Today's agenda

Before ("interview") Wrapping up Prob (Theory) → important distribution

After (hands-on) Stat (next class) → some dist in stats next week

Bernoulli
Binomial
Uniform
Gaussian / Normal

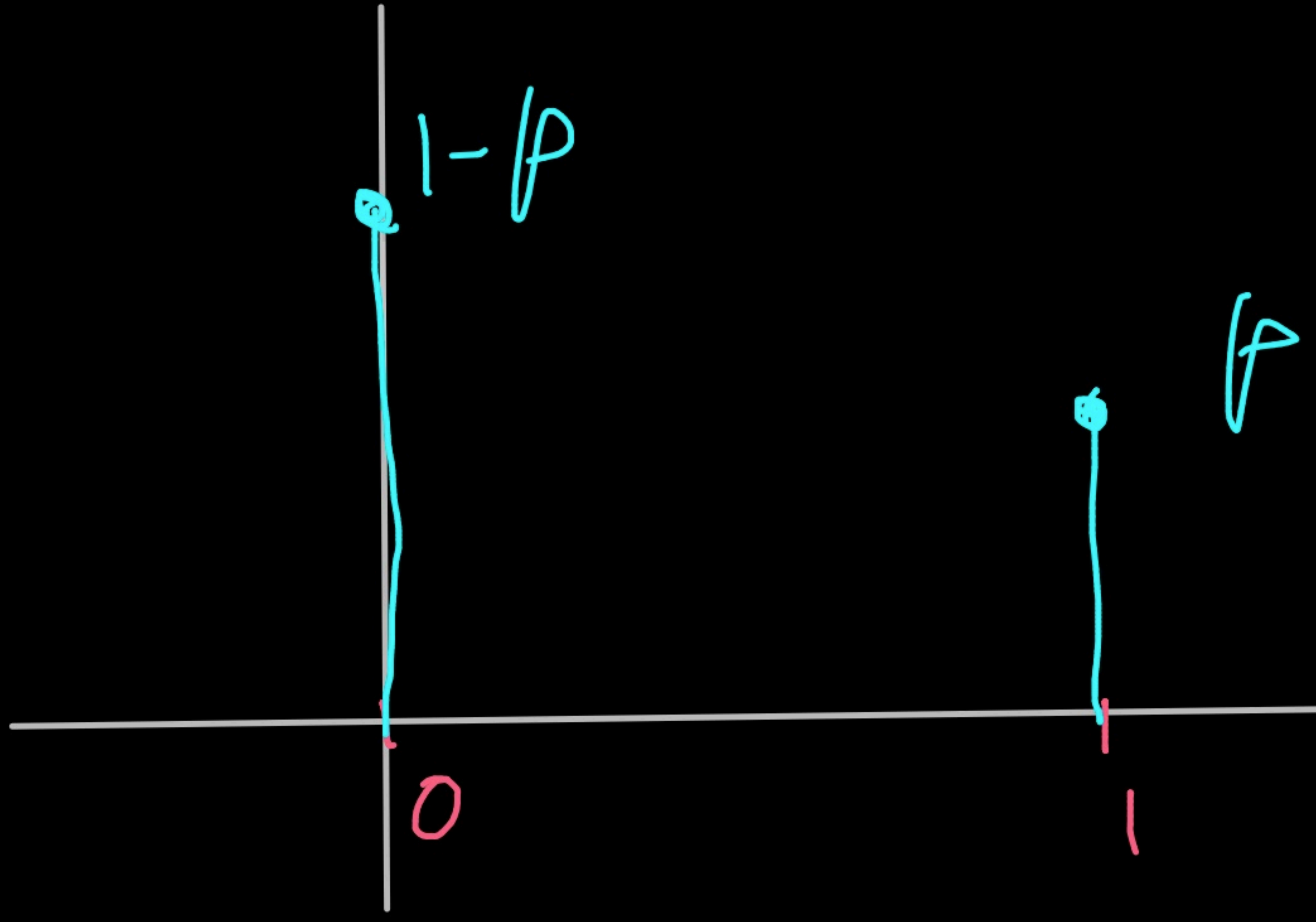
→ in context
Scenario-based

Geometric distribution → How many rejects before 1st offer

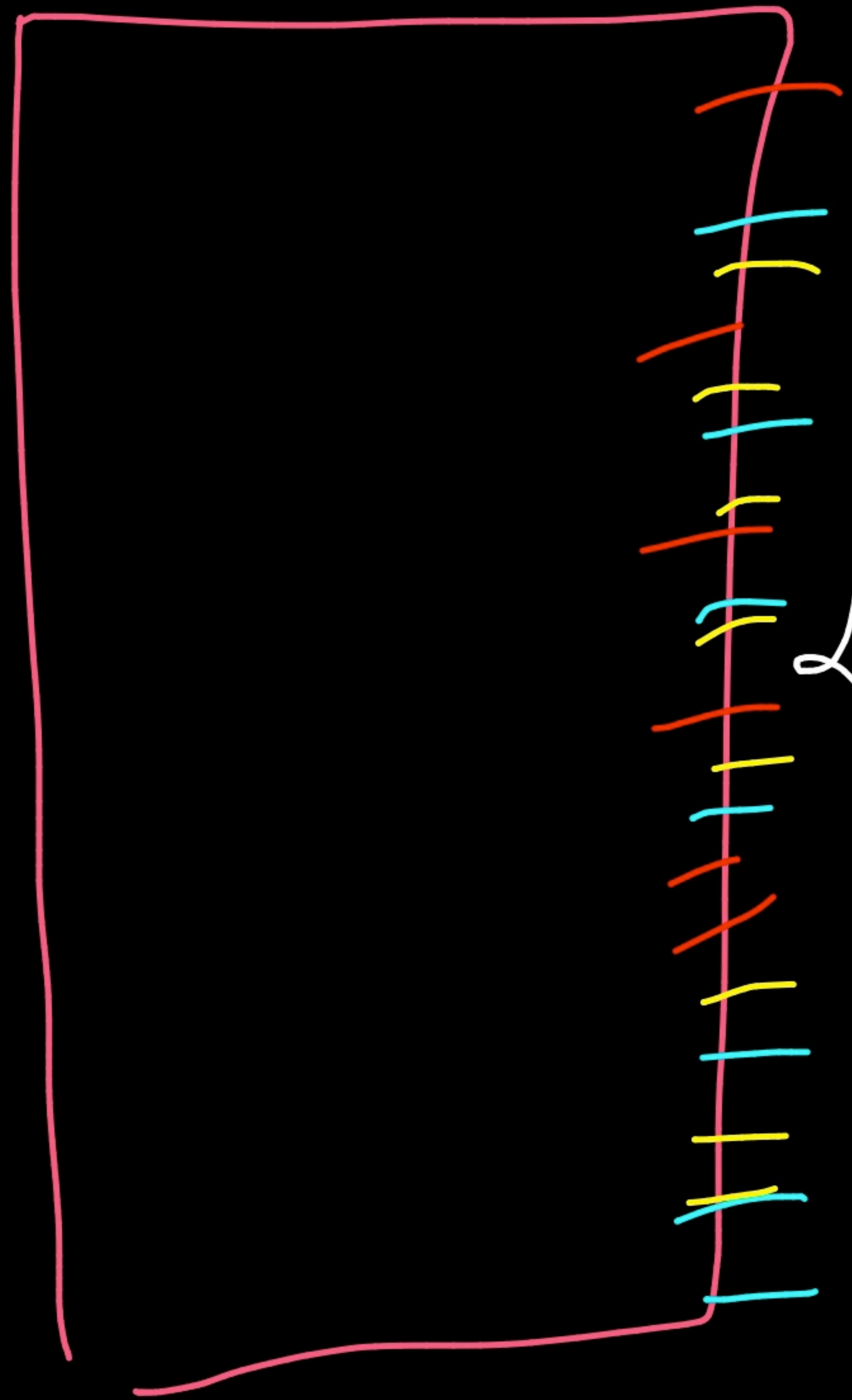
→ Poisson & Exponential → football / customers entering

Log-Normal → waiting time / Swiggy / Zomato

Bernoulli



frac of people who clears



30 people Anurag
30 people Aditya
30 people Ajay

3 chard
5 chard
4 choc
2
1
7
6
2
4

Counts[3] = 1
Counts[2] = 2
Counts[4] = 2
Counts[15] = 0

1.8 Lakhs

On average, how many interviews are to be taken till we find the first accept?

Let X : # interviews till first accept p : prob of clearing

$$P[X = 7] = (1-p)^6 \, p$$

1st guy rejected : $1-p$

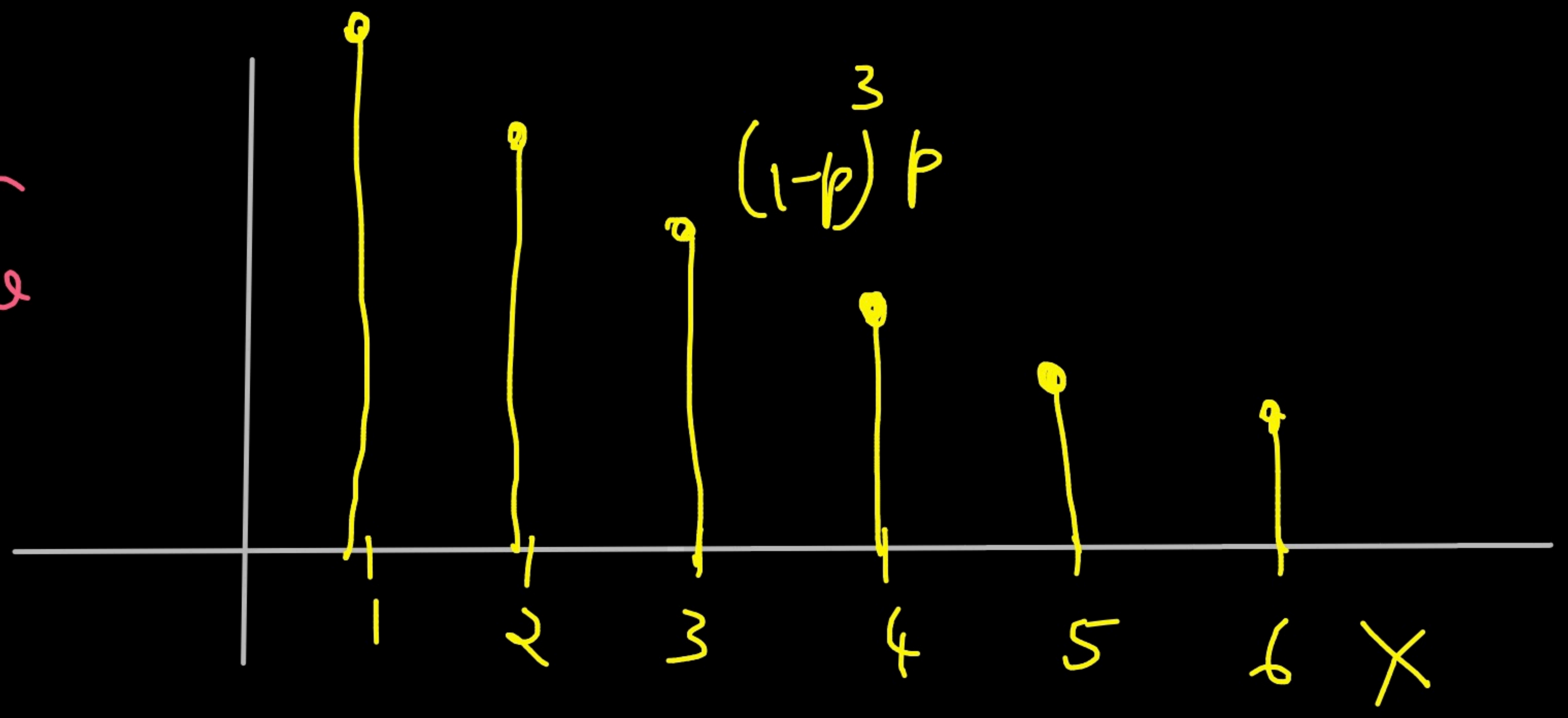
2nd guy rejected : $1-p$

⋮

6th guy rejected : $1-p$

7th guy accepted : p

Geometric Random variable



$$P\{X=4\} = (1-p)^3 p$$

$$P\{X=5\} = (1-p)^4 p$$

$$P[X=k] = (1-p)^{k-1} p$$


Jan 1
1
2
3
4
0
0
0
1
0
0
0
4th
Jan 2
1
2
3
7
0
0
0
0
0
0
1
0
7th
[4, 7, - - - -]

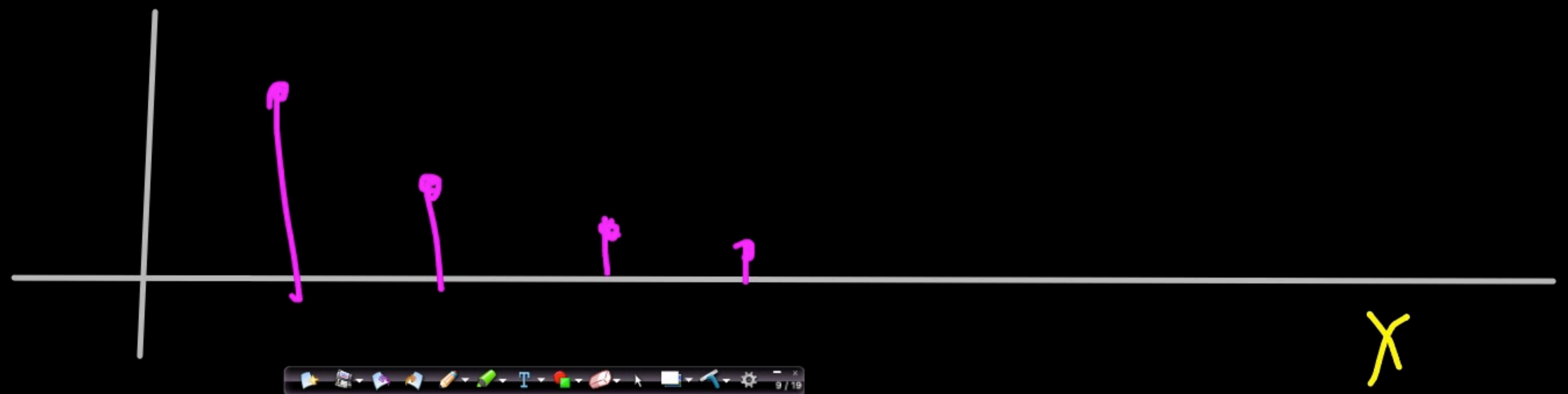
Coin toss

Prob heads is "p"

Toss till you get first heads

$$S = \left\{ \begin{array}{cccc} X=1 & X=2 & X=3 & \\ H, & TH, & TTH, & TTTH \;\cdots \\ p & (1-p)p & (1-p)^2 p & (1-p)^3 p \end{array} \right\}$$

$\boxed{10:20}$



Poisson "Counting" → discrete

Eg: 1) Football scores

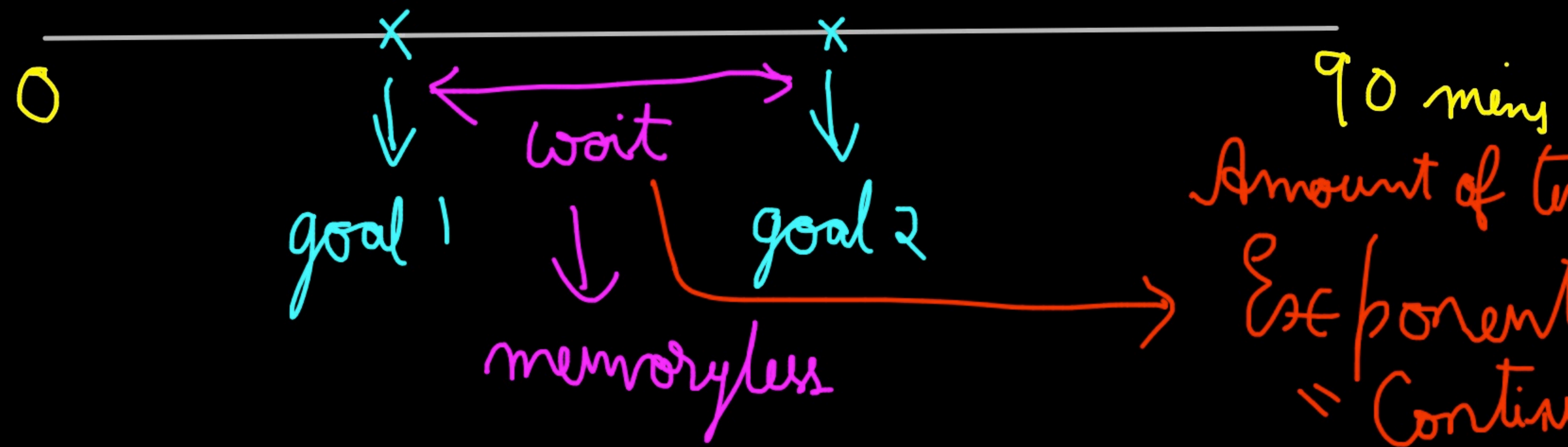
2) Customers entering a store

3) Call centre → no of phone

4) Manufacturing → no. defective

Football

Goal → event : Count no. of "events"



Amount of time we wait

Exponential "λ"

"Continuous"

"Wait for the next goal" → independent of when the previous goal happened

If any "counting" process satisfies the above, we call "Poisson RV"

\# "events"

Waiting for next goal → independent of when previous goal happened

no. events → Poisson distribution

$$P[X = k] = \frac{\lambda^{k} e^{-\lambda}}{k!}$$

mean : λ

When we wait for a goal : Parameter λ

- Discrete : counting # of goals in 90 min
  "Poisson" mean $= \lambda$
- Continuous: Time b/w goals
  "Exponential" mean $= \frac{1}{\lambda}$

The past does <u>not</u> impact how much you have to wait

2 weeks

→ Gaussian — "most important"

→ If data is not Gaussian,
what is the best we can do
make it as close to Gaussian

log / Box-Cox

→ If looks bell-shaped may not be enough

→ Confirm that its Gaussian

→ QQ Plot

Geometric : run till first success $(1-p)^{k-1} p$

Poisson : # events under independence assumption on waiting time

(Parameter: λ)

Exponential: Actual waiting time

Log-normal: log of data is normal (μ, σ)

$\boxed{\text{Remedial}}$ session

→ list out topics to recap

→ problem → easy / interview } topics

↓

full session

Distribution:

Helps generate more (fake) data that looks like real data