

DISTRIBUTIONS

CLASS
101



HOW IS DATA
SPREAD OUT?



FABIAN WERKMEISTER

WHAT IS THAT?

A distribution tries to show how different things or events are grouped or spread out



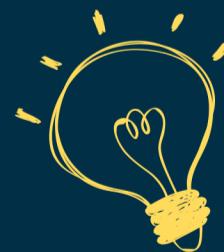
It is a representation of occurrences of something specific in the real world



REAL WORLD EXAMPLE



Let's take **football** data
as an example to learn
what distributions are



WE WILL LOOK AT SOME OF THE
MOST COMMON DISTRIBUTIONS



NORMAL

The normal distribution is often found in many natural occurring variables

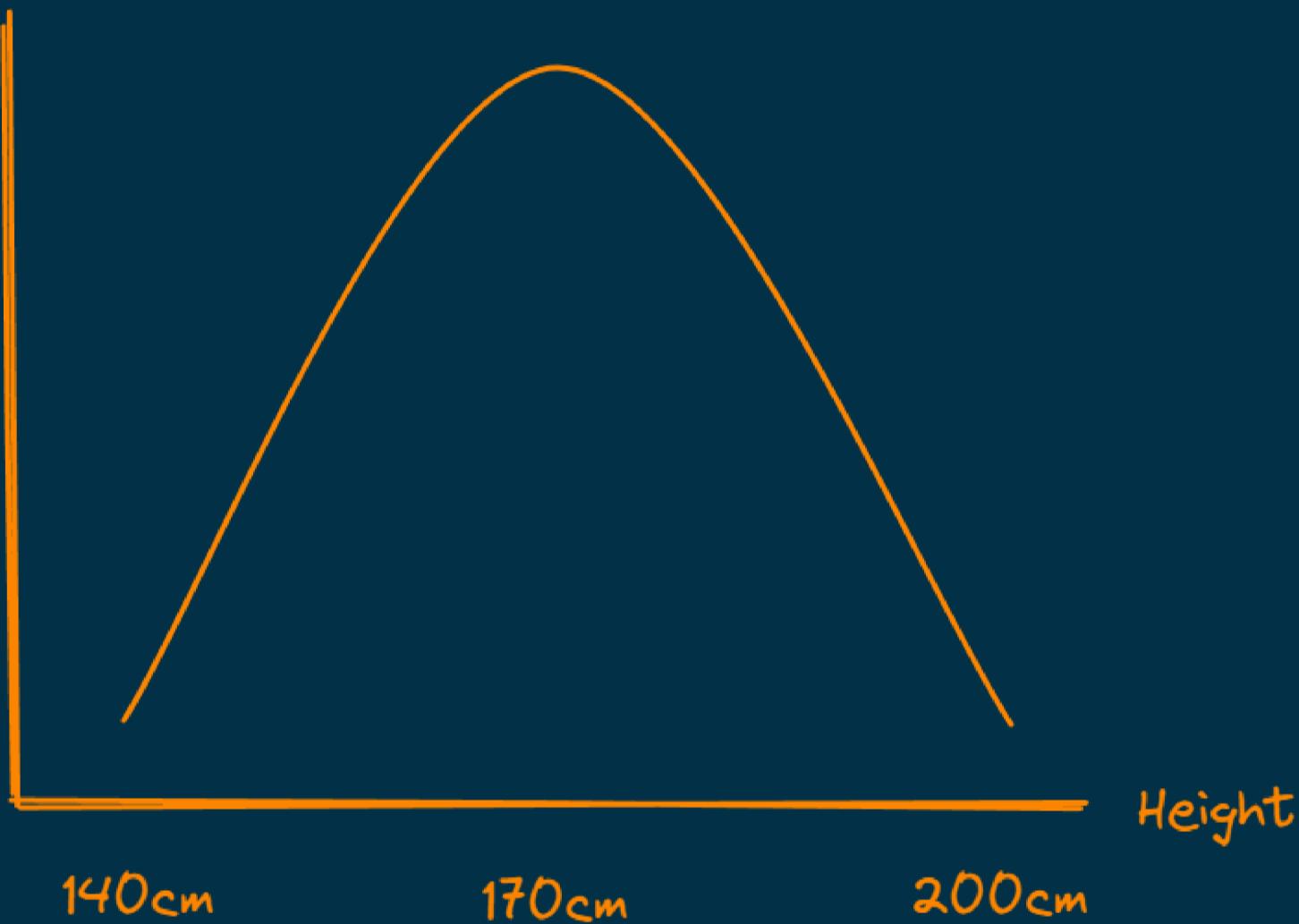
It has an average and is symmetrical, because data is equally likely to be below or above the average

This gives it a bell-shape, which is why it is often called "bell-curve"



NORMAL

Occurence



Height of different players

There are more players around **average height (1,70m)** and less players, who are either very small (**1,40m**) or very tall (**2,00m**)



FABIAN WERKMEISTER

BERNOULLI

The bernoulli distribution describes an event, which is only done once & has only two possible outcomes

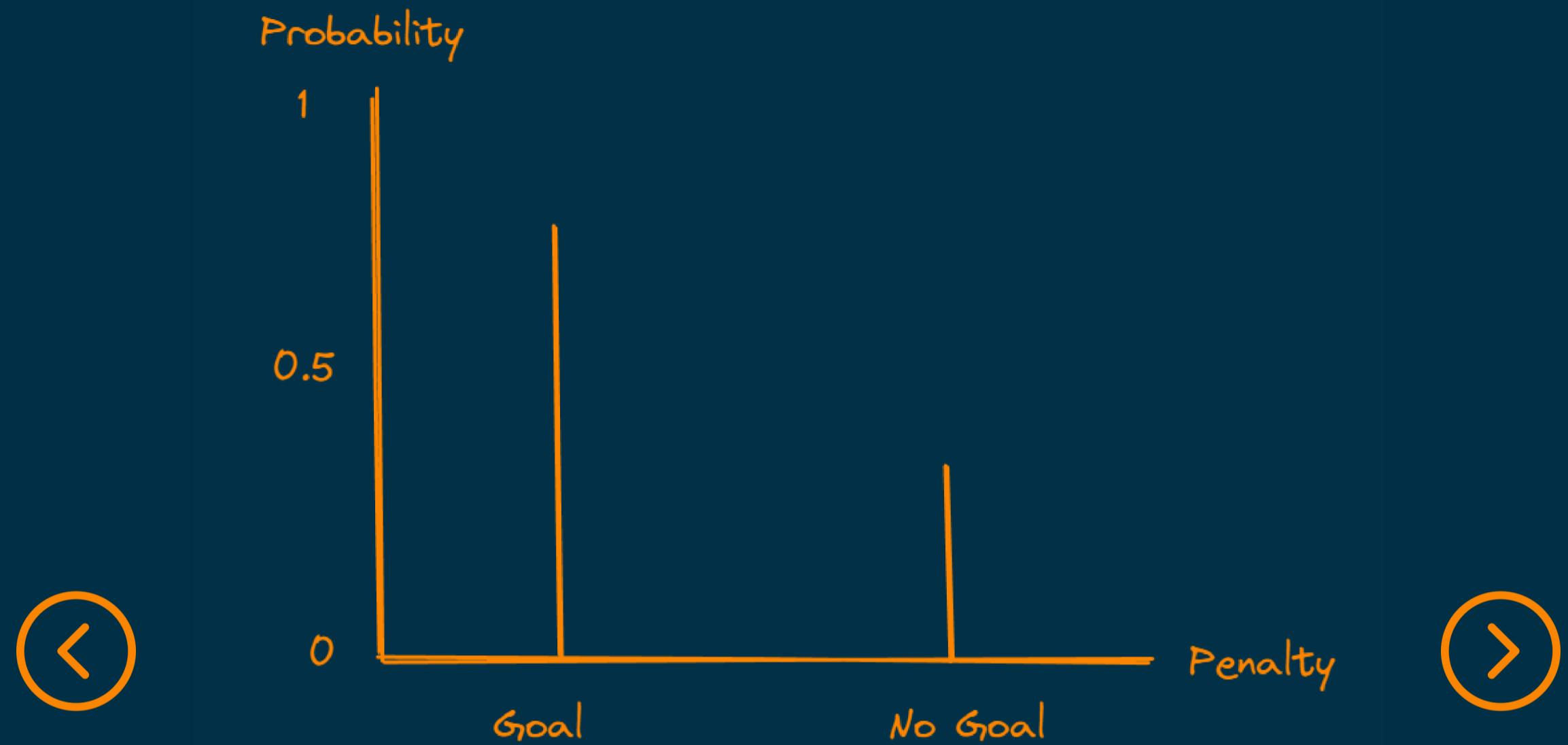


→ success or failure



Each outcome has a specific probability of occurring

BERNOULLI



Football penalty shot

Only 2 possible scenarios with a specific probability

- Goal with probability of 70%
- No goal with probability of 30%



BINOMIAL

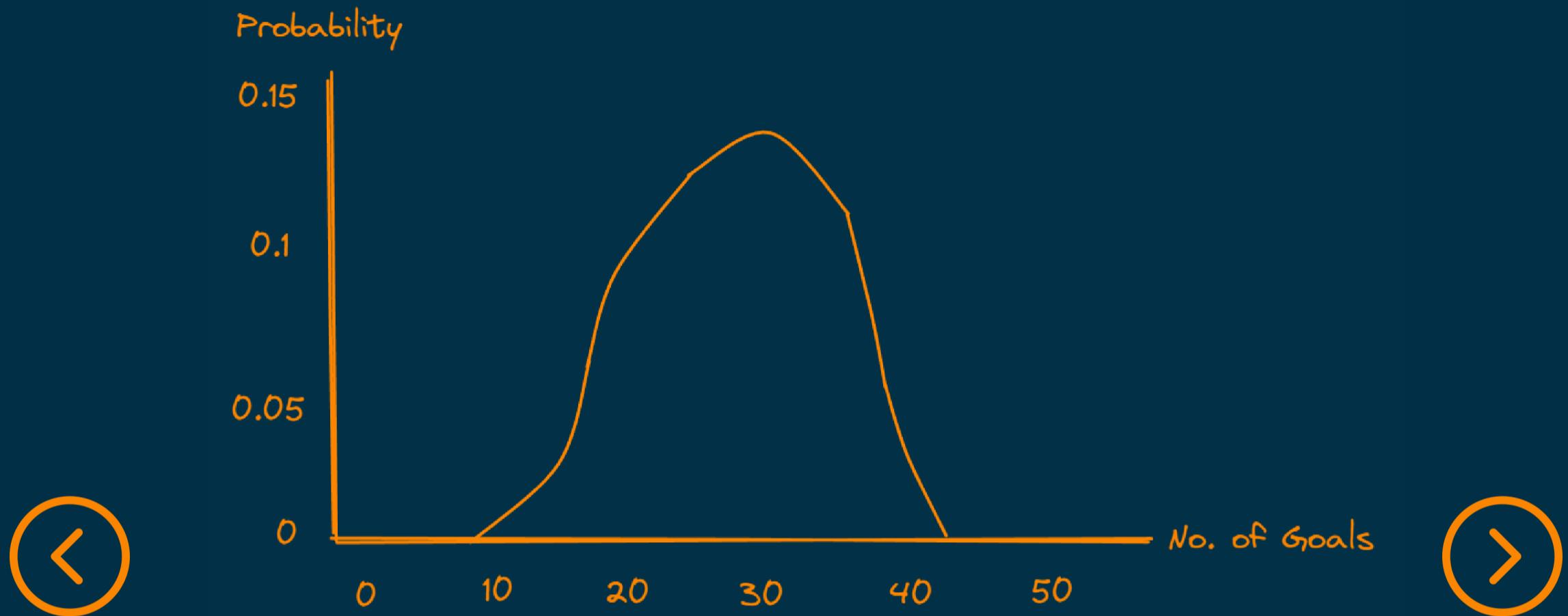
The binomial distribution is like the bigger brother of the bernoulli distribution

It also has only two possible outcomes with a specific probability

But it models multiple repetitions of the same event



BINOMIAL



Taking a penalty 50 times

It is more likely to score a goal around 30 to 35 times and almost impossible to never (0) score or score all the time (50)

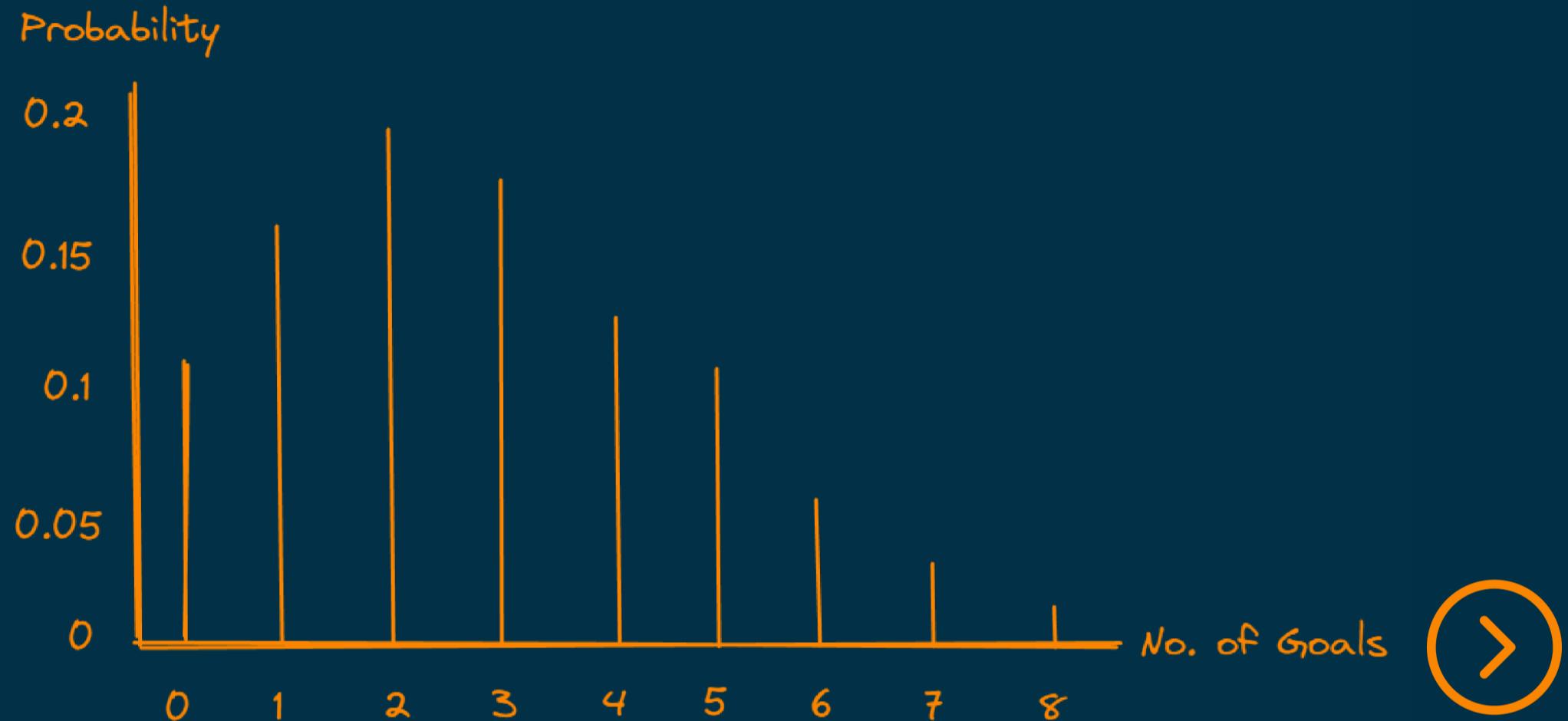
POISSON

The poisson distribution is often used to count the occurrence of events in a fixed time frame



It assumes that the number of events is **discrete** (full numbers like 3 and not 3.25) and that the events happen randomly and independent of each other

POISSON



Goals in a 90min game

It is likely that some goals will be scored, but it gets increasingly unlikely that more than 2 or 3 goals are scored in one single game

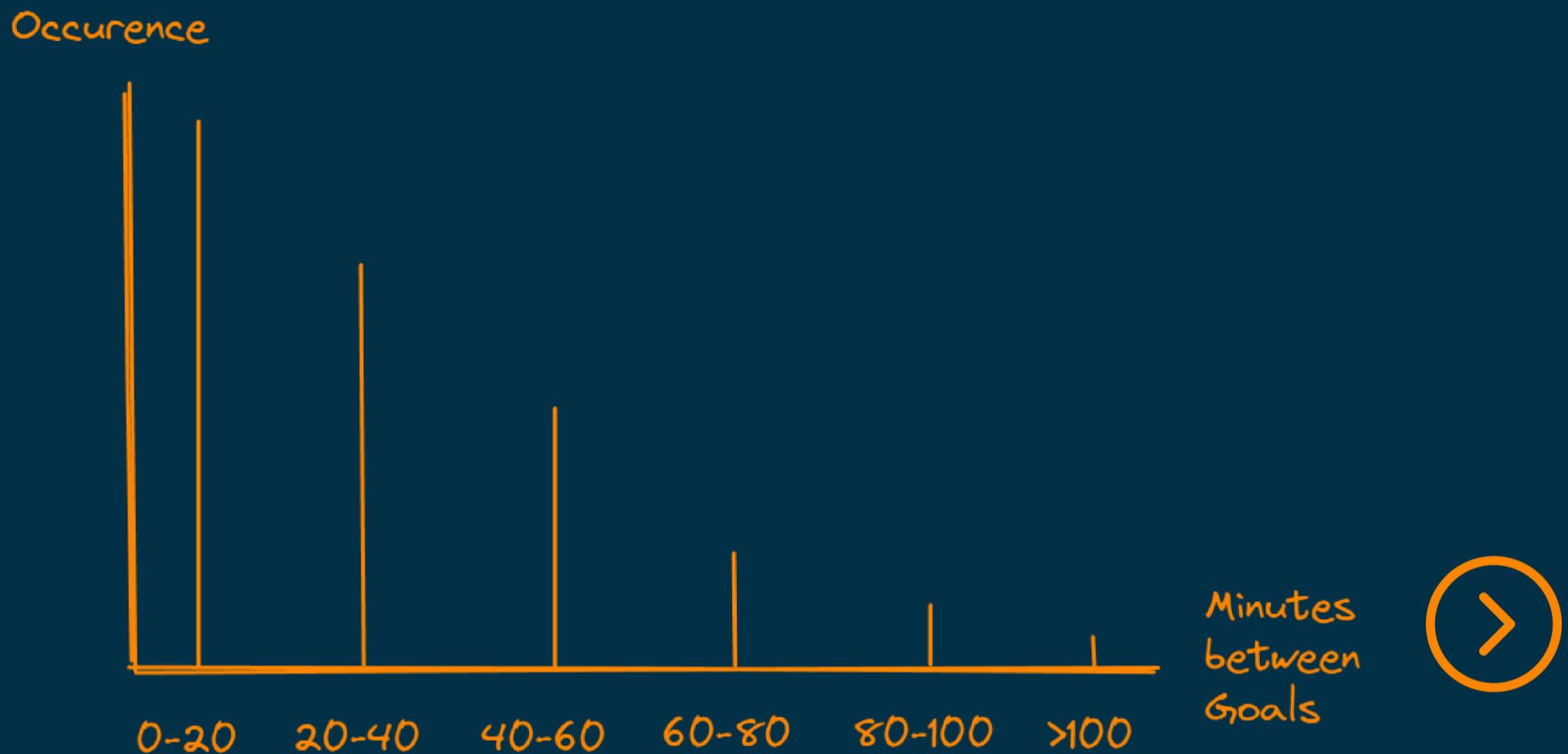
EXPONENTIAL

The exponential distribution is similar to the poisson distribution

 **The poisson distribution expresses the number of events in a time frame, whereas the exponential distribution describes the time between the events** 



EXPONENTIAL



Minutes between goals

If we look at several games it is very likely that the next goal is scored in the next 40 to 60 minutes

→ it is very rare or unlikely, that no goals are scored (by any team) for a longer time period



TAKEAWAY

**Any series of data can
follow a specific distribution
to some degree**

**Looking at the distribution
of your data is most often
the first step of any analysis
or machine learning
workflow you want to do**

**So be sure to know your
distributions well!**



DO YOU MISS ANY OTHER DISTRIBUTIONS?



SEE YOU IN
THE COMMENTS



FABIAN WERKMEISTER

HI, I'M FABIAN



**I HELP YOU
TO BECOME
FLUENT IN DATA**



IF YOU LIKE THIS POST

+ Follow

& hit the 



FABIAN WERKMEISTER

PS: I HAVE A FREE DATA NOTION SPACE FOR YOU

- ✓ Learning Resources
- ✓ Data Projects & Tools



Fabian Werkmeister

I help you to become fluent in Data and enable you to translate between business & tech ★ Data Scientist @FUNKE ★ Python Data Science Data Engineering

Talks about #data, #python, #mindset, #datascience, and #productivity

Hamburg, Hamburg, Germany · [Contact info](#)

[FREE Data Notion Space](#)



FABIAN WERKMEISTER