دورهٔ آموزشی «علم داده»

Data Science Course

جلسهٔ چهارم:

توزیع‌های پیوسته

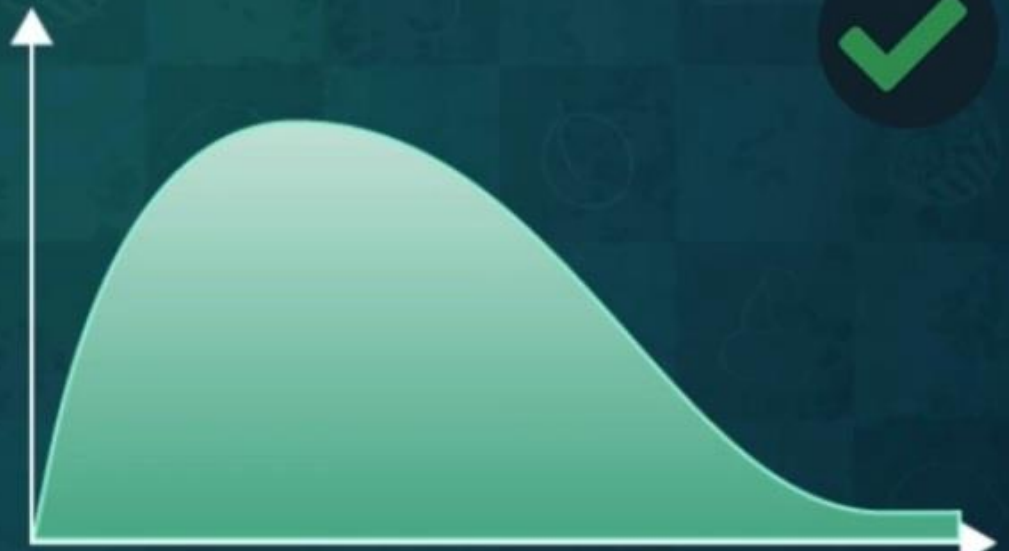
مدرس: محمد فزونی
عضو هیات علمی دانشگاه گنبدکاووس

پائیز ۱۳۹۹

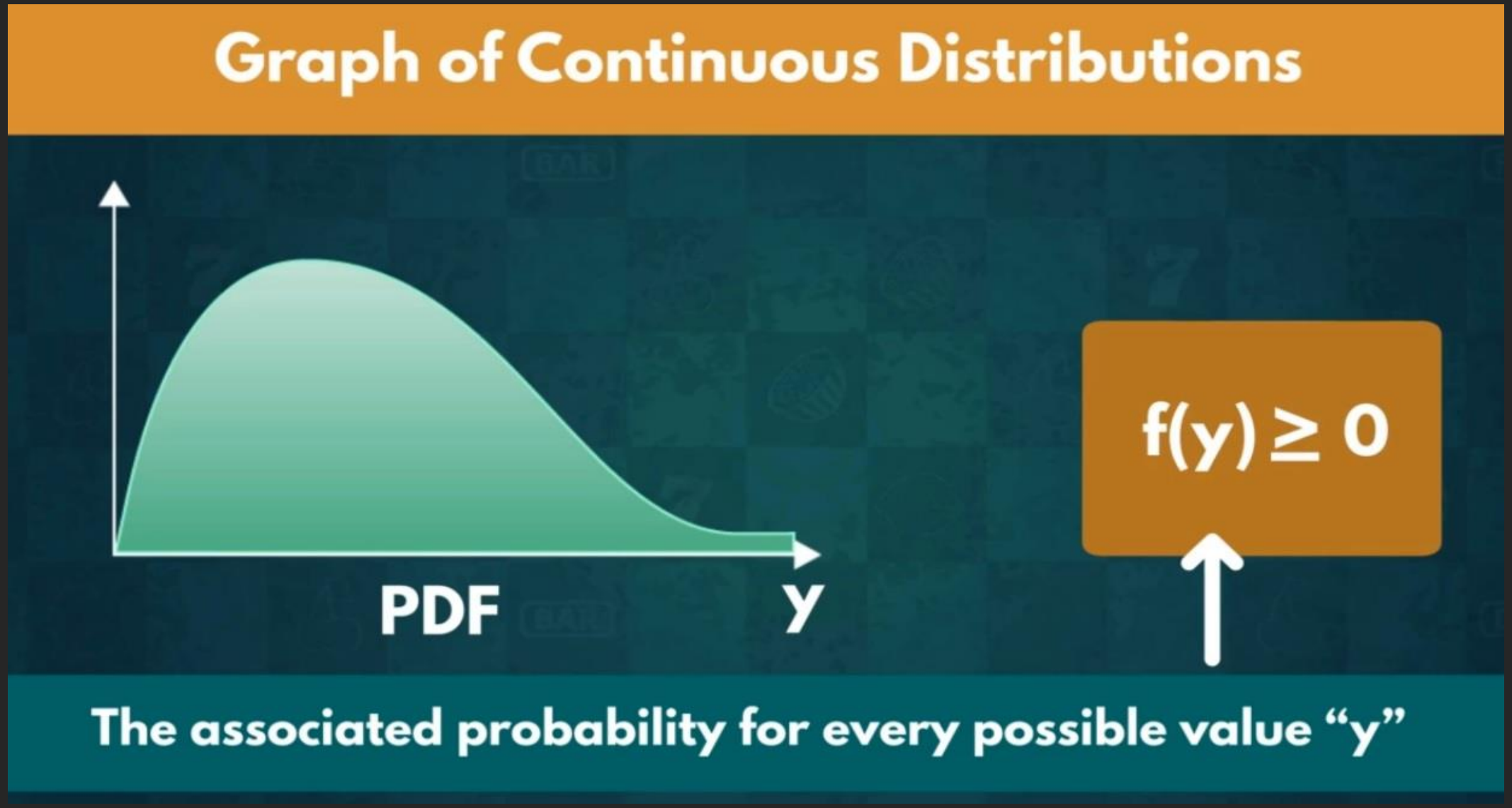
Continuous Distributions

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

Y	P(y)



PDF stands for "*Probability Density Function*"



Discrete vs Continuous

The probability for any individual value equal to 0

$$P(X) = 0$$

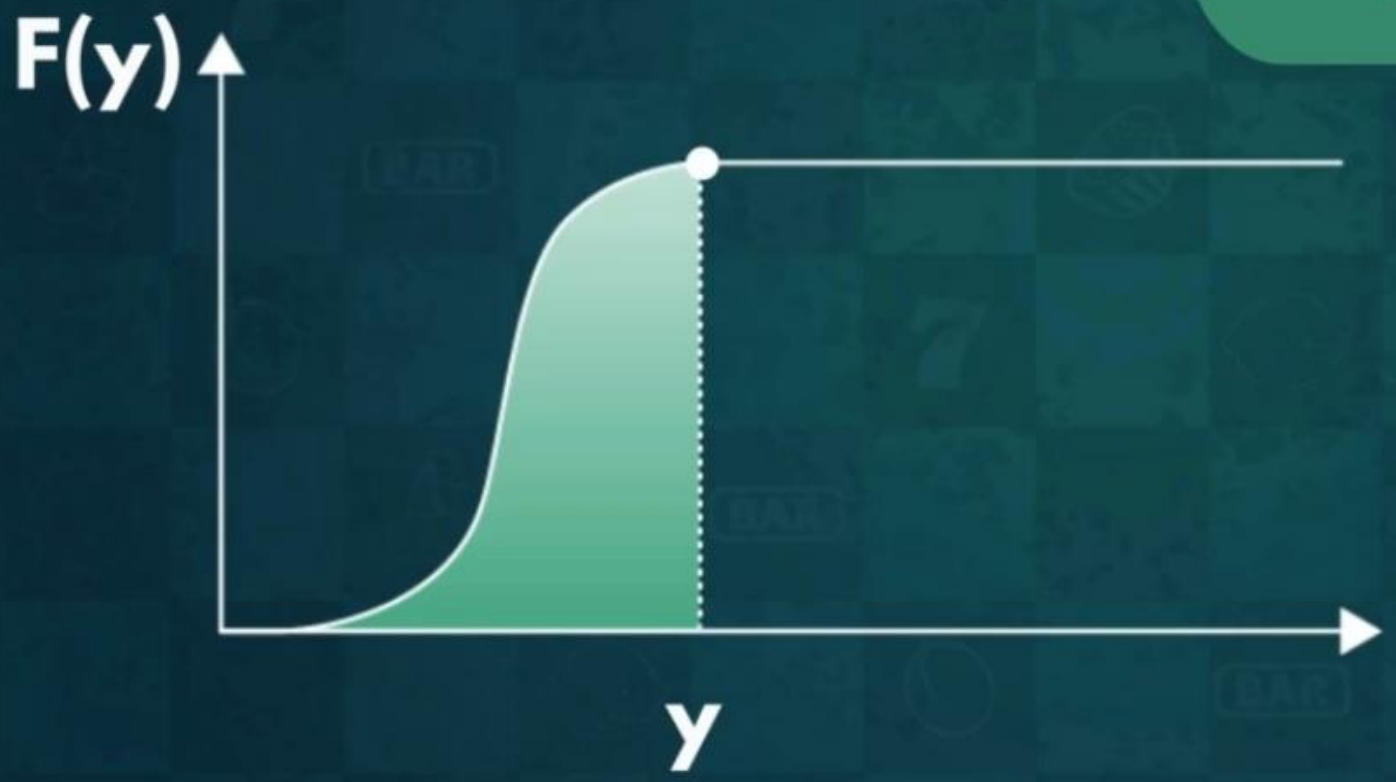
$$P(x > X) = P(x \geq X)$$

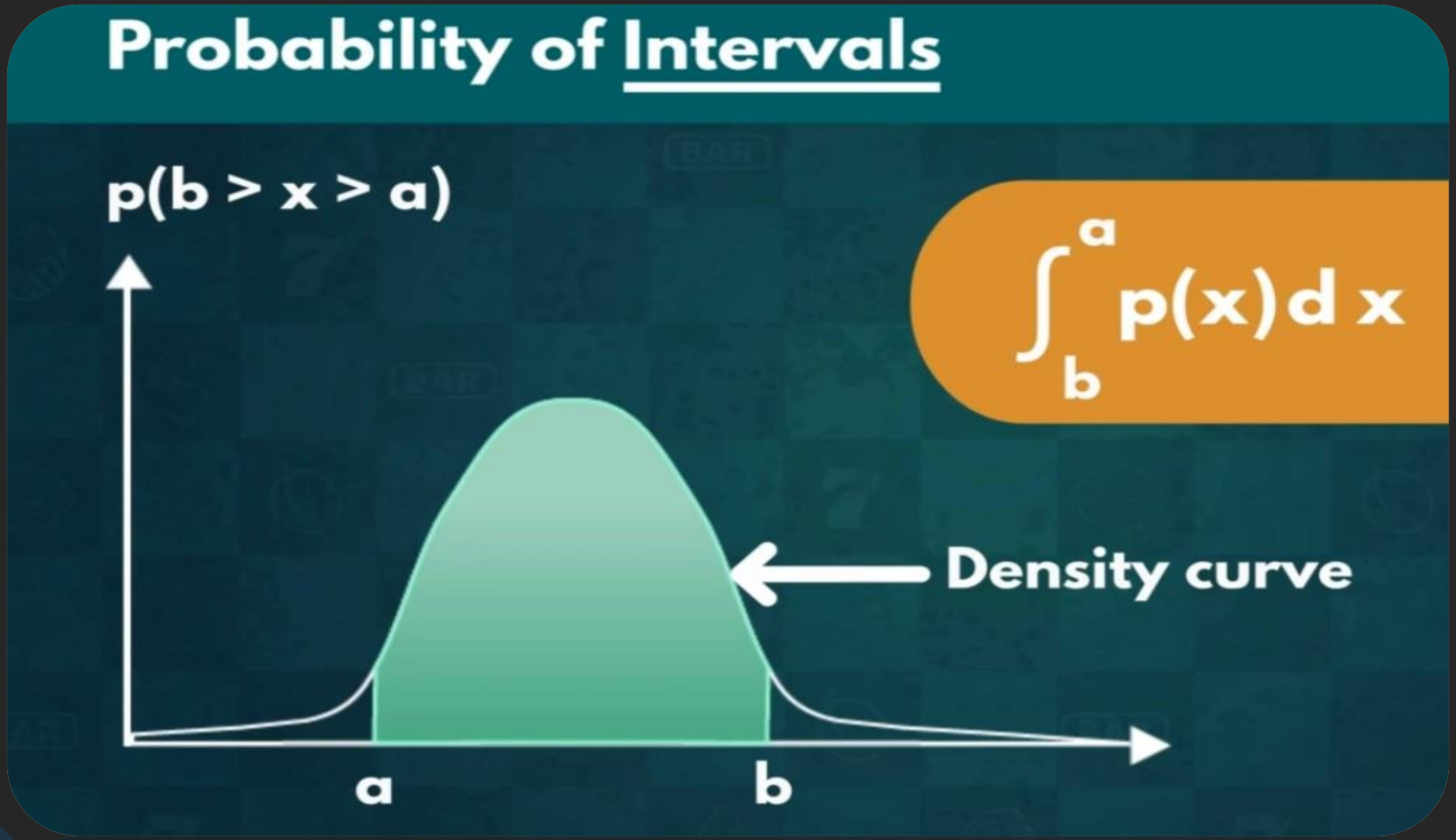
$$P(x < 6 \text{ min}) = P(x \leq 6 \text{ min})$$

$$P(x = 6) = 0$$

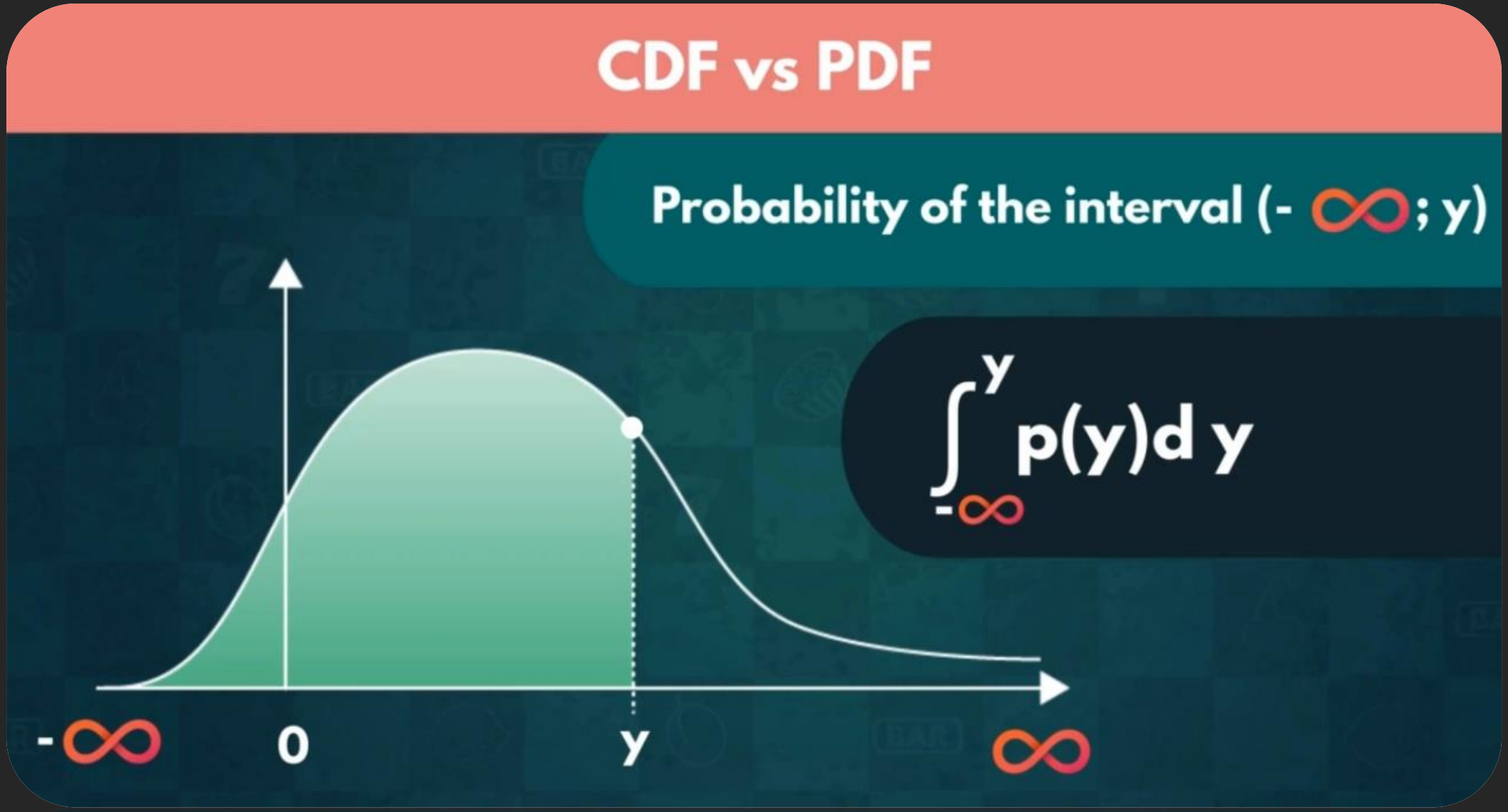
Cumulative Distribution Function (CDF)

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




Probability of Intervals
p(b > x > a)
$\int_b^a p(x)\,dx$
Density curve
a
b

CDF vs PDF



CDF vs PDF

PDF —Integral→ CDF

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

PDF —Derivative→ CDF

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

E(y)

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

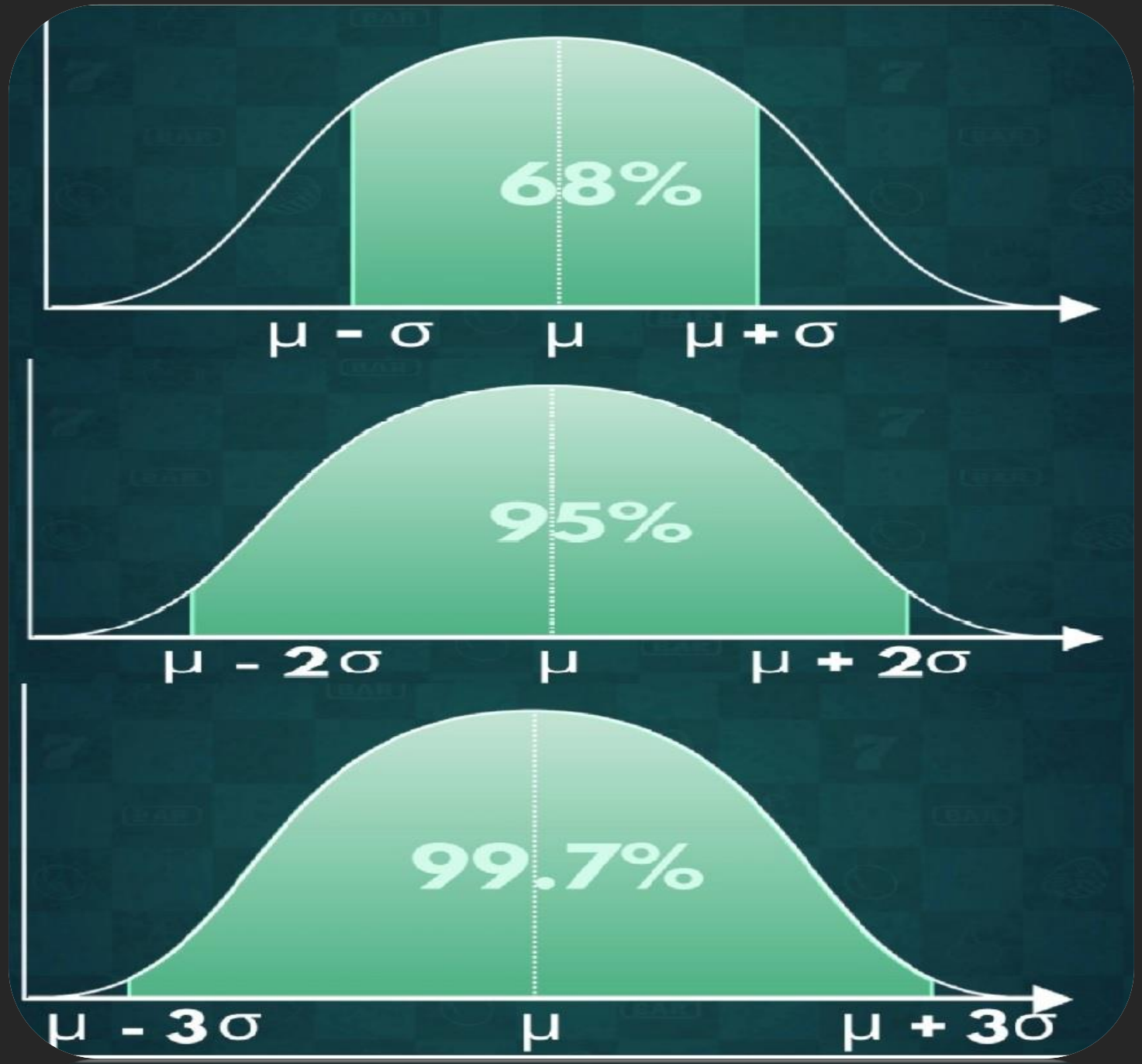
$$E(y) = \int_{-\infty}^{\infty} y\, p(y)\, dy$$

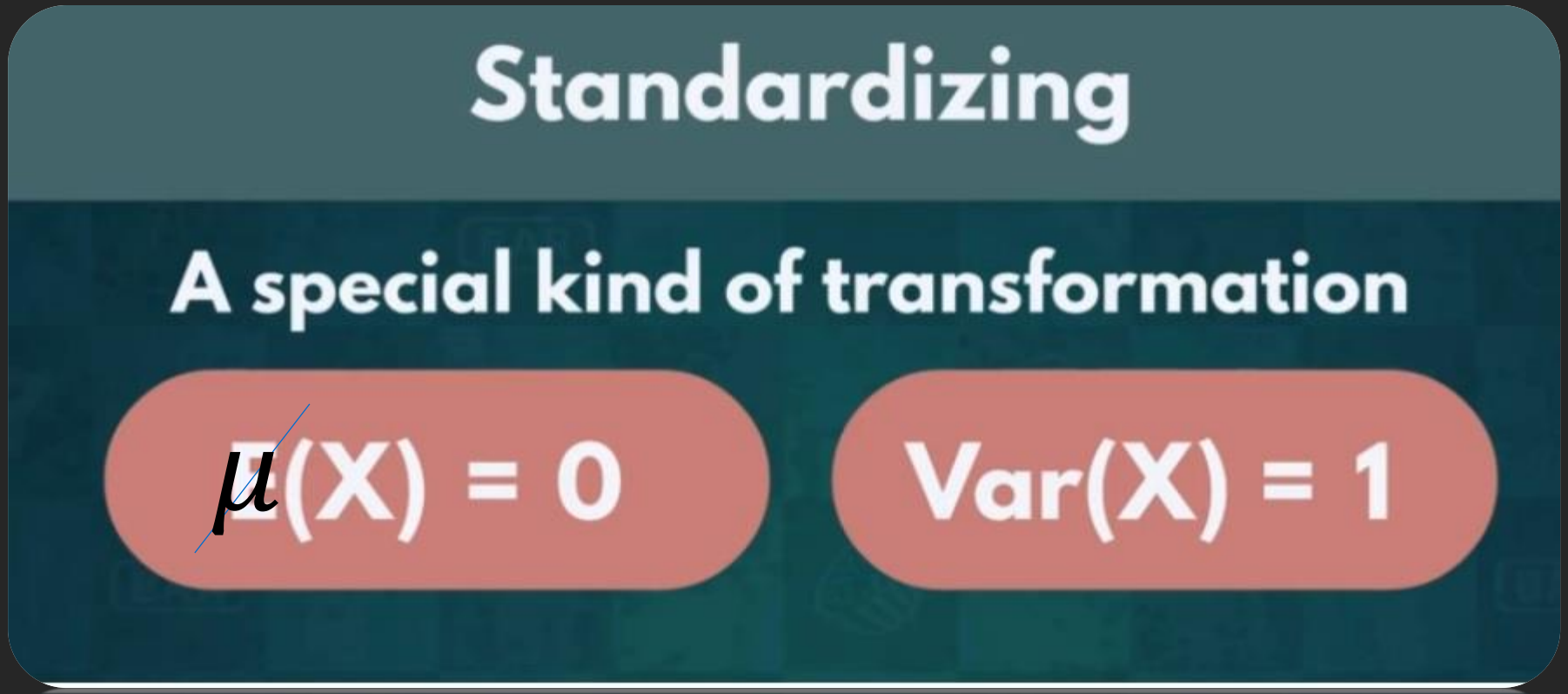
Expected Value and Variance

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

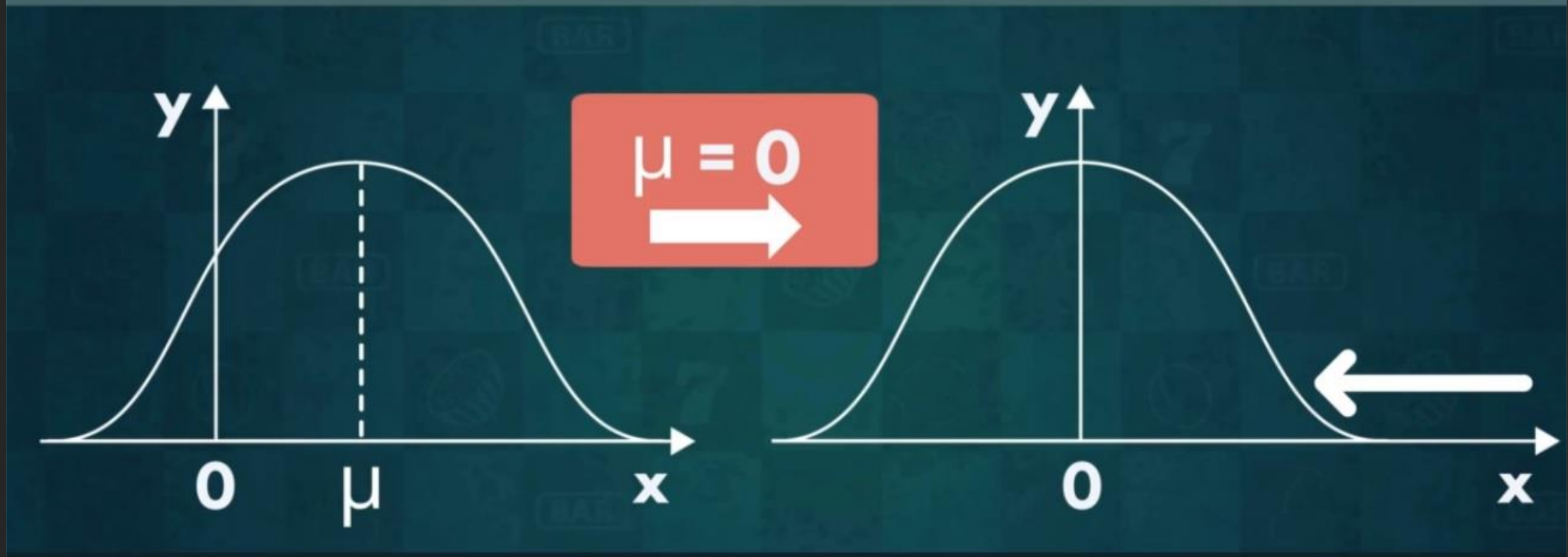
$$E(y) = \int_{-\infty}^{\infty} y\, p(y)\, dy$$

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


68%
μ - σ
μ
μ + σ
95%
μ - 2σ
μ
μ + 2σ
99.7%
μ - 3σ
μ
μ + 3σ


Standardizing
A special kind of transformation
μ(X) = 0
Var(X) = 1

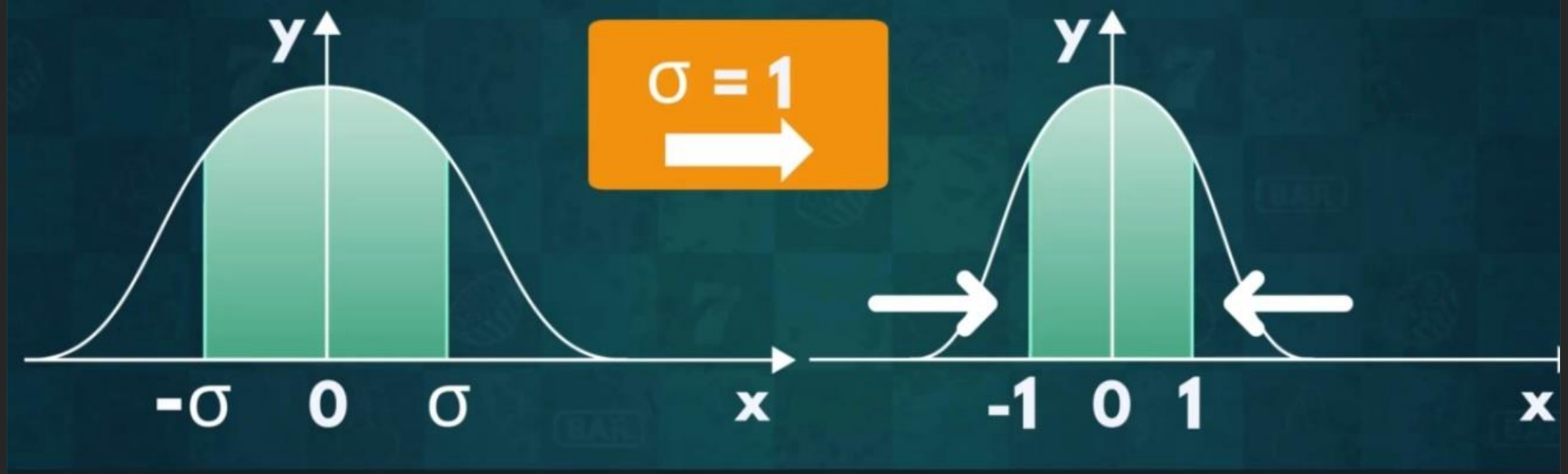
Standardizing



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

$$\lambda = t(x) \qquad \lambda = t(x - h)$$

Standardizing



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

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

$$\lambda = t(x - h)$$

$$\lambda = t\left(\frac{\alpha}{x - h}\right)$$

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**

Applications *of chi-squared*

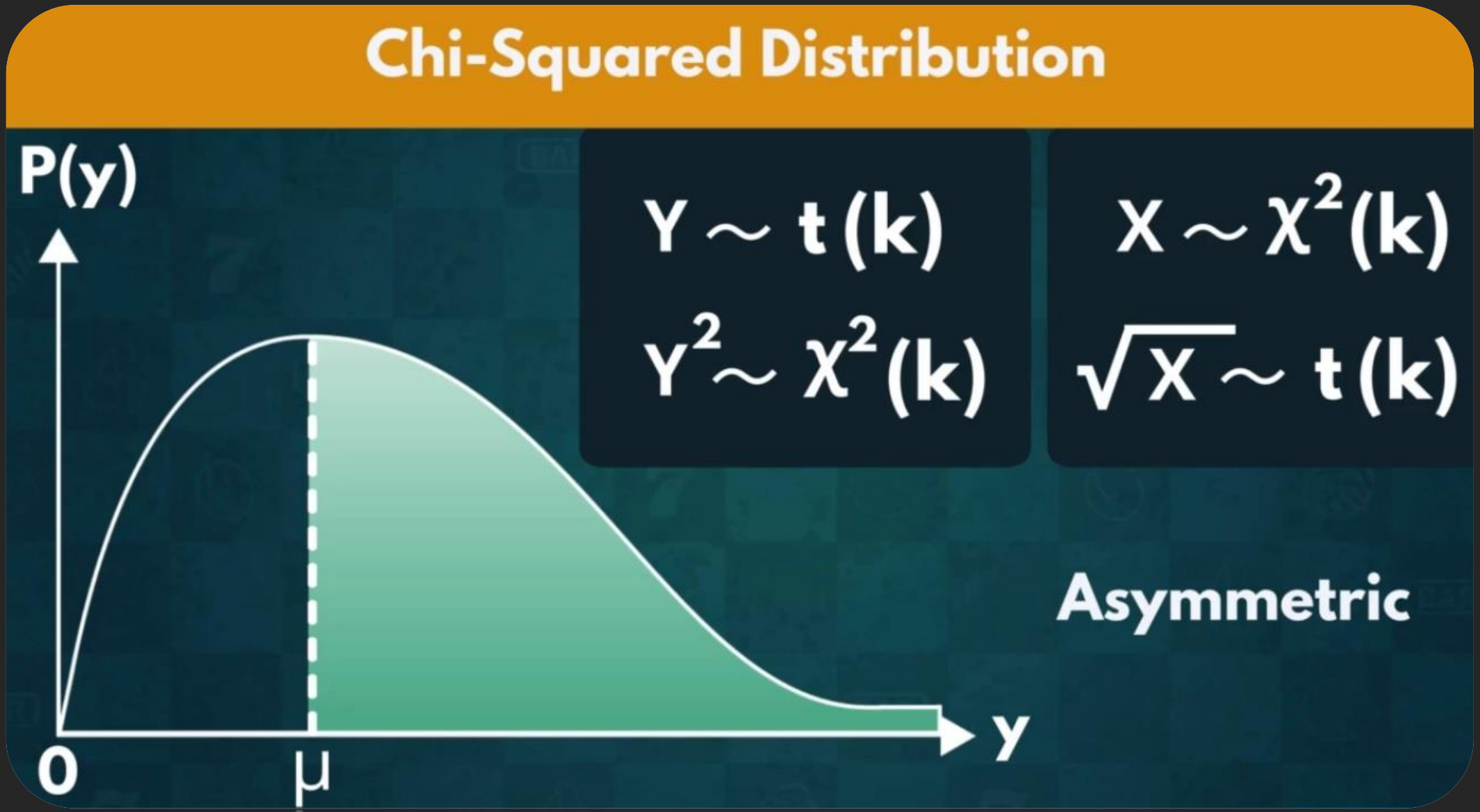
Few events in real life

Statistical analysis

- Hypothesis testing
- Computing confidence intervals

Goodness of fit

Chi-Squared Distribution

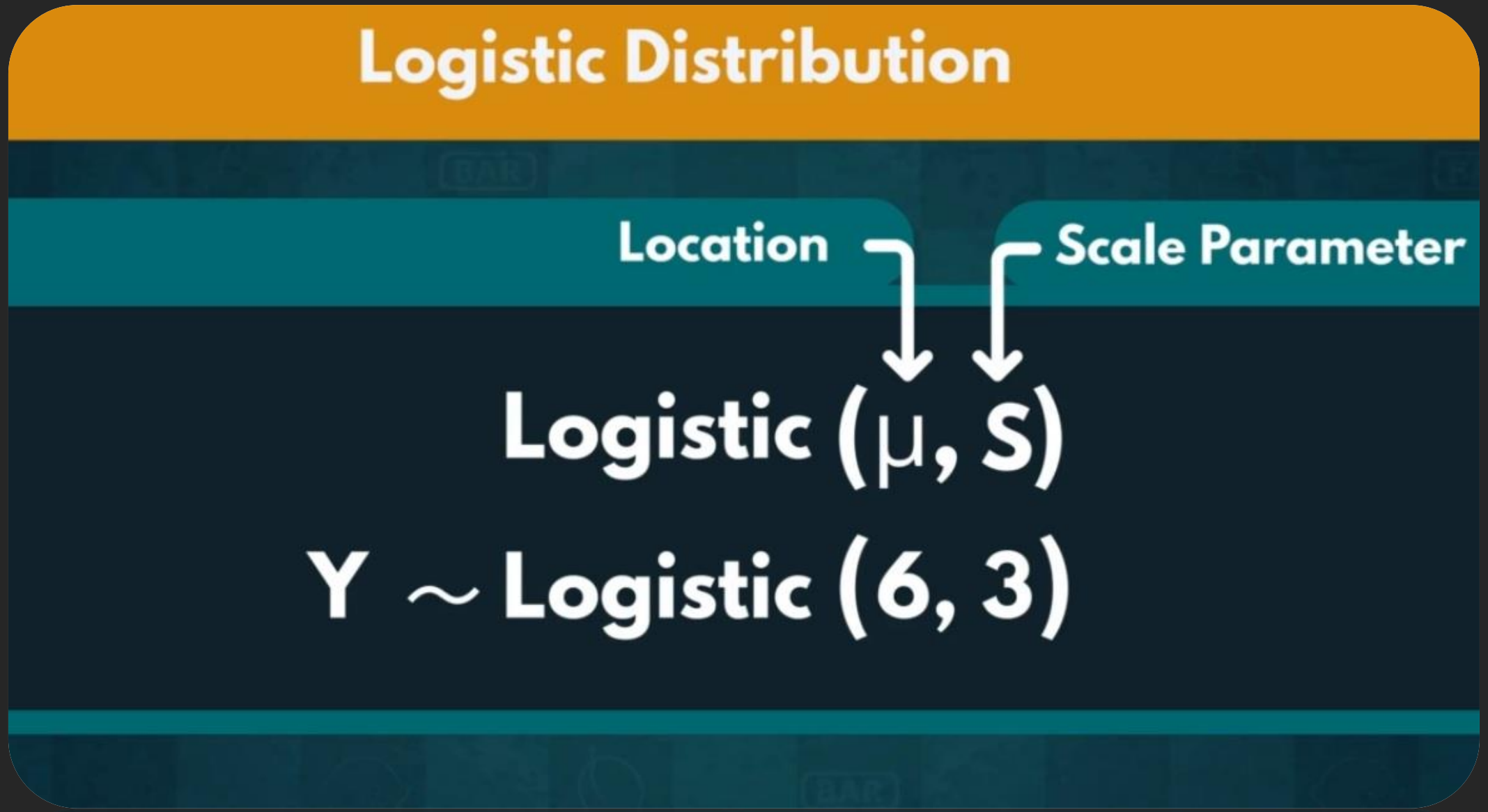


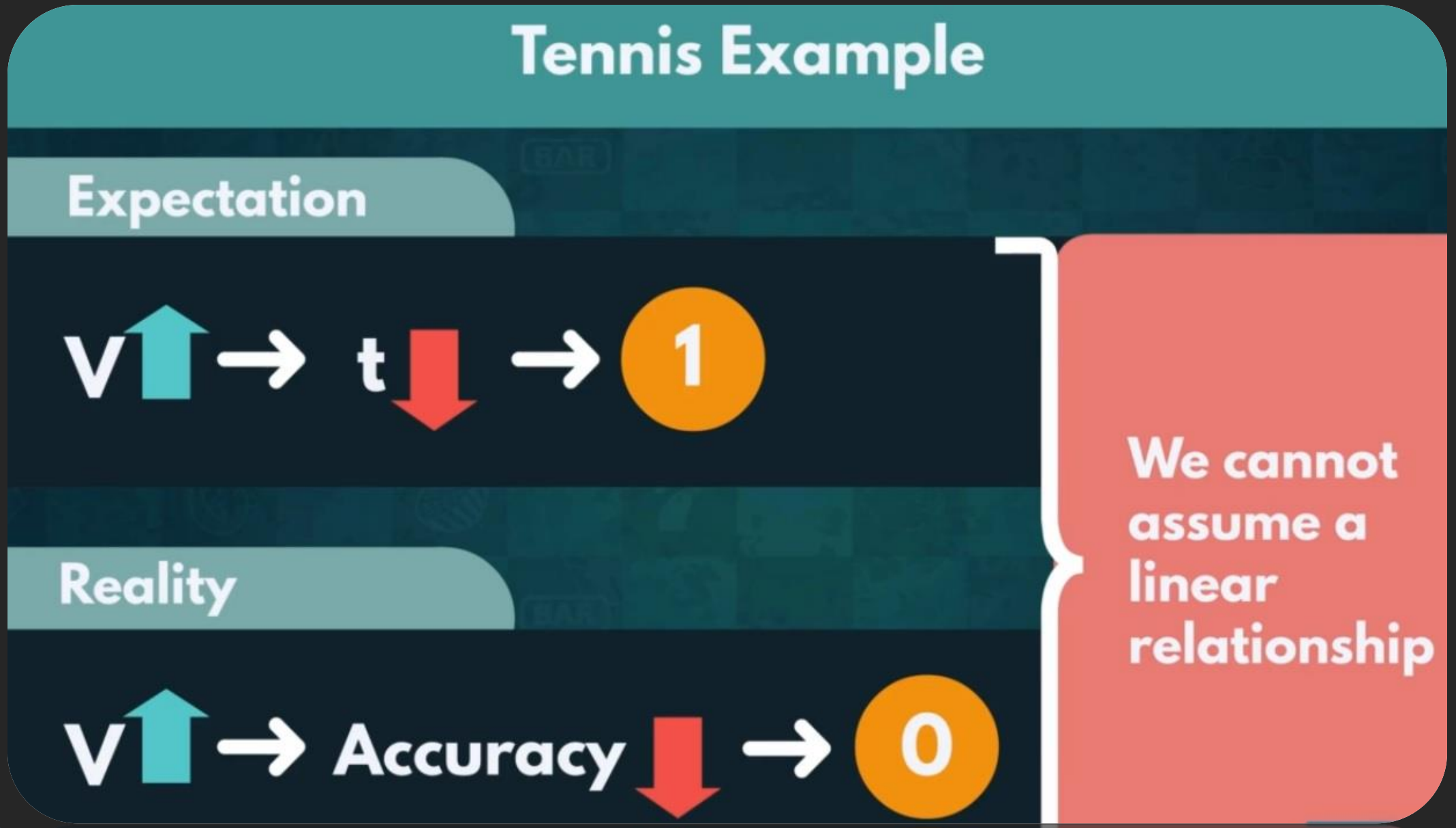
Transformation

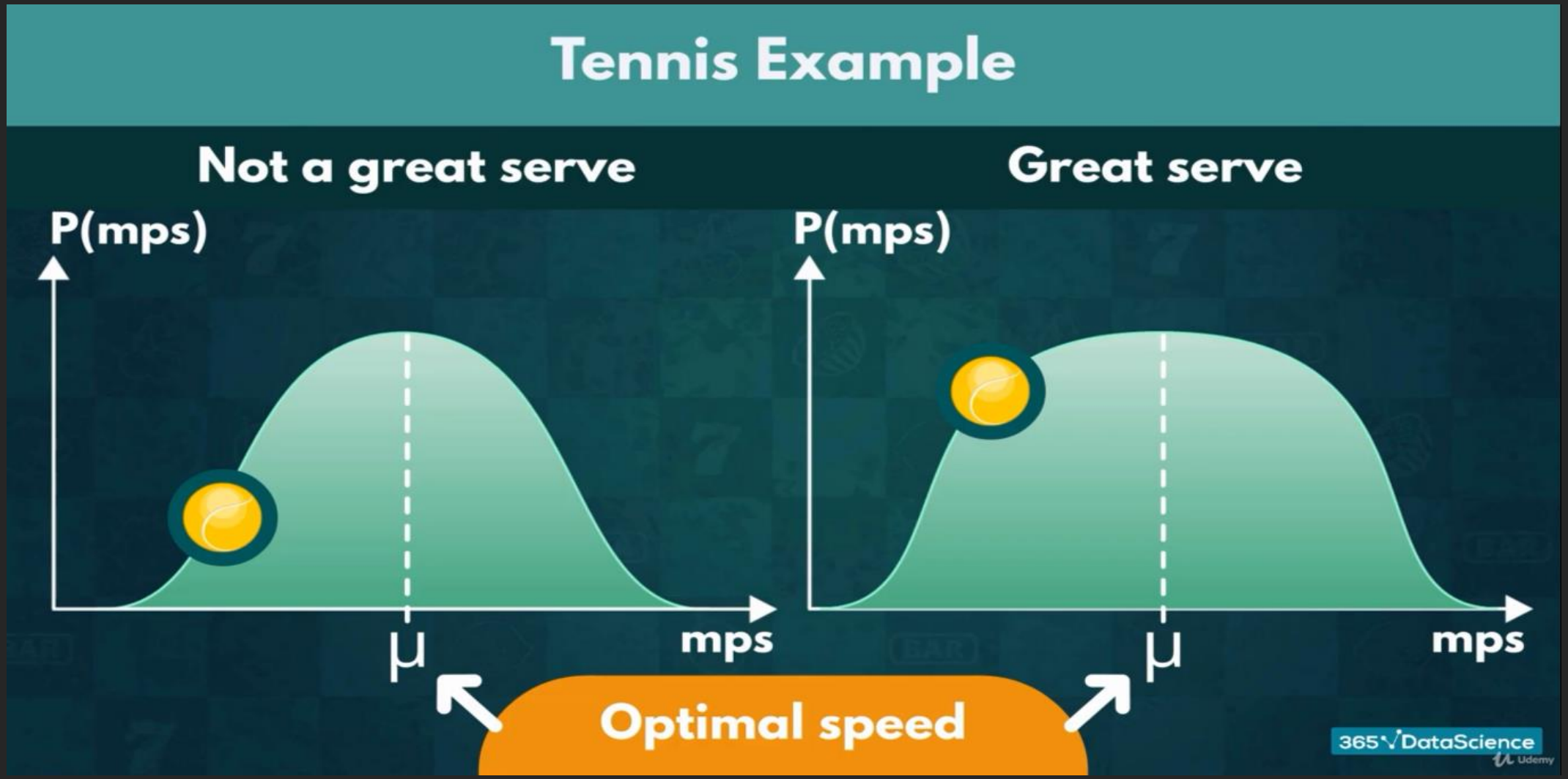
$$\left.\begin{array}{l} Y \sim Exp(\lambda) \\ X = \ln(Y) \end{array}\right\} \Rightarrow X \sim N(\mu, \sigma^2)$$

- One of the most common transformations

Logistic Distribution




Tennis Example
Expectation
V → t → 1
Reality
V → Accuracy → 0
We cannot assume a linear relationship


Tennis Example
Not a great serve
P(mps)
μ
mps
Great serve
P(mps)
μ
mps
Optimal speed
365 DataScience
Udemy

SUBJECT OF THE NEXT VIDEO

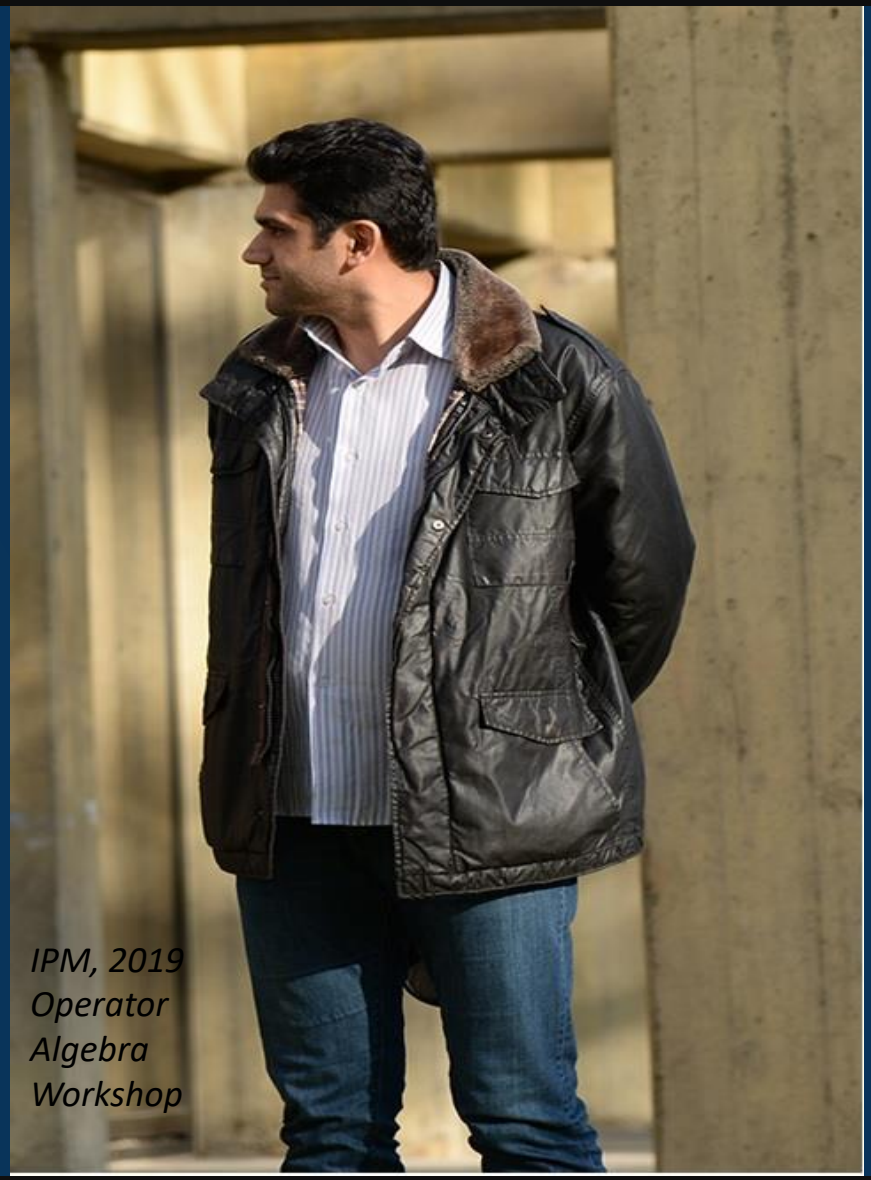
Conditionals in Python

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