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

True, regardless of the expected value

Variance

$$\sigma^2 = p(1-p)$$

$$\sigma = \sqrt{p(1-p)}$$

They bring us little value

Coin Example

Unfair coin → P() = 0.6

 → 1

 → 0

$E(X) = p = 0.6$

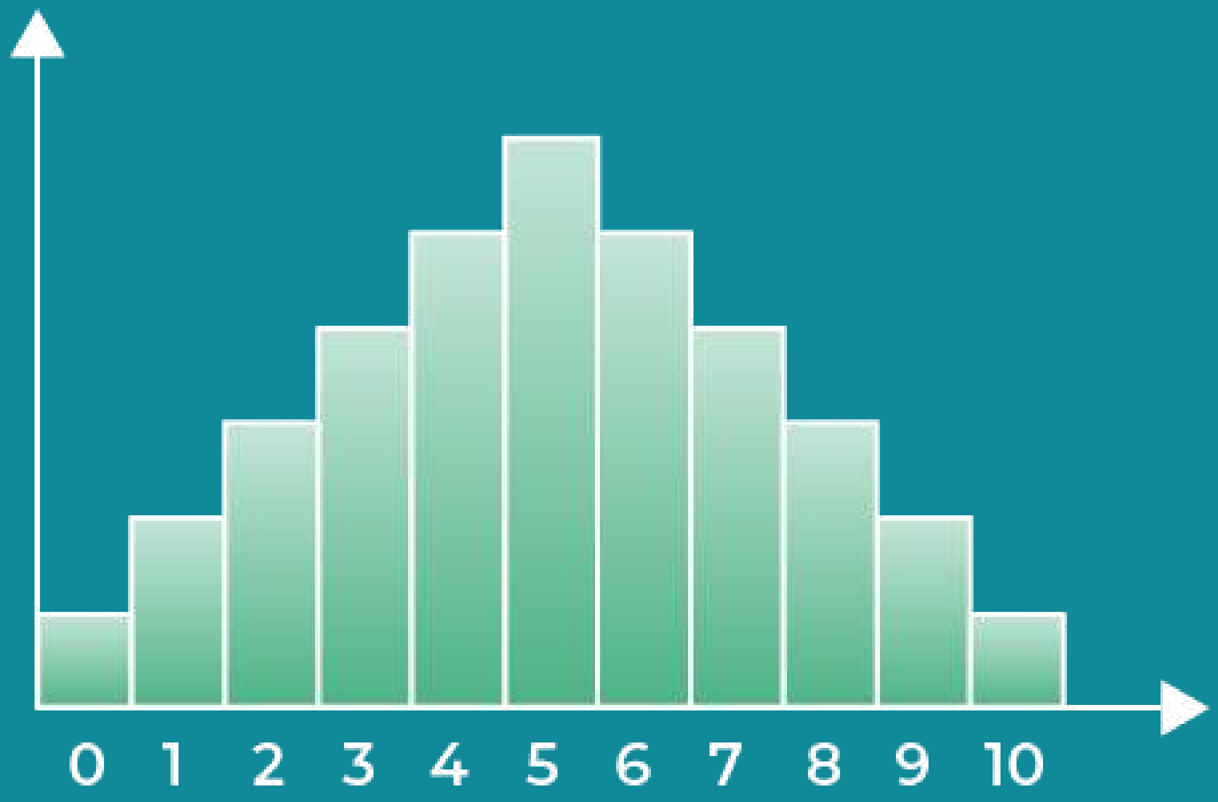
Variance

$$\sigma^2 = p(1 - p) = 0.6 \times 0.4 = 0.24$$

Binomial Distribution

Sometimes instead of wanted to know which of two outcomes is more probable, we want to know how often if would occur over several trials.

A sequence of identical Bernoulli events is called Binomial and follows a **Binomial Distribution**.



Notation:
- $Y \sim B(n, p)$

Key characteristics
- Measures the frequency of occurrence of one of the possible outcomes over the n trials.
- $P(Y = y) = C(y, n) \times p^y \times (1 - p)^{n-y}$
- $E(Y) = n \times p$
- $Var(Y) = n \times p \times (1 - p)$

Example and uses:
- Determining how many times we expect to get a heads if we flip a coin 10 times.
- Often used when trying to predict how likely an event is to occur over a series of trials.

365√DataScience

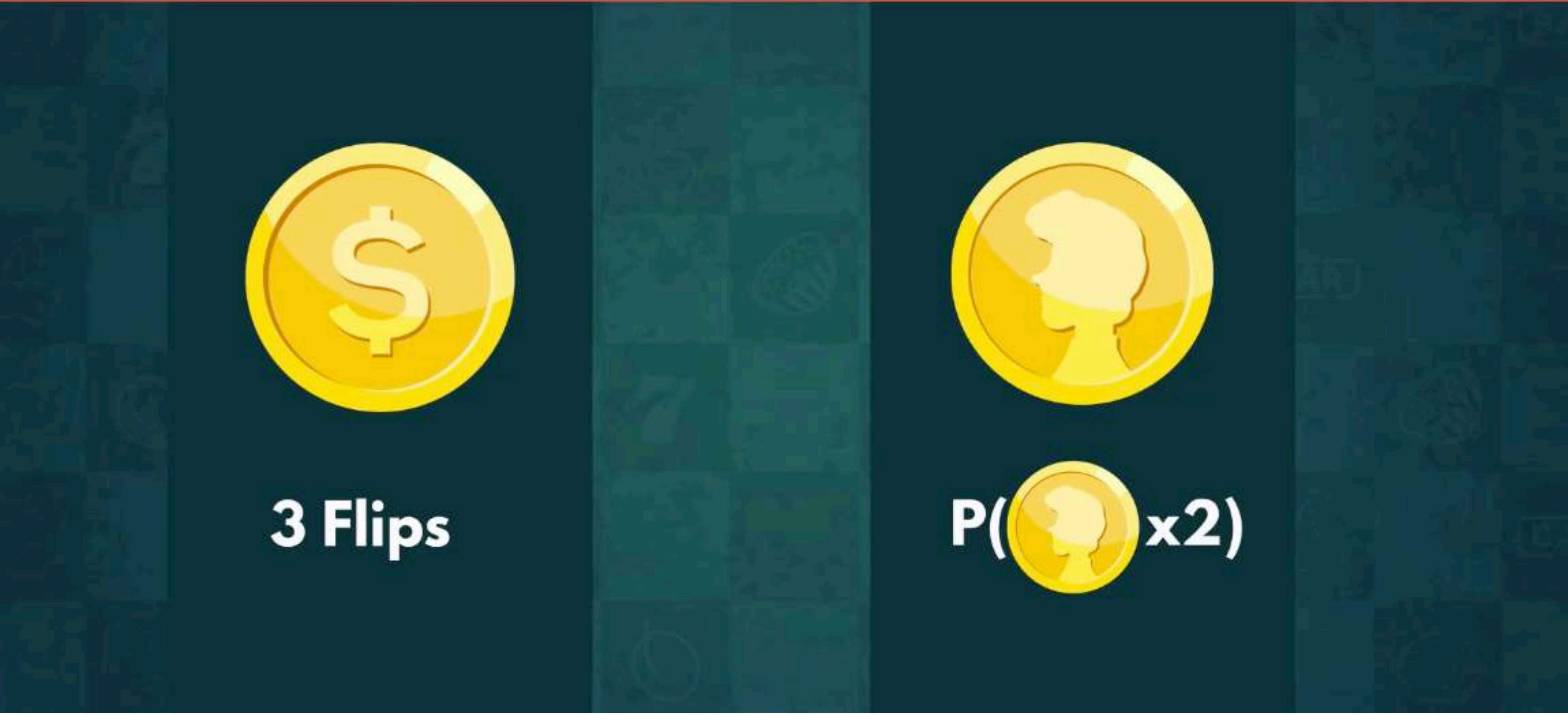
Discrete Distributions

Carrying out a similar experiment several times in a row

Binomial Distribution

- Two outcomes per iteration
- Many iterations

Binomial Distribution



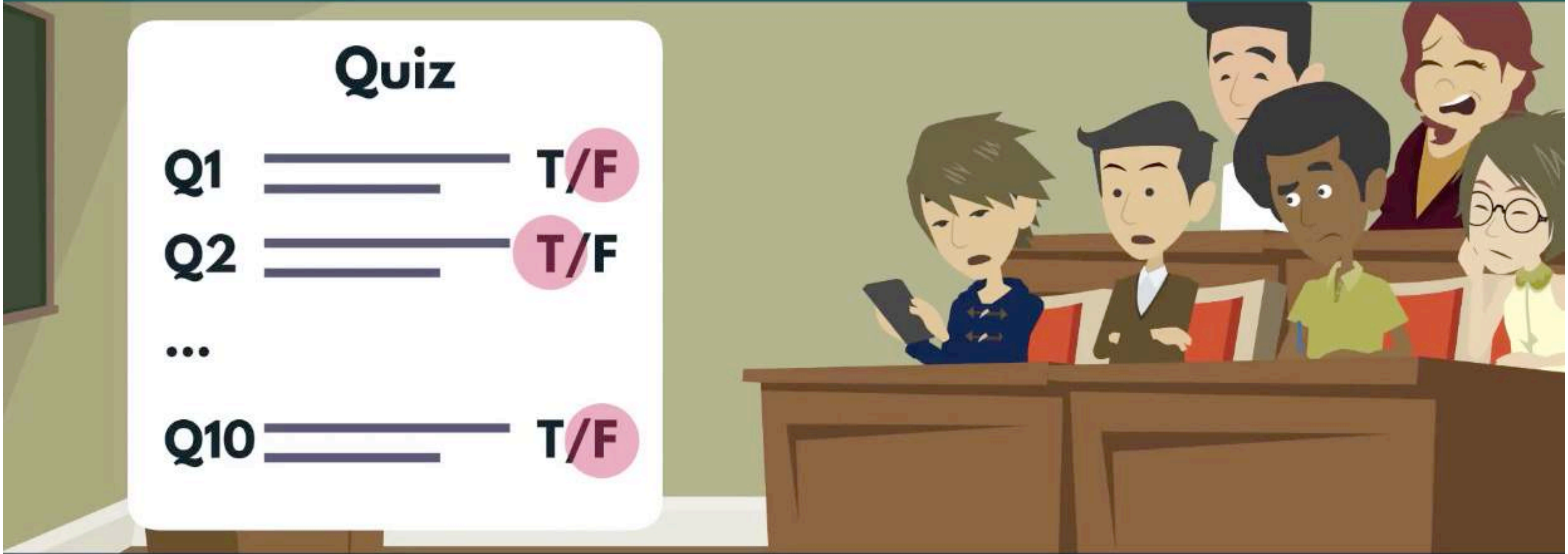
Binomial Distribution

Notation

$$B\,(n,\ p)$$

$$X \sim B\,(10, 0.6)$$

Bernoulli vs Binomial Distribution



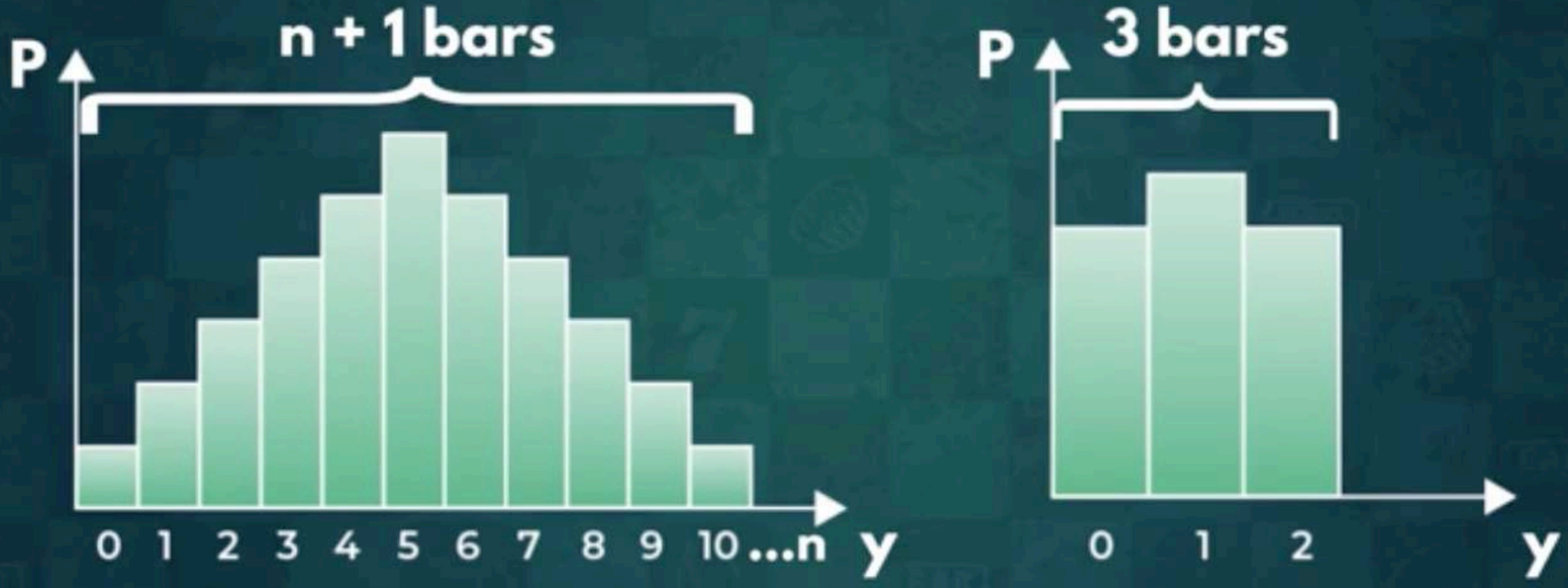
- Guessing 1 question → **Bernoulli event**
- Guessing the entire quiz → **Binomial Eve**

Bernoulli vs Binomial Distribution

E(Bernoulli event) → Which outcome we expect for a single trial

E(Binomial Event) → The number of times we expect to get a specific outcome

Graph of Binomial Distribution



The Probability Function

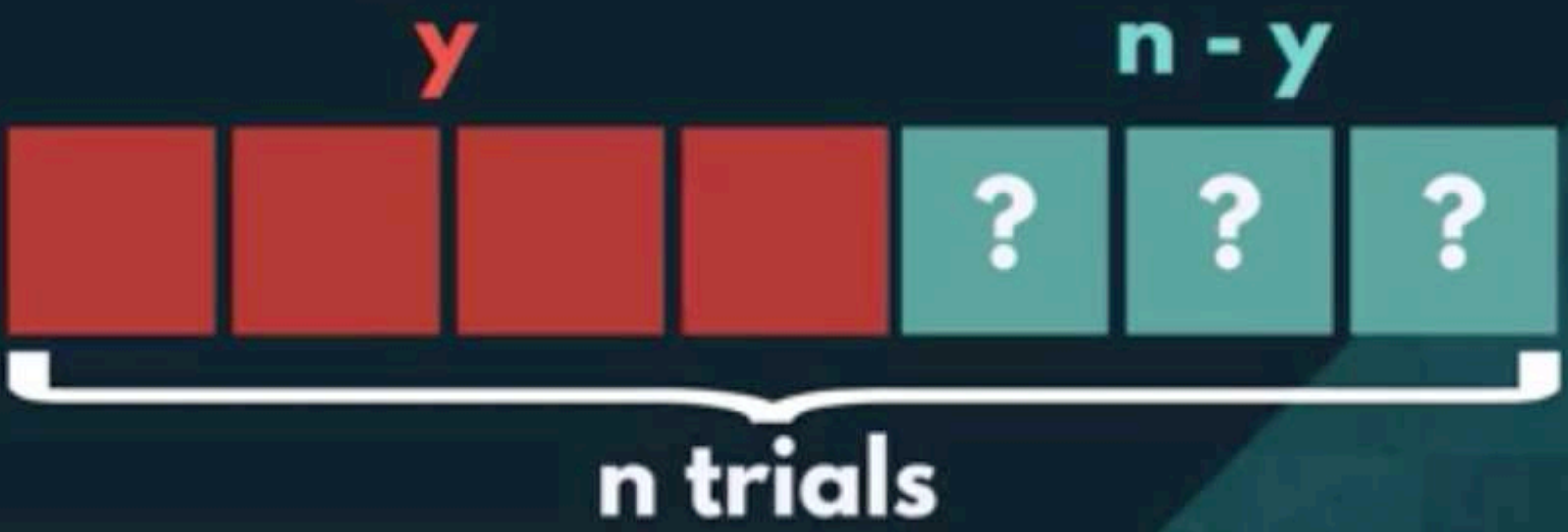
The likelihood of getting a given outcome a precise number of times

p(y)

The Probability Function

P(desired outcome) = p

P(alternative outcome) = 1 - p



Exactly y-many times
over the n trials

The Probability Function

The number of ways in which 4 out of the 6 trials can be successful

=

Picking 4 elements out of a sample space of 6

$$C_4^6$$

The Probability Function

$$p(y) = \binom{n}{y} \cdot p^{y} \cdot (1 - p)^{n - y}$$

GM Example

- A single stock of General Motors

$$p(\uparrow) = 60\% = 0.6$$

$$p(\downarrow) = 40\% = 0.4$$

- 3 Increases in 5 days

$$p(y) = \binom{n}{y} \cdot p^{y} \cdot (1 - p)^{n - y}$$

GM Example

$y \Rightarrow 3$

$n \Rightarrow 5$

$p \Rightarrow 0.6$

$$p(y) = \binom{n}{y} \cdot p^{y} \cdot (1-p)^{n-y} =$$

$$= C_3^5 \times 0.6^3 \times 0.4^2 =$$

$$= 0.3456$$

Expected Values

$$E(X) = x_0 \cdot p(x_0) + x_1 \cdot p(x_1) + \ldots x_n \cdot p(x_n)$$

$$Y \sim B(n, p)$$

$$E(Y) = p \, . \, n$$

Formula for computing the expected values for categorical variables

Variance and Standard Deviation

$$\sigma^2 = E(y^2) - E(y)^2 =$$

$$= n \cdot p \cdot (1 - p) =$$

$$= 1.2$$

$$\sigma \sim 1.1$$

$y \Rightarrow 3$
$n \Rightarrow 5$
$p \Rightarrow 0.6$

GM Example

$$p(y) = \binom{n}{y} \cdot p^{y} \cdot (1-p)^{n-y} =$$

$$= C_3^5 \times 0.6^3 \times 0.4^2 =$$

= 34.56%
of getting exactly 3 increases

Binomial Distribution

Notation

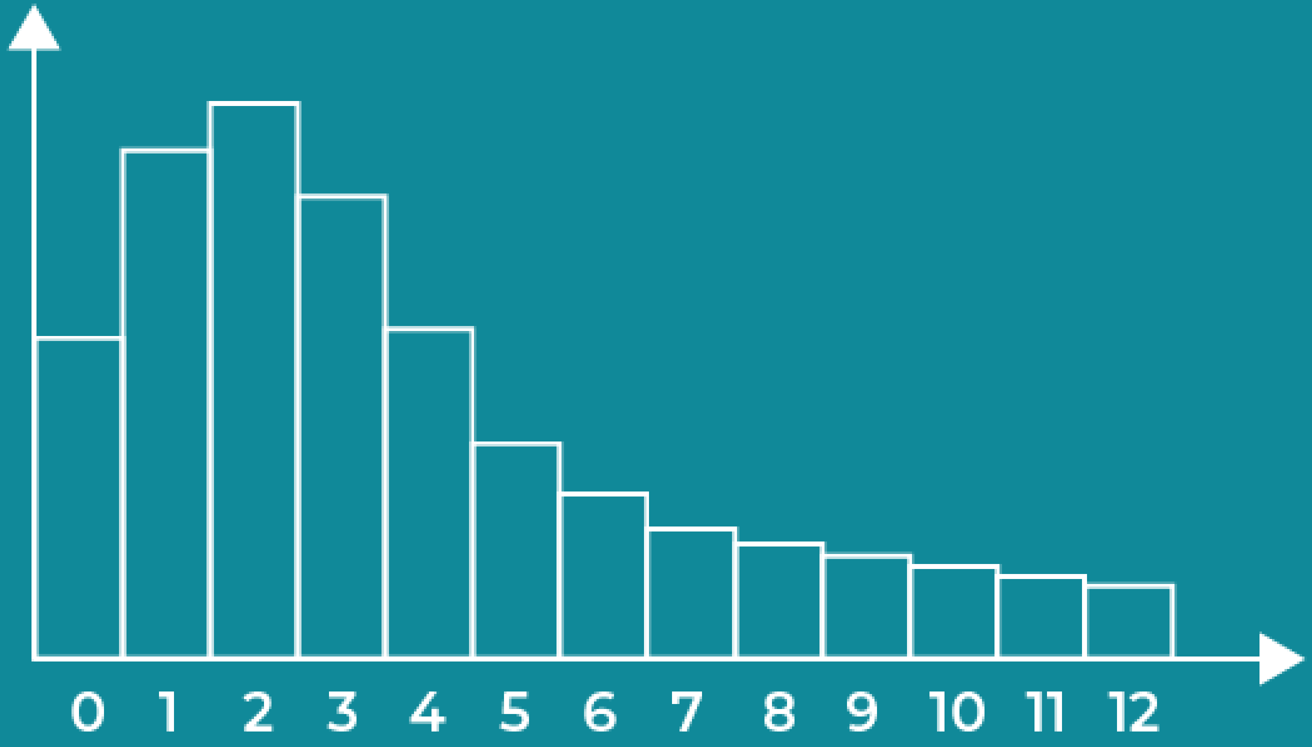
$$B\,(n,\ p)$$

$$X \sim B\,(10, 0.6)$$

$$\Rightarrow \mathrm{Bern}\,(p) = B\,(1,\ p)$$

Poisson Distribution

When we want to know the likelihood of a certain event occurring over a given interval of time or distance we use a **Poisson Distribution**.



Notation:

- $Y \sim Po(\lambda)$

Key characteristics

- Measures the frequency over an interval of time or distance. (Only non-negative values.)
- $P(Y = y) = \frac{\lambda^y}{y! e^{-\lambda}}$
- $E(Y) = \lambda$
- $Var(Y) = \lambda$

Example and uses:

- Used to determine how likely a specific outcome is, knowing how often the event **usually** occurs.
- Often incorporated in marketing analysis to determine whether above average visits are out of the ordinary or not.

Poisson Distribution

$$Po(\lambda)$$

$$Y \sim Po(4)$$

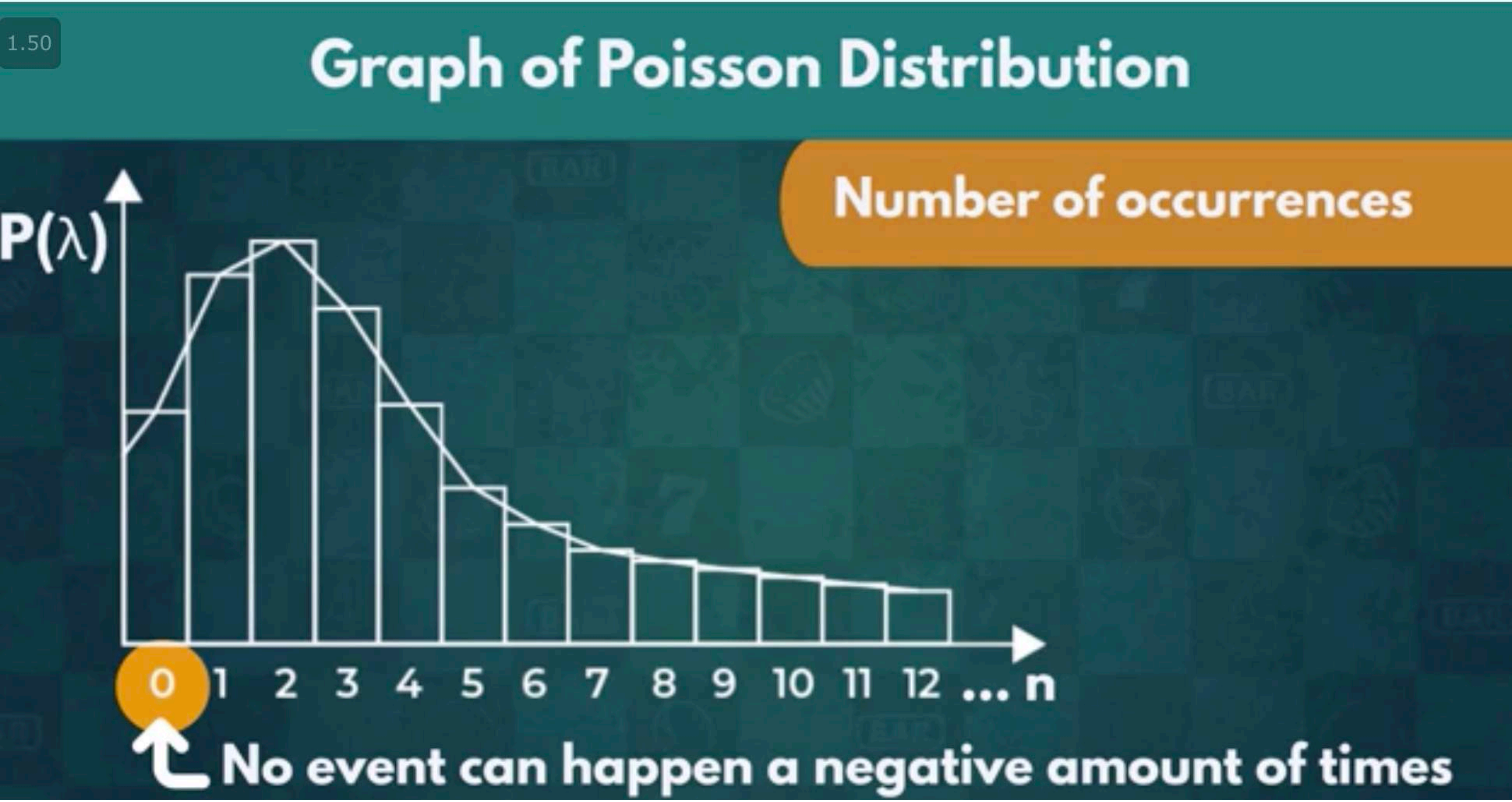
Firefly Example

3 times in 10 sec

$$Y \sim Po(3)$$

8 times in 20 sec

365√DataScience


1.50
Graph of Poisson Distribution
Number of occurrences
P(λ)
0 1 2 3 4 5 6 7 8 9 10 11 12 ... n
No event can happen a negative amount of times

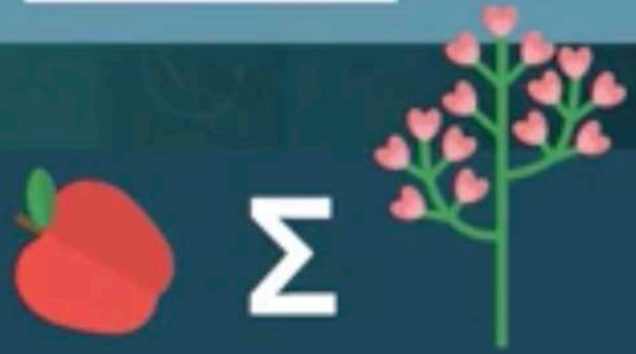
Formula

Poisson Distribution is wildly different

$$P(Y) = \frac{\lambda^{y} e^{-\lambda}}{y!}$$

e → Euler's number (Napier's constant)

$e \approx 2.72$



1.50

Formula

Poisson Distribution is wildly different

$$P(Y) = \frac{\lambda^{y} e^{-\lambda}}{y!}$$

$$x^{-n} = \frac{1}{x^{n}}$$

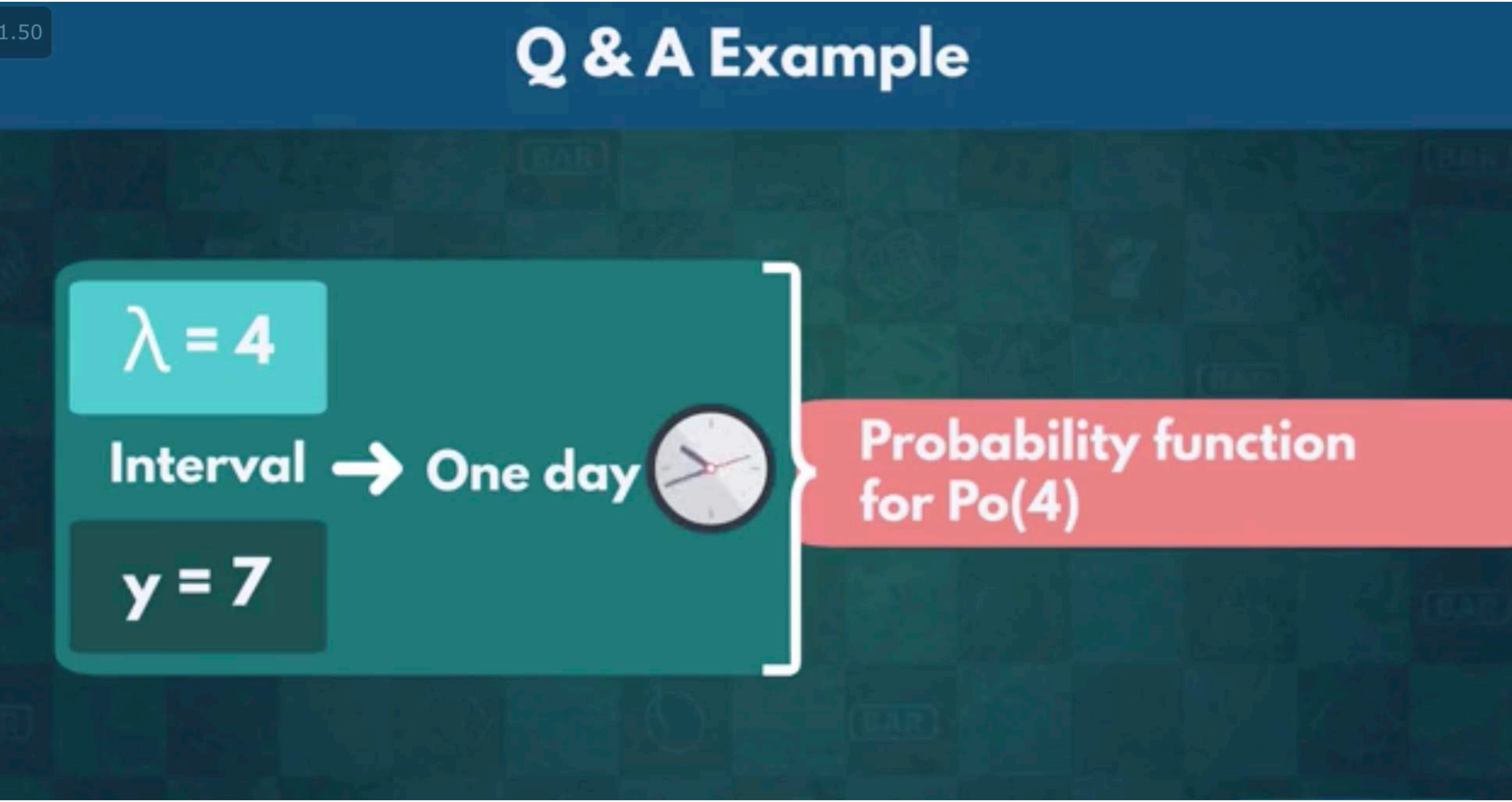
$$e^{-\lambda} = \frac{1}{e^{\lambda}}$$

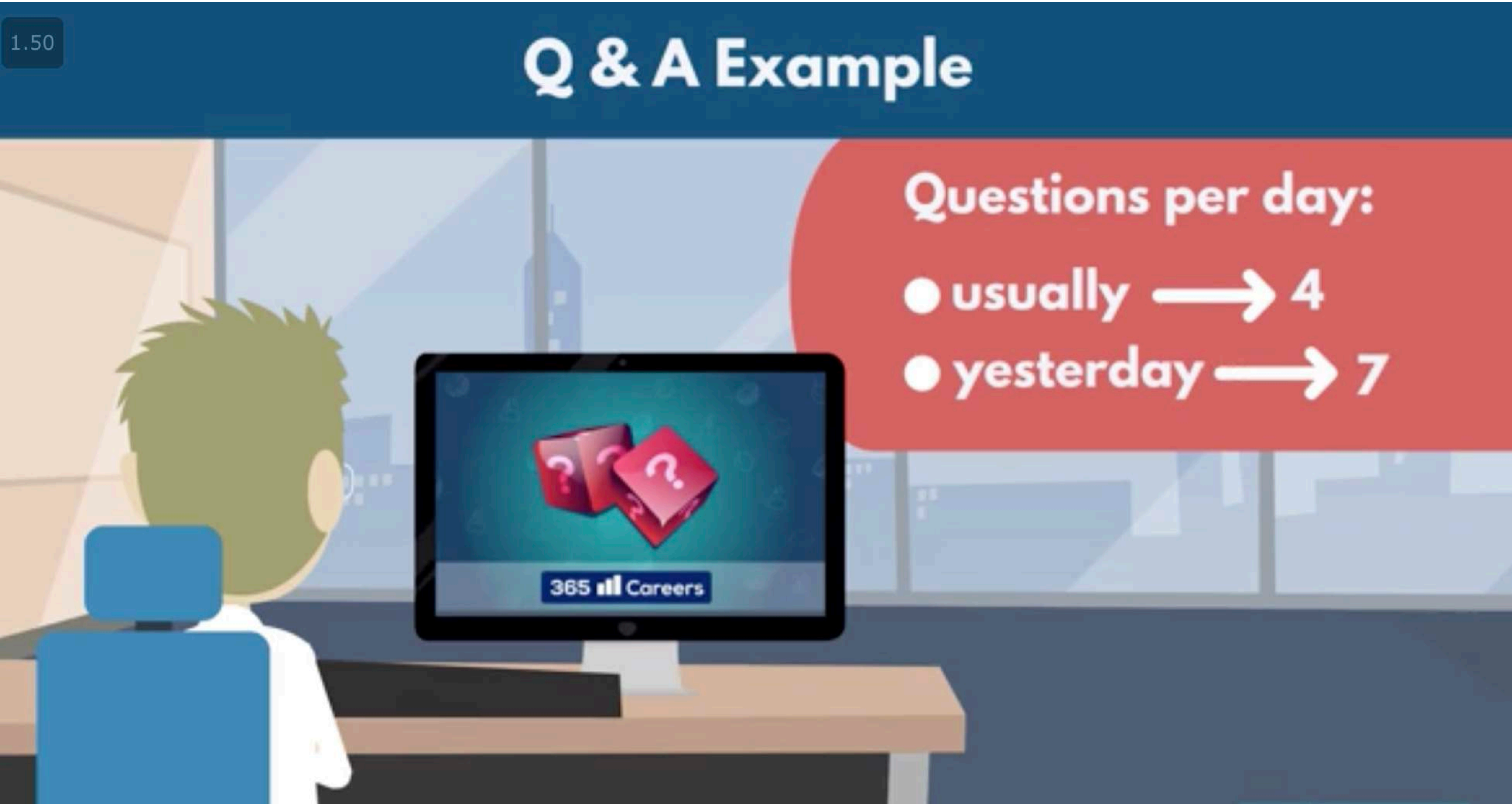
Q & A Example

Questions per day:

- usually $\longrightarrow$ 4
- yesterday $\longrightarrow$ 7

$P(y = 7) = ?$


1.50
Q & A Example
λ = 4
Interval → One day
y = 7
Probability function for Po(4)


1.50
Q & A Example
Questions per day:
usually → 4
yesterday → 7
365 Careers

Expected Value

$p(y) \Rightarrow E(y)$

$$E(y) = y_0 \cdot p(y_0) + y_1 \cdot p(y_1) + \dots =$$

Expected Value

$p(y) \Rightarrow E(y)$

$$E(y) = y_0 \frac{\lambda^{y_0} e^{-\lambda}}{y_0!} + y_1 \frac{\lambda^{y_1} e^{-\lambda}}{y_1!} + \dots =$$

$= \lambda$

*Check the additional materials

Variance

$$\sigma^2 = (y_0 - \mu)^2 + (y_1 - \mu)^2 + \dots =$$

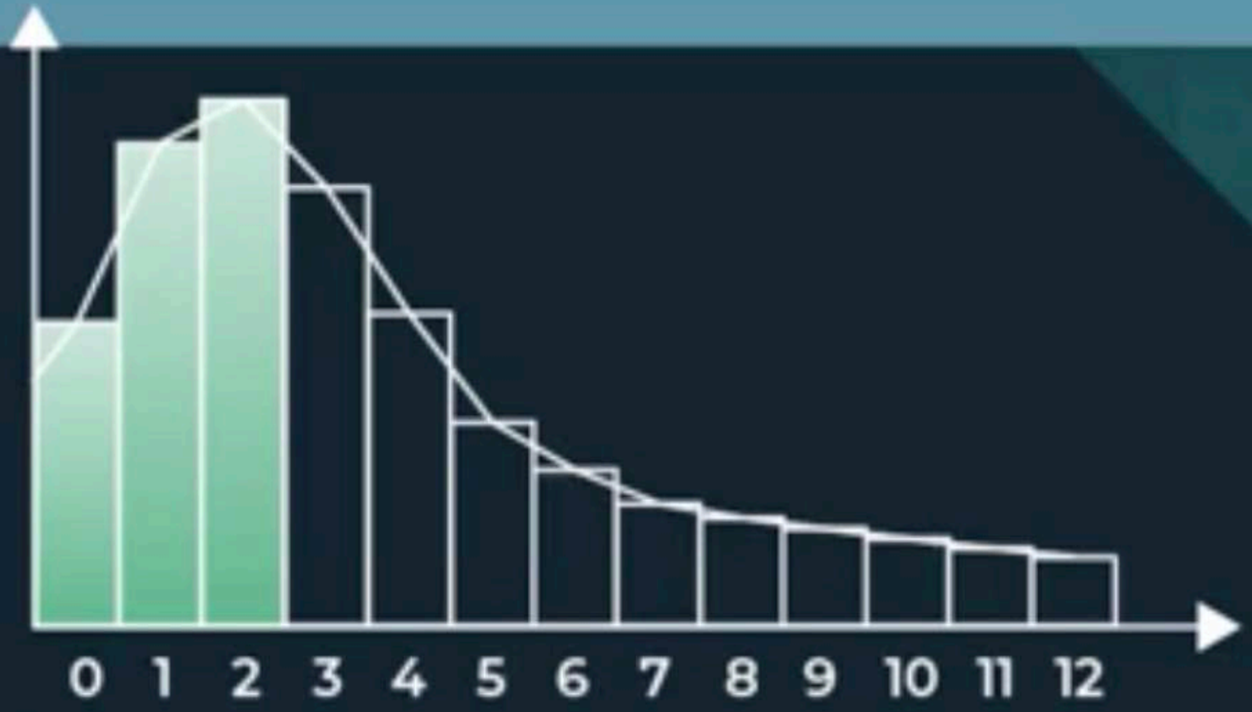
$$= \lambda$$

Mean and Variance

$$\mu = \sigma^2 = \lambda$$

Elegant Statistics

Interval



- Same steps as discrete distributions
- The joint probability of all individual elements

$\lambda \Rightarrow 4$

$y \Rightarrow 7$

Q & A Example

$$P(7) = \frac{4^7 e^{-4}}{7!} \approx \frac{16384 \times 0.0183}{5040} \approx 0.06$$

Only a 6% chance of receiving exactly 7 questions

Continuous Distributions

Infinitely many consecutive outcomes

Continuous Distributions

- Sample space is infinite
- We cannot record the frequency of each distinct value

Q & A Example

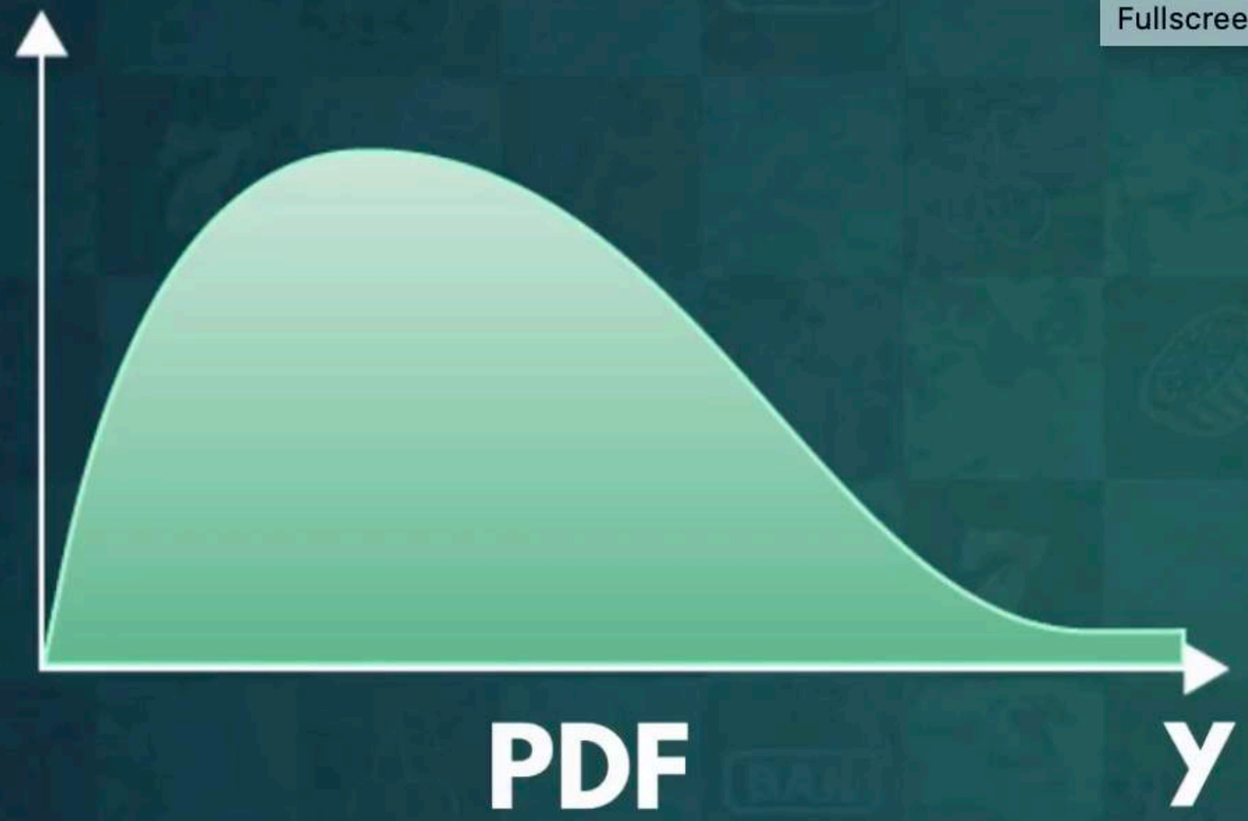
$\lambda = 4$

Interval → One day

$y = 7$

Probability function for Po(4)

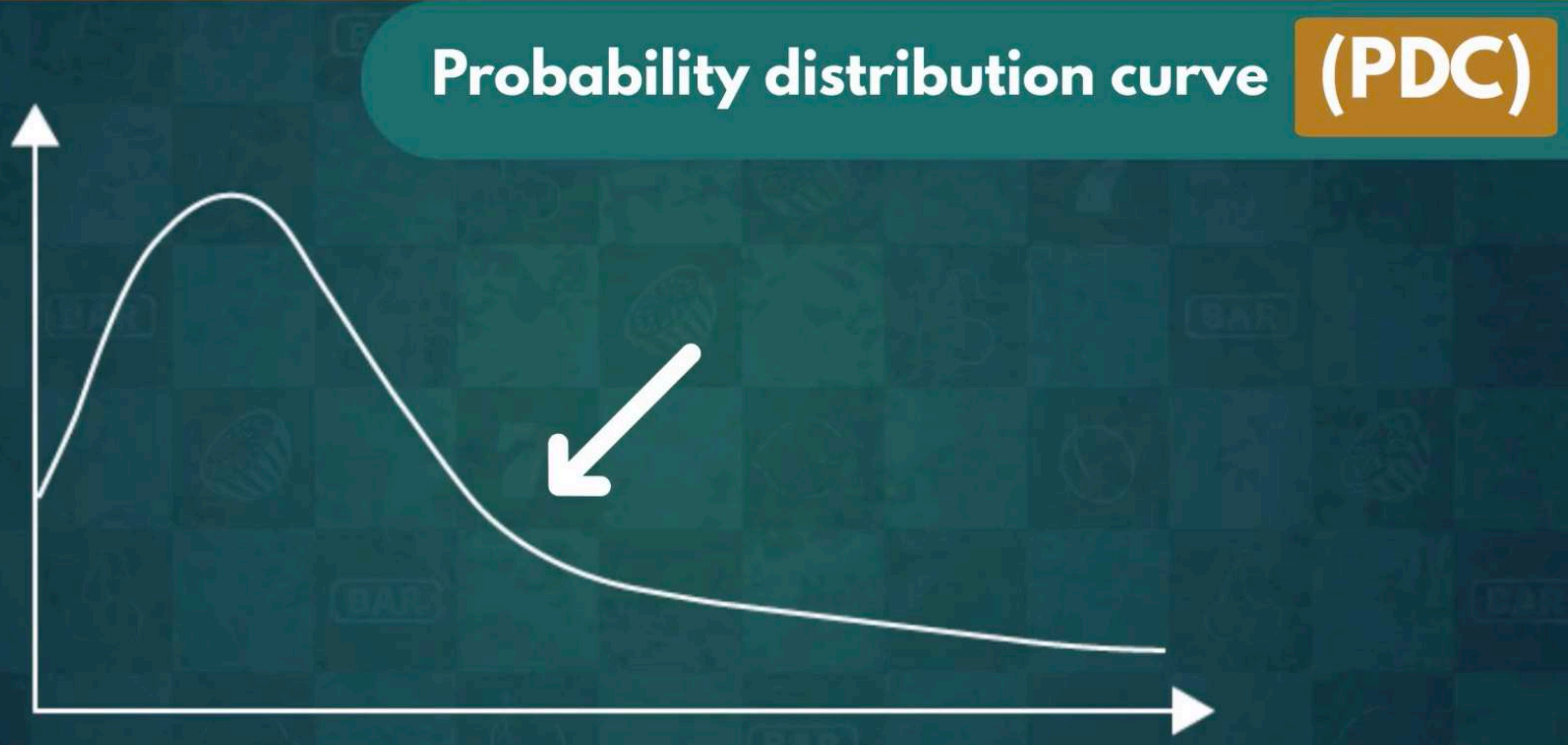
Graph of Continuous Distributions



$$f(y) \geq 0$$

The associated probability for every possible value "y"

Graph of Continuous Distributions



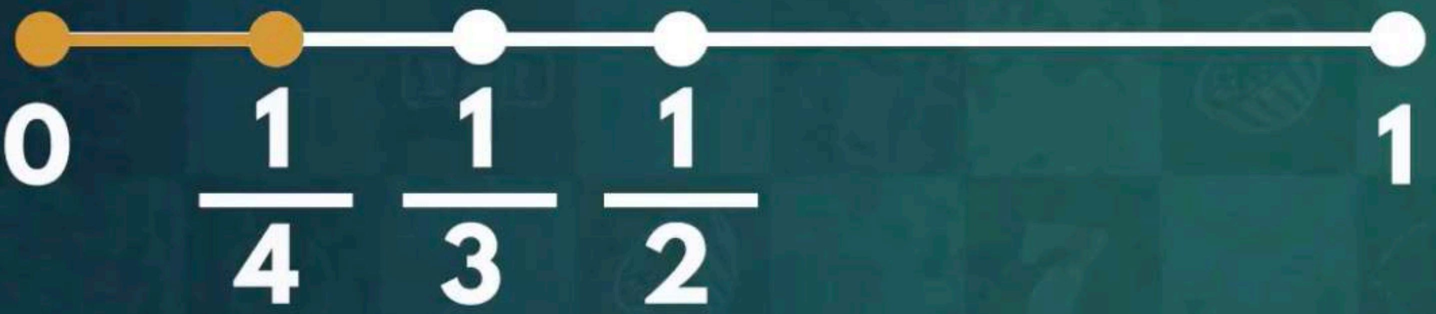
Discrete vs Continuous

$$P(y) = \frac{\text{favourable}}{\text{sample space}} = \frac{1}{\infty}$$

The likelihood of each individual one would be extremely small

Discrete vs Continuous

The greater the denominator becomes, the closer the fraction is to 0



$$P(y) = \frac{1}{\infty} = 0 \leftarrow$$ **Extremely insignificant**

Discrete vs Continuous

The probability for any individual value equal to 0

$$P(X) = 0$$

$$P(x > X) = P(x \geq X)$$

$$P(x < 6 \text{ min}) = P(x \leq 6 \text{ min})$$

$$\boxed{P(x = 6) = 0}$$

1.60

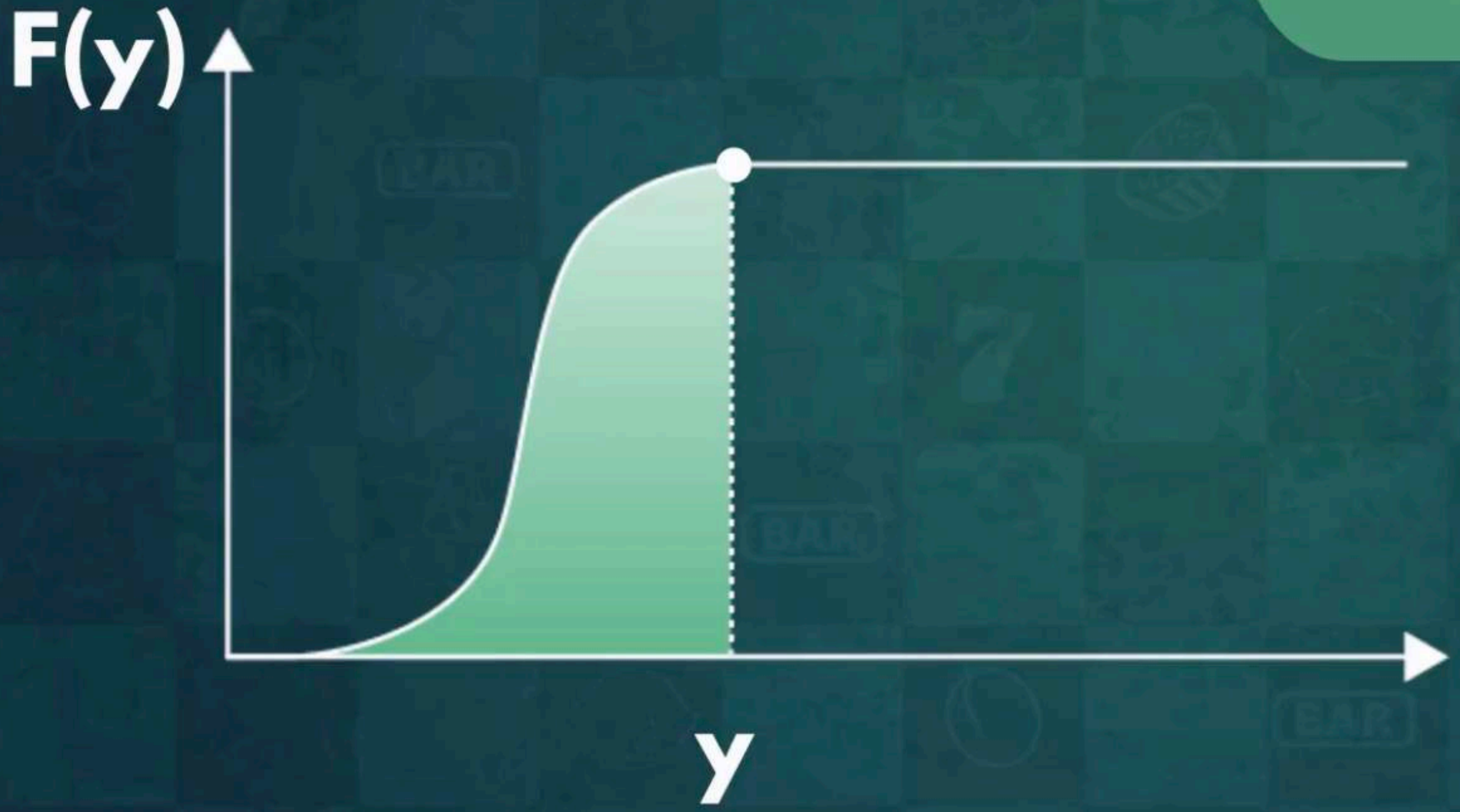
Probability Density Function (PDF)

Probability function

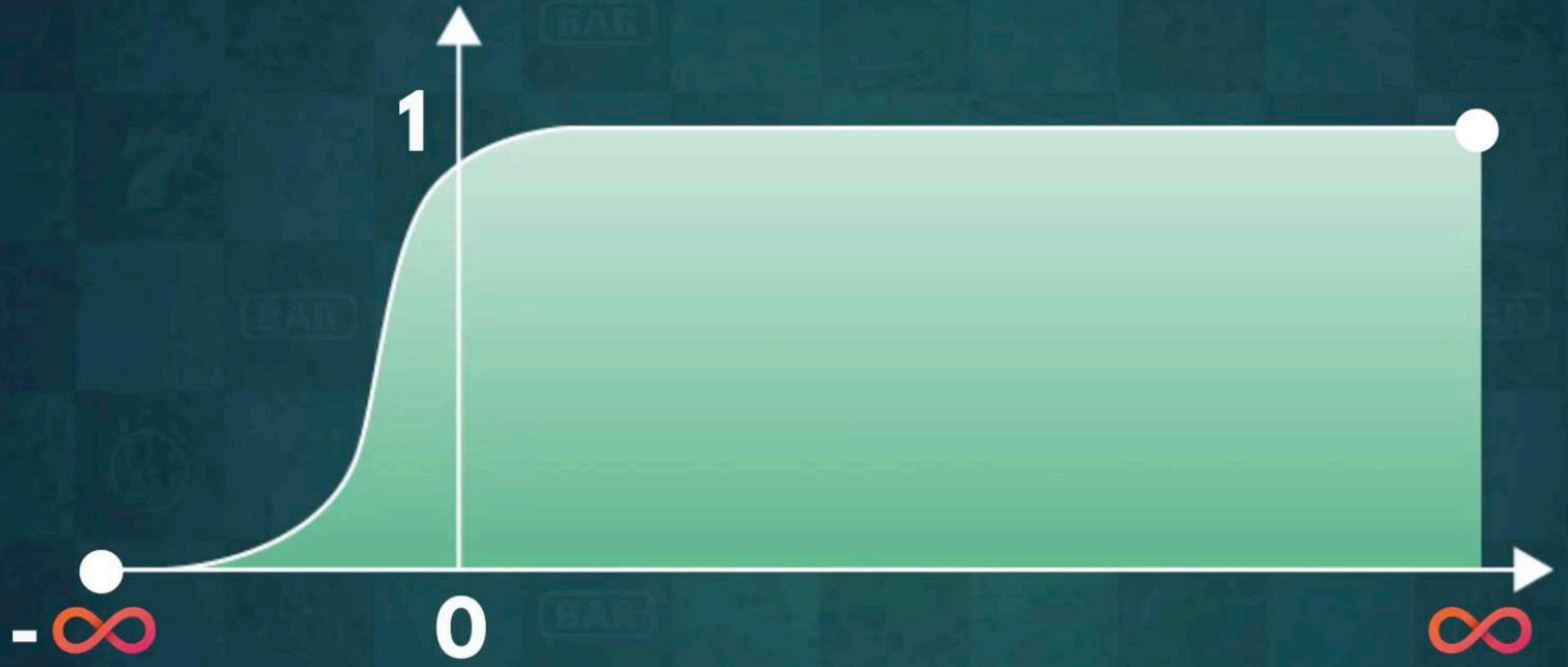
1.60

Cumulative Distribution Function (CDF)

$$F(y) = P(Y \leq y)$$



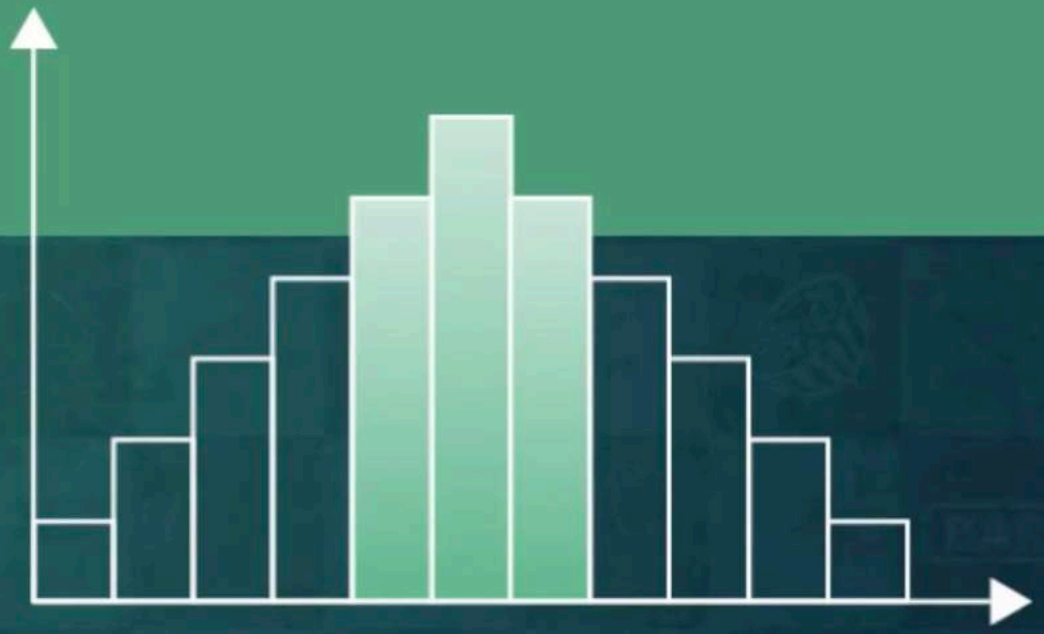
Cumulative Distribution Function (CDF)



$F(-\infty) = 0$ $\qquad$ $F(\infty) = 1$

Discrete CDFs

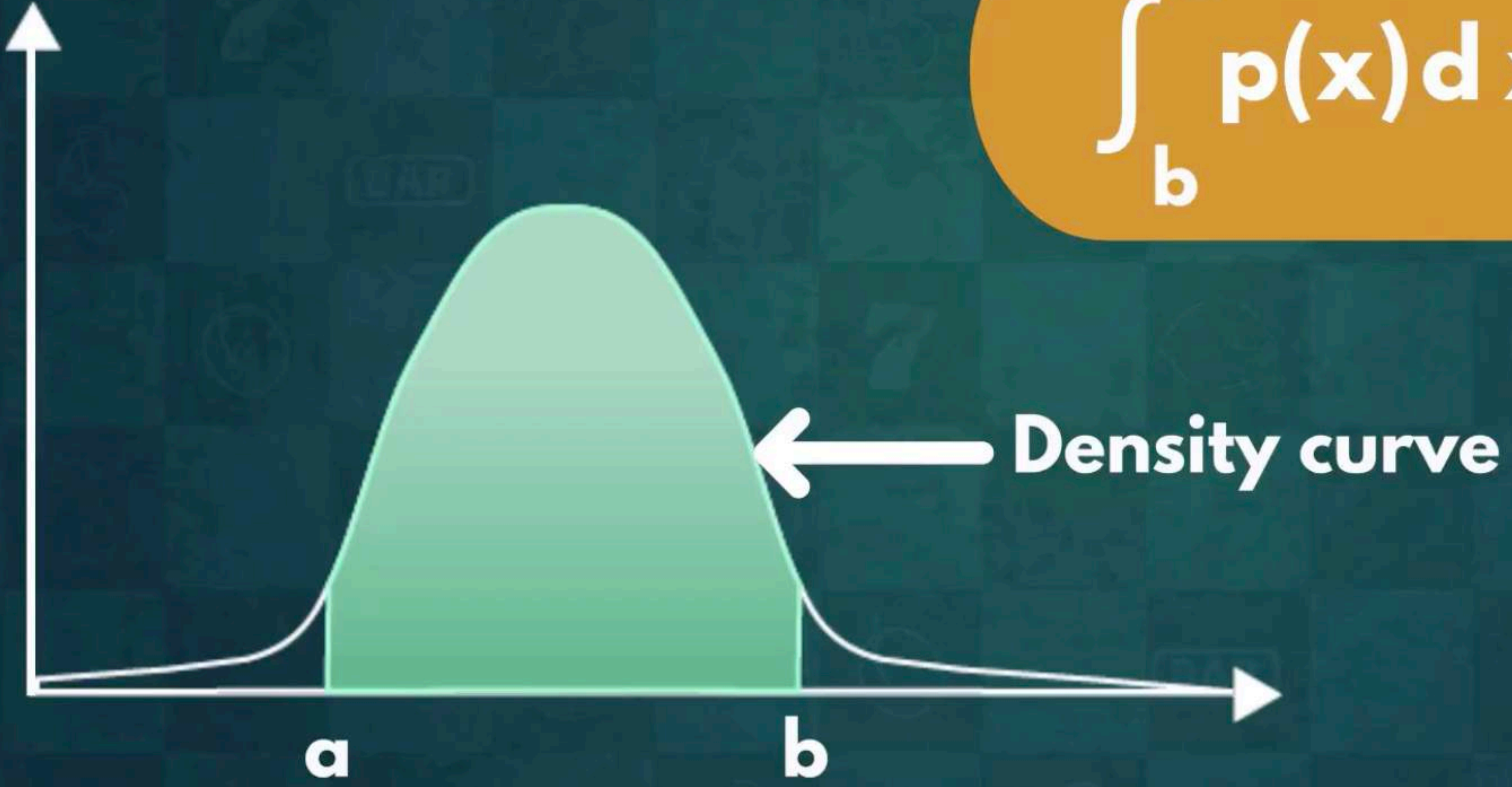
- Far less frequently used
- We can add up the PDF values



Probability of Intervals

$$p(b > x > a)$$

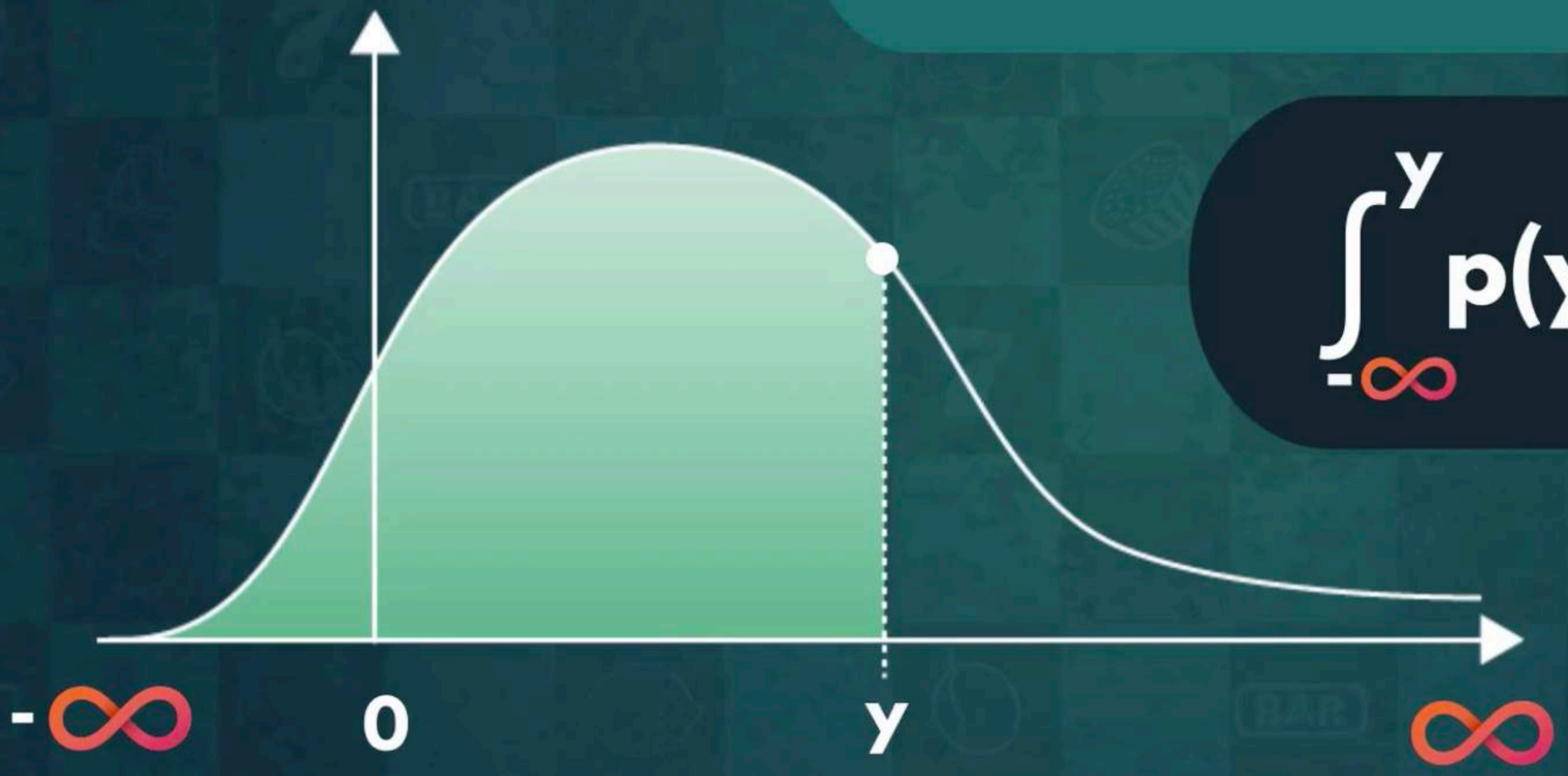
$$\int_{b}^{a} p(x)\,dx$$



CDF vs PDF

Probability of the interval ($-\infty$; y)

$$\int_{-\infty}^{y} p(y)\,dy$$



CDF vs PDF

PDF $\xrightarrow{\text{Integral}}$ CDF

$$\int_{-\infty}^{y} p(y)\,d\,y = F(y)$$

PDF $\xleftarrow{\text{Derivative}}$ CDF

$$p(y) = F(y)\frac{d}{dy}$$

1.60

PDF Graph

Only given PDF

Compute:

$E(y) = ?$

$Var(y) = ?$

1.60

E(y)

$P(y) = 0 \rightarrow$ **We can't apply the summation formula**

$$E(y) = \int_{-\infty}^{\infty} y\,p(y)\,dy$$

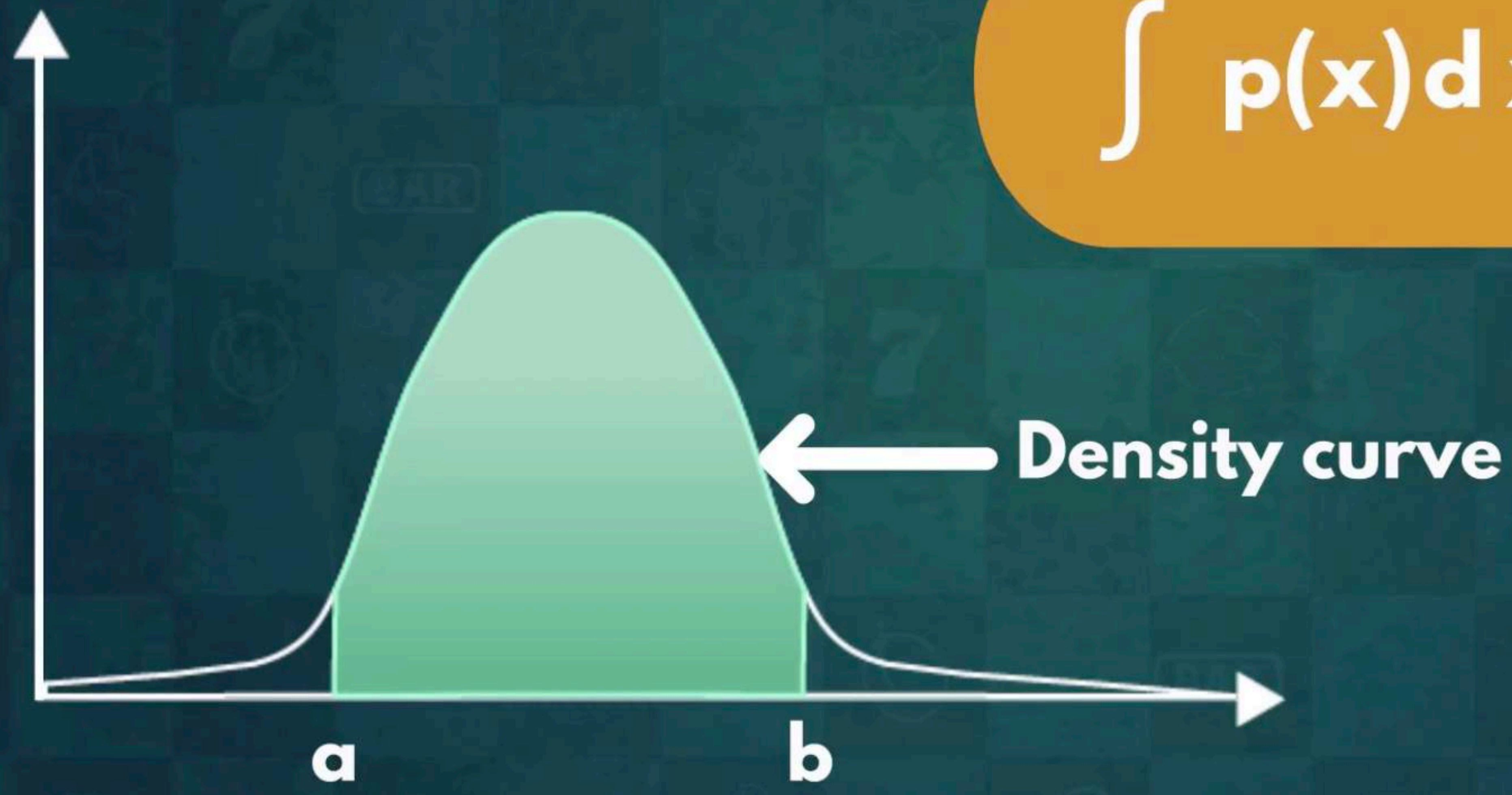
Var(y)

Same variance formula $Var(y) = E(y^2) - E(y)^2$

Probability of Intervals

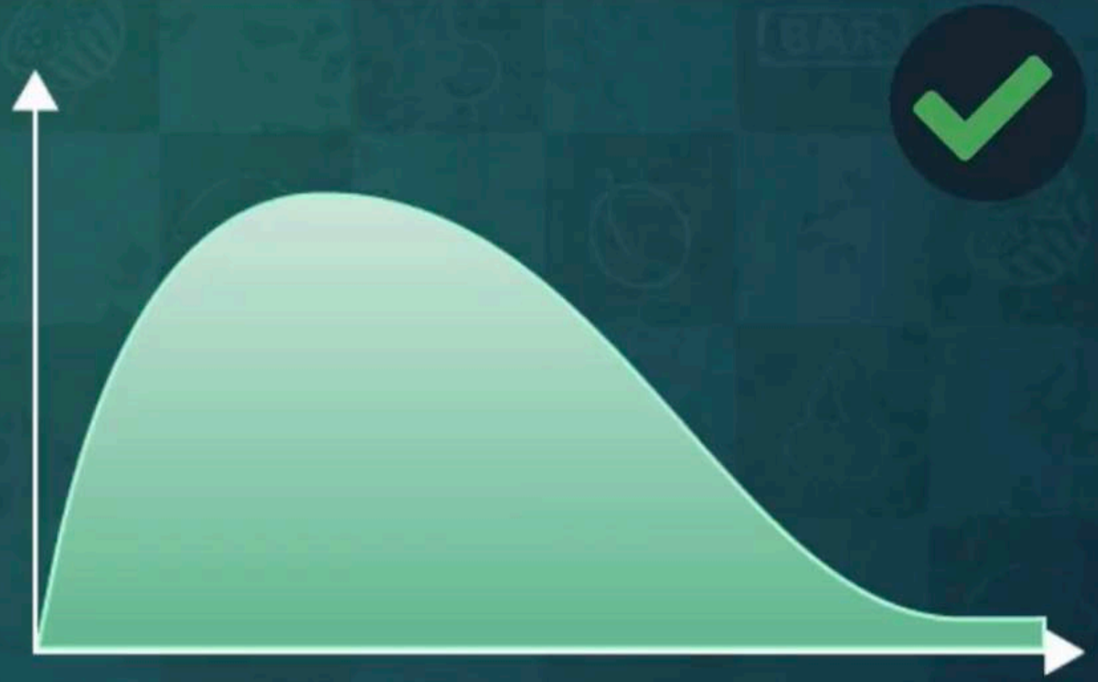
$p(b > x > a)$

$$\int p(x)\,dx$$



Continuous Distributions

- Sample space is infinite
- We cannot record the frequency of each distinct value



1.50

Q & A Example

Questions per day:

- usually $\longrightarrow$ 4
- yesterday $\longrightarrow$ 7

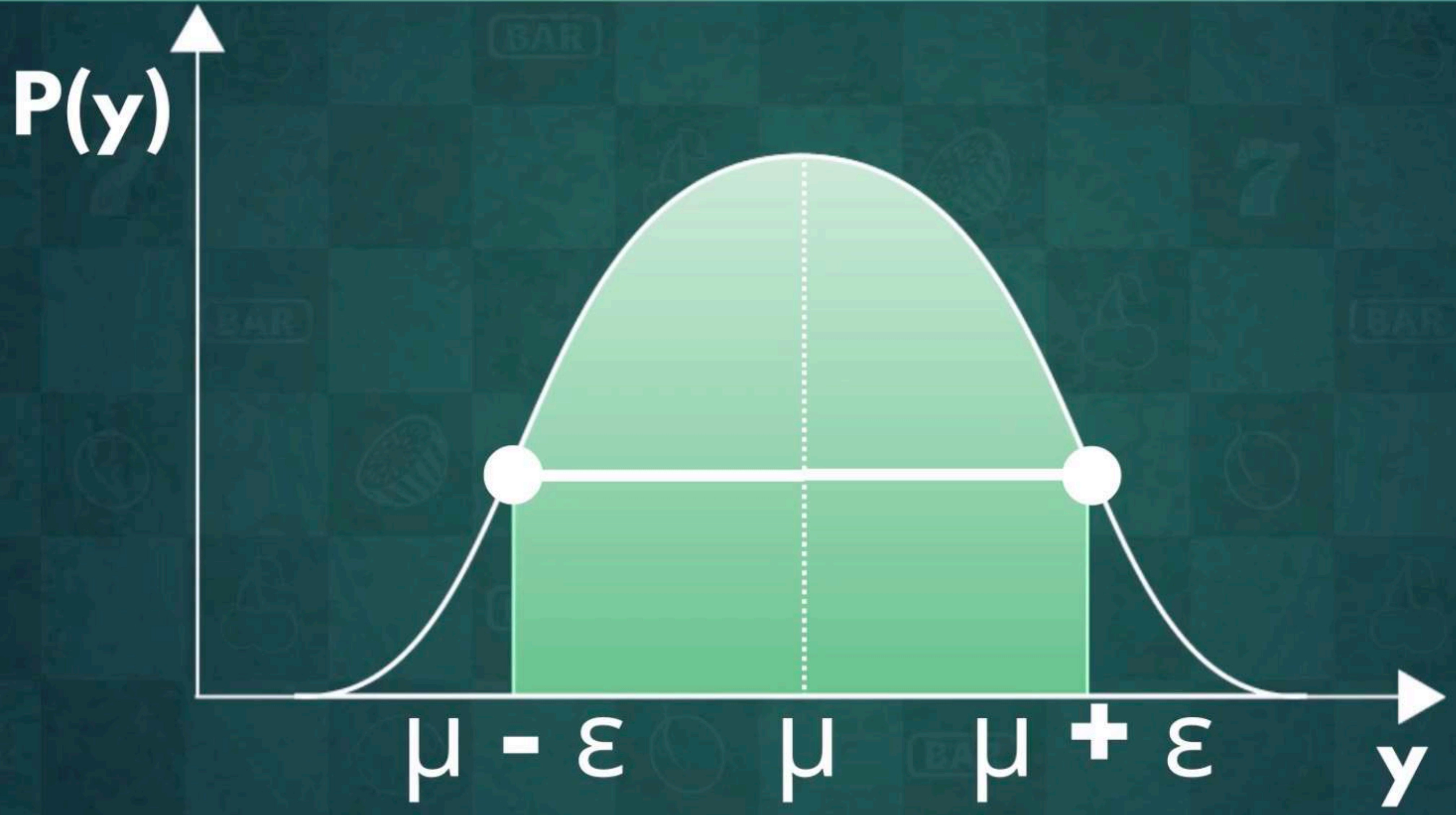
$$P(y = 7) = ?$$

Normal Distribution

$$N(\mu, \sigma^2)$$

$$X \sim N(\mu, \sigma^2)$$

Distinct Characteristics



CDF and PDF

*Check the additonal materials

Expected Value

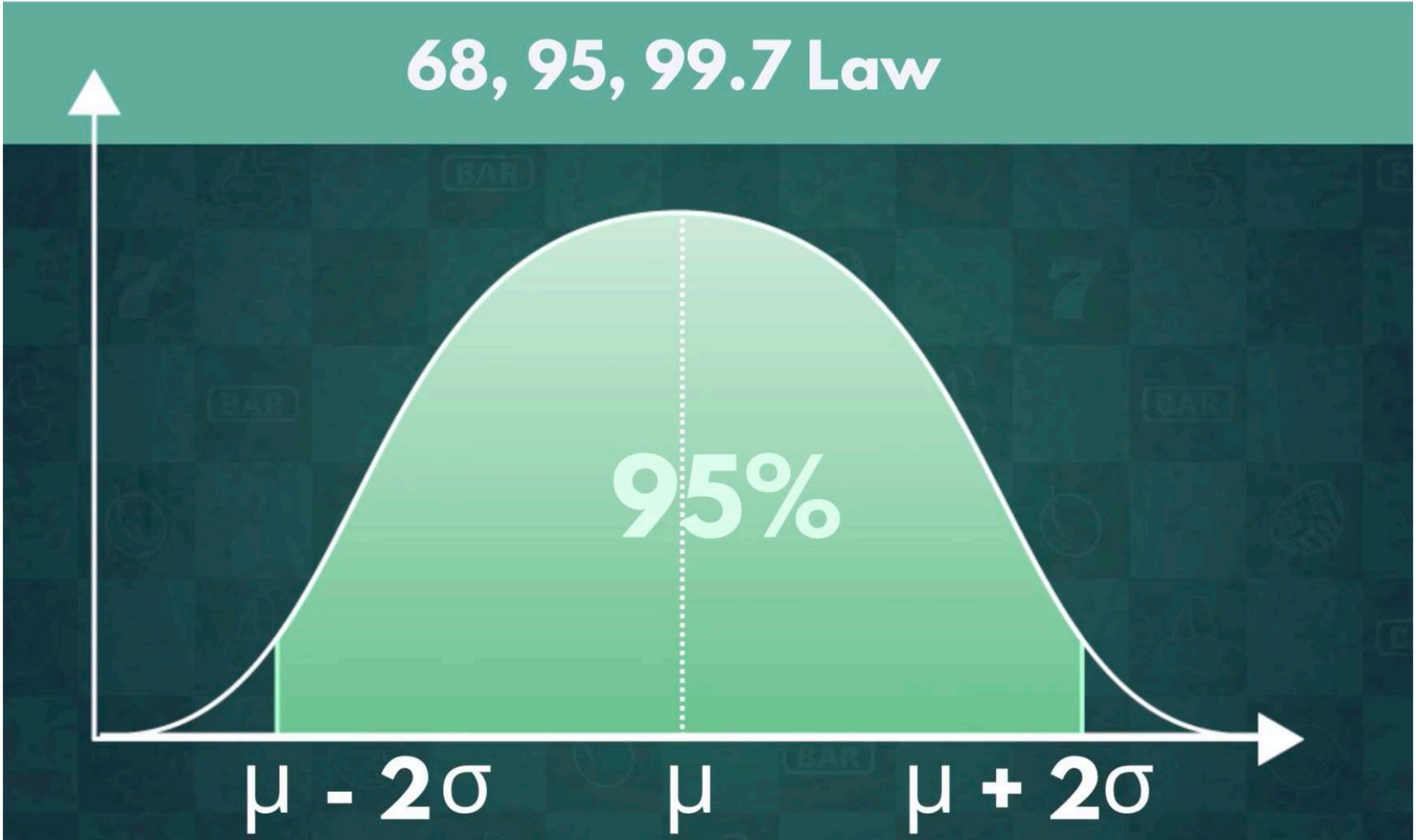
$E(X) = \mu$

$Var(X) = \sigma^2$

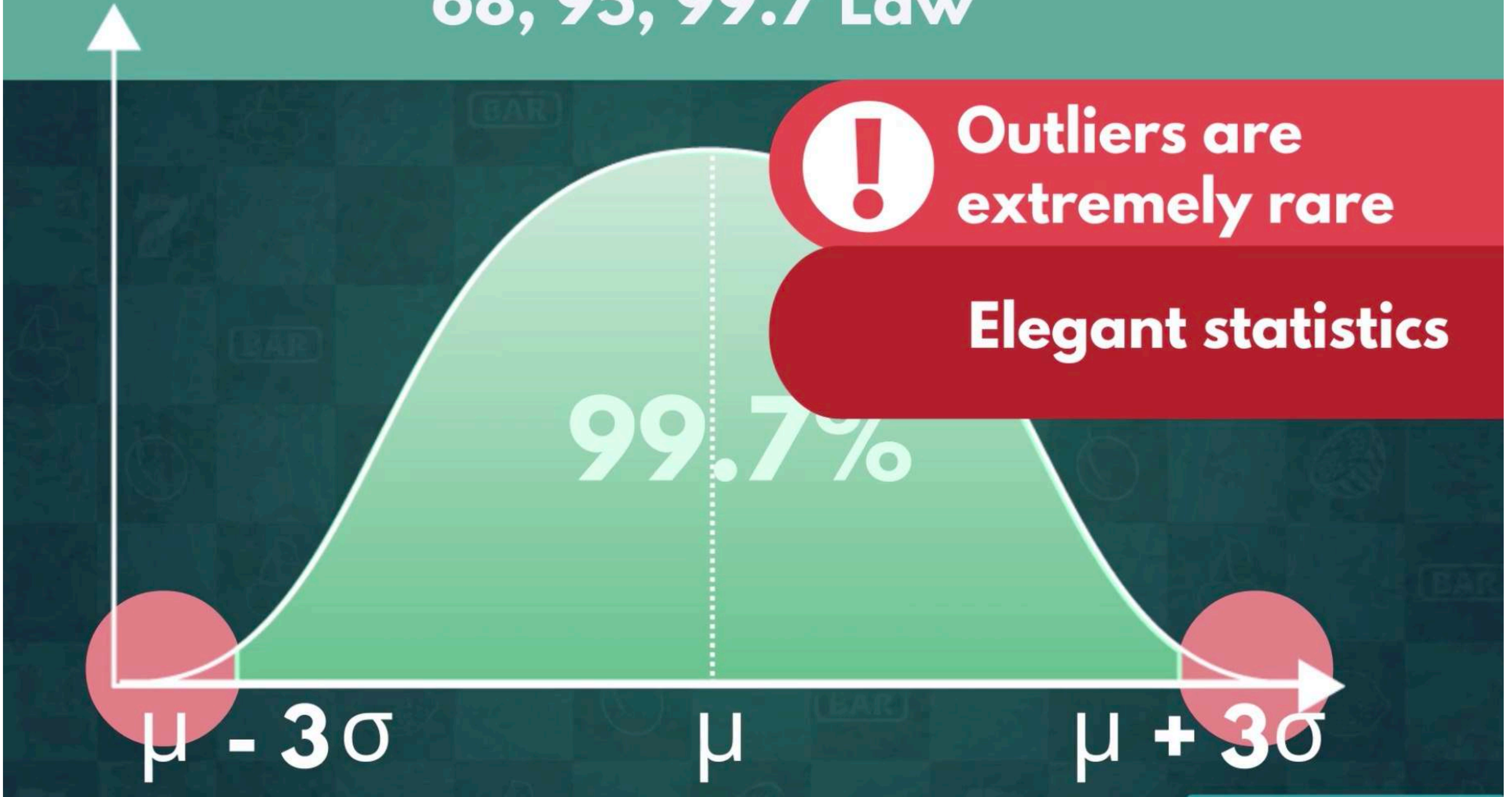
← Usua given

$$Var(X) = E(X^2) - E(X)^2$$

68, 95, 99.7 Law


68, 95, 99.7 Law
95%
μ - 2σ
μ
μ + 2σ

68, 95, 99.7 Law


Outliers are extremely rare
Elegant statistics
99.7%
μ - 3σ
μ
μ + 3σ

$\lambda \Rightarrow 4$

$y \Rightarrow 7$

Q & A Example

$$P(7) = \frac{4^7 e^{-4}}{7!} \approx \frac{16384 \times 0.0183}{5040} \approx 0.06$$

Standardizing

What is "standardizing"?

Why do we use "standardizing"?

Transformation

Transformation

A way in which we can alter every element of a distribution to get a new distribution

Transformation

$+$ $-$ $\times$ $\div$ $X \sim N(\mu, \sigma^2)$

If $X \sim N(\mu_1, \sigma_1^2)$,

then $X + 3 \sim N(\mu_2, \sigma_2^2)$

Transformation

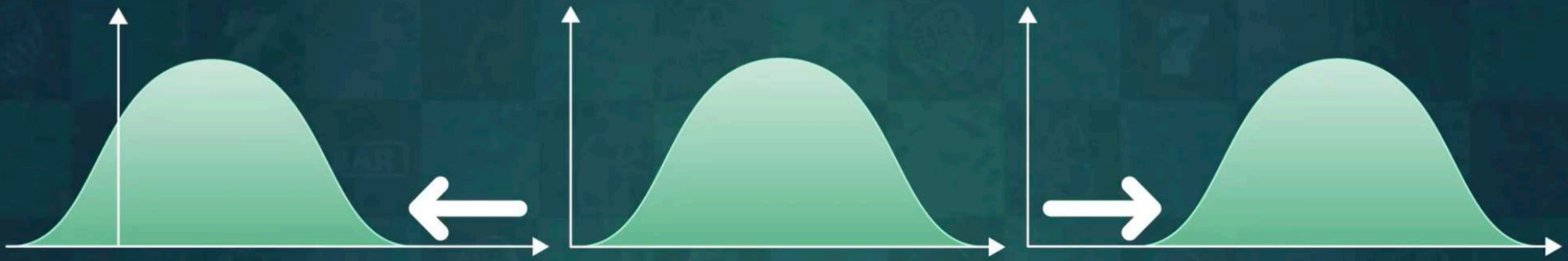
$+$ $-$ $\times$ $\div$ $X \sim N(\mu, \sigma^2)$

If $X \sim N(\mu_1, \sigma_1^2)$,

then $X + 3 \sim N(\mu_2, \sigma_2^2)$

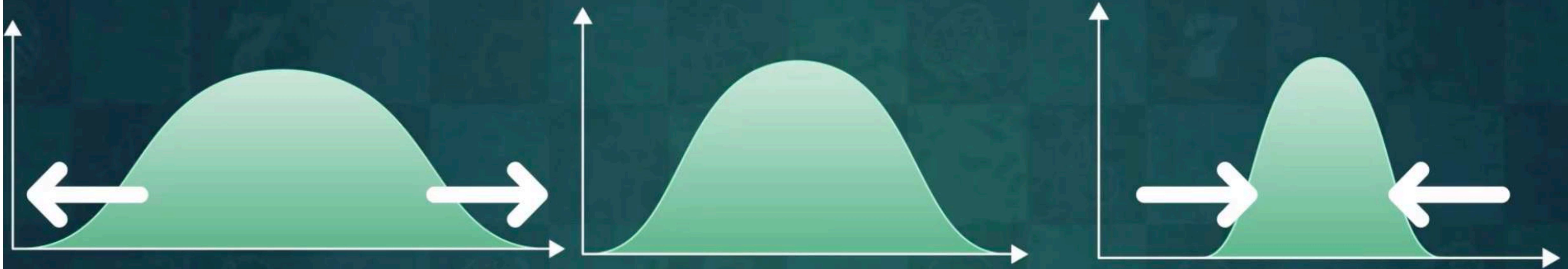
1.50

Addition and Subtraction



$y = f(x - 3)$ $\quad$ $y = f(x)$ $\quad$ $y = f(x + 3)$

Multiplication and Division



$y = f\left(\frac{x}{c}\right)$ $y = f(x)$ $y = f(x \cdot c)$

if $1 > C > 0$

$\frac{x}{1/2} = 2x$

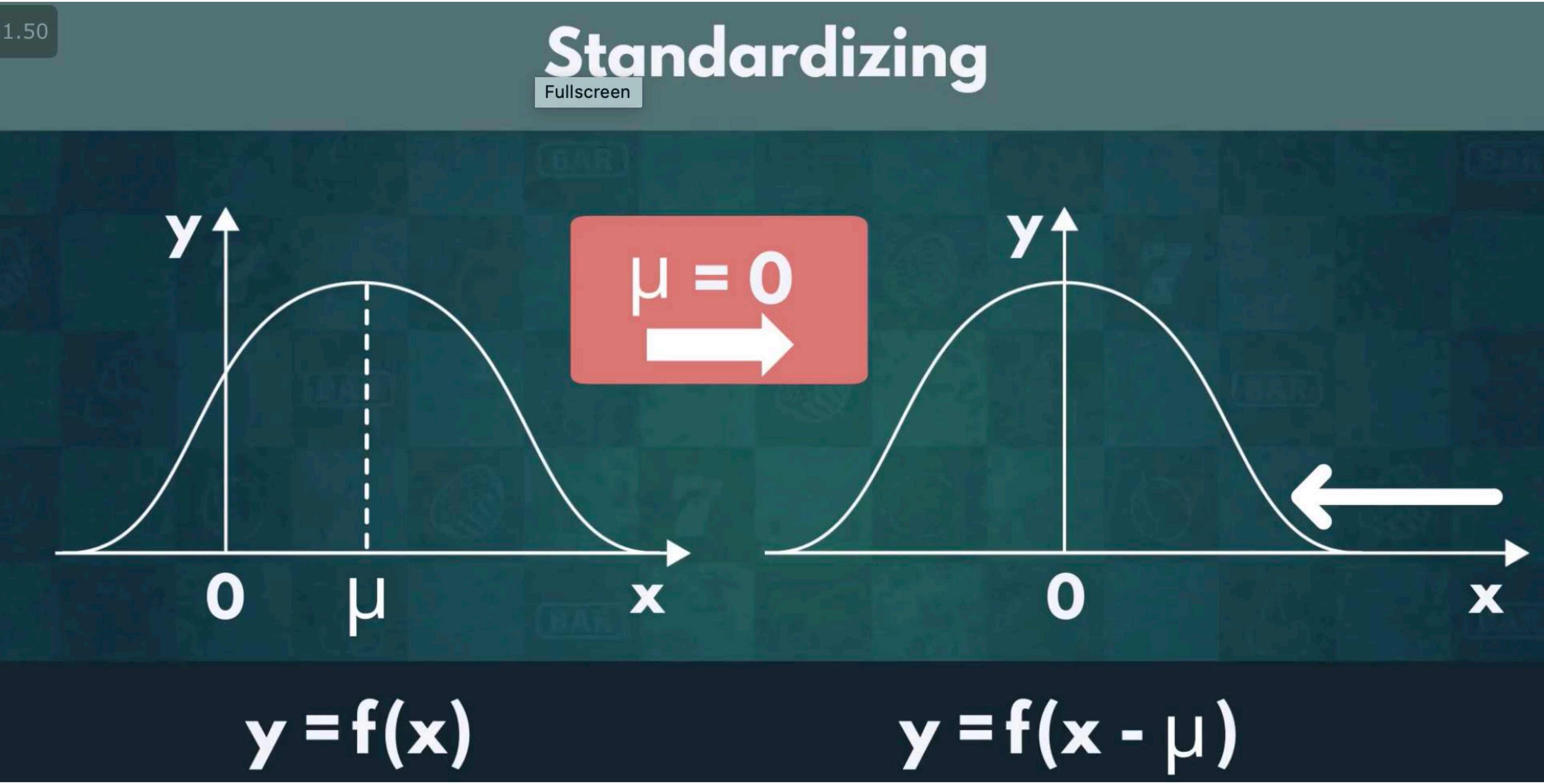
Standardizing

A special kind of transformation

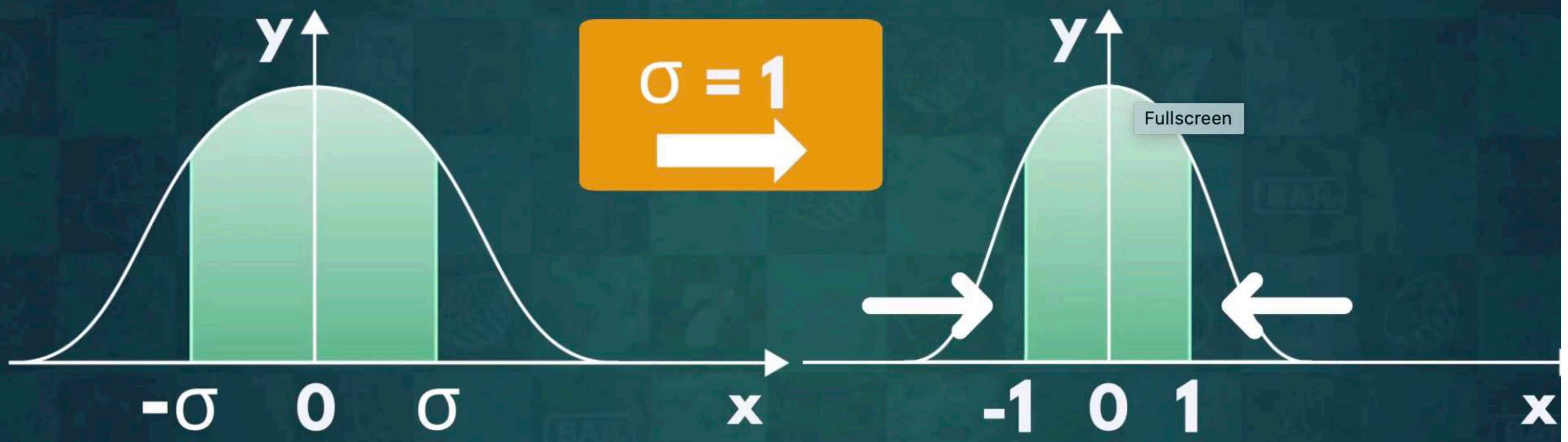
$E(X) = 0$ $Var(X) = 1$

Standard Normal Distribution

68, 95, 99.7% Rule → CDF table (Z- score table)


1.50
Standardizing
Fullscreen
y
0
μ
x
μ = 0
y
0
x
y = f(x)
y = f(x - μ)

Standardizing



$$y = f(x - \mu)$$

$$y = f\left(\frac{x - \mu}{\sigma}\right)$$

Notation

$$\left.\begin{array}{l} Z \sim N(0, 1) \\ Y \sim N(\mu, \sigma^2) \end{array}\right\} => Z = \frac{Y - \mu}{\sigma}$$

Transformation

From Normal To Standard Normal

A single transformation

$$Z = \frac{Y - \mu}{\sigma} \quad \Rightarrow \quad Y \rightarrow Z$$

The number of standard deviations away from the mean

$$y = \mu + 2.3\sigma \quad \Rightarrow \quad z = 2.3$$

Fullscreen

Drawbacks

Useful when we have a Normal Distribution

Not always the case

It requires a lot of data

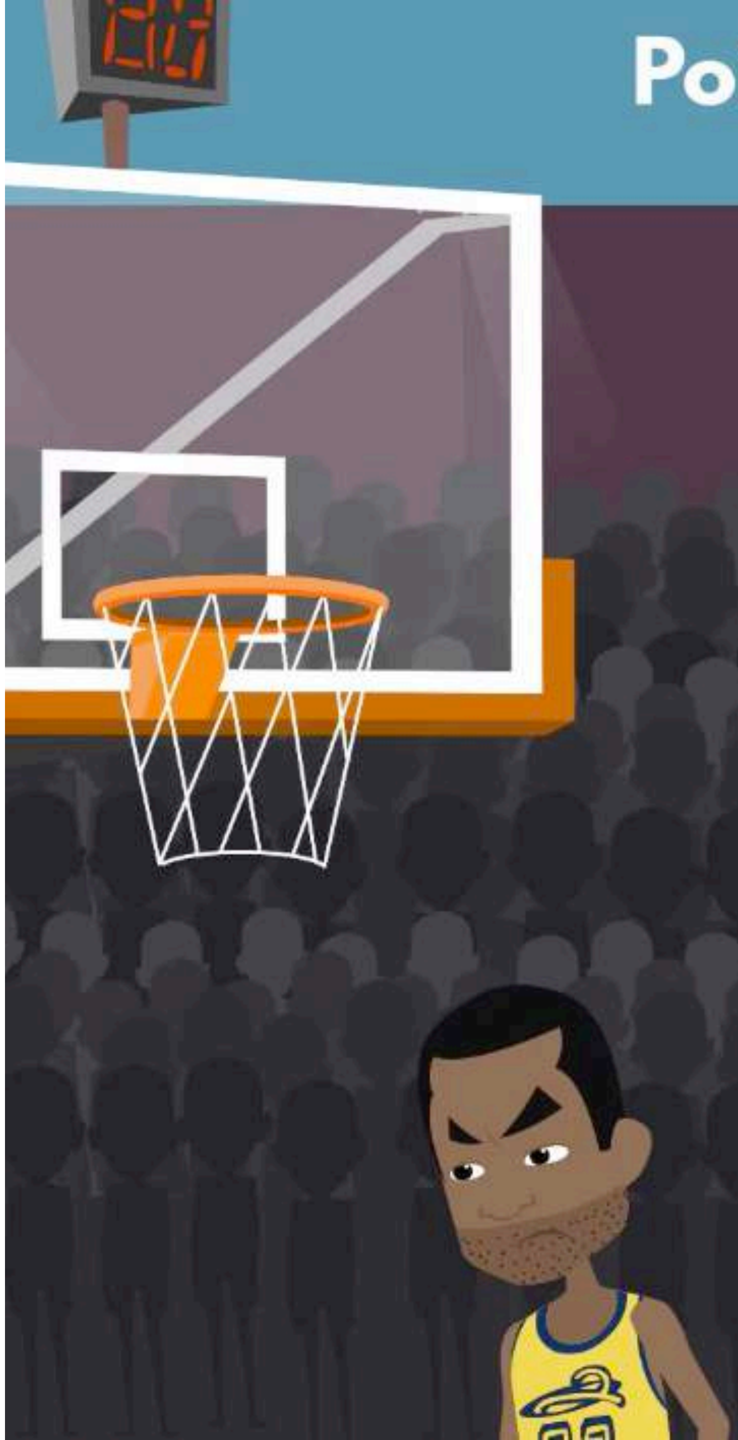
Risk of outliers drastically affecting our analysis

Discrete Distributions

Test out how unusual an event frequency is for a given interval

Poisson Distribution

Poisson Distribution



35 points per game

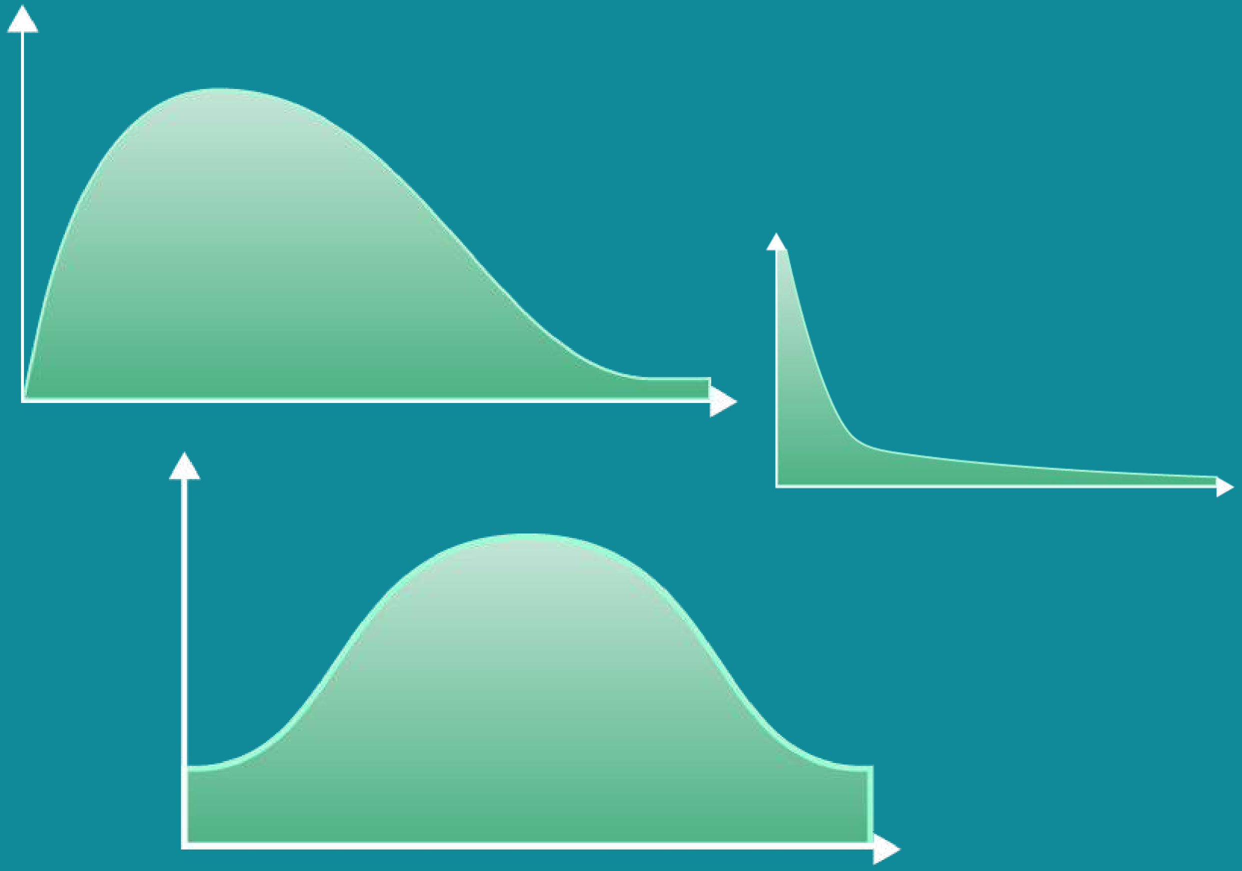
12 points in the first quarter of his next game

P(12)

365√DataScience

Continuous Distributions

If the possible values a random variable can take are a sequence of infinitely many consecutive values, we are dealing with a **continuous distribution**.

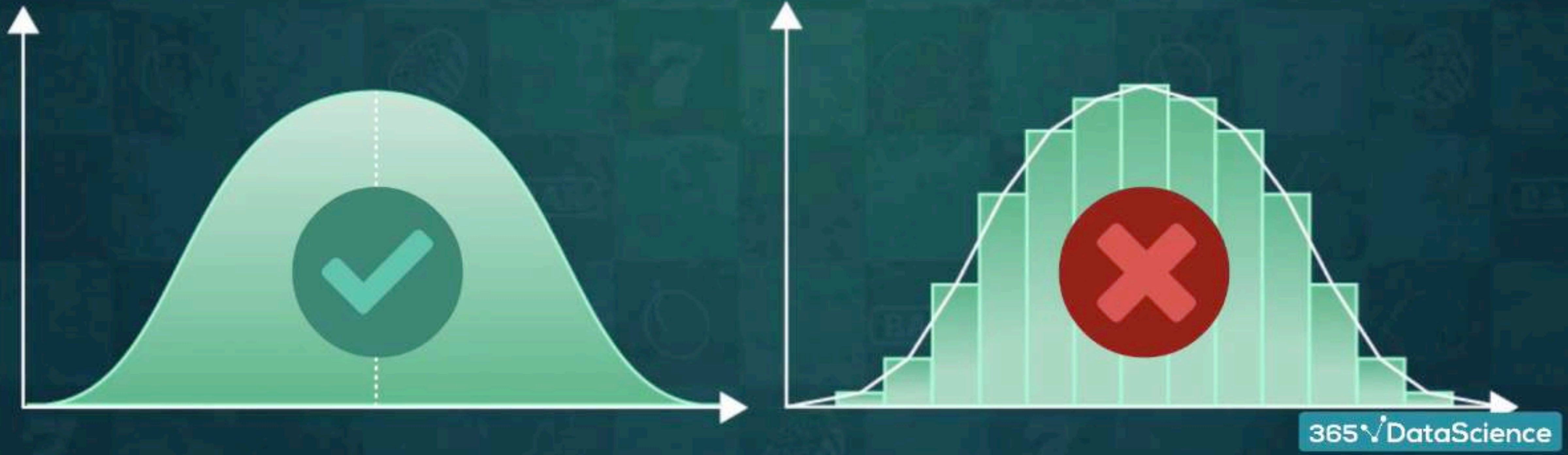


Key characteristics

- Have infinitely many consecutive possible values.
- Cannot add up the individual values that make up an interval because there are **infinitely many** of them.
- Can be expressed with a graph or a continuous function. Cannot use a table, be
- Graph consists of a smooth curve.
- To calculate the likelihood of an interval, we need integrals.
- They have important CDFs.
- $\boldsymbol{P(Y = y)} = 0$ for any individual value y.
- $\boldsymbol{P(Y < y) = P(Y \leq y)}$

Continuous Distributions

The probability distribution would be a curve



Student's T Distribution

Fullscreen

$$t(k)$$

$$Y \sim t(3)$$

Student's T Distribution

Small sample size approximation of a Normal Distribution

Certain characteristics + Sufficient data = **Normal distribution**

Certain characteristics + ~~Suffic~~ data = **Student's T distribution**

1.50

Normal vs Student's T Distribution

Average lap times for the entire season = Normal distribution

Lap times for the first lap of the  GP = Student's T distribution

1.50

Normal vs Student's T Distribution

Mean μ

Variance σ^2

k Degrees of freedom

Normal vs Student's T Distribution

If $k \geq 2$

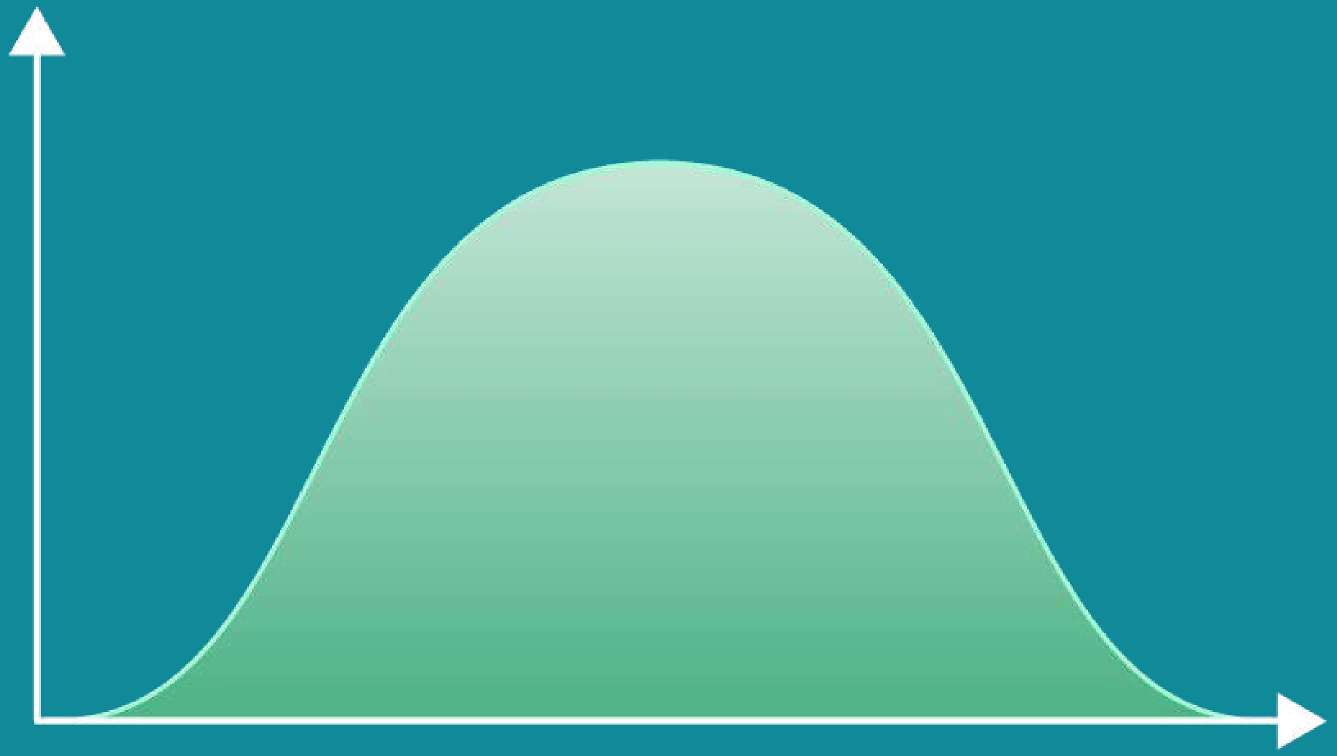
- $E(Y) = \mu$
- $Var(Y) = \dfrac{S^2 . k}{k - 2}$

Application

- Frequently used when conducting statistical analysis
- Hypothesis testing with limited data
- CDF table (T- table)

Normal Distribution

A **Normal Distribution** represents a distribution that most natural events follow.



Notation:

- $\boldsymbol{Y} \sim \boldsymbol{N}(\mu, \sigma^2)$

Key characteristics

- Its graph is bell-shaped curve, symmetric and has thin tails.
- $\boldsymbol{E}(\boldsymbol{Y}) = \mu$
- $\boldsymbol{Var}(\boldsymbol{Y}) = \sigma^2$
- 68% of all its values should fall in the interval:
  - $(\mu - \boldsymbol{\sigma},\ \boldsymbol{\mu} + \boldsymbol{\sigma})$

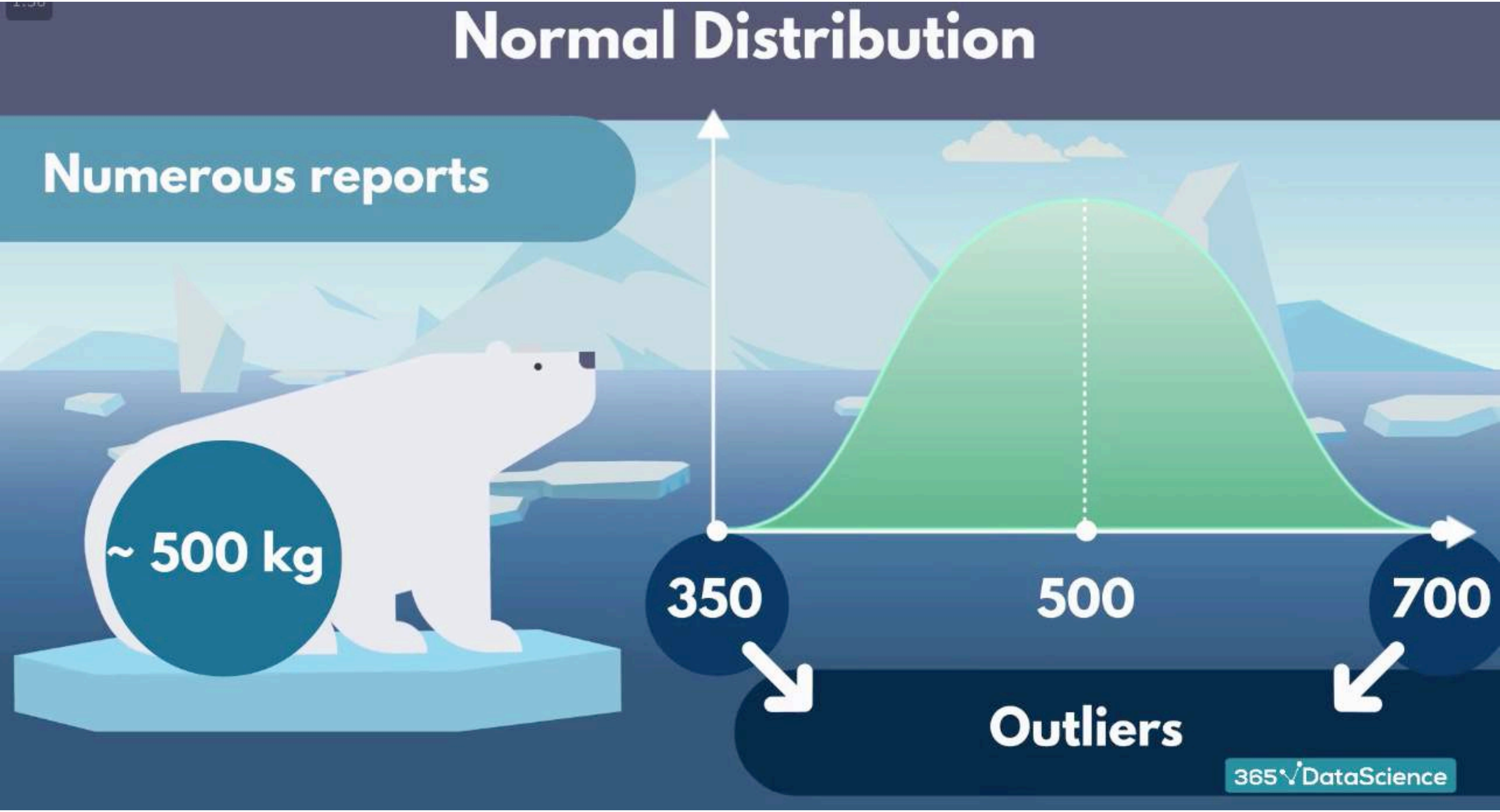
Example and uses:

- Often observed in the size of animals in the wilderness.
- Could be standardized to use the Z-table.

Continuous Distributions

Normal distributions

Often observed in nature


Normal Distribution
Numerous reports
~ 500 kg
350
500
700
Outliers
365√DataScience

Chi-Squared Distribution

$$\chi^2(k)$$

$$Y \sim \chi^2(3)$$

Applications

Few events in real life

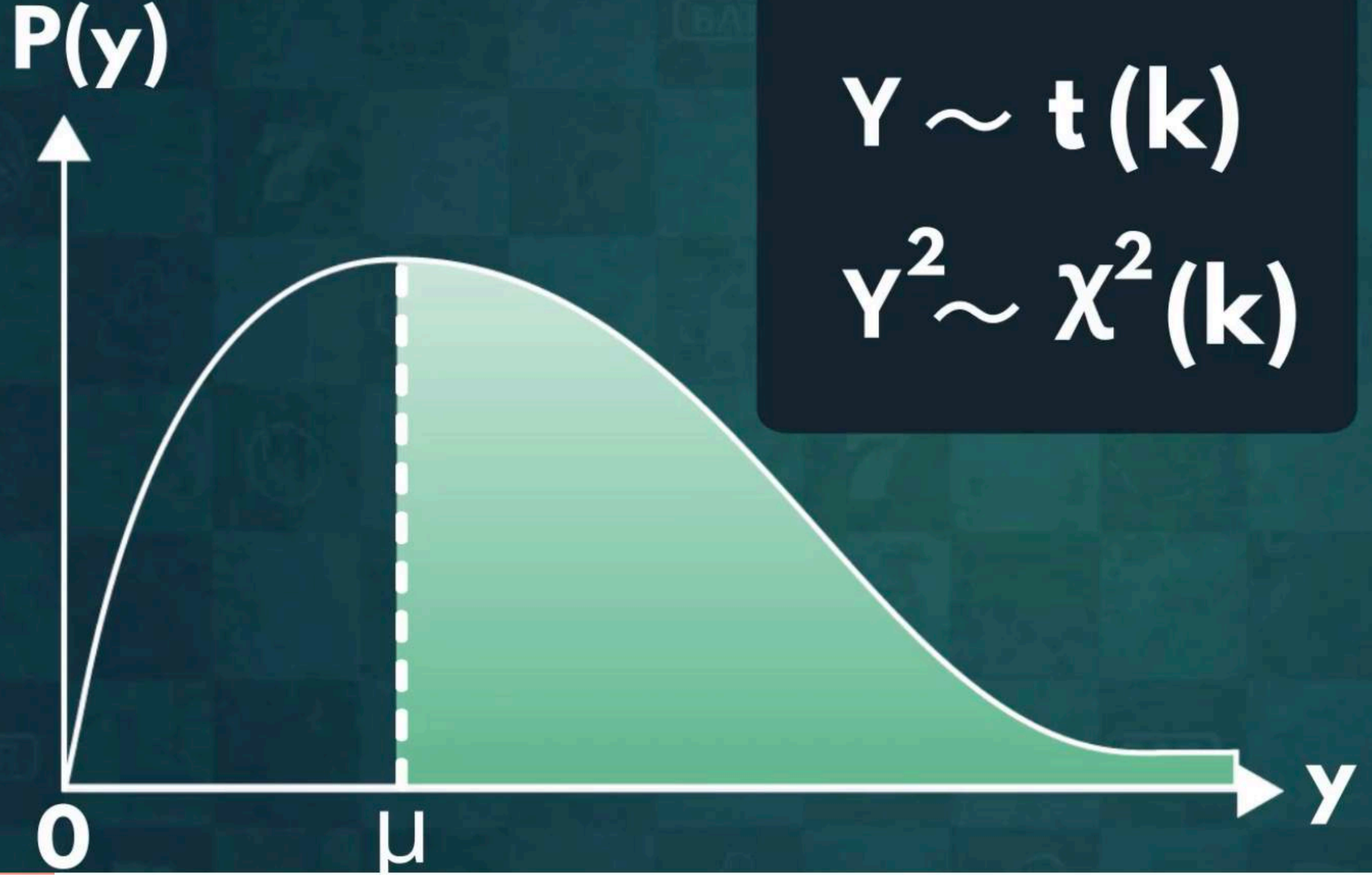
Statistical analysis

- Hypothesis testing
- Computing confidence intervals

Goodness of fit

1.50

Chi-Squared Distribution



$$Y \sim t(k)$$

$$Y^2 \sim \chi^2(k)$$

$$X \sim \chi^2(k)$$

$$\sqrt{X} \sim t(k)$$

Asymmetric

365 DataScience

Convenience

A table of known values: N, T

$E(X) = k$

$Var(X) = 2k$

Standardizing a Normal Distribution

To standardize any normal distribution we need to transform it so that the mean is 0 and the variance and standard deviation are 1.

Using a transformation to create a new random variable z.

$$z = \frac{y - \mu}{\sigma}$$

Ensures mean is 0.

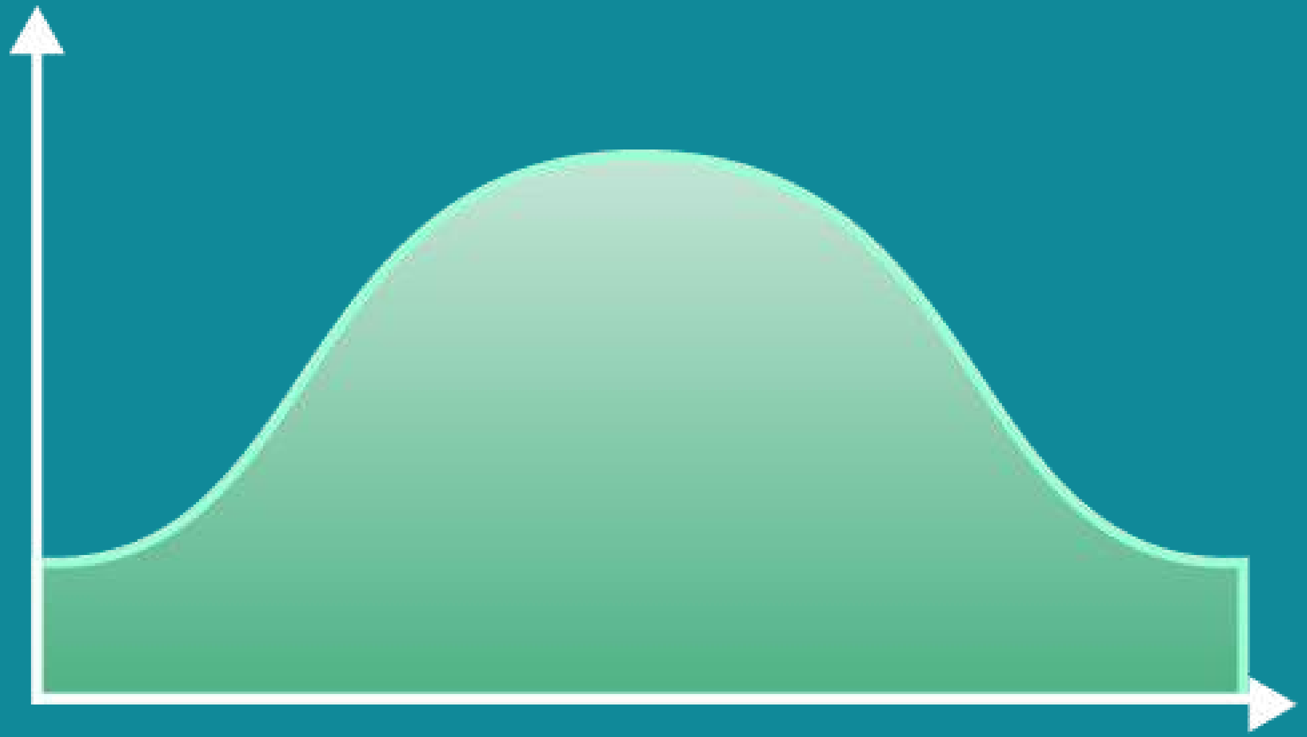
Ensures standard deviation is 1.

Importance of the Standard Normal Distribution.

- The new variable z, represents how many standard deviations away from the mean, each corresponding value is.
- We can transform any Normal Distribution into a Standard Normal Distribution using the transformation shown above.
- Convenient to use because of a table of known values for its CDF, called the Z-score table, or simply the Z-table.

Students' T Distribution

A **Normal Distribution** represents a small sample size approximation of a Normal Distribution.



Notation:

- $\boldsymbol{Y \sim t\ (k)}$

Key characteristics

- A small sample size approximation of a Normal Distribution.
- Its graph is bell-shaped curve, symmetric, but has **fat** tails.
- Accounts for extreme values better than the Normal Distribution.
- If k>1: $\boldsymbol{E(Y)} = \mu$ and $\boldsymbol{Var(Y) = s^2} \times \boldsymbol{\frac{k}{k-2}}$

Example and uses:

- Often used in analysis when examining a small sample of data that usually follows a Normal Distribution.

Exponential Distribution

$$\mathrm{Exp}(\lambda)$$

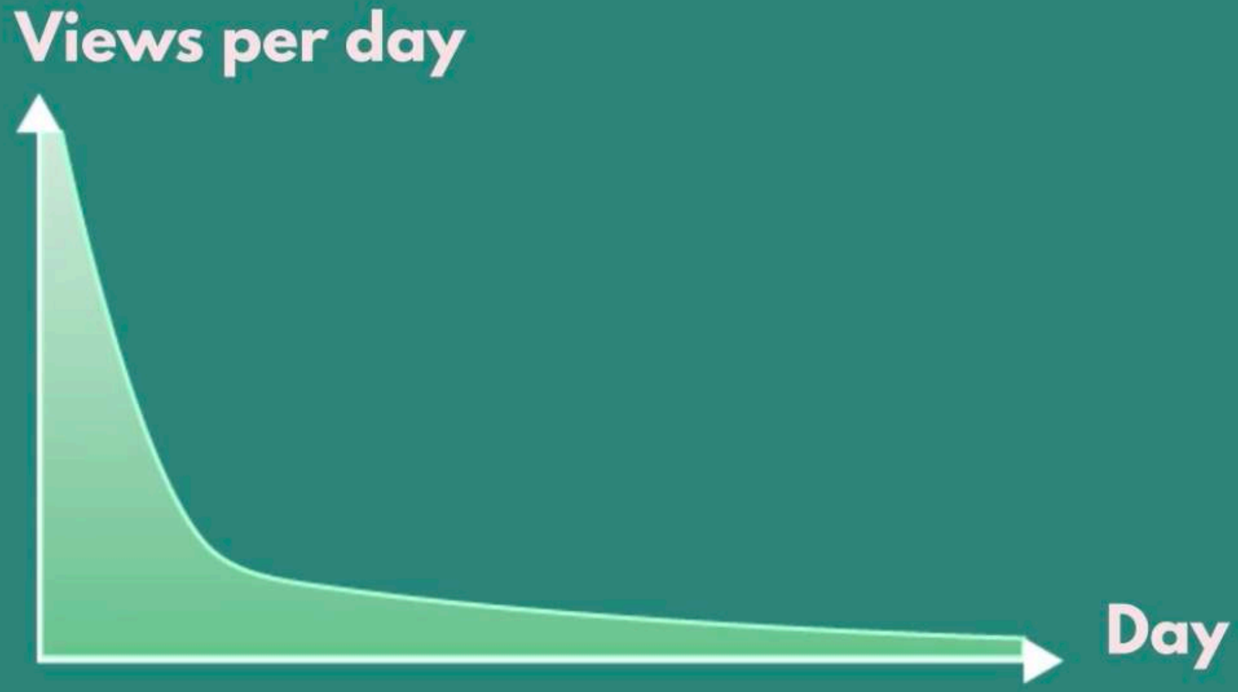
$$X \sim \mathrm{Exp}\left(\frac{1}{2}\right)$$

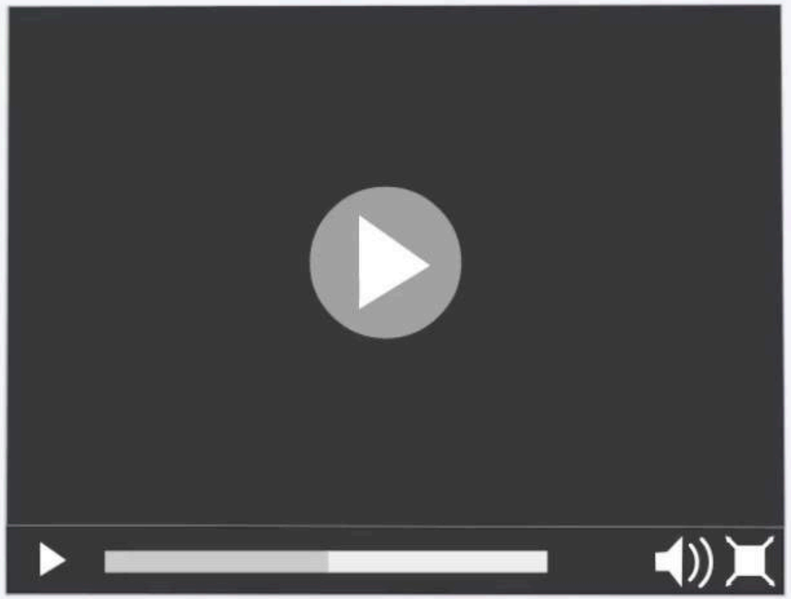
Exponential Distribution

- Starts off high
- Initially decreases
- Eventually plateauing

Youtube Video Example

- Aggregate amount of views keeps increasing
- Number of new ones diminishes
- Viewership focus shifts away



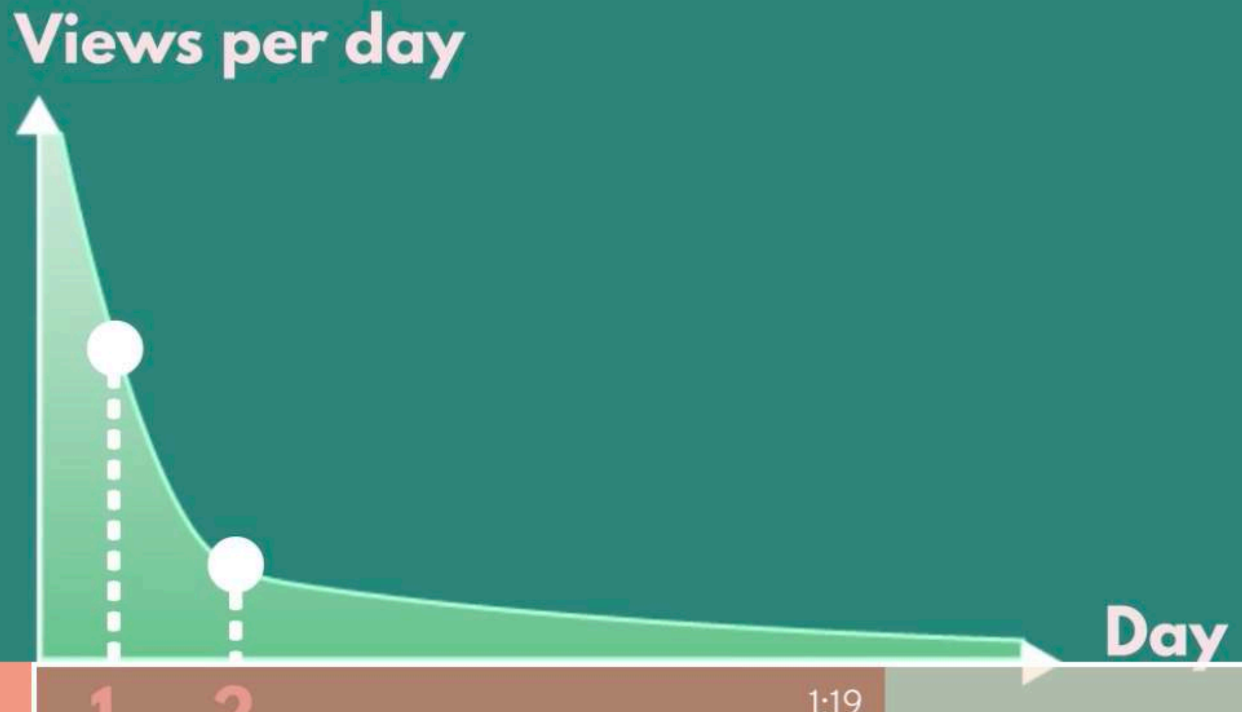


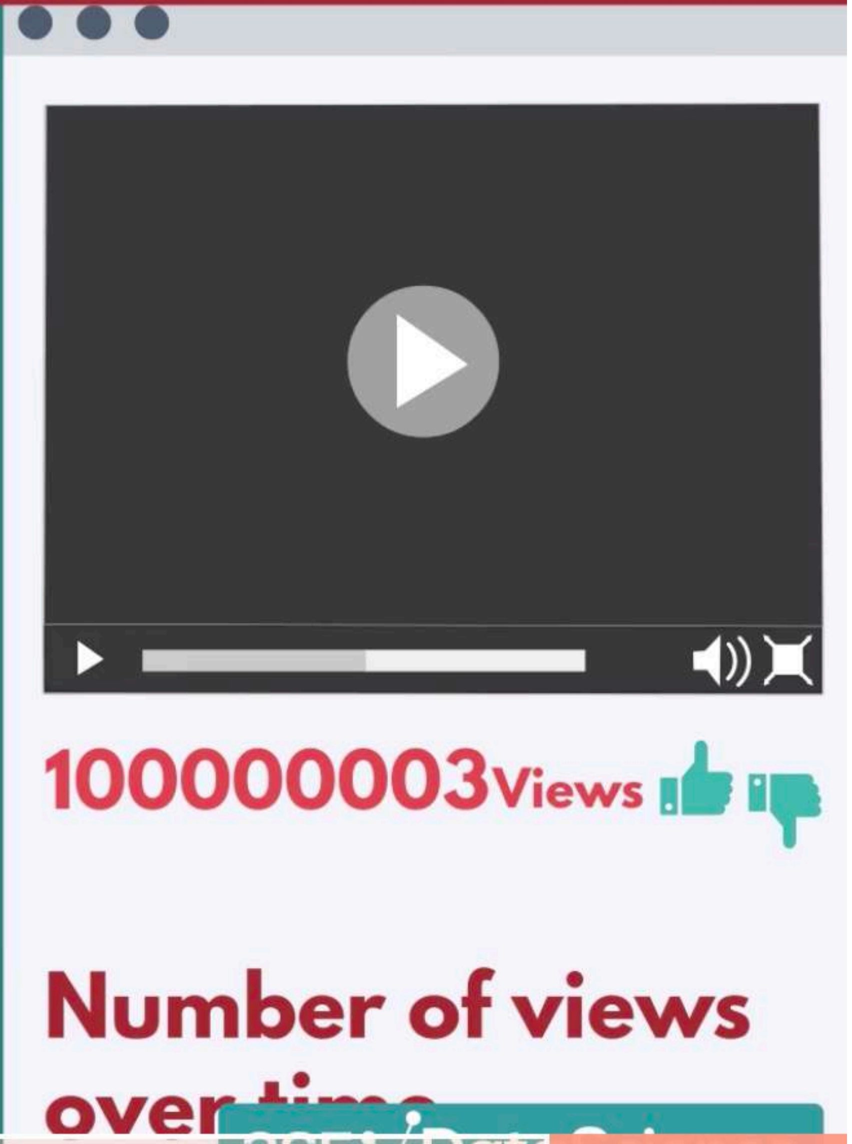
100000003 Views

Number of views over time

Youtube Video Example

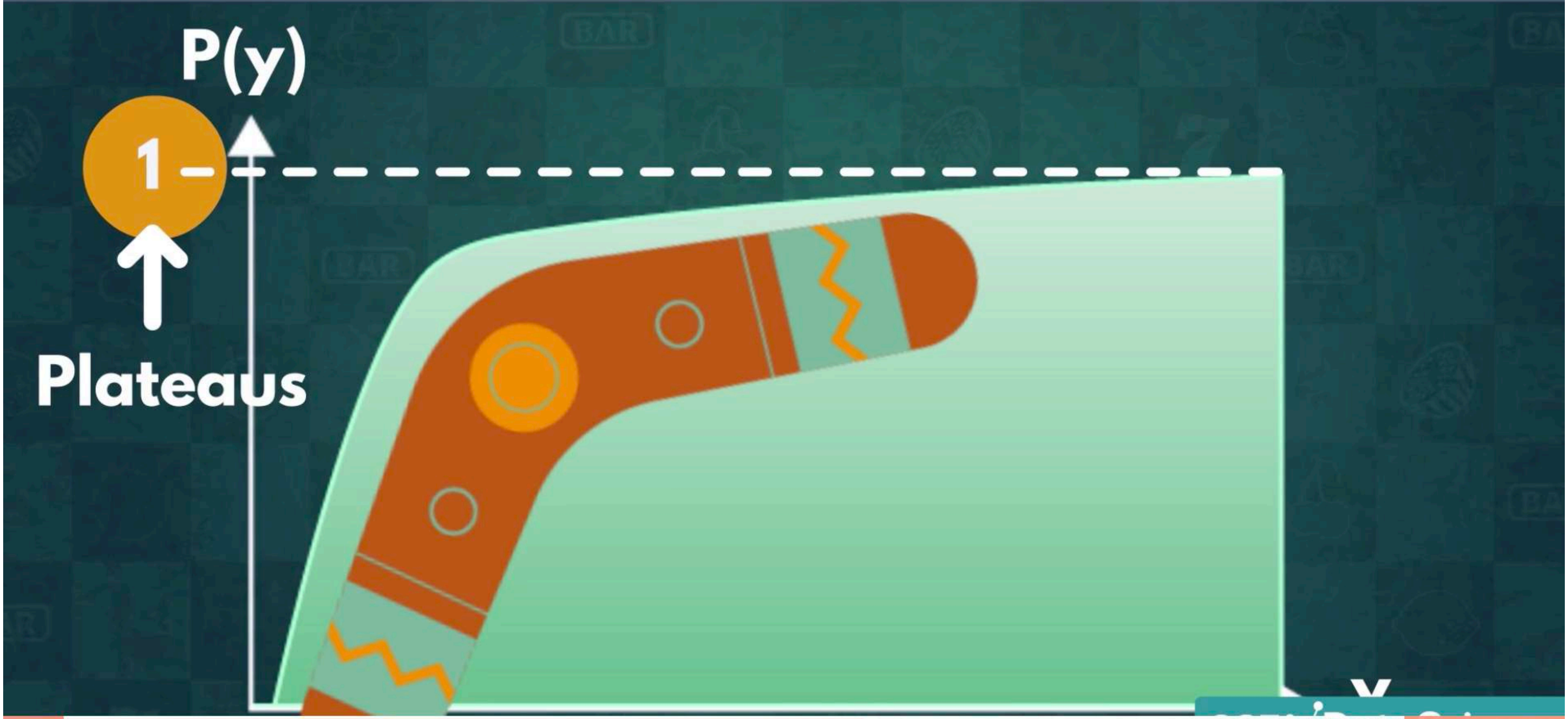
=> Most likely for a random viewing to have occured close to the video's intial release





PDF

CDF



1.50

Rate Parameter

Rate parameter → λ

How fast the CDF/PDF curve reaches the point of plateauing

How spread out the graph

E(Y), Var(Y)

$$E(Y) = \frac{1}{\lambda}$$

$$Var(Y) = \frac{1}{\lambda^2}$$

Drawbacks

- Unlike N or χ^2
- No table of known variables
(Z - table; χ^2- table)
- Transform it

Transformation

$$\left.\begin{array}{r} Y \sim \mathrm{Exp}(\lambda) \\ X = \ln(Y) \end{array}\right\} \Rightarrow X \sim N(\mu, \sigma^2)$$

- One of the most common transformations

Continuous Distributions

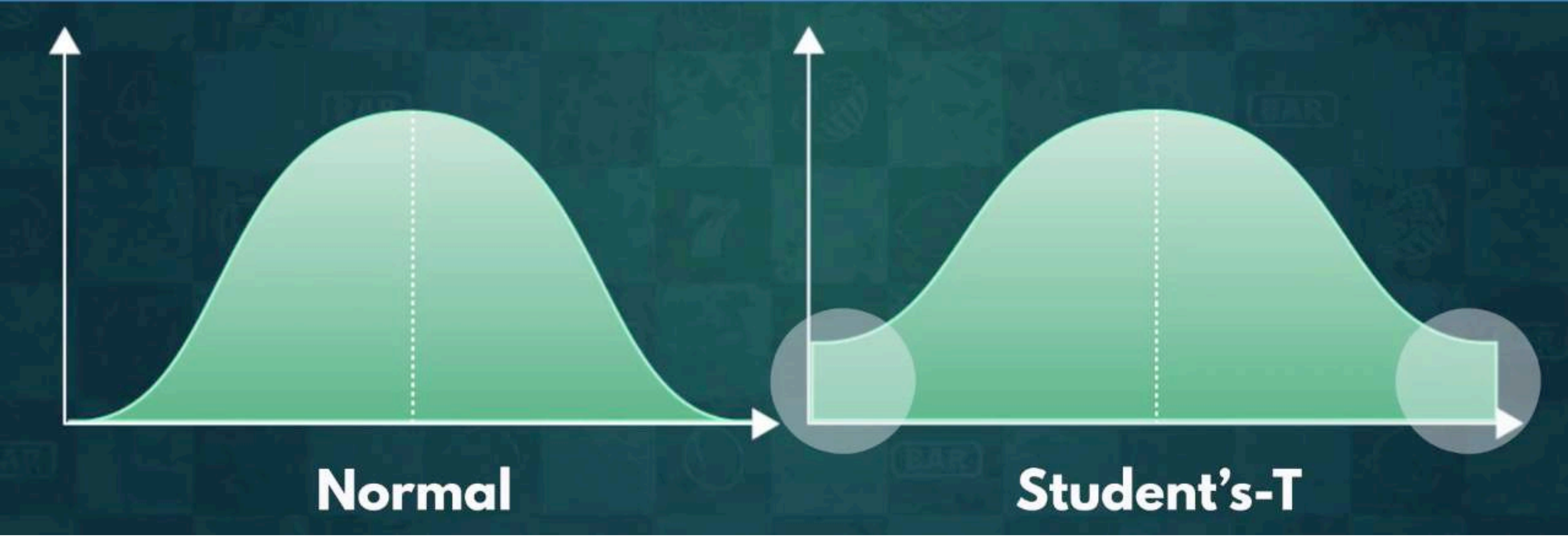
Limited data

Student's-T distribution

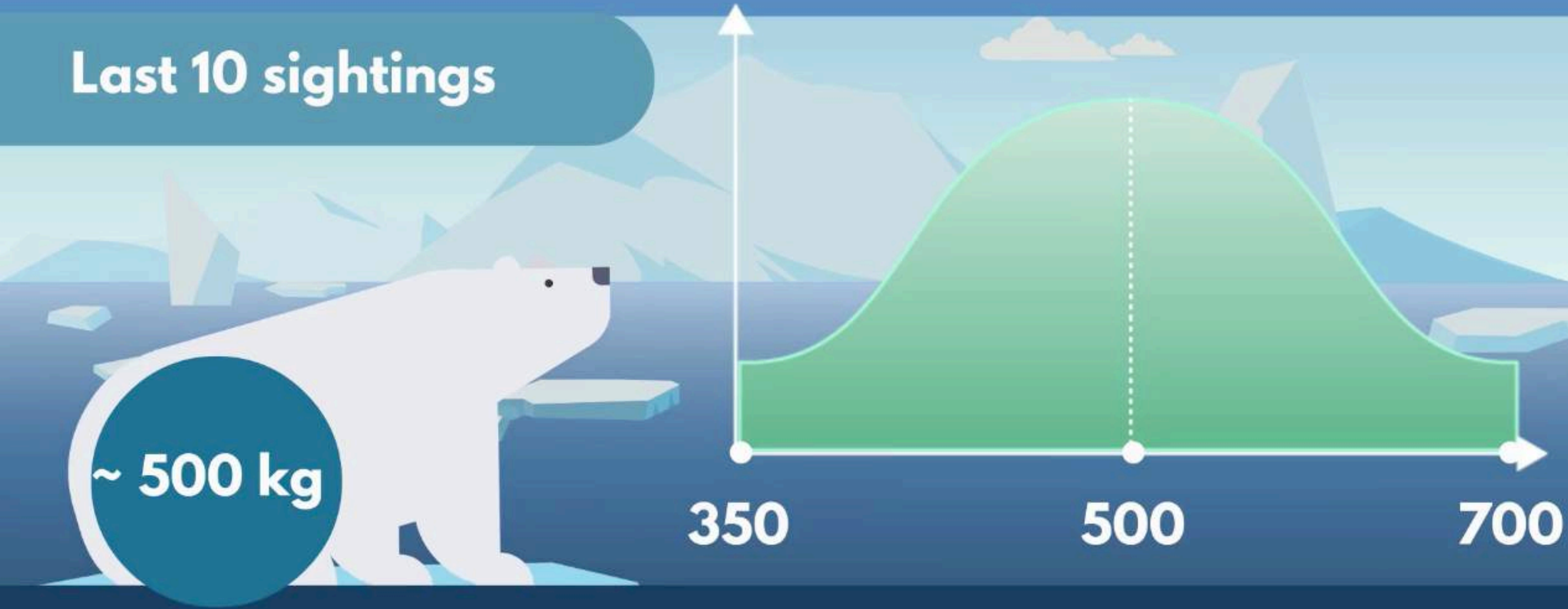
A small sample approximation of a Normal distribution

Continuous Distributions

Student's-T accommodates extreme values significantly better



Student's T Distribution



Any extreme value represents a much bigger part of the population

Logistic Distribution

Location

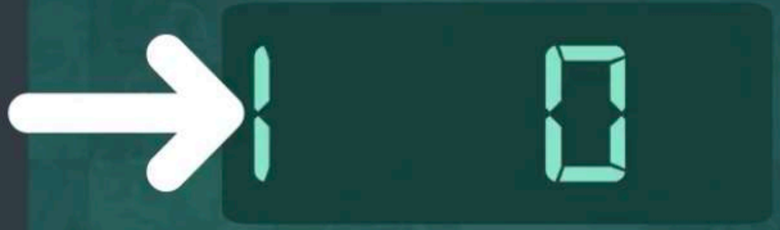
Scale Parameter

Logistic (μ, S)

$$X \sim \text{Logistic}(6, 3)$$

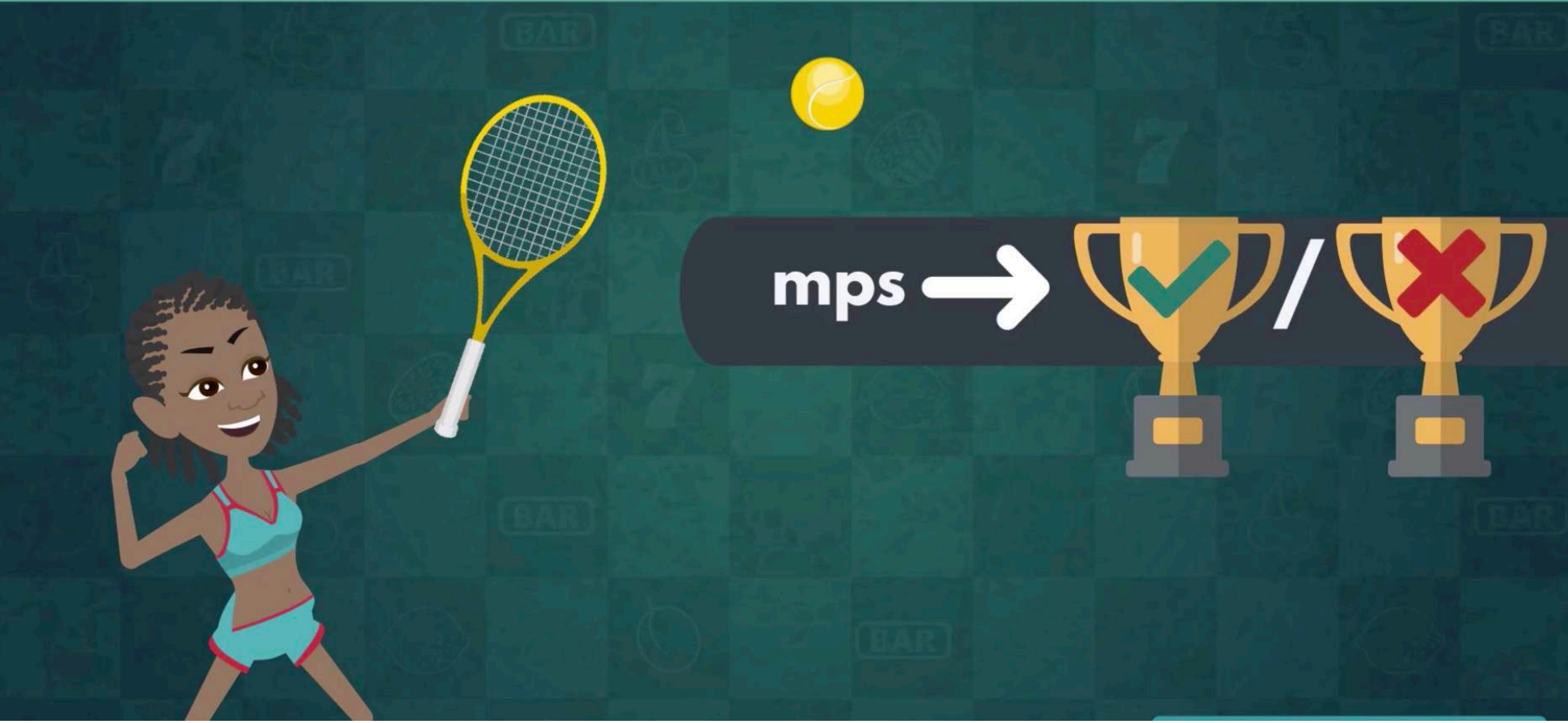
1.50

Application



Forecasting: Victory or defeat

Tennis Example



1.50

Tennis Example

Expectation

$V\uparrow \rightarrow t\downarrow \rightarrow 1$

Reality

$V\uparrow \rightarrow \text{Accuracy}\downarrow \rightarrow 0$

We cannot assume a linear relationship

Tennis Example

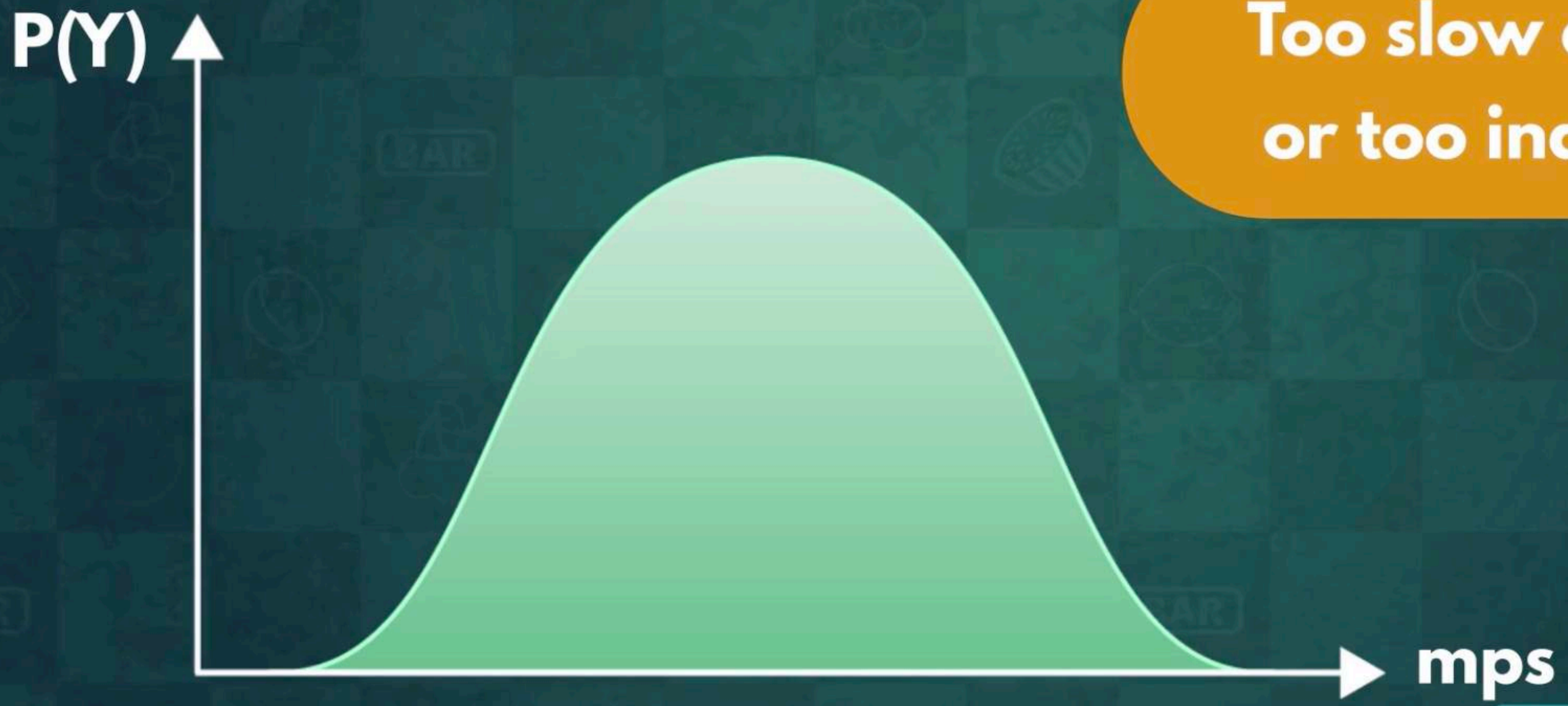
Theory:

Optimal speed → Still accurate

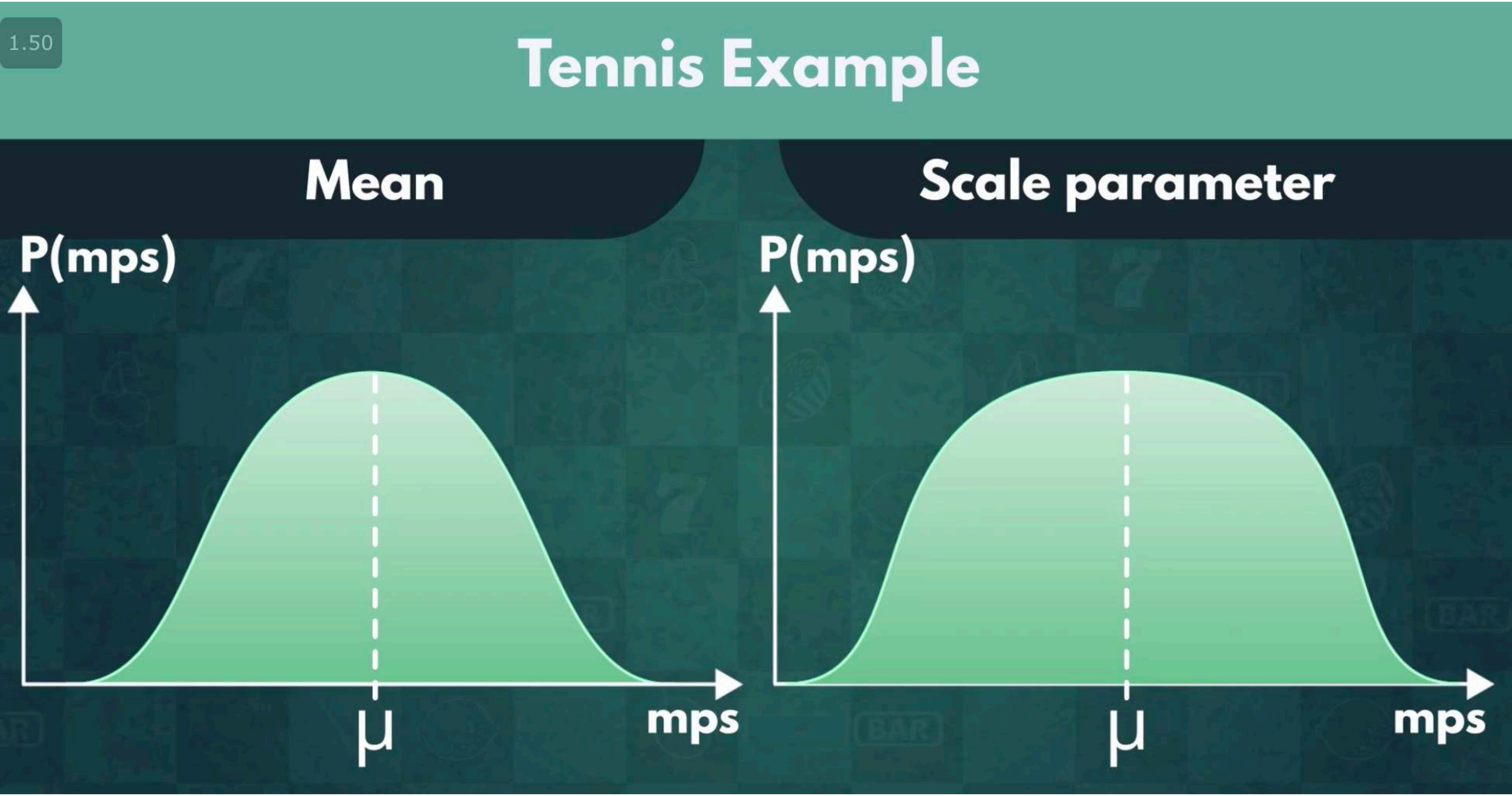
Most of the shots we convert into points will likely have similar velocities

Tennis Example

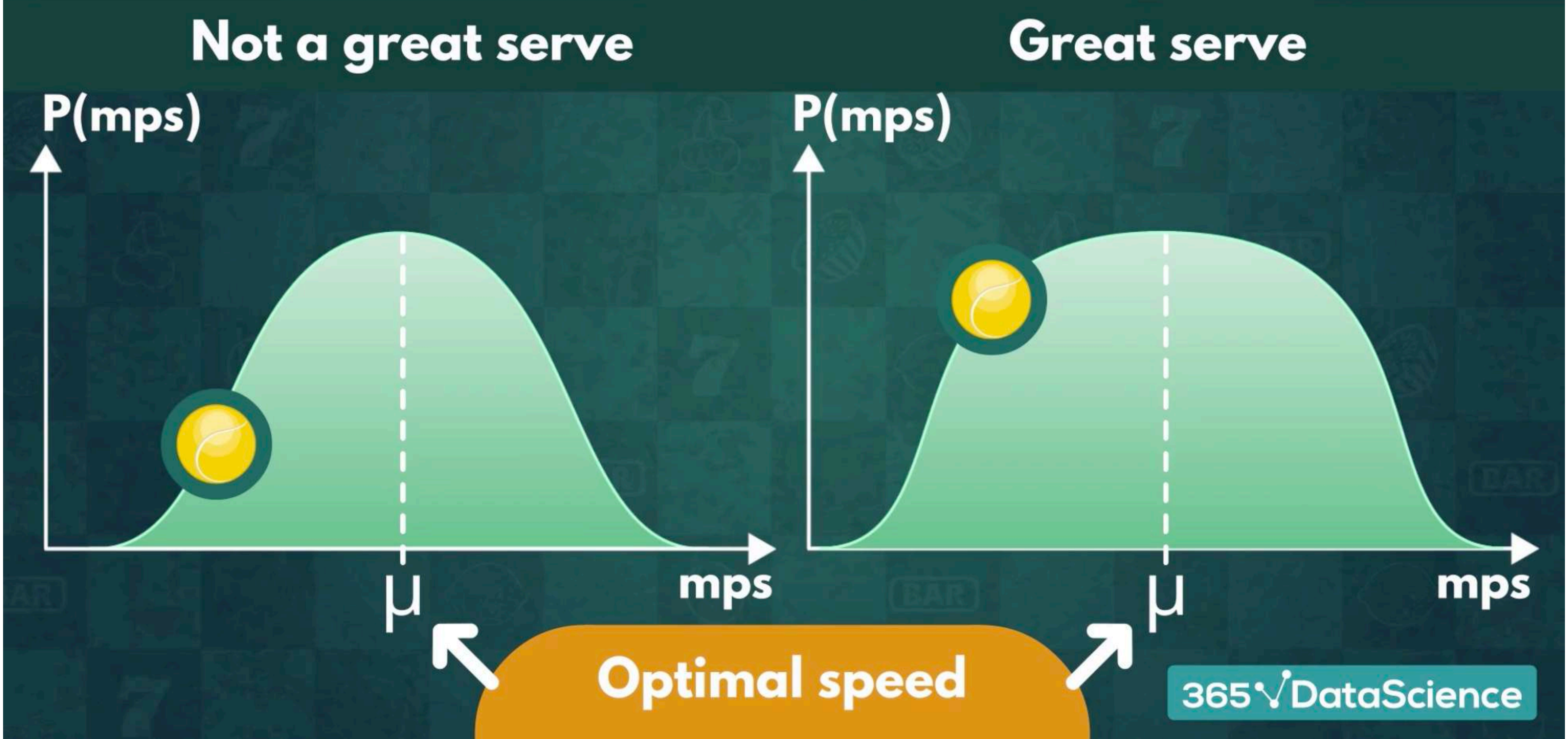
Shots further from optimal speed become:



Too slow and easy or too inaccurate


1.50
Tennis Example
Mean
Scale parameter
P(mps)
μ
mps
P(mps)
μ
mps

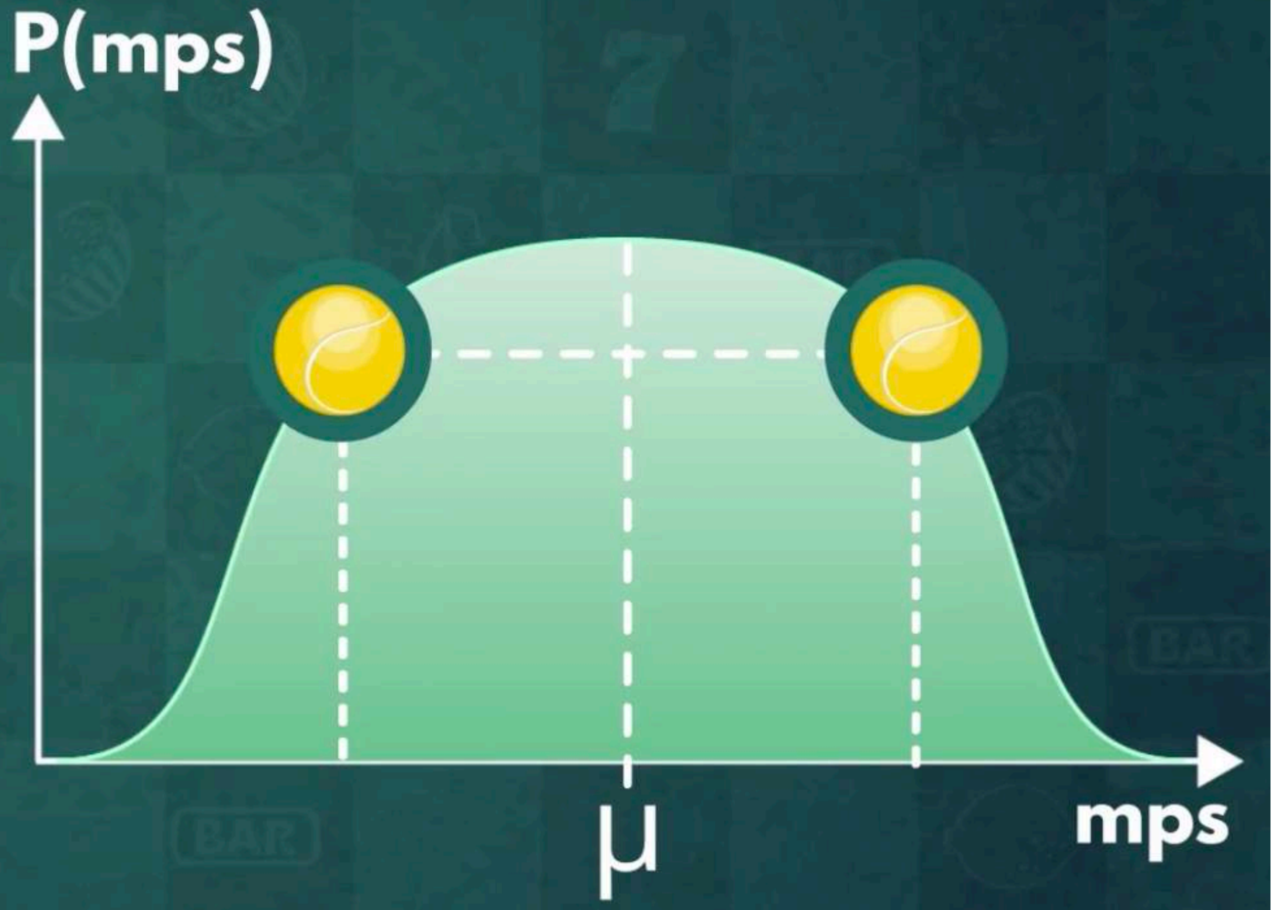
Tennis Example



Tennis Example

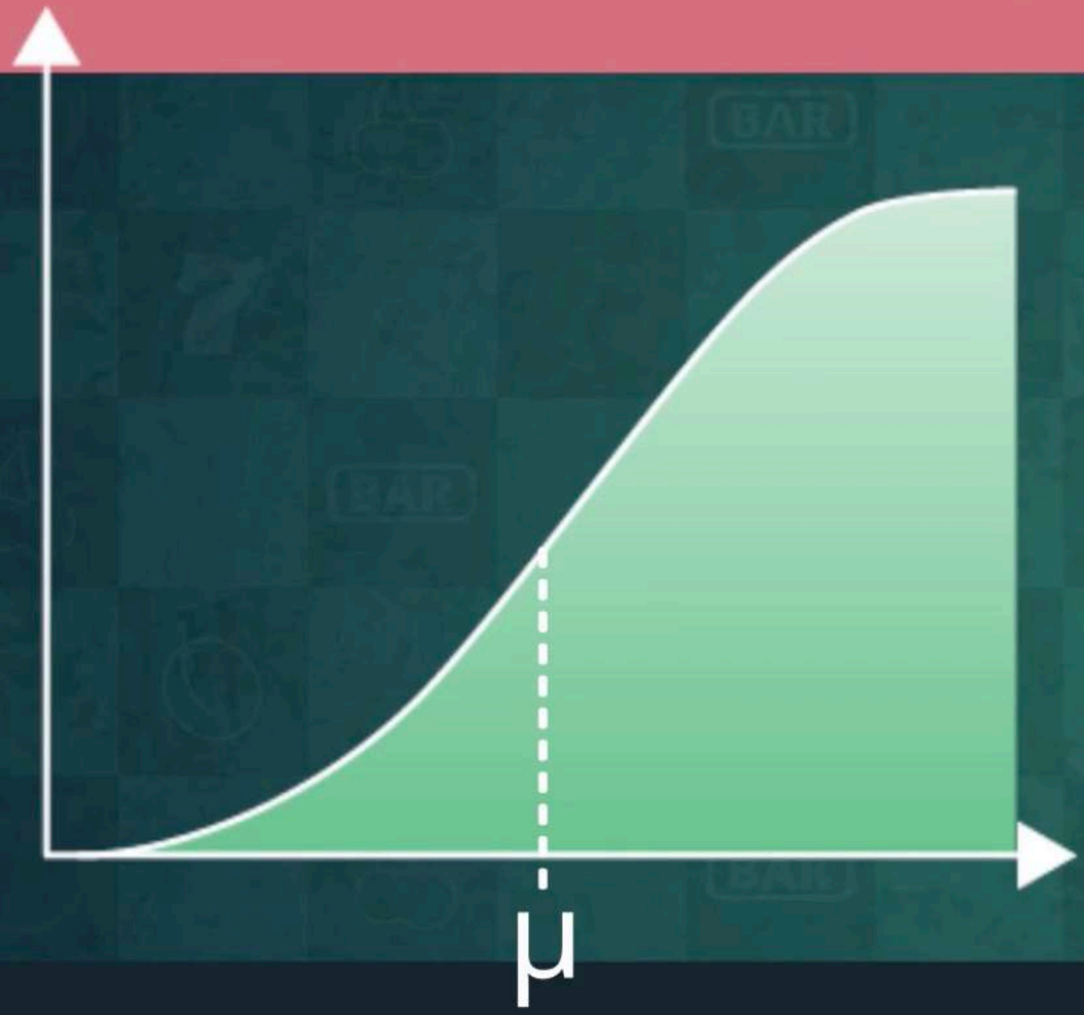
Serena Williams

↳ Fantastic serves even if the ball moves much faster or slower than it optimally should



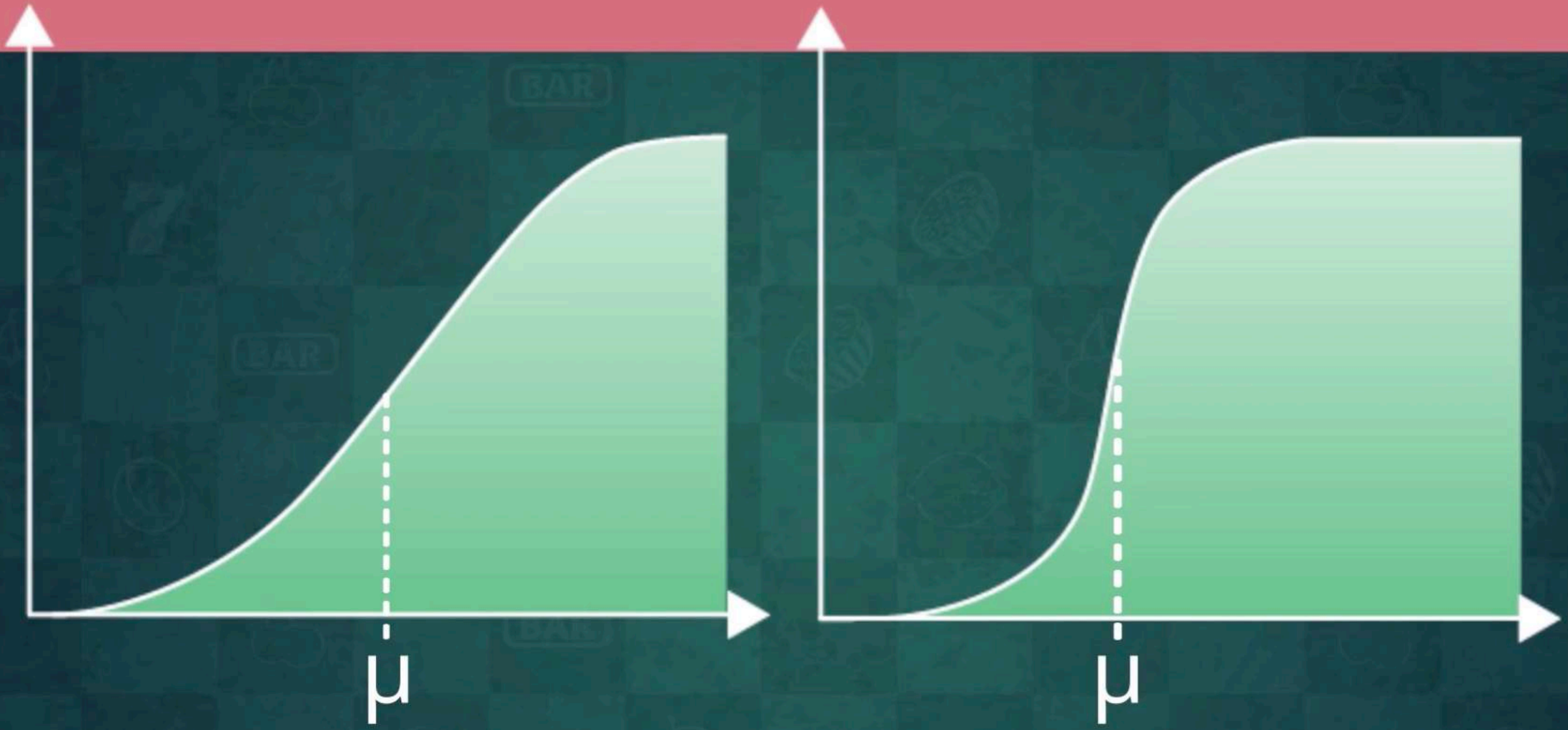
1.50

CDF



Once we reach values near the mean, the probability drastically goes up

CDF



The later the graph starts to pick up, but the quicker it reaches values close to 1

365√DataScience

Expected Value and Variance

Fullscreen

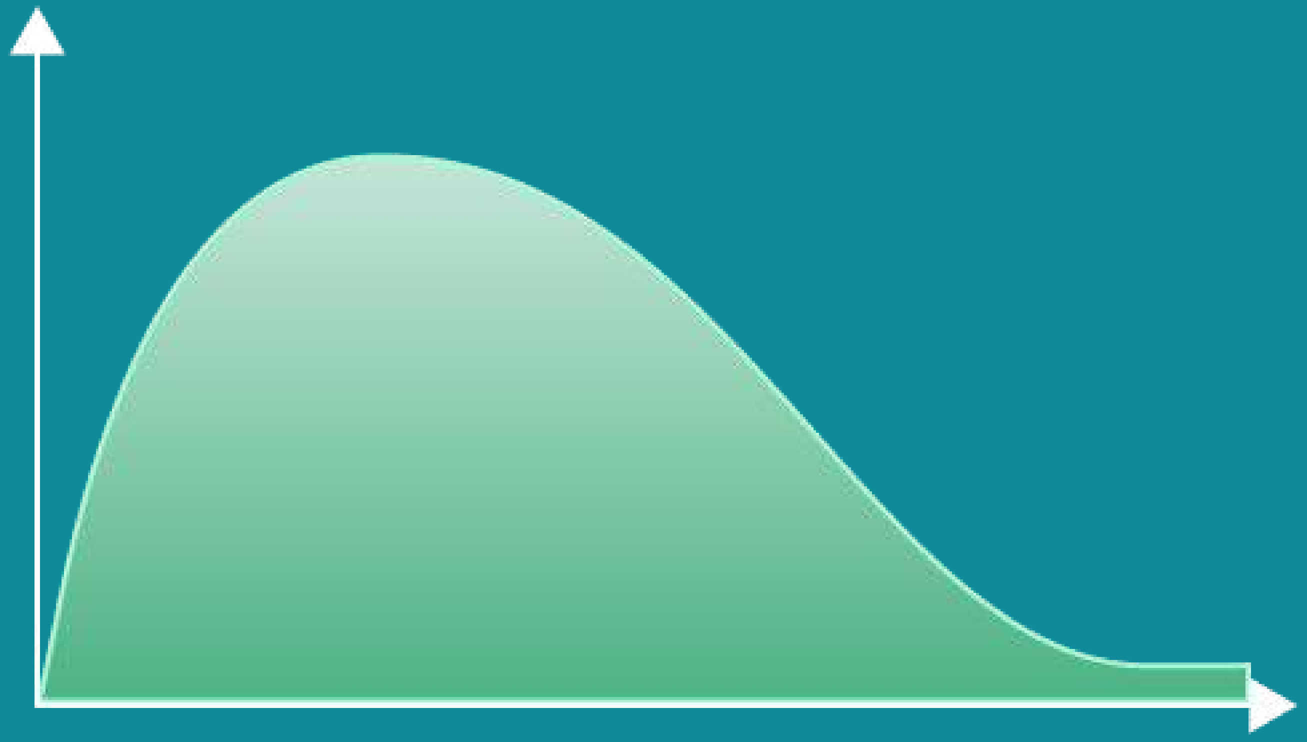
$$E(Y) = \mu$$

$$Var(Y) = \frac{s^2\pi^2}{3}$$

Chi-Squared Distribution

/kai/

A **Chi-Squared** distribution is often used.



Notation:

- $Y \sim \chi^2(k)$

Key characteristics

- Its graph is asymmetric and skewed to the right.
- $E(Y) = k$
- $Var(Y) = 2k$
- The Chi-Squared distribution is the square of the t-distribution.

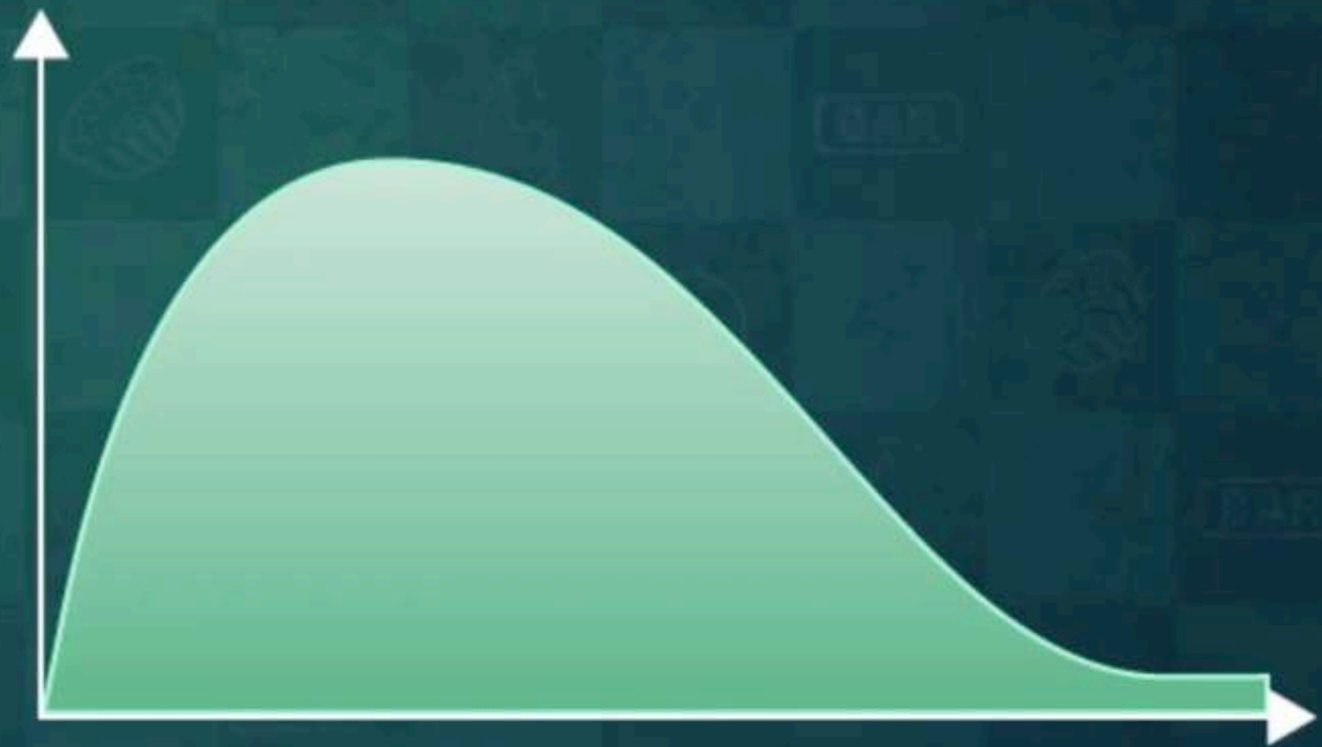
Example and uses:

- Often used to test goodness of fit.
- Contains a table of known values for its CDF called the χ^2-table. The only difference is the table shows what part of the table

Continuous Distributions

Chi-Squared

- Asymmetric
- Only consists of non-negative values



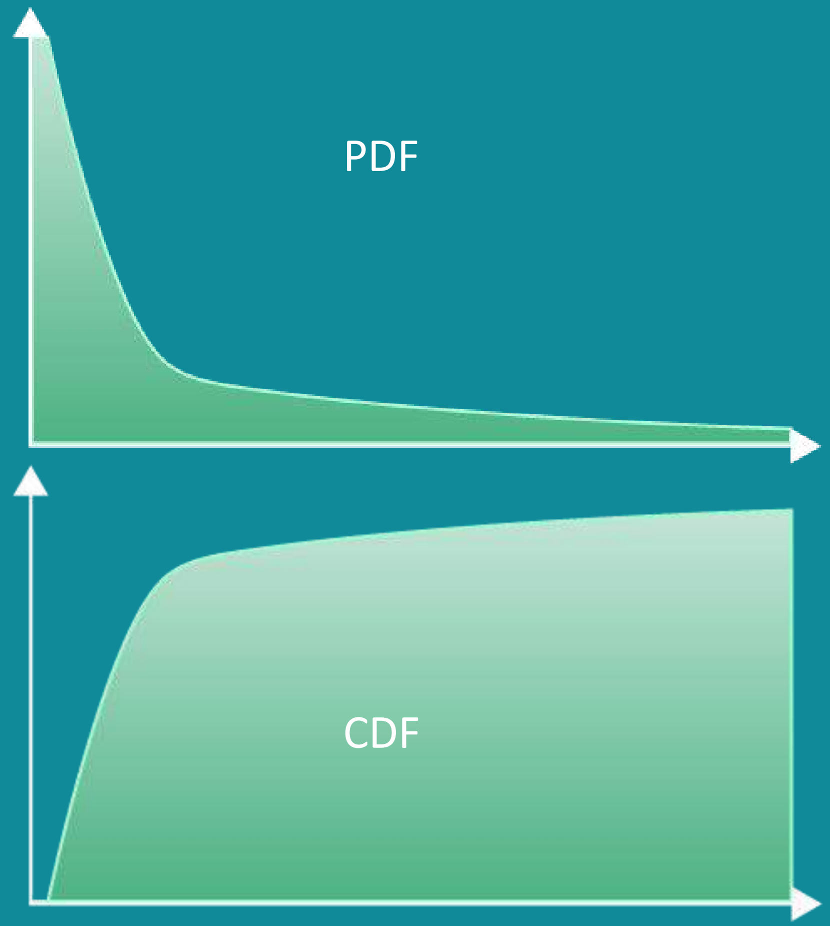
Chi-Squared

The Chi-Squared does not often mirror real life events

- Used in Hypothesis Testing
- Goodness of fit

Exponential Distribution

The **Exponential Distribution** is usually observed in events which significantly change early on.



Notation:

- $Y \sim Exp\ (\lambda)$

Key characteristics

- Both the PDF and the CDF plateau after a certain point.
- $E(Y) = \frac{1}{\lambda}$
- $Var(Y) = \frac{1}{\lambda^2}$
- We often use the natural logarithm to transform the values of such distributions since we do not have a table of known values like the Normal or Chi-Squared.

Example and uses:

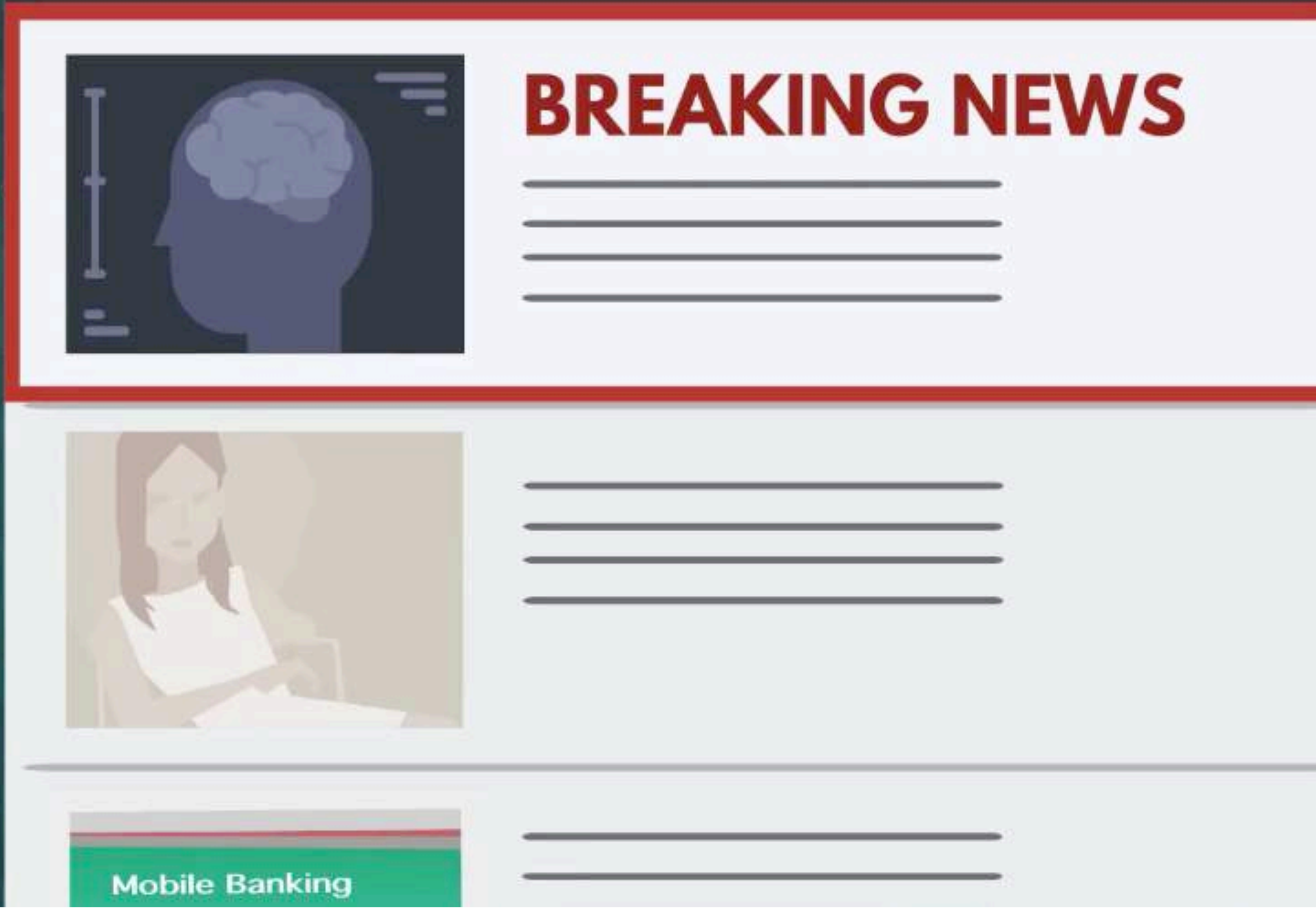
- Often used with dynamically changing variables, like online website traffic or radioactive decay.

Exponential distribution

Events that are rapidly changing early on

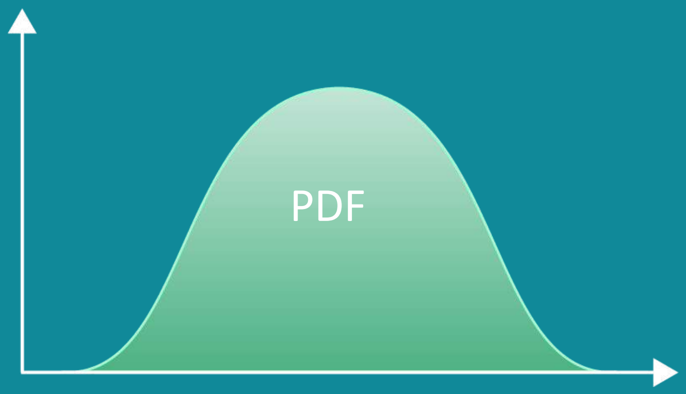
Fresh

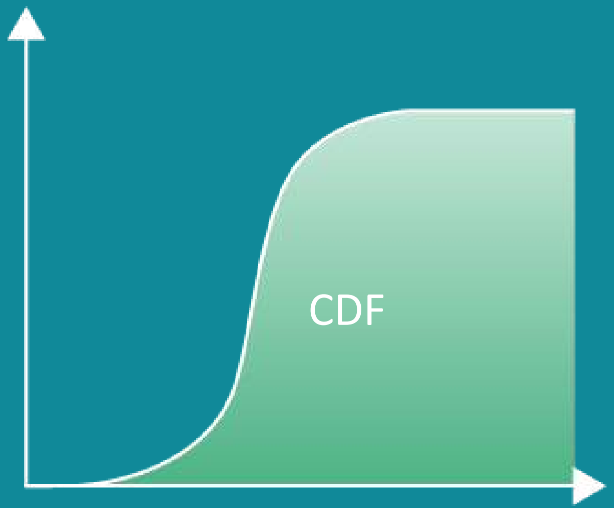
Online news articles



Logistic Distribution

The **Continuous Logistic Distribution** is observed when trying to determine how continuous variable inputs can affect the probability of a binary outcome.





Notation:

- $Y \sim Logistic\ (\mu, s)$

Key characteristics.

- $E(Y) = \mu$
- $Var(Y) = \frac{s^2 \times \pi^2}{3}$
- The CDF picks up when we reach values near the mean.
- The smaller the scale parameter, the quicker it reaches values close to 1.

Example and uses:

- Often used in sports to anticipate how a player's or team's performance can determine the outcome of the match.

Continuous Distributions

Logistic distribution

- Useful in forecast analysis
- Useful for determining a cut-off point for a successful outcome

Logistic distribution

- How much of an in-game advantage is necessary to predict victory
- Our predictions would never reach true certainty

DOTA 2