دورهٔ آموزشی «علم داده»

Data Science Course

جلسهٔ چهارم:

توزیع‌های احتمال

مدرس: محمد فزونی

عضو هیات علمی دانشگاه گنبدکاووس

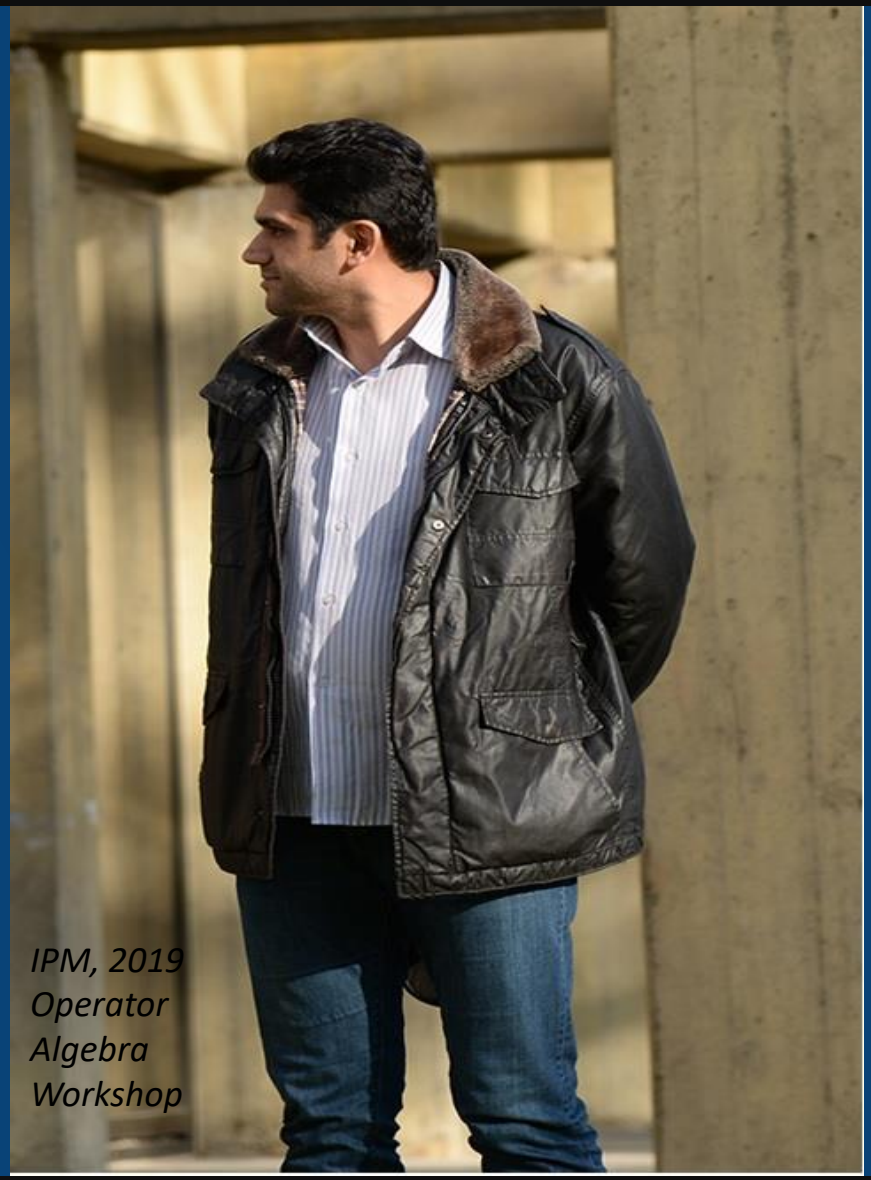
پائیز ۱۳۹۹

دربارهٔ من

Mohammad Fozouni (Ph.D.)
Dep. of Math & Statistics
Gonbad Kavous University

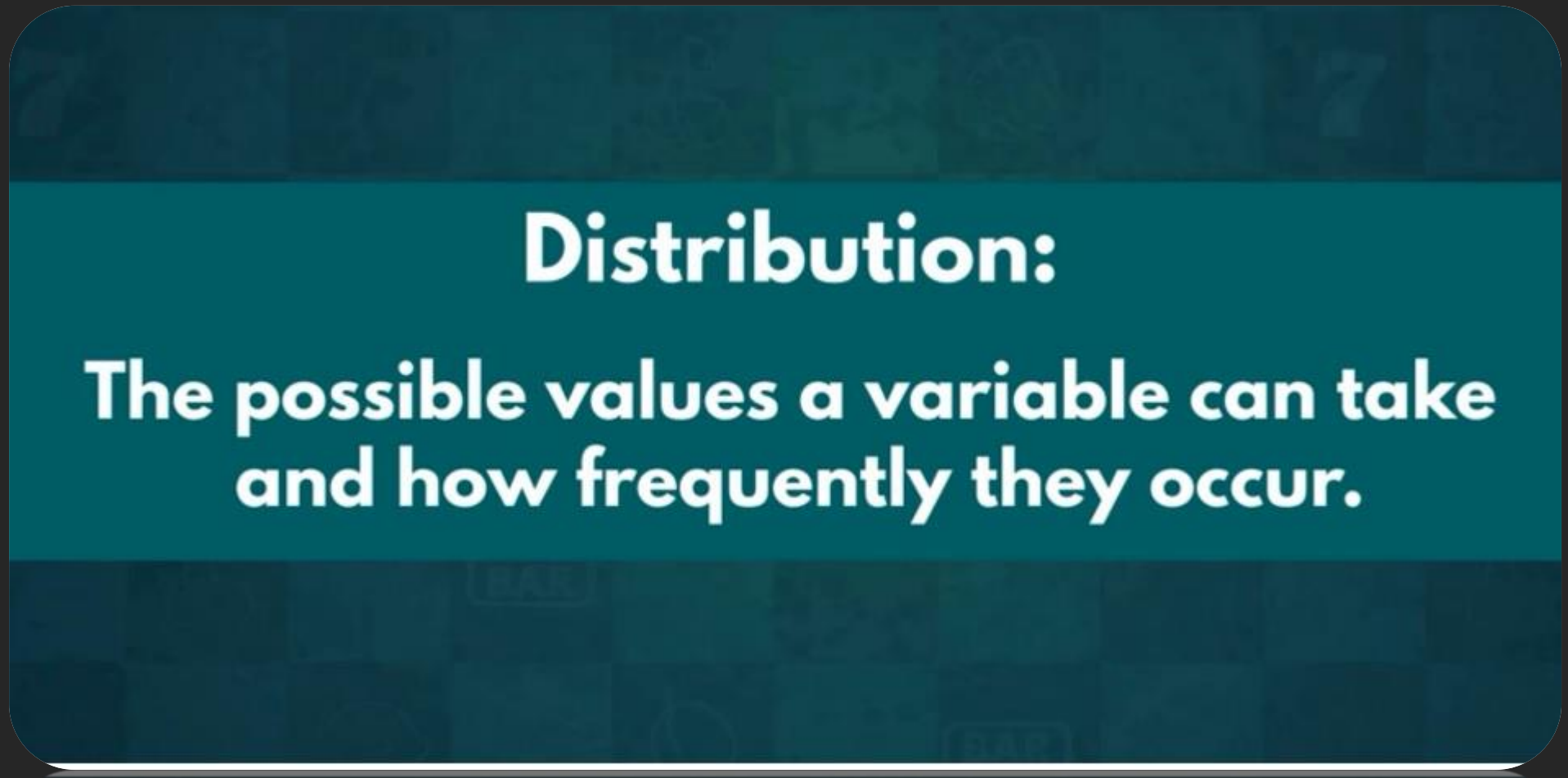
- fozouni@hotmail.com
- https://m-fozouni.ir
- http://profs.gonbad.ac.ir/fozouni/en
- https://www.aparat.com/elmedade

#data_science_fozouni

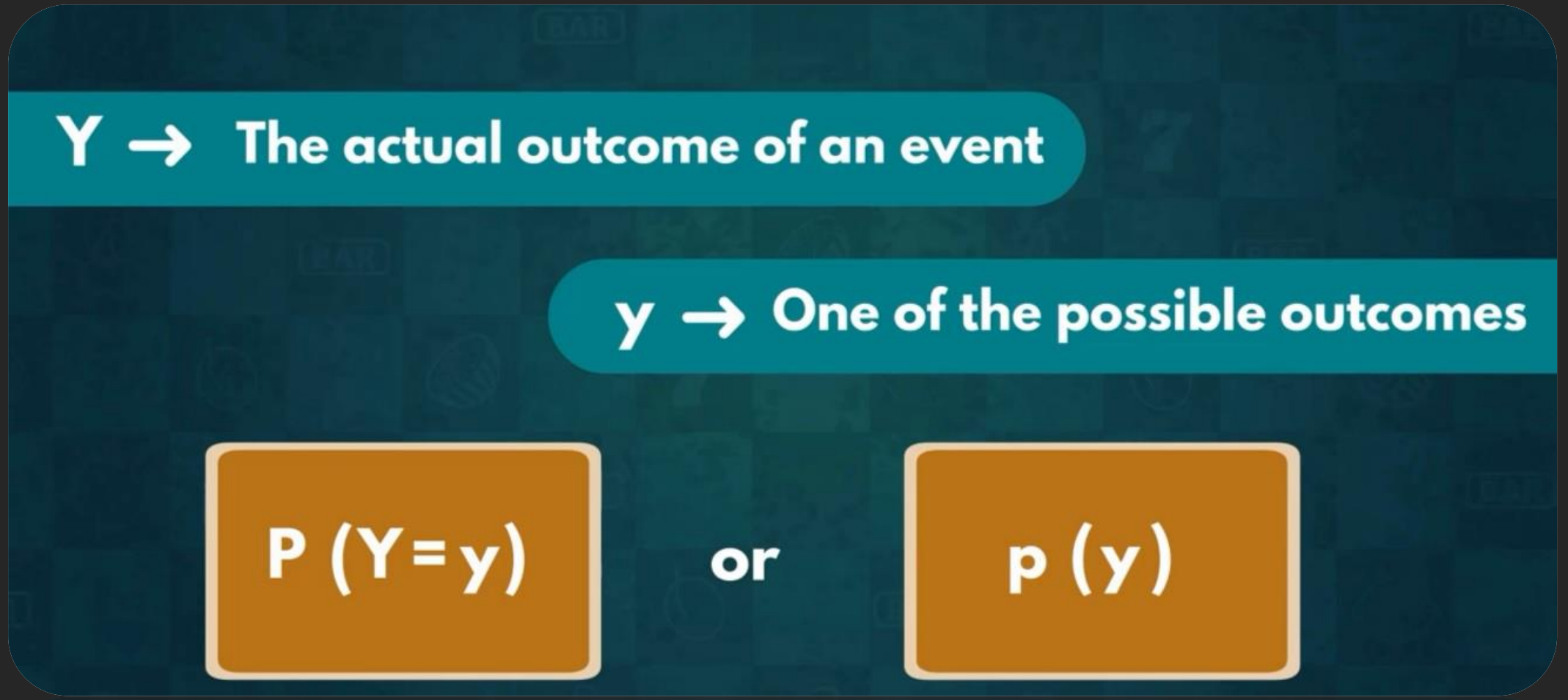


IPM, 2019
Operator Algebra Workshop

What is a Probability Distribution?



One Notation



Example

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

y → 5

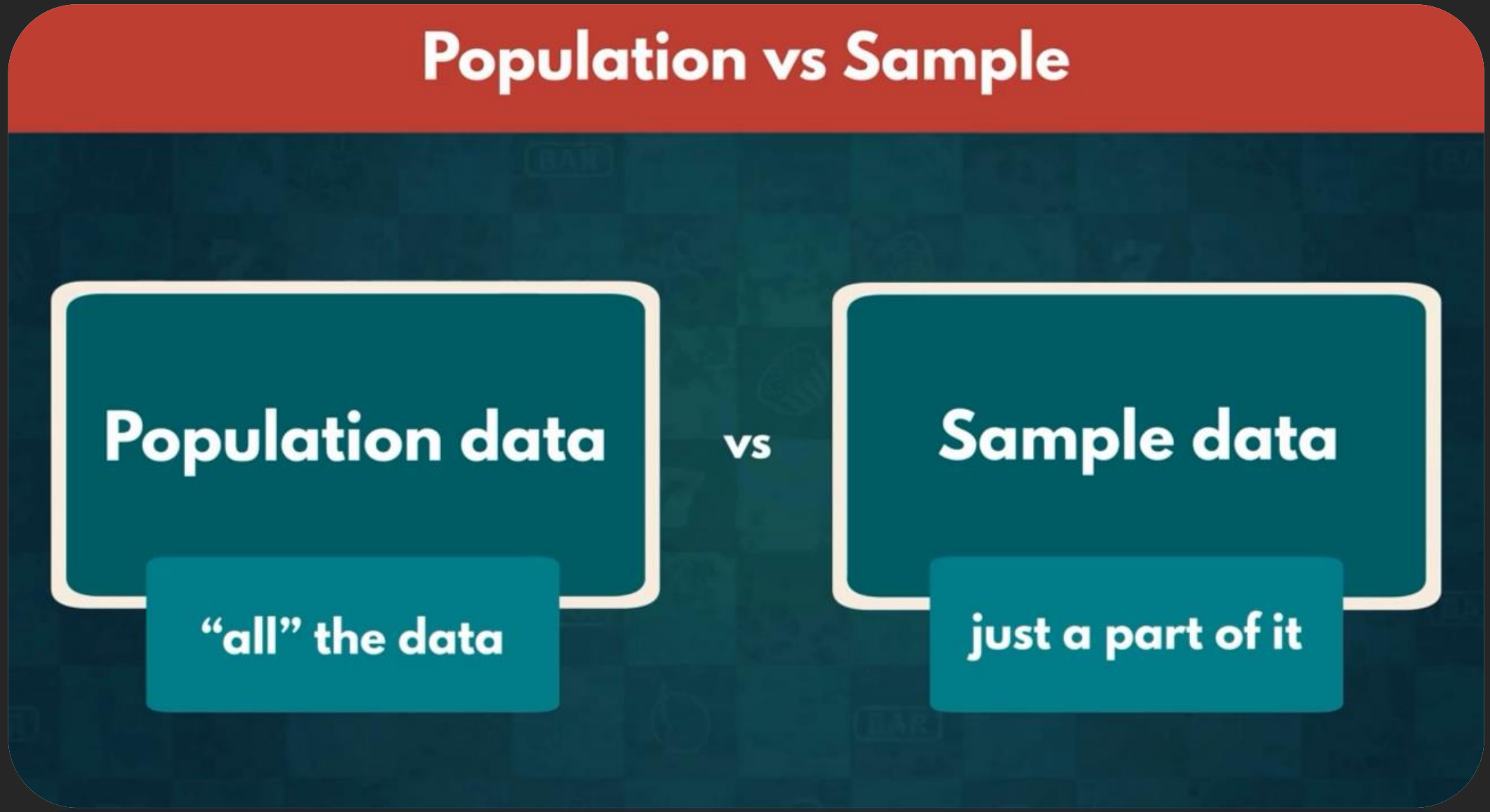
P (Y = 5) or p (5)

Definitions

Two characteristics:

mean → average value → μ

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


Population vs Sample
Population data
vs
Sample data
"all" the data
just a part of it

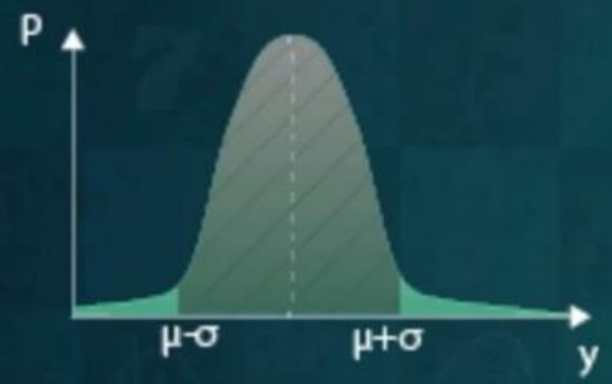
Standard Deviation

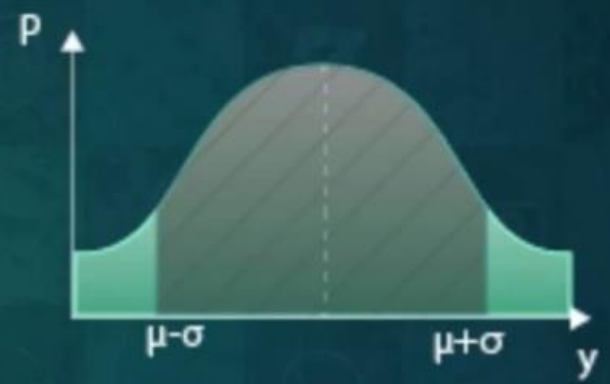
Standard deviation → square root of variance $\sqrt{\sigma^2}$

- σ — population
- s — sample

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

Standard Deviation





the more congested the middle of the distribution

the less data falls within the interval

the more data falls within that interval

the more dispersed the data is

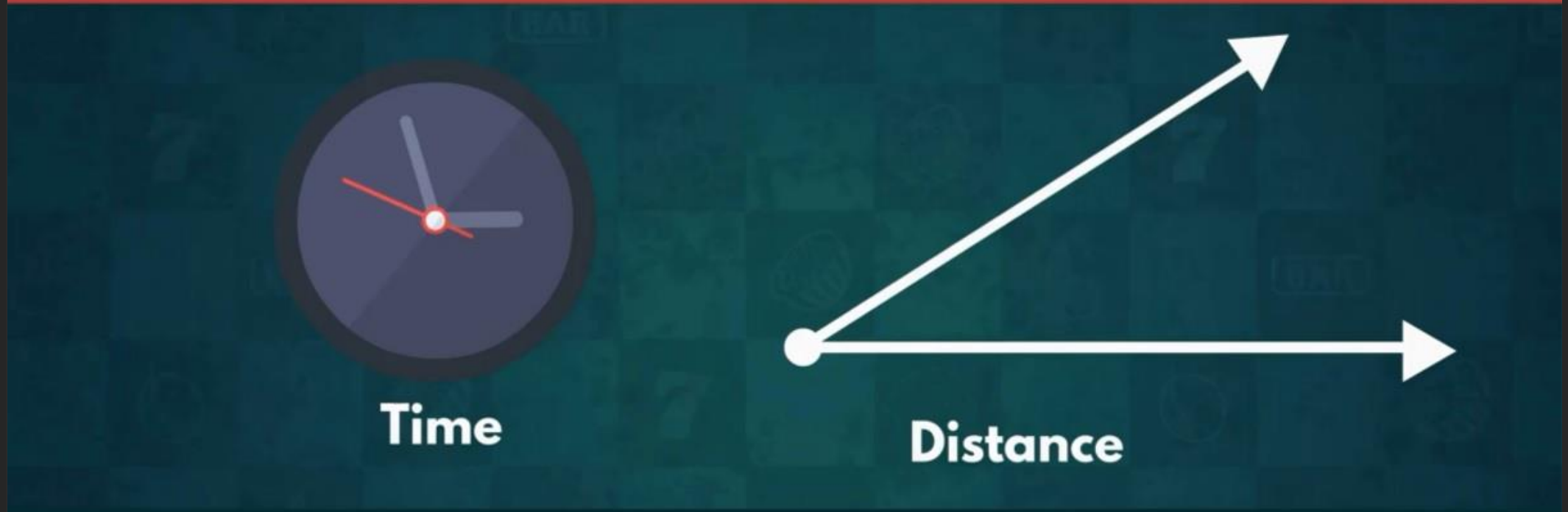
Types of Probability Distributions

Die

Picking a card

Finite number of outcomes ➡ Discrete distributions

Types of Probability Distributions



Infinitely many outcomes → Continuous distributions

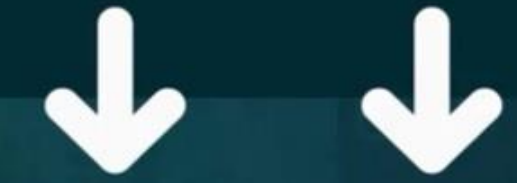
Discrete Distributions

All outcomes are equally likely ➡ Equiprobable

Uniform Distribution

Discrete Distributions

Events with only two possible outcomes



Bernoulli Distribution

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

Discrete Distributions

Carrying out a similar experiment several times in a row

- **Binomial Distribution**
  - Two outcomes per iteration
  - Many iterations

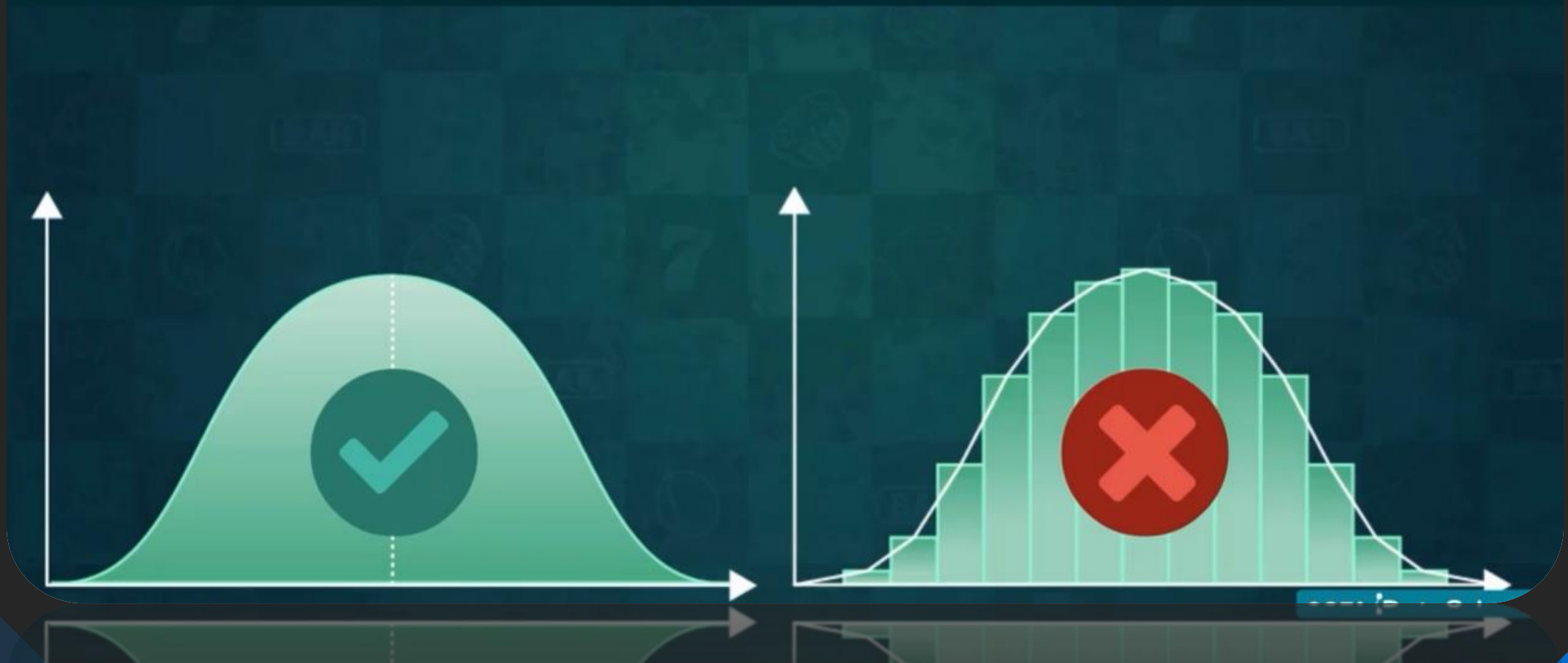
Discrete Distributions

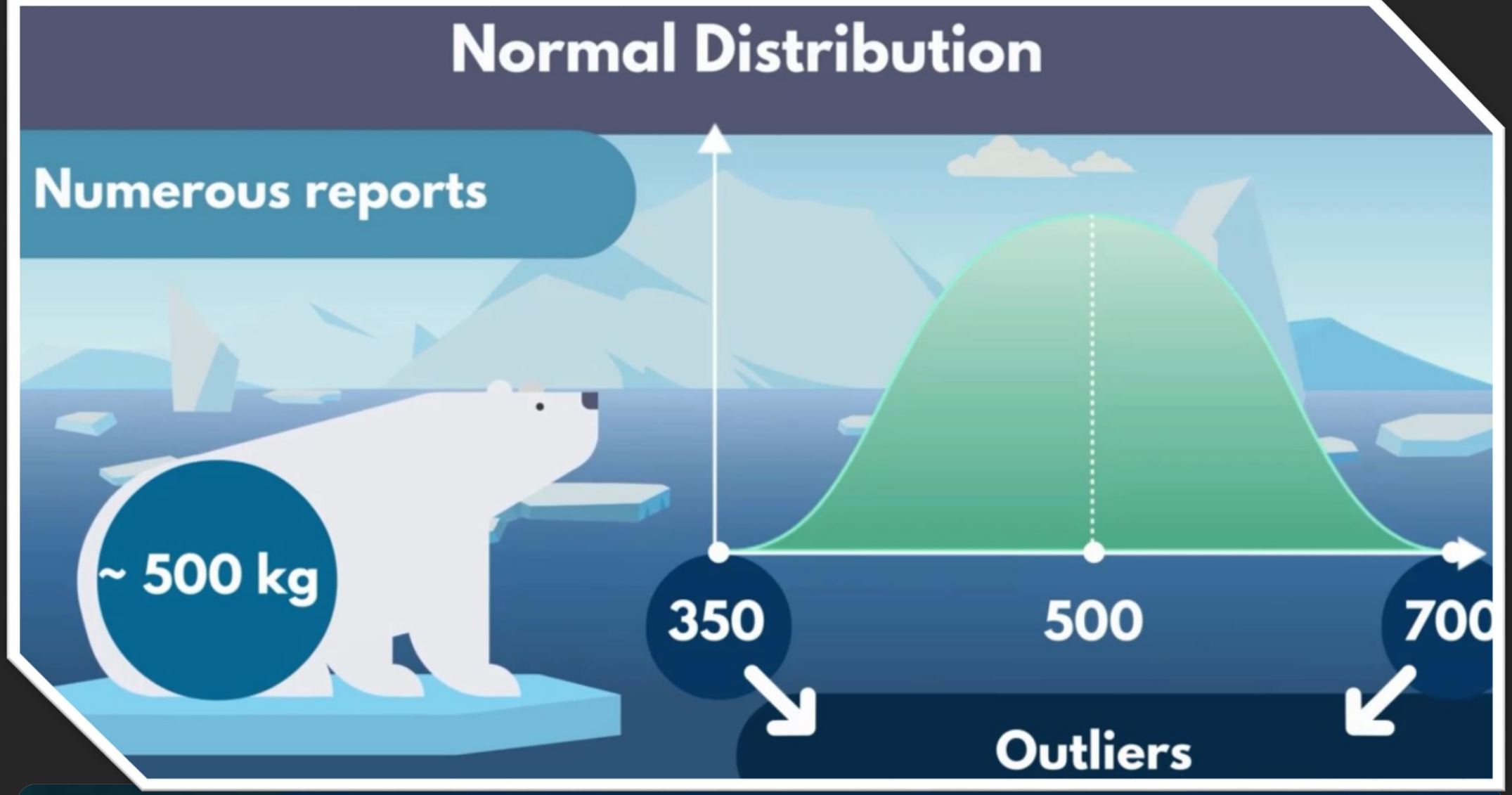
Test out how unusual an event frequency is for a given interval

Poisson Distribution

Continuous Distributions

The probability distribution would be a curve

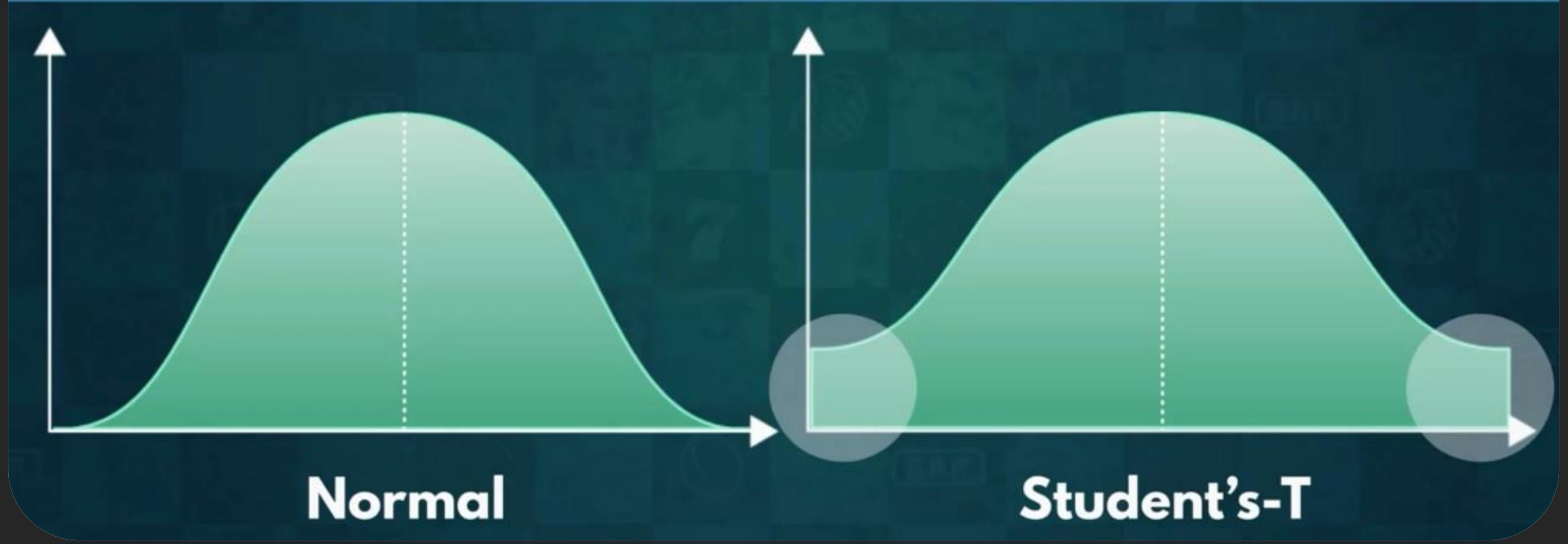




Normal distributions — Often observed in nature

Continuous Distributions

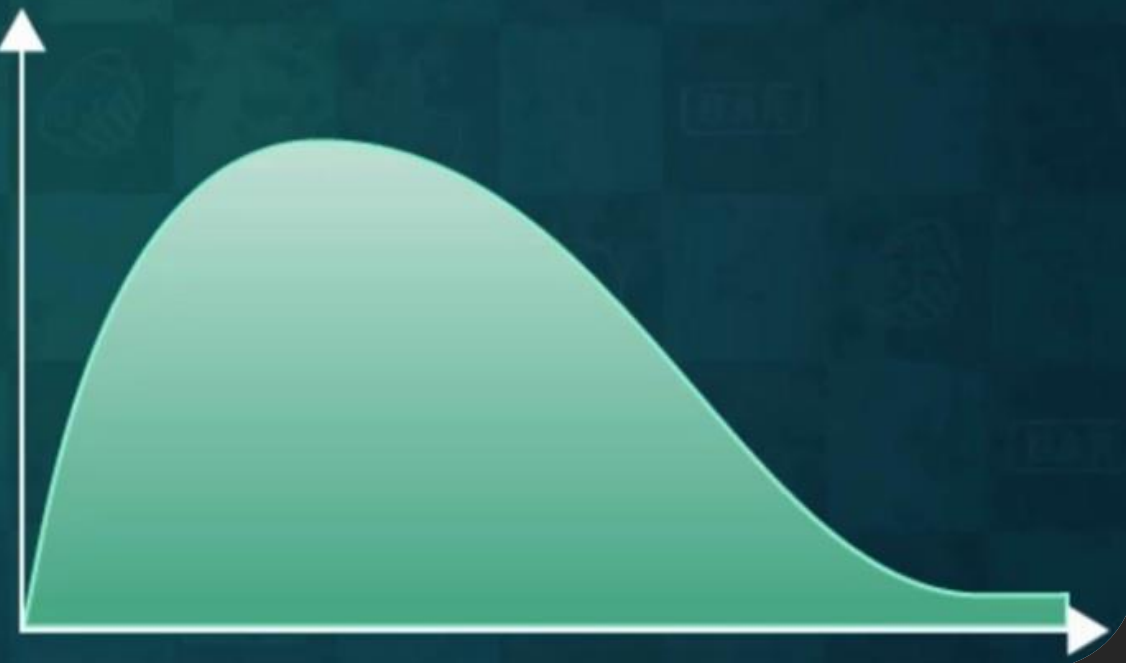
Student's-T accommodates extreme values significantly better



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

Events that are rapidly changing early on

Fresh

Online news articles

BREAKING NEWS

Continuous Distributions

Logistic distribution

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

SUBJECT OF THE NEXT VIDEO

Discrete Distributions
In details