

CS6046 : Multi Armed Bandits

Pre - req: Probability , Linear Algebra, coding in Python

Grading Policy

Quiz I : 20

} as per institute
calendar

Quiz II : 20

(last 2 weeks class timings)

(Group of 3) Paper Presentation : 15 (understand, implement
and replicate experiments)

(Depending on class strength)

Individual
Effort

{ Coding Assignment : 20

(4 or 5 algorithms)

Coding Project : 25

(5 questions)

uploaded
after Quiz I

Python (any requests on Matlab, Mathematica)
will not be allowed

Coding Assignment : 30 April (Tuesday, 5.00 - 8.00 PM)

Coding Project : One day before Sunday
f - slot final exam date (12 May 2024)

Exam dates cannot be changed for sake of internships
9.00 - 12.00, 2.00 - 5.00

Book: Bandit Algorithms

Csaba Szepesvari and Tomáš Lattimore

What is Multi Armed Bandit?

Artificial Intelligence, Machine learning, Deep learning,
Reinforcement learning, Supervised learning, Unsupervised learning

Define **Intelligence**:

- solutions to problems
- efficient way to solve
- learning task based on previous experience
- logical reasoning
- correct action taking environmental cues
- acquire and apply knowledge.
- remember past experience.

- memory
- new ways of solving
- Food reward
- predicting, penalty
- Different definitions are capturing different requirements
- we need a unified framework that covers most of the requirements

Artificial Intelligence

- Programming**
- computers are better than humans in solving big linear systems of equations, PDE, sorting, shortest path
 - " IRCTC ticket booking, bank transaction, tax calculation
- AI**
- humans are better than computers
 - walking, speaking, driving car, playing chess
- (Smart Programs)**

Program vs Smart Program

IRCTC Ticket Booking

- get source, destination, # tickets
- if available - tickets > # tickets
 - issue tickets
 - available - tickets \leftarrow available - tickets - issued tickets
- else
 - issue W/L tickets

can be coded up as an if-then-else

- alpha go

$$3^{N^2}$$

more combination than number of atoms in the universe



so many if-then-else \Rightarrow deadend

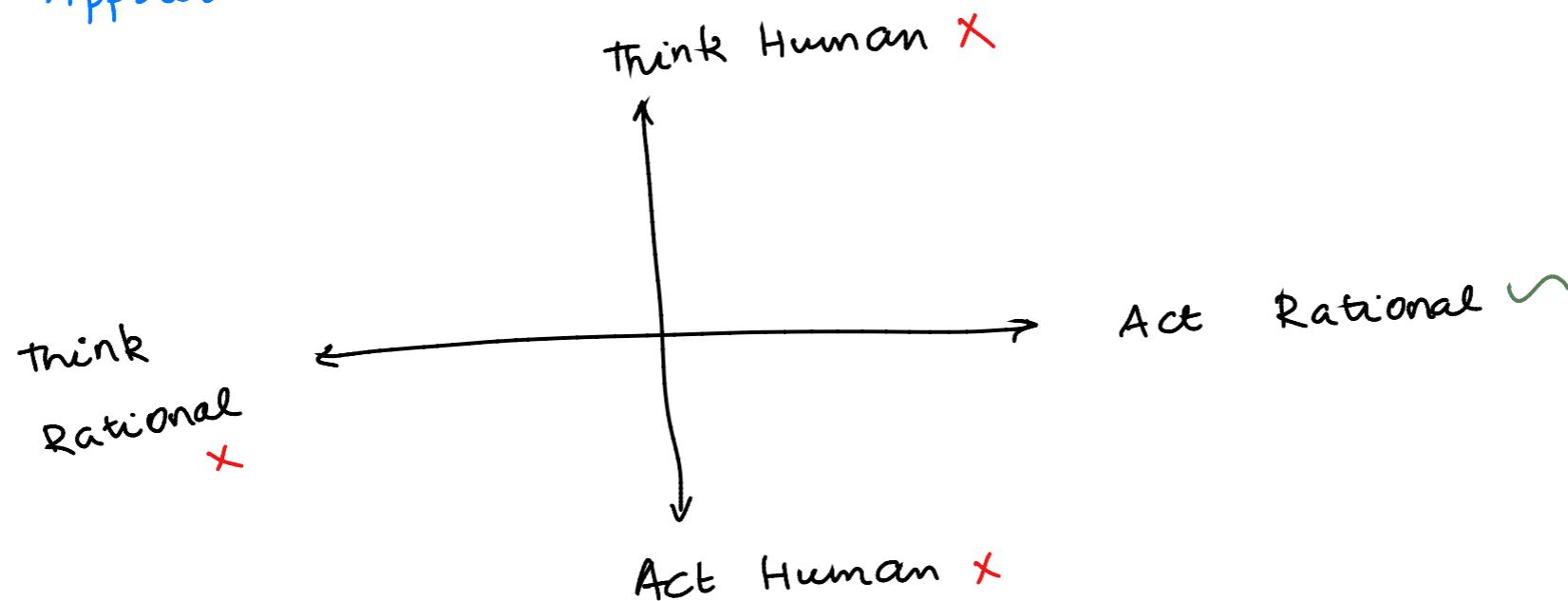
Walking

"learn"
we know how to walk
,

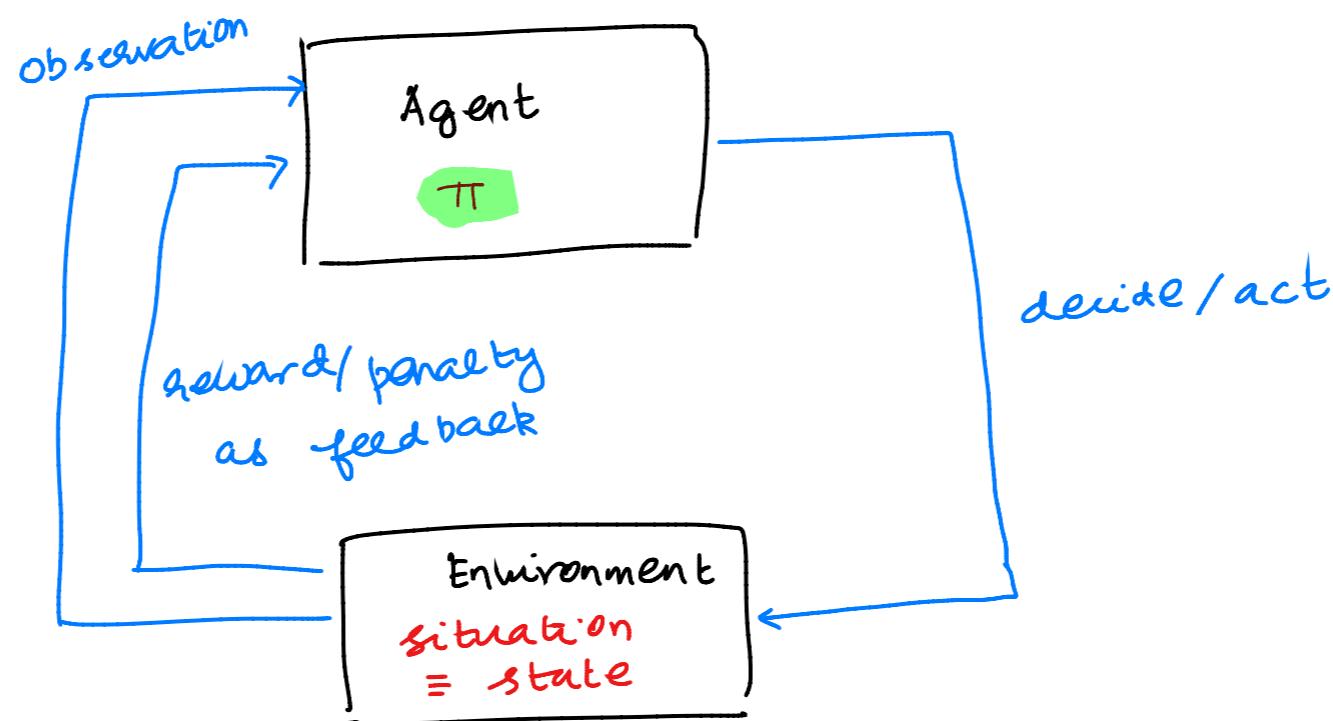
but we ourselves cannot break the process down into a set of if then else type commands

this is why a "learning" paradigm is required

Approaches to Artificial Intelligence



Rational Agents



AI Task: Describe, Predict, Prescribe

- state at time t : $s_t \in S$ (state space ; the set of all possible states)
- observation " : $o_t \in O$ (observation space)
- action " : $a_t \in A$ (action space)
- reward " : $r_t = R(s_t, a_t)$

Agent has a decision rule $\pi : O \rightarrow A$

Agent's Goal

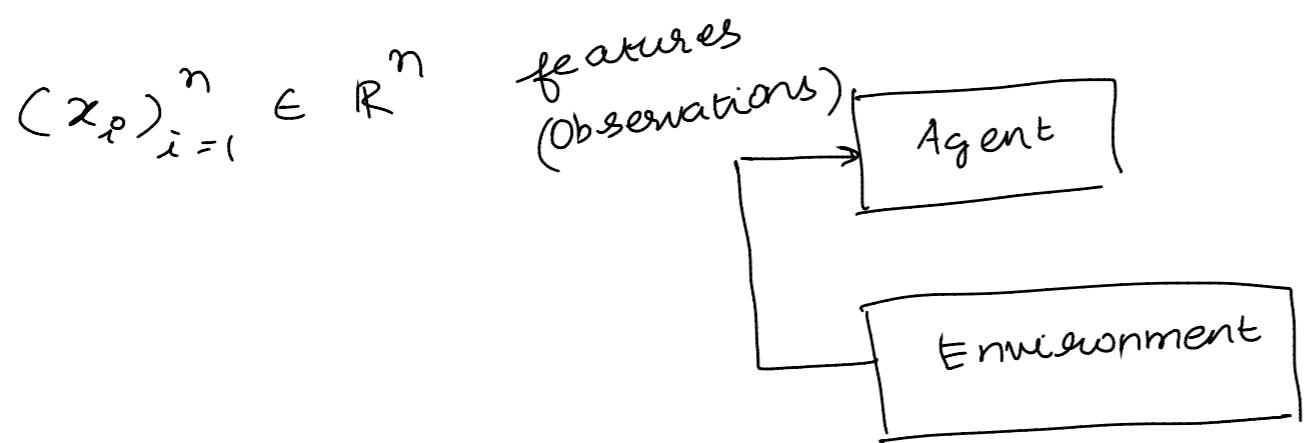
$$\max_{\pi} \mathbb{E} \left[\sum_{t=1}^{+} r_t \right]$$

Agent wants to find that decision rule which maximises its cumulative reward

In short, agent wants to behave optimally.

Next Goal: Add math to the above diagram.

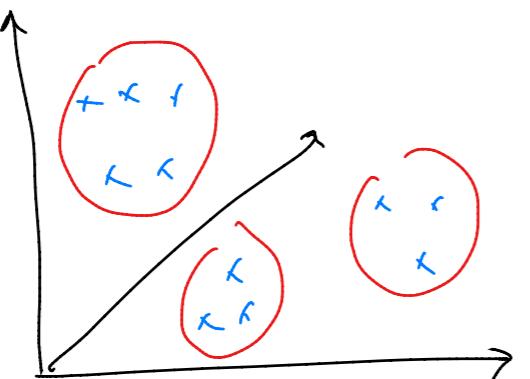
Describe (\Leftarrow No decision to be made)



Data : $(x_i)_{i=1}^n \in \mathbb{R}^d$

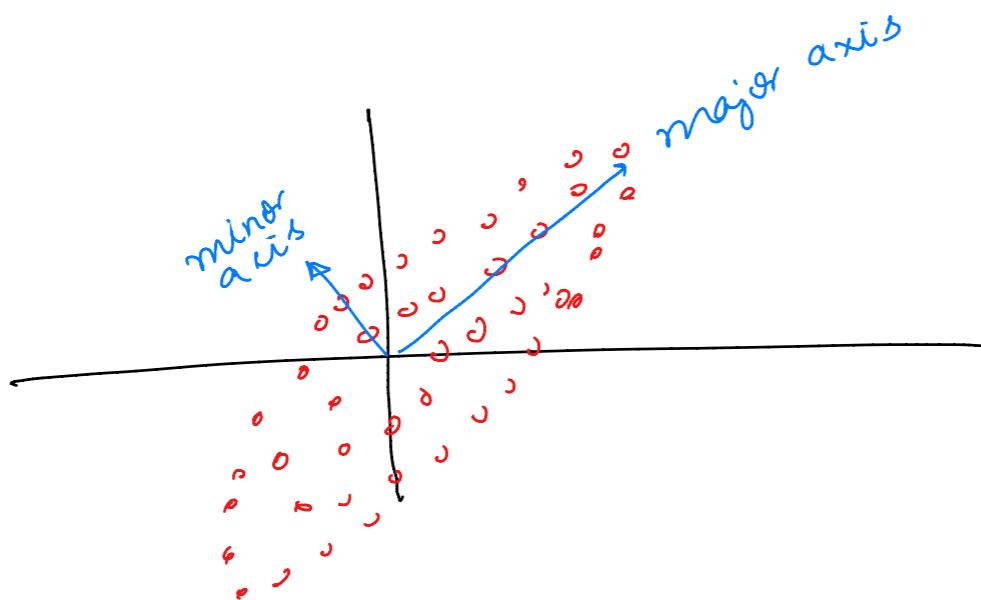
Examples of Describe Tasks

+ clustering a bunch of news articles

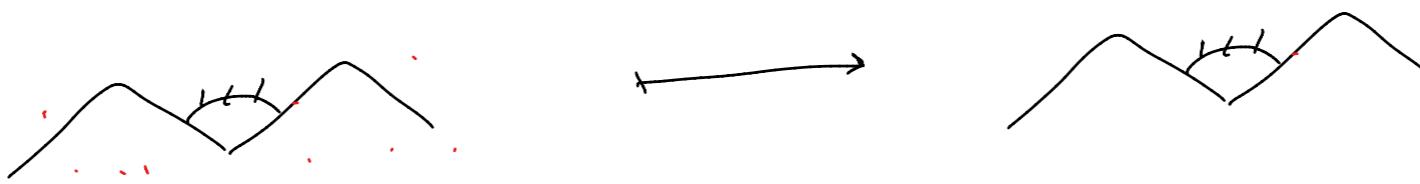


vocab
 \mathbb{R}
(freq₁, ..., freq_{#vocab})

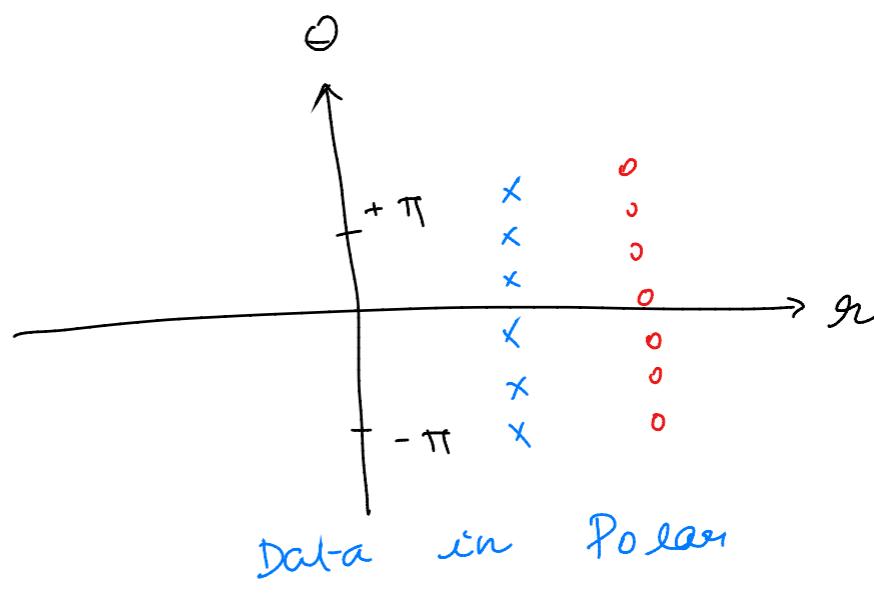
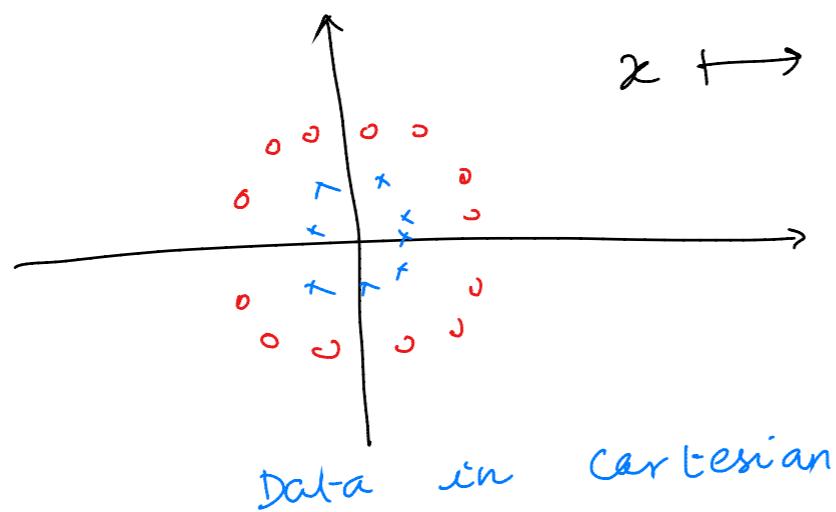
+ Dimensionality Reduction



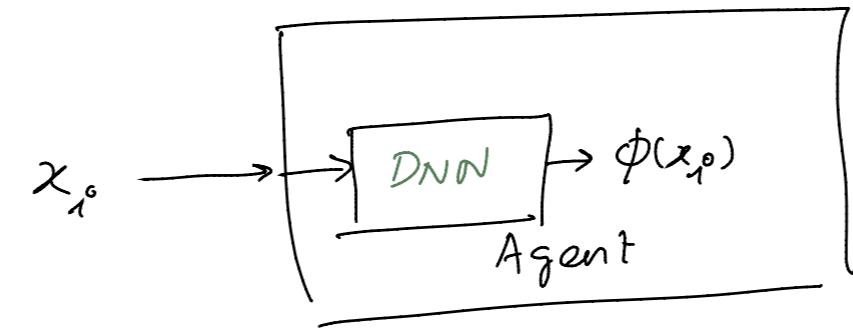
+ Image denoising



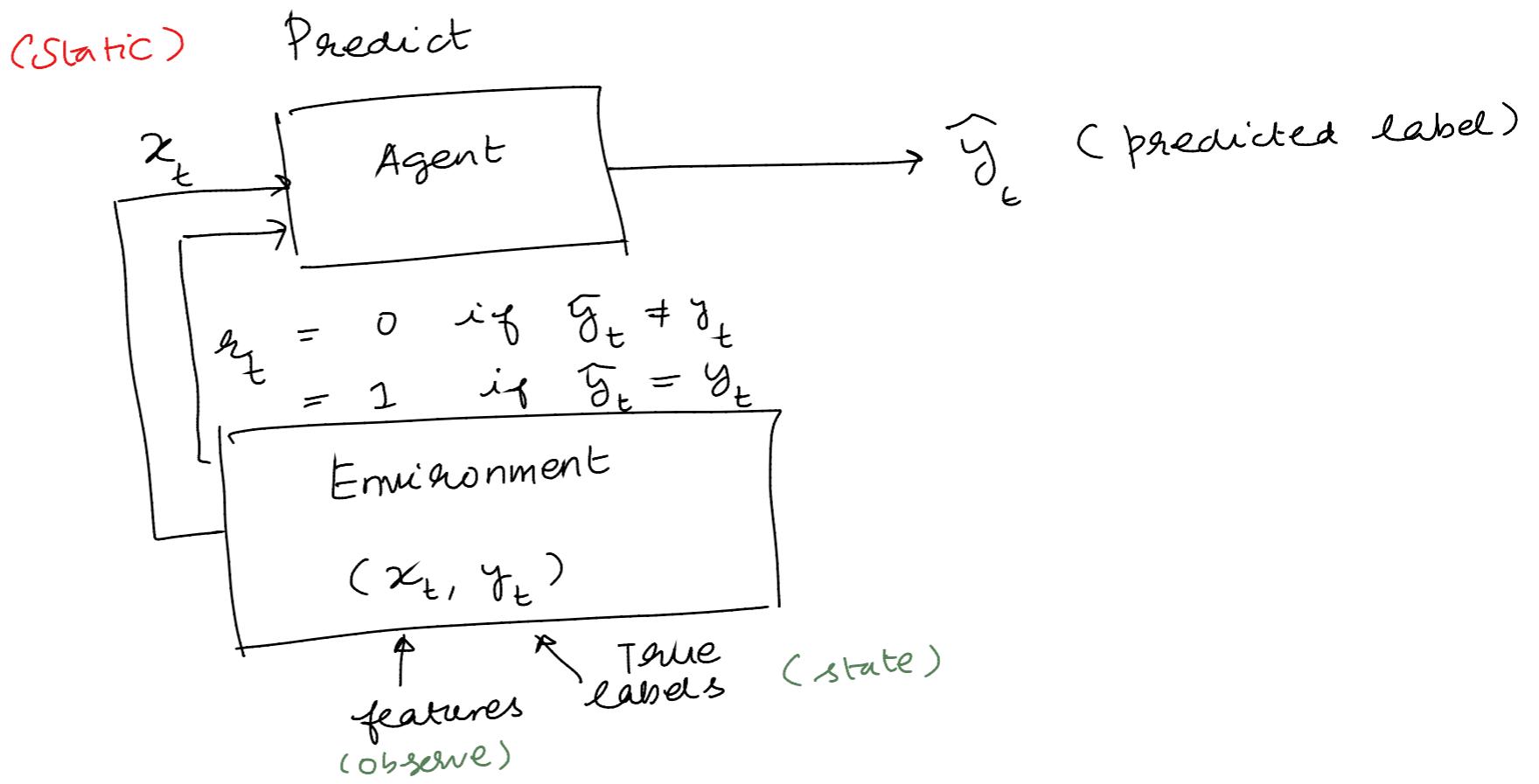
+ Representation learning or Feature learning



Deep Neural Networks are so useful in representation learning



Machine learning algorithms for descriptive tasks are known as unsupervised learning algorithms.



- Data : $(x_i, y_i)_{i=1}^n \in \mathbb{R}^d \times \{1, \dots, c\}$

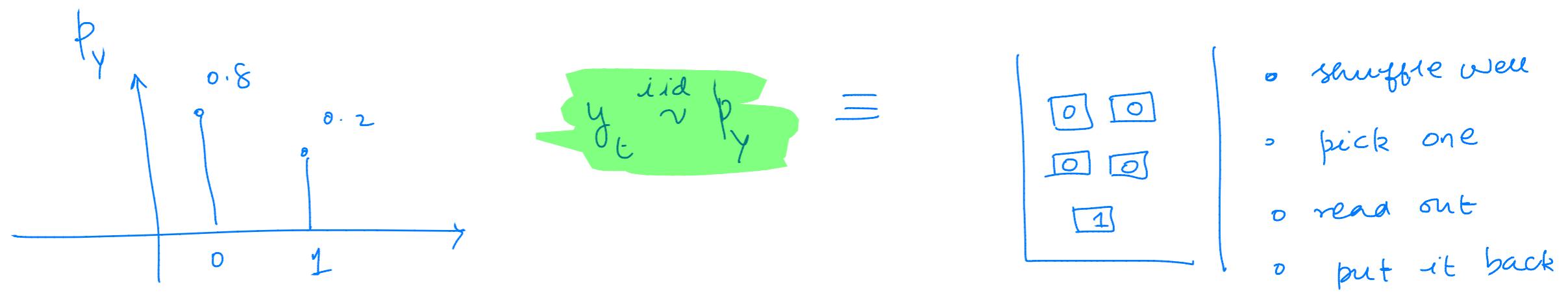
Probability Law:
Toy problem: Design a world with 360 days and 20% of days it rains

Model 1: $y_t = 1, \underbrace{1, \dots, 1}_{72 \text{ days}}, \underbrace{0, 0, \dots, 0}_{360 - 72}, 0$

Model 2 : $y_t = 0, 0, 0, 1, 0, 0, 0, 0, 1$

Model 3 :





Data model for prediction

$$y_t \sim p_y, \quad x_t \sim p_{x|y}$$

↑
class

↑ class conditionals.

Image classification

$S = Y = \{1, \dots, c\} = \{\text{cat, dog, human, \dots, house, bus, \dots}\}$

$$y_t \sim p_y, \quad x_t \sim p_{x|y},$$

↑
category

image

• $O = X = \text{Feature Space} \subseteq \mathbb{R}^d = 3 \times 10^6$

• $A = Y = \{1, \dots, c\}$

classification

Data : $(x_i, y_i)_{i=1}^n \sim P_{XY}$

• $S = Y = \{1, \dots, c\} = A$

• $O = X \subseteq \mathbb{R}^d$

Regression

$(x_i, y_i)_{i=1}^n \sim P_{XY}, x_i \in \mathbb{R}^d, y_i \in \mathbb{R}^m$

Dynamic Prediction

$$y_{t+1} \sim p_{\theta_{t+1}}(\cdot | y_t)$$

↑
next

Speech Recognition

My name is Chandru. I am a faculty in the department

w_1 w_2 w_3
 x_1 x_2 x_3

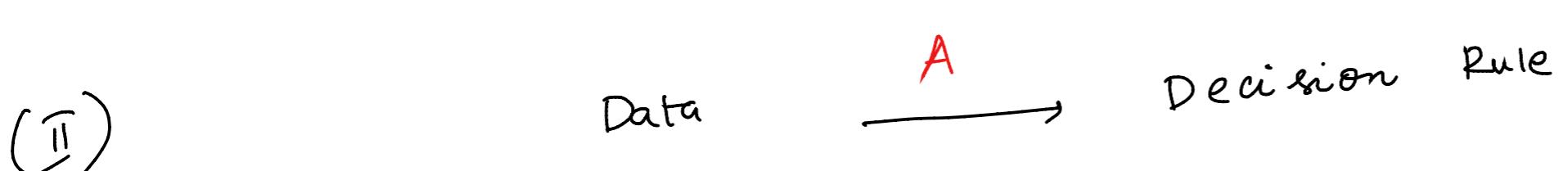
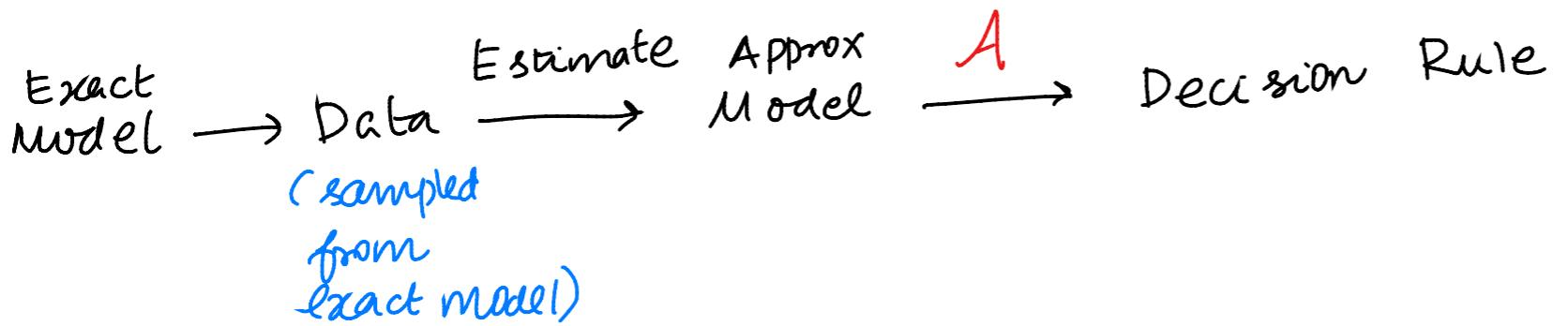
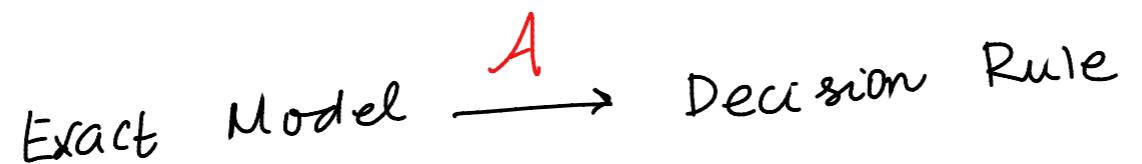
AI Tasks: Describe, Predict, Prescribe

Model : Probability Law that dictates how data gets generated

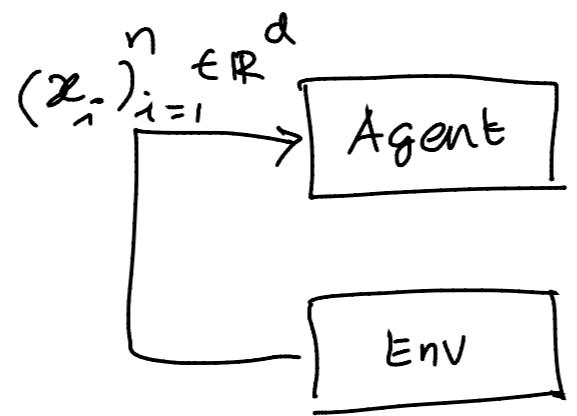
Machine learning : Data driven (model is not given by data sampled from the model is given)

algorithms to solve AI tasks.

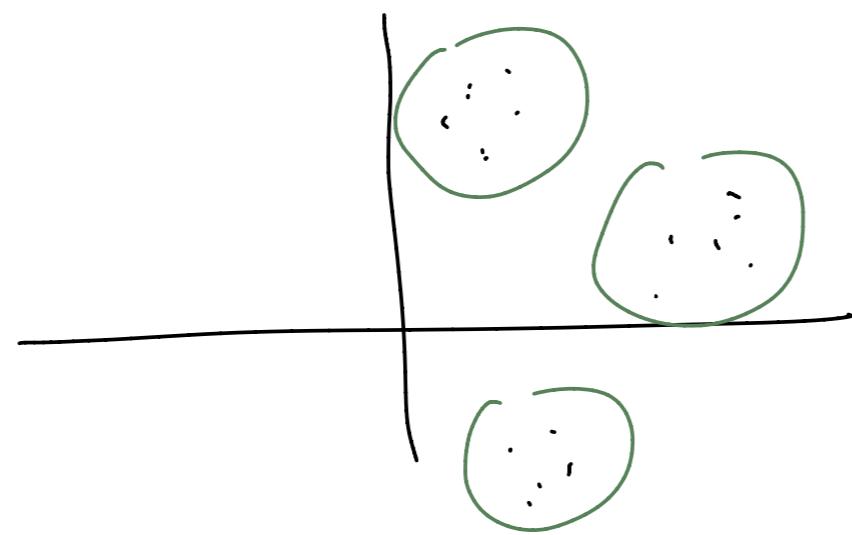
Overall philosophy of how to design algorithms.



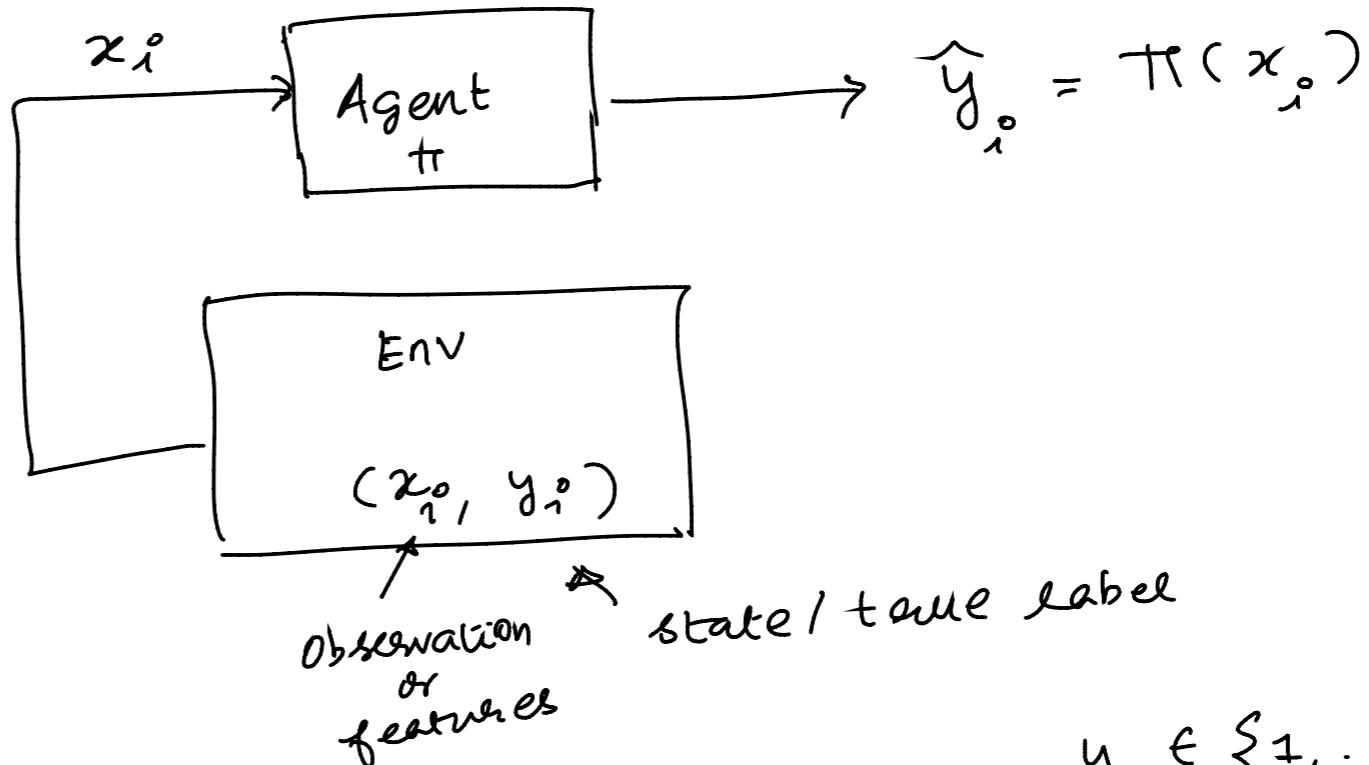
Describe Task:



$(x_i) \stackrel{iid}{\sim} p_x$
(mixture of gaussians)



(Static) Predict (the true state / label of the environment is not known to the agent; i.e.; needs to predict this)



- Model : $(x_i, y_i) \sim P_{XY}$, $x_i \in \mathbb{R}^d$, $y_i \in \{1, \dots, C\}$ (Classification)

- Bayesian Decision Theory

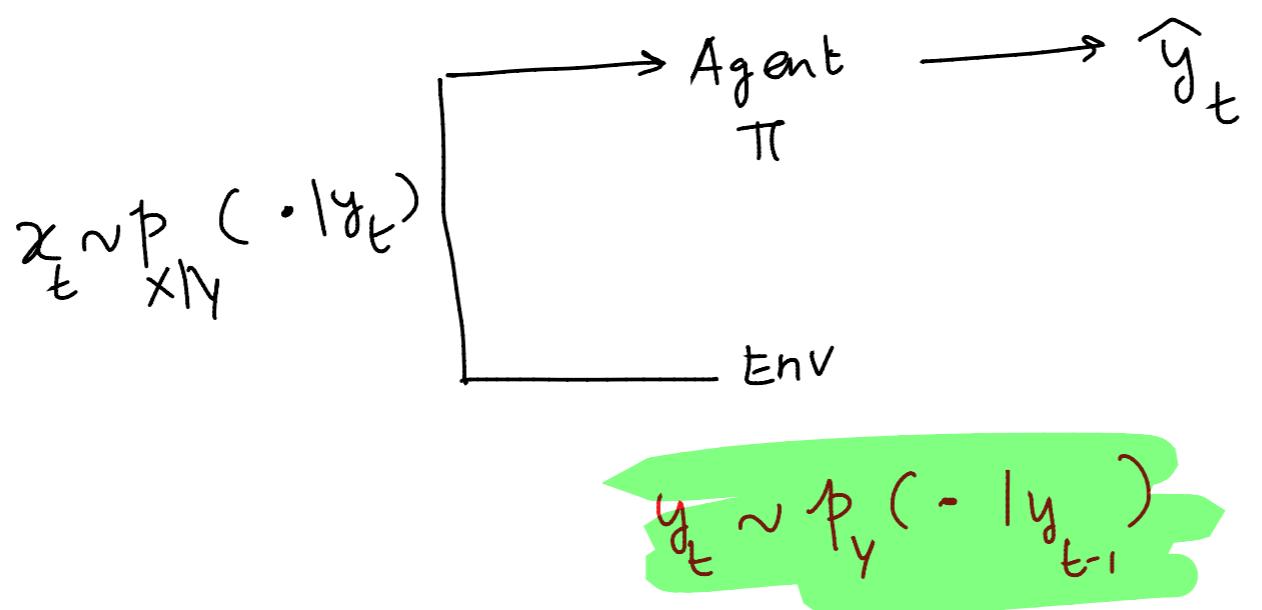
classification : $\hat{y}(x_i) = \arg \max_{y \text{ dummy}} P_{Y|X}(y \text{ dummy})$

(Model: P_Y , P_{XY})
prior class conditional

Regression : $\hat{y}(x_i) = \mathbb{E}_{y \sim P_{Y|X}} [y \text{ dummy}]$

- Supervised learning
 - $(x_i^o, y_i^o)_{i=1}^n$
 - Binary classification
 - Regression
- $$\hat{y}(x_i^o) = \theta^T x_i^o, \theta \in \mathbb{R}^d$$
- Logistic Regression
- Linear Regression

Dynamic Prediction (Model: Hidden Markov Model)



Speech Recognition

y_t : sequence of words in the speakers mind
 x_t : corresponding sound signals as heard by the listener

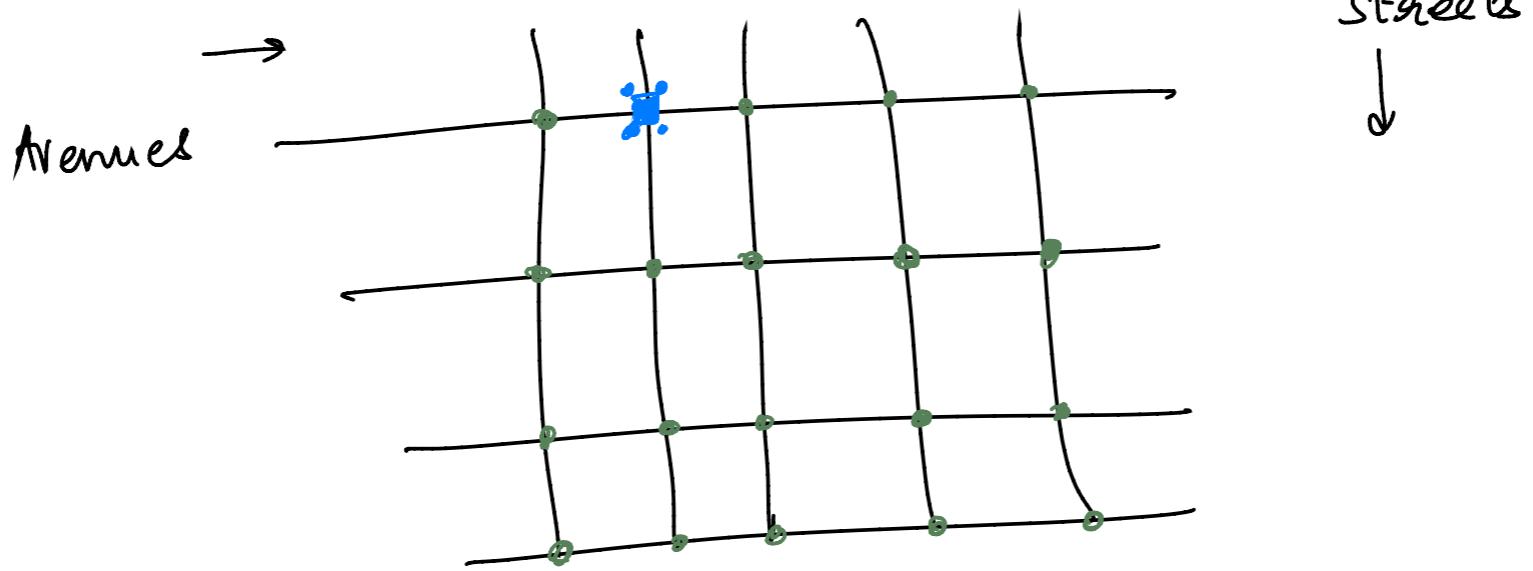
Filtering Problem

Given : $x_{1:t} = x_1, \dots, x_t \rightarrow \epsilon^{\text{Grid} \times \text{Grid}}$

Predict : $p(y_{t+R} = y), R \geq 0$

Algo : Forward Algorithm

Example: Sensor network to monitor movement of say vehicles



Smoothing Problem

Given: $x_{1:t} = x_1, \dots, x_t$

Predict: $p(y_k=y) , 1 \leq k \leq t , t = 10 \text{ min}$
 $k = 5 \text{ min}$

A algo : Forward Backward A algo

Maximum likely Sequence

Given: $x_{1:t} = x_1, \dots, x_t$

Predict $\arg \max_{\tilde{y}_{1:t}} p(\tilde{y}_{1:t} | x_{1:t})$

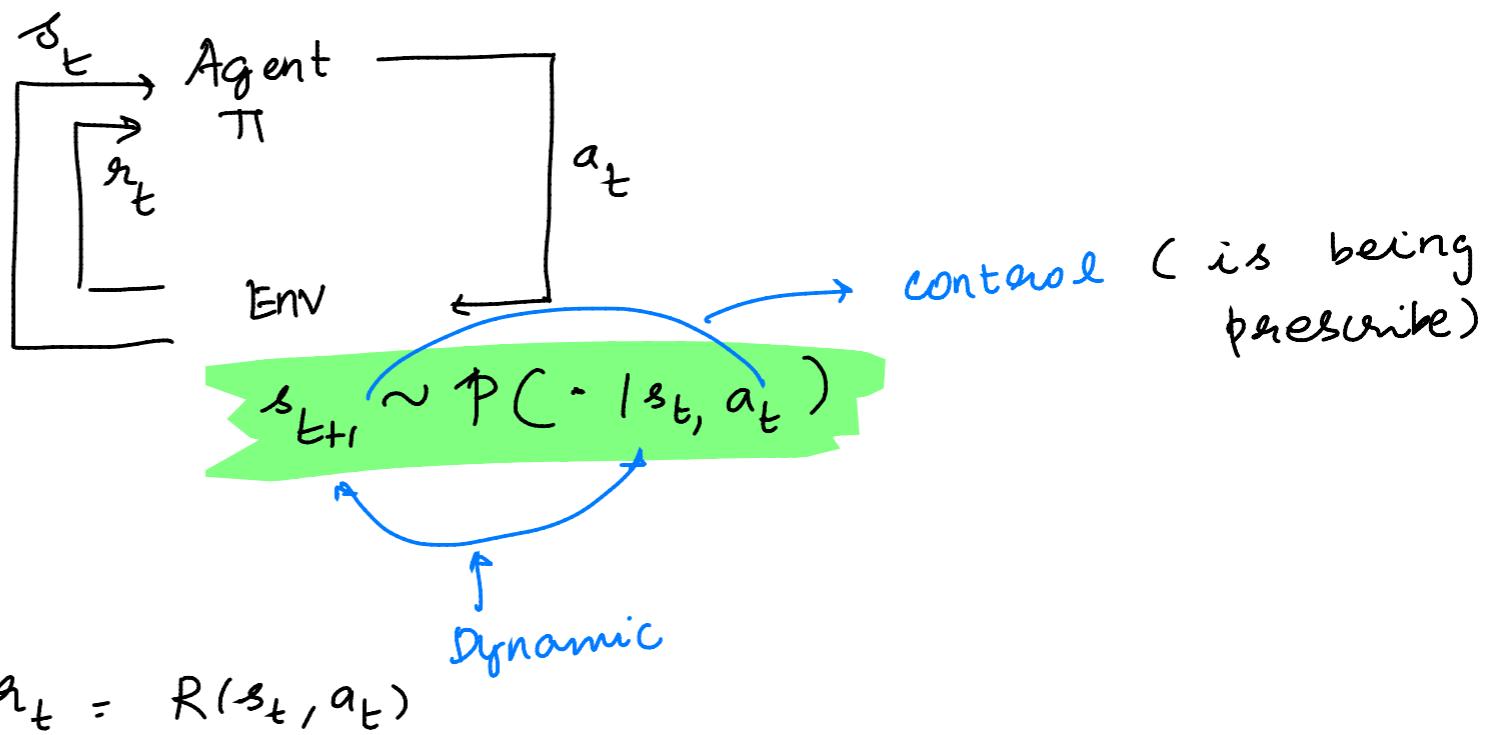
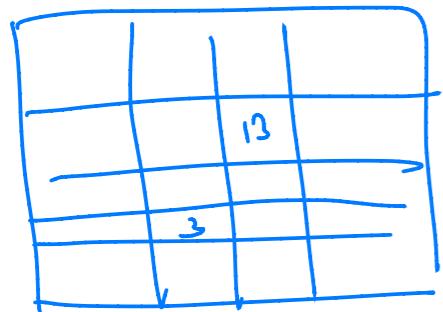
A algo : Viterbi

Data driven: RNNs, LSTMs, Auto Regression, Marked Language Models

Machine
Methods

Prescribe = Dynamic Control (agent knows the state fully; how to change the state for good)

Example: Playing chess



Model Based
Algorithm: Value Iteration, Policy Iteration, Linear Programming

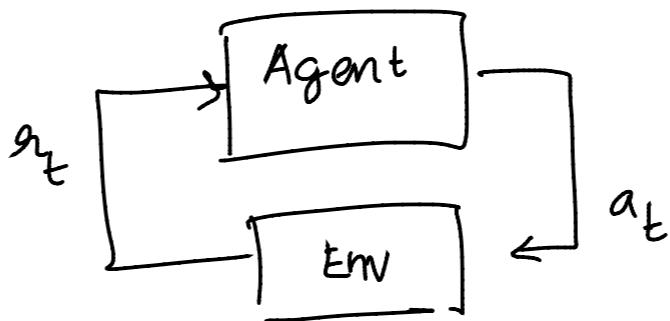
Data

: $(s_t, a_t, r_t, s_{t+1})_{t=1}^+$

↑
was here
↑ did Env's
↑ got this
↑ this happened next
because of what I did

Algo : Q-learning
Actor critic

Static Control
(Focus of our course)



$$a_t \in A = \{1, \dots, k\}$$

$$a_t \sim p(\cdot)$$

$$a_t$$

Example: Google Ad Placement

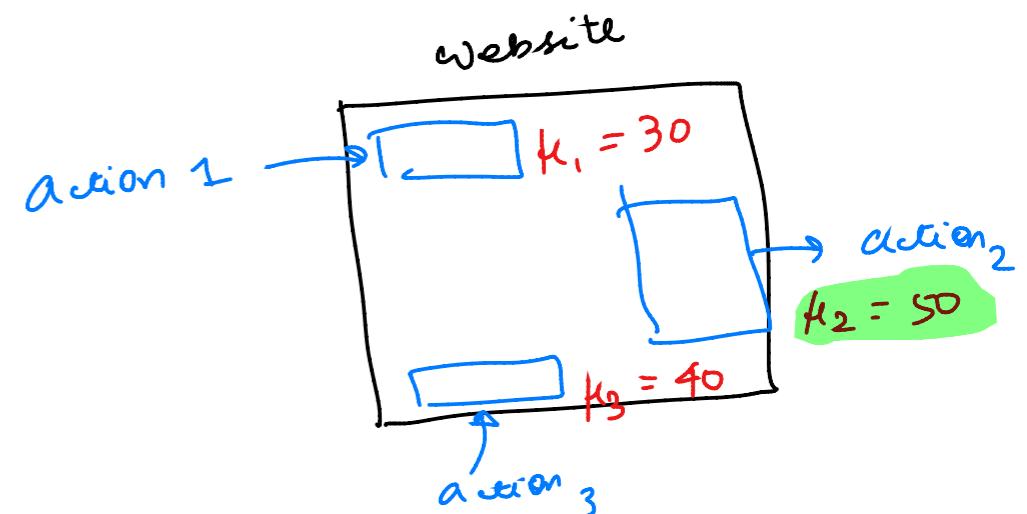
If Model is known

$$p_1, \dots, p_k$$

just compute

$$\mu_i = \mathbb{E}_{r_t \sim p_i} [r_t]$$

$$i^* = \arg \max_i \mu_i$$



r_t : click through rate
or
clicks / sec

This course: • p_1, \dots, p_k are not known

• at time t , $a_t \sim p(\cdot)$ can be sampled.

AI Problem
Task

Descriptive

$$\phi_x$$

Predictive
(Static)

classification
 $p_x, p_{x|y}$

Regression
 p_{xy}

Predictive
(Dynamic)

$$p_{y_{t+1}|y_t} / p_{x|y}$$

Dynamic
control

$$s_{t+1} \sim p(s_{t+1} | s_t, a_t)$$

$$h_t = R(s_t, a_t)$$

Static
control

$$p_1, \dots, p_k$$

Model
Driven

NA

Data
Driven

$$(x_i)_{i=1}^n$$

ML Method

Unsupervised learning
(PCA, clustering, GMM)
auto-encoders
Courses: PRML, DL

Bayesian
Decision
theory

$$(x_i, y_i)_{i=1}^n$$

Supervised Learning

(linear / logistic
regression, SVM,
DNN, CNN, ResNet)

Courses: PRML, DL

Hidden
Markov
Models
(filtering, smoothing)
Maximum likely
sequence
FBN, FW-BN, n-terbi

$$(x_i, y_i)_{i=1}^n$$

 $y_{i+1} \text{ depends}$
on y_i

LSTMS, RNNs
Transformers

Courses: PRML, DL

Markov
Decision
processes (MDP)

$$(s_t, a_t, r_t, s_{t+1})_{t=1}^T$$

TD, Q Learning,
Actor critic
Courses: RL,
Topics in RL

Stateless MDP

Trivial

$$f_u := \mathbb{E}_{a_t \sim p_i} [r_t]$$

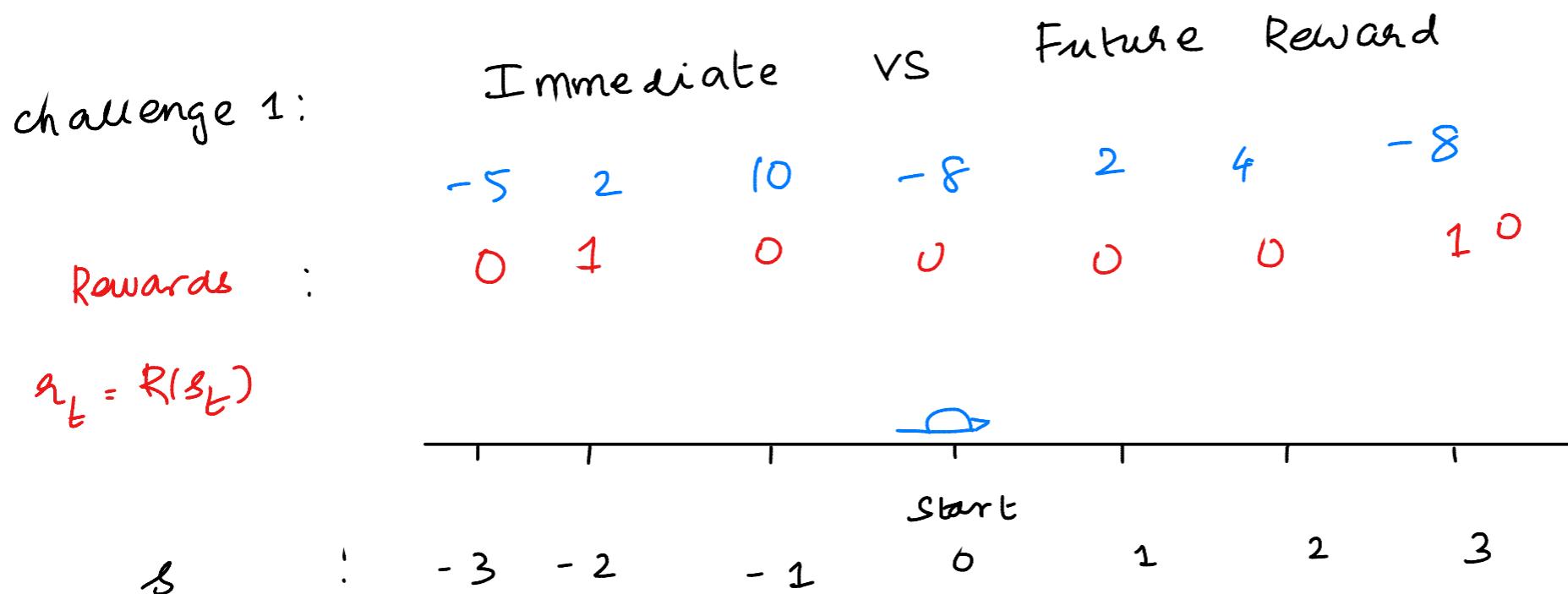
$$i_* = \arg\max_k f_k$$

$$(a_t, r_t)_{t=1}^T$$

Multi-Armed
Bandits
This course

Interesting aspects of Control / Prescription Problems

Examples: Playing chess/Go, Autonomous Driving, Robotics (Mujoco)
 NeurIPS competition Learning by Doing challenge, Finance (portfolio)
 2021 FinRL Management
 inventory control, online Ad placement, dynamic pricing



Goal: $\max \sum_{t=0}^T a_t \quad | s_0 = 0$

say $T = 0$, no action

$t = 1$, any action is fine

$$A = \{\leftarrow, \rightarrow, \cdot\}$$

$$T=2, \quad a_0 = \text{left}, \quad a_1 = \text{left}$$

$$q_0 = 0, \quad q_1 = 0, \quad q_2 = 1$$

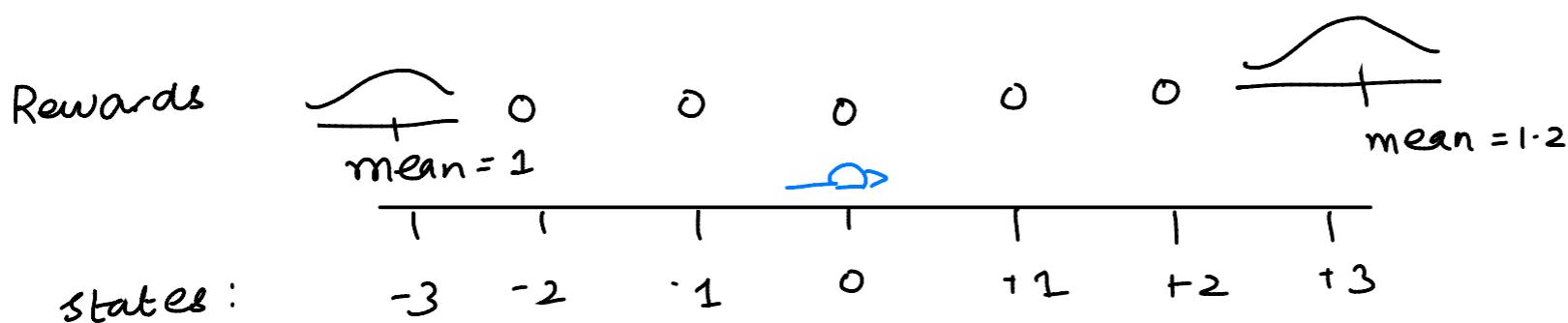
$$T=3, \quad a_0 = \text{right}, \quad a_1 = \text{right}, \quad a_2 = \text{right}$$

$$q_0 = 0, \quad q_1 = 0, \quad q_2 = 0, \quad q_3 = 10$$

challenge 2: Temporal Credit Assignment
 Say I play a game of chess, and I win. Is it possible to pinpoint the move that caused the win (in general)?

Jan 25

challenge 3: Randomness in Environment



- you have to visit / sample each reward multiple times
 - ideally one needs infinitely many samples to learn
- explore →
- as you keep sampling, you get a decent estimate (after point of time) of the means. Tendency is to choose the option which has given the best so far
- exploit →
- Moral:
- cannot explore for a finite amount time and then commit (i.e., lock the option)
 - Continuously balance explore and exploit

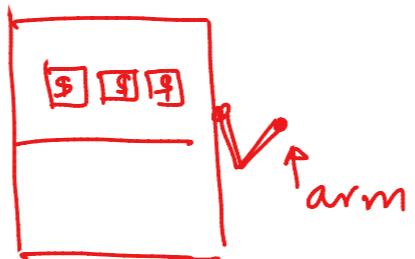
Key Question:

What is the right way to balance explore and exploit?

Formal Multi Armed Bandit Setup

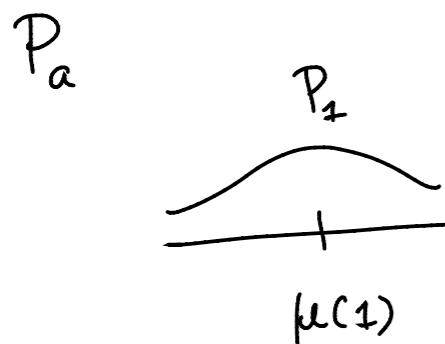
- action space : $A = \{1, \dots, K\}$ (each action is known as an arm)

book : Martingale

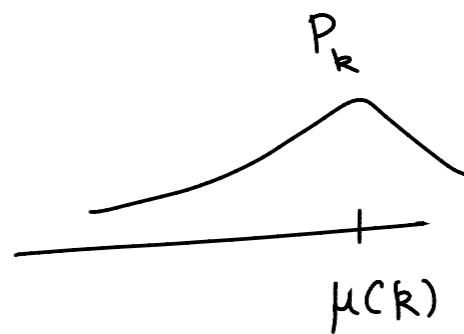


examples: ad placement, alpha Go / chess

- each arm is associated with a probability distribution



...



- at time t, we play/pull/choose arm $a_t \in A$

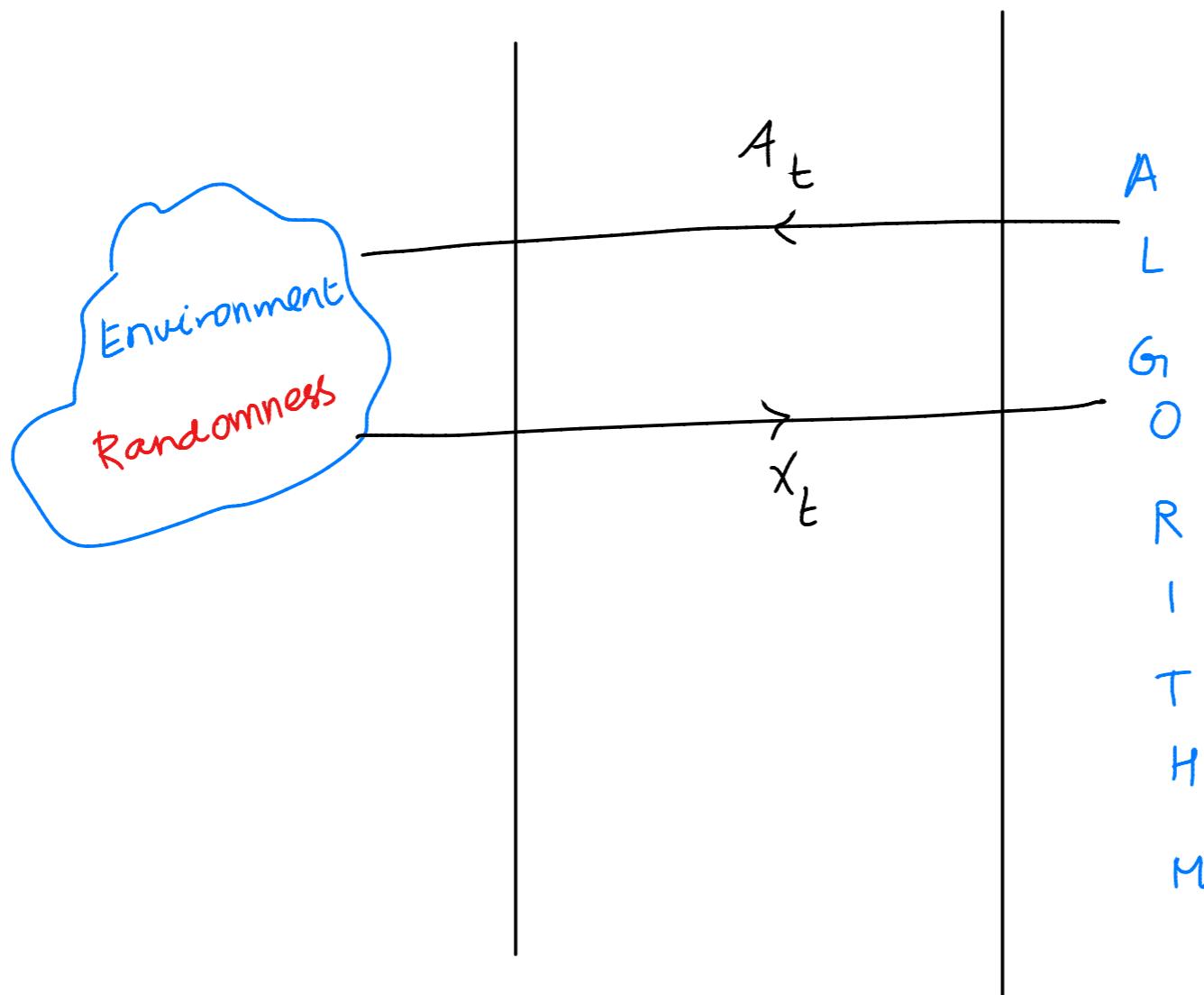
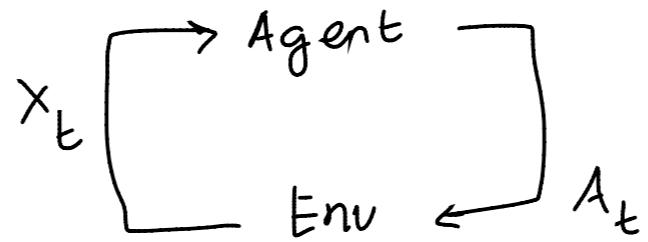
- Reward obtained

$$x_t \sim P_{a_t}$$

- Mean reward

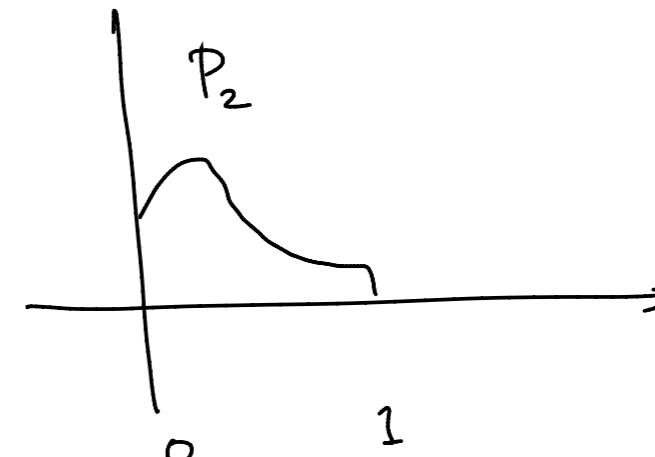
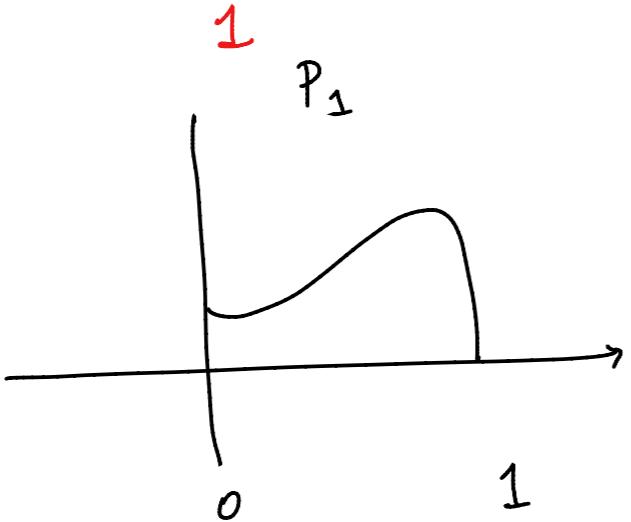
$$\mu(a) = \mathbb{E}_{P_a} [x_t]$$

- Best reward $\mu_* = \max_a \mu(a)$
 - Best Arm $a_* = \arg \max_a \mu(a)$
 - Sub-Optimality Gaps $\Delta(a) = \mu_* - \mu(a)$
 - Goal : Minimise Regret $R(T)$
- Regret (T) = $R(T) = \mathbb{E} \left[\sum_{t=1}^T (\mu_* - x_t) \right]$
- ↑
how much did we lose
w.r.t best possible
- $x_t \sim P_{a_t}$
- A logarithmic Goal: To pick a_t such that we incur less regret, i.e.,
- regret is Sub-linear $\Rightarrow \frac{R(T)}{T} \rightarrow 0$ (loss per round)



- say algorithm is deterministic
- has some internal variables
- due to randomness from environment
internal variables are also random

Problem Instance: $A = \{1, 2\}$



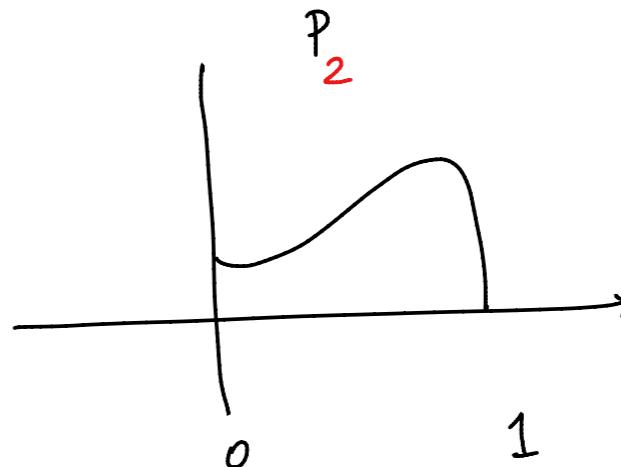
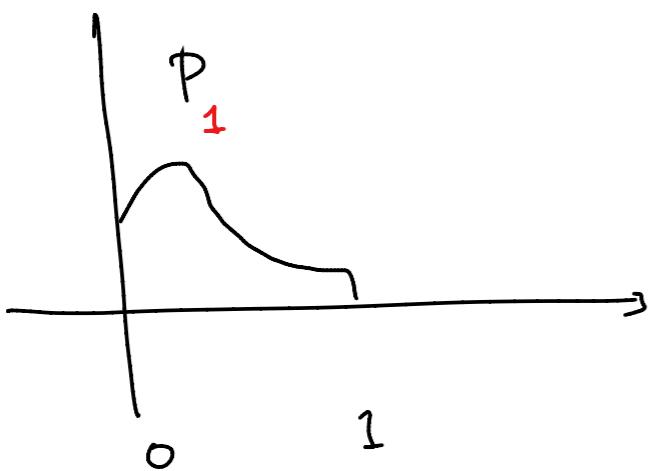
Algorithm is deterministic

$$A_1 = \text{arm 1}, \quad x_1 = 0.9$$

$$A_2 = \text{arm 2}, \quad x_2 = 0.1$$

$$A_3 = \text{blah}$$

Problem Instance: $A = \{1, 2\}$



Algorithm is deterministic

$$A_1 = \text{arm 1}, \quad x_1 = 0.9$$

$$A_2 = \text{arm 2}, \quad x_2 = 0.1$$

$A_3 = \text{blah}$

Probability Space $\equiv (\Omega, \mathcal{F}, P)$

Omega

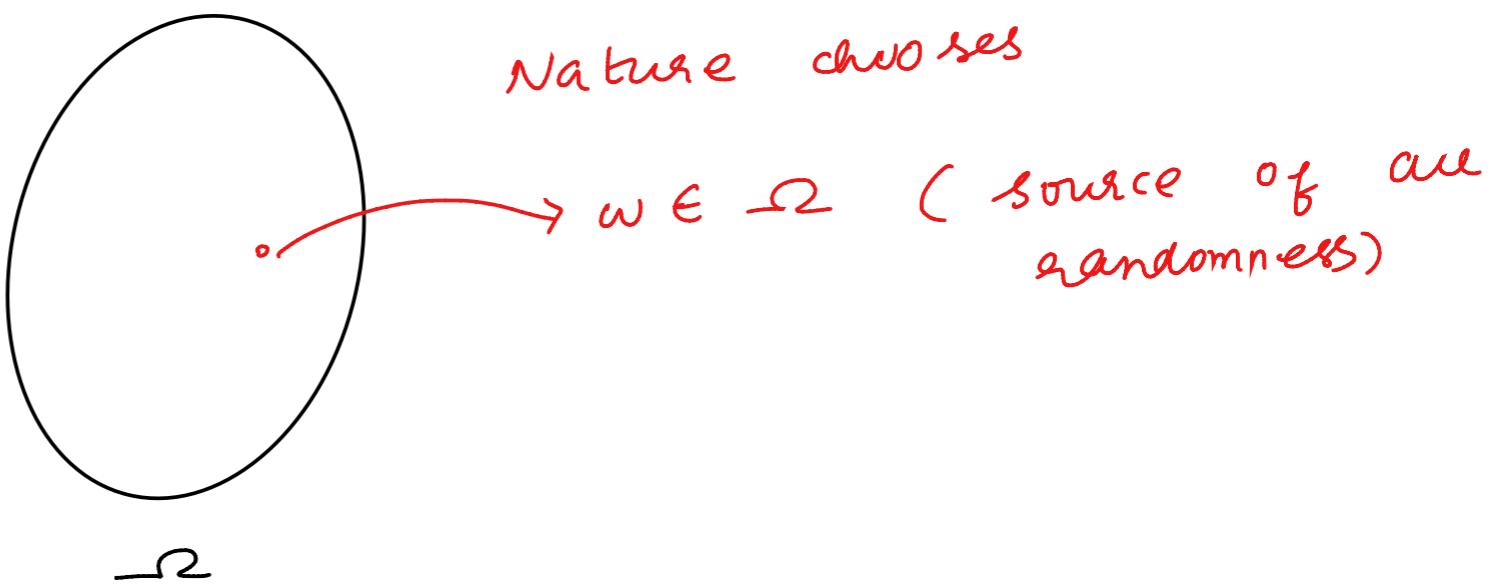
↑
script F

↑
script P

Ω : Sample space

\mathcal{F} : sigma algebra of events

P : Probability assignment



Jan 30

$$X: \Omega \rightarrow \mathbb{R}$$

- Nature picks $w \in \Omega$

- The random variable X 's value is $X(w)$

Toy Example :

ω and X for roll of a die

$$\omega = \{1, 2, 3, 4, 5, 6\}$$

$$X(\omega) = \omega$$

Toy Example :

ω and X_1 for roll of a die

X_2 is toss of a coin

$$\omega = \{(1, H), (2, H), \dots, (6, H), (1, T), (2, T), \dots, (6, T)\} \rightarrow \omega = (\omega_1, \omega_2)$$

Die

$$X_1(\omega) = X_1((\omega_1, \omega_2)) = \omega_1 \Rightarrow X_1(\omega) = \omega_1$$

Coin

$$X_2(\omega) = X_2((\omega_1, \omega_2)) = \omega_2$$

$$X(\omega) = \omega \times$$

ω and X_1 for roll of a die

X_2 is toss of a coin

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Customer 1

Die₁

$$X_1(\omega) = \omega$$

win₁

$$X_2(\omega) = \begin{cases} H & , 1 \leq \omega \leq 3 \\ T & , 4 \leq \omega \leq 6 \end{cases}$$

win₂

$$X_3(\omega) = \begin{cases} H & , \omega \text{ is odd} \\ + & , \omega \text{ is even} \end{cases}$$

Customer 4

Die₂

$$X_4(\omega) = 7 - \omega$$

$$\omega = \{1, 2, 3, 4, 5, 6\}$$

x_1, x_2 independent of each other

$A = \{1, 3, 5\}$ and $B = \{2, 4, 6\}$ are mutually exclusive

$A = \{2, 3, 5\}$ and $B = \{2, 4, 6\}$ are not related

\Rightarrow independence ?

$$P(A \cap B) = P(A) \cdot P(B)$$

$$P(\{\}) = P(\{1, 3, 5\}) \cdot P(\{2, 4, 6\})$$

$$\frac{1}{6} \neq \frac{1}{2} \cdot \frac{1}{2}$$

$$\frac{1}{6} \neq \frac{1}{4}$$

$$A = \{1, 2, 3, 4, 5, 6\}$$

$$B = \{2\}$$

} correct by trivial

$$P(A \cap B) = \frac{1}{6}$$

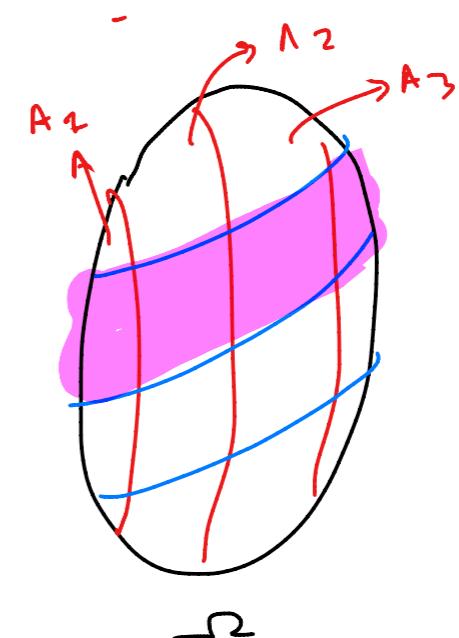
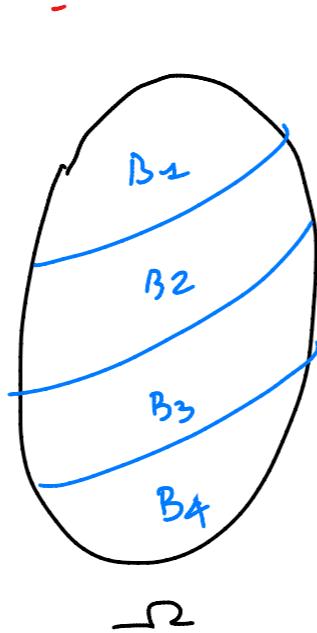
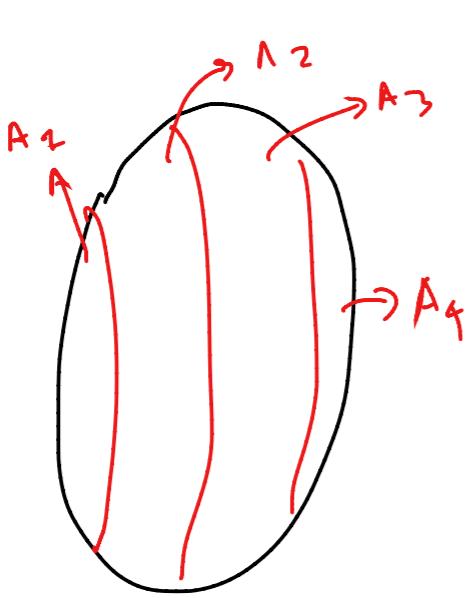
$$A = \{2, 3\}, \quad B = \{2, 4, 6\}$$

$$P(A \cap B) = P(\{2\}) = \frac{1}{6}$$

$$P(A) = P(\{2, 3\}) = \frac{1}{3}$$

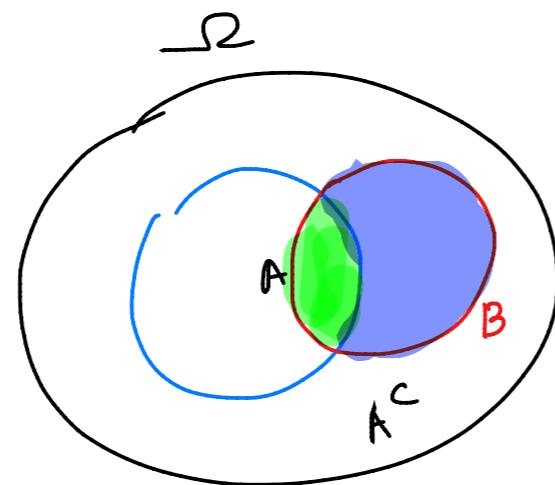
$$P(B) = P(\{2, 4, 6\}) = \frac{1}{2}$$

$$\Rightarrow P(A)P(B) = \frac{1}{3} \cdot \frac{1}{2} = \frac{1}{6}$$



$$X_1 = \begin{cases} 1, & \text{if } \omega = 2 \text{ or } 3 \\ 0, & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{if } \omega = 2 \text{ or } 4 \text{ or } 6 \\ 0, & \text{otherwise} \end{cases}$$



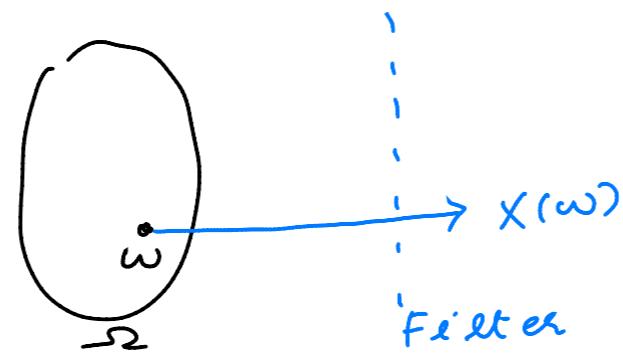
$$P(A) : 1 - P(A)$$

Step 1 :

$\omega \in \Omega$ is the outcome of random experiment
nature chooses ω

Step 2 !

Real Valued Random variable $X : \Omega \rightarrow \mathbb{R}$

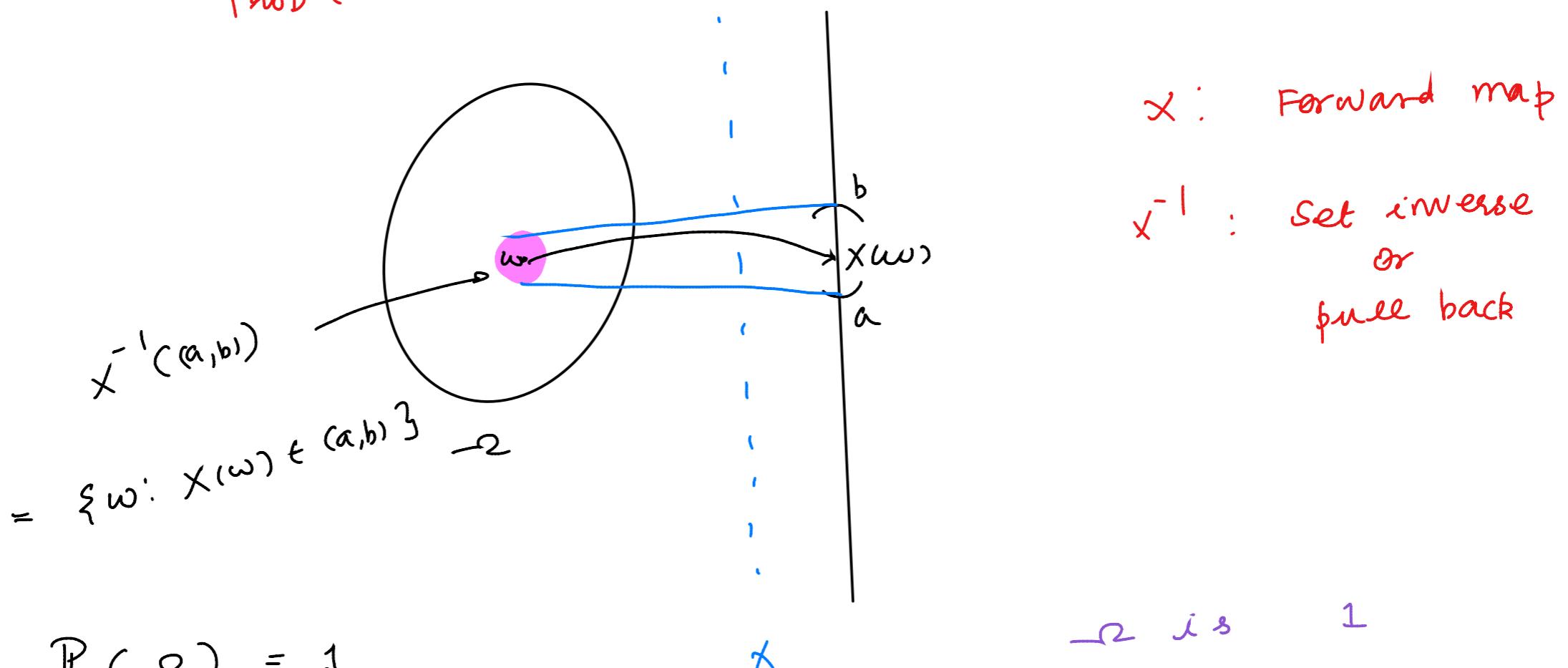


- we cannot see ω
- we can only observe $X(\omega)$

We are interested in knowing "chances" (informally but)

Step 3:

$$\text{Prob}(a < X < b) = \text{Prob}(X \in (a, b))$$



$$P(\Omega) = 1$$

$$P(x^{-1}(a, b)) = ?$$

we need

$$\text{Prob}(a < x < b)$$



we need

to know "how much" is

$$A = x^{-1}(a, b)$$

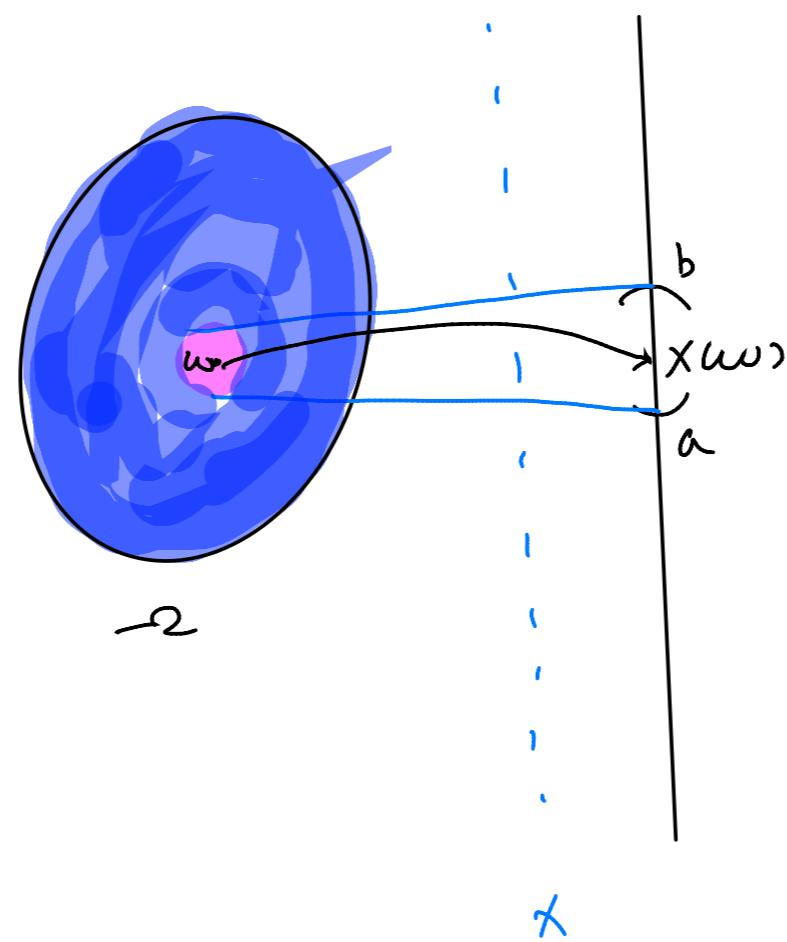
$$(\mathcal{R}, \mathcal{F}, P)$$

$$P(A)$$

↑
Probability assignment

• we also need $\text{Prob}(X \notin (a, b))$

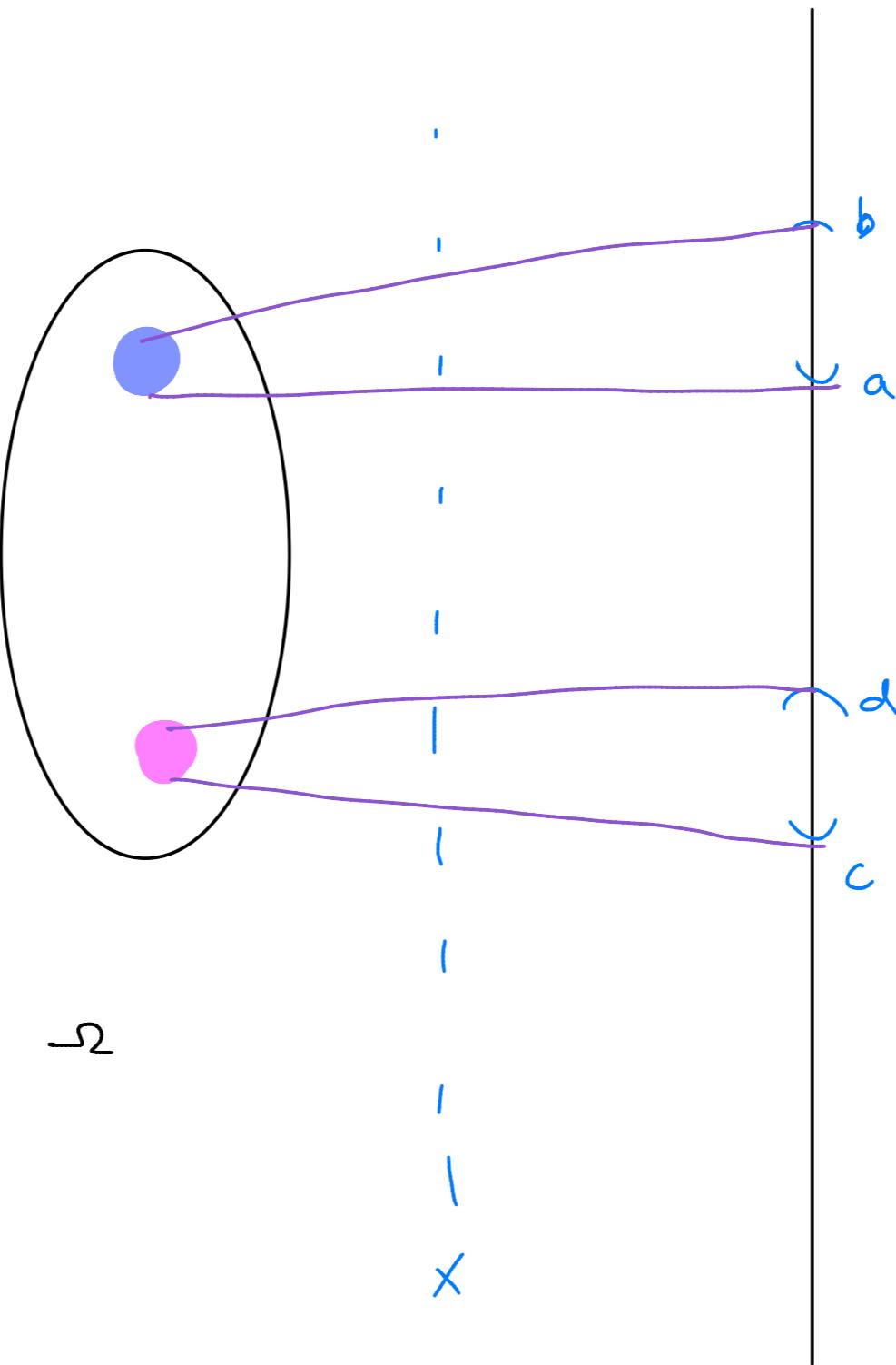
$$A^c = \{\omega : X(\omega) \notin (a, b)\}$$

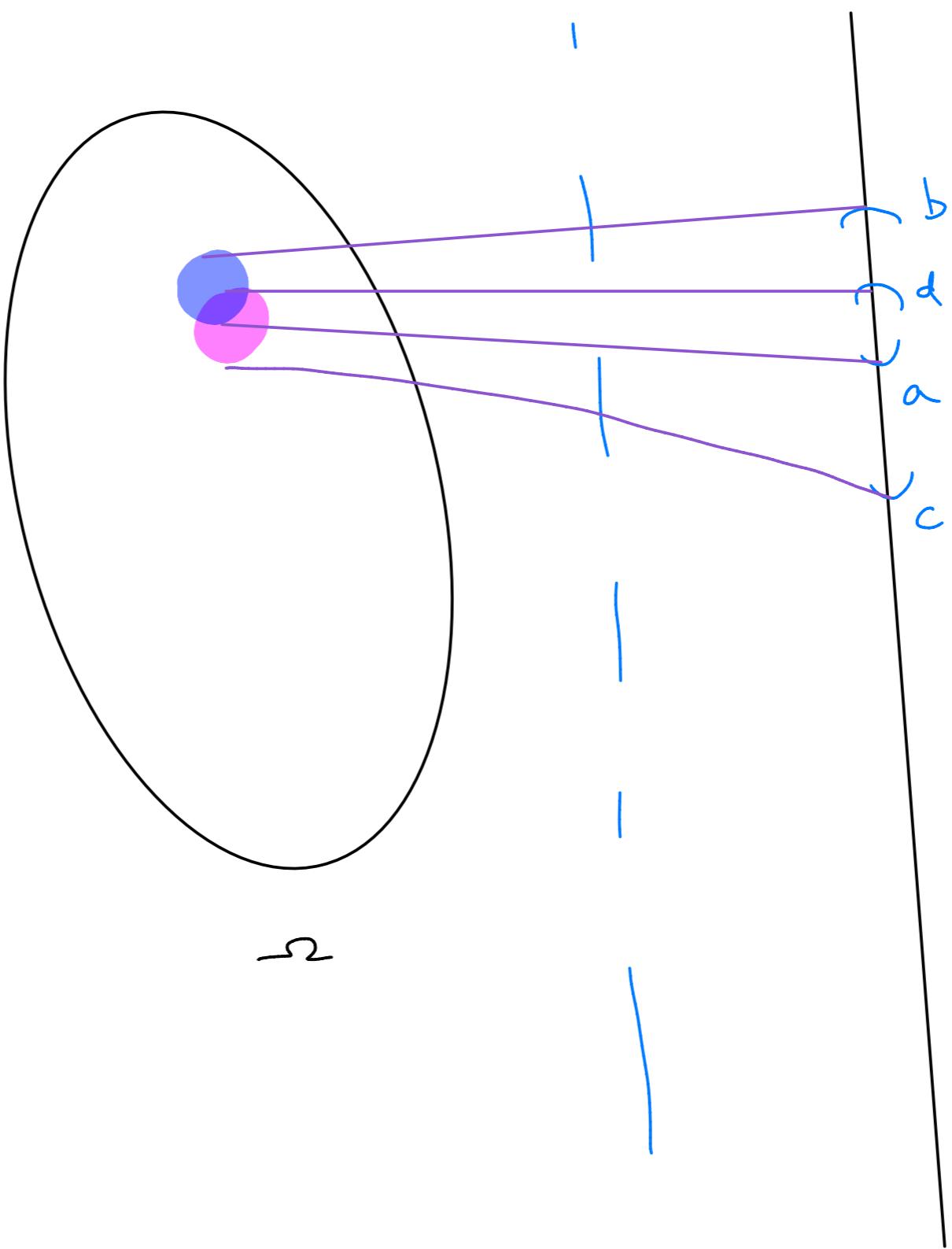


I need $P((x^{-1}(a, b))^c)$

• we also need

$$\text{Prob } (x \in (a, b) \text{ or } x \in (c, d))$$





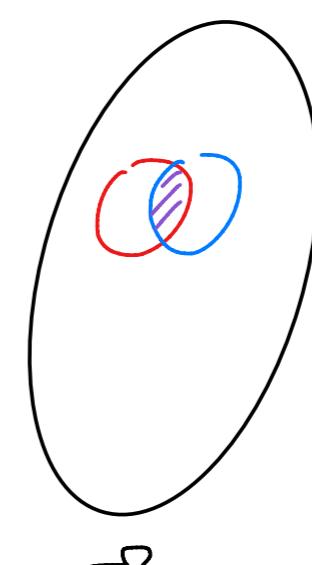
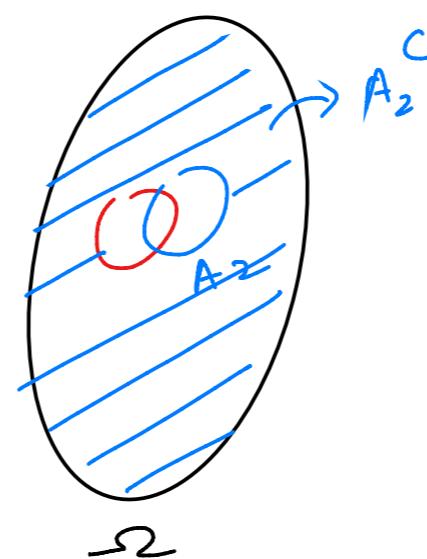
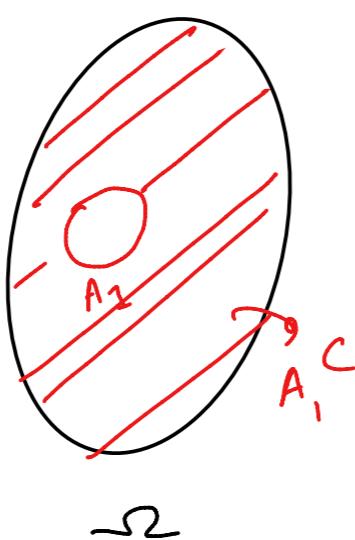
$$\pi_P(x^{-1}(a,b) \cup x^{-1}(cc,d))$$

Step 3: \mathcal{F} is a collection of sets such that we pick any $A \in \mathcal{F}$, then we should be able to tell $P(A)$

\mathcal{F} contains all sets / shapes / areas / regions for which we can tell $P(A)$

\mathcal{F} is a σ -algebra

- * $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ (*Closed under complements*)
- * $A_1, A_2, \dots \in \mathcal{F} \Rightarrow \bigcup_i A_i \in \mathcal{F}$ (*Closed under countable union*)



$$(A_1^c \cup A_2^c)^c$$

Step 4:

Probability Assignment

$$P: \mathcal{F} \rightarrow [0, 1]$$

$$\ast P(\Omega) = 1 \quad (P(\emptyset) = 1 - P(\Omega) = 0)$$

$$\ast A_1 \cap A_2 = \emptyset$$

$$\Rightarrow P(A_1 \cup A_2) = P(A_1) + P(A_2)$$

Finite Additivity

Countable additivity

$$\ast \{A_i\}_{i \geq 0}, \quad A_i \cap A_j = \emptyset, \quad i \neq j \quad (\text{pairwise disjoint})$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Step 5: Generalisation Of Step 3

we are okay with any $x: \Omega \rightarrow \mathbb{R}$ as long as
 $P(x^{-1}(a,b))$ can be told $\Rightarrow x^{-1}(a,b) \in \mathcal{F}$ for all open intervals (a,b)

then we say that x is measurable w.r.t (Ω, \mathcal{F}, P)

Tidy example:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$\begin{aligned} \mathcal{F} &= 2^\Omega \text{ (power set)} \\ &= \{\emptyset, \Omega\} \end{aligned} \quad \left. \right\} \text{ trivial example}$$

$$= \{\emptyset, \Omega, \{2, 4, 6\}, \{1, 3, 5\}\}$$

$$\begin{array}{ll} \bullet P(\emptyset) = 0 & \bullet P(A) = \frac{1}{2} \\ \bullet P(\Omega) = 1 & \bullet P(A^c) = \frac{1}{2} \end{array}$$

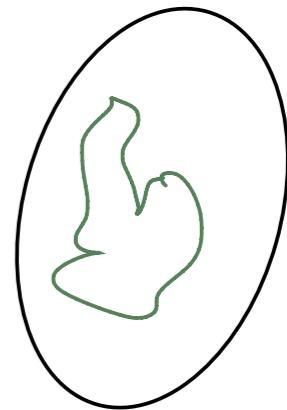
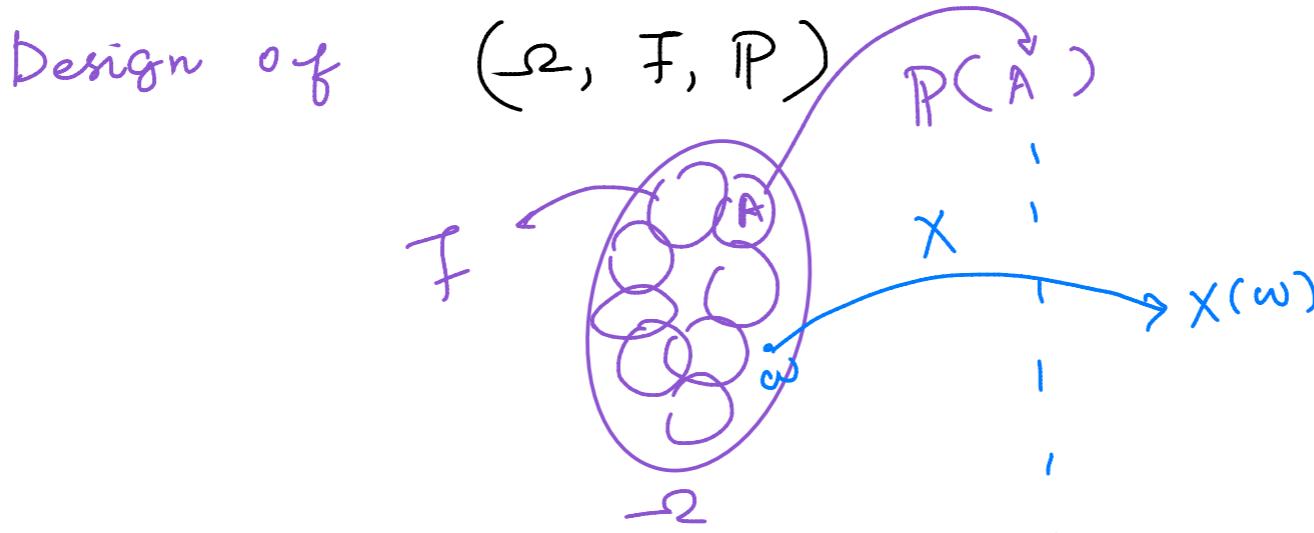
$$x_1 = \begin{cases} 1, & \text{if } \omega \text{ even} \\ 0, & \text{if } \omega \text{ odd} \end{cases}$$

$$x_2 = \begin{cases} 1, & \text{if } \omega > 3 \\ 0, & \text{if } \omega \leq 3 \end{cases}$$

$$\text{Prob}(x_1 = 1) = P(x_1^{-1}(1)) = P(\{2, 4, 6\}) = \frac{1}{2}$$

$$\text{Prob}(x_2 = 1) = P(x_2^{-1}(1)) = P(\{4, 5, 6\}) = ???$$

L
notin F



Example 1: Toss of a coin

$$\Omega = \{H, T\}$$

$$\mathcal{F} = \{\emptyset, \Omega, \{H\}, \{T\}\}$$

Pick any $A \in \mathcal{F}$

- $P(\emptyset) = 0$

- $P(\Omega) = 1$

- $P(\{H\}) = \frac{1}{2}$

- $P(\{T\}) = \frac{1}{2}$

$$P(A) = \frac{|A|}{2}$$

$|A| = \# \text{ elements in } A$

Exercise

$$\Omega = \{1, 2, 3, 4\}$$

Fill up equivalent entries

$$X(\omega) = \omega$$

Example 2:

Roll of a die

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$\mathcal{F}_{\text{simple}} = \{\emptyset, \Omega, \{1, 2, 3\}, \{4, 5, 6\}\}$$

$$\mathcal{F} = 2^{\Omega}$$

$$P(A) = \frac{|A|}{6}$$

$$X(\omega) = \omega$$

$$\text{Prob}(X(\omega) = 1) = P(\{1\}) = \frac{1}{6}$$

$$\mathcal{C} = \mathcal{F} = \{\emptyset, \Omega, \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$$

$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\} \in \mathcal{F}_{\text{Monan}}$

$\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\} \notin \mathcal{F}_{\text{Monan}}$

$$\text{Issue 1: } \text{Prob}(2 \leq X \leq 3) = P(\{2, 3\})$$

Issue 2: σ -algebra generated by a collection of sets C

$$\mathcal{F}_C$$

- \mathcal{F}_C is the smallest σ -algebra that

contains C

In this example

σ -algebra generated by C_{mohan}

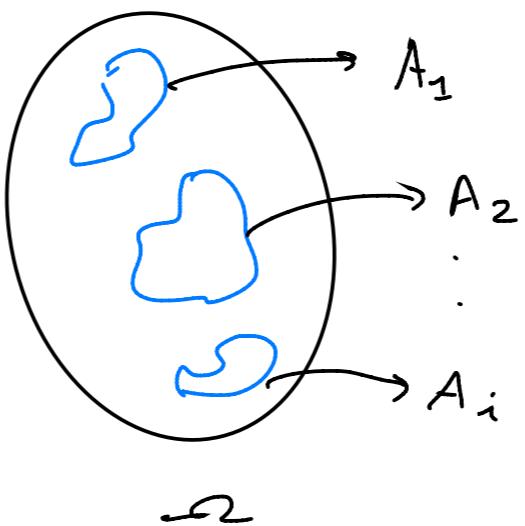
$$\cup \Omega$$

$$C_1 = \{\emptyset, \omega\} \Rightarrow \mathcal{F}_{C_1} = C_1$$

$$C_2 = \{\emptyset, \omega, \{1\}, \{2, 3, 4, 5, 6, 3\}\} \Rightarrow \mathcal{F}_{C_2} = \{\emptyset, \omega, \{1\}, \{2, 3, 4, 5, 6, 3\}\}$$

Goal 1: P will always not be as simple as $\frac{|A|}{2}$ or $\frac{|A|}{6}$
 \Rightarrow Non-uniform assignments

Goal 2: Till now $\mathcal{F} = 2^{\Omega}$ (power set = collection of all subsets)
 (cards, dice
 coins)



(identical
 shape/
 size)

Create a jigsaw of sets A_1, \dots, A_i, \dots

A_i, A_j such that $i \neq j$ $A_i \cap A_j = \emptyset$

$$\Rightarrow P(\bigcup_i A_i) = P(\Omega) = 1$$

$$= \sum_{i=1}^{\infty} P(A_i)$$

(Ω, \mathcal{F}, P)

- Toss of a coin and roll of a die
- These two are independent

unbiased

$$\omega = (\omega_1, \omega_2)$$
$$\Omega = \{ (1, H), (2, H), \dots, (6, H) \\ (1, T), (2, T), \dots, (6, T) \}$$

$$\mathcal{F} = 2^\Omega$$

$$A = \{ (1, H), (2, T) \}$$

$$x_{\text{coin}}(\omega) = \omega_2$$

$$x_{\text{die}}(\omega) = \omega_1$$

$$P(A) = \frac{|A|}{|\Omega|}$$

$$\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

$$T = 2^{\Omega}$$

$$X_{\text{coin}}(\omega) = \begin{cases} H, & \omega \text{ is odd} \\ T, & \omega \text{ is even} \end{cases}$$

or

$$P(A) = \frac{|A|}{12}$$

$$X_{\text{coin}}(\omega) = \begin{cases} H, & 1 \leq \omega \leq 6 \\ T, & 7 \leq \omega \leq 12 \end{cases}$$

$$X_{\text{dice}}(\omega) = \begin{cases} 1, & \omega = 1 \text{ or } 7 \\ 2, & \omega = 2 \text{ or } 8 \end{cases}$$

$$= \begin{cases} 3, & \omega = 3 \text{ or } 9 \\ 4, & \omega = 4 \text{ or } 10 \\ 5, & \omega = 5 \text{ or } 11 \\ 6, & \omega = 6 \text{ or } 12 \end{cases}$$

.

.

$$\Omega = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$$

$$\mathcal{F} = 2^\Omega$$

Die is unbiased, coin is biased $\text{Prob}(H) = 0.6$

Die and coin are independent

P ???

$$P(\{1, \dots, 6\}) = 0.6$$

$$P(\{7, \dots, 12\}) = 0.4$$

$$P(\{1\}) = \frac{0.6}{6}$$

$$P(\{7\}) = \frac{0.4}{6}$$

$$P(\{6\}) = \frac{0.6}{6}$$

$$P(\{12\}) = \frac{0.4}{6}$$

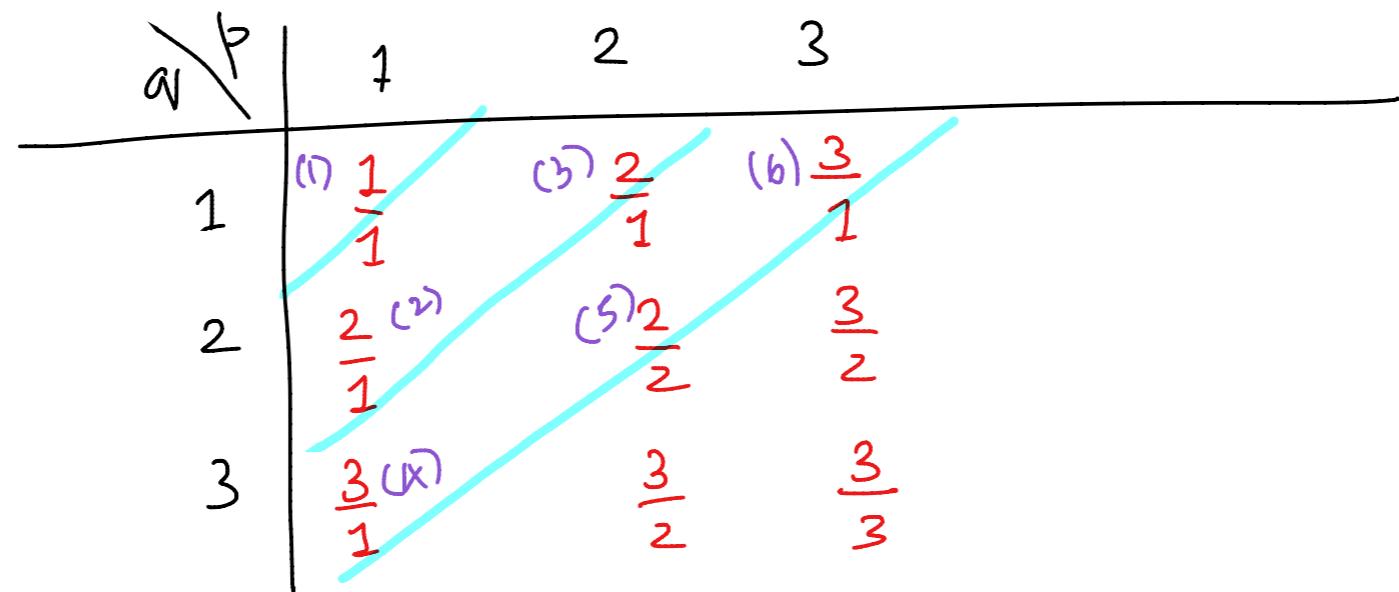
$$\begin{aligned}
 \text{Prob}(\text{Die} = 1) &= P(\{1, 3\}, \{7\}) = P(1, 3) + P(7) \\
 &= \frac{0.6}{6} + \frac{0.4}{6} \\
 &= \frac{1}{6}
 \end{aligned}$$

Exercise: Independent, both unbiased.

Feb 2

$x = p/q$
 $q \neq 0$

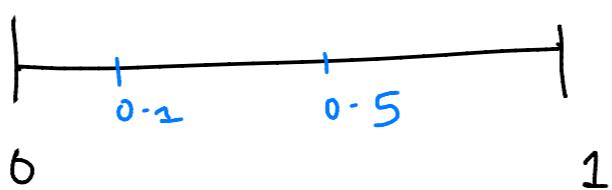
Property of rational numbers = they can be made to stand in a line/queue



Take $\Omega = [0, 1]$

$f \neq 2^{-x}$ (we want to demonstrate)

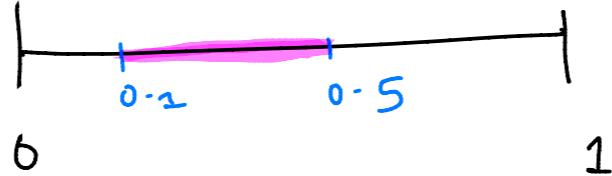
- $P([a, b]) = b - a$ (measure of an interval is its length) common sense



$$P([0.1, 0.5]) = 0.4$$

- $P(S + \alpha) = P(S)$ (translation invariance)

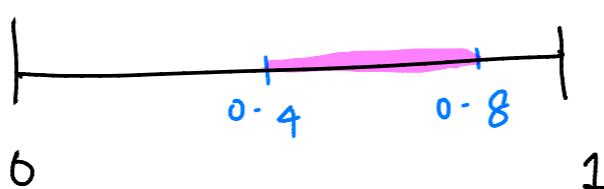
modulo 1



$$S = [0.1, 0.5]$$

$$\alpha = 0.3$$

$$S + \alpha = [0.1, 0.5] + 0.3$$

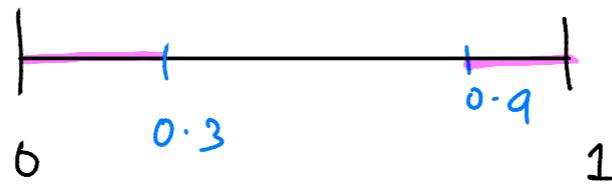


$$= [0.4, 0.8]$$

$$S = [0.1, 0.5]$$

$$\alpha = 0.8$$

$$S + \alpha = [0.9, 1] \cup [0, 0.3]$$



Definition of
\$\delta + \alpha\$

$$\begin{aligned}\delta + \alpha &= -\delta + \alpha && \text{if } \delta + \alpha < 1 \\ &= 1 - \delta + \alpha\end{aligned}$$

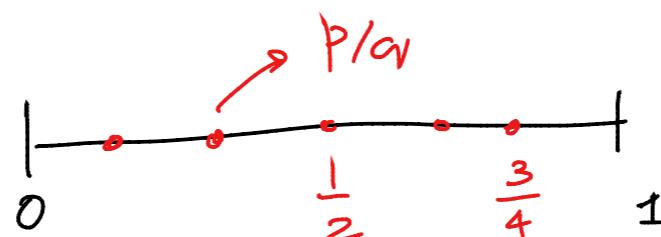
Goal: Create a jigsaw of $[0, 1]$ that is troublesome

Equivalence class (a set with the following property)

Set of things
that look alike

$$A_x = \{ y \in [0, 1] \mid y - x \text{ is a rational number} \}$$

$$A_0 = \{ y \in [0, 1] \mid y - 0 = y \text{ is a rational number} \}$$

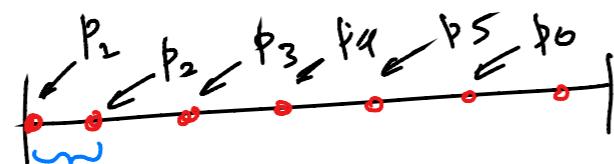


$A_{\frac{1}{2}} = \{ y \in [0, 1] \mid y - \frac{1}{2} \text{ is a rational number} \}$

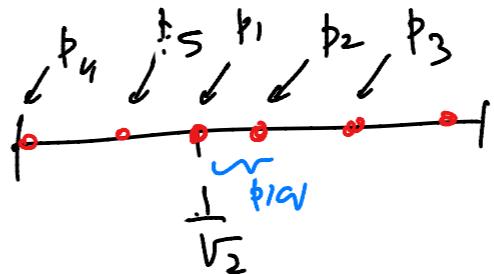
$$A_{\frac{1}{2}} = A_0 = A_{\frac{p}{q}}$$

$A_{\frac{1}{\sqrt{2}}} = \{ y \in [0, 1] \mid y - \frac{1}{\sqrt{2}} \text{ is rational} \}$

$$A_{\frac{1}{\sqrt{2}}} \cap A_0 = \emptyset$$



A_0



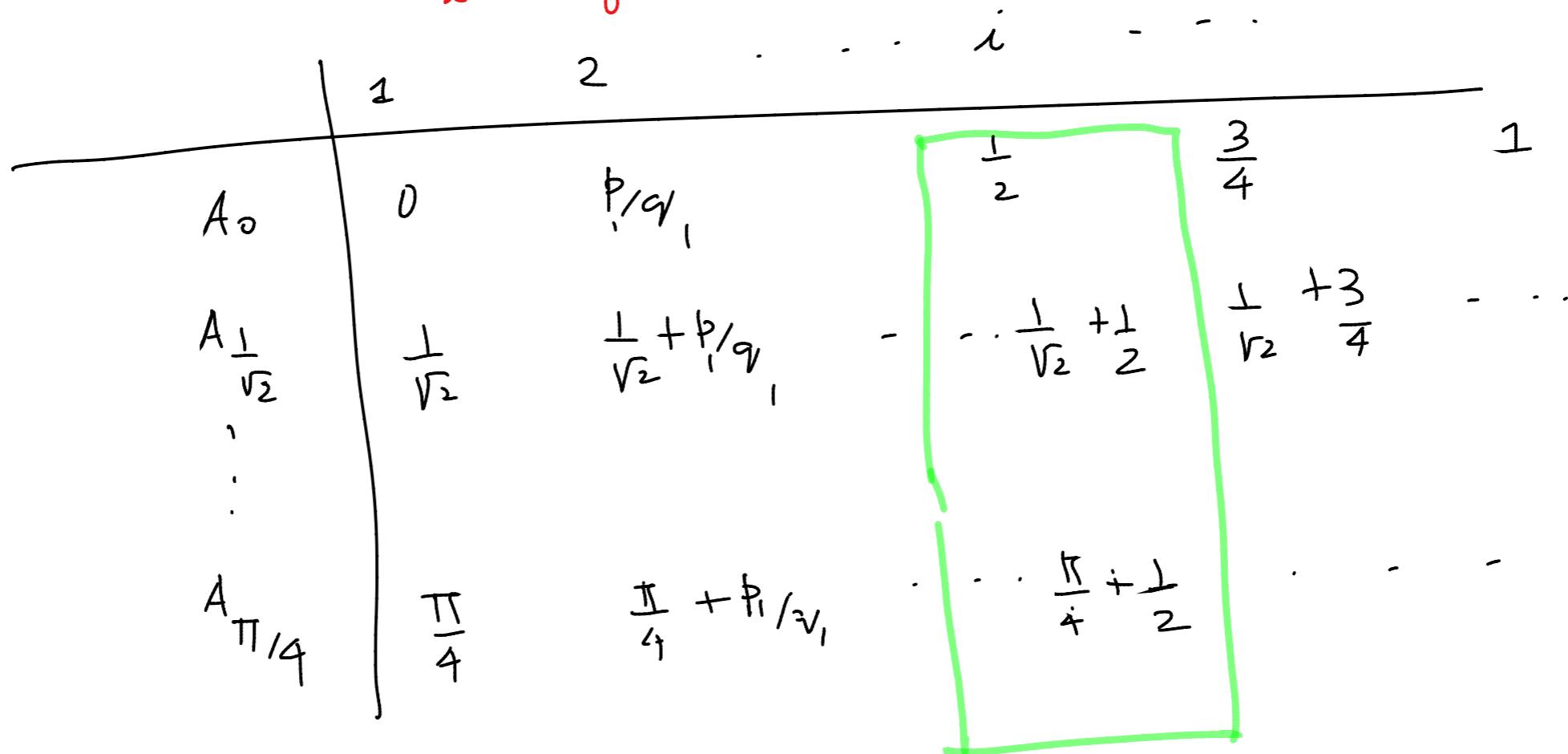
- $A_x = \text{irrational } x + \text{all rational numbers}$
- (base point)
- if you exceed 1
then we come back from 0
- see \oplus

- $A_{\frac{\pi}{4}}$, $A_{\frac{\pi}{6}}$, $A_e \frac{1}{4}$

- $A_{\frac{\pi}{4} + \frac{1}{2}} = A_{\frac{\pi}{4}}$

Take x and y such that $x - y$ is irrational

$$A_x \cap A_y = \emptyset$$



Pick the i^{th} element from each A_x and form B_i

$$\bullet \quad \bigcup_{i=1}^{\infty} B_i = [0, 1]$$

$$\bullet \quad B_i \cap B_k = \emptyset, \quad i \neq k \quad (\text{jigsaw})$$

$$\bullet \quad P(B_i) = M = F(B_k)$$

$$\bullet \quad P\left(\bigcup_{i=1}^{\infty} B_i\right) = P(\Omega) = 1$$

$$\bullet \quad \sum_{i=1}^{\infty} P(B_i) = \sum_{i=1}^{\infty} M = \infty \cdot M \quad \cancel{+}$$

$B_i \subseteq 2^{\omega}$ but B_i are not measurable

Worksheet 1 :

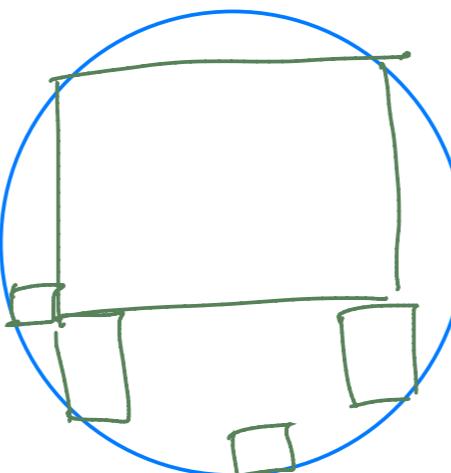
Goal : To construct a (Ω, \mathcal{F}, P) which supports two random variables

Given : $F_{X_1, X_2}(x_1, x_2) = \text{Prob}(X_1 \leq x_1, X_2 \leq x_2)$

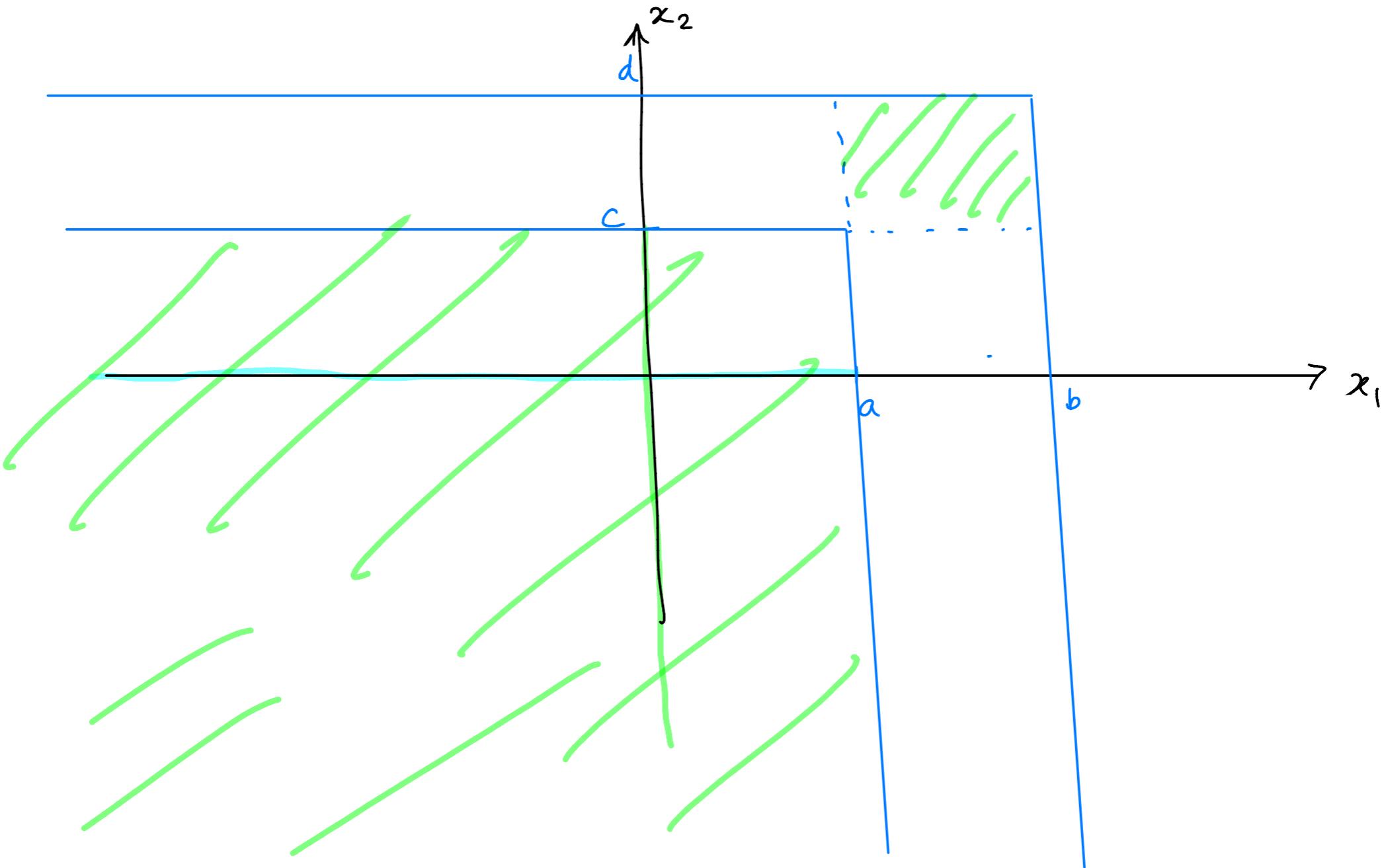
$(-\infty, a] = \bigcup_{n=1}^{\infty} (-\infty, a - \frac{1}{n}]$

function name function arguments

- Pick $\Omega = \mathbb{R} \times \mathbb{R} = \mathbb{R}^2$, $\omega \in \Omega \rightarrow (\omega_1, \omega_2)$
- Borel sets : $(-\infty, a] \times (-\infty, c]$

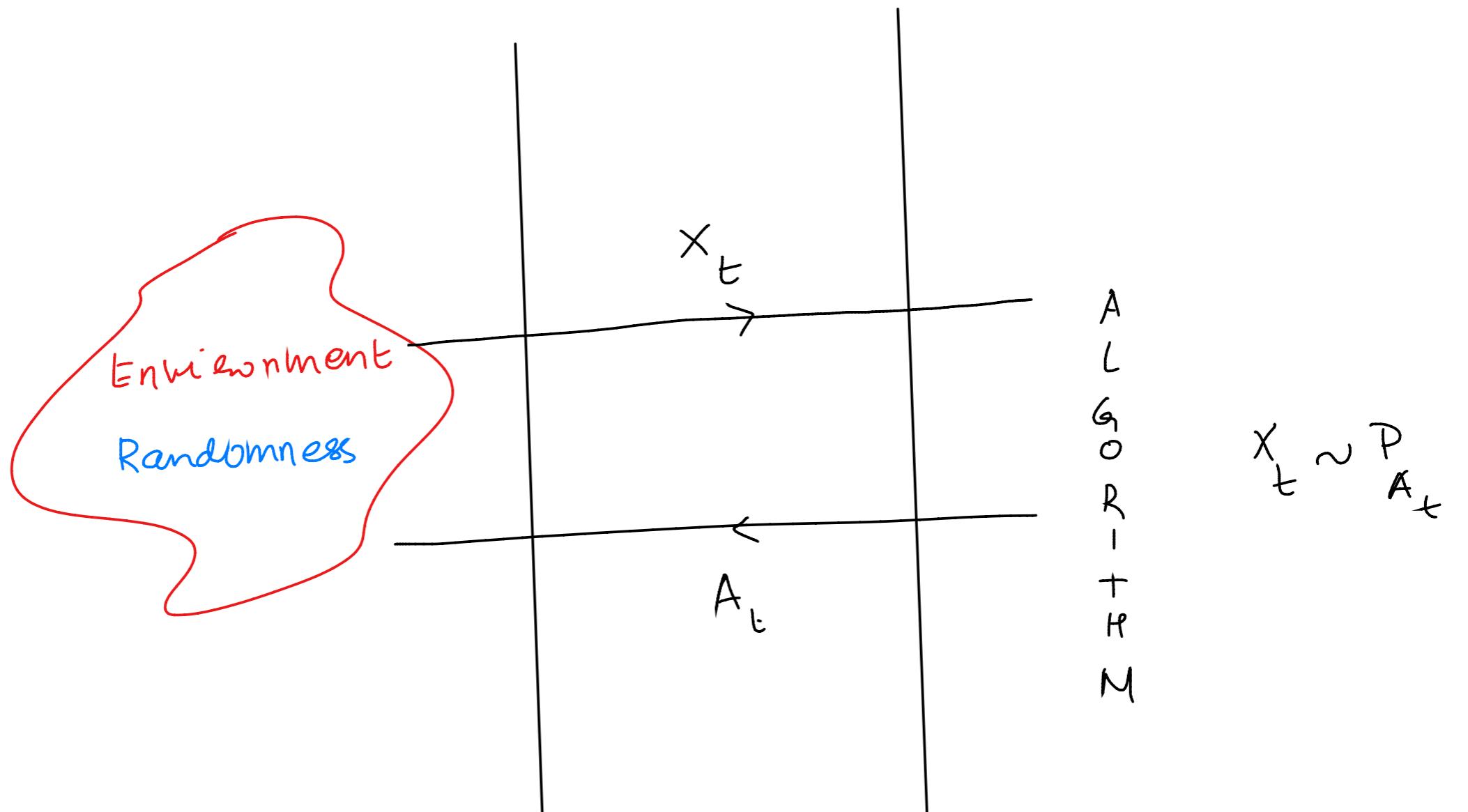


- $X_1(\omega) = \omega_1, X_2(\omega) = \omega_2$
- $\text{Prob}(X_1 \leq a, X_2 \leq c) = P((-\infty, a] \times (-\infty, c])$
 $= F_{X_1, X_2}(a, c)$



- $\text{Prob}(a < x_1 \leq b, c < x_2 \leq d) = P(a, b] \times (c, d])$
 $= F_{x_1 x_2}(b, d) - F_{x_1 x_2}(b, c) - F_{x_1 x_2}(a, d) + F_{x_1 x_2}(a, c)$

Given x_1, \dots, x_n and F_{x_1, x_2, \dots, x_n} we can construct
 (Ω, \mathcal{F}, P) that supports these random variables

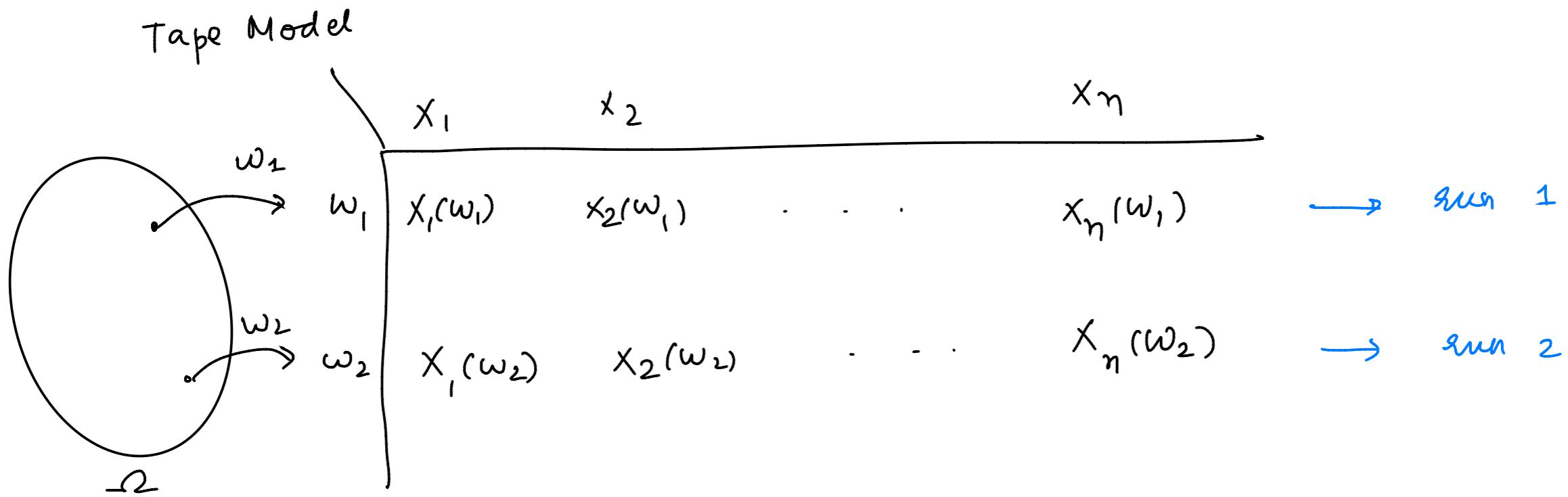


Goal: Even before we look x_t generated via
at algorithmic interaction we want to look at

X_t as a sequence without any algorithm for se

Sequence of random variables

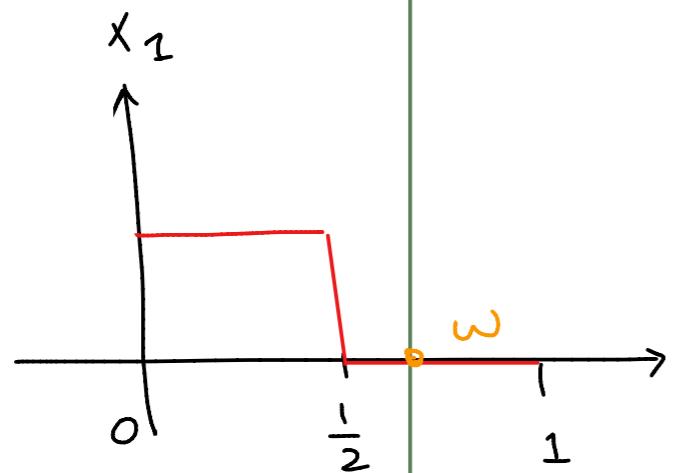
$$\{x_n\}_{n \geq 0} = x_1, \dots, x_n$$



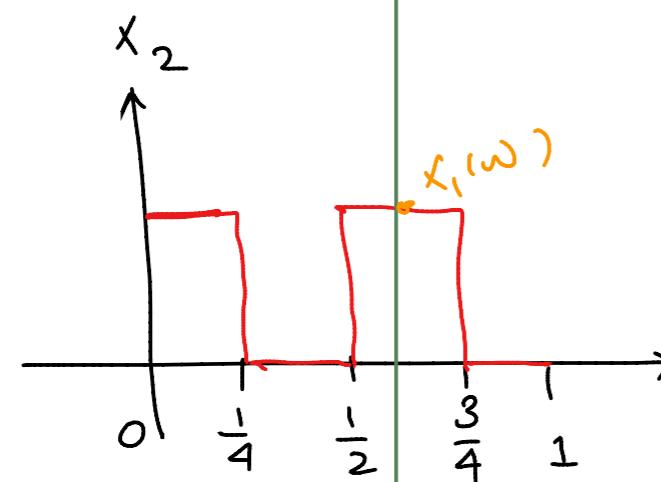
Nature realises a $w \in \Omega$ and reads out from
the tape $x_1(w), \dots, x_n(w)$

Pick $\Omega = [0, 1]$, \mathcal{F} = Borel sigma algebra, P = length measure

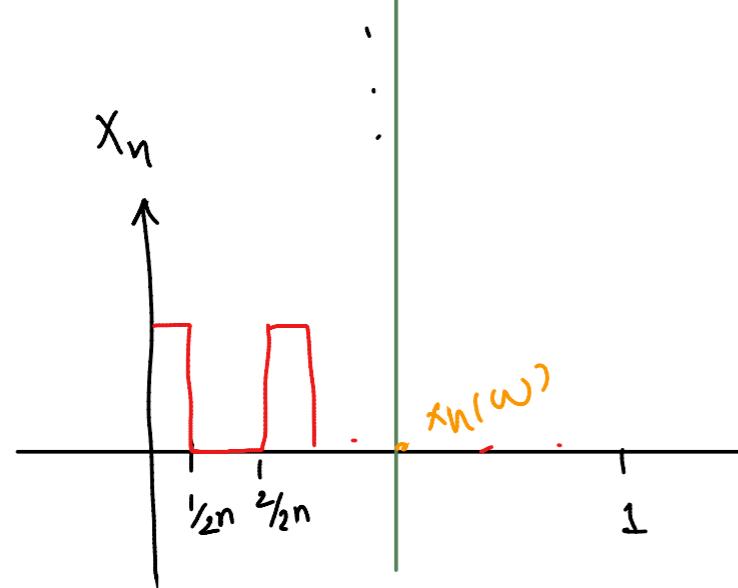
$$P(a, b) = b - a$$



$$\begin{aligned} x_1(\omega) &= 1, \quad \omega \in [0, \frac{1}{2}) \\ &= 0, \quad \omega \in [\frac{1}{2}, 1] \end{aligned}$$

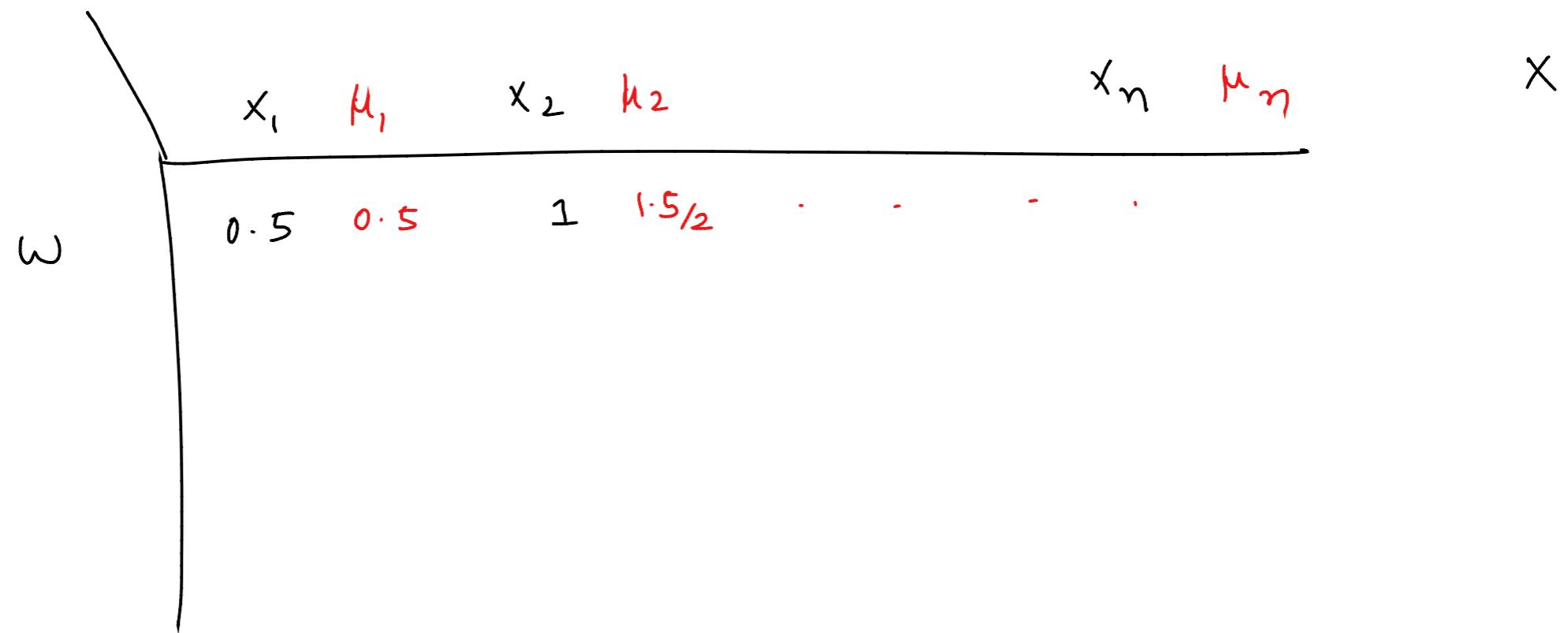


$$\begin{aligned} x_2(\omega) &= 1, \quad \omega \in [0, \frac{1}{4}) \\ &= 0, \quad \omega \in [\frac{1}{4}, \frac{1}{2}) \\ &= 1, \quad \omega \in [\frac{1}{2}, \frac{3}{4}) \\ &= 0, \quad \omega \in (\frac{3}{4}, 1] \end{aligned}$$



$$\begin{aligned} x_n(\omega) &= 1, \quad \omega \in [0, \frac{1}{2^n}) \\ &= 0, \quad \omega \in [\frac{1}{2^n}, \frac{2}{2^n}) \end{aligned}$$

Sample Mean $\bar{x}_n = \frac{x_1 + \dots + x_n}{n}$



Convergence of random variables

- Almost sure (a.s.) / Pointwise convergence

$\{X_n\}_{n \geq 0}$ be a sequence of random variables

$$X_n \xrightarrow{\text{a.s.}} X$$

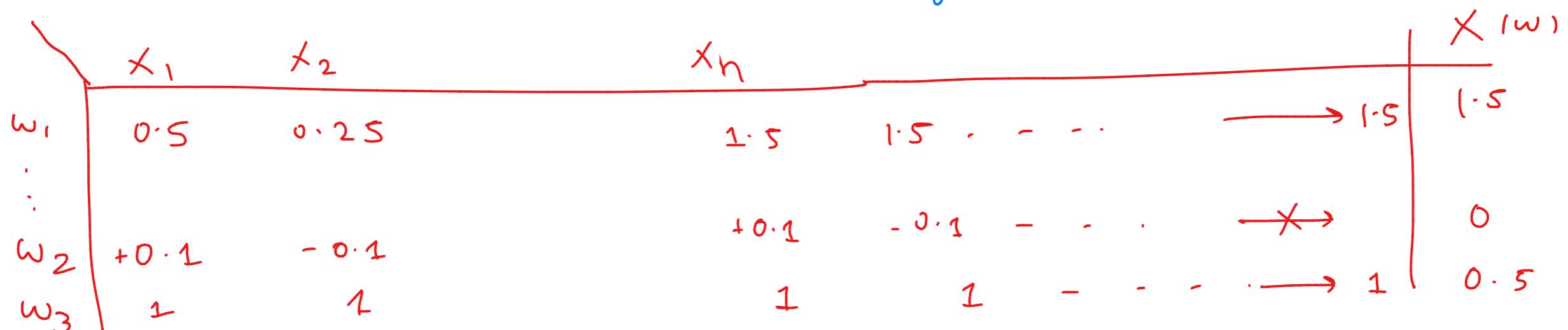
(X_n converges to X almost surely)

$$\text{Prob} \left(\lim_{n \rightarrow \infty} X_n = X \right) = 1$$

|||

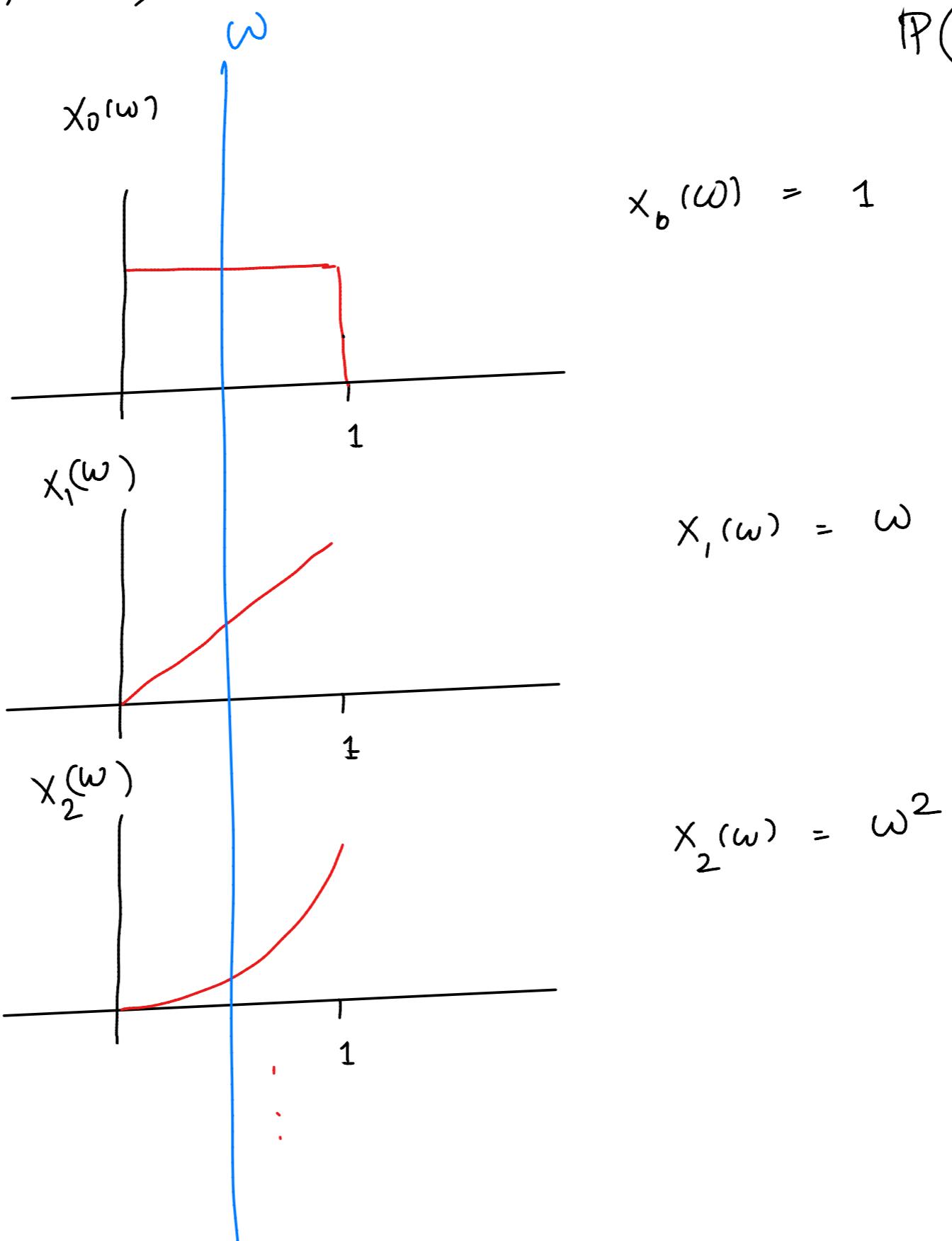
$$P \left(\{ \omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega) \} \right) = 1$$

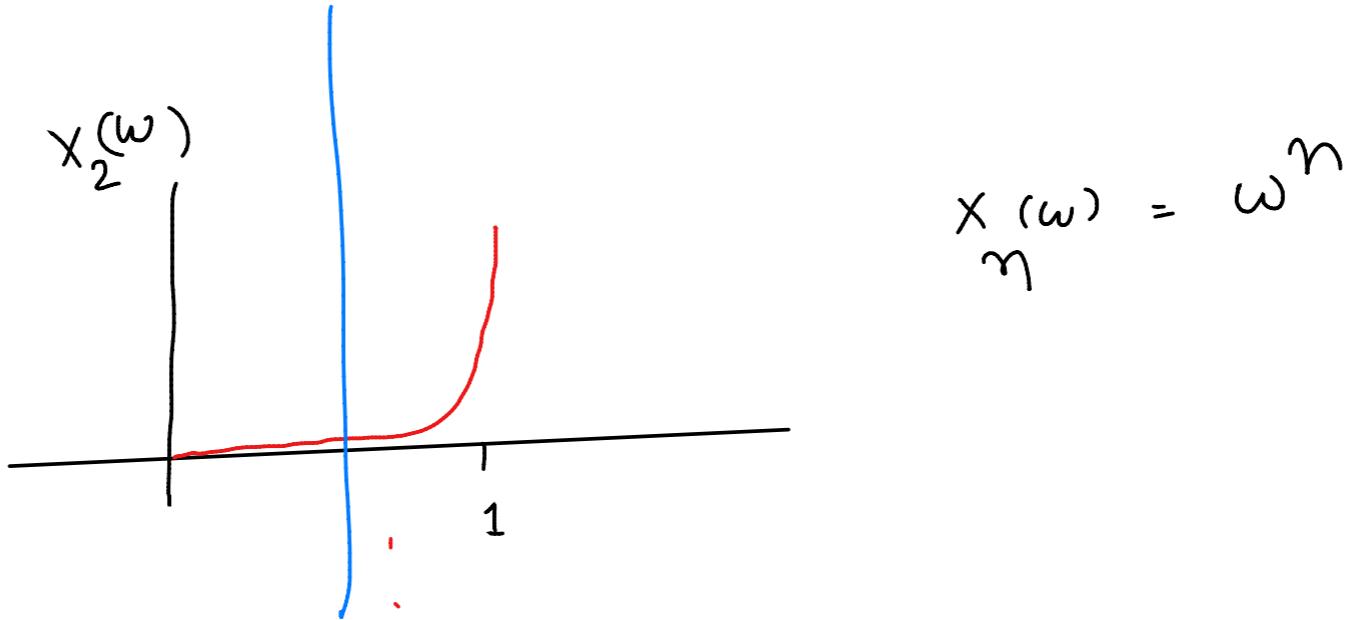
agreement



Example of almost sure convergence

$\Omega = [0, 1]$, \mathcal{F} = Borel sigma algebra, P = length measure
 $P([a, b]) = P((a, b)) = b - a$





$$x_n(\omega) = \omega^n$$

$$x_0(\omega), x_1(\omega), x_2(\omega) \cdots \rightarrow 0$$

$$\omega = 1$$

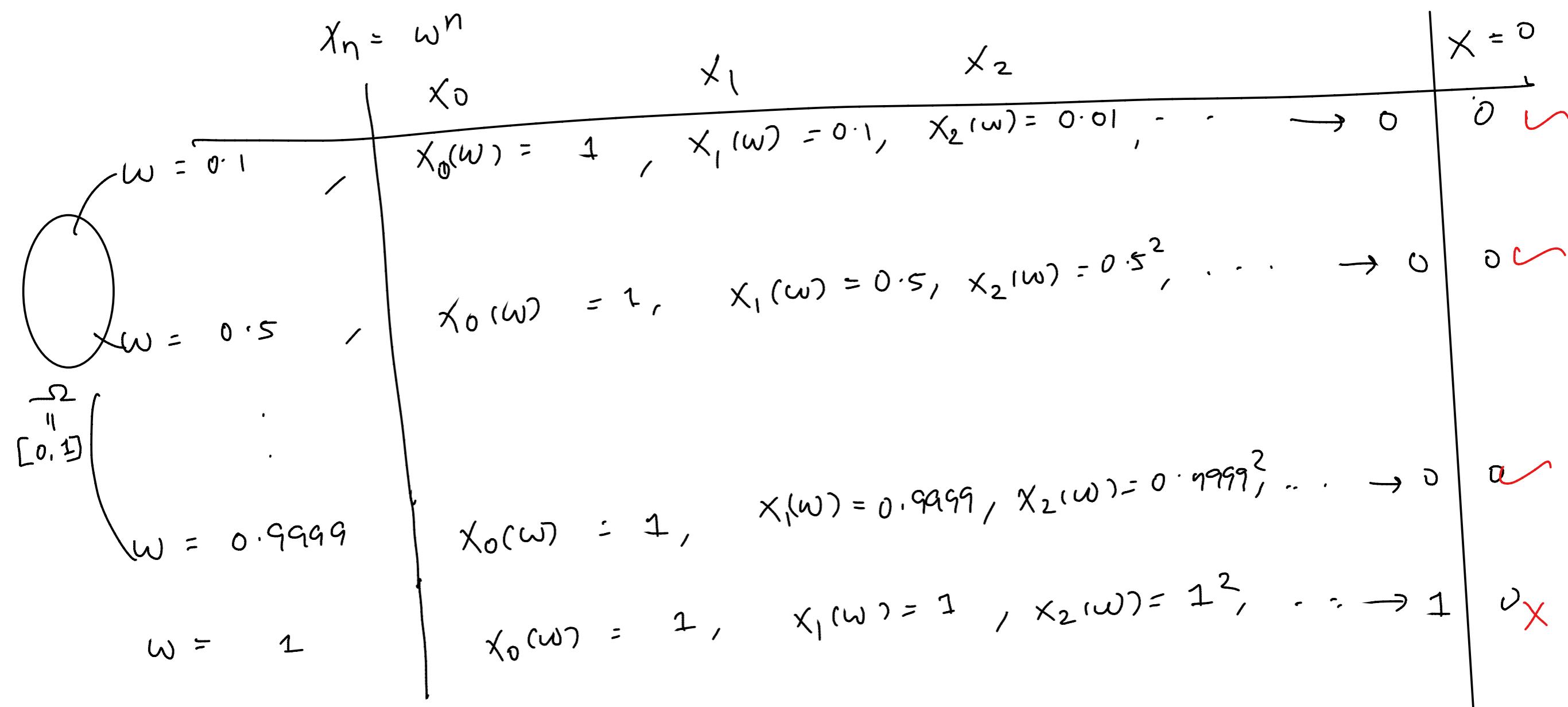
$\not\rightarrow 0$

$$x_n \xrightarrow{a.s.} 0$$

$$P(\{\omega\}) = 0$$

For every $\omega \in \Omega$, $\{X_n(\omega)\}$ is a sequence of real numbers

$$\lim_{n \rightarrow \infty} X_n(\omega) \longrightarrow X(\omega)$$

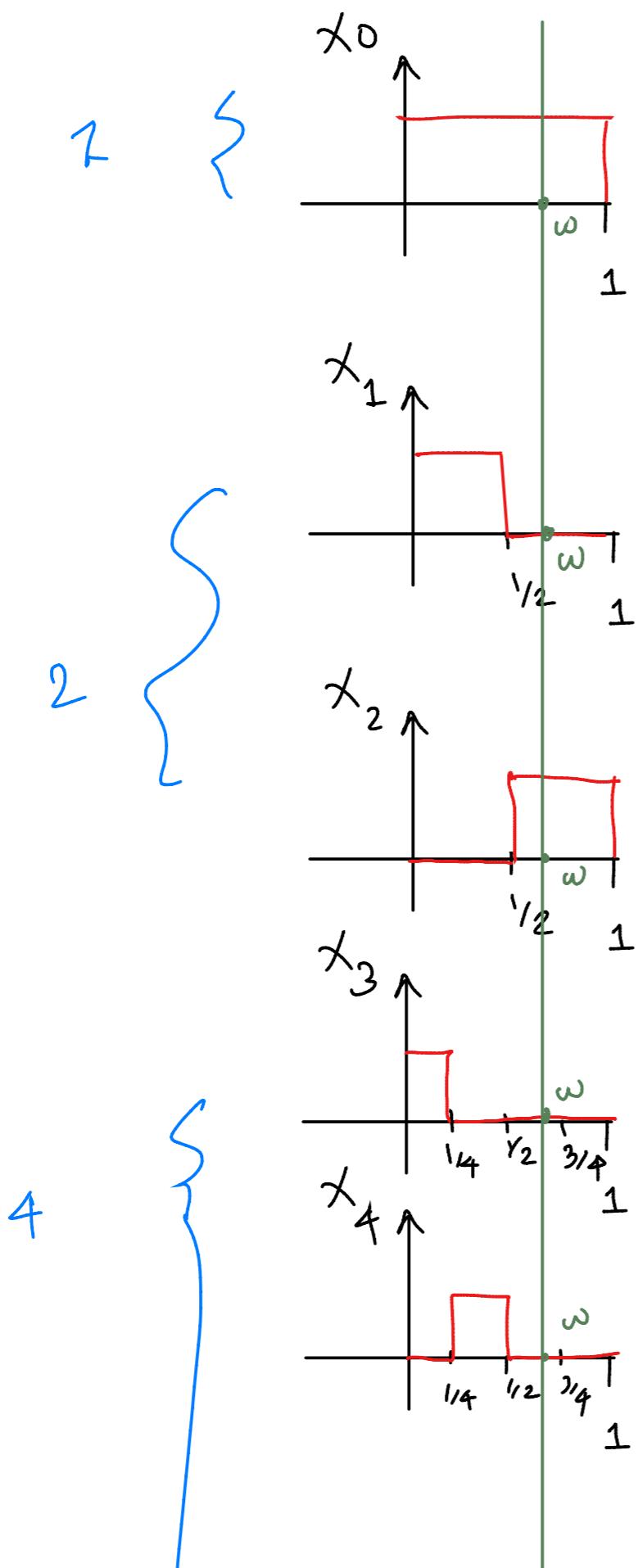


Example 2 : Shrinking Moving Rectangles

$$\Omega = [0, 1]$$

\mathcal{F} = Borel - sigma algebra

P : length

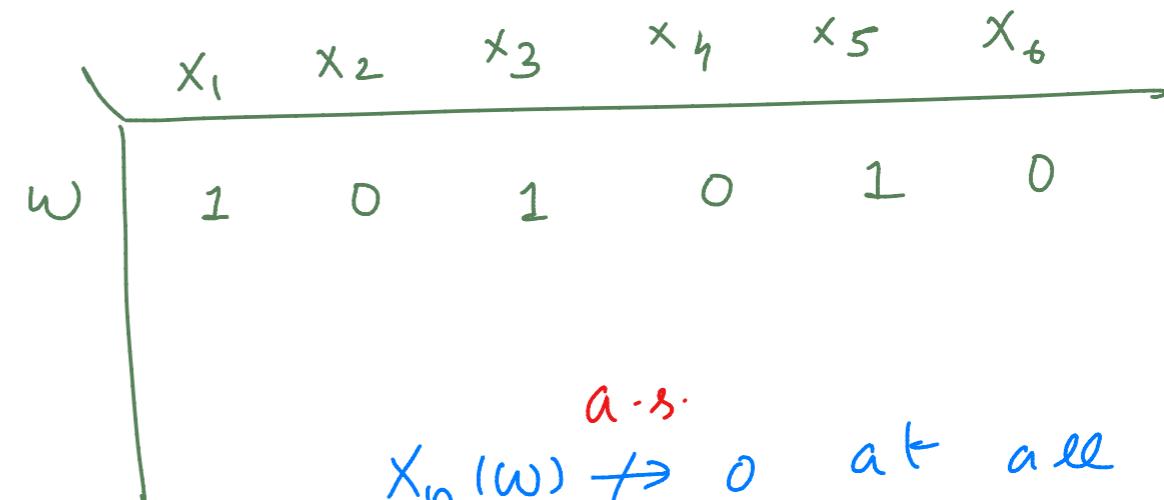


$$k = 0, 1, 2, \dots$$

$$i = 0, 1, \dots, 2^k - 1$$

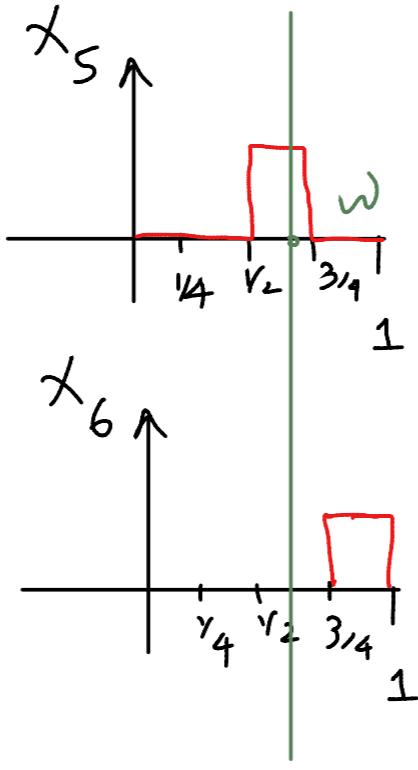
$$x_{\frac{k}{2}+i}(\omega) = 1, \omega \in \left[\frac{i}{2^k}, \frac{i+1}{2^k}\right]$$

$$= 0, \text{ elsewhere}$$

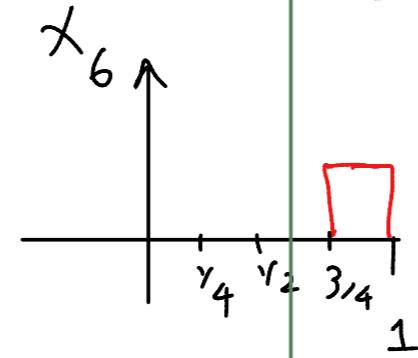


a.s.
 $x_n(\omega) \not\rightarrow 0$ at all for any ω

$x_n(\omega)$ does not converge because it takes value 1 infinitely often.



$x_n \xrightarrow{P} x$



2^k

For given
 k and ω

$$\text{Prob} \left(X_{2^k+i} = 1 \right) = \frac{1}{2^k}$$

as $k \rightarrow \infty$

$$\text{Prob} \left(X_{2^k+i} = 0 \right) = 1 - \frac{1}{2^k} \rightarrow 1 \quad \text{as } k \rightarrow \infty$$

$$\text{P} \left(\{ \omega : \lim_{n \rightarrow \infty} X_n(\omega) = 0 \} \right) = \text{P}(\varnothing) = 0$$

convergence in Probability

$\{X_n\}_{n \geq 0}$ are said to converge in probability
to a random variable X

$$X_n \xrightarrow{P} X$$

for any $\epsilon > 0$,

$$\text{Prob}(|X_n - X| > \epsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

$\lim_{n \rightarrow \infty}$

$$P(\{\omega: \lim_{n \rightarrow \infty} x_n(\omega) \neq x(\omega)\}) = 0$$

place of deviates

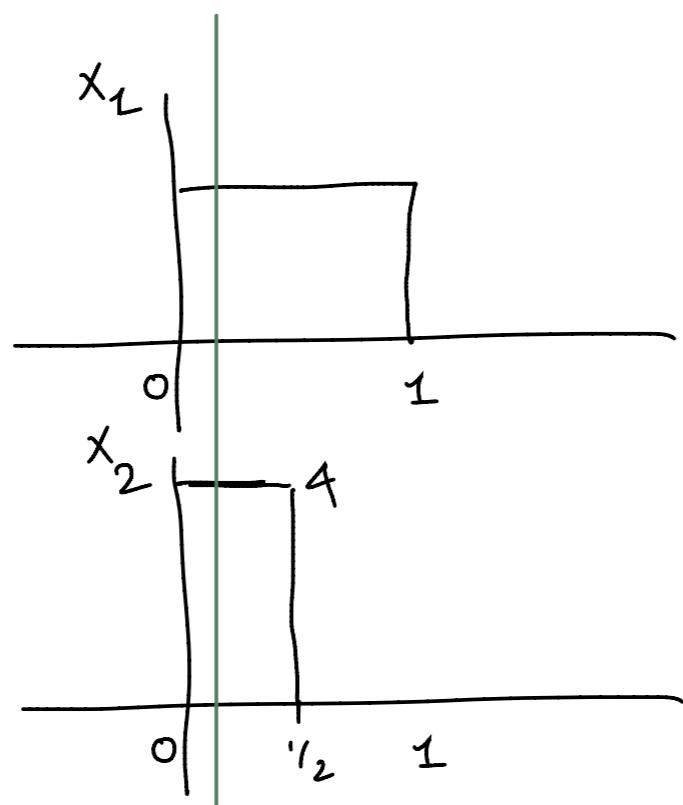
Mean Square Convergence

$$x_n \xrightarrow{m.s} x$$

$$\mathbb{E}[|x_n - x|^2] \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

amount of deviation

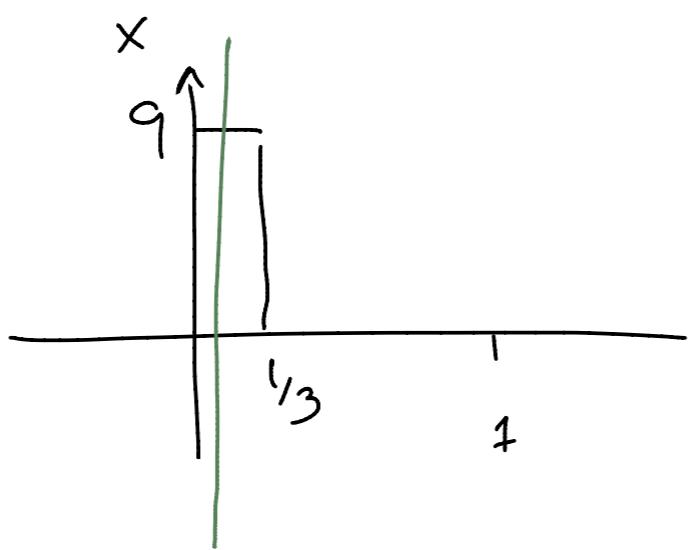
$$x_n(\omega) = n^2, \quad \omega \in [0, \frac{1}{n}]$$



$$x_n \xrightarrow{a.s} 0$$

$$x_n(\omega) \rightarrow 0 \quad \forall \omega \in (0, 1]$$

$$x_n(\omega) = 1 \quad \omega = 0$$



$$E [|k_n - b|^2] = (n^2)^2 \cdot \frac{1}{n} = n^3 \not\rightarrow 0$$

Convergence in distribution

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_x(x)$$

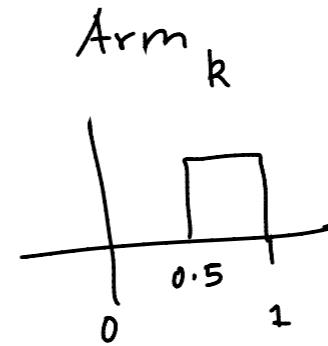
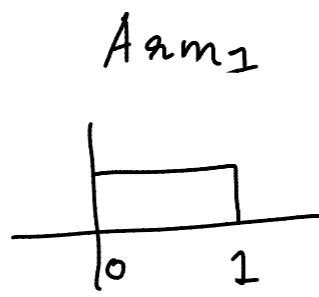
$\{X_n\}_{n>0}$ is iid

Gwal : x_1, x_2, \dots, x_n iid with $\mathbb{E}[x_n] = \mu$

$$\widehat{\mu}_n = \frac{x_1 + \dots + x_n}{n}$$

$\widehat{\mu}_n \rightarrow \mu$ how fast does this happen.

Bandit Problem

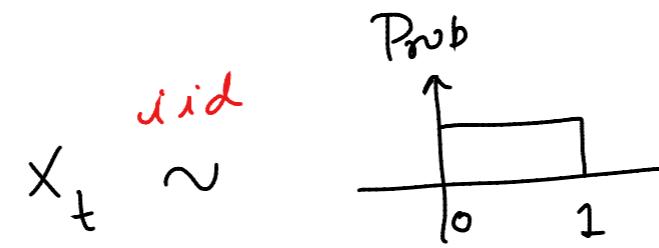


at t , we choose arm A_t

we observe $X_t \sim P_{A_t}$ (every time say $A_t = a$
we see an independent sample from distribution P_a)

for making things simple let us say only one arm is there

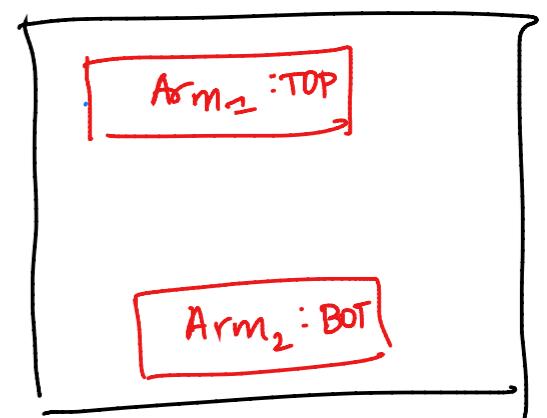
$$A_t = 1, \quad X_t \sim P_1$$

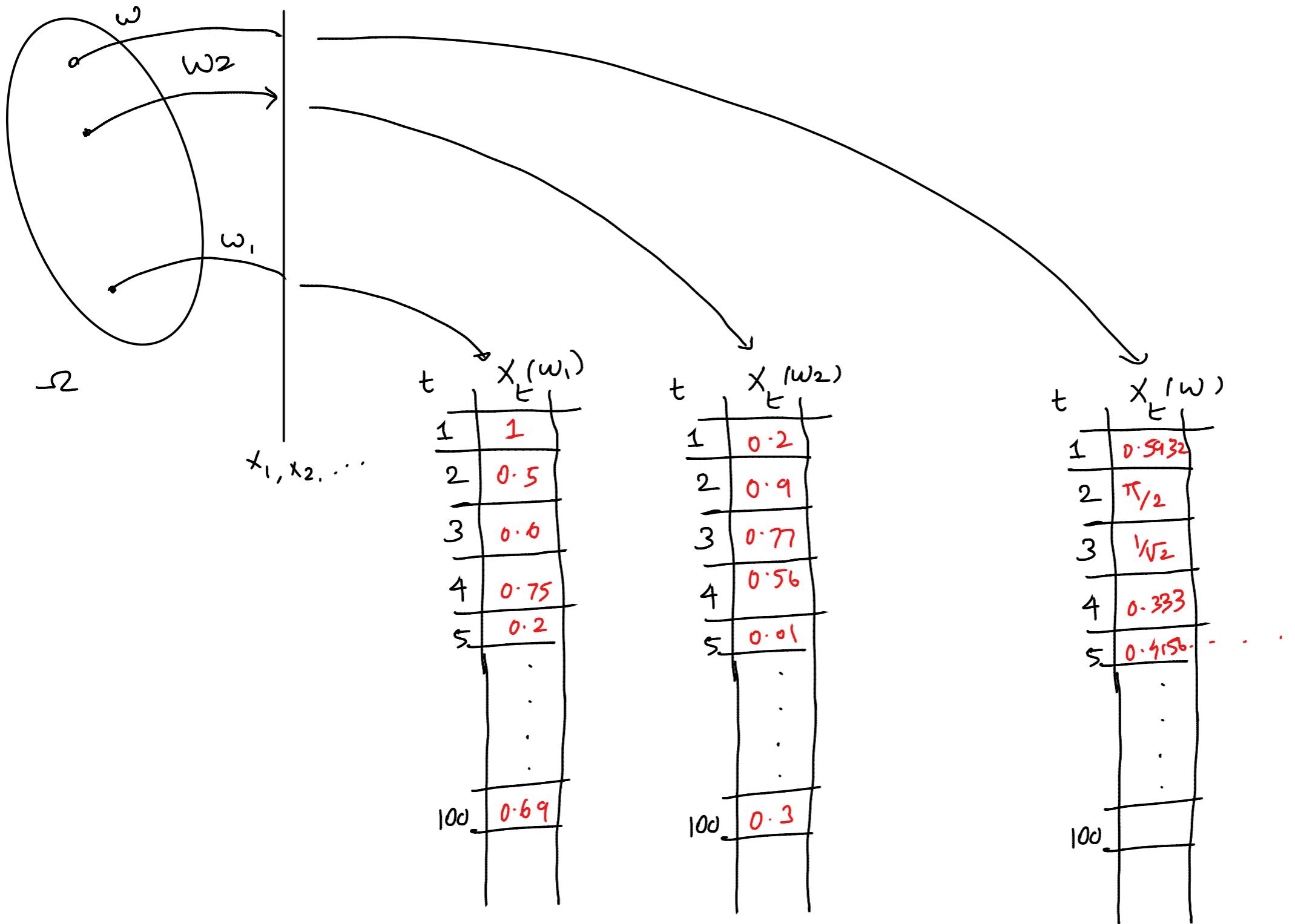


for a real world example:
pretend that this is
the revenue for placing
ad in top location
of a website

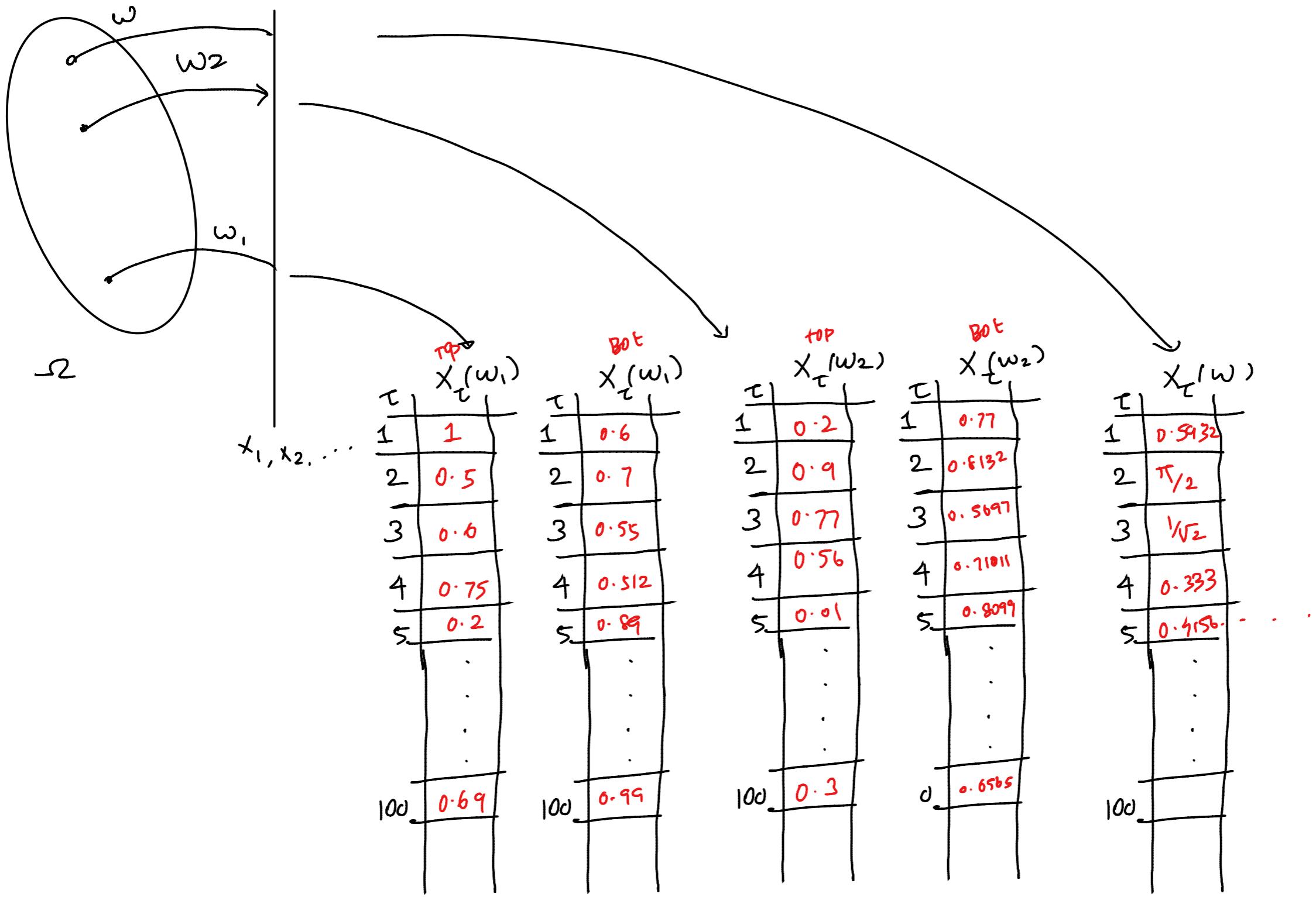
Mental Model

- Stack Model
- Tape Model





t	1	2	3	4	5			100
ω_1	x_t	1	0.5	0.6	0.75	0.2	-	-
ω_2	x_t	0.2	0.9	0.77	0.56	0.01		0.3



current time

↓

t

	1	2	3	4	5	6
ω_1	x_t^{TOP}	1	0.5	0.6	0.75	
	x_t^{BOT}	0.6	0.7	0.55	0.512	
	A_t	1				
ω_2	x_t^{TOP}	0.2	0.9	0.77	0.58	
	x_t^{BOT}	0.77	0.8132	0.5697	0.71011	
	A_t					

Current time

Time

\downarrow

t

w

x_t^{TOP}

x_t^{BOT}

A_t

x_t^{TOP}

x_t^{BOT}

A_t

w_1

w_2

Scenario

$1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6$

	1	2	3	4	5	6
x_t^{TOP}	1	0.5	0.6	0.75		
x_t^{BOT}	0.6	0.7	0.55	0.512		
A_t	1					
x_t^{TOP}	0.2	0.9	0.77	0.58		
x_t^{BOT}	0.77	0.8132	0.5697	0.71011		
A_t	1	2				

Current time

Time



Scenarios

t

	1	2	3	4	5	6
ω_1	x_t^{TOP} 1	0.5	0.6	0.75		
	x_t^{BOT} 0.6	0.7	0.55	0.512		
ω_2	A_t 1					
	x_t^{TOP} 0.2	0.9	0.77	0.58		
	x_t^{BOT} 0.77	0.8132	0.5697	0.7101		
	A_t <u>0.2</u> 1	<u>0.2</u> 2	<u>0.77</u>			

current time

Time



t

	1	2	3	4	5	6
ω						
x_t^{TOP}	1	0.5	0.6	0.75		
x_t^{BOT}	0.6	0.7	0.55	0.512		
A_t	1					
ω_1						
x_t^{TOP}	0.2	0.9	0.77	0.58		
x_t^{BOT}	0.77	0.8132	0.5697	0.71011		
A_t	<u>0.2</u>	<u>0.2</u> <u>0.77</u>	—	—		
ω_2		1	2	2		

Scenario



current time

↓

t	1	2	3	4	5	6
ω_1	x_t^{TOP}	1	0.5	0.6	0.75	
	x_t^{BOT}	0.6	0.7	0.55	0.512	
ω_2	A_t	1	2			
	x_t^{TOP}					
	x_t^{BOT}					
	A_t					

Current time

time

Scenarios

t

ω

	1	2	3	4	5	6
ω	x_t^{TOP}	1	0.5	0.6	0.75	
	x_t^{BOT}	0.6	0.7	0.55	0.512	
ω_1	v_{t+1}^{top}	1	0.6	0.75	0.6	--
	A_t	1	2	1	1	
ω_2	x_t^{TOP}					
	x_t^{BOT}					
	A_t					

$$S_n = x_1 + \dots + x_n$$

$$\hat{\mu}_n = \frac{S_n}{n} \xrightarrow{?} \mu \text{ (true expected value / true mean)}$$

Convergence in Probability: $X_n \xrightarrow{P} X$

for any $\varepsilon > 0$,

$$\text{Prob}(|X_n - X| > \varepsilon) \rightarrow 0 \quad \text{as } n \rightarrow \infty$$

$$P(\{w : |X_n(w) - X(w)| > \varepsilon\}) \rightarrow 0$$

~~lim~~
 $n \rightarrow \infty$

$$X_n \xrightarrow{a.s.} X$$

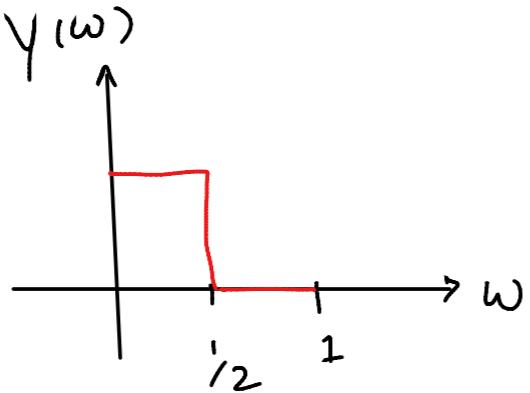
Convergence in
almost surely :

$$P(\{w : \lim_{n \rightarrow \infty} X_n(w) \neq X(w)\}) = 0$$

~~lim~~
 $n \rightarrow \infty$ place of deviation "across time"

keep w fixed, $\{X_n(w)\} \rightarrow X(w)$
real
number real
number

(Ω, \mathcal{F}, P)
 \downarrow length
 $[0,1]$ Borel measure



$$y(\omega) = \begin{cases} 1, & 0 \leq \omega \leq \frac{1}{2} \\ 0, & \frac{1}{2} < \omega \leq 1 \end{cases}$$

$$Y_1 = Y, \quad Y_2 = 1 - Y, \quad Y_3 = Y, \quad Y_4 = 1 - Y, \quad \dots$$

$$Y_1(\frac{1}{2}) = 1, \quad Y_2(\frac{1}{2}) = 0, \quad Y_3(\frac{1}{2}) = 1, \quad Y_4(\frac{1}{2}) = 0, \quad \dots \rightarrow$$

checking
for convergence
in almost
surely

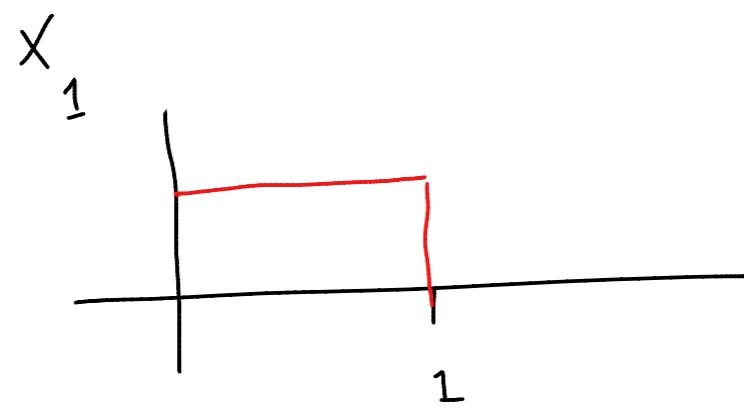
w: $\lim_{n \rightarrow \infty} Y_n(\omega)$ does not have a limit
n → ∞ ↳ checking across time

$$\mathbb{P}([0,1]) = 1$$

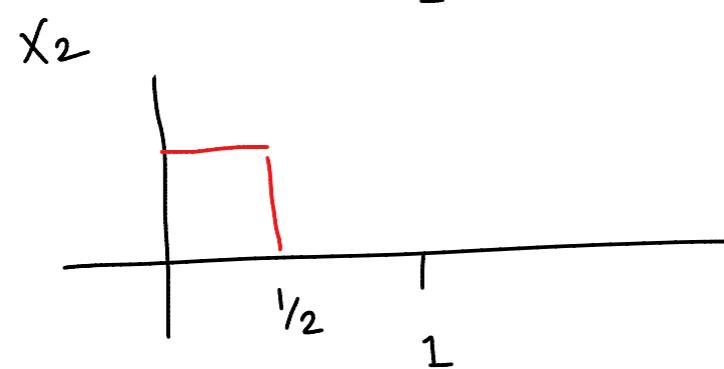
Fix $\epsilon = 0.1$

$$\omega: |\gamma_n(\omega) - 0| > \epsilon \quad n \text{ is odd}$$

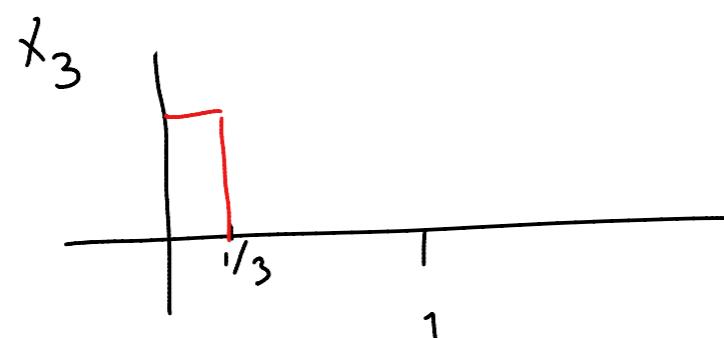
$$P([0, \frac{1}{2}]) = \frac{1}{2}$$



$$x_n = \begin{cases} 1, & \omega \in [0, \frac{1}{n}] \\ 0, & \text{elsewhere} \end{cases}$$



$$(Q_n) \text{ Does } x_n \xrightarrow{P} 0 \quad ?$$



$$\omega: |x_n(\omega) - 0| > \epsilon$$

$$P([0, \frac{1}{n}]) = \frac{1}{n}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} = 0$$

An) Yes

Qn) Does $X_n \xrightarrow{a.s.} 0$

$$\omega: \lim_{n \rightarrow \infty} X_n(\omega) \neq 0$$

$$\mathbb{P}(\{\omega\}) = 0$$

Weak Law of Large Numbers: $\{X_t\} \stackrel{iid}{\sim} P_X$, $\mathbb{E}_{P_X}[X_t] = \mu$
 (assume: $\mathbb{E}|X_t|^2 < \infty$)

$$\hat{\mu}_n = \frac{S_n}{n} = \frac{\overbrace{X_1 + \dots + X_n}^n}{n} \xrightarrow{P} \mu, \text{ as } n \rightarrow \infty$$

$$\frac{S_n - n\mu}{n} \xrightarrow{P} 0$$

$S_n - n\mu = (X_1 - \mu) + \dots + (X_n - \mu)$

Strong Law of Large Numbers:

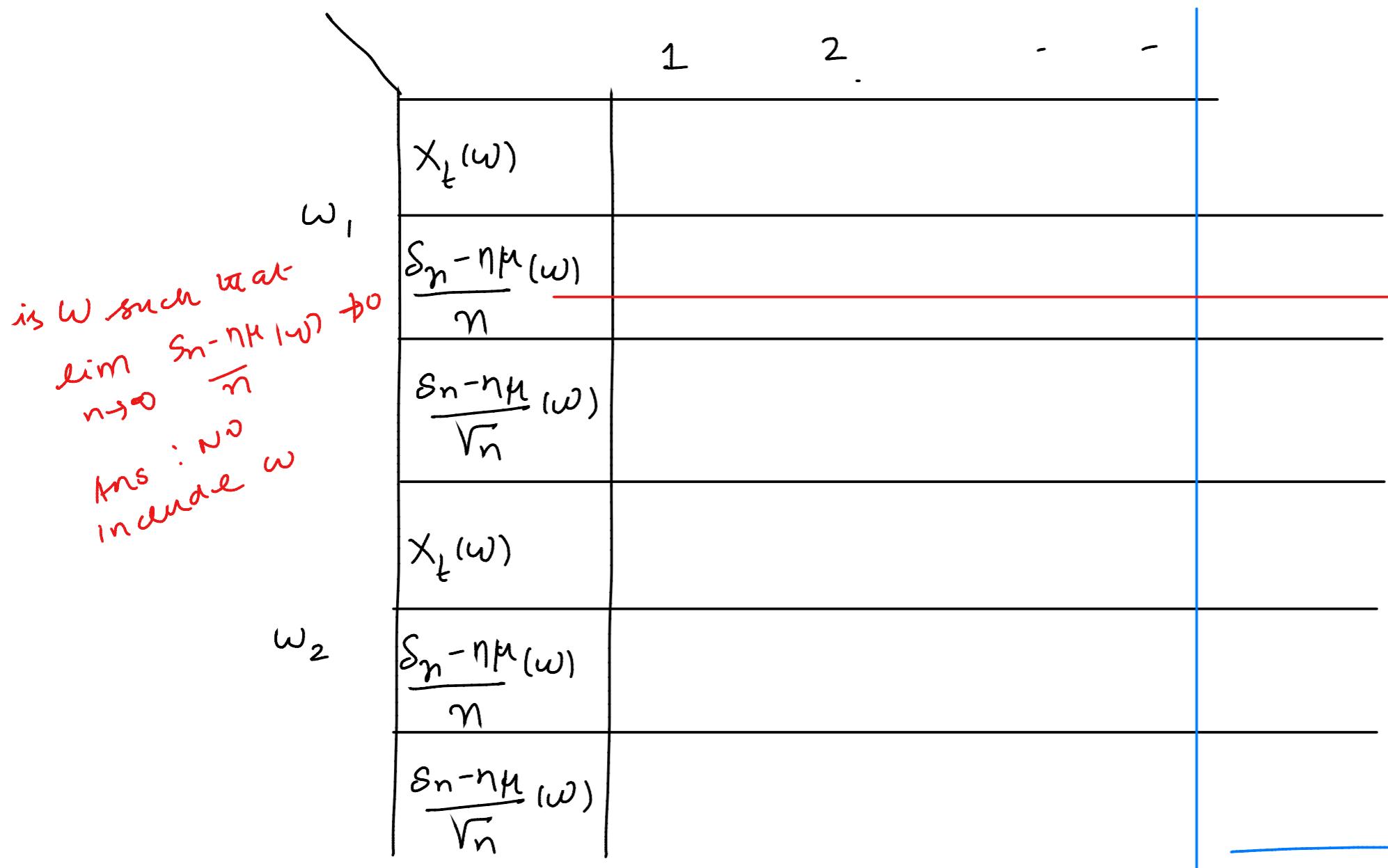
(assume: $\mathbb{E}[|X_t|^4] < \infty$)

$$\hat{\mu}_n = \frac{S_n}{n} = \frac{\overbrace{X_1 + \dots + X_n}^n}{n} \xrightarrow{a.s.} \mu, \text{ as } n \rightarrow \infty$$

$$\frac{S_n - n\mu}{n} \xrightarrow{a.s.} 0$$

Central Limit Theorem

$$\frac{S_n - n\mu}{\sqrt{n}} \xrightarrow{\text{dist}} \text{Normal}(0, \sigma^2)$$



is ω such that
 $\lim_{n \rightarrow \infty} \frac{S_n - n\mu(\omega)}{\sqrt{n}} \neq 0$
 Ans: No
 include ω

$P(\omega : \lim_{n \rightarrow \infty} X_n(\omega) \neq 0)$

$$P_n = P\left(\omega : \left| \frac{S_n - n\mu}{\sqrt{n}} \right| > \varepsilon\right)$$

$$\lim_{n \rightarrow \infty} P_n$$

We are interested in looking at convergence of sample mean to the true mean

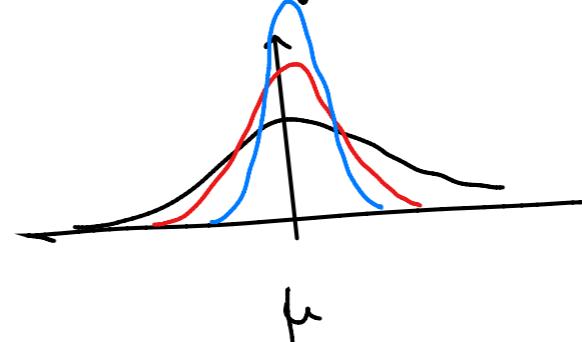
$$\{x_n\}_{n \geq 0} \stackrel{iid}{\sim} P \quad \underset{P}{E}[x_n] = \mu$$

Random variable
 $\rightarrow S_n = x_1 + \dots + x_n \quad (\text{sum})$

$\rightarrow \hat{\mu}_n = \frac{x_1 + \dots + x_n}{n} = \frac{S_n}{n} \quad (\text{sample mean})$

$$\hat{\mu}_n = \frac{S_n}{n} \rightarrow \mu$$

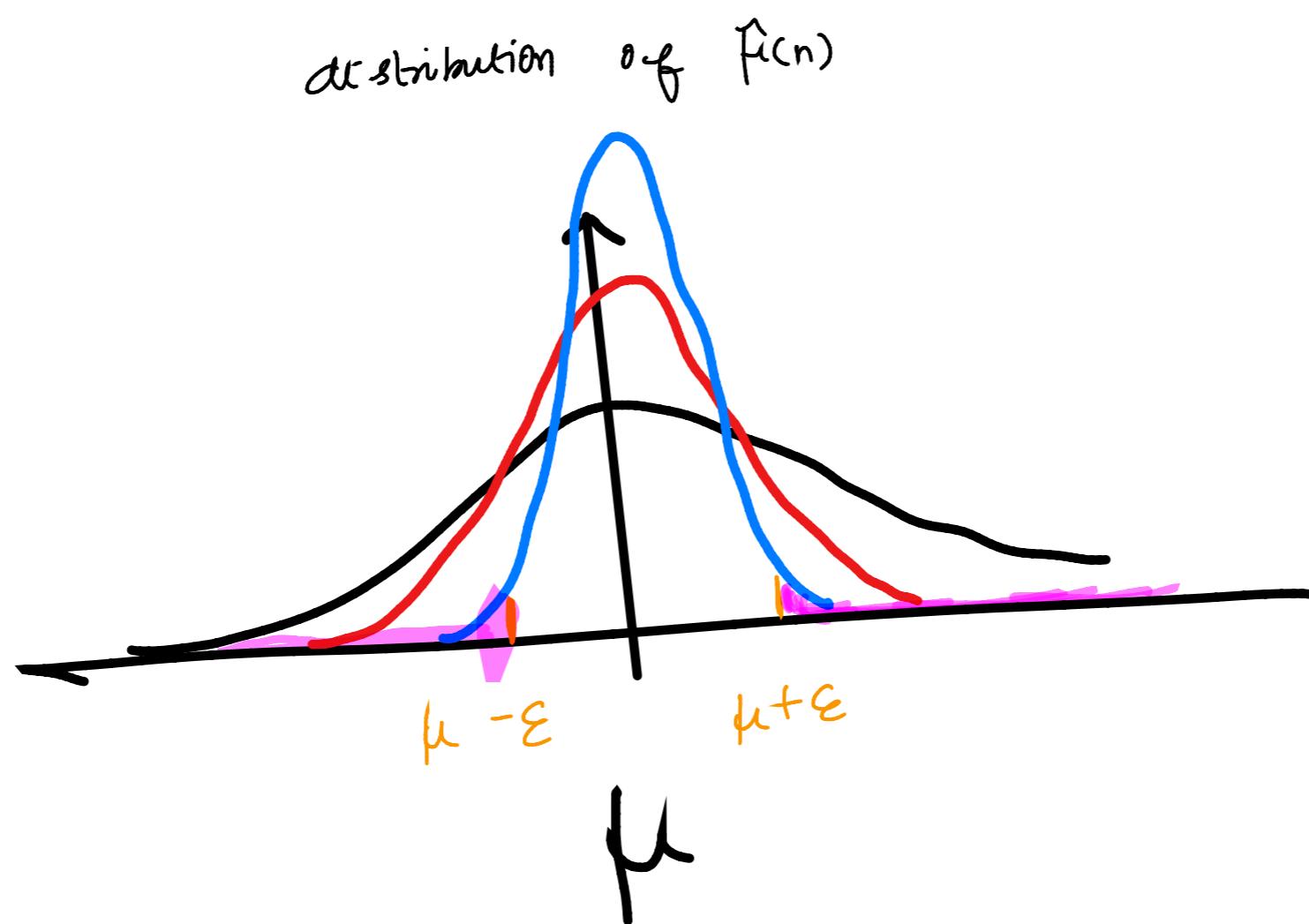
distribution of $f(n)$



as $n \rightarrow \infty$, more and more mass concentrates around true mean

$$\text{Prob}(|\hat{\mu}_n - \mu| > \varepsilon)$$

Tail Mass





- Label events happening in the tail to "low probability" events. For these just assume worst case happens.
- Rest of events we analyse deterministic as though $\hat{\mu} \in [\mu - \epsilon, \mu + \epsilon]$

$$\text{Regret}(n) = R_n = \mathbb{E} \left[\sum_{t=1}^n (\mu_* - x_t) \right], \quad x_t \sim P_{A_t}$$

$$= n\mu_* - \mathbb{E} \left[\sum_{t=1}^n x_t \right]$$

- Sub-optimality $\Delta_a = \mu_* - \mu_a$ ($\mu_a = \mathbb{E}_{x \sim P_a}[x]$)

- $T_a(n) = \sum_{t=1}^n \mathbb{I}_{\{A_t = a\}}$ (# times arm a was picked in n plays)

($T_a(n) \downarrow w$)

$$\begin{aligned} \mathbb{I}_{\{\text{condition}\}} &= 1 && \text{if condition is true} \\ &= 0 && \text{" false} \end{aligned}$$

$$\mathbb{E} \left[\sum_{t=1}^n (\mu_* - x_t) \right] = \sum_{t=1}^n \mathbb{E} [(\mu_* - \mu_t)]$$

Note at any given time one of the arms is definitely getting played

$$\mathbb{I}_{\{A_t = 1\}} + \mathbb{I}_{\{A_t = 2\}} + \dots + \mathbb{I}_{\{A_t = k\}} = 1$$

$$\mathbb{E}[(\mu_* - x_t)] = \mathbb{E}[(\mu_* - x_t) \cdot 1]$$

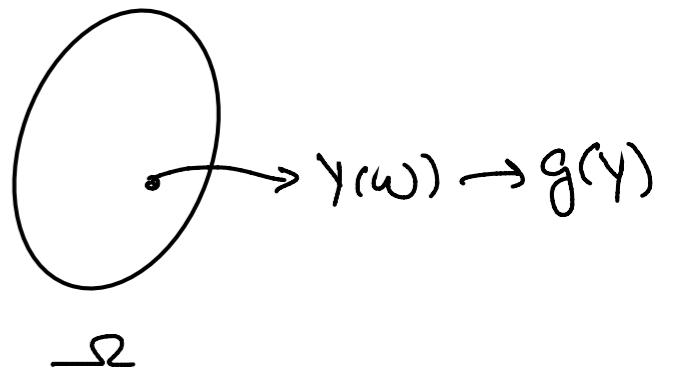
$$= \mathbb{E}[(\mu_* - x_t) \cdot (\mathbb{I}_{\{A_t = 1\}} + \dots + \mathbb{I}_{\{A_t = k\}})]$$

$$= \mathbb{E}\left[\sum_{a \in A} (\mu_* - x_t) \mathbb{I}_{\{A_t = a\}}\right]$$

$$R_n = \sum_{t=1}^n \mathbb{E}\left[\sum_{a \in A} (\mu_* - x_t) \mathbb{I}_{\{A_t = a\}}\right]$$

Tower Property of Expectation

$$\mathbb{E}[x] = \mathbb{E} \left[\underbrace{\mathbb{E}[x|y]}_{\text{function of } y} \right]$$



$$\mathbb{E}[\cdot] = \mathbb{E} [\mathbb{E} [\cdot | A_t]]$$

$$\mathbb{E} \left[\sum_{a \in A} (\mu_a - x_t) \mathbb{I}_{\{A_t = a\}} \right] = \mathbb{E} \left[\mathbb{E} \left[\sum_{a \in A} (\mu_a - x_t) \mathbb{I}_{\{A_t = a\}} \mid A_t \right] \right]$$

$$= \mathbb{E} \left[\sum_{a \in A} \mathbb{E} \left[(\mu_* - x_t) \mathbb{I}_{\{A_t=a\}} \mid A_t \right] \right]$$

$$\mathbb{E}[c \cdot g(x)] = c \cdot \mathbb{E}[g(x)]$$

$$\mathbb{E}[g(x) \cdot f(y) | y] = f(y) \cdot \mathbb{E}[g(x) | y]$$

↑ y is frozen

expectation averages out x given y.

$$\mathbb{E} \left[\sum_{a \in A} \underbrace{\mathbb{E}[(C\mu_* - x_t) \mathbb{I}_{\{A_t=a\}}]}_{g(x)} \underbrace{| A_t}_{f(y)} \right]$$

$$= \mathbb{E} \left[\sum_{a \in A} \mathbb{I}_{\{A_t=a\}} \mathbb{E}[(C\mu_* - x_t) | A_t] \right]$$

$$= \mathbb{E} \left[\sum_{a \in A} \mathbb{I}_{\{A_t=a\}} (\mu_* - \mu_{A_t}) \right]$$

$$= \mathbb{E} \left[\sum_{a \in A} \mathbb{I}_{\{A_t=a\}} (\mu_* - \mu_a) \right]$$

(so much trouble to reach this step so that we can fix the arm chosen)

$$= \mathbb{E} \left[\sum_{a \in A} \mathbb{I}_{\{A_t = a\}} \Delta_a \right]$$

$$R_n = \sum_{t=1}^n \mathbb{E} \left[\sum_{a \in A} (\mu_* - x_t) \mathbb{I}_{\{A_t = a\}} \right]$$

$$= \sum_{t=1}^n \mathbb{E} \left[\sum_{a \in A} \mathbb{I}_{\{A_t = a\}} \Delta_a \right]$$

$$= \mathbb{E} \left[\sum_{a \in A} \Delta_a \sum_{t=1}^n \mathbb{I}_{\{A_t = a\}} \right]$$

$$= \sum_{a \in A} \Delta_a \mathbb{E} \left[\sum_{t=1}^n \mathbb{I}_{\{A_t = a\}} \right]$$

$$= \sum_{a \in A} \Delta_a \mathbb{E} [T_a(n)]$$

- we pay a penalty Δ_a everytime we touch/play arm a
- we want to keep $T_{a(n)}$ as low as possible for
sub-optimal arms.
 - finite $T_{a(n)}$ will not give full knowledge of μ_a
 - question is at what rate should $T_{a(n)} \uparrow \infty$ as $n \uparrow \infty$
(sublinear is desired)

Tail Bound from CLT

$$\begin{aligned}
 \text{Prob} (| \bar{\mu}_n - \mu | > \varepsilon) &= \text{Prob} (| \frac{s_n}{n} - \mu | > \varepsilon) \\
 &= \text{Prob} (| \frac{s_n - n\mu}{\sqrt{n}} | > \varepsilon) \\
 &= \text{Prob} (| \frac{s_n - n\mu}{\sqrt{n}} | > \sqrt{n} \varepsilon)
 \end{aligned}$$

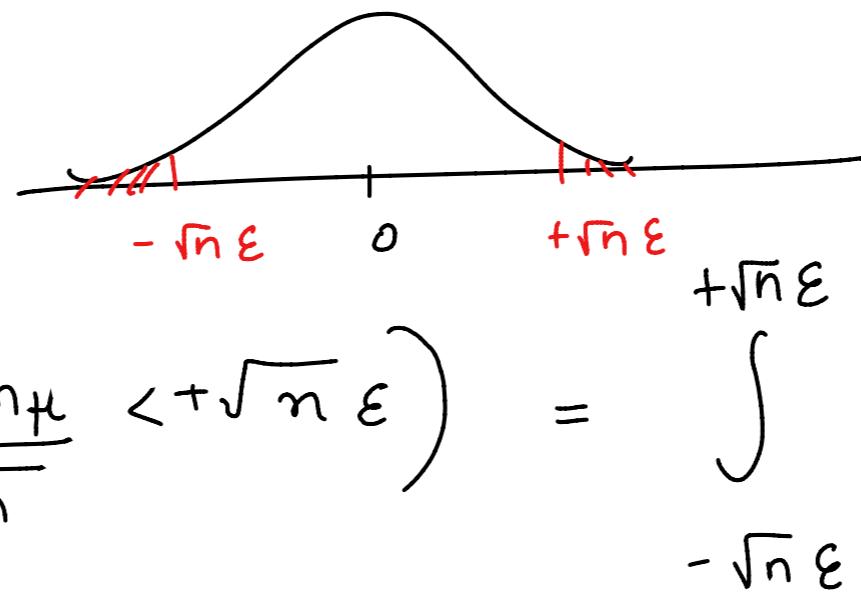
deviation of sample mean
 away from the true mean

$$\text{Prob} \left(\left| \frac{S_n - n\mu}{\sqrt{n}} \right| > \sqrt{n}\varepsilon \right) = 1 - \text{Prob} \left(-\sqrt{n}\varepsilon < \frac{S_n - n\mu}{\sqrt{n}} < +\sqrt{n}\varepsilon \right)$$

If we knew the exact density

$$f_{\frac{S_n - n\mu}{\sqrt{n}}}^{\text{true}}$$

diff ($f^{\text{true}}, f^{\text{normal}}$)
 $\frac{S_n - n\mu}{\sqrt{n}}$
 ↓
 not known



$$\text{Prob} \left(-\sqrt{n}\varepsilon < \frac{S_n - n\mu}{\sqrt{n}} < +\sqrt{n}\varepsilon \right) = \int_{-\sqrt{n}\varepsilon}^{+\sqrt{n}\varepsilon} f(x) dx \quad (*)$$

$f_{\frac{S_n - n\mu}{\sqrt{n}}}^{\text{true}}(x)$ is not known. For n sufficient large, then we know

large, then we know $f_{\frac{S_n - n\mu}{\sqrt{n}}}^{\text{true}}(x)$ looks more and more like

the density of $N(0, \sigma^2)$. So plug in the expression for $N(0, \sigma^2)$ in $(*)$ in place of $f_{\frac{S_n - n\mu}{\sqrt{n}}}^{\text{true}}(x)$

$$\text{Prob} \left(\left| \frac{S_n - n\mu}{\sqrt{n}} \right| > \sqrt{n}\varepsilon \right) = \int_{-\infty}^{-\sqrt{n}\varepsilon} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx + \int_{\sqrt{n}\varepsilon}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx$$

left tail mass *right tail mass*

*due to symmetry
(left = right)*

$$= 2 \int_{\sqrt{n}\varepsilon}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx$$

$$\leq 2 \int_{\sqrt{n}\varepsilon}^{\infty} \left(\frac{x}{\sqrt{n}\varepsilon} \right) \cdot \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} dx$$

$$= \frac{2}{\sqrt{2\pi\sigma^2}} \cdot \frac{\sigma^2}{\sqrt{n}\varepsilon} \int_{\sqrt{n}\varepsilon}^{\infty} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} dx$$

$$= \frac{2}{\sqrt{2\pi\sigma^2}} \cdot \frac{\sigma^2}{\sqrt{n}\varepsilon} \int_{\sqrt{n}\varepsilon}^{\infty} e^{-\frac{x^2}{2\sigma^2}} d\left(\frac{x^2}{2\sigma^2}\right)$$

$$= \frac{2}{\sqrt{2\pi\sigma^2}} \cdot \frac{\sigma^2}{\sqrt{n}\varepsilon} \left[e^{-\frac{x^2}{2\sigma^2}} \right]_{-\infty}^{\sqrt{n}\varepsilon}$$

$$= \frac{2}{\sqrt{2\pi\sigma^2}} \cdot \frac{\sigma^2}{\sqrt{n}\varepsilon} \cdot e^{-\frac{(\sqrt{n}\varepsilon)^2}{2\sigma^2}}$$

$$= \left(\sqrt{\frac{2}{\pi}} \right) \cdot \frac{1}{\sqrt{n}} \cdot \left(\frac{\sigma}{\varepsilon} \right) \cdot e^{-\frac{1}{2} \left(\frac{\varepsilon}{\sigma} \right)^2 \cdot n}$$

Approximate Tail Bound

$$\text{Prob} \left(|\bar{\mu}_n - \mu| > \varepsilon \right) \leq \left(\sqrt{\frac{2}{\pi}} \right) \cdot \frac{1}{\sqrt{n}} \cdot \left(\frac{\sigma}{\varepsilon} \right) \cdot e^{-\frac{1}{2} \left(\frac{\varepsilon}{\sigma} \right)^2 \cdot n}$$



approximate
issue: not an inequality

- we cannot assume knowledge of P_1, \dots, P_k
- we can make certain structural assumption (common sense based)
 - + random variables are bounded $\pm B$
 - + variance is bounded.

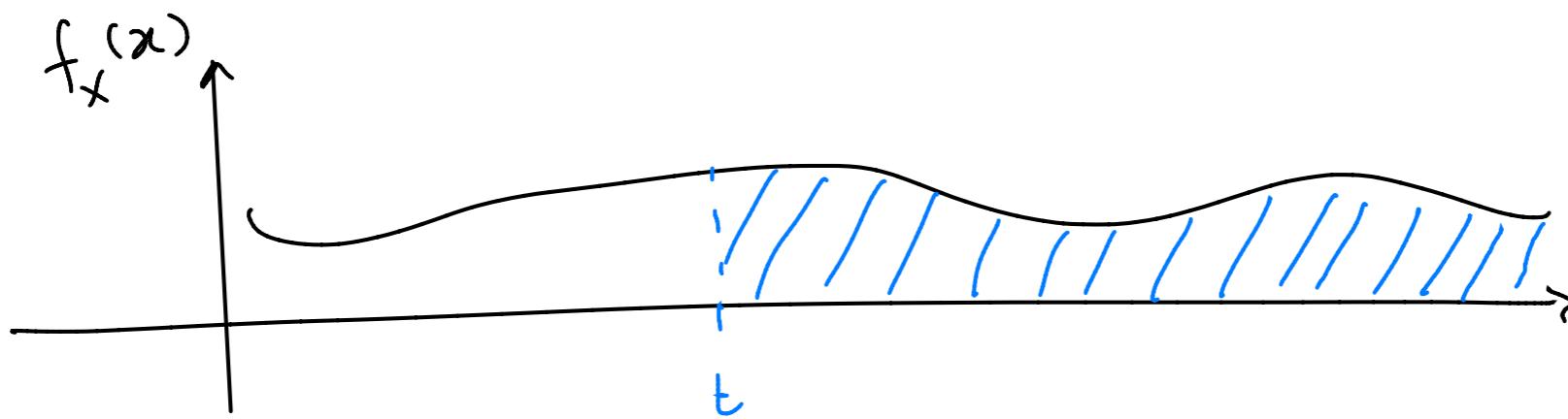
Markov Inequality

Let X be a positive random variable $X: \Omega \rightarrow \mathbb{R}_+$

$$\text{Prob}(X > t) \leq \frac{\mathbb{E}[X]}{t}$$

inequality

Note: we have implicitly assumed that $\mathbb{E}[X]$ is finite



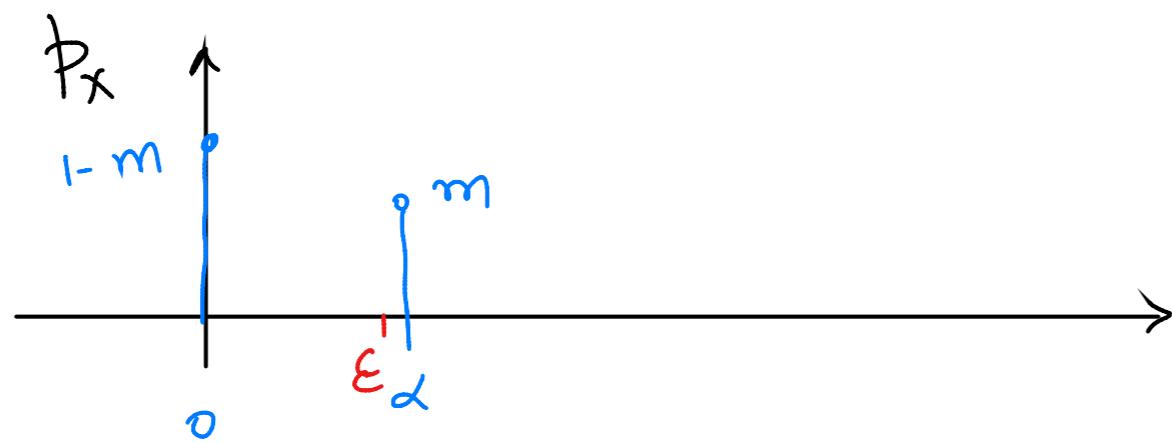
$$P_{\text{prob}}(x > t) = \int_t^{\infty} 1 \cdot f_x(x) dx$$

$$\leq \int_t^{\infty} \left(\frac{x}{t}\right) \cdot f_x(x) dx$$

$$= \frac{1}{t} \int_t^{\infty} x \cdot f_x(x) dx$$

$$\leq \frac{1}{t} \int_0^{\infty} x \cdot f_x(x) dx$$

$$= \frac{\mathbb{E}[x]}{t}$$



$$\begin{aligned} \text{Prob}(X=0) &= 1-m \\ \text{Prob}(X=\alpha) &= m \end{aligned}$$

$$\begin{aligned} \mathbb{E}[X] &= 0 \cdot (1-m) + \alpha \cdot m \\ &= \alpha m \end{aligned}$$

Markov Inequality:

$$\text{Prob}(X > t) \leq \frac{\mathbb{E}[X]}{t} = \frac{\alpha m}{t}$$

Bound is tight only for t close to α

$$\text{Prob}(X > \alpha - \varepsilon) = m$$

$$\frac{\mathbb{E}[X]}{t} = \frac{\alpha m}{t} = \left(\frac{\alpha}{\alpha - \varepsilon} \right) \cdot m$$

Exact

Markov Inequality

For $t < \alpha$

$$\text{Prob}(X > t) = m \quad \text{vs}$$

$$\left(\frac{\alpha}{t}\right) \cdot m$$

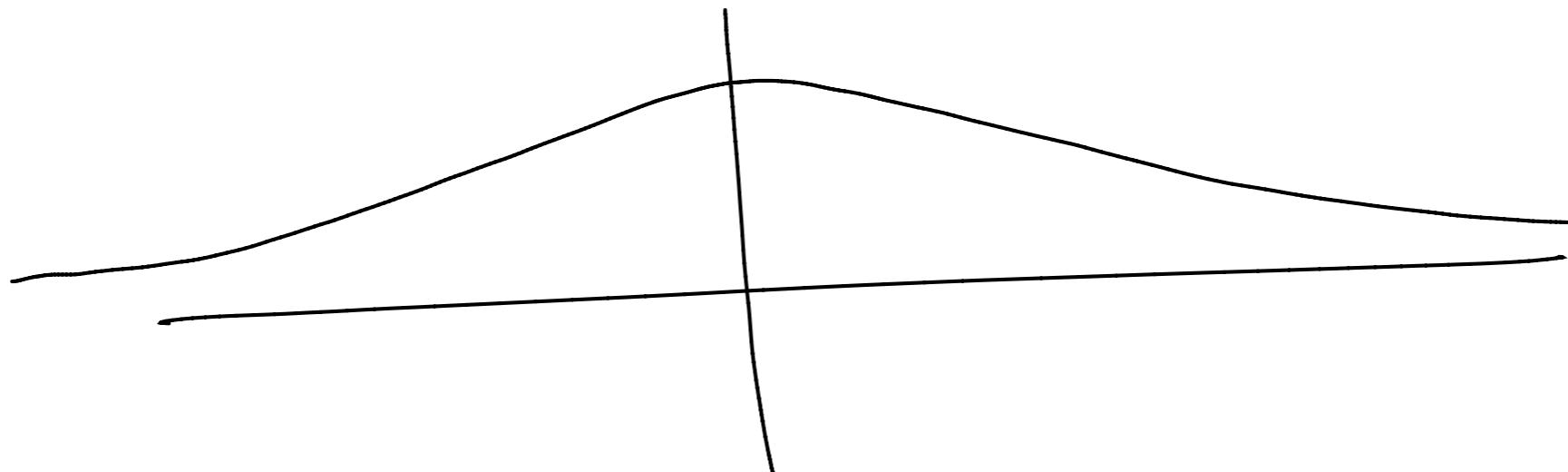
$$\left(\frac{\alpha}{t}\right)^{-1} \cdot m$$

For $t > \alpha$

$$\text{Prob}(X > t) = 0 \quad \text{vs}$$

Example 2: When we cannot apply Markov inequality

Take Cauchy Density $f_x(x) = \frac{1}{\pi(1+x^2)}$



$$\mathbb{E}[|x|] = \int_0^\infty x \cdot \frac{2}{\pi(1+x^2)} dx \quad \text{increasing} \quad \text{decreasing}$$

$$= \int_0^\infty \frac{d(1+x^2)}{\pi(1+x^2)} = \frac{1}{\pi} \left[\log(1+x^2) \right]_0^\infty = \infty$$

Fact I: $\sum_{n>0} \frac{1}{n} = \infty$

$$= 1 + \underbrace{\frac{1}{2}}_{\text{blue brace}} + \underbrace{\frac{1}{3} + \frac{1}{4}}_{\text{blue brace}} + \underbrace{\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}}_{\text{blue brace}}$$

$$\geq \frac{1}{2} + \frac{1}{2} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2}$$

$$\text{Fact II: } \sum_{n \geq 1} \frac{1}{n^2} = 1 + \frac{1}{2^2} + \underbrace{\frac{1}{3^2} + \frac{1}{4^2}}_{\downarrow} + \underbrace{\frac{1}{5^2} + \frac{1}{6^2} + \frac{1}{7^2} + \frac{1}{8^2}}_{\downarrow}$$

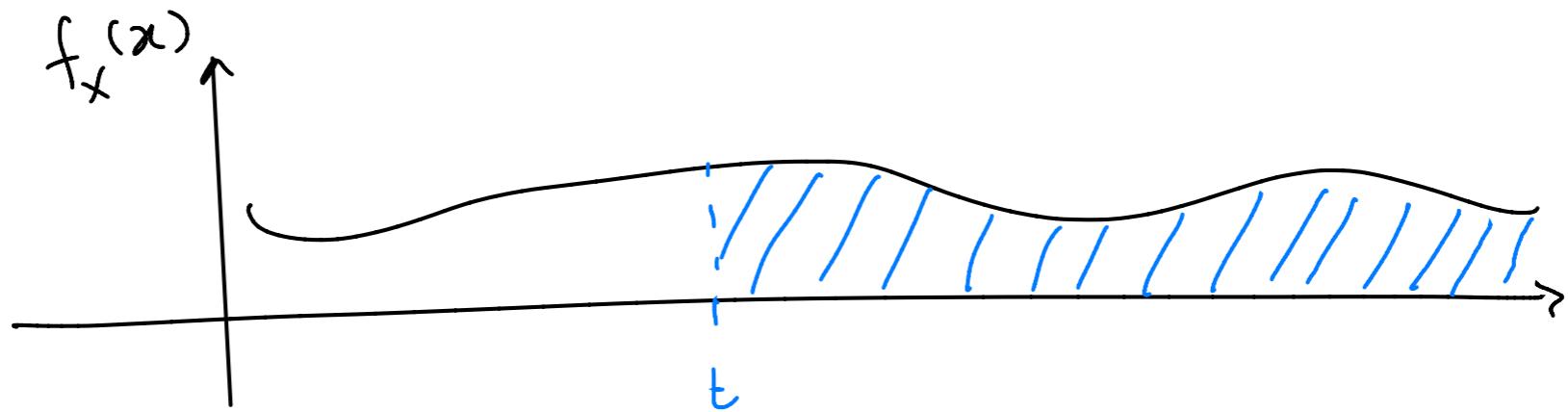
$$\leq 1 + 1 + \frac{1}{2^2} + \frac{1}{2^2} + \frac{1}{4^2} + \frac{1}{4^2} + \frac{1}{4^2} + \frac{1}{8^2} + \dots$$

$$= 1 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \dots$$

$$\leq 3$$

$$\sum_{n \geq 1} \frac{1}{n^2} = \frac{\pi^2}{6}$$

Try to extend



$$P_{\text{prob}}(x > t) = \int_t^{\infty} 1 \cdot f_x(x) dx$$

$$\leq \int_t^{\infty} \left(\frac{x}{t}\right)^m \cdot f_x(x) dx$$

$$= \frac{1}{t^m} \int_t^{\infty} x^m \cdot f_x(x) dx$$

$$P_{\text{prob}}(x > t) = P_{\text{prob}}(x^m > t^m) \leq \frac{\mathbb{E}[x^m]}{t^m}$$

same quantities

↑ perhaps this fails and kills x^m
steeper function
and faster

$$\text{Prob}(X > t) \leq \frac{\mathbb{E}[X]}{t}$$

"

$$\text{Prob}(X^2 > t^2) \leq \frac{\mathbb{E}[X^2]}{t^2}$$

"

$$\text{Prob}(X^{100} > t^{100}) \leq \frac{\mathbb{E}[X^{100}]}{t^{100}}$$

Insight on why $\text{Prob}(X > t)$ has $\mathbb{E}[X]$

Consider a discrete (positive) random variable $X \sim p_x$
 Supported on integers

$$\mathbb{E}[X] = \sum_{x \geq 0} x \cdot p_x(x) = 0 \cdot p_x(0) + 1 \cdot p_x(1) + 2 \cdot p_x(2) + 3 \cdot p_x(3) + \dots$$

$$\begin{aligned}
 &= p_x(1) \\
 &\quad + p_x(2) + p_x(2) \\
 &\quad + \boxed{p_x(3)} + \boxed{p_x(3)} + \boxed{p_x(3)} + p_x(4) \\
 &\quad + p_x(4) + p_x(4) + p_x(4) + p_x(4)
 \end{aligned}$$

$p_{\text{rob}}(x_7, 3)$ appears 3 times in $\#[\mathbf{x}]$.

$$\text{Prob}(X > t) \leq \frac{\mathbb{E}[x^m]}{t^m}$$

(Q1) We wanted to understand smaller deviations of the sample mean away from true mean. But Prob($X > t$) means larger and larger t away from origin

(Q2) Do we really see t^m effect

(Q3) Do we always keep winning by increasing m

$$(A1) S_n = x_1 + \dots + x_n, \quad \hat{\mu}_n = \frac{S_n}{n}$$

$$\text{Prob}(|\hat{\mu}_n - \mu| > \varepsilon) = \text{Prob}\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right)$$

$$= \text{Prob}(|S_n - n\mu| > n\varepsilon)$$

$\underbrace{x}_{\sim t}$

(Q1) seems like a contradiction
but it is not

(A2)

Chebyshov Inequality

$$\begin{aligned} \text{Prob} \left(\left| \frac{S_n}{n} - \mu \right| > \varepsilon \right) &= \text{Prob} \left(|S_n - n\mu| > n\varepsilon \right) \\ &\leq \frac{\mathbb{E} [|S_n - n\mu|^m]}{(n\varepsilon)^m} \end{aligned}$$

Pick $m = 2$

$$\begin{aligned} \mathbb{E} [(S_n - n\mu)^2] &= \mathbb{E} [(\{x_1 - \mu\} + \{x_2 - \mu\} + \dots + \{x_n - \mu\})^2] \\ &= \mathbb{E} [(\{x_1 - \mu\} + \{x_2 - \mu\} + \dots + \{x_n - \mu\}) \\ &\quad \times \\ &\quad (\{x_1 - \mu\} + \{x_2 - \mu\} + \dots + \{x_n - \mu\})] \\ &= \sum_{i,j=1}^n \mathbb{E} [(x_i - \mu)(x_j - \mu)] \end{aligned}$$

$$\sum_{i,j=1}^n \mathbb{E} [(x_i - \mu)(x_j - \mu)] = \sum_{i=1}^n \mathbb{E} [(x_i - \mu)^2] +$$

$\sim \sigma^2$

$$\sum_{\substack{i,j=1 \\ i \neq j}} \mathbb{E} [(x_i - \mu)(x_j - \mu)]$$

$\parallel 0$

$$\begin{aligned} \mathbb{E} [(x_i - \mu)(x_j - \mu)] &= \mathbb{E} [x_i - \mu] \mathbb{E} [x_j - \mu] \\ &= (\mathbb{E}[x_i] - \mu) (\mathbb{E}[x_j] - \mu) \\ &= (\mu - \mu) (\mu - \mu) \\ &= 0 \end{aligned}$$

$$\mathbb{E} [(S_n - n\mu)^2] = n\sigma^2$$

$$\text{Prob} \left(\left| \frac{s_n}{n} - \mu \right| > \varepsilon \right) \leq \frac{n \sigma^2}{(n \varepsilon)^2}$$

$$= \left(\frac{\sigma}{\varepsilon} \right)^2 \cdot \frac{1}{n}$$

We get a rate of $\frac{1}{n}$

Pick $m = 4$

$$\text{Prob} \left(\left| \frac{s_n}{n} - \mu \right| > \varepsilon \right) = \text{Prob} \left(\underbrace{|s_n - n\mu|}_{x} > \underbrace{n\varepsilon}_{t} \right) \leq \frac{\mathbb{E}[x^4] t^4}{(n\varepsilon)^4}$$

$$\mathbb{E}[(s_n - n\mu)^4] = \mathbb{E} \left[\{ (x_1 - \mu) + (x_2 - \mu) + \dots + (x_n - \mu) \}^4 \right]$$

$$= \mathbb{E} \left[\{ (x_1 - \mu) + (x_2 - \mu) + \dots + (x_n - \mu) \} \right.$$

$$\times \{ (x_1 - \mu) + (x_2 - \mu) + \dots + (x_n - \mu) \}$$

$$\times \{ (x_1 - \mu) + (x_2 - \mu) + \dots + (x_n - \mu) \}$$

$$\times \{ (x_1 - \mu) + (x_2 - \mu) + \dots + (x_n - \mu) \}$$

$$\times \{ (x_1 - \mu) + (x_2 - \mu) + \dots + (x_n - \mu) \} \left. \right]$$

$$= \mathbb{E} \left[\sum_{i,j,k,l=1}^n (x_i - \mu) (x_j - \mu) (x_k - \mu) (x_l - \mu) \right]$$

$$= \sum_{i,j,k,l=1}^n \mathbb{E} [(x_i - \mu) (x_j - \mu) (x_k - \mu) (x_l - \mu)]$$

all distinct

For $i \neq j \neq k \neq l$ terms do not survive

$$\mathbb{E} [(x_i - \mu) (x_j - \mu) (x_k - \mu) (x_l - \mu)] =$$

$$\mathbb{E} [(x_i - \mu)] \mathbb{E} [(x_j - \mu)] \mathbb{E} [(x_k - \mu)] \mathbb{E} [(x_l - \mu)]$$

$$(\mathbb{E}[x_i] - \mu) (\mathbb{E}[x_j] - \mu) (\mathbb{E}[x_k] - \mu) (\mathbb{E}[x_l] - \mu) \\ (\mu - \mu) = 0$$

*three
distinct* $i = j \neq k = \ell$

$$\mathbb{E} [(x_i - \mu)^2 (x_k - \mu) (x_\ell - \mu)]$$

$$(\mathbb{E} [(x_i - \mu)^2]) \underset{!!}{=} (\mathbb{E} [x_k - \mu]) \underset{!!}{=} (\mathbb{E} [x_\ell - \mu]) = 0$$

two distinct

$$i = j = k \neq \ell$$

$$\mathbb{E} [(x_i - \mu)^3 (x_\ell - \mu)] = \mathbb{E} [(x_i - \mu)^3] \underset{!!}{=} \mathbb{E} [x_\ell - \mu]$$
$$= 0$$

$$i = j \neq k = \ell$$

$$\mathbb{E} [(x_i - \mu)^2 (x_k - \mu)^2] = \mathbb{E} [(x_i - \mu)^2] \underset{\sigma^2}{\mathbb{E} [(x_k - \mu)^2]}$$
$$= \sigma^4$$

none
distinct

$$\mathbb{E}[(x_i - \mu)^4] \neq 0$$

$\mathbb{E}[(S_n - n\mu)^4]$ contains $\sum_{i=1}^n \mathbb{E}[(x_i - \mu)^4]$ plus all terms of

form $\mathbb{E}[(x_i - \mu)^2 (x_j - \mu)^2]$ choice of i, j from

$$\mathbb{E}[(S_n - n\mu)^4] = \sum_{i=1}^n \mathbb{E}[(x_i - \mu)^4] + \frac{4}{C_2} \binom{n}{2} \sigma^4$$

\parallel

$$n \mathbb{E}[(x_1 - \mu)^4]$$

$$= n \mathbb{E}[(x_1 - \mu)^4] + \frac{4 \times 3}{1 \times 2} \frac{n \times (n-1)}{1 \times 2} \sigma^4$$

$$= n \mathbb{E}[(x_1 - \mu)^4] + 3n(n-1) \sigma^4$$

$$\text{Prob}(|\hat{\mu}_n - \mu| \geq \varepsilon) = \text{Prob}\left(\left|\frac{s_n}{n} - \mu\right| \geq \varepsilon\right) = \text{Prob}(|s_n - n\mu| \geq \varepsilon)$$

$$\leq \frac{\mathbb{E}[(s_n - n\mu)^4]}{(n\varepsilon)^4}$$

$$= \frac{3n(n-1)\sigma^4 + n\mathbb{E}[(x_1 - \mu)^4]}{(n\varepsilon)^4}$$

$$= \frac{\frac{3n(n-1)\sigma^4}{n^2} + \frac{n}{n^2}\mathbb{E}[(x_1 - \mu)^4]}{n^2\varepsilon^4}$$

$$\leq \frac{\frac{3\sigma^4}{n^2\varepsilon^4}}{n^2\varepsilon^4} + \frac{\mathbb{E}[(x_1 - \mu)^4]/n}{n^2\varepsilon^4}$$

$$\text{Prob}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \frac{\sigma^2}{n\varepsilon^2}$$

- (Bound 1)

$$\text{Prob}\left(\left|\frac{S_n}{n} - \mu\right| \geq \varepsilon\right) \leq \boxed{\frac{3 + 1}{n^2 \varepsilon^4}} + \frac{\mathbb{E}[(X_i - \mu)^4]/n}{n^2 \varepsilon^4}$$

- (Bound 2)

dominating term

Pick $\varepsilon = 0.1$, $\sigma = 1$, for bound 1 to be non-vacuous we need

$$\frac{1}{n\varepsilon^2} < 1, \quad n \geq 100 \text{ samples}$$

For bound 2 to be non-vacuous

$$\frac{3}{n^2 \times (0.1)^4} < 1, \quad n^2 > 3 \times 10^9$$

$$n \geq \sqrt{3} \cdot (100)$$

Qn: when we go from $m=2$ to $m=4$ we are winning

check for $m=6$

Moment Generating Function for a random variable X

$$M_X(\lambda) = \mathbb{E}[e^{\lambda X}]$$

$$= \mathbb{E}\left[1 + \frac{\lambda X}{1!} + \frac{\lambda^2 X^2}{2!} + \frac{\lambda^3 X^3}{3!} + \dots\right]$$

$M_X(x)$ for Gaussian random variable $N(0, \sigma^2)$

$$M_X(\lambda) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\lambda x} e^{-\frac{1}{2} \frac{x^2}{\sigma^2}} dx$$

mixes faster

$$= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left(\frac{x^2}{\sigma^2} - 2\lambda x + \lambda^2 \sigma^2 \right)} \cdot e^{\frac{1}{2} \lambda^2 \sigma^2} dx$$

$$= \frac{e^{\frac{-\frac{1}{2}x^2\sigma^2}{\lambda^2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\left(\frac{x}{\sigma} - \lambda\sigma\right)^2} dx$$

$$= \frac{e^{\frac{-\frac{1}{2}x^2\sigma^2}{\lambda^2\sigma^2}}}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2\sigma^2}(x - \lambda\sigma^2)^2} dx$$

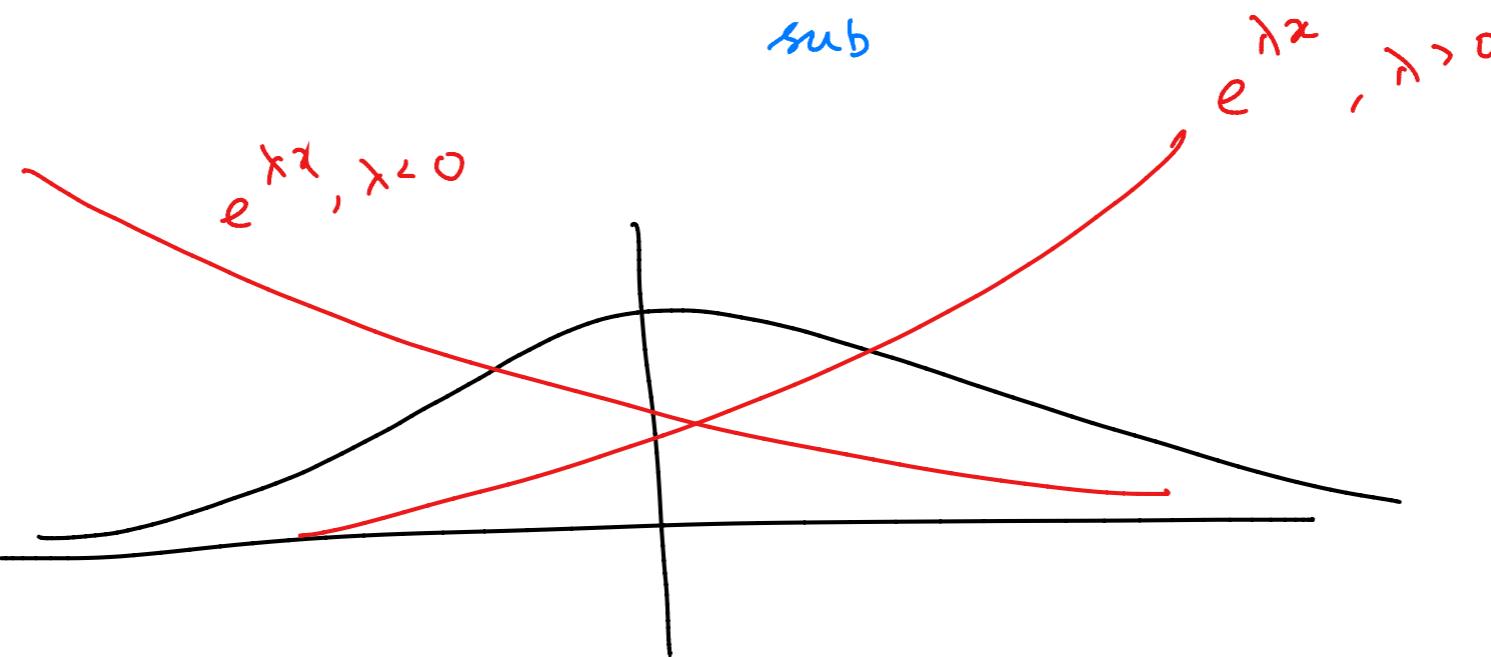
$$= e^{\frac{-\frac{1}{2}x^2\sigma^2}{\lambda^2\sigma^2}}$$

Sub-Gaussian Random Variable

A random variable X is said to be σ -sub Gaussian

$$M_X(\lambda) = \mathbb{E}[e^{\lambda X}] \leq e^{\lambda^2 \sigma^2 / 2}, \quad \lambda \in \mathbb{R}$$

↑
sub



Informally put, the tail falls as fast as $N(0, \sigma^2)$ is tail

Properties of Sub-Gaussian R.Vs

(1) X is σ -sub Gaussian $\Rightarrow \mathbb{E}[X] = 0$

Proof:

$$\mathbb{E}[e^{\lambda x}] \leq e^{\frac{\lambda^2 \sigma^2}{2}} \quad (\text{expand on both sides})$$

$$\mathbb{E}\left[1 + \frac{\lambda x}{L} + \frac{\lambda^2 x^2}{L^2} + \dots\right] \leq 1 + \frac{\lambda^2 \sigma^2 / 2}{L} + \frac{(\lambda^2 \sigma^2 / 2)^2}{L^2} + \dots$$

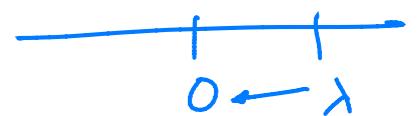
Push $\mathbb{E}[\cdot]$ inside

$$1 + \frac{\mathbb{E}[\lambda x]}{L} + \frac{\mathbb{E}[\lambda^2 x^2]}{L^2} + \dots \leq 1 + \frac{\lambda^2 \sigma^2 / 2}{L} + \frac{(\lambda^2 \sigma^2 / 2)^2}{L^2} + \dots$$

Pull λ terms out of $\mathbb{E}[\cdot]$,

$$\lambda \frac{\mathbb{E}[x]}{L} + \lambda^2 \frac{\mathbb{E}[x^2]}{L^2} \leq \frac{\lambda^2 \sigma^2 / 2}{L} + \frac{(\lambda^2 \sigma^2 / 2)^2}{L^2} + \dots$$

Case 1: $\lambda > 0$, divide both sides by λ and let $\lambda \rightarrow 0$



$$\frac{\mathbb{E}[X]}{L^1} + \lambda \frac{\mathbb{E}[X^2]}{L^2} \leq \frac{\frac{\lambda \sigma^2}{2}}{L^1} + \frac{1}{\lambda} \frac{(\lambda^2 \sigma^2 / 2)^2}{L^2} + \frac{1}{\lambda} \dots$$

$\xrightarrow{0}$ $\xrightarrow{0}$ $\xrightarrow{0}$

$$\frac{\mathbb{E}[X]}{L^1} \leq -\lambda \frac{\mathbb{E}[X^2]}{L^2} + \frac{\frac{\lambda \sigma^2}{2}}{L^1} + \frac{1}{\lambda} \frac{(\lambda^2 \sigma^2 / 2)^2}{L^2} + \frac{1}{\lambda} \dots$$

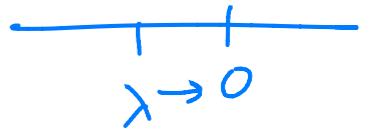
$\xrightarrow{0}$ $\xrightarrow{0}$ $\xrightarrow{0}$

$$\frac{\mathbb{E}[X]}{L^1} \leq + \frac{\frac{\lambda \sigma^2}{2}}{L^1} + \frac{1}{\lambda} \frac{(\lambda^2 \sigma^2 / 2)^2}{L^2} + \frac{1}{\lambda} \dots$$

$\xrightarrow{0}$ $\xrightarrow{0}$ $\xrightarrow{0}$

$$\mathbb{E}[X] \leq 0 \quad -(1)$$

case 2: $\lambda < 0$, divide by λ and let $\lambda \rightarrow 0_-$



$$\frac{\mathbb{E}[X]}{L^1} + \lambda \frac{\mathbb{E}[X^2]}{L^2} \xrightarrow{\lambda \rightarrow 0^-}$$

negative x

$$\geq \frac{\lambda \sigma^2/2}{L^1} + \frac{1}{\lambda} \frac{(\lambda^2 \sigma^2/2)^2}{L^2} + \frac{1}{\lambda} \cdots \xrightarrow{\lambda \rightarrow 0^-} 0 \xrightarrow{\lambda \rightarrow 0} 0$$

$$\mathbb{E}[X] \geq 0 \quad - (2)$$

$$\textcircled{1} \text{ \& } \textcircled{2} \Rightarrow \mathbb{E}[X] = 0$$

$$(2) \quad X \text{ is } \sigma\text{-subGaussian} \Rightarrow \text{Var}[X] \leq \sigma^2$$

$$\cancel{\frac{\lambda \mathbb{E}[x]}{L^1}} + \frac{\lambda^2 \mathbb{E}[x^2]}{L^2} \leq \frac{\frac{\lambda^2 \sigma^2}{2}}{L^1} + \frac{(\frac{\lambda^2 \sigma^2}{2})^2}{L^2} + \dots$$

$\cancel{\frac{\lambda \mathbb{E}[x]}{L^1}}$
 $\cancel{+}$
 $\cancel{\frac{\lambda^2 \mathbb{E}[x^2]}{L^2}}$

$\cancel{0}$

$$\frac{\lambda^2 \mathbb{E}[x^2]}{L^2} + \frac{\lambda^3 \mathbb{E}[x^3]}{L^3} \leq \frac{\frac{\lambda^2 \sigma^2}{2}}{L^1} + \frac{(\frac{\lambda^2 \sigma^2}{2})^2}{L^2} + \dots$$

divide by λ^2 and let $\lambda \rightarrow 0$

$$\frac{\mathbb{E}[x^2]}{L^2} + \lambda \frac{\mathbb{E}[x^3]}{L^3} \leq \frac{\frac{\sigma^2}{2}}{L^1} + \frac{1}{\lambda^2} \frac{(\frac{\lambda^2 \sigma^2}{2})^2}{L^2} + \frac{1}{\lambda^2} \dots$$

$\rightarrow 0 \qquad \qquad \qquad \rightarrow 0$

$$\mathbb{E}[x^2] \leq \sigma^2$$

$$\text{Var}[x] = \mathbb{E}[x^2] - (\mathbb{E}[x])^2 \leq \sigma^2$$

(3) X is σ -subGaussian $\Rightarrow cX$ is $|c|\sigma$ subGaussian

$$M(\lambda) = \mathbb{E}[e^{\lambda cX}] \leq e^{(\lambda c)^2 \sigma^2 / 2} = e^{\lambda^2 c^2 \sigma^2 / 2}$$

(4) If x_1 and x_2 are independent and say
 σ_1 -subGaussian and σ_2 -subGaussian then $x_1 + x_2$
 is $\sqrt{\sigma_1^2 + \sigma_2^2}$ -subGaussian

$$\begin{aligned} M_{x_1+x_2}(\lambda) &= \mathbb{E}[e^{\lambda(x_1+x_2)}] \\ &= \mathbb{E}[e^{\lambda x_1} \cdot e^{\lambda x_2}] \quad (\text{invoke independence}) \\ &= \mathbb{E}[e^{\lambda x_1}] \cdot \mathbb{E}[e^{\lambda x_2}] \\ &\leq e^{\lambda^2 \sigma_1^2 / 2} \cdot e^{\lambda^2 \sigma_2^2 / 2} \end{aligned}$$

$$\begin{aligned}
 & \lambda^2 (\sigma_1^2 + \sigma_2^2) / 2 \\
 = & e \\
 & \lambda^2 \left(\sqrt{\sigma_1^2 + \sigma_2^2} \right)^2 / 2 \xrightarrow{\sigma_{\text{effective}}} \\
 = & e
 \end{aligned}$$

Cramer - Chernoff Bound

If x is σ -sub Gaussian
 $- (\epsilon^2 / 2\sigma^2)$

$$\text{Prob}(x \geq \epsilon) \leq e$$

$$\text{Prob}(y \geq t) \leq \frac{\mathbb{E}[y]}{t}$$

Proof:

$$\begin{aligned}
 \text{Prob}(x \geq \epsilon) &= \text{Prob}(e^{\lambda x} \geq e^{\lambda \epsilon}) \quad \left(e^{\lambda x} \text{ is a r.v.} \right) \\
 &\leq \frac{\mathbb{E}[e^{\lambda x}]}{e^{\lambda \epsilon}}
 \end{aligned}$$

$$\leq e^{\frac{x^2\sigma^2/2}{x\varepsilon}}$$

$$= e^{(\frac{x^2\sigma^2/2 - \lambda\varepsilon}{x^2\sigma^2/2 - \lambda\varepsilon})}$$

$$\text{Prob}(X > \varepsilon) \leq \min_x e^{-\lambda\varepsilon + \frac{x^2\sigma^2/2}{x}}$$

$$\min_{\lambda} -\lambda\varepsilon + \frac{x^2\sigma^2/2}{x} = \min_x h(x)$$

diff $h(x)$ w.r.t λ

$$-\varepsilon + \frac{2\lambda\sigma^2}{x} = 0$$

$$\lambda_* = \frac{\varepsilon}{\sigma^2}$$

$$h(x_*) = -\frac{\varepsilon}{\sigma^2} \cdot \varepsilon + \left(\frac{\varepsilon}{\sigma^2}\right)^2 \frac{\sigma^2}{2}$$

$$= -\frac{\varepsilon^2}{\sigma^2} + \frac{\varepsilon^2}{2\sigma^2}$$

$$= -\frac{\varepsilon^2}{2\sigma^2}$$

$$\text{Prob}(X \geq \varepsilon) \leq e^{-\frac{\varepsilon^2}{2\sigma^2}}$$

we need deviation of sample mean from true mean

$$\text{Prob}(|\hat{\mu}_n - \mu| \geq \varepsilon) = \text{Prob}\left(\left|\frac{s_n}{n} - \mu\right| \geq \varepsilon\right)$$

$$= \text{Prob}\left(\left|\frac{s_n - n\mu}{n}\right| \geq \varepsilon\right)$$

$$\frac{s_n - n\mu}{n} = \frac{(x_1 - \mu) + (x_2 - \mu) + \dots + (x_n - \mu)}{n}$$

$$= \frac{(x_1 - \mu)}{n} + \frac{(x_2 - \mu)}{n} + \dots + \frac{(x_n - \mu)}{n}$$

If $(X_i - \mu)$ is σ -sub Gaussian

$\frac{X_i - \mu}{n}$ is $\frac{\sigma}{n}$ -sub Gaussian (3)

(4) $\frac{S_n - n\mu}{\sqrt{n}}$ is $\sqrt{\underbrace{\left(\frac{\sigma}{n}\right)^2 + \dots + \left(\frac{\sigma}{n}\right)^2}_{n\text{-time}}}$ sub Gaussian

$$\text{is } \sqrt{n \frac{\sigma^2}{n^2}} = \sqrt{\frac{\sigma^2}{n}} \text{ sub Gaussian}$$

$$\Pr \left(\frac{S_n}{n} - \mu \geq \varepsilon \right) \leq e^{-\left(\frac{\varepsilon^2}{2} \left(\frac{\sigma^2}{n}\right)\right)}$$

$$\text{Prob}\left(\frac{s_n - n\mu}{n} > \varepsilon\right) \leq e^{-\left(\frac{n\varepsilon^2}{2\sigma^2}\right)}$$

$\underbrace{x}_{\text{X}}$

$$\text{Prob}\left(\left|\frac{s_n - n\mu}{n}\right| > \varepsilon\right) \leq 2e^{-\left(\frac{n\varepsilon^2}{2\sigma^2}\right)}$$

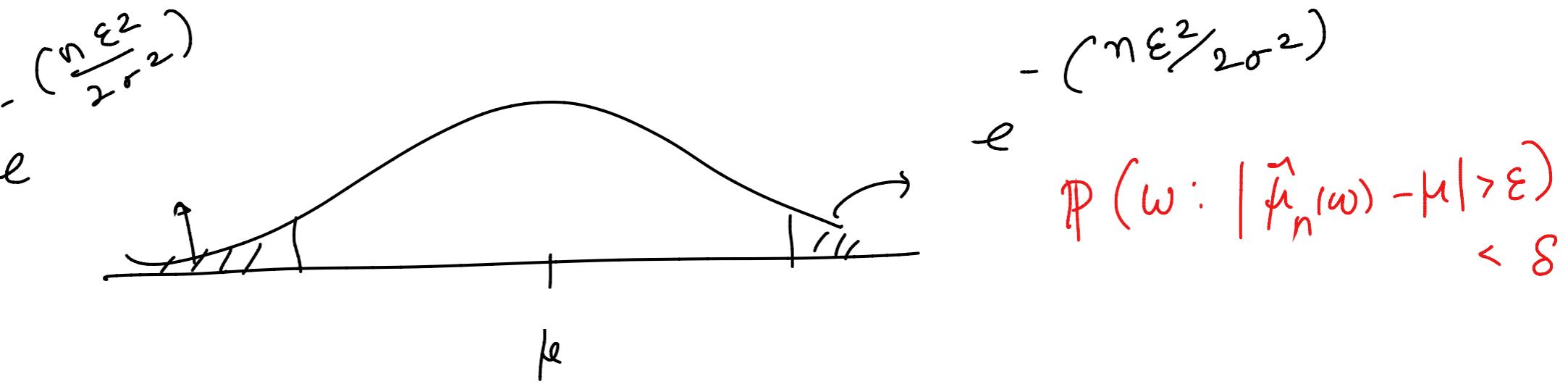
(one side tail)

(both sides
of the
tail)

For those who are in the center

Please don't disturb the class

- others sorry on behalf



Say I fix a probability $\delta > 0$, then with probability $1-\delta$, \tilde{f}_n falls within $\pm \sqrt{\frac{2\sigma^2 \ln(2/\delta)}{n}}$

$$-(\frac{n\epsilon^2}{2\sigma^2})$$

$$\delta = 2e$$

$$-(\frac{n\epsilon^2}{2\sigma^2})$$

$$\frac{\delta}{2} = e$$

$$e^{\frac{n\epsilon^2}{2\sigma^2}} = \frac{2}{\delta}$$

$$\frac{n\epsilon^2}{2\sigma^2} = \ln(\frac{2}{\delta}), \quad \epsilon^2 = \frac{2\sigma^2 \ln(2/\delta)}{n}$$

$$\epsilon = \sqrt{\frac{2\sigma^2 \ln(2/\delta)}{n}}$$

Deterministic statement is with prob $\geq 1 - \delta$

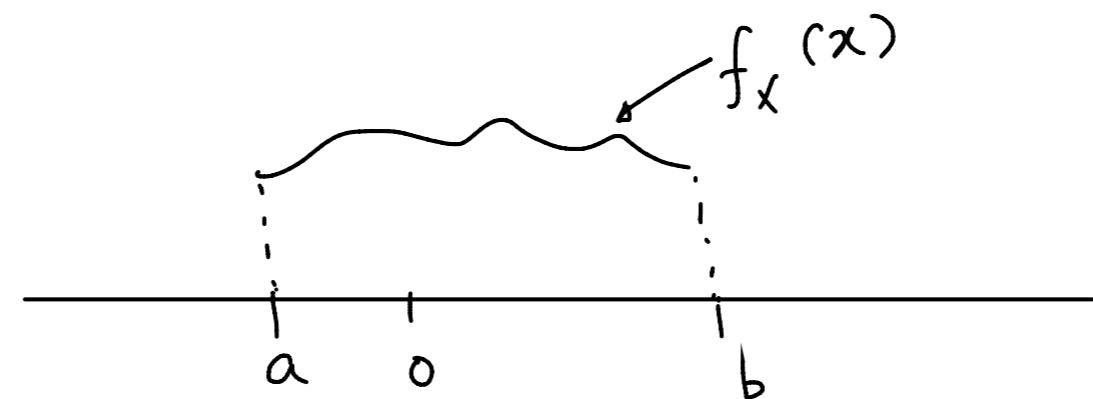
$$\hat{f}_n \in [\mu - \sqrt{\frac{2\sigma^2 \ln(2/\delta)}{n}}, \mu + \sqrt{\frac{2\sigma^2 \ln(2/\delta)}{n}}]$$

\uparrow
statistical rate

Chebychev

$$\hat{f}_n \in [\mu - \sqrt{\frac{\sigma^2}{n\delta}}, \mu + \sqrt{\frac{\sigma^2}{n\delta}}]$$

Let X be a bounded random variable taking values in interval $[a, b]$ s.t $E[X] = 0 \Rightarrow X$ is sub Gaussian



$$\min_c \mathbb{E} [\underbrace{(x-c)^2}_{\text{centered}}] = \text{Var}[x]$$

$$\begin{aligned}\mathbb{E}[(x-c)^2] &= \mathbb{E}[x^2 - 2cx + c^2] \\ &= \mathbb{E}[x^2] - 2c\mathbb{E}[x] + \mathbb{E}[c^2]\end{aligned}$$

diff w.r.t c

$$-2\mathbb{E}[x] + 2c = 0 \Rightarrow c = \mathbb{E}[x]$$

$$\text{Var}[x] = \mathbb{E}[(x - \mathbb{E}[x])^2]$$

Let x be a bounded random variable between $[a, b]$

$$\text{Var}[x] \leq \frac{(b-a)^2}{4}$$

$$\begin{aligned}
 \text{Var}[x] &= \min_c \mathbb{E}[(x-c)^2] \\
 &\leq \mathbb{E}\left[\left(x - \frac{a+b}{2}\right)^2\right] \\
 &= \int_a^b \left(x - \frac{a+b}{2}\right)^2 f_x(x) dx \\
 &\leq \int_a^b \max\{e_1, e_2\} f_x(x) dx \\
 &= \max\{e_1, e_2\} \int_a^b f_x(x) dx
 \end{aligned}$$

$$\text{var}[x] \leq \max \left\{ \left(a - \frac{a+b}{2} \right)^2, \left(b - \frac{a+b}{2} \right)^2 \right\}$$

$$= \max \left\{ \left(\frac{2a - (a+b)}{2} \right)^2, \left(\frac{2b - (a+b)}{2} \right)^2 \right\}$$

$$= \max \left\{ \left(\frac{a-b}{2} \right)^2, \left(\frac{b-a}{2} \right)^2 \right\}$$

$$= \frac{(b-a)^2}{4}$$

• consider $\gamma(\lambda) = \ln M_x(\lambda)$

$$= \ln \mathbb{E}[e^{\lambda x}]$$

Expand $\gamma(x)$ around 0 using Taylor series (reminder version)

$$\gamma(0) = \ln \mathbb{E}[e^{0x}] = \ln \mathbb{E}[e^0] = \ln \mathbb{E}[1] = \ln 1 = 0$$

$$\gamma'(\lambda) = \frac{d}{d\lambda} \ln \mathbb{E}[e^{\lambda x}]$$

$$\begin{aligned} \gamma'(\lambda) &= \frac{1}{\mathbb{E}[e^{\lambda x}]} \mathbb{E}[x e^{\lambda x}] \\ &= \int_{-\infty}^{\infty} x \frac{e^{\lambda x} f_x(x)}{\mathbb{E}[e^{\lambda x}]} dx \\ &\quad \text{is } g_\lambda(x) \end{aligned}$$

$$\gamma'(0) = \frac{\mathbb{E}[x e^0]}{\mathbb{E}[e^{0x}]}$$

$$= \frac{\mathbb{E}[x]}{\mathbb{E}[1]} = \frac{\mathbb{E}[x]}{1} = 0$$

$$= \mathbb{E}_{g_\lambda} [x]$$

is $g_\lambda(x)$ a density?

$$\int_{-\infty}^{\infty} g_\lambda(x) dx = 1$$

$$\int_{-\infty}^{\infty} \frac{e^{\lambda x} f_x(x)}{\mathbb{E}[e^{\lambda x}]} dx = \frac{1}{\mathbb{E}[e^{\lambda x}]} \int_{-\infty}^{\infty} e^{\lambda x} f_x(x) dx = \frac{\mathbb{E}[e^{\lambda x}]}{\mathbb{E}[e^{\lambda x}]} = 1$$

$$\gamma'(\lambda) = \frac{1}{\mathbb{E}[e^{\lambda x}]} \mathbb{E}[x e^{\lambda x}]$$

$$\gamma''(\lambda) = \frac{\mathbb{E}[x^2 e^{\lambda x}]}{\mathbb{E}[e^{\lambda x}]} - \frac{\mathbb{E}[x e^{\lambda x}] \mathbb{E}[x e^{\lambda x}]}{(\mathbb{E}[e^{\lambda x}])^2} dv$$

$$= \frac{\mathbb{E}[x^2 e^{\lambda x}]}{\mathbb{E}[e^{\lambda x}]}$$

$$= \mathbb{E}_{g_\lambda}[x^2]$$

$$- \left(\frac{\mathbb{E}[x e^{\lambda x}]}{\mathbb{E}[e^{\lambda x}]} \right)^2$$

$$- \left(\mathbb{E}_{g_\lambda}[x] \right)^2$$

$$= \text{Var}_{g_\lambda}[X]$$

Invoke remainder version of Taylor expansion

$$\begin{aligned} \gamma(x) &= \gamma(0) + \gamma'(0) \lambda + \gamma''(\tilde{x}) \frac{\lambda^2}{L^2}, \quad \tilde{x} \in [0, \lambda] \\ &= 0 + 0 + \text{Var}_{g_{\tilde{x}}}[X] \cdot \frac{\lambda^2}{2} \end{aligned}$$

$$\leq \frac{(b-a)^2}{4} \cdot \frac{\lambda^2}{2}$$

$$\begin{aligned} \gamma_x(x) &= \ln \mathbb{E}[e^{Xx}], \quad M_x(\lambda) = e^{\gamma_x(\lambda)} \\ &\leq e^{\frac{(b-a)^2}{4} \cdot \frac{\lambda^2}{2}} \end{aligned}$$

$$\Rightarrow X \text{ is sub Gaussian with } \sigma = \frac{(b-a)^2}{4}$$

Design and Analysis of MAB algorithm ($\tau = 1$)

Basic Design / Algorithm : Explore-Then-Commit (ETC)

Exploration

Phase

- play each arm 'm' times and obtain the rewards
- calculate the sample mean
- always play the arm with best sample mean

$$\hat{\mu}_i(t) = \frac{\sum_{s=1}^t x_s \mathbb{I}_{\{A_s=i\}}}{T_i(t)}$$

sample mean
of i^{th} arm at time 't'

$$T_i(t) = \sum_{s=1}^t \mathbb{I}_{\{A_s=i\}}$$

(total times the i^{th}
arm was played
in ' t ' rounds)

for ETC

$$\bullet \quad A_1 = A_2 = \dots = A_m = 1$$

$$A_{m+1} = A_{m+2} = \dots = A_{2m} = 2$$

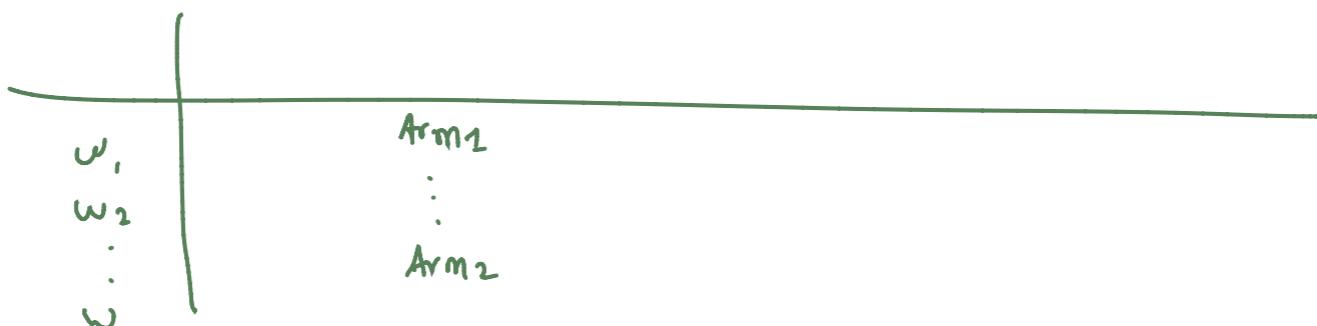
⋮

$$A_{(k-1)m+1} = A_{(k-1)m+2} = \dots = A_{km} = k$$

for $t > m k$

$$A_t = \arg \max_i f_i^{(mk)}$$

(ties are broken
arbitrarily)



Analysis of ETC: Let $n > mk$ (we have explored and then we are playing the arm with best sample mean)

$$\text{Regret}(n) = R_n \leq \sum_{i=1}^k m \Delta_i + \sum_{i=1}^k (n-mk) \Delta_i - \frac{(m \Delta_i)^2}{4}$$

$\brace{ \text{Explore} }$

exploit / commit phase

Proof: For sake of analysis $(\mu_1, \mu_2, \dots, \mu_k)$
 true means of arms

We know from regret decomposition

$$R_n = \sum_{i=1}^k \Delta_i \mathbb{E}[T_i(n)]$$

\uparrow
arms gap

fix any arm ' i '

$$T_i(n) = m + (n-mk), \text{ with } \text{Prob}(i \text{ got picked})$$

$$= m, \text{ with } 1 - \text{Prob}(i \text{ got picked})$$

\uparrow
explore phase

we pick it deterministically

$$\mathbb{E}[T_i(n)] = m + (n-mk) \text{Prob}(i \text{ got picked}) + m(1 - \text{Prob}(i \text{ got picked}))$$

= $m + (n-mk) \text{Prob}(\text{ } i^{\text{th}} \text{ arm turned out to be the best}$
 at the end of explore phase)

$$\leq m + (n-mk) \text{Prob}(\hat{\mu}_i(mk) > \max_{j \neq i} \hat{\mu}_j(mk))$$

arbitrary
ties

$$\text{Prob}(\text{arm } i \text{ beats all other arms}) \leq \text{Prob}(\text{arm } i \text{ beats arm 1})$$

event A event B

$$\text{Prob}(\text{arm } i \text{ beats arm 1})$$

$$= \text{Prob}(\hat{\mu}_i(mk) > \hat{\mu}_1(mk))$$

$$= \text{Prob}(\hat{\mu}_i(mk) - \hat{\mu}_1(mk) > 0)$$

$$= \text{Prob}(\underbrace{\hat{\mu}_i(mk)}_{\text{- sample mean}} - \underbrace{\mu_i}_{\text{- true mean}} + \underbrace{\mu_i - \mu_1}_{-\Delta_i} + \underbrace{\mu_1 - \hat{\mu}_1(mk)}_{\text{- sample mean - true mean}} > 0)$$

$$= \text{Prob} \left(\underbrace{\hat{\mu}_i(mk) - \mu_{i^*}}_Y + \underbrace{\mu_{i^*} - \hat{\mu}_1(mk)}_{Y'} > \Delta_{i^*} \right)$$

Fact 1: Y and Y' are $\frac{1}{\sqrt{m}}$ sub-Gaussian

$$\sqrt{\underbrace{\frac{1}{m^2} + \dots + \frac{1}{m^2}}_{m \text{ times}}}$$

$$Z = Y + Y' \text{ is } \sqrt{\frac{2}{m}} \text{ sub-Gaussian}$$

$$- \left(\frac{\Delta_{i^*}^2}{2 \cdot 2/m} \right) - \left(\frac{m \Delta_{i^*}^2}{4} \right)$$

$$\text{Prob}(Z \geq \Delta_{i^*}) \leq e^{-\left(\frac{m \Delta_{i^*}^2}{4}\right)} = e$$

$$R_n \leq m \underbrace{\sum_{i=1}^R \Delta_i}_{\text{explore}} + (n-mk) \underbrace{\sum_{i=1}^k \Delta_i}_{\text{commit}} \geq e$$

Basic Tools for design and analysis

(Prob of failure)

Tail Bound

$$(\varepsilon, m); \quad \delta_{(\varepsilon, m)} = 2e^{-(\frac{m\varepsilon^2}{2})}$$

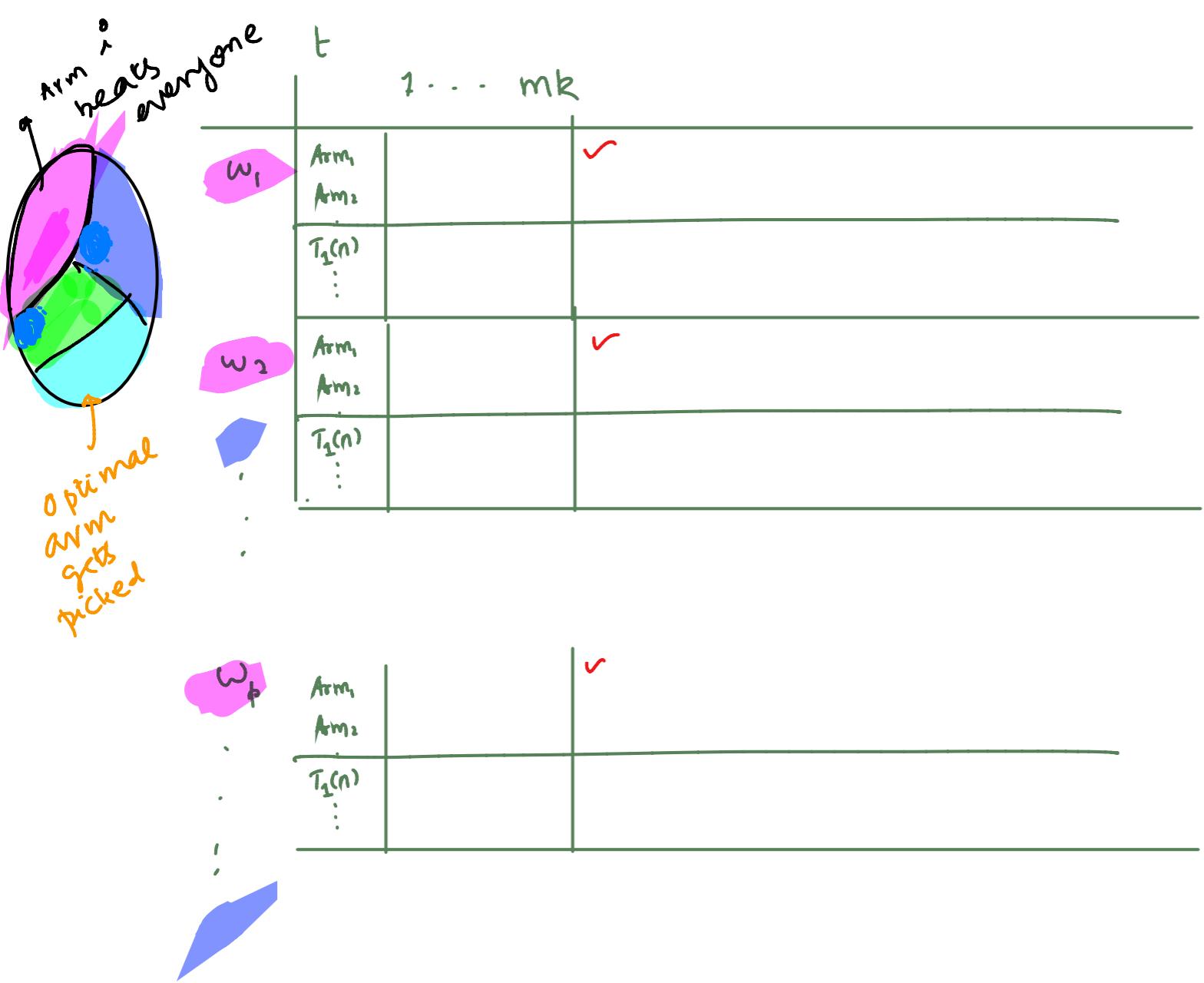
Confidence Interval

$$(m, \delta), \quad \varepsilon_{(m, \delta)} = \sqrt{\frac{2 \ln(2/\delta)}{m}}$$

Number of samples

$$(\varepsilon, \delta), \quad m_{(\varepsilon, \delta)} = \frac{2 \ln(2/\delta)}{\varepsilon^2}$$

ETC : we did not use above tools for design
we used it only for analysis



For this illustration
say Arm k is the best

$P(\{\omega : \text{Arm 1 wins for trial } \omega\})$



for sake of simplicity let us consider 2 arms

$$k=2, \Delta_1 = 0, \Delta_2 = \Delta$$

only design variable is 'm'

$$R_n \leq m\Delta + (n-2m)\Delta e^{-\left(\frac{m\Delta^2}{4}\right)} \quad (\text{say for } n > 2m)$$

$$\leq m\Delta + n\Delta e^{-\left(\frac{m\Delta^2}{4}\right)}$$

Suppose I know Δ , how many samples are optimal

$$-\left(\frac{m\Delta^2}{4}\right) \quad (\text{diff w.r.t } m)$$

$$\min_m m\Delta + n\Delta e^{-\left(\frac{m\Delta^2}{4}\right)}$$

$$\cancel{\Delta} + (n \cancel{\Delta}) \left(-\frac{\Delta^2}{4}\right) e^{-\left(\frac{m_*\Delta^2}{4}\right)} = 0$$

$$\frac{n\Delta^2}{4} e^{-\left(\frac{m_*\Delta^2}{4}\right)} = 1$$

$$e^{\frac{m_*\Delta^2}{4}} = \frac{n\Delta^2}{4}$$

$$m_* = \frac{4}{\Delta^2} \ln \left(\frac{n \Delta^2}{4} \right)$$

$$m_* = \max \left\{ 1, \left\lceil \frac{4}{\Delta^2} \ln \left(\frac{n \Delta^2}{4} \right) \right\rceil \right\}$$

Plug in m_*

$$- \left(m_* \frac{\Delta^2}{4} \right)$$

$$R_m \leq \underbrace{m_* \Delta}_{\text{Term 1}} + \underbrace{n \Delta e}_{\text{Term 2}}$$

$$\text{Term 1} \quad m_* \Delta = \Delta \max \left\{ 1, \left\lceil \frac{4}{\Delta^2} \ln \left(\frac{n \Delta^2}{4} \right) \right\rceil \right\}$$

$$\lceil x \rceil \leq 1 + x$$

$$m_* \Delta \leq \Delta \max \left\{ 1, 1 + \frac{4}{\Delta^2} \ln \left(\frac{n \Delta^2}{4} \right) \right\}$$

$$= \Delta \left(1 + \max \left\{ 0, \frac{4}{\Delta^2} \ln \left(\frac{n\Delta^2}{4} \right) \right\} \right)$$

$$= \Delta \left(1 + \frac{4}{\Delta^2} \max \left\{ 0, \ln \left(\frac{n\Delta^2}{4} \right) \right\} \right)$$

$$= \Delta + \frac{4}{\Delta} \max \left\{ 0, \ln \left(\frac{n\Delta^2}{4} \right) \right\}$$

(pay during commit)

Term 2

$$\begin{aligned} & - \left(m_* \frac{\Delta^2}{4} \right) \\ n\Delta e & = n\Delta e - \left(\frac{\Delta^2}{4} \max \left\{ 1, \left\lceil \frac{4}{\Delta^2} \ln \left(\frac{n\Delta^2}{4} \right) \right\rceil \right\} \right) \end{aligned}$$

$$b \leq a, \quad e^{-a} \leq e^{-b}$$

say

$$\max \left\{ 1, \left\lceil \frac{4}{\Delta^2} \ln \left(\frac{n\Delta^2}{4} \right) \right\rceil \right\} \geq \left\lceil \frac{4}{\Delta^2} \ln \left(\frac{n\Delta^2}{4} \right) \right\rceil$$

$$\geq \frac{4}{\Delta^2} \ln \left(\frac{n\Delta^2}{4} \right)$$

$$\text{Term 2} \leq n \Delta e^{-\left(\frac{\Delta^2}{4} + \frac{4}{\Delta^2} \ln\left(\frac{n\Delta^2}{4}\right)\right)}$$

$$= n \Delta \frac{\frac{4}{\Delta^2}}{n\Delta^2}$$

$$= \frac{4}{\Delta}$$

$$R_n \leq \text{Term 1} + \text{Term 2}$$

$$\leq \Delta + \underbrace{\frac{4}{\Delta} \max\left\{1_0, \ln\left(\frac{n\Delta^2}{4}\right)\right\}}_{\text{explore}} + \underbrace{\frac{4}{\Delta}}_{\text{commit}}$$

$$= \Delta + \frac{4}{\Delta} \left(1 + \max\left\{1_0, \ln\left(\frac{n\Delta^2}{4}\right)\right\}\right) \quad -(1) \rightarrow \infty$$

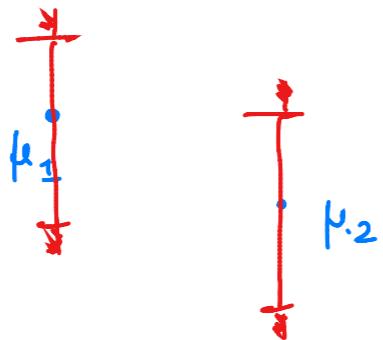
we know as $\Delta \rightarrow 0$ (1) $\rightarrow \infty$, so fact that was missing

$$R_n \leq n \Delta$$

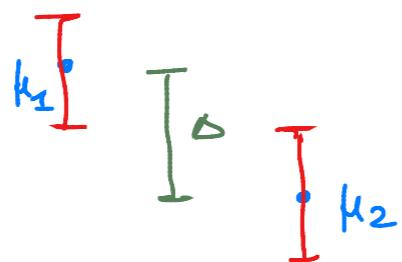
Fix prob of failure to be δ , to ease our case

let us consider $R=2$ (only 2 arms $\mu_1 > \mu_2$)

δ prob event
I am writing off
in my books



situation is not okay
because we don't have
resolution



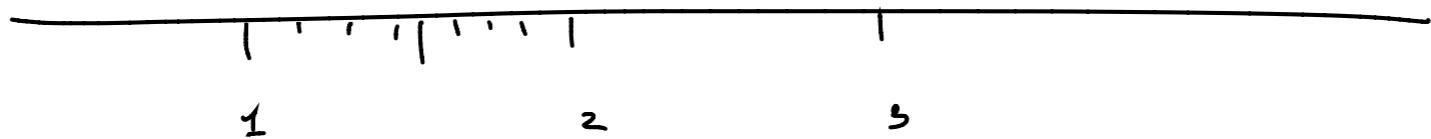
best case because we
have resolved the two
arms

Say we know Δ, δ we need roughly

$$\epsilon = \frac{\Delta}{2}$$

$$2 \frac{\ln(2/\delta)}{(\Delta/2)^2}$$

$$= \frac{\delta \ln(2/\delta)}{\Delta^2}$$



$$R_n \leq \min \left\{ n\Delta, \Delta + \frac{4}{\Delta} \left(1 + \max \left\{ \xi_0, \ln \left(\frac{n\Delta^2}{4} \right) \right\} \right)^2 \right\}$$

what is the worst Δ ? or what is the worst rate?

Player 1 : A ego Designer

Player 2 : Environment

Theorem : $R_n \leq \Delta + C\sqrt{n}$

Proof :

case I : Pick $\Delta \leq \frac{1}{\sqrt{n}}$

$$R_n \leq n\Delta \Rightarrow R_n \leq n \cdot \frac{1}{\sqrt{n}} \Rightarrow R_n \leq \sqrt{n}$$

case II : Pick $\Delta > \frac{1}{\sqrt{n}} \Rightarrow \frac{1}{\Delta} < \sqrt{n}$

$$R_n \leq \Delta + \frac{4}{\Delta} \left(1 + \max \{ \xi_0, \ln \left(\frac{n\Delta^2}{4} \right) \} \right)$$

$$= \Delta + \frac{4}{\Delta} + \frac{4}{\Delta} \max \{ \xi_0, \ln \left(\frac{n\Delta^2}{4} \right) \}$$

$$< \Delta + 4\sqrt{n} + \frac{4}{\Delta} \max \{ \xi_0, \ln \left(\frac{n\Delta^2}{4} \right) \}$$

Make ' Δ ' a variable and differentiate

$$R_n \leq \Delta + 4\sqrt{n} + 4 \cdot \max_{x>0} \frac{1}{x} \ln_+ \left(\frac{nx^2}{4} \right),$$

where $\ln_+(z) = \max \{ \xi_0, \ln(z) \}$

$$\frac{d}{dx} \left(\frac{1}{x} \ln_+ \left(\frac{nx^2}{4} \right) \right) = -\frac{1}{x^2} \ln_+ \left(\frac{nx^2}{4} \right) + \frac{1}{x} \frac{1}{\ln_+ \left(\frac{nx^2}{4} \right)} \frac{n(2x)}{x} = 0$$

$$\Rightarrow -\ln_+ \left(\frac{nx^2}{4} \right) + 2 = 0$$

$$\ln_+ \left(\frac{nx^2}{4} \right) = 2$$

$$\frac{n x_*^2}{4} = e^2$$

$$x_* = \sqrt{\frac{4}{n} e^2}$$

Plug x_*

$$\frac{1}{x_*} \ln\left(\frac{n x_*^2}{4}\right) = \frac{1}{\sqrt{\frac{4}{n} e^2}} \ln(e^2)$$

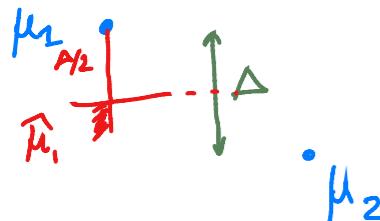
$$= \frac{2}{\sqrt{\frac{4}{n} e^2}} = \sqrt{n} e^{-1}$$

$$R_n \leq \Delta + 4\sqrt{n} + 4\sqrt{n}e^{-1}$$

$$R_n \leq \Delta + C\sqrt{n}$$

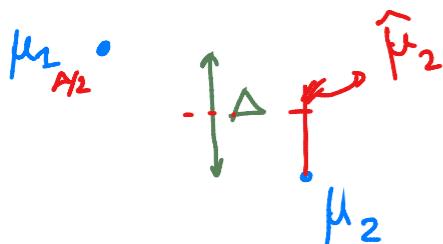
Till now, only design parameter was m , we found m
 An alternative way is to set aside budget for bad events
 best arm

Bad Event 1: the sample mean of arm 1 slips below its
 true mean by $\frac{\Delta}{2}$
 $- \left(\frac{m}{2} \left(\frac{\Delta}{2} \right)^2 \right)$



$$\text{Prob} \left(\hat{\mu}_1 < \mu_1 - \frac{\Delta}{2} \right) \leq e$$

Bad Event 2: the sample mean of arm 2 jumps above its
 true mean by $\frac{\Delta}{2}$
 $- \left(\frac{m}{2} \left(\frac{\Delta}{2} \right)^2 \right)$



$$\text{Prob} \left(\hat{\mu}_2 > \mu_2 + \frac{\Delta}{2} \right) \leq e$$

Total Budget = δ (fixing δ : prob of bad event)
 $- \left(\frac{m}{2} \left(\frac{\Delta}{2} \right)^2 \right)$ (interval is also fixed)

$$\delta = 2e$$

$$m = \max \left\{ 1, \left\lceil \frac{8}{\Delta^2} \ln \left(\frac{2}{\delta} \right) \right\rceil \right\}$$

Regret

$$\begin{aligned} R_m &\leq \underbrace{m \Delta}_{\text{explore}} + \text{Good Event} + \text{Bad Event} \\ &\leq m \Delta + (1-\delta) \cdot 0 + \delta (n-2m) \cdot \Delta \\ &\leq m \Delta + n \delta \Delta \\ &= \Delta \left\{ 1, \left\lceil \frac{8}{\Delta^2} \ln \left(\frac{2}{\delta} \right) \right\rceil \right\} + n \delta \Delta \\ &\leq \Delta + \cancel{\Delta} \frac{8}{\Delta^2} \ln \left(\frac{2}{\delta} \right) + n \delta \Delta \end{aligned}$$

Now our design parameter is '8'

$$\frac{8}{\Delta} \left(\frac{s_*}{2} \right) \left(\frac{2}{-s_*} \right) + n \Delta = 0$$

$$\frac{8}{\Delta s_*} = n \Delta$$

$$s_* = \frac{8}{n \Delta^2}$$

$$R_n \leq \Delta + \frac{8}{\Delta} \ln \left(\frac{2}{\delta_*} \right) + n \delta_x \Delta$$

$$= \Delta + \frac{8}{\Delta} \ln \left(\frac{2}{48/n\Delta^2} \right) + n \frac{8}{n\Delta^2} \cdot \Delta$$

$$= \Delta + \frac{8}{\Delta} \ln \left(\frac{n\Delta^2}{4} \right) + \frac{8}{\Delta}$$

$$R_n \leq \min \left\{ n\Delta, \Delta + \frac{8}{\Delta} \ln \left(\frac{n\Delta^2}{4} \right) + \frac{8}{\Delta} \right\}$$

$$R_n \leq \min \left\{ n\Delta, \Delta + \frac{4}{\Delta} \left(1 + \max \left\{ \xi_0, \ln \left(\frac{n\Delta^2}{4} \right) \right\} \right) \right\}$$

ETC

Variant 1) start with 'm' samples \rightarrow Looked at the bound \rightarrow optimise for m_*
(assuming Δ known)

Variant 2) we first set aside budget for failure (Δ known) \rightarrow decide number of samples $m_* (\delta)$
Optimal for a given (Δ, δ) skip 2 jumps

\downarrow

optimised for ' δ ' to find δ_*

\downarrow

$m_* (\delta_*)$

ETC: ' Δ ' is not known

(building upon variant 2's idea)
 $\bullet \sigma \leq 1, \Delta \leq 1$

confidence interval, number of samples required, Prob failure
(ϵ) (m) (s)

$$- \left(\frac{m\epsilon^2}{2} \right)$$

-- Previously
we had $\frac{\Delta}{2}$

Bad Event 1:

$$\text{Prob}(\hat{\mu}_1 < \mu_1 - \epsilon) \leq e$$

skip of best
arm

$$- \left(\frac{m\epsilon^2}{2} \right)$$

Bad Event 2:

$$\text{Prob}(\hat{\mu}_2 > \mu_2 + \epsilon) \leq e$$

jump of the
bad arm

Total Budget for failure prob = δ

$$- \left(\frac{m\epsilon^2}{2} \right)$$

$$\delta = 2e$$

$$\epsilon = \sqrt{\frac{2 \ln(2/\delta)}{m}}$$

Moral: Set aside $\delta \cdot \Delta \cdot (n-2m)$ as a write off / subscription fee

Having written off bad events, let's analyse good event

case 1: **Somehow** for the (m, δ) we chose
and ' Δ' that the environment chose

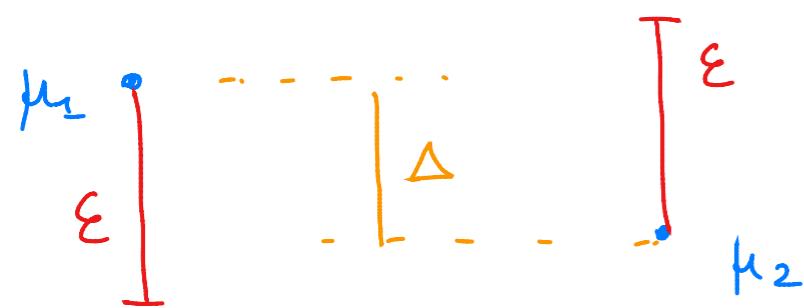


this is a good thing → During commit phase we
don't pay any penalty

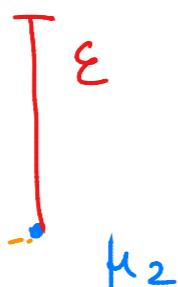
case 2: (m, δ) we chose and ' Δ' environment chose

(new case)
are such that gap does not get
when compared to
' Δ' known)
resolved

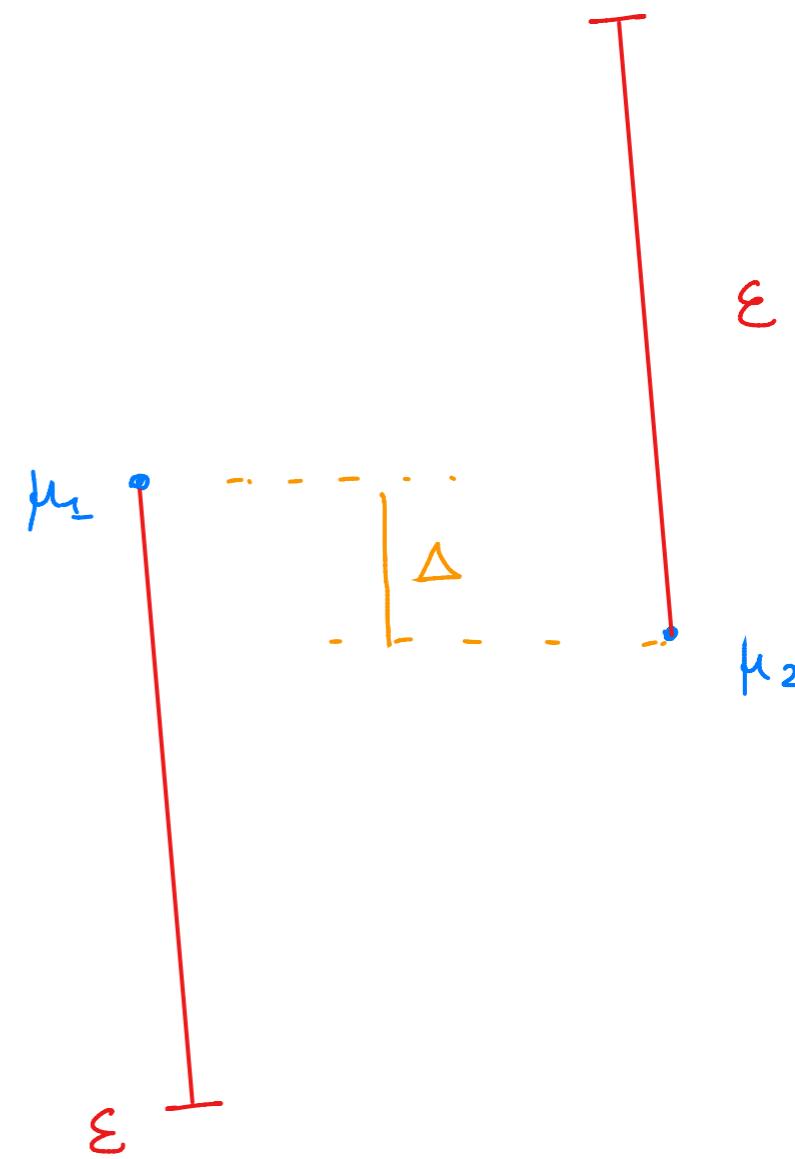
Arm 1



Arm 2



Arm 1



Arm 2

ϵ

Moral: Potentially one could choose the 'bad' arm even during the 'good' event. But note that penalty per round is not more than 2ϵ

$$R_m \leq \underbrace{m\Delta}_{\text{explore}} + \underbrace{\text{Penalty @ Good Event} + \text{Penalty @ Bad Event}}_{\substack{\text{case 1} \\ \text{case 2}}} \quad (\text{loose way of writing})$$

$$R_m \leq m\Delta + \underbrace{(n-2m)}_{\text{exploit}} \left(\underbrace{(1-\delta)}_{\text{Good}} \left[\underbrace{0 + 2\varepsilon}_{\text{WST @ Good}} \right] + \underbrace{\delta \cdot \Delta}_{\substack{\text{Bad cost @ Bad}}} \right)$$

$$R_m \leq m\Delta + n^2\varepsilon + n\delta \cdot \Delta \quad (n \gg 2m, 1-\delta < 1)$$

$$R_m \leq m + n^2\varepsilon + n\delta \quad (\Delta < 1)$$

$$\leq m + n^2 \sqrt{\frac{2 \ln(2/\delta)}{m}} + n \cdot \delta$$

To minimise the LHS, diff w.r.t m

$$1 + n \cancel{\sqrt{2 \ln(2/\delta)}} \left(-\frac{1}{2}\right) \frac{1}{m_*^{3/2}} = 0$$

$$1 = \frac{n \sqrt{2 \ln(2/\delta)}}{m_*^{3/2}}$$

$$m_*^{3/2} = n \sqrt{2 \ln(2/\delta)} = n (2 \ln(2/\delta))^{1/2}$$

$$m_* = n^{2/3} (2 \ln(2/\delta))^{1/3}$$

Note $n^{2/3} \propto \ln(\frac{m_\Delta^2}{4})$

$$R_n \leq n^{2/3} (2 \ln(2/\delta))^{1/3} + n^2 \sqrt{\frac{2 \ln(2/\delta)}{n^{2/3} (2 \ln(2/\delta))^{1/3}}} + n \cdot \delta$$

$$R_n \leq n^{2/3} (2 \ln(2/\delta))^{1/3} + 2n^{2/3} (2 \ln(2/\delta))^{1/3} + n \cdot \delta$$

$$R_n \leq 3n^{2/3} (2 \ln(2/\delta))^{1/3} + n \cdot \delta \quad \delta = \frac{1}{n}$$

$$R_n \leq 3n^{2/3} (2 \ln(2n))^{1/3} + 1$$

Giving forward: We will need to use the intervals also in our design.

Upper Confidence Bound (UCB) Algorithm

Things that went wrong with Explore-then-Commit

- Should not commit based on finite samples. Common sense reasoning tells us that we need infinite number of samples to learn the true mean.

We should continually obtain new samples
↓
Explore and Exploit simultaneously

$t_i(t)$ = total number of times we pick arm i
in 't' round

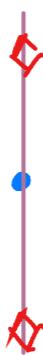
Should not

max but at ' m '.

- we need to use the confidence interval explicitly in the algorithm

statement about sample means

Arm 1



Arm 2



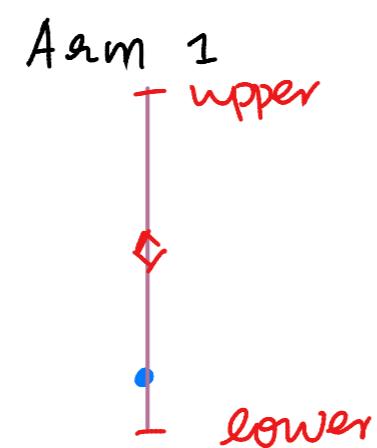
- True mean
- ◊ Sample mean

w.p. $> 1 - \delta$
after ' m'
samples

$$\hat{\mu} \in \left[\mu - \sqrt{\frac{2 \ln(2/\delta)}{m}}, \mu + \sqrt{\frac{2 \ln(2/\delta)}{m}} \right]$$

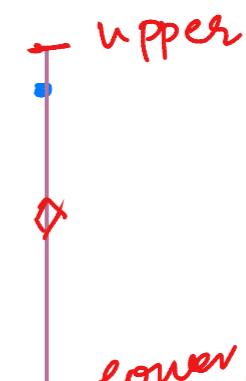
- we need to use the confidence interval explicitly in the algorithm

Statement about True means



Arm 2

- True mean
- ◊ Sample mean



w.p. $> 1 - \delta$
after ' m'
samples

$$\mu \in [\hat{\mu} - \sqrt{\frac{2 \ln(2/\delta)}{m}}, \hat{\mu} + \sqrt{\frac{2 \ln(2/\delta)}{m}}]$$

- we have located the true mean to belong to an interval
- Deciding with true is ideal
- we don't have true \Rightarrow we have to substitute in place with approximate value
- Approximate value belongs to the interval
- Qn : Where in the interval should we pick
An : Definitely not bang in the middle, i.e.,
the sample mean (picking sample mean = ETC)
- choices left probe up or probe down

Why is it a bad idea to use the lower bound?

Arm 1



What the
environment gave

Arm 2



Arm 1

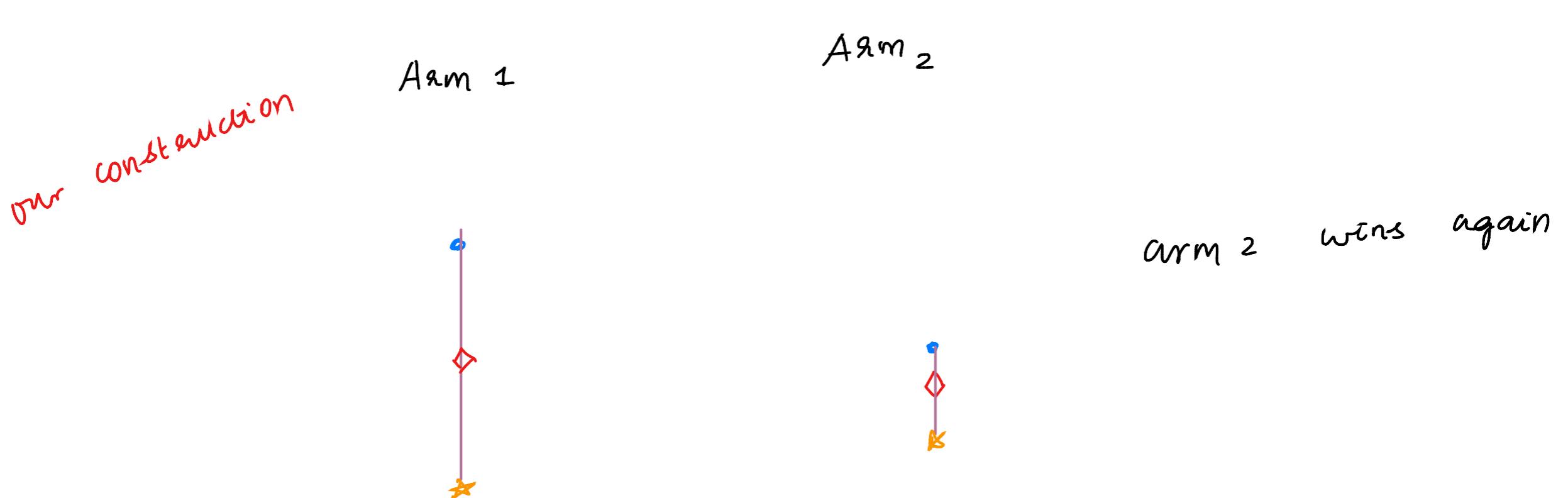
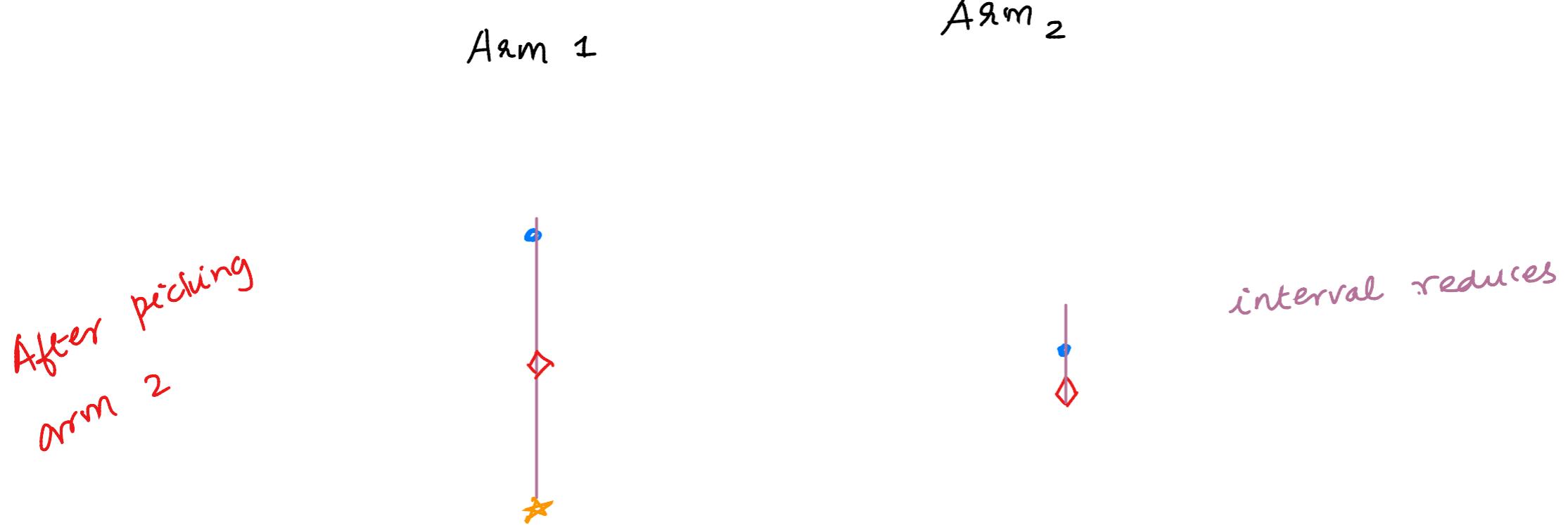


our construction
• not known

Arm 2



If we pick the
best lowest value
arm₂ wins



Moral : Lower confidence bound has two negatives going under the hood.

Why is it a good idea to use the upper bound?

Arm 1



•

◊

What the
environment gave

Arm 2



•

◊

our construction
• not known

Arm 1



•

◊

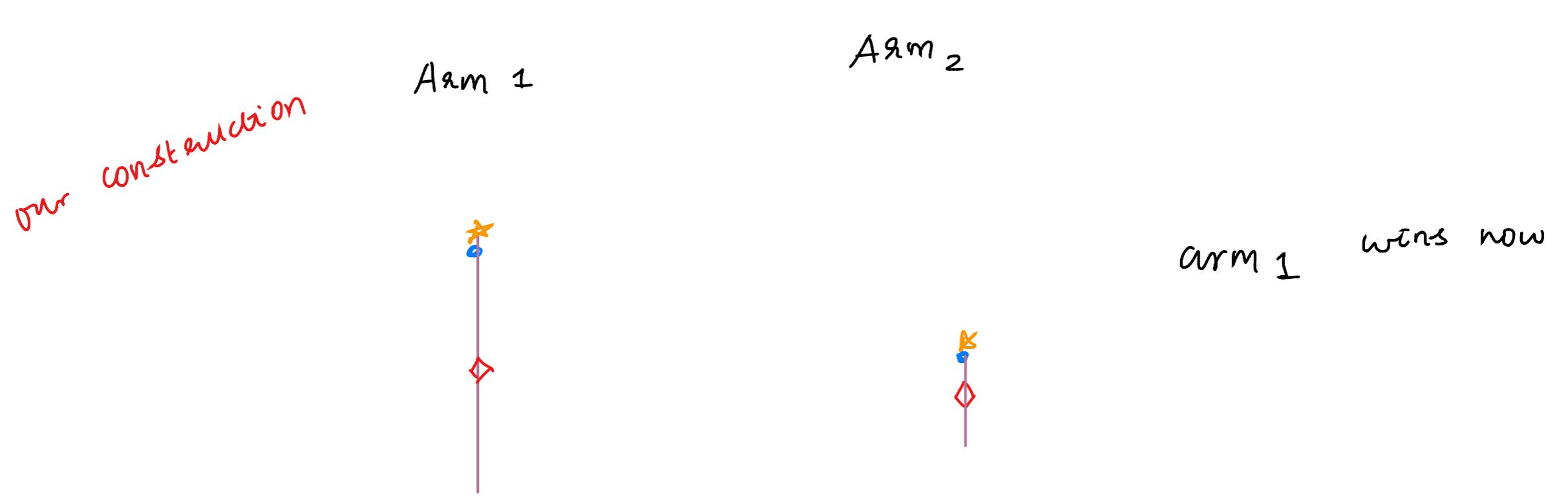
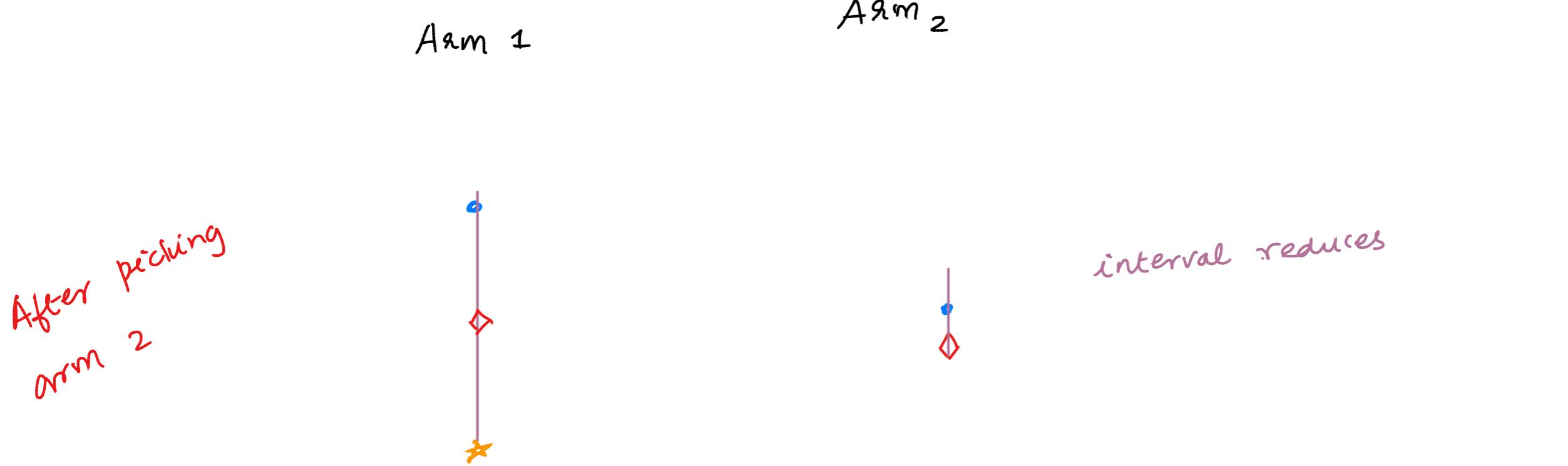
Arm 2



•

◊

If we pick the
best upper value
Arm₂ wins



Moral: Lower confidence bound has two negatives going under the hood.

UCB Algorithm

- Play each arm once

$$\hat{\mu}_i(t-1) = \frac{\sum_{s=1}^{t-1} X_s \mathbb{I}_{\{A_s = i\}}}{T_i(t-1)}, \text{ recall } T_i(t-1) = \sum_{s=1}^{t-1} \mathbb{I}_{\{A_s = i\}}$$

$$UCB_i(t-1) = \hat{\mu}_i(t-1) + \sqrt{\frac{2 \ln(1/\delta)}{T_i(t-1)}}$$

$$\text{Play } A_t = \arg \max_i UCB_i(t-1)$$

$\delta = \frac{1}{(n+1)^2}$

$$\sqrt{\frac{4 \ln(n+1)}{T_i(t-1)}}$$

Issue:

$$\sqrt{\frac{2 \ln(2/\delta)}{T_i(t-1)}}$$

vs

$$\sqrt{\frac{2 \ln(2/\delta)}{m}}$$

\uparrow
random variable

\uparrow
fixed

Say there are two arms at $t = 11$, $T_i(t-1) = T_i(10) \in \underbrace{\{1, \dots, 9\}}_{\leq t \text{ cases}}$

at time t

Bad Event 1 : Best Arm (say Arm 1 w.l.o.g) slips at time t

$$\text{Prob} \left(\hat{\mu}_1(t) < \mu_1 - \sqrt{\frac{2 \ln(1/\delta)}{T_1(t)}} \right) \leq t^\delta$$

as opposed to fixed samples ' m '

$$\text{Prob} \left(\hat{\mu}_m < \mu - \sqrt{\frac{2 \ln(1/\delta)}{m}} \right) \leq \delta$$

\uparrow t such bounds
invoked for
the varying
number of
samples that
participated
in $\hat{\mu}_1(t)$

Bad Event 2: Arm₂ jumps at time t

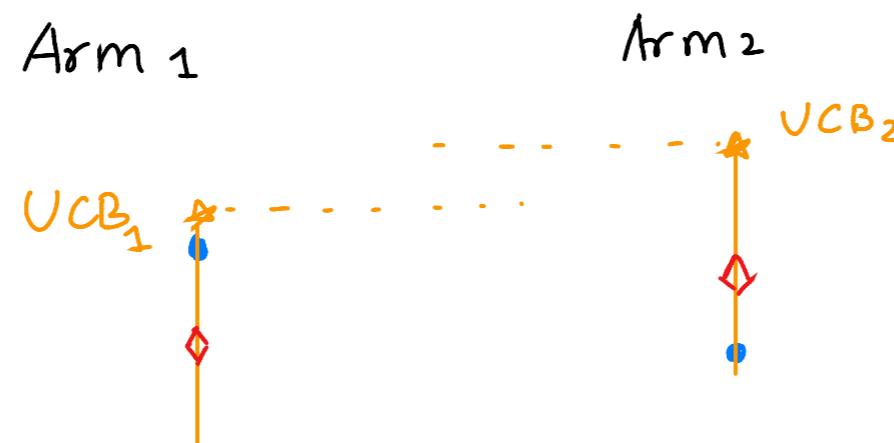
$$\text{Prob} \left(\hat{\mu}_2(t) > \mu_2 + \sqrt{\frac{2 \ln(1/\delta)}{T_2(t)}} \right) \leq t \delta$$

total penalty we pay over $t=1, \dots, n$ rounds due to bad events

$$= 2\delta + 2\delta \cdot 2 + 2\delta \cdot 3 + \dots + 2\delta n$$

$$= 2\delta (1+2+\dots+n) = 2\delta \frac{n(n+1)}{2} \leq (n+1)^2 \delta$$

On Good Event: How many times do I pick Arm₂

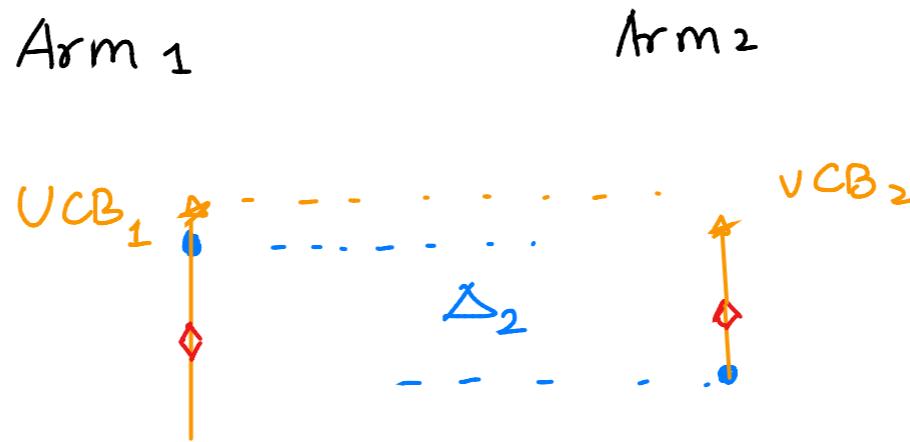


on good event , $\hat{\mu}_1(t) \geq \mu_1 - \sqrt{\frac{2 \ln(1/\delta)}{T_1(t)}}$

$$UCB_1(t) = \hat{\mu}_1(t) + \sqrt{\frac{2 \ln(1/\delta)}{T_1(t)}}$$

$$\geq \mu_1$$

we want the following picture



$$2 \sqrt{\frac{2 \ln(1/\delta)}{T_2(t-1)}} \leq \Delta_2$$

$$\frac{8 \ln(1/\delta)}{T_2(t-1)} \leq \Delta_2^2$$

$$T_2(t-1)$$

$$\geq \left[\frac{8 \ln(1/\delta)}{\Delta_2^2} \right]$$
$$\geq 1 + \frac{8 \ln(1/\delta)}{\Delta_2^2}$$

$$\text{Regret}_n = R_n \leq \Delta_2 + (n+1)^2 \delta \Delta_2 + \Delta_2 \left(1 + \frac{8 \ln(1/\delta)}{\Delta_2^2} \right)$$

↑
 exploration
 each arm gets
 picked once

↑
 penalty
 for bad
 events

↑
 during further
 execution of UCB

$$\delta = \frac{1}{(n+1)^2}$$

$$R_n \leq \Delta_2 + \Delta_2 + \Delta_2 + \frac{8 \ln \left(\frac{1}{\delta(n+1)^2} \right)}{\Delta_2}$$

$$= \Delta_2 \left(3 + 16 \frac{\ln(n+1)}{\Delta_2^2} \right)$$

For k arms

$$R_n \leq 3 \sum_{i=2}^k \Delta_i + 16 \ln(n+1) \sum_{i=2}^k \frac{1}{\Delta_i}$$

Markov Decision Process (MDP)

Finite State
Action Horizon

$$MDP = \langle S, A, P, R \rangle$$

- S : state space

- A : action space

- P : probability transition

$$P(s_{\text{cur}}, a, s_{\text{next}})$$

- R : reward

$$R: S \rightarrow \mathbb{R}$$

or

$$S \times A \rightarrow \mathbb{R}$$

or

$$S \times A \times S \rightarrow \mathbb{R}$$

Horizon

\downarrow
H

$$\mathbb{E} \left[\sum_{t=1}^H R(s_t, a_t, s_{t+1}) \right]$$

Grid World

s^{13}	s^{14}	s^{15}	s^{16}
7	1	6	2
s^1	s^{10}	s^{11}	s^{12}
0	0	-4	8
s^5	s^6	s^7	s^8
1	-10	10	5
s^1	s^2	s^3	s^4
2	0	-5	3

- $S = \{s^1, \dots, s^{16}\}$
- $A = \{\leftarrow, \rightarrow, \uparrow, \downarrow, \textcircled{0}\}$
stay
- $\text{Prob}(s_{t+1} = s^j \mid s_t = s^i, a_t = a) = P(s^i, a, s^j)$

Probability of moving from s^i to s^j by action a

Each action is successful with $p = 0.9$

$$P(s^2, \uparrow, s^5) = 0.9, \quad P(s^1, \uparrow, s^1) = 0.1$$

$$P(s^2, \textcircled{0}, s^2) = P(s^2, \textcircled{0}, s^5) = \frac{0.1}{2}$$

- Bandit = Stateless MDP

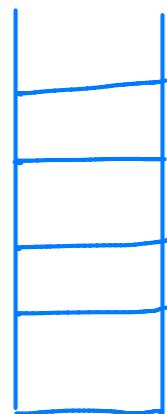
Policy $\pi = \{\pi_1, \dots, \pi_H\}$, where each $\pi_t : S \rightarrow A$

A policy specifies how to act at time 't' in any given state.

When one 'acts' according to π , at t $a_t = \pi_t(s_t)$

Toy Example : Inventory management one day before end of sale

Stock size = 5



cost price = CP

selling price = SP

transportation price = TP

next day demand forecast probabilities

$$d_t = [d(0) \ d(1) \ \dots \ d(5)]$$

$$\sum_{i=0}^5 d_t(i) = 1$$

$$\text{Action} = \{0, 1, 2, 3, 4, 5\}$$

new item to be procured

$$S = \{0, 1, 2, 3, 4, 5\}$$

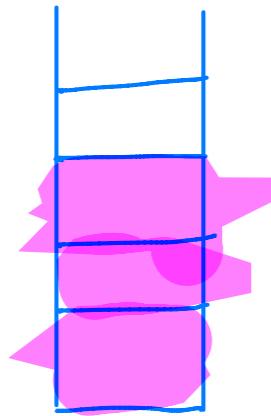
no item
in stack

↑
stack
is full

Suppose at the end of penultimate day (one day before)

3 items left

$$\text{actions } \in \{0, 1, 2\}$$



$Q(s, a)$: The expected value on end of last day
for ordering 'a' items, when we have
's' items in the stack one day
before

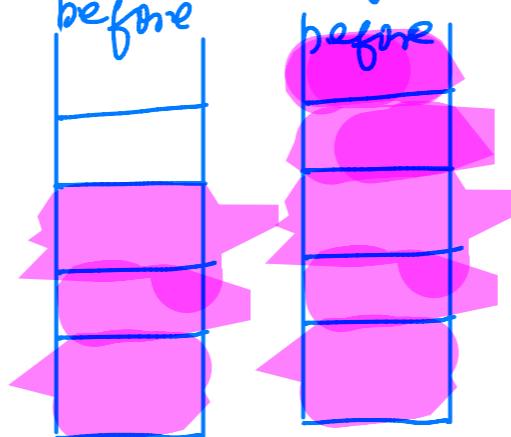
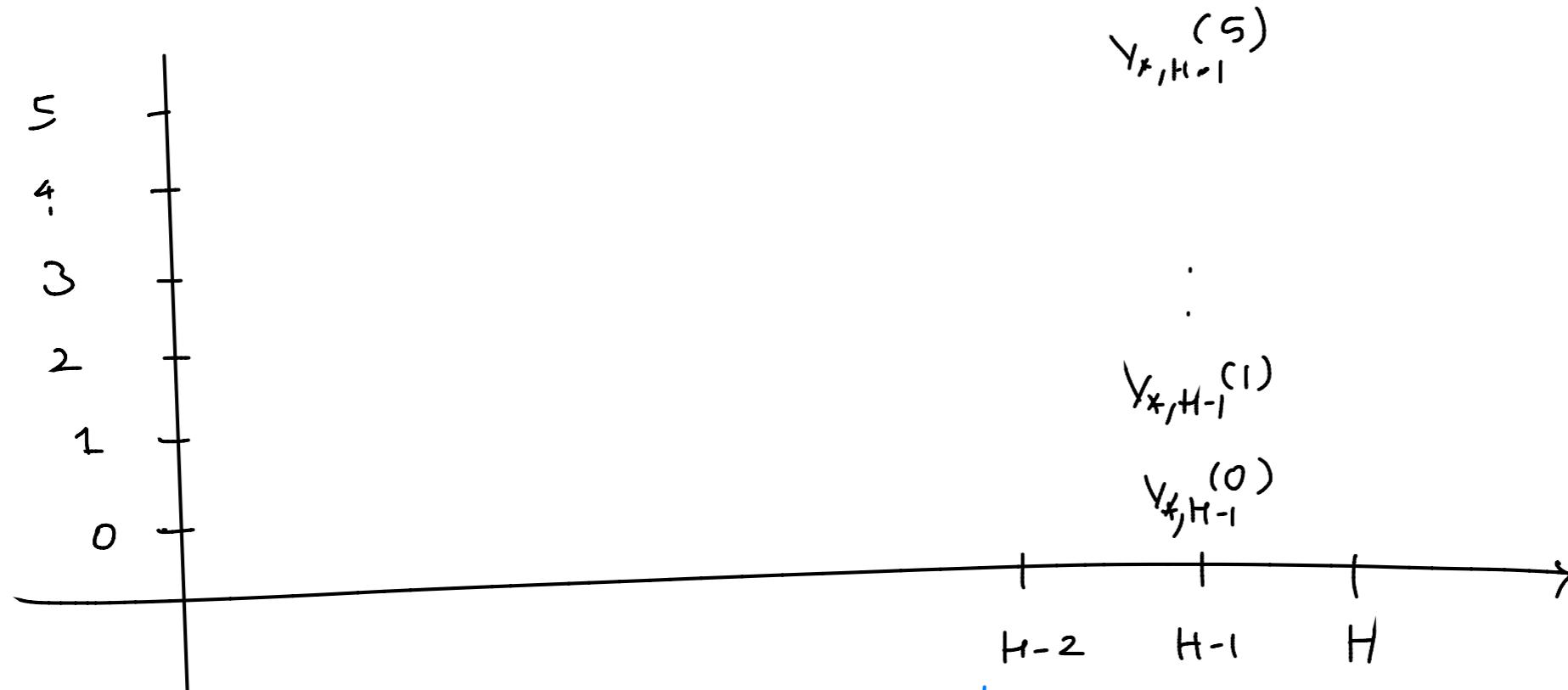
$$Q(s=3, a=0) = 0 \times d(0) + sp \times d(1) + 2.sp \times d(2) + 3.sp (d(3) + d(4) + d(5))$$

$$Q(s=3, a=1) = -c_f - t_p + 0 \times d(0) + sp \times d(1) + 2.sp \times d(2) + 3.sp \times d(3) \\ + 4.sp (d(4) + d(5))$$

$$Q(s=3, a=2) = -2C_F - t_P + 0 \times d(0) + 3p \times d(1) + 2.sp \times d(2) + 3.sp \times d(3) \\ + 4.sp \times d(4) + 5.sp \times d(5)$$

(policy) $\pi_{\star}(s) = \arg \max_a Q(s, a)$

$$V_{\star}(s) = \max_a Q(s, a) = Q(s, \pi_{\star})$$



$$V_{*,H-2}(s) = \max_a Q_{*,H-2}(s, a)$$

$$\pi_{*,H-2}(s) = \arg \max_a Q_{*,H-2}(s, a)$$

$$\begin{aligned}
Q_{*,H-2}(3,2) &= -2sp - bp + d_{H-1}(0) [0 + V_{*,H-1}(5)] \\
&\quad + d_{H-1}(1) [sp + V_{*,H-1}(4)] \\
&\quad + d_{H-1}(2) [2sp + V_{*,H-1}(3)] \\
&\quad + d_{H-1}(3) [3sp + V_{*,H-1}(2)] \\
&\quad + d_{H-1}(4) [4sp + V_{*,H-1}(1)] \\
&\quad + d_{H-1}(5) [5sp + V_{*,H-1}(0)]
\end{aligned}$$

$\pi_{*,H-1}$, $Q_{*,H-1}$, $V_{*,H-1}$

$\pi_{*,H-2}$, $Q_{*,H-2}(s, a)$, $V_{*,H-2}$

Bellman Equation

$$Q_{*,t}(s,a) = \sum_{s'} P(s,a,s') \left(R(s,a,s') + \underbrace{V_{*,t+1}(s')}_{\text{next state}} \right)$$

$$V_{*,t}(s) = \max_a Q_{*,t}(s,a)$$

$$\pi_{*,t}(s) = \arg \max_a Q_{*,t}(s,a)$$

$$V_{*,t}(s) = \max_a \sum_{s'} P(s,a,s') \left(R(s,a,s') + V_{*,t+1}(s') \right)$$

$$Q_{*,t}(s,a) = \sum_{s'} P(s,a,s') \left(R(s,a,s') + \max_{a'} Q_{*,t+1}(s',a') \right)$$

Deterministic Grid World

s^1	s^8	s^9
-2	2	0
s^4	s^5	s^6
3	-1	2
s^1	s^2	s^3
1	0	-2

$$R: S \rightarrow A$$

terminating time

$$t = 1, 2, \dots, H$$

$$S = \{s^1, \dots, s^9\}$$

$$V_*(s) = \max \sum_{t=1}^H R(s_t) \mid s_1 = s$$

$$A = \{\leftarrow, \rightarrow, \uparrow, \downarrow, \circlearrowright\}$$

$$\Pi = \{\pi_1, \dots, \pi_{H-1}\}$$

$$\text{Say } H = 1$$

$$V_{*,1} =$$

s^1	s^8	s^9
-2	2	0
s^4	s^5	s^6
3	-1	2
s^1	s^2	s^3
1	0	-2

$$\pi_* = \{\text{empty}\}$$

No actions to make / actions made does not matter because game ends at $H = 1$

Say $H = 2$

$$\pi_x = \{ \pi_{*,1} \}$$

$$V_{*,2} =$$

s^1	s^2	s^3
s^4	s^5	s^6
s^7	s^8	s^9
s^1	s^2	s^3
s^1	s^2	s^3

s^1	s^2	s^3
s^4	s^5	s^6
s^7	s^8	s^9
s^1	s^2	s^3
s^1	s^2	s^3

Reward

$$V_{*,1} =$$

s^1	s^2	s^3
s^4	s^5	s^6
s^7	s^8	s^9
s^1	s^2	s^3
s^1	s^2	s^3

$$\pi_{*,1} =$$

s^1	s^2	s^3
s^4	s^5	s^6
s^7	s^8	s^9
s^1	s^2	s^3
s^1	s^2	s^3

$$V_{*,2}(s^1) = \max_a Q_{*,1}(s^1, a) = \max_{\uparrow, \rightarrow, \ominus} \{ Q_{*,1}(s^1, \uparrow), Q_{*,1}(s^1, \rightarrow), Q_{*,1}(s^1, \ominus) \}$$

$$\pi_{*,1}(s^1) = \arg \max_a Q_{*,1}(s^1, a) = \max \{ 1+3, 1+0, 1+1 \} = 4$$

Say $t = 3$

$$\pi_k = \{ \pi_{k,1}, \pi_{k,2} \}$$

$$V_{k,3} =$$

s^1	s^2	s^3
-2	2	0
s^4	s^5	s^6
3	-1	2
s^1	s^2	s^3
1	0	-2

s^1	s^2	s^3
-2	2	0
s^4	s^5	s^6
3	-1	2
s^1	s^2	s^3
1	0	-2

Reward

$$V_{k,2} =$$

s^1	s^2	s^3
1	4	2
s^4	s^5	s^6
6	2	4

$$\pi_{k,2} =$$

\downarrow	s^1	s^2	s^3
s^4	s^5	s^6	\leftarrow
0	\leftarrow	0	\downarrow
s^1	s^2	s^3	

$$V_{k,1} =$$

s^1	s^2	s^3
4	6	4
s^4	s^5	s^6
9	5	6

$$\pi_{k,1} =$$

\downarrow	s^1	s^2	s^3
s^4	s^5	s^6	\leftarrow
0	\leftarrow	0	\downarrow
s^1	s^2	s^3	

$$V_{k,1}(s^1) = \max \{ Q_{k,1}(s^1, \uparrow), Q_{k,1}(s^1, \rightarrow), Q_{k,1}(s^1, \Theta) \}$$

$$= \max \{ 1 + b, 1 + 1, 1 + 4 \}$$

Say $k = 4$
 $\pi_k = \{\pi_{k,1}, \pi_{k,2}, \pi_{k,3}\}$

$v_{k,4} =$

s^1	s^2	s^3
s^4	$\frac{1}{2}$	$\frac{1}{2}$
s^5	$\frac{1}{3}$	$\frac{1}{3}$
s^6	$\frac{1}{3}$	$\frac{1}{3}$
s^7	$\frac{1}{4}$	$\frac{1}{4}$

$v_{k,3} =$

s^1	s^2	s^3
s^4	$\frac{1}{2}$	$\frac{1}{2}$
s^5	$\frac{1}{3}$	$\frac{1}{3}$
s^6	$\frac{1}{3}$	$\frac{1}{3}$
s^7	$\frac{1}{4}$	$\frac{1}{4}$

$v_{k,2} =$

s^1	s^2	s^3
s^4	$\frac{1}{2}$	$\frac{1}{2}$
s^5	$\frac{1}{3}$	$\frac{1}{3}$
s^6	$\frac{1}{3}$	$\frac{1}{3}$
s^7	$\frac{1}{4}$	$\frac{1}{4}$

$v_{k,1} =$

s^1	s^2	s^3
s^4	$\frac{1}{2}$	$\frac{1}{2}$
s^5	$\frac{1}{3}$	$\frac{1}{3}$
s^6	$\frac{1}{3}$	$\frac{1}{3}$
s^7	$\frac{1}{4}$	$\frac{1}{4}$

$\pi_{k,3} =$

s^1	s^2	s^3
s^4	$\frac{1}{2}$	$\frac{1}{2}$
s^5	$\frac{1}{3}$	$\frac{1}{3}$
s^6	$\frac{1}{3}$	$\frac{1}{3}$
s^7	$\frac{1}{4}$	$\frac{1}{4}$

$\pi_{k,2} =$

s^1	s^2	s^3
s^4	$\frac{1}{2}$	$\frac{1}{2}$
s^5	$\frac{1}{3}$	$\frac{1}{3}$
s^6	$\frac{1}{3}$	$\frac{1}{3}$
s^7	$\frac{1}{4}$	$\frac{1}{4}$

$\pi_{k,1} =$

s^1	s^2	s^3
s^4	$\frac{1}{2}$	$\frac{1}{2}$
s^5	$\frac{1}{3}$	$\frac{1}{3}$
s^6	$\frac{1}{3}$	$\frac{1}{3}$
s^7	$\frac{1}{4}$	$\frac{1}{4}$

Exercise : Find R such that $\pi_{k,1} \neq \pi_{k,2}$

s^1	s^2	s^3
s^4	$\frac{1}{2}$	$\frac{1}{2}$
s^5	$\frac{1}{3}$	$\frac{1}{3}$
s^6	$\frac{1}{3}$	$\frac{1}{3}$
s^7	$\frac{1}{4}$	$\frac{1}{4}$

Reward

0	0	10
0	0	0
1	0	0

Reward

Model Cricket as Finite Horizon MDP (WASP)
Scott Brooker (NZ)

Batting First

- assume : no extras (wide, no ball), one batter at a time

- state : w_t - wickets left at time t

- action : $A = \{1, 2, 3, 4, 6\}$

- Total Balls = 300 ($= H$)

at each ball (time), we pick $a_t \in A$

- either a wicket falls $p_{out}(w_t, a_t)$ (average)

- if the wicket does not fall run = a_t
gets scored with $p_{run}(w_t, a_t)$ (strike rate)

$$w = 10 \rightarrow \text{Opener} \rightarrow p_{\text{out}}^{\min} = [0.01, 0.02, 0.03, 0.1, 0.3]$$

$$w = 1 \rightarrow \text{tail ender} \rightarrow p_{\text{out}}^{\max} = [0.1, 0.2, 0.3, 0.5, 0.7]$$

For middle order we just interpolate

$$p_{\text{out}}(w, a) = p_{\text{out}}^{\max}(a) + \left(p_{\text{out}}^{\min}(a) - p_{\text{out}}^{\max}(a) \right) \frac{(w-1)}{9}$$

opener

$$p_{\text{out}}^{\max} = [1, 0.9, 0.85, 0.8, 0.75, 0.7]$$

tail

$$p_{\text{out}}^{\min} = [1, 0.4, 0.35, 0.3, 0.25, 0.2]$$

For middle order

$$p_{\text{out}}(w, a) = p_{\text{out}}^{\min}(a) + \left(p_{\text{out}}^{\max}(a) - p_{\text{out}}^{\min}(a) \right) \frac{(w-1)}{9}$$

Last Ball

Q-values

$$Q_{*,1}(\omega, a) = \underset{\text{out}}{p}(\omega, a)[0] + (1 - p_{\text{out}}(\omega, a)) [p_{\text{run}}(\omega, a) \cdot a]$$

Values

$$V_{*,1}(\omega) = \max_a Q_{*,1}(\omega, a) = \max_a \underset{\text{out}}{p}(\omega, a)[0] + (1 - p_{\text{out}}(\omega, a)) [p_{\text{run}}(\omega, a) \cdot a]$$

Policy

$$\pi_{*,1}(\omega) = \operatorname{argmax}_a Q_{*,1}(\omega, a) = \operatorname{argmax}_a \underset{\text{out}}{p}(\omega, a)[0] + (1 - p_{\text{out}}(\omega, a)) [p_{\text{run}}(\omega, a) \cdot a]$$

Any other ball

$$Q_{*,t+1}(\omega, a) = \underset{\text{out}}{p}(\omega, a)[0 + V_{*,t}^{(\omega-1)}] + (1 - p_{\text{out}}(\omega, a)) [p_{\text{run}}(\omega, a) \cdot a + V_{*,t}(\omega)]$$

$$V_{*,t+1}(\omega) = \max_a Q_{*,t+1}(\omega, a)$$

$$\pi_{*,t+1}(\omega) = \operatorname{argmax}_a Q_{*,t+1}(\omega, a)$$

One Step Bellman
Operator

Worst Case Regret Bound for UCB

(coming up with an environment that will be challenging for algorithm)
 ↓
 struggles the most

$$\text{Regret in } n \text{ rounds} = R_n = \sum_i \Delta_i \mathbb{E}[T_i(n)]$$

Split the arms into two groups

Group 1: $\{i : \Delta_i < \Delta\}$ (we will optimise for Δ later)

Group 2: $\{i : \Delta_i \geq \Delta\}$

$$R_n = \sum_{i: \Delta_i < \Delta} \Delta_i \mathbb{E}[T_i(n)] + \sum_{i: \Delta_i \geq \Delta} \Delta_i \mathbb{E}[T_i(n)]$$

$$\leq n \Delta + \sum_{i: \Delta_i \geq \Delta} \left(3 \Delta_i + \frac{16 \ln(n\tau)}{\Delta_i} \right)$$

$$\leq n \Delta + 3 \sum_i \Delta_i + \frac{16 \ln(n\tau)}{\Delta}$$

Balance $n \Delta$ vs $\frac{16 \ln(n\tau)}{\Delta}$

$$n \Delta_* = \frac{16k \ln(n+1)}{\Delta_*}$$

$$\Delta_* = \sqrt{\frac{16k \ln(n+1)}{n}} = 4 \sqrt{k \frac{\ln(n+1)}{n}}$$

$$R_n \leq n \Delta_* + \frac{16k \ln(n+1)}{\Delta_*} + 3 \sum_i \Delta_i^*$$

$$= 2n \Delta_* + 3 \sum_i \Delta_i^*$$

$$= 2 \cdot n \cdot 4 \sqrt{k \frac{\ln(n+1)}{n}} + 3 \sum_i \Delta_i^*$$

$$= 8 \sqrt{n k \ln(n+1)} + 3 \sum_i \Delta_i^*$$

But for the $\ln(n)$ factor the regret grows \sqrt{n}

We used $S = \frac{1}{(n+1)^2}$ which implicitly assume we know n .

so we can let $\delta_t = \frac{1}{t^3}$

$$(\text{per round}): 2 + S = 2 \cdot t \cdot \frac{1}{t^3} = \frac{2}{t^2}$$

$$(\text{all rounds}): \sum_{t=1}^{\infty} \frac{2}{t^2} = \frac{\pi^2}{3}$$

$$UCB_i(t-1) = \hat{f}_i(t-1) + \sqrt{\frac{2 \ln(1/\delta)}{T_i(t-1)}}, \text{ put } \delta_t = \frac{1}{t^3}$$

$$UCB_i(t-1) = \hat{f}_i(t-1) + \sqrt{\frac{6 \ln(t)}{T_i(t-1)}}$$

(*)

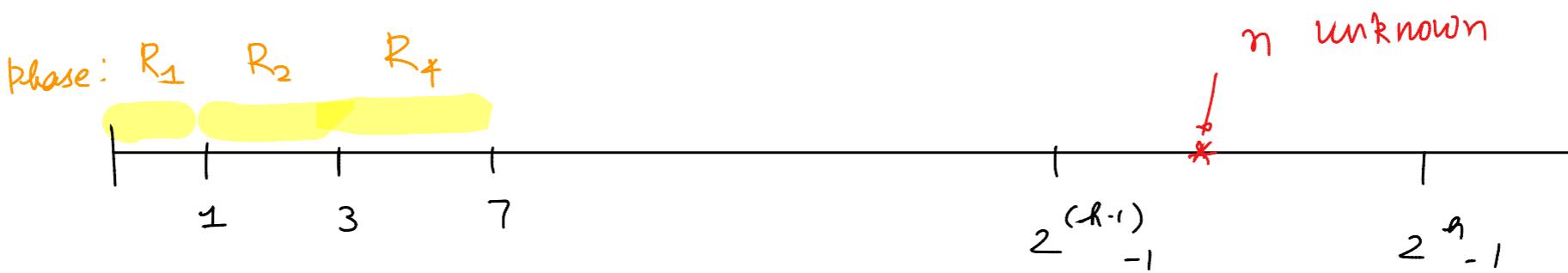
Exercise: Work out the regret bound for (*)

Refined UCB

$$UCB_i(t-1) = \hat{\mu}_i(t-1) + \sqrt{2 \frac{\ln(1+t \ln^2(t))}{T_i(t-1)}}$$

$$(\delta_t = \frac{1}{1+t \ln^2(t)})$$

Meta Approach: Doubling Trick



$$\begin{aligned} R_n &\leq R_1 + R_2 + R_4 + \dots \\ &\leq C\sqrt{2} + C\sqrt{2} + C\sqrt{4} + \dots + C\sqrt{2^k} \\ &= C\sqrt{2^k} \left(\dots + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} + 1 \right) \end{aligned}$$

$$\leq C\sqrt{2^k} \frac{1}{1 - \frac{1}{\sqrt{2}}}$$

$$= C' \sqrt{n}$$

Thompson Sampling

(Bayesian approach)

Frequentist: mean of a random variable is unknown. Needs to be estimated from samples.

Uncertainty quantifier: confidence Interval

$$U \subset B$$

Arms₁ . . . Arm_K



we have "belief" distribution

$$Q_x(t), \quad (\hat{\mu}_x(t) \sim Q_x(t)), \quad \text{we}$$

keep updating $Q_x(t)$

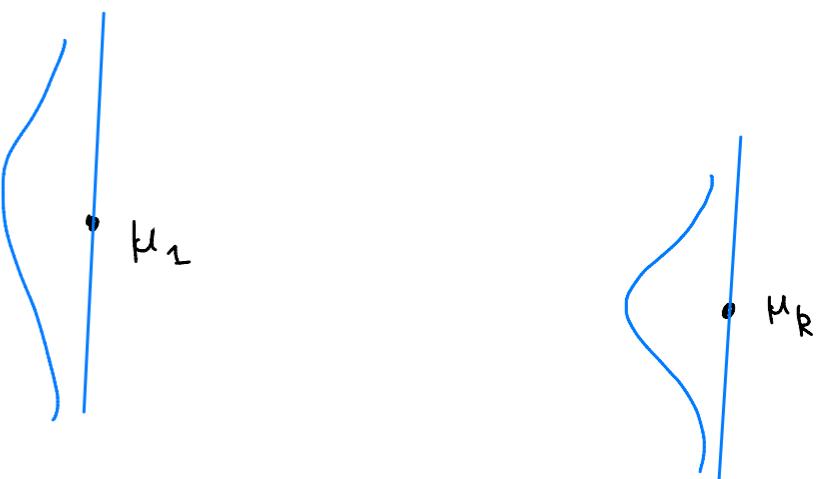
Bayesian:

- associate a "belief" with any unknown. (prior)
- based on data change the "belief" (posterior)

Spread of the belief distribution

$$+ S$$

Arms₁ . . . Arm_K



Thompson Sampling

- Sample

$$\hat{\mu}_i(t) \sim Q_i(t)$$

algorithms
mind

(Exploration happens because of uncertainty / spread in the distribution)

- Play

$$A_t = \arg \max_i \hat{\mu}_i(t)$$

- Obtain reward

$$x_t \sim P_{A_t}$$

actual environment

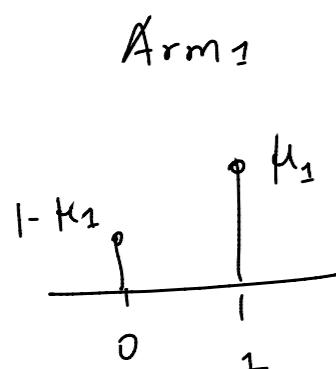
- we use x_t to update

$$Q_{A_t}^{(t)} \rightarrow Q_{A_t}^{(t+1)}$$

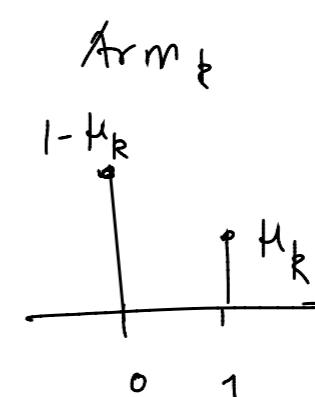
(recalibrate / update belief)

Concrete Example : Bernoulli Bandits

Environment



$$x_t \sim P_i$$



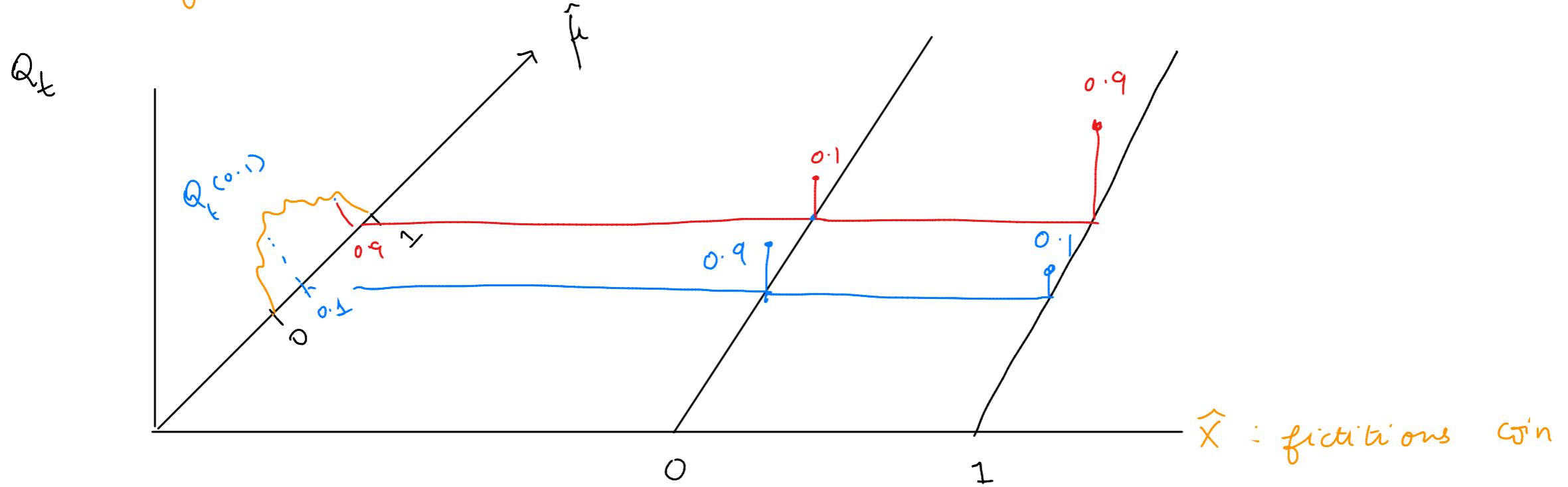
$$\text{Prob}(x_t = 0) = 1 - \mu_i$$

$$\text{Prob}(x_t = 1) = \mu_i$$

$$\mathbb{E}[x] = \mu_i$$

P_i

Belief



How to update the belief \hat{Q}_t

- Environment
 - I play arm i , then environment gives me $x_t \sim P_i$

- Algorithm thinks
 - i th fictitious coin, I tossed it and I obtained x_t

Case I: say I obtained $x_t = 0$

$$\text{Posterior} \mid \text{Data} = \frac{\text{Evidence in favour}}{\text{Total Evidence}}$$

$$Q_{\ell,t+1}(\hat{\mu}) = \frac{\text{Prob}(\hat{X}_t = 0 | \hat{\mu}) \cdot Q_{i,t}(\hat{\mu})}{\int_0^1 \text{Prob}(\hat{X}_t = 0 | \hat{\mu}) \cdot Q_{i,t}(\hat{\mu}) \cdot d\hat{\mu}}$$

$$= \frac{(1 - \hat{\mu}) \cdot Q_{i,t}(\hat{\mu})}{\int_0^1 (1 - \hat{\mu}) \cdot Q_{i,t}(\hat{\mu}) \cdot d\hat{\mu}} \quad - \text{Eq (1)}$$

case II: say I obtained $x_t = 1$

$$Q_{\ell,t+1}(\hat{\mu}) = \frac{\text{Prob}(\hat{X}_t = 1 | \hat{\mu}) \cdot Q_{i,t}(\hat{\mu})}{\int_0^1 \text{Prob}(\hat{X}_t = 1 | \hat{\mu}) \cdot Q_{i,t}(\hat{\mu}) \cdot d\hat{\mu}}$$

$$= \frac{\hat{\mu} \cdot Q_{i,t}(\hat{\mu})}{\int_0^1 \hat{\mu} \cdot Q_{i,t}(\hat{\mu}) \cdot d\hat{\mu}} \quad - \text{Eq (2)}$$

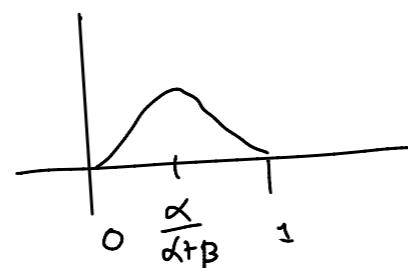
Encoding both Eq (1) & (2)

$$Q_{i,t+1}(\hat{\mu}) = \frac{(1-\hat{\mu})^{[1-x_t]} \cdot (\hat{\mu})^{[x_t]} \cdot Q_{i,t}(\hat{\mu})}{\int_0^1 (1-\hat{\mu})^{[1-x_t]} \cdot (\hat{\mu})^{[x_t]} \cdot Q_{i,t}(\hat{\mu}) \cdot d\hat{\mu}}$$

It is key that $Q_{i,t+1}$ and $Q_{i,t}$ have the same functional form

Beta Distribution

$$Q_t(\hat{\mu}) = \text{Beta}(\alpha, \beta) = \frac{(\hat{\mu})^{[\alpha-1]} \cdot (1-\hat{\mu})^{[\beta-1]}}{\int_0^1 (\hat{\mu})^{[\alpha-1]} \cdot (1-\hat{\mu})^{[\beta-1]} \cdot d\hat{\mu}}$$

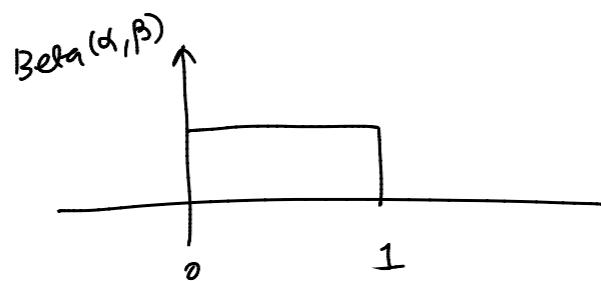


Properties of Beta Distribution

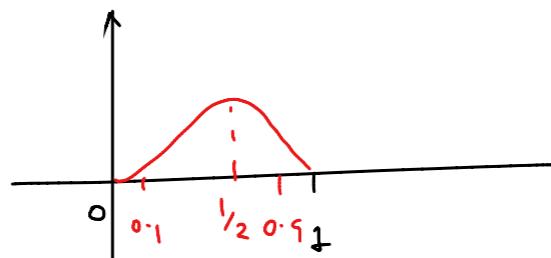
- Expectation is $\frac{\alpha}{\alpha + \beta}$

- For $\alpha = \beta = 1$

$$\text{Beta}(1, 1) = (\hat{\mu})^{[1-1]} (1-\hat{\mu})^{[1-1]} \\ = (\hat{\mu})^0 (1-\hat{\mu})^0 = 1$$



- Plugging $\alpha = \beta = \gamma$, $\text{Beta}(\gamma, \gamma) = (\hat{\mu})^{[\gamma-1]} (1-\hat{\mu})^{[\gamma-1]}$

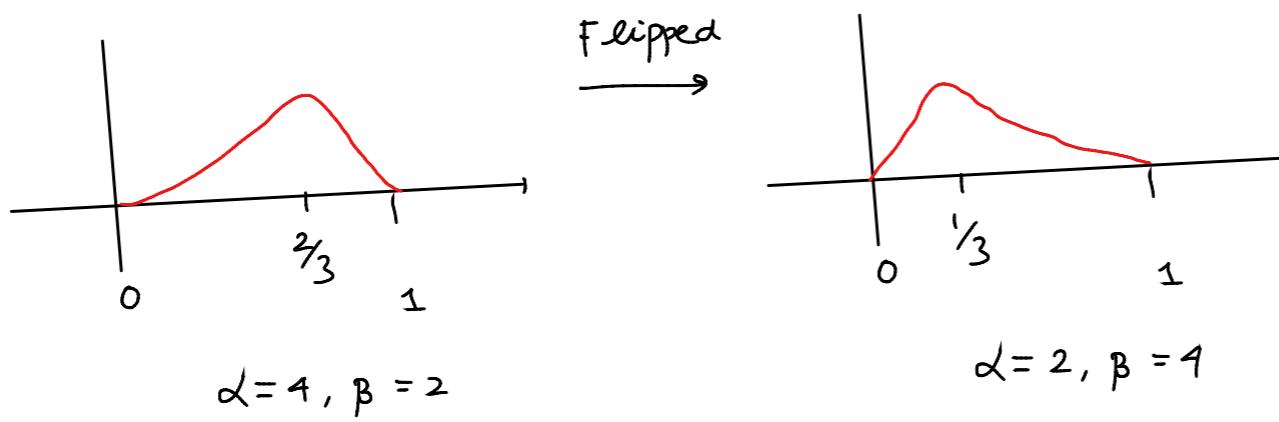


- For $\alpha = 4, \beta = 2$

@ $\hat{\mu} = 0.01^{[\alpha-1]} (0.99)^{[\beta-1]}$
 $\text{Beta}(4, 2) \propto (0.01)^3 (0.99)^1$

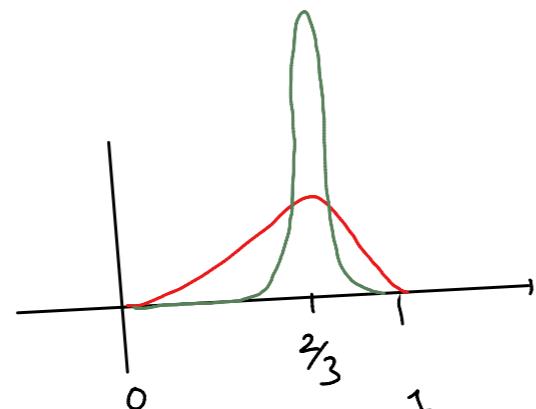
$\propto (0.01)^3 (0.99)^1 = 10^{-6} \times 0.99$

@ $\hat{\mu} = 0.99$
 $\text{Beta}(4, 2) \propto (0.99)^3 (0.01)^1$
 $\propto 0.01$



$$\frac{\alpha}{\alpha + \beta} = \frac{k\alpha}{k\alpha + k\beta}, \quad \alpha = 40, \quad \beta = 20$$

$\textcircled{c} \quad \mu = 0.01$
 $\text{Beta}(40, 20) \propto (0.01)^{39} (0.99)^{19}$



$\alpha = 40, \beta = 20$

Define $S_i(t) = \sum_{s=1}^t \mathbb{I}_{\{A_s = i\}} X_t$

(Total reward accumulated from arm i)
= total wins or 1s

$$T_i(t) = \sum_{s=1}^t \mathbb{I}_{\{A_s = i\}}$$

- For all arms $i=1, \dots, k$ start with $\alpha_{i,t}^0 = 1$, $\beta_{i,t} = 1$, $t=1$

at t , sample $\hat{\mu}_i(t) \sim Q_i(\alpha_{i,t}, \beta_{i,t})$

Play arm, $A_t = \arg \max_i \hat{\mu}_i(t)$

observe $x_t \sim P_{A_t}$

update $Q_{i,t+1} = \text{Beta} \left(\underbrace{1 + S_i(t)}_{\text{on heads}} + \underbrace{x_t}_{X_t = 1}, \underbrace{1 + T_i(t)}_{\text{on tails}} - \underbrace{S_i(t)}_{\text{success}} + \underbrace{1 - x_t}_{X_t = 0} \right)$

$$Q_{i,t+1}(\hat{\mu}) = \frac{Q_{i,t}(\hat{\mu}) \cdot (\hat{\mu})^{x_t} (1-\hat{\mu})^{1-x_t}}{\int_0^1 Q_{i,t}(\hat{\mu}) \cdot (\hat{\mu})^{x_t} (1-\hat{\mu})^{1-x_t} \cdot d\hat{\mu}}$$

$$= \begin{bmatrix} [1+s_i(t)] & [1+\tau_i(t)-s_i(t)] \\ (\hat{\mu}) & (1-\hat{\mu}) \end{bmatrix} \frac{1}{\int_0^1 (\tilde{\mu})^{[1+s_i(t)]} (1-\tilde{\mu})^{[1+\tau_i(t)-s_i(t)]} \cdot d\tilde{\mu}}$$

1

$$\begin{bmatrix} [1+s_i(t)] & [1+\tau_i(t)-s_i(t)] \\ (\hat{\mu}) & (1-\hat{\mu}) \end{bmatrix} \frac{1}{\int_0^1 (\tilde{\mu})^{[1+s_i(t)]} (1-\tilde{\mu})^{[1+\tau_i(t)-s_i(t)]} \cdot d\tilde{\mu}}$$

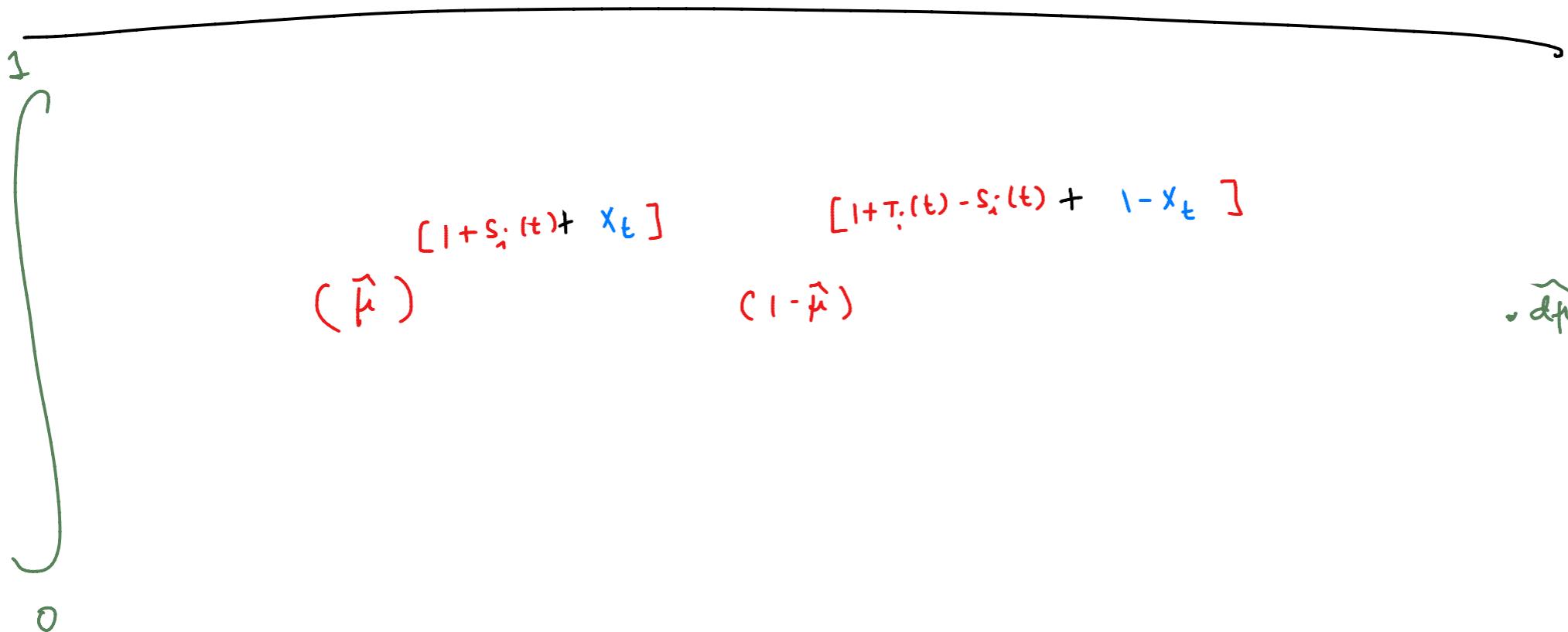
0

$$[1 + s_i(t) + x_t] \quad [1 + T_i(t) - s_i(t) + 1 - x_t]$$

($\hat{\mu}$)

($1 - \hat{\mu}$)

=



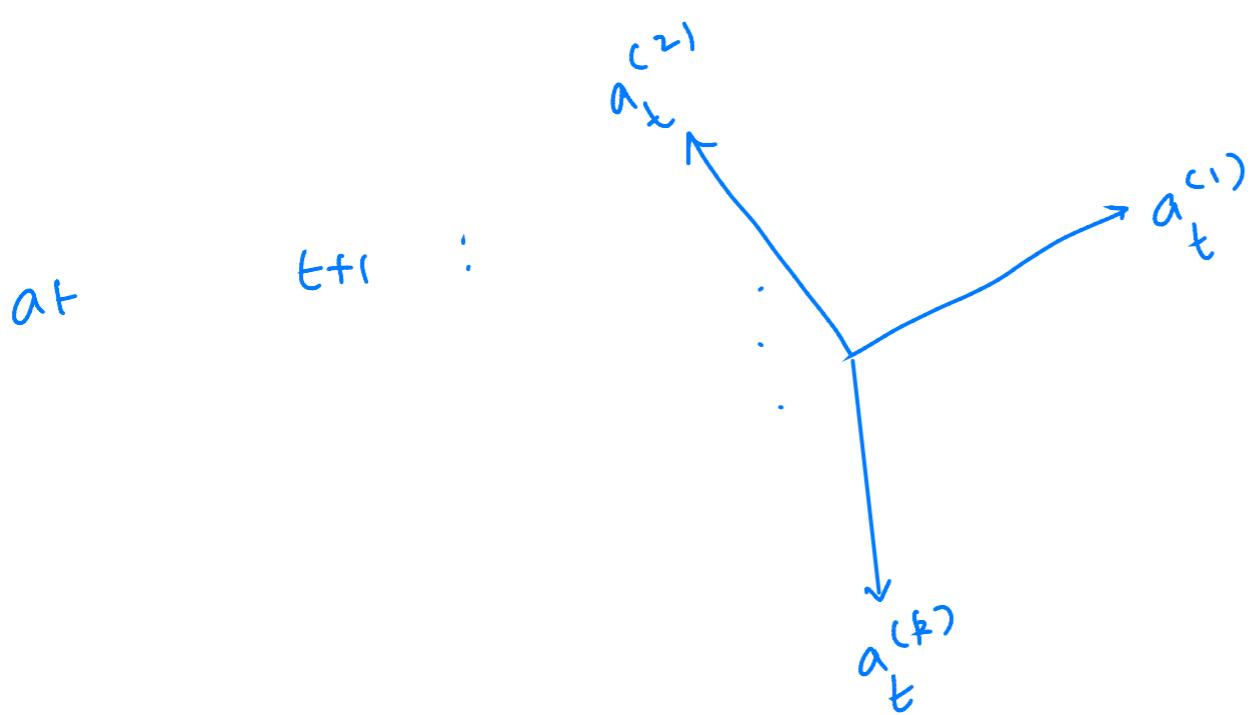
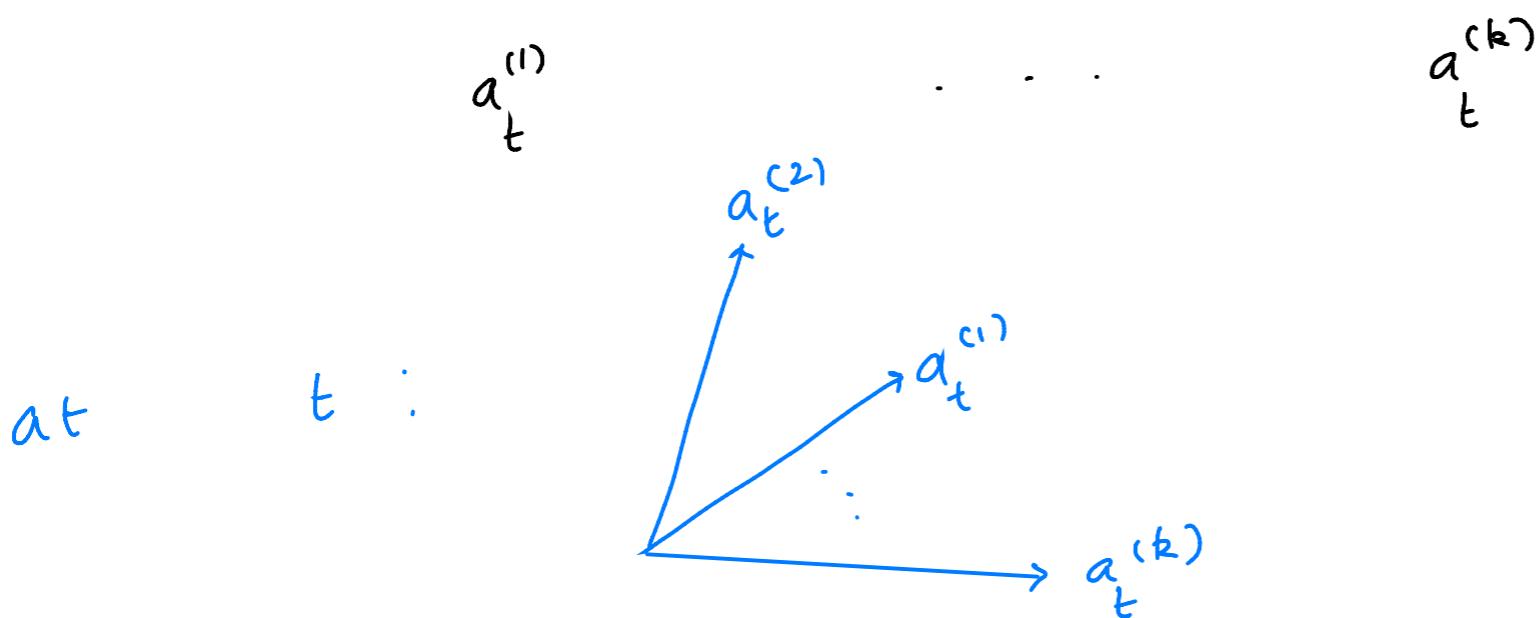
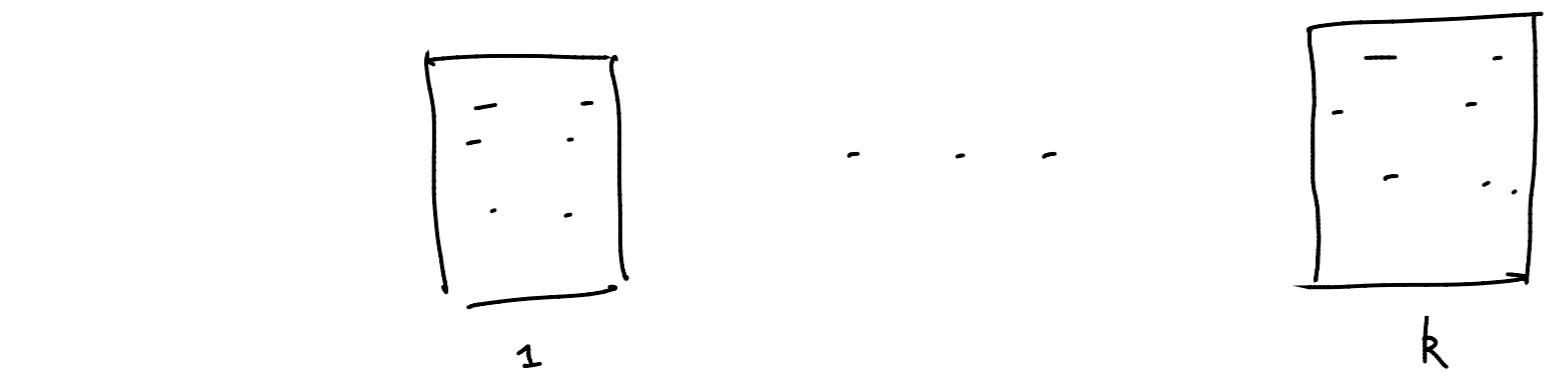
$$= \text{Beta} (1 + s_i(t) + x_t, 1 + T_i(t) - s_i(t) + 1 - x_t)$$

Linear Bandits

Motivating Scenario : Say we want to recommend news items

- $1, \dots, k$ news articles
- user visits page and clicks on a particular item.
- the preference is user dependent
 ↓
(original / vanilla bandits: $x_t \sim P_{A_t}^A$, does not hold)
 across users
 ↓
 no single best arm
- assumption
 $\psi : \text{user} \times \text{news items} \rightarrow \mathbb{R}^d$

at time t (each time we have a different user)



The arms keep changing with respect to time , either

- user changed
- new item itself changed

Assumption: $\mathcal{A}_t = \{a_t^{(1)}, \dots, a_t^{(k)}\} \in \mathbb{R}^d$

script A

at time t , play $A_t \in \mathcal{A}_t$

$$u, v \in \mathbb{R}^d$$

$$\langle u, v \rangle = u^T v$$

$$= \sum_{i=1}^d u(i)v(i)$$

$$x_t = \langle A_t, \theta_* \rangle + \eta_t$$

$\theta_* \in \mathbb{R}^d$ is an unknown parameter

$A_t \in \{a_t^{(1)}, \dots, a_t^{(k)}\}$ is the arm (feature) played vector

$\{\eta_t\}_{t \geq 0}$ are iid 1 - subGaussian random variables

Recovering Vani Ila "Independent arms" case

$$k = d, \quad A_t = \{a_t^{(1)} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad a_t^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots, \quad a_t^{(k)} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}\}$$

standard basis : e_1, e_2, \dots, e_k

On playing arm 1, reward is $\langle a_t^{(1)}, \theta_* \rangle + \eta_t$

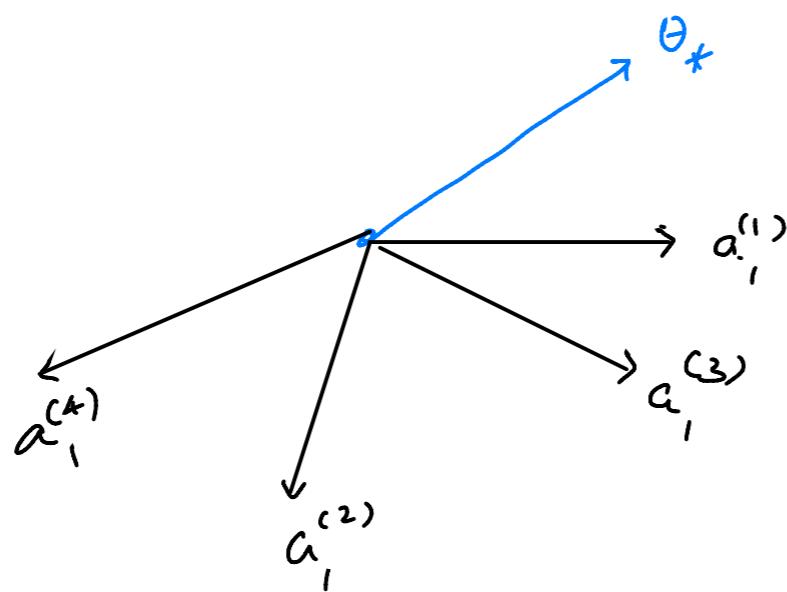
$$= \theta_{*1}^{(1)} + \eta_t$$

True parameter →

$$\hat{i}_* = \operatorname{argmax}_i \theta_{*1}^{(i)}$$

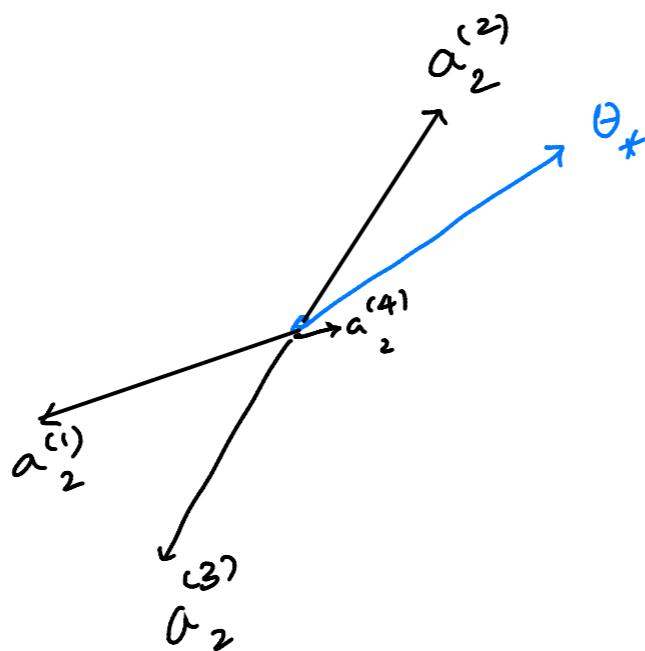
Linear Bandits , pick $d = 2$, $k = 4$

user 1
at $t = 1$



$a_1^{(1)}$ is better

user 2
at $t = 2$



$a_2^{(2)}$ is better

Fixed Design: a_1, \dots, a_t (fixed sequence) $\equiv ETC$

Say $A_t = A = \{a^{(1)}, \dots, a^{(k)}\}$

How to recover θ^* ?

Say $\gamma_t = 0, \forall t$

- it is straightforward in the vanilla case
play arm $a^{(i)} = e_i$, observe $\langle e_i, \theta^* \rangle = \theta^*(i)$
- say $k=4$ arms, $d=2$, $\theta^* = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, environment knows this

$$A = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0.5 \\ 0.75 \end{bmatrix}, \begin{bmatrix} 0.75 \\ 0.5 \end{bmatrix} \right\}$$

play arm 1 and arm 2

Arm 1 : $\left\langle \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \theta^* \right\rangle = 2\theta^*(1) + 1\theta^*(2) = 3 + \eta_1$

Arm 2 : $\left\langle \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \theta^* \right\rangle = 1\cdot\theta^*(1) + 2\cdot\theta^*(2) = 3 + \eta_2$

Arm 3 : $\left\langle \begin{bmatrix} 0.5 \\ 0.75 \end{bmatrix}, \theta^* \right\rangle = 0.5\theta^*(1) + 0.75\theta^*(2) = 1.25 + \eta_3$

If we have noise these equations will not be consistent
 ↓
 Set up a linear regression problem to recover θ^*

$$2\theta_x(1) + 1\theta_x(2) = 3 + \eta_1$$

$$1\cdot\theta_x(1) + 2\cdot\theta_x(2) = 3 + \eta_2$$

$$\equiv \begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 0.5 & 0.75 \end{bmatrix} \begin{bmatrix} \theta_x(1) \\ \theta_x(2) \end{bmatrix} = \begin{bmatrix} 3 + \eta_1 \\ 3 + \eta_2 \\ 1.25 + \eta_3 \end{bmatrix}$$

$$0.5\theta_x(1) + 0.75\theta_x(2) = 1.25 + \eta_3$$

$$\eta_1 = -0.1, \eta_2 = 0.2, \eta_3 = -0.25$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \\ 0.5 & 0.75 \end{bmatrix} \begin{bmatrix} \theta(1) \\ \theta(2) \end{bmatrix} = \begin{bmatrix} 2.9 \\ 3.2 \\ 1 \end{bmatrix}$$

fixed Design : $A_E : A = \{a^{(1)}, \dots, a^{(k)}\} \in \mathbb{R}^{d^2}$

a_1, \dots, a_t have been played already

- x_1, \dots, x_t have been observed $x_t = \langle a_t, \theta_t \rangle + \eta_t$
 - $V_t = \sum_{s=1}^t a_s a_s^\top$ is invertible
- Goal : Find $\widehat{\theta}$ = $\arg \min_{\theta} L(\theta) = \arg \min_{\theta} \sum_{s=1}^t (\langle a_s, \theta \rangle - x_s)^2$
- $\widehat{\theta}$ estimate

at $\widehat{\theta}$

$$\frac{\partial L(\theta)}{\partial \theta^{(1)}} \Big|_{\theta=\widehat{\theta}} = 0$$

$$\frac{\partial L(\theta)}{\partial \theta^{(2)}} \Big|_{\theta=\widehat{\theta}} = 0$$

.

.

$$\frac{\partial L(\theta)}{\partial \theta^{(d)}} \Big|_{\theta=\widehat{\theta}} = 0$$

}

d . equations

$$\text{Note: } \langle a_s, \theta \rangle = a_{s(1)} \theta(1) + a_{s(2)} \theta(2) + \dots + a_{s(d)} \theta(d)$$

$$\frac{\partial L(\theta)}{\partial \theta(1)} = \frac{\partial}{\partial \theta(1)} \left(\sum_{s=1}^t (\langle a_s, \theta \rangle - x_s)^2 \right) = \sum_{s=1}^t 2 (\langle a_s, \theta \rangle - x_s) a_{s(1)} = 0$$

$$\frac{\partial L(\theta)}{\partial \theta(2)} = \frac{\partial}{\partial \theta(2)} \left(\sum_{s=1}^t (\langle a_s, \theta \rangle - x_s)^2 \right) = \sum_{s=1}^t 2 (\langle a_s, \theta \rangle - x_s) a_{s(2)} = 0$$

$$\vdots$$

$$\vdots$$

$$\frac{\partial L(\theta)}{\partial \theta(d)} = \frac{\partial}{\partial \theta(d)} \left(\sum_{s=1}^t (\langle a_s, \theta \rangle - x_s)^2 \right) = \sum_{s=1}^t 2 (\langle a_s, \theta \rangle - x_s) a_{s(d)} = 0$$

$$\begin{bmatrix} - & a_1^\top & - \\ \vdots & \vdots & \\ - & a_t^\top & - \end{bmatrix} \begin{bmatrix} \theta(1) \\ \vdots \\ \theta(d) \end{bmatrix} - \begin{bmatrix} x_1 \\ \vdots \\ x_t \end{bmatrix} = \begin{bmatrix} \langle a_1, \theta \rangle - x_1 \\ \vdots \\ \langle a_t, \theta \rangle - x_t \end{bmatrix}$$

$$\begin{bmatrix} a_1^{(1)} & a_2^{(1)} \dots a_s^{(1)} \dots a_t^{(1)} \\ a_1^{(2)} & a_2^{(2)} \dots a_s^{(2)} \dots a_t^{(2)} \\ \vdots & \vdots \\ a_1^{(d)} & a_2^{(d)} \dots a_s^{(d)} \dots a_t^{(d)} \end{bmatrix} = \begin{bmatrix} \langle a_1, \theta \rangle - x_1 \\ \vdots \\ \langle a_s, \theta \rangle - x_s \\ \vdots \\ \langle a_t, \theta \rangle - x_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

} d-equations

$$\begin{bmatrix} a_1^{(1)} & a_2^{(1)} \dots a_s^{(1)} \dots a_t^{(1)} \\ a_1^{(2)} & a_2^{(2)} \dots a_s^{(2)} \dots a_t^{(2)} \\ \vdots & \vdots \\ a_1^{(d)} & a_2^{(d)} \dots a_s^{(d)} \dots a_t^{(d)} \end{bmatrix} - \begin{bmatrix} a_1^+ \\ \vdots \\ a_t^+ \end{bmatrix} - \begin{bmatrix} \theta^{(1)} \\ \vdots \\ \theta^{(d)} \end{bmatrix} = \begin{bmatrix} x_1 \\ \vdots \\ x_t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$\nabla_{\theta} \hat{\theta} - \underbrace{\sum_{s=1}^t a_s x_s}_{d \times 1} = 0$

$$Y_t = \begin{bmatrix} a_1^{(1)} & a_2^{(1)} & \dots & a_s^{(1)} & \dots & a_t^{(1)} \\ a_1^{(2)} & a_2^{(2)} & & a_s^{(2)} & \dots & a_t^{(2)} \\ \vdots & & & & & \\ a_1^{(d)} & a_2^{(d)} & & a_s^{(d)} & \dots & a_t^{(d)} \end{bmatrix}_{d \times t} - \begin{bmatrix} a_1^+ \\ \vdots \\ a_t^+ \end{bmatrix}_{t \times d}$$

dxd matrix

$$= \begin{bmatrix} a_1^+ \\ a_2^+ \\ \vdots \\ a_t^+ \end{bmatrix}^T \begin{bmatrix} -a_1^+ \\ \vdots \\ -a_t^+ \end{bmatrix} = \sum_{s=1}^t a_s a_s^T$$

$$\begin{bmatrix} a_1^{(1)} & a_2^{(1)} & \dots & a_s^{(1)} & \dots & a_t^{(1)} \\ a_1^{(2)} & a_2^{(2)} & & a_s^{(2)} & \dots & a_t^{(2)} \\ \vdots & & & & & \\ a_1^{(d)} & a_2^{(d)} & & a_s^{(d)} & \dots & a_t^{(d)} \end{bmatrix}_{d \times t} \begin{bmatrix} x_1 \\ \vdots \\ x_t \end{bmatrix}_{t \times 1} = \sum_{s=1}^t a_s x_s \in \mathbb{R}^d$$

$$y_t \hat{\theta} = \sum_{s=1}^t a_s x_s \quad \left. \begin{array}{l} \text{d-equations in} \\ \text{d-variables} \end{array} \right\}$$

$$\hat{\theta} = V_t^{-1} \left(\sum_{s=1}^t a_s x_s \right)$$

→ Expression to calculate the time parameter

Let us calculate $\theta_* - \hat{\theta}$

$$\hat{\theta} = V_t^{-1} \left[\sum_{s=1}^t a_s \left(\langle a_s, \theta_* \rangle + \eta_s \right) \right]$$

x_s

$$= V_t^{-1} \left[\sum_{s=1}^t a_s \left(\underbrace{a_s^\top \theta_*}_{\substack{\text{number} \\ \text{a-dim vector}}} + \eta_s \right) \right]$$

$$= V_t^{-1} \left[\sum_{s=1}^t \underbrace{a_s a_s^\top}_{\substack{\text{dxd matrix} \\ \text{dxd matrix}}} \theta_* + \sum_{s=1}^t \underbrace{a_s \eta_s}_{\substack{\text{d-dim vector} \\ \text{number}}} \right]$$

$$= V_t^{-1} \left[\left(\sum_{s=1}^t a_s a_s^\top \right) \theta_* + \sum_{s=1}^t a_s \eta_s \right]$$

$\underbrace{\text{dxd matrix}}_{\text{dxd matrix}} = Y_t$

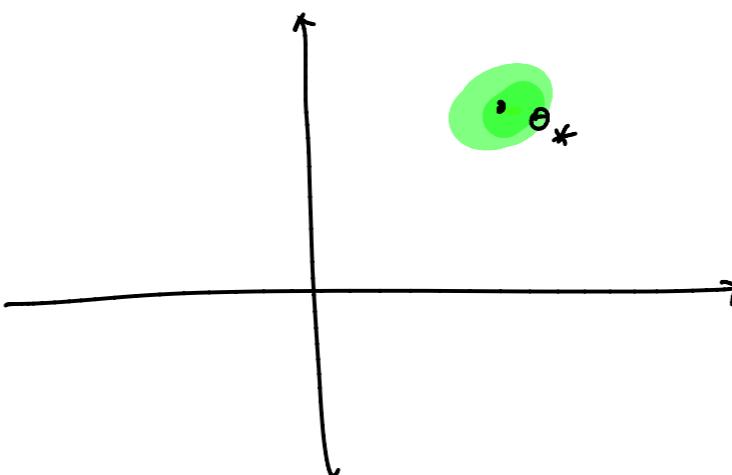
$$= \mathbf{V}_t^{-1} [\mathbf{V}_t \theta_* + \sum_{s=1}^t a_s \eta_s]$$

$$= \theta_* + \mathbf{V}_t^{-1} \sum_{s=1}^t a_s \eta_s$$

Expression for noise \rightarrow

$$\hat{\theta} - \theta_* = \mathbf{V}_t^{-1} \sum_{s=1}^t a_s \eta_s$$

$a \cdot \text{dim}$



How to make sense of $\mathbf{V}_t^{-1} \sum_{s=1}^t a_s \eta_s$

Simplifying our problem, pick $A = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$

$$a^{(1)} a^{(1)T} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$a^{(2)} a^{(2)T} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

$$a_1 = a^{(1)}, a_2 = a^{(2)}, a_3 = a^{(1)}, a_4 = a^{(1)}, a_5 = a^{(2)}, a_6 = a^{(1)}, a_7 = a^{(2)}, a_8 = a^{(2)}$$

\rightarrow Bottom line: each arm 4 times

$$V_t = \sum_{s=1}^t a_s a_s^\top = 4 a^{(1)} a^{(1)\top} + 4 a^{(2)} a^{(2)\top}$$

$$= \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \rightarrow \text{Counting number of times each arm was played was modified into effective number of times each of the principal directions gets played}$$

$$\sum_{s=1}^t a_s \eta_s = \begin{bmatrix} \eta_1 + \eta_3 + \eta_4 + \eta_6 \\ \eta_2 + \eta_5 + \eta_7 + \eta_8 \end{bmatrix} \rightarrow \text{noise affects each arm separately}$$

} principal direction gets played

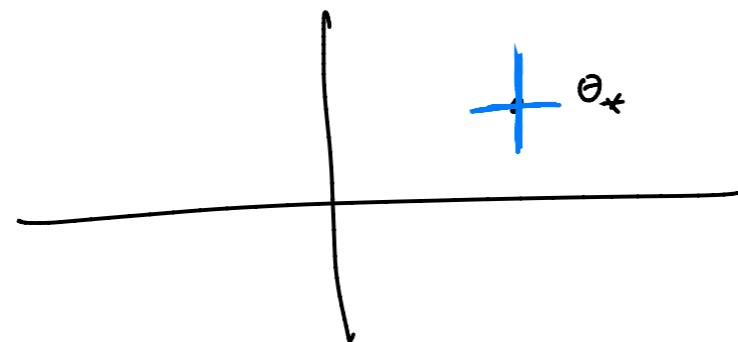
every coordinate will see all noise η_s

$$\hat{\theta} - \theta_* = V_t^{-1} \sum_{s=1}^t a_s \eta_s = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/4 \end{bmatrix} \begin{bmatrix} \eta_1 + \eta_3 + \eta_4 + \eta_6 \\ \eta_2 + \eta_5 + \eta_7 + \eta_8 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{\eta_1 + \eta_3 + \eta_4 + \eta_6}{4} \\ \frac{\eta_2 + \eta_5 + \eta_7 + \eta_8}{4} \end{bmatrix} \rightarrow \frac{\sqrt{4}}{4}$$

$$\hat{\theta} = Y_t^{-1} \left(\sum_{s=1}^t a_s x_s \right) = \begin{bmatrix} 1/4 & 0 \\ 0 & 1/4 \end{bmatrix} \begin{bmatrix} x_1 + x_3 + x_4 + x_6 \\ x_2 + x_5 + x_7 + x_8 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{x_1 + x_3 + x_4 + x_6}{4} \\ \frac{x_2 + x_5 + x_7 + x_8}{4} \end{bmatrix}$$



Consider a slightly case $A = \left\{ \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \right\}$

$a^{(1)}$ $a^{(2)}$

$$a_1 = a_2 = a_3 = a_4 = a^{(1)} . \quad a_5 = a_6 = a_7 = a_8 = a^{(2)}$$

$$\hat{\theta} = \gamma_t^{-1} \sum_{s=1}^t \alpha_s x_s$$

$$\hat{\theta} \cdot \theta_* = V_t^{-1} \sum_{s=1}^t \alpha_s \eta_s$$

2x2 2-dim vector
 mat. matrix

$$V_t = \sum_{s=1}^t \alpha_s \alpha_s^T = 4 [a^{(1)}] [a^{(1)}]^T + 4 [a^{(2)}] [a^{(2)}]^T$$

$$= 4 \frac{1}{\sqrt{5}} \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \end{bmatrix} + 4 \cdot \frac{1}{\sqrt{5}} \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \end{bmatrix}$$

$$= \frac{1}{5} \left(4 \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} + 4 \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} \right)$$

$$= \frac{1}{5} \begin{bmatrix} 20 & 16 \\ 16 & 20 \end{bmatrix} \rightarrow \text{Symmetric Positive Definite Matrix}$$

$$\Sigma^{\text{asy}} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} (\eta_1 + \eta_2 + \eta_3 + \eta_4) + \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix} (\eta_5 + \eta_6 + \eta_7 + \eta_8)$$

Real Symmetric Matrix : $M_{n \times n}$

Real symmetric + Positive Definite

$$M_x = V \Sigma V^T x,$$

$$\Sigma = \begin{bmatrix} \Sigma_1 & & 0 \\ 0 & \Sigma_2 & & \\ & & \ddots & \Sigma_n \end{bmatrix}$$

is $n \times n$ diagonal matrix
 $\Sigma_i > 0$

V is a unitary matrix

$V = [u_1 \ u_2 \ \dots \ u_n]$, u_i is the eigenvector with eigen value Σ_i

$$V^T V = I$$

u_i are orthonormal, they form a basis
 $\langle u_i, u_j \rangle = 0, \forall i \neq j$

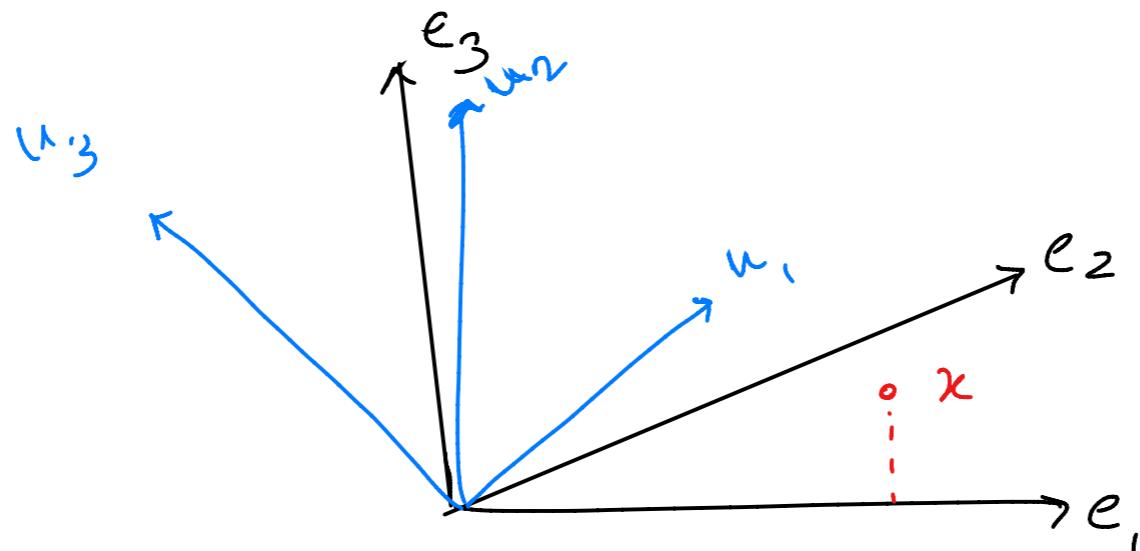
$$I = \begin{bmatrix} 1 & 0 & \cdots \\ 0 & 1 & \cdots \\ \vdots & \vdots & \ddots \\ 0 & 0 & \cdots \end{bmatrix}$$

$e_1 \ e_2 \ \dots$

e_i are orthonormal

$$\langle e_i, e_j \rangle = 0 \quad \forall i \neq j$$

$$M = U M^T$$



Proj of x on e_i

$$\langle x, e_i \rangle$$

$$x = \sum_{i=1}^n \langle x, e_i \rangle e_i$$

$$= \sum_{i=1}^n (e_i^T x) e_i$$

$$x = \sum_{i=1}^n \langle x, u_i \rangle u_i$$

$$= \sum_{i=1}^n (u_i^T x) u_i$$

$\underbrace{U \Sigma U^T}_{\text{recover}} \underbrace{x}_{\text{measure}}$
 measure x is projection
 SCNE
 the coordinates different by

$$U \begin{bmatrix} S & 0 \\ 0 & S \end{bmatrix} U^T x = U S I U^T x \\ = S x$$

Symmetric Positive Definite matrix : M

$$\underbrace{x^T M x}_{\text{distance measure}} \geq 0, \quad \forall x \neq 0$$

distance measure

$$x^T x = [x^{(1)} \ x^{(2)} \ \dots \ x^{(n)}] \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{bmatrix} = x^{(1)2} + x^{(2)2} + \dots + x^{(n)2}$$

$$x^T \begin{bmatrix} D_1 & & & \\ & D_2 & & 0 \\ & & \ddots & \\ 0 & & & D_n \end{bmatrix} x = D_1 x^{(1)2} + D_2 x^{(2)2} + \dots + D_n x^{(n)2}$$

$$D_i > 0$$

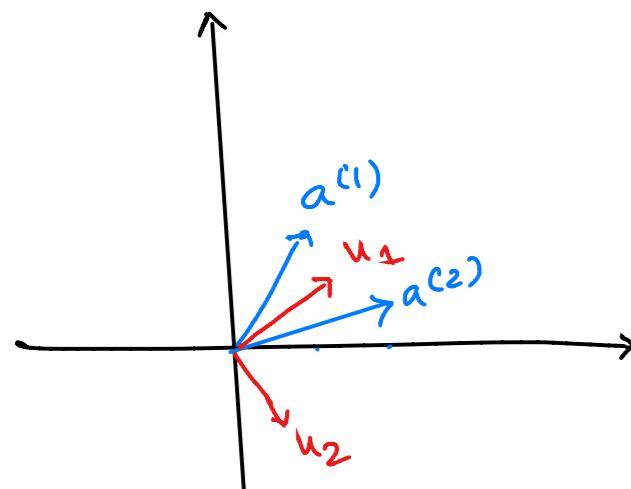
$$x^T M x = \underbrace{x^T}_{y^T} U \Sigma \underbrace{U^T x}_{y}$$

$$= y^T \Sigma y$$

$$\|x\|_M^2 = x^T M x$$

$$V_t = \frac{1}{\sqrt{5}} \begin{bmatrix} 2.0 & 1.0 \\ 1.0 & 2.0 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 36 & 0 \\ 0 & 4 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$$

U_1 U_2



$$\hat{\theta} \cdot \theta_k = V_t^{-1} \sum_{s=1}^t a_s \eta_s$$

2x2
mat. matrix 2-dim vector

$$\sum a_s \eta_s = \begin{bmatrix} 1 \\ 2 \end{bmatrix} (\eta_1 + \eta_2 + \eta_3 + \eta_4) + \begin{bmatrix} 2 \\ 1 \end{bmatrix} (\eta_5 + \eta_6 + \eta_7 + \eta_8)$$

$$\frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = (\langle u_1, a^{(1)} \rangle \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + \langle u_2, a^{(1)} \rangle \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix})$$

$$u_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \quad u_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \quad a^{(1)} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

$$\frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 3 \\ -1 \end{bmatrix} + \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{5}} (-1) \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$= \frac{3}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} - \frac{1}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \frac{3}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + \frac{1}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\sum a_s \eta_s = \frac{1}{\sqrt{5}} \left(\begin{bmatrix} 1 \\ 2 \end{bmatrix} (\eta_1 + \eta_2 + \eta_3 + \eta_4) + \begin{bmatrix} 2 \\ 1 \end{bmatrix} (\eta_5 + \eta_6 + \eta_7 + \eta_8) \right)$$

$$= \left(\frac{3}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} - \frac{1}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right) (\eta_1 + \eta_2 + \eta_3 + \eta_4)$$

$$+ \left(\frac{3}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} + \frac{1}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right) (\eta_5 + \eta_6 + \eta_7 + \eta_8)$$

$$= \frac{3}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} (\gamma_1 + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5 + \gamma_6 + \gamma_7 + \gamma_8) \rightarrow \xi_1$$

$$\frac{1}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} (-\gamma_1 - \gamma_2 - \gamma_3 - \gamma_4 + \gamma_5 + \gamma_6 - \gamma_7 - \gamma_8) \rightarrow \xi_2$$

$$V_t^{-1} = 5 V \begin{bmatrix} \frac{1}{36} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} V^+$$

$$V_t^{-1} (\xi_1 + \xi_2) = \underbrace{V_t^{-1} \xi_1}_{+} + V_t^{-1} \xi_2$$

$$5 \quad U \begin{bmatrix} \frac{1}{36} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} U^+ \frac{3}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} (\eta_1 + \eta_2 + \eta_3 + \eta_4 + \eta_5 + \eta_6 + \eta_7 + \eta_8)$$

$\underbrace{\hspace{1cm}}_{u_1}$

$$= \frac{5 \times 3}{\sqrt{10}} \quad U \begin{bmatrix} \frac{1}{36} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 \\ 3 \end{bmatrix} (\eta_1 + \dots + \eta_8)$$

$$= \frac{5 \times 3}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \frac{(\eta_1 + \dots + \eta_8)}{36} + \frac{5}{\sqrt{10}} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -1 \\ \frac{1}{\sqrt{2}} \end{bmatrix} \frac{(-\eta_1 + \dots + \eta_5 + \dots + \eta_8)}{4}$$

$\underbrace{\hspace{1cm}}_{\text{Term } 1}$ $\underbrace{\hspace{1cm}}_{\text{Term } 2}$

Sub Gaussianity factor calculation

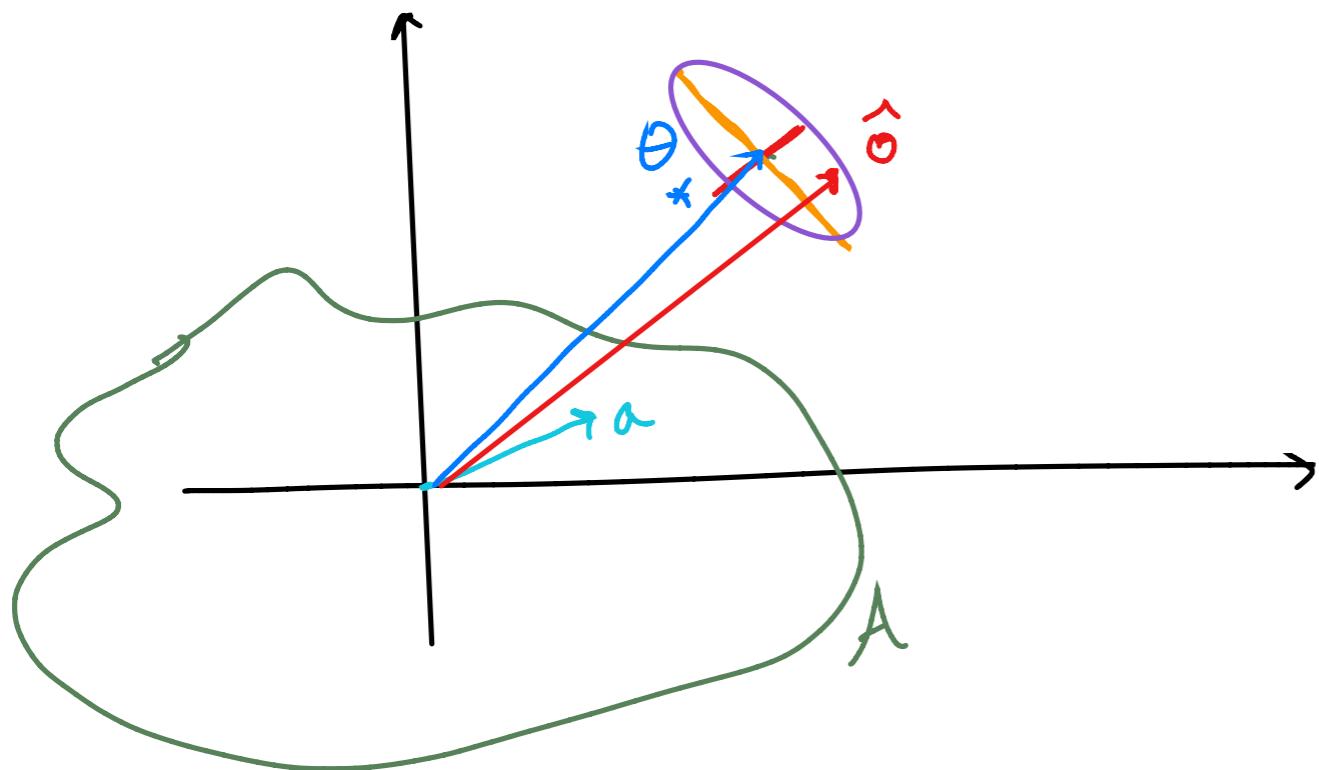
$$\frac{\sqrt{4}}{4} = \frac{1}{2}$$

$$\frac{5 \times 3}{\sqrt{10}} \times \frac{\sqrt{8}}{36} = \sqrt{\frac{4}{5}} \times \frac{5}{12}$$

$$\frac{5}{\sqrt{10}} \times \frac{\sqrt{8}}{4} = \sqrt{\frac{8}{10}} \times \frac{5}{4}$$

$$= \sqrt{\frac{5}{4}}$$

independent arms case



40 samples each:

$$\frac{5 \times 3}{\sqrt{10}} \cdot \frac{\sqrt{80}}{360} = \sqrt{\frac{4}{5}} \times \frac{5}{12} \cdot \frac{1}{\sqrt{10}}$$

$$\frac{5}{\sqrt{10}} \times \frac{\sqrt{80}}{40} = \sqrt{\frac{5}{4}} \cdot \frac{1}{\sqrt{10}}$$

Say I play an arm $a \in \mathbb{R}^d$

$$\text{Prob} (\langle a, \hat{\theta} \rangle - \langle a, \theta_* \rangle > \varepsilon)$$

In the vanilla case $a = e_i$ for some $i \in \{1, \dots, k\}$

$$\text{Prob} (\underbrace{\hat{\theta}(i)}_{\text{Sample - Time Mean}} - \underbrace{\theta_*(i)}_{\text{Mean}} > \varepsilon)$$

$$\text{Prob} (\langle a, \hat{\theta} - \theta_* \rangle \geq \varepsilon)$$

$$= \text{Prob} (\langle a, V_t^{-1} \sum_{s=1}^t a_s \eta_s \rangle \geq \varepsilon)$$

under the assumption that $\{\eta_s\}_{s \geq 0}$ are iid 1-subgaussian

we can calculate the sub-Gaussianity of

$$\begin{aligned} \langle a, V_t^{-1} \sum_{s=1}^t a_s \eta_s \rangle &= \sum_{s=1}^t \langle a, V_t^{-1} a_s \eta_s \rangle = \sum_{s=1}^t \underbrace{\langle a, V_t^{-1} a_s \rangle}_{\alpha_s} \eta_s \\ &= \sum_{s=1}^t \alpha_s \eta_s \end{aligned}$$

Sub-Gaussianity factor is $\sqrt{\alpha_1^2 + \dots + \alpha_t^2}$

$$\alpha_s = \langle a, V_t^{-1} a_s \rangle \Rightarrow \alpha_s^2 = (\langle a, V_t^{-1} a_s \rangle)^2$$

$$\alpha_s^2 = (a^\top V_t^{-1} a_s)^2 = (a^\top V_t^{-1} a_s) (a^\top V_t^{-1} a_s)^T$$

$$= a^\top V_t^{-1} a_s a_s^\top V_t^{-1} a$$

$$\sum_{s=1}^t \alpha_s^2 = \sum_{s=1}^t a^\top V_t^{-1} a_s a_s^\top V_t^{-1} a = a^\top V_t^{-1} \left(\sum_{s=1}^t a_s a_s^\top \right) V_t^{-1} a$$

$$= a^\top V_t^{-1} V_t V_t^{-1} a$$

$$= a^\top V_t^{-1} a$$

$$= \|a\|_V^2 \quad = a^\top U \Sigma^{-1} U^\top a$$

$$M = U \Sigma U^\top$$

$$M^{-1} = U \Sigma^{-1} U^\top$$

$$MM^{-1} = U \underbrace{\Sigma}_{I} U^\top \Sigma^{-1} U^\top = I$$

For σ -SubGaussian

$$\text{Prob}(\hat{\mu} - \mu > \sqrt{2\sigma^2 \ln(1/\delta)}) \leq \delta$$

↓

$$\text{Prob}(\langle a, \hat{\theta} - \theta_* \rangle > \sqrt{2 \|a\|_{V_t^{-1}}^2 \ln(1/\delta)}) \leq \delta$$

- error depends on length of $a \in \mathbb{R}^d$

- specifically length measured w.r.t V_t^{-1} i.e. $\|a\|_{V_t^{-1}}^2$

Say $|A_t|$ is finite, i.e. $A_t = \{a^{(1)}, \dots, a^{(k)}\}$

$$\text{Prob}(\langle a^{(1)}, \hat{\theta} - \theta_* \rangle > \sqrt{2 \|a\|_{V_t^{-1}}^2 \ln(1/\delta)}) \leq \delta$$

⋮

$$\text{Prob}(\langle a^{(k)}, \hat{\theta} - \theta_* \rangle > \sqrt{2 \|a\|_{V_t^{-1}}^2 \ln(1/\delta)}) \leq \delta \Rightarrow \text{Total } k\delta$$

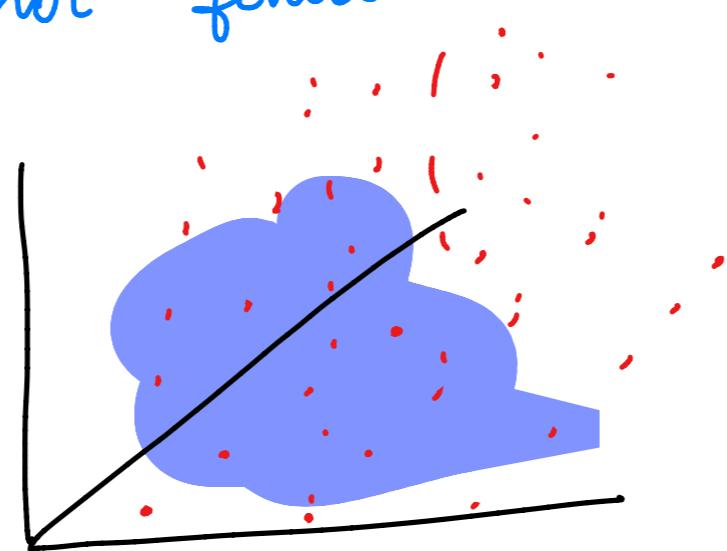
Union Bound

Now if we need total to be δ then

Pick any $a \in A_t = \{a^{(1)}, \dots, a^{(k)}\}$

$$\text{Prob} (| \langle a, \hat{\theta} - \theta_* \rangle | \geq \sqrt{2 \|a\|_{V_t^{-1}}^2 \ln(\frac{|A_t|}{\delta})}) \leq \delta$$

Suppose A_t is not finite



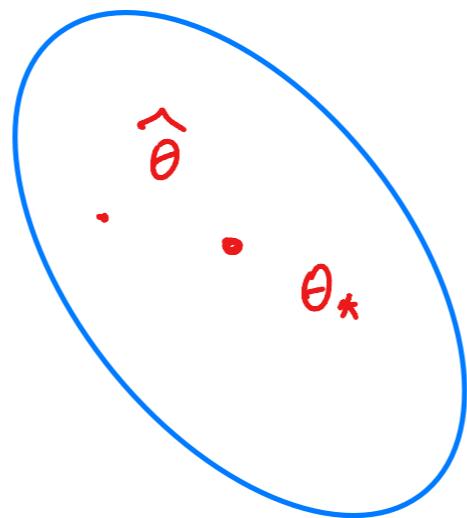
6 grid point
per dimension

after gridding $|A_t^G| \leq 6^d$

$$\Pr_{\text{rob}} \left(\langle a, \hat{\theta} - \theta_* \rangle \geq \sqrt{2 \|a\|^2_{V_t^{-1}} \ln\left(\frac{G^d}{\delta}\right)} \right) \leq \delta$$

$$\Pr_{\text{rob}} \left(\langle a, \hat{\theta} - \theta_* \rangle \geq \sqrt{d^2 \|a\|^2_{V_t^{-1}} \ln\left(\frac{G}{\delta}\right)} \right) \leq \delta$$

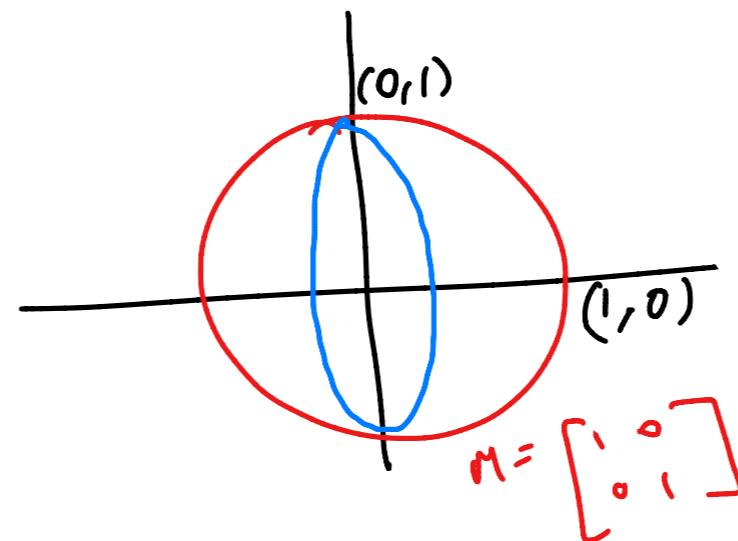
Alternative way of constructing error bound



The natural confidence set to look at is

$$B_\varepsilon = \left\{ \theta : \|\hat{\theta}_t - \theta\|_{V_t^{-1}}^2 \leq \varepsilon \right\}$$

$$B_1 = \{ \theta : \| \theta \|_M^2 \leq 1 \} \quad M = \begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\| \hat{\theta}_t - \theta_* \|_{v_t}^2 = (\hat{\theta}_t - \theta_*)^\top v_t (\hat{\theta}_t - \theta_*) \\ = \langle v_t (\hat{\theta}_t - \theta_*), \hat{\theta}_t - \theta_* \rangle$$

$$\| \hat{\theta}_t - \theta_* \|_{v_t} = \left\langle v_t \frac{(\hat{\theta}_t - \theta_*)}{\| \hat{\theta}_t - \theta_* \|_{v_t}}, \hat{\theta}_t - \theta_* \right\rangle$$

$$= \underbrace{\left\langle \sqrt{v_t} \frac{y_t^{1/2} (\hat{\theta}_t - \theta_*)}{\|\hat{\theta}_t - \theta_*\|_{V_t}}, \right.}_{\text{deterministic}} \left. \hat{\theta}_t - \theta_* \right\rangle$$

a : random

random
 $\hat{\theta}_t - \theta_$*

random variable

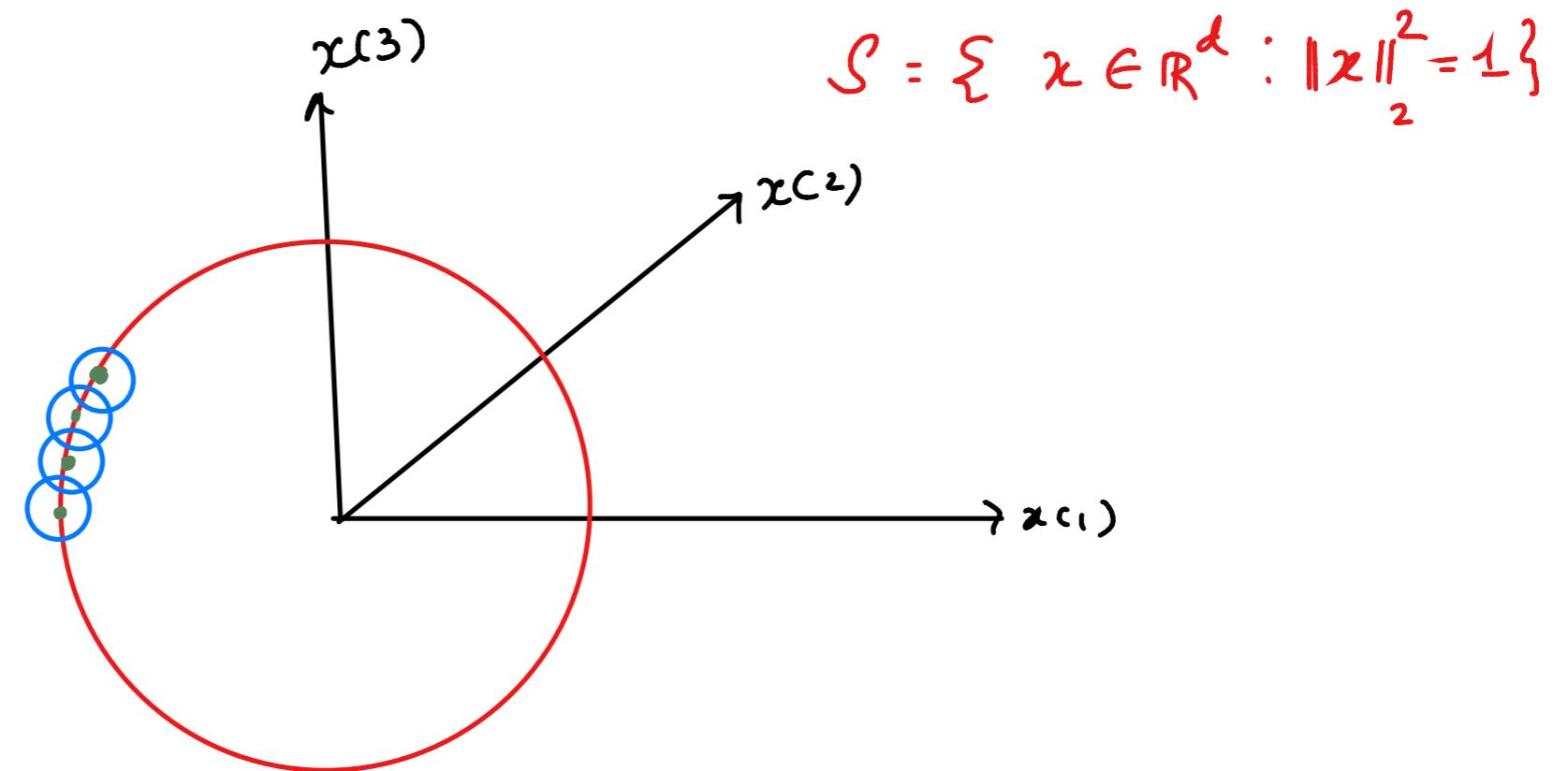
z : random variable

$$z^T z = \begin{pmatrix} y_t^{1/2} (\hat{\theta}_t - \theta_*) \\ \|\hat{\theta}_t - \theta_*\|_{V_t} \end{pmatrix}^T \begin{pmatrix} y_t^{1/2} (\hat{\theta}_t - \theta_*) \\ \|\hat{\theta}_t - \theta_*\|_{V_t} \end{pmatrix}$$

$$= \frac{(\hat{\theta}_t - \theta_*)^T v_t (\hat{\theta}_t - \theta_*)}{\|\hat{\theta}_t - \theta_*\|_{V_t}^2} = \frac{\|\hat{\theta}_t - \theta_*\|_{V_t}^2}{\|\hat{\theta}_t - \theta_*\|_{V_t}^2} = 1$$

$$\Pr_{\text{rob}} \left(\langle a, \hat{\theta} - \theta_* \rangle \geq \sqrt{2 \|a\|^2 \frac{\ln(\frac{|\mathcal{A}_t|}{\delta})}{V_t^{-1}}} \right) \leq \delta$$

↓ ↓
 fixed random



Cover the unit sphere by ϵ balls

red circle : unit sphere

blue circles : epsilon balls

green centers : centers of these balls

Let C_ϵ be the set of these centers

Fact : the number of centers $|C_\epsilon| \leq \left(\frac{3}{\epsilon}\right)^d$
(Theorem)

E is a low probability event. there exists a

$y \in C_\epsilon$ which violates the bound

$$E = \{ \text{exists } y: C_\epsilon : \langle V_t^{1/2} y, \hat{\theta}_t - \theta_* \rangle \geq \sqrt{2 \|V_t^{1/2} y\|_{V_t^{-1}}^2 \ln\left(\frac{|C_\epsilon|}{8}\right)} \}$$

$$\|V_t^{1/2} y\|_{V_t^{-1}}^2 = (V_t^{1/2} y)^T V_t^{-1} (V_t^{1/2} y)$$

$$= y^T V_t^{1/2} V_t^{-1} V_t^{1/2} y$$

$$= y^T y$$

$$= 1$$

$$E = \left\{ \text{exists } y \in C_\varepsilon : \underbrace{\langle v_t^{1/2} y, \hat{\theta}_t - \theta_* \rangle}_{\geq} \geq \sqrt{2 \ln \left(\frac{|C_\varepsilon|}{\delta} \right)} \right\}$$

$$\text{Prob}(E) \leq \delta$$

$$\|\hat{\theta}_t - \theta_*\|_{v_t} = \left\langle v_t^{1/2} \underbrace{\frac{v_t^{1/2} (\hat{\theta}_t - \theta_*)}{\|\hat{\theta}_t - \theta_*\|_{v_t}}}_{z}, \hat{\theta}_t - \theta_* \right\rangle$$

$$\leq \max_{y \in \text{unit sphere}} \left\langle v_t^{1/2} y, \hat{\theta}_t - \theta_* \right\rangle$$

$$= \max_{y \in \text{unit sphere}} \left\langle v_t^{1/2} (y + y - y), \hat{\theta}_t - \theta_* \right\rangle$$

$$= \max_{\bar{y} \in \text{unit sphere}} \left[\langle V_t^{1/2}(\bar{y} - y), \hat{\theta}_t - \theta_* \rangle + \langle V_t^{1/2}y, \hat{\theta}_t - \theta_* \rangle \right]$$

$$= \max_{\bar{y} \in \text{unit sphere}} \min_{y \in C_E} \left[\langle V_t^{1/2}(\bar{y} - y), \hat{\theta}_t - \theta_* \rangle + \langle V_t^{1/2}y, \hat{\theta}_t - \theta_* \rangle \right]$$

$$= \max_{\bar{y} \in \text{unit sphere}} \min_{y \in C_E} \left[\underbrace{\langle (\bar{y} - y), V_t^{1/2} \hat{\theta}_t - \theta_* \rangle}_{\langle u, v \rangle} + \underbrace{\langle V_t^{1/2}y, \hat{\theta}_t - \theta_* \rangle}_{\langle u, v \rangle} \right]$$

$$\langle u, v \rangle \leq \|u\| \|v\|$$

$$\|\bar{y} - y\| \|V_t^{1/2}(\hat{\theta}_t - \theta_*)\|$$

$$= (\hat{\theta}_t - \theta_*)^T V_t^{1/2} V_t^{1/2} (\hat{\theta}_t - \theta_*)$$

$$= \|\hat{\theta}_t - \theta_*\|_{V_t}$$

$$\leq \max_{\substack{y \in \text{unit} \\ \text{sphere}}} \min_{y \in C_\varepsilon} \left[\|y - g\| \frac{\|\hat{\theta}_t - \theta_*\|}{\nu_t} + \sqrt{2 \ln \left(\frac{|C_\varepsilon|}{8} \right)} \right]$$

$$\|\hat{\theta}_t - \theta_*\|_{\nu_t} \leq \varepsilon \frac{\|\hat{\theta}_t - \theta_*\|}{\nu_t} + \sqrt{2 \ln \left(\frac{|C_\varepsilon|}{8} \right)}$$

$$\|\hat{\theta}_t - \theta_*\|_{\nu_t} (1-\varepsilon) \leq \sqrt{2 \ln \left(\frac{|C_\varepsilon|}{8} \right)}$$

$$\text{or } \|\hat{\theta}_t - \theta_*\|_{\nu_t} \leq \frac{1}{1-\varepsilon} \sqrt{2 \ln \left(\frac{|C_\varepsilon|}{8} \right)}$$

$$\text{pick } \varepsilon = \frac{1}{2}, \quad |C_\varepsilon| \leq \left(\frac{3}{\gamma_1 \gamma_2}\right)^d = 6^d$$

$$\text{Prob} \left(\|\hat{\theta}_t - \theta_*\| \geq 2 \sqrt{2d \ln \left(\frac{6}{\delta} \right)} \right) \leq \delta$$

*V
E*

y_t is already known : deterministic

Need a mechanism to talk about history dependent noise...

Algorithm picks arms A_1, A_2, \dots, A_t (Capital A_t because these are random variables)

(For the sake of analysis)
To keep things simple, pick $d=1$, $\theta_* = 0$ (true parameter)

$$A = \{ 1, 0.5 \}$$

a⁽¹⁾ a⁽²⁾

$$x_t = \langle A_t, \theta_* \rangle + \eta_t = \langle A_t, 0 \rangle + \eta_t = \eta_t$$

we want to look at

$$\sum_{s=1}^t A_s \eta_s$$

Picked by algorithm.

For fixed design this was

$$\sum_{s=1}^t a_s \eta_s$$

Very simplistic algorithm, if x_t was positive I will pick

$$A_{t+1} = a^{(2)}, \text{ else } A_{t+1} = a^{(1)}$$

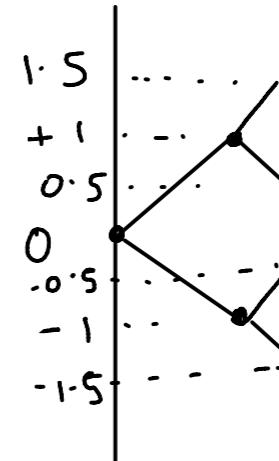
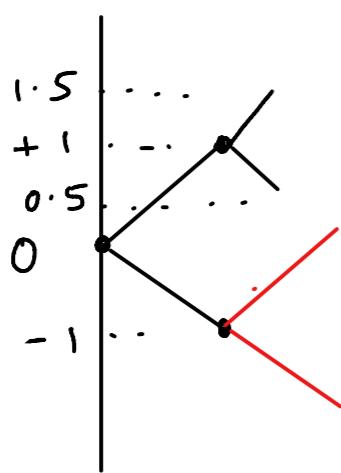
say we start with $A_1 = a^{(1)}$, let $\eta_t \sim_{\text{iid}} \text{Bernoulli}\{-1, +1\}$

$$\sum_{s=1}^2 A_s \eta_s$$

$a^{(1)} \eta_1 + a^{(2)} \eta_2$ if $\eta_1 = +1$
 $a^{(1)} \eta_1 + a^{(1)} \eta_2$ if $\eta_1 = -1$

vs

Fixed Design
in which
 $A_1 = a^{(1)}, A_2 = a^{(2)}$
 $a^{(1)} \eta_1 + a^{(2)} \eta_2$



We need a way to collect the historical information



Filtrations, Martingales

A bit of detour into measure theoretic probability
Integration and Expectation

Focus on positive random variables, i.e. $X \geq 0$ (Ω, \mathcal{F}, P)

$$\mathbb{E}[X] = \int_{\Omega} X(\omega) dP(\omega)$$

$$\Omega = \{1, 2, \dots, 6\}, \quad \mathcal{F} = 2^\Omega, \quad P(A) = \frac{|A|}{6}$$

$$x_1(\omega) = 1, \quad \omega = 1 \\ = 0, \quad \omega = 2, \dots, 6$$

$$\begin{aligned}\mathbb{E}[x_1] &= \int_{\Omega} x_1(\omega) dP(\omega) \\ &= \sum_{\omega=1}^6 x_1(\omega) P(\omega) \\ &= 1 \cdot \frac{1}{6} + 0 \cdot \frac{1}{6} + \dots + 0 \cdot \frac{1}{6} \\ &= \frac{1}{6}\end{aligned}$$

$$x_2(\omega) = 5, \quad \omega = 1 \\ = 0, \quad \omega = 2, \dots, 6$$

$$\begin{aligned}\mathbb{E}[x_2] &= 5 \cdot \frac{1}{6} + 0 \cdot \frac{1}{6} + \dots + 0 \cdot \frac{1}{6} \\ &= \frac{5}{6}\end{aligned}$$

$$x_3 = 2 \cdot 5, \quad \omega = 1, 3, 6$$

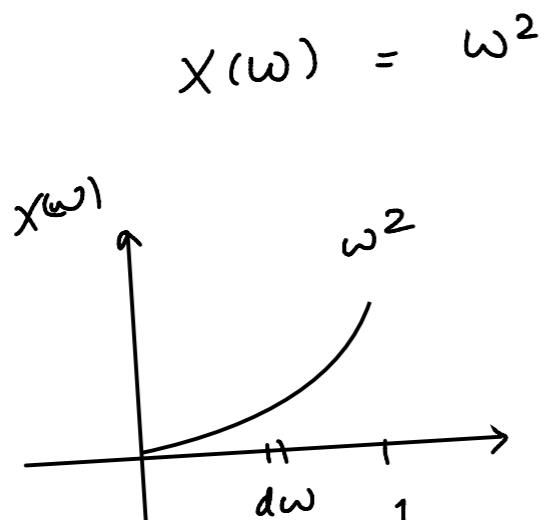
$$= 3, \quad \omega = 2, 4, 5$$

$$\mathbb{E}[x_3] = 2 \cdot 5 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 2 \cdot 5 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 2 \cdot 5 \cdot \frac{1}{6}$$

$$= 2 \cdot 5 \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6} \right) + 3 \cdot \left(\frac{1}{6} + \frac{1}{6} + \frac{1}{6} \right)$$

$$= 2.75$$

$\Omega = [0, 1]$, \mathcal{F} = Borel σ -algebra of open sets, $P([a, b]) = b - a$



$$\mathbb{E}[x] = \int x(\omega) dP(\omega)$$

$$= \int_0^1 \omega^2 d\omega$$

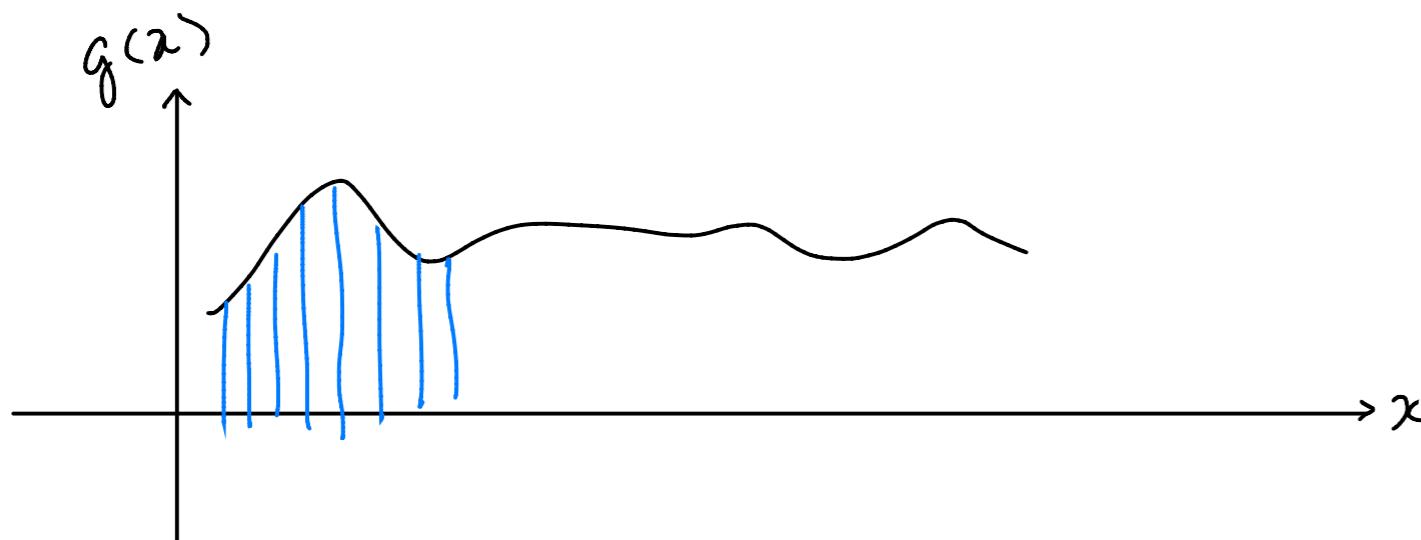
$$= \frac{1}{3} [\omega^3]_0^1$$

$$= \frac{1}{3}$$

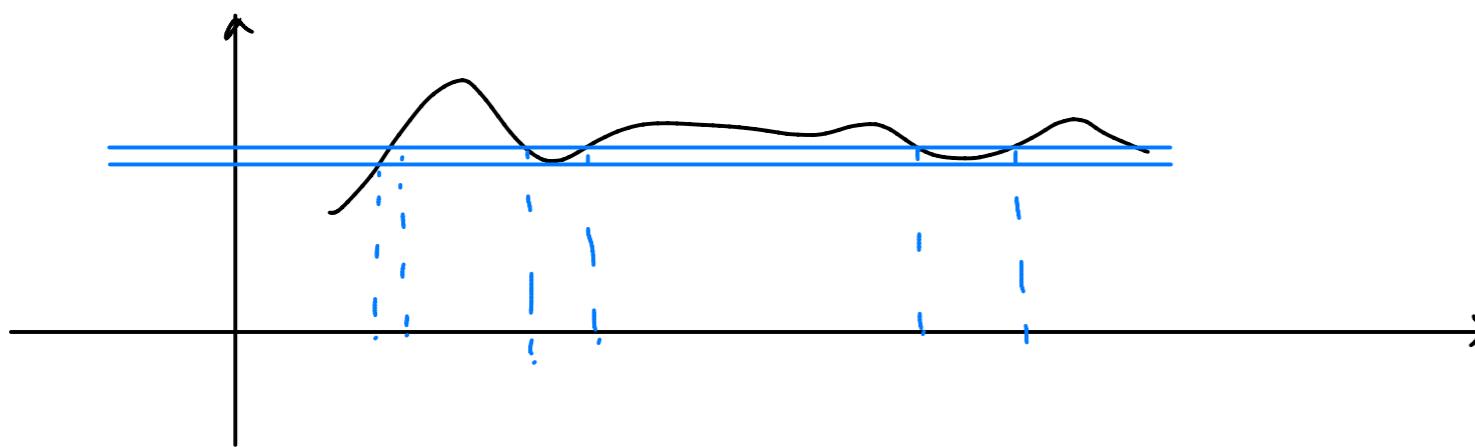
Exercise : Compute $F_x(x) = \text{Prob}(X \leq x)$

$$f_x(x) = \frac{dF}{dx}$$

$$\mathbb{E}[x] = \int_{-\infty}^{\infty} x f_x(x) dx$$



vs



- Define $E[X]$ for indicator random variables

$$x(\omega) = \begin{cases} 1 & = 1, \omega \in A \\ 0 & = 0, \omega \notin A \end{cases}$$

$$E[1_{\{A\}}] = P(A)$$

- Simple random variables

$$x(\omega) = \sum_{i=1}^n \alpha_i 1_{\{\omega \in A_i\}}$$

$$E[x] = \sum_{i=1}^n \alpha_i P(A_i)$$

- For any general $x \geq 0$

$$\int_{\Omega} x dP = \text{Supremum } \left\{ \int_{\Omega} h dP : h \text{ is simple and } 0 \leq h \leq x \right\}$$

For general x

$$x^+ = \max(0, x)$$

$$x^- = \max(0, -x)$$

$$x = x^+ - x^-$$

$$\mathbb{E}[x] = \mathbb{E}[x^+] - \mathbb{E}[x^-]$$

Generated σ -algebra

(Ω, \mathcal{F}, P)

- (Ω, \mathcal{F}, P)
- A random variable is measurable w.r.t (Ω, \mathcal{F}, P) if

$$x^{-1}((a, b)) \in \mathcal{F}$$

- $\sigma(x) :$ σ -algebra generated by x is the smallest σ -algebra such that

$$x^{-1}(a, b) \in \sigma(x)$$

$(\Omega, \mathcal{F}, \mathbb{P})$

$(\Omega, \sigma(x), \mathbb{P})$

Coarse / simple

Fuel / Fine grained

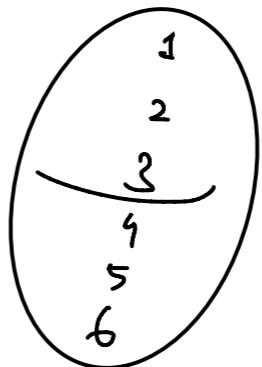
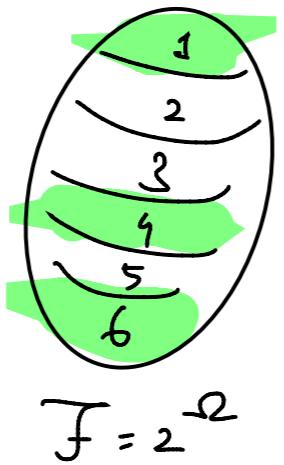
$$\Omega = \{1, \dots, 6\}, \quad \mathcal{F} = 2^\Omega, \quad \mathbb{P}(A) = \frac{|A|}{6}$$

$$y_1(\omega) = -1, \quad \omega = 1, 2, 3 \\ = +1, \quad \omega = 4, 5, 6$$

$$\mathcal{F} = 2^\Omega = \{\emptyset, \Omega, \{1\}, \{2\}, \dots, \{6\}, \{1, 2\}, \dots, \dots\}$$

$$\sigma(y_1) = \{\emptyset, \Omega, \{1, 2, 3\}, \{4, 5, 6\}\}$$

$$y_2(\omega) = +1, \quad \omega = 1, 2 \\ = 0, \quad \omega = 3 \\ = -1, \quad \omega = 4, 5, 6$$



Exercise $\sigma(Y_2) = ?$ $\sigma(Y_2)$ is more finer than $\sigma(Y_1)$

$$\sigma(Y_2) = \{\emptyset, \omega, \{1, 2\}, \{3\}, \{4, 5, 6\}, \{1, 2, 3\}, \{1, 2, 4, 5, 6\}, \{3, 4, 5, 6\}\}$$

$$\sigma(Y_1) = \{\emptyset, \omega, \{1, 2, 3\}, \{4, 5, 6\}\}$$

$$\sigma(Y_2) > \sigma(Y_1)$$

Finer

Coarse

$$Y_3 = +1, \quad \omega = 1, 2, 3, 4 \\ = -1, \quad \omega = 5, 6$$

$$\sigma(Y_3) = \{ \phi, \omega, \{1, 2, 3, 4\}, \{5, 6\} \}$$

$$\sigma(Y_2) = \{ \phi, \omega, \{1, 2\}, \{3\}, \{4, 5, 6\}, \{1, 2, 3\}, \{1, 2, 4, 5, 6\}, \{3, 4, 5, 6\} \} \\ = \mathcal{F}_2$$

$$\sigma(Y_1) = \{ \phi, \omega, \{1, 2, 3\}, \{4, 5, 6\} \} \\ = \mathcal{F}_1$$

One cannot compare $\sigma(Y_3)$ with $\sigma(Y_1)$ and $\sigma(Y_2)$

for this pair of Y_1 and Y_2

$$\sigma(Y_1, Y_2) = \sigma(Y_2)$$

$$\sigma(Y_1, Y_3) = \{ \phi, \omega, \{1, 2, 3\}, \{4, 5, 6\}, \{1, 2, 3, 4\}, \{5, 6\}, \{1, 2, 3, 5, 6\}, \{4\} \}$$

Exercise

$$\sigma(Y_2, Y_3), \quad \sigma(Y_1, Y_2, Y_3) = \mathcal{F}_3$$

$(\omega, \sigma(Y_1, Y_2, Y_3), P)$ is enough to talk about Y_1, Y_2 and Y_3

conditional expectation with respect to a σ -algebra

Let y_1, y_2, y_3 be defined as before.

$$\begin{aligned}y_1(\omega) &= -1, \quad \omega = 1, 2, 3 \\&= +1, \quad \omega = 4, 5, 6\end{aligned}\quad \left| \begin{array}{ll} y_2(\omega) = +1, & \omega = 1, 2 \\ & \\ & = 0, \quad \omega = 3 \\ & \\ & = -1, \quad \omega = 4, 5, 6 \end{array} \right. \quad \left| \begin{array}{ll} y_3 = +1, & \omega = 1, 2, 3, 4 \\ & \\ & = -1, \quad \omega = 5, 6 \end{array} \right.\end{aligned}$$

$$x(\omega) = \omega$$

$$\mathbb{E}[x] = 3.5$$

$$\begin{aligned}\mathbb{E}[x | y_1 = -1] &= 1 \cdot \frac{1}{3} + 2 \cdot \frac{1}{3} + 3 \cdot \frac{1}{3} \\&= 2\end{aligned}$$

$$\begin{aligned}\mathbb{E}[x | y_1 = +1] &= 4 \cdot \frac{1}{3} + 5 \cdot \frac{1}{3} + 6 \cdot \frac{1}{3} \\&= 5\end{aligned}$$

$\mathbb{E}[x | y]$ is a random variable

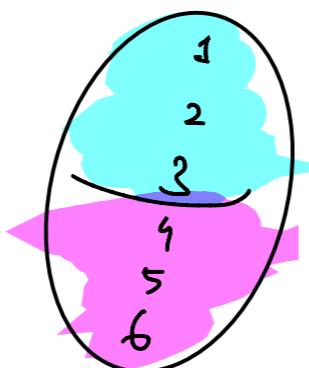
$$\mathbb{E}[x | y]$$

Moral: we are looking
at random variable
 x through the
lens of y

$$(\mathbb{E}[X|Y])(\omega) = 2, \quad \omega = 1, 2, 3$$

$$= 5, \quad \omega = 4, 5, 6$$

This is X seen through the granularity of Y ,



$$X^{-1}(1) = \{1\} \notin \sigma(Y)$$

$$\mathbb{E}[X|Y] = \mathbb{E}[X|\sigma(Y)]$$

this can be generalized into $\mathbb{E}[X|H]$ where H is a

σ -algebra $H \subseteq \mathcal{F}$

Exercise $\mathbb{E}[X|\sigma(Y_1)], \mathbb{E}[X|\sigma(Y_2)], \mathbb{E}[X|\sigma(Y_3)]$

$$\mathbb{E}[X|\sigma(Y_1, Y_2, Y_3)]$$

Formally, pick any $h \in \mathcal{H}$

$$\int_{\mathcal{H}} \mathbb{E}[X|\mathcal{H}] dP = \int_{\mathcal{H}} X dP \quad (\text{local averaging property})$$

and $\mathbb{E}[X|\mathcal{H}]$ is \mathcal{H} measurable

$$\mathcal{H} = \sigma(Y_1) \quad \text{and} \quad \mathcal{H} = \{1, 2, 3\}$$

$$\int_{\{1, 2, 3\}} \mathbb{E}[X|\sigma(Y_1)] dP = \int_{\{1, 2, 3\}} X dP$$

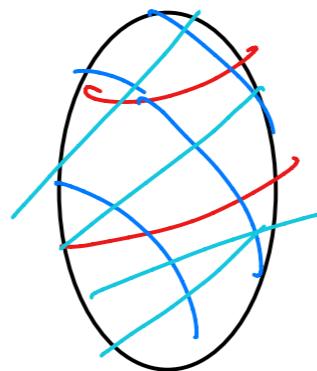
$$2 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} = 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6}$$

$$1 = 1$$

Filtration: $\{\mathcal{F}_t\}_{t \geq 0}$ a sequence of increasing σ -algebras

$$\mathcal{F}_1 \supseteq \mathcal{F}_2 \supseteq \mathcal{F}_3 \dots \supseteq \mathcal{F}_t \supseteq \mathcal{F}_{t+1} \dots \supseteq \mathcal{F} = \mathcal{F}_\infty$$

$$\mathcal{F}_1 = \sigma(Y_1), \quad \mathcal{F}_2 = \sigma(Y_1, Y_2), \quad \mathcal{F}_3 = \sigma(Y_1, Y_2, Y_3)$$



Filtration: Gridding / meshing / breaking down of Ω
finer and finer as we get more and more information

Martingale: (Doob)

$\{X_t\}_{t \geq 0}$ be a sequence of real valued random variables.

$\{\mathcal{F}_t\}_{t \geq 0}$ be a filtration, such that X_t are \mathcal{F}_t measurable
(X_3 is \mathcal{F}_3 measurable, X_4 may/may not be \mathcal{F}_3 measurable)

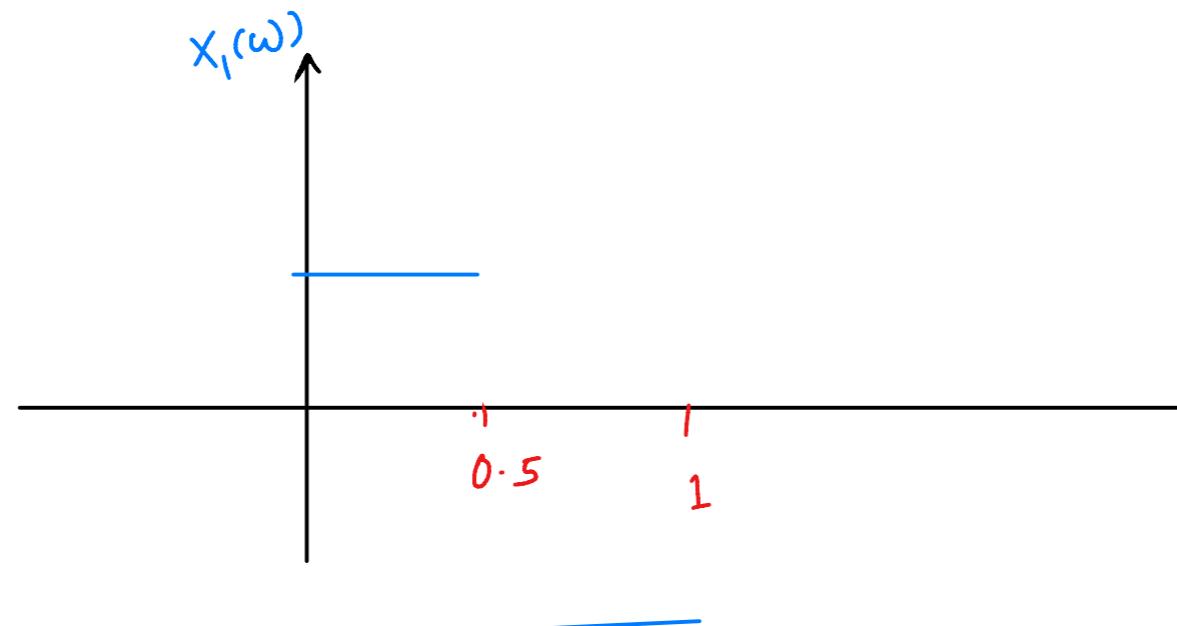
We say $\{X_t\}_{t \geq 0}$ to be a martingale if

$$\mathbb{E}[X_t | \mathcal{F}_{t-1}] = X_{t-1}$$

Symmetric Random Walk $\sum_{s=1}^t \eta_s$, η_s iid Bernoulli $\{-1, +1\}$

$\Omega = [0, 1]$, \mathcal{F} = Borel σ -algebra of open sets, P = length measure

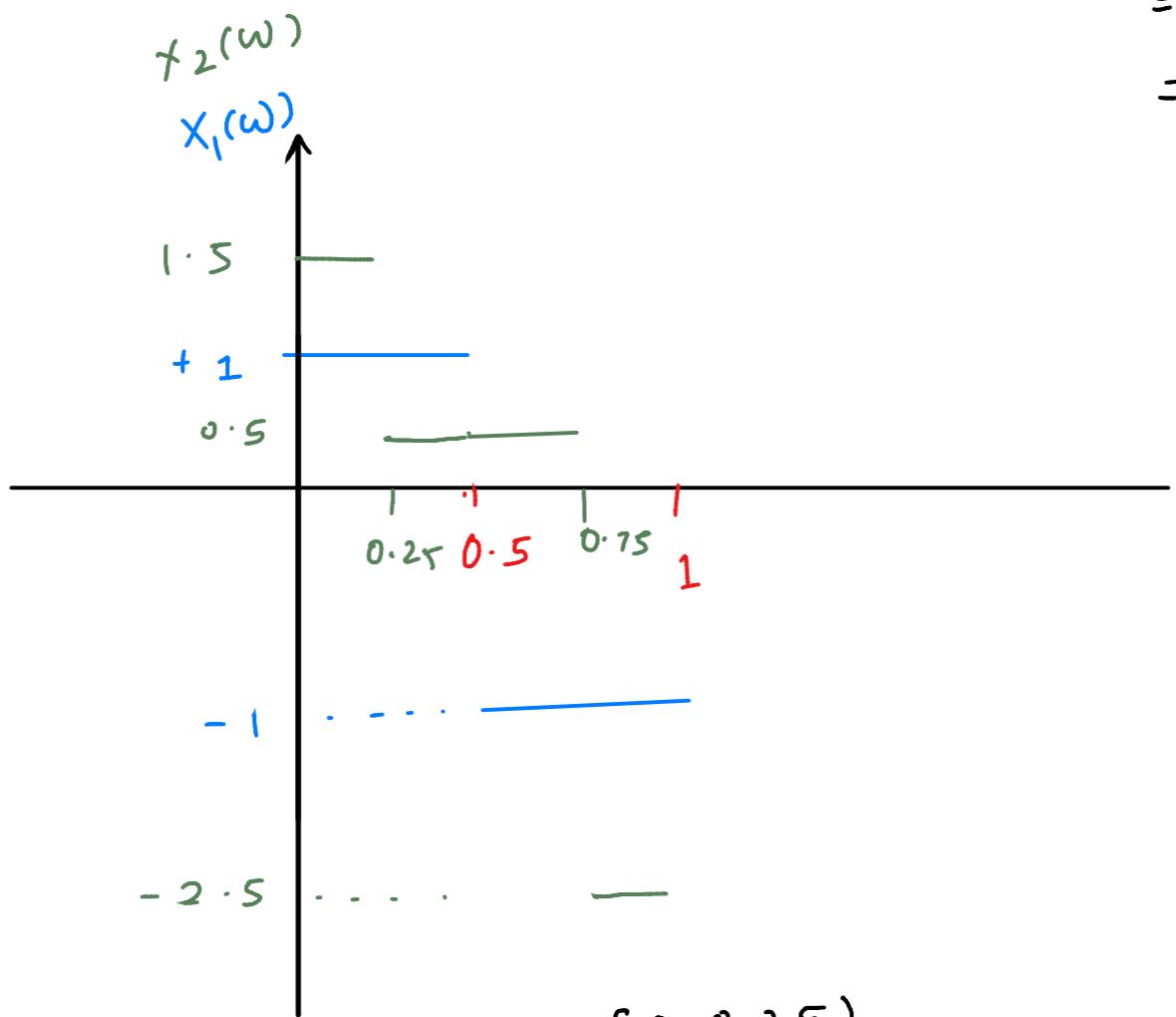
$$\mathcal{F}_1 = \{\emptyset, \Omega, [0, 0.5], [0.5, 1]\}, \quad X_1 = \begin{cases} +1, & \omega \in [0, 0.5) \\ -1, & \omega \in [0.5, 1] \end{cases}$$



$$\mathbb{E}[x_2 | \mathcal{F}_1] = x_1$$

$$\mathcal{F}_2 = \sigma\left(\{\phi, \omega, [0, 0.25), [0.25, 0.5) \cup [0.5, 0.75), [0.75, 1]\}\right)$$

$$x_2 = \begin{cases} +2 & , \omega \in [0, 0.25) \\ 0 & , \omega \in [0.25, 0.5) \\ 0 & , \omega \in [0.5, 0.75) \\ -2 & , \omega \in [0.75, 1] \end{cases}$$



Asymmetric

$$x_2 = \begin{cases} 1.5 & , \omega \in [0, 0.25) \\ 0.5 & , \omega \in [0.25, 0.5) \\ 0.5 & , \omega \in [0.5, 0.75) \\ -2.5 & , \omega \in [0.75, 1] \end{cases}$$

$$\mathbb{E}[x_2 | \mathcal{F}_1] = x_1$$

$\mathcal{F}_2 = \sigma\text{-algebra containing } ([0, 1/8), [1/8, 1/4), [1/4, 3/8), \dots, [7/8, 1])$

$$X_3 = 2.25, \omega \in [0, 1/8)$$

$$0.75, \omega \in [1/8, 1/4)$$

$$-0.25, \omega \in [1/4, 3/8)$$

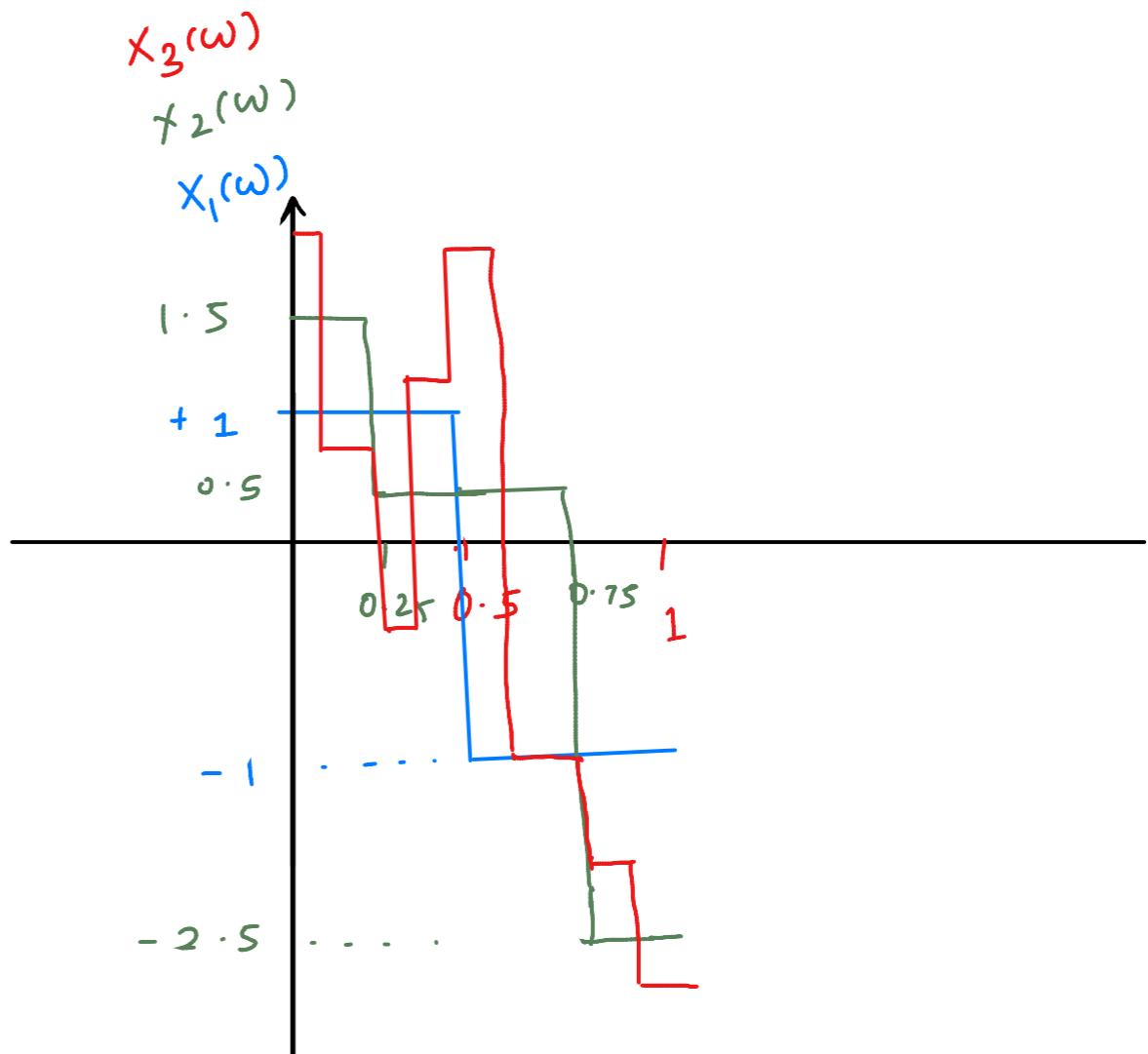
$$1.25, \omega \in [3/8, 1/2)$$

$$2, \omega \in [1/2, 5/8)$$

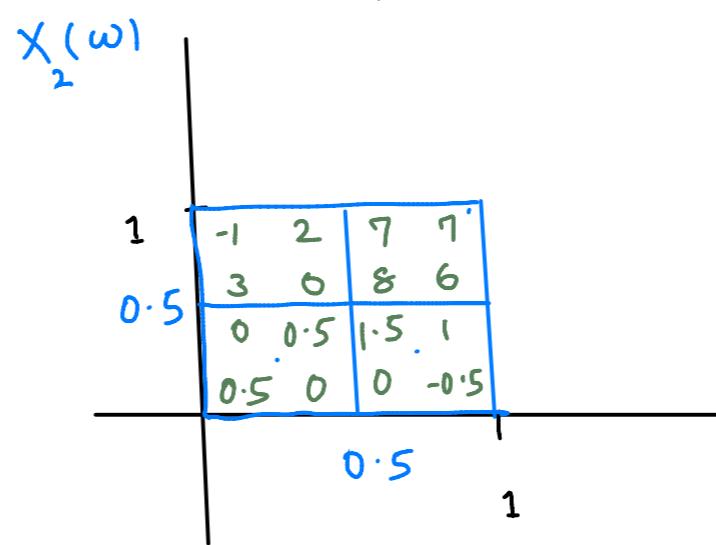
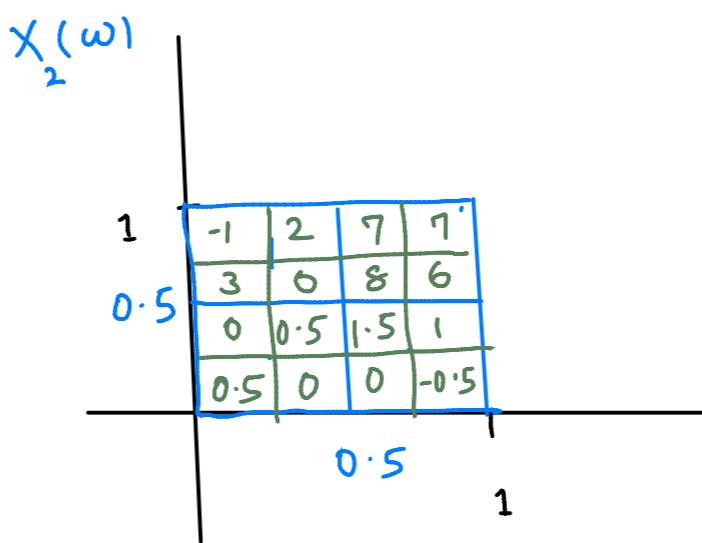
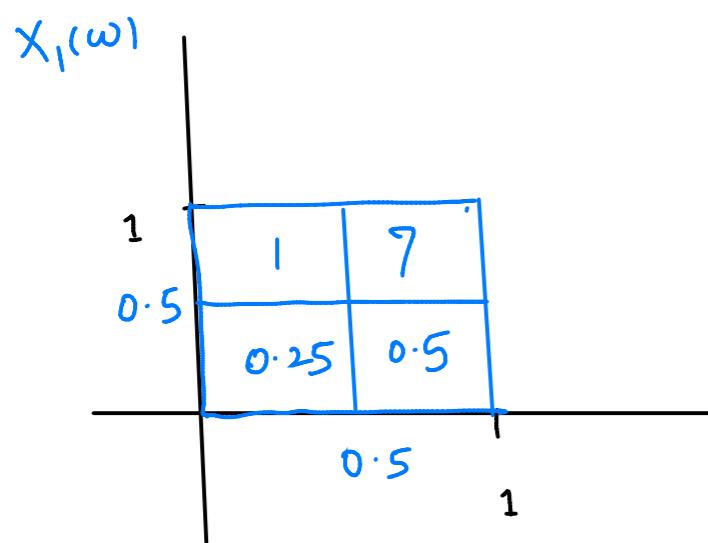
$$-1, \omega \in [5/8, 3/4)$$

$$-2, \omega \in [3/4, 7/8)$$

$$-3, \omega \in [7/8, 1]$$



$\Omega = [0, 1] \times [0, 1]$, \mathcal{F} = Borel σ -algebra
of open rectangles P = area measure



$$\mathbb{E}[X_{10} | \mathcal{F}_5] = X_5 \quad \mathbb{E}[X_{10} | \mathcal{F}_9] = X_9$$

Stopping Time: τ is a random variable taking values in
 $\tau : \Omega \rightarrow \{1, 2, 3, \dots\} \cup \{\infty\}$ (denotes a random time instance)

$\tau(\omega)$

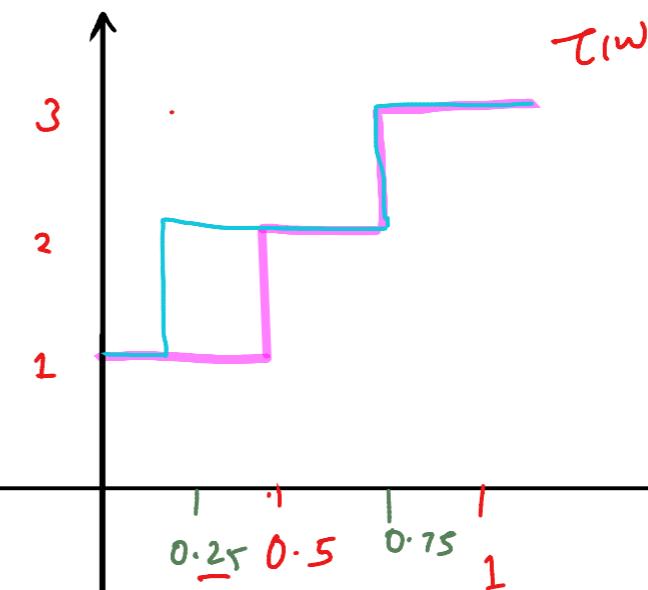
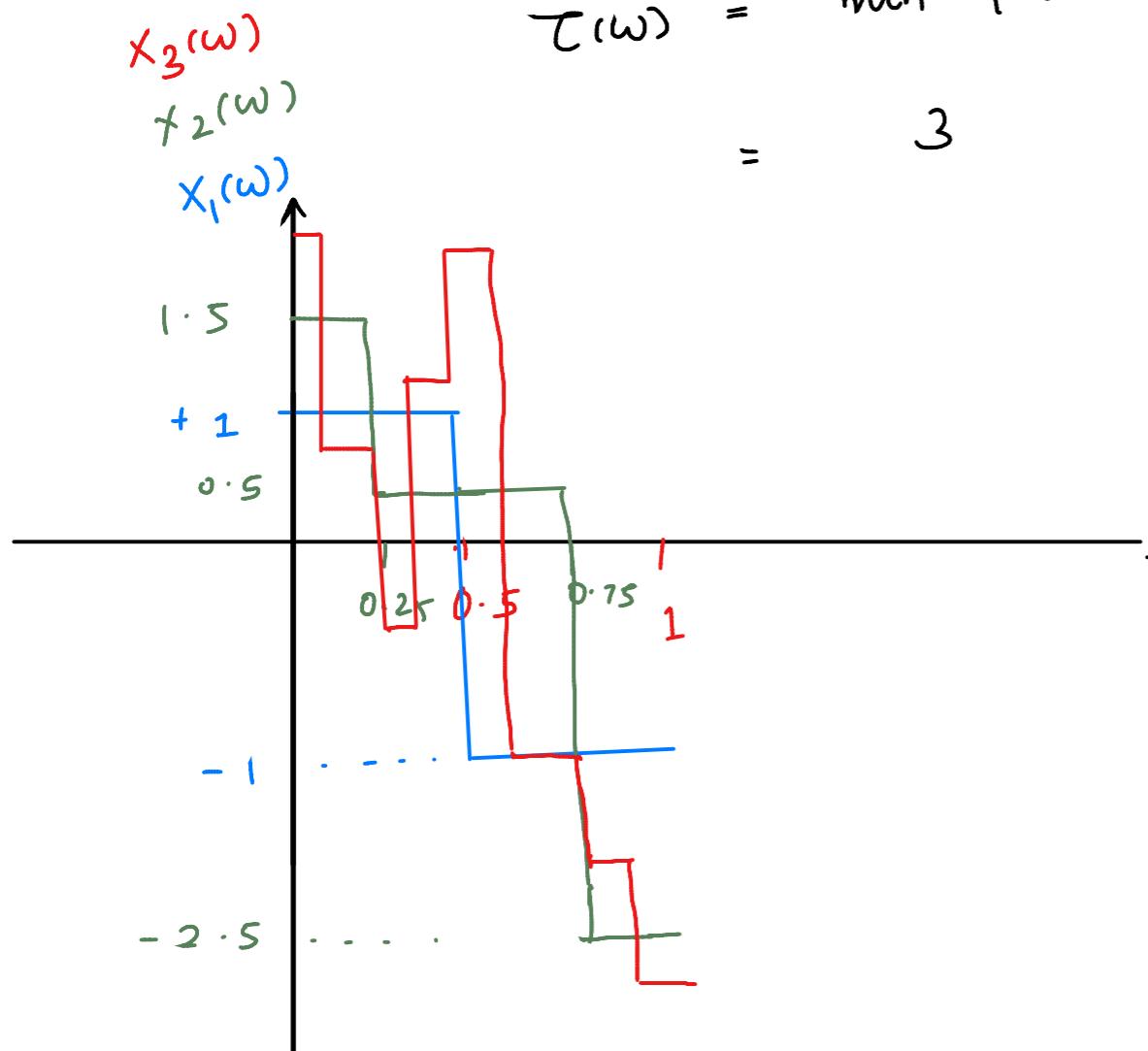
↓
history dependent decisions are made

Idea: If I am stopping at some time $T(w) = t$ then it should be based on information available upto time t

$$\{w: T(w) \leq t\} \in \mathcal{F}_t$$

Example: Take x_1, x_2, x_3 as above (1-dim case)

$$T(w) = \min \{t: x_t^{(w)} > 0\}, \text{ if such } t \text{ exists}$$



How do you get an example of T but is not a stopping time?

not a stopping time
is a stopping time

Stopped Random Variable $X_{\tau} : \Omega \rightarrow \mathbb{R}$

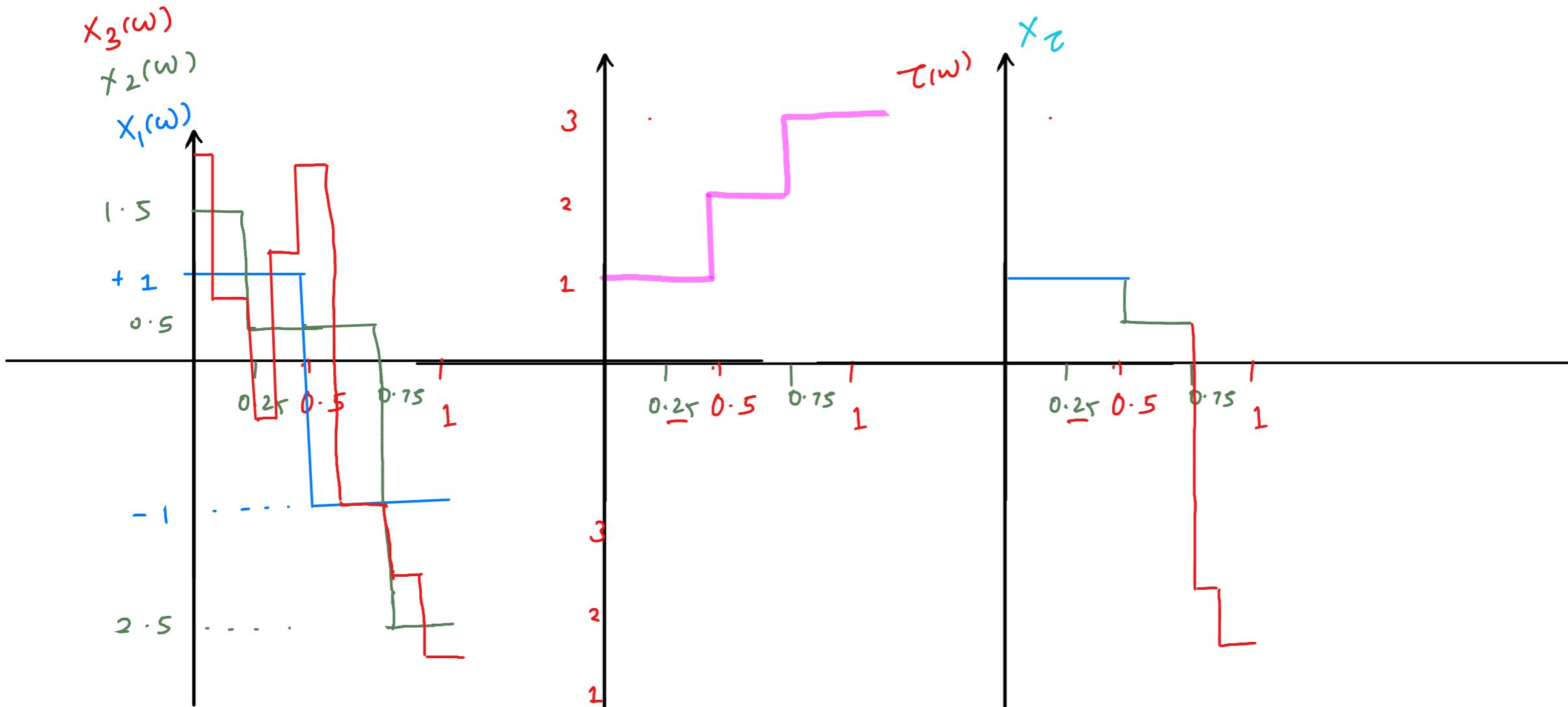
$$X_{\tau}^{(\omega)} = X(\omega)$$

↑
name ↑
T(w)
function

- Pick any $\omega \in \Omega$, $x_1(\omega), \dots, x_{100}(\omega), \dots, x_t(\omega)$
is a sequence of real numbers

- Read out $\tau(\omega)$ (which is a time instance), say $\tau(\omega) = 95$

$$X_{\tau}(\omega) = X_{95}(\omega)$$

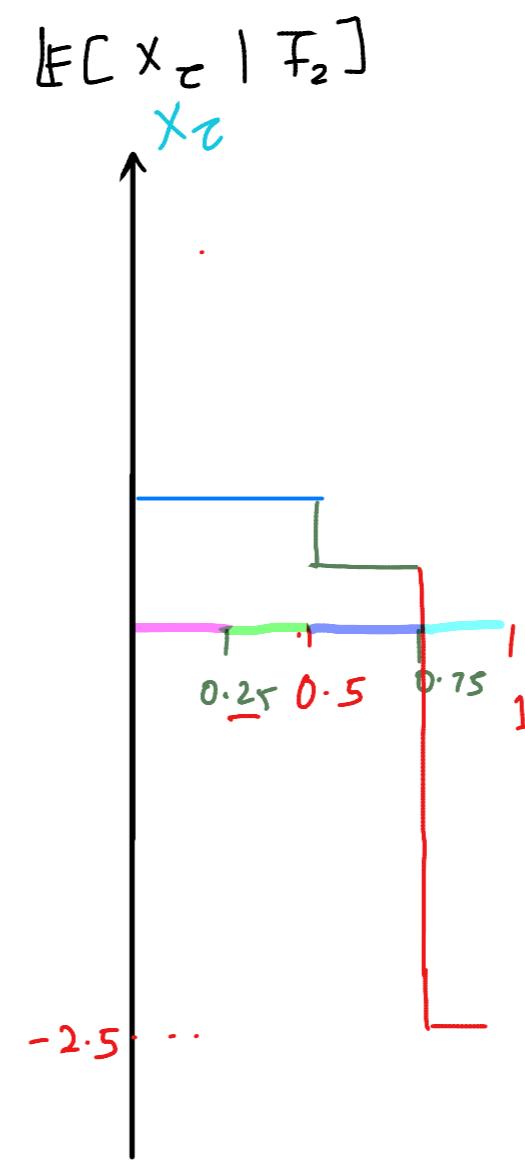
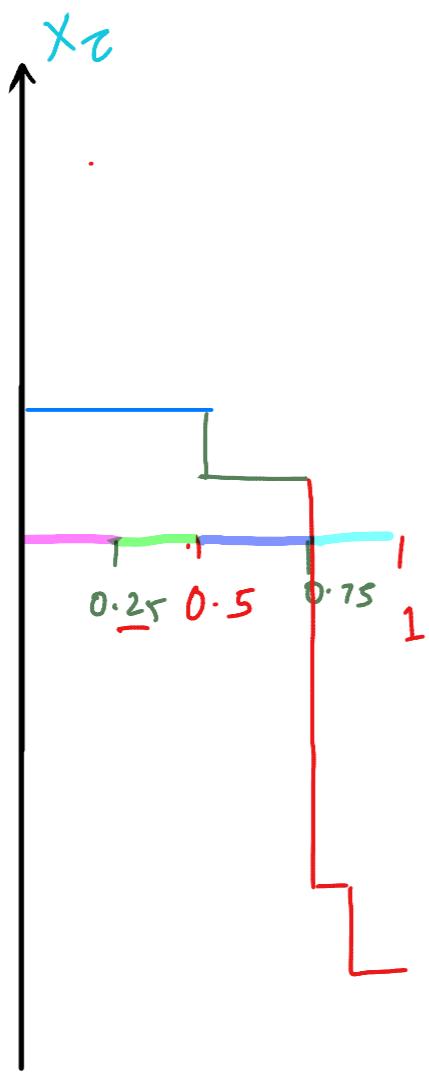


Fact: Say τ is a stopping time such that there exists a $n \in \mathbb{N}$
such that

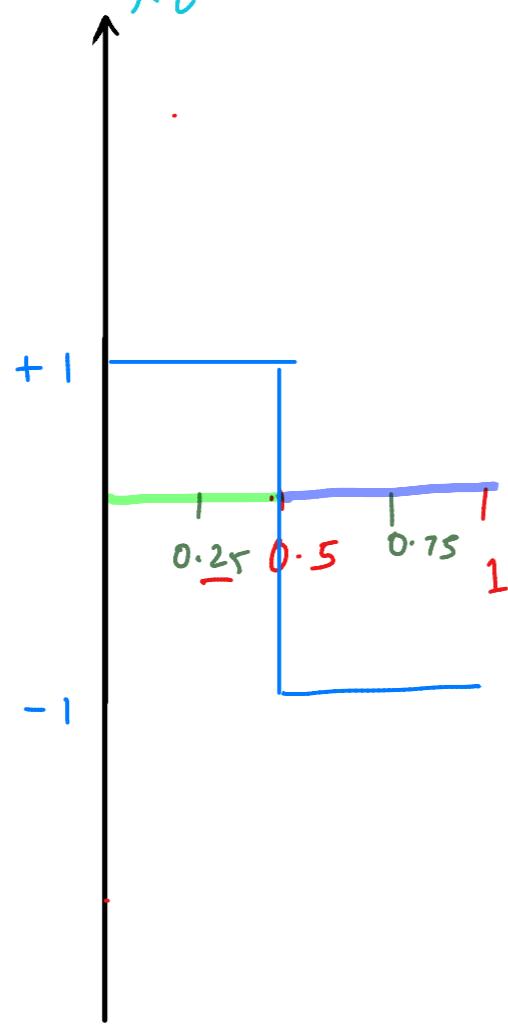
$$\text{Prob}(\tau > n) = 0$$

in this case, we can say

$$\mathbb{E}[X_\tau] = \mathbb{E}[X_0] = \mathbb{E}[\mathbb{E}[\mathbb{E}[X_\tau | \mathcal{F}_2] | \mathcal{F}_1]]$$



$$\mathbb{E} \left[\mathbb{E}[x_2 | \mathcal{F}_2] | \mathcal{F}_1 \right] = \mathbb{E}[x_1]$$



Sub - Martingales : $\mathbb{E}[X_t | \mathcal{F}_{t-1}] \geq X_{t-1}$ (Modify X_1, X_2, X_3 such that)
 they are
 (i) Sub - Martingales

Super - Martingales : $\mathbb{E}[X_t | \mathcal{F}_{t-1}] \leq X_{t-1}$
 (ii) Super - Martingales

Maximum Principle (Extension of Markov Inequality)

$$\text{Markov} : \text{Prob}(X > \varepsilon) \leq \frac{\mathbb{E}[X]}{\varepsilon}$$

(For a positive random variable x)

$$\text{Prob}(\sup_{t \in \mathbb{N}} X_t > \varepsilon) \leq \frac{\mathbb{E}[X_0]}{\varepsilon} \quad (\text{Prob}(X_0 > \varepsilon) \leq \frac{\mathbb{E}[X_0]}{\varepsilon})$$

Maximum

Let $\{X_t\}_{t \geq 0}$ be a sequence of positive random variables

Let $\{X_t\}_{t \geq 0}$ also be a super martingale

$$\mathbb{E}[X_0] \geq \mathbb{E}[X_1] \geq \mathbb{E}[X_2]$$

(Exercise: say you have access to a black-box to which if you give some $\alpha > 0$ it will give a sample from uniform $(0, \alpha)$. Use this black-box to construct a supermartingale)

Proof: Consider $A_n = \{ \omega: \sup_{t \leq n} X_t(\omega) > \varepsilon \}$

$$\begin{aligned}\tau(\omega) &= \min \{ t \leq n : X_t > \varepsilon \}, \text{ if such a } t \text{ exists} \\ &\quad , \text{ otherwise} \\ &= n+1\end{aligned}$$

$$\mathbb{E}[X_0] = \mathbb{E}[X_\tau] \quad (\text{for martingale})$$

$$\mathbb{E}[X_0] \geq \mathbb{E}[X_\tau] \quad (\text{for super-martingale})$$

$$\geq \mathbb{E}[X_\tau \cdot \mathbb{1}_{\{\tau \leq n\}}]$$

$$\geq \mathbb{E}[\varepsilon \cdot \mathbb{1}_{\{\tau \leq n\}}]$$

$$= \varepsilon \cdot \text{Prob}(\tau \leq n)$$

$$= \varepsilon \cdot \text{Prob}(A_n)$$

$$\text{Prob}(A_n) \leq \frac{\mathbb{E}[X_0]}{\varepsilon}$$

$$A_1 \subseteq A_2 \subseteq A_3 \dots \subseteq A_n \subseteq A_{n+1} \dots$$

$$\text{Prob} \left(\sup_{t \in \mathbb{N}} X_t \geq \varepsilon \right) \leq \frac{\mathbb{E}[X_0]}{\varepsilon}$$

Chernoff method

$$\begin{aligned}\text{Prob}(X \geq \varepsilon) &= \text{Prob}(e^{\lambda X} \geq e^{\lambda \varepsilon}) \\ &\leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda \varepsilon}} \\ &\leq e^{(\lambda^2 \sigma^2 / 2 - \lambda \varepsilon)}\end{aligned}$$

\uparrow
find the optimal λ

Method of mixtures

idea is to put a density on X and mix the bounds.

Method of Mixtures

$$\text{Prob}(x \geq \varepsilon) \leq e^{-\lambda\varepsilon + \lambda^2\sigma^2/2}$$

$$L.H.S \leq R.H.S, \text{ consider } \alpha_1, \alpha_2 > 0$$

\downarrow \downarrow
 α_1 L.H.S α_1 R.H.S

$$\alpha_2 L.H.S \leq \alpha_2 R.H.S$$

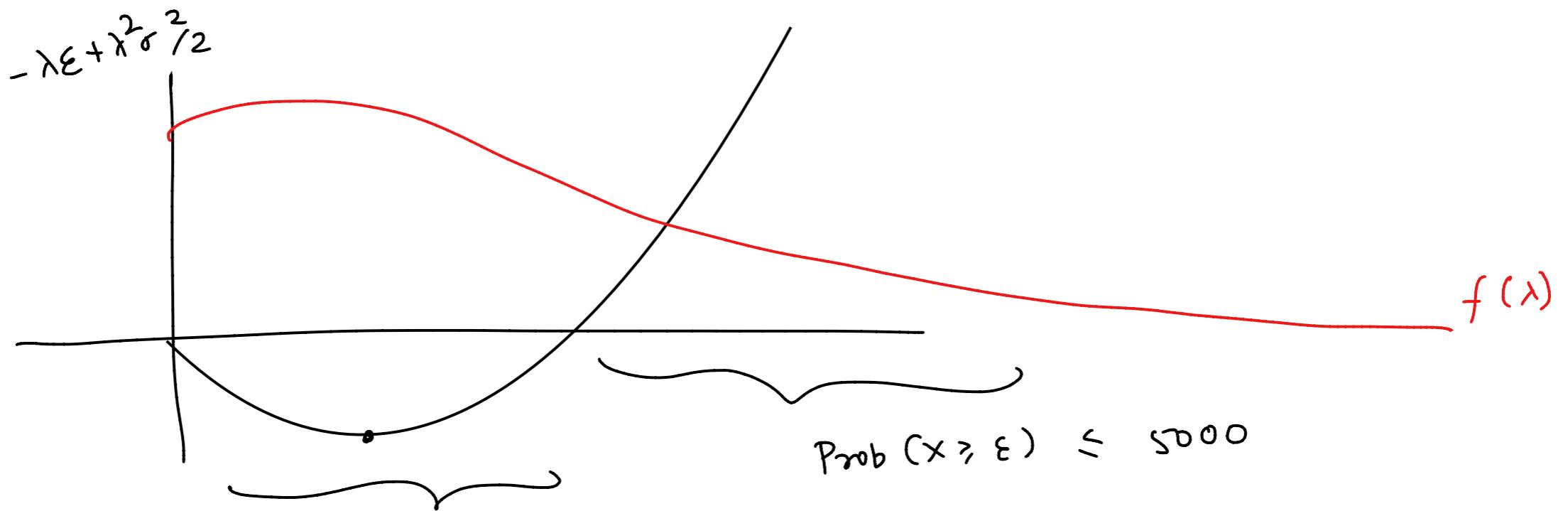
$$\alpha_2 L.H.S \leq \alpha_2 R.H.S$$

$$(\alpha_1 + \alpha_2) L.H.S \leq (\alpha_1 + \alpha_2) R.H.S$$

Re-arranging

choice 1: $\int_0^\infty \text{Prob}(x \geq \varepsilon) \cdot e^{\lambda\varepsilon - \lambda^2\sigma^2/2} \cdot f(x) dx \leq \int_0^\infty 1 \cdot f(x) dx$

choice 2: $\int_0^\infty \text{Prob}(x \geq \varepsilon) \cdot f(x) dx \leq \int_0^1 e^{-\lambda\varepsilon + \lambda^2\sigma^2/2} \cdot f(x) dx$



$$\text{Prob}(x > \varepsilon) \leq 0.1 \text{ or } 0.9$$

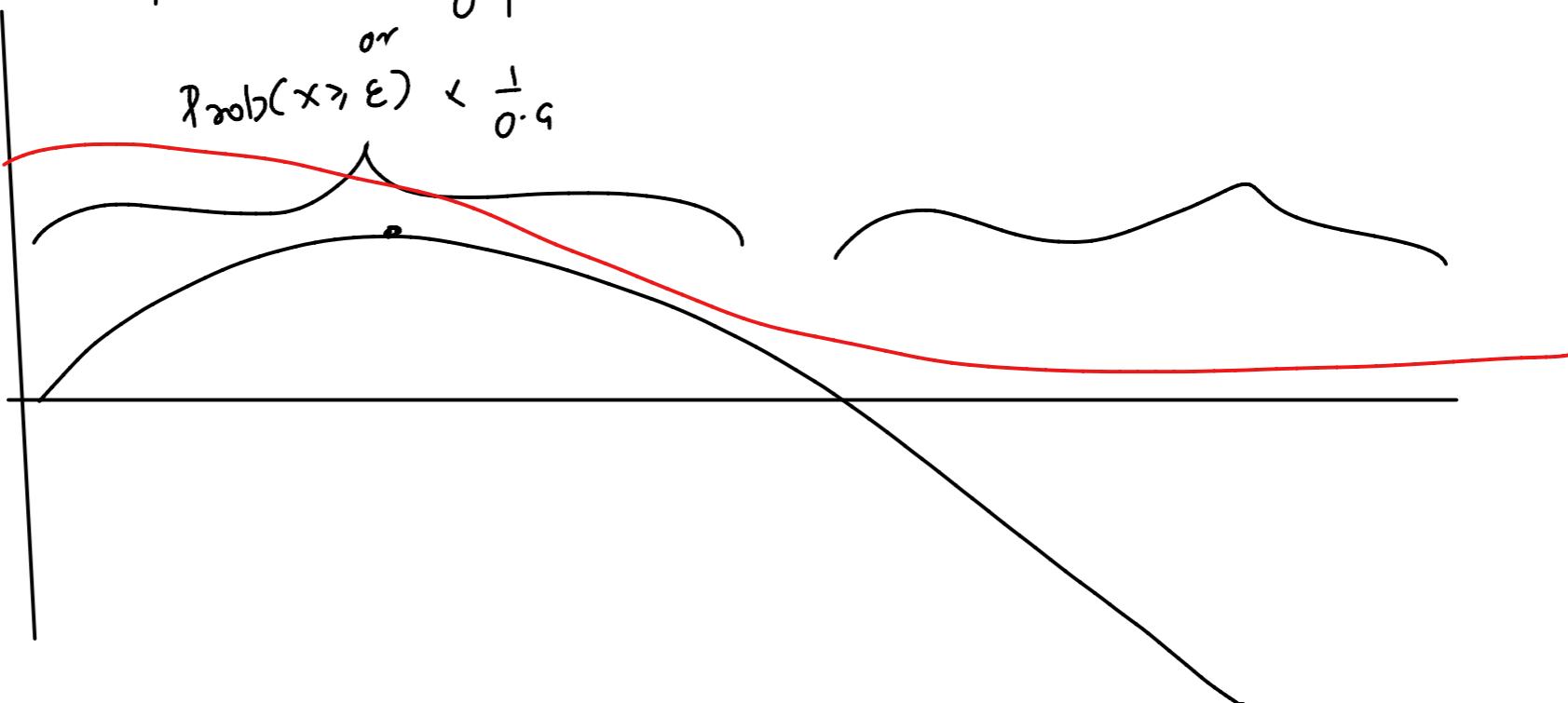
L.H.S

$$\text{Prob}(x > \varepsilon) \perp 0.1$$

or

$$\text{Prob}(x > \varepsilon) < \frac{1}{0.9}$$

$$\text{Prob}(x > \varepsilon) \frac{1}{5000}$$



choice 2:

$$L.H.S \leq 0.1 \times 0.5$$

$$L.H.S \leq 1000 \times 0.5$$

$$L.H.S \leq 500.05$$

choice 1:

$$L.H.S \frac{1}{0.1} \leq 1 \times 0.5$$

$$L.H.S \frac{1}{5000} \leq 1 \times 0.5$$

$$L.H.S \left(\frac{0.5}{0.1} + \frac{0.5}{5000} \right) \leq 1$$

0.2

0.8

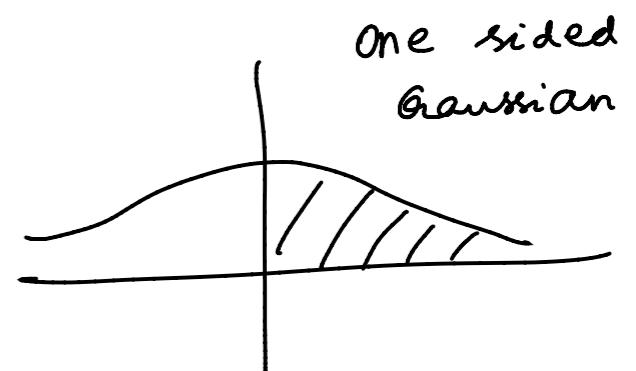
L.H.S ≤ 0.5

$$L.H.S (5 + \text{small value}) \leq 1$$

$$L.H.S \leq \frac{1}{5 + \text{small value}} \leq 0.2$$

$$\int_0^{\infty} \text{Prob}(X > \varepsilon) \cdot e^{\lambda\varepsilon - \frac{\lambda^2 \sigma^2}{2}} \cdot f(x) dx \leq \int_0^{\infty} 1 \cdot f(x) \cdot dx$$

choose $f(x) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{x^2}{2\sigma_c^2}}$



Looking at only L.H.S (that too leaving out
 $\text{Prob}(X > \varepsilon)$)

$$\text{Prob}(X > \varepsilon)$$

$$\text{Prob}(X > x > \varepsilon)$$

$$\text{Prob}(e^{\lambda x} > e^{\lambda \varepsilon})$$

$$\frac{1}{\sqrt{2\pi\sigma_c^2}} \int_0^{\infty} e^{\lambda\varepsilon - \frac{\lambda^2 \sigma^2}{2}} e^{-\frac{x^2}{2\sigma_c^2}} dx$$

$$= \frac{1}{\sqrt{2\pi\sigma_c^2}} \int_0^{\infty} e^{-\frac{1}{2} \left(\frac{\lambda^2 \sigma^2}{\sigma_c^2} + \frac{x^2}{\sigma_c^2} - 2\lambda\varepsilon \right)} dx$$

$$= \frac{2}{\sqrt{2\pi\sigma_c^2}} \int_0^\infty e^{-\frac{1}{2} \left(\underbrace{\lambda^2(\sigma^2 + \sigma_c^{-2})}_{a^2} - \frac{2\lambda\varepsilon}{2ab} \right)} d\lambda$$

Completing the squares

$$= \frac{2}{\sqrt{2\pi\sigma_c^2}} \int_0^\infty e^{-\frac{1}{2} \left(\lambda \sqrt{\sigma^2 + \sigma_c^{-2}} - \frac{\varepsilon}{\sqrt{\sigma^2 + \sigma_c^{-2}}} \right)^2} \cdot e^{\frac{1}{2} \frac{\varepsilon^2}{(\sigma^2 + \sigma_c^{-2})}} \cdot d\lambda$$

$$= \frac{2}{\sqrt{2\pi\sigma_c^2}} \cdot e^{\frac{1}{2} \frac{\varepsilon^2}{(\sigma^2 + \sigma_c^{-2})}} \cdot \int_0^\infty e^{-\frac{1}{2} \left(\lambda \sqrt{\sigma^2 + \sigma_c^{-2}} - \frac{\varepsilon}{\sqrt{\sigma^2 + \sigma_c^{-2}}} \right)^2} \cdot d\lambda$$

$$= \frac{2}{\sqrt{2\pi\sigma_c^2}} \cdot e^{\frac{1}{2} \frac{\varepsilon^2}{(\sigma^2 + \sigma_c^{-2})}} \cdot \int_0^\infty e^{-\frac{1}{2} (\sigma^2 + \sigma_c^{-2}) \cdot \left(\lambda - \frac{\varepsilon}{\sigma^2 + \sigma_c^{-2}} \right)^2} \cdot d\lambda$$

$$= \frac{2}{\sqrt{2\pi\sigma_c^2}} \cdot e^{-\frac{1}{2} \frac{\varepsilon^2}{(\sigma^2 + \sigma_c^{-2})}} \int_0^\infty e^{-\frac{(\lambda - \frac{\varepsilon}{\sigma^2 + \sigma_c^{-2}})^2}{2(\sigma^2 + \sigma_c^{-2})^{-1}}} d\lambda$$

$$= \frac{2}{\sqrt{2\pi\sigma_c^2}} e^{-\frac{1}{2} \frac{\varepsilon^2}{(\sigma^2 + \sigma_c^{-2})}} \cdot \frac{\sqrt{2\pi(\sigma^2 + \sigma_c^{-2})^{-1}}}{2}$$

$$= e^{-\frac{1}{2} \frac{\varepsilon^2}{(\sigma^2 + \sigma_c^{-2})}} \cdot \sqrt{\frac{\sigma_c^{-2}}{(\sigma^2 + \sigma_c^{-2})}}$$

$$\text{Prob}(x \geq \varepsilon) \cdot \left[e^{-\frac{1}{2} \frac{\varepsilon^2}{(\sigma^2 + \sigma_c^{-2})}} \cdot \sqrt{\frac{\sigma_c^{-2}}{(\sigma^2 + \sigma_c^{-2})}} \right] \leq 1$$

$$\begin{aligned} \text{Prob}(x \geq \varepsilon) &\leq \sqrt{\frac{\sigma^2 + \sigma_c^{-2}}{\sigma_c^{-2}}} \cdot e^{-\frac{1}{2} \frac{\varepsilon^2}{\sigma^2 + \sigma_c^{-2}}} \\ &= \sqrt{1 + \left(\frac{\sigma}{\sigma_c}\right)^2} \cdot e^{-\frac{1}{2} \frac{\varepsilon^2}{\sigma^2} \cdot \frac{1}{1 + (\sigma_c^{-1}/\sigma)^2}} \end{aligned}$$

Say we chose $\sigma_c = \sigma^{-1}$

$$\text{Prob}(x \geq \varepsilon) \leq \sqrt{2} \cdot e^{-\frac{1}{2} \frac{\varepsilon^2}{\sigma^2} \cdot \frac{1}{2}}$$

For dim = 1

Fixed Design $\xrightarrow{\text{deterministic}}$

$$S_t = \sum_{s=1}^t a_s \eta_s$$

$$\mathbb{E}[e^{\lambda S_t}] = \mathbb{E}[e^{\lambda \sum_{s=1}^t a_s \eta_s}]$$

$$= \prod_{s=1}^t \mathbb{E}[e^{\lambda a_s \eta_s}]$$

Sequential Design $\xrightarrow{\text{random Variable}}$

$$S_t = \sum_{s=1}^t A_s \gamma_s$$

$$\mathbb{E}[e^{\lambda S_t}] = \mathbb{E}[e^{\lambda \sum_{s=1}^t A_s \gamma_s}]$$

$$= \mathbb{E}[\mathbb{E}[e^{\lambda \sum_{s=1}^{t-1} A_s \gamma_s} | \mathcal{F}_{t-1}]]$$

$$= \mathbb{E}[e^{\lambda \sum_{s=1}^{t-1} A_s \gamma_s} \mathbb{E}[e^{\lambda A_t \gamma_t} | \mathcal{F}_{t-1}]]$$

$$\begin{aligned} &\leq \prod_{s=1}^t e^{\lambda^2 a_s^2 / 2} \\ &= e^{\lambda^2 / 2 \sum_{s=1}^t a_s^2} \\ &\leq e^{\lambda^2 / 2 \sum_{s=1}^{t-1} A_s^2} \end{aligned}$$

claim: $M_t^\lambda = e^{\lambda S_t - \lambda^2 / 2 \sum_{s=1}^t A_s^2}$ is a positive super Martingale

(λ has no restriction)

$$\mathbb{E}[M_t^\lambda | \mathcal{F}_{t-1}] = \mathbb{E}[e^{\lambda S_t - \lambda^2 / 2 \sum_{s=1}^{t-1} A_s^2} | \mathcal{F}_{t-1}]$$

$$\begin{aligned} &= e^{\lambda S_{t-1} - \lambda^2 / 2 \sum_{s=1}^{t-1} A_s^2} \mathbb{E}[e^{\lambda A_t n_t - \lambda^2 A_t^2 / 2} | \mathcal{F}_{t-1}] \\ &\quad \text{---} \\ &\quad M_{t-1}^\lambda \end{aligned}$$

$$\leq M_{t-1}^\lambda$$

Invoke Method of mixtures

$$M_t = \int_{-\infty}^{\infty} M_t^\lambda f(\lambda) d\lambda$$

$$M_t = \int_{-\infty}^{\infty} e^{(\lambda S_t - \lambda^2/2 \sum_{s=1}^t A_s^2)} f(\lambda) d\lambda$$

$f(\lambda)$ be Gaussian with variance $\sigma_c^2 = \frac{1}{\gamma}$,

let $\sigma^2 = \sum_{s=1}^t A_s^2 = V_t$

$$M_t = e^{\frac{1}{2} S_t^2 / (V_t + \gamma)}$$

$$\sqrt{\frac{\gamma}{(V_t + \gamma)}}$$

Compare it with

$$e^{\frac{1}{2} \frac{\epsilon^2}{(\sigma^2 + \sigma_c^{-2})}}$$

$$\sqrt{\frac{\sigma_c^{-2}}{(\sigma^2 + \sigma_c^{-2})}}$$

$$S_t = \sum_{s=1}^t A_s \eta_s, \quad \text{we want to bound } \|S_t\|_{V_t^{-1}}$$

- S_t is the total noise injected in the system

- algorithm dependent

- we project S_t onto $\lambda \in \mathbb{R}^d \Rightarrow$ look $\langle \lambda, S_t \rangle = \lambda^T S_t$

$$\langle \lambda, S_t \rangle = \left\langle \lambda, \sum_{s=1}^t A_s \eta_s \right\rangle = \sum_{s=1}^t \underbrace{\langle \lambda, A_s \rangle}_{\alpha_s} \eta_s$$

$$\sum_{s=1}^t \alpha_s^2 = \sum_{s=1}^t (\langle \lambda, A_s \rangle)^2$$

$$= \sum_{s=1}^t \lambda^T A_s A_s^T \lambda$$

$$= \lambda^T \sum_{s=1}^t A_s A_s^T \lambda = \lambda^T V_t \lambda$$

$$= \|\lambda\|_{V_t}^2$$

Martingale $M_t^\lambda = e^{\langle \lambda, s_t \rangle - \frac{1}{2} \|\lambda\|_{V_t}^2}$

$$M_t^\lambda = \int_{\substack{\text{Over} \\ \mathbb{R}^d}} M_t^\lambda f(\lambda) d\lambda$$

$f(\lambda)$ is a Gaussian with variance γ

$$= \sqrt{\frac{\det(\gamma I)}{\det(\gamma I + V_t)}} e^{\frac{1}{2} \|s_t\|_{V_t^{-1}}^2}$$

where $V_t^{(z)} = \gamma I + V_t$

$$e^{\frac{1}{2} \frac{\epsilon^2}{(\sigma^2 + \sigma_c^{-2})}} \cdot \sqrt{\frac{\sigma_c^{-2}}{(\sigma^2 + \sigma_c^{-2})}}$$

Problem

$$x_t = \langle A_t, \theta_x \rangle + \eta_t$$

$$Y_t = \sum_{s=1}^t A_s A_s^\top$$

$$\hat{\theta}_t = Y_t^{-1} \sum_{s=1}^t A_s x_s$$

Y_t may not be invertible

$$V_t(\gamma) = \gamma I + \sum_{s=1}^t A_s A_s^\top = \gamma I + V_t$$

$$\hat{\theta}_t = V_t(\gamma)^{-1} \sum_{s=1}^t A_s x_s$$

$$\hat{\theta}_t = V_t(\gamma)^{-1} \sum_{s=1}^t A_s (\langle A_s, \theta_x \rangle + \eta_s)$$

$$= V_t(\gamma)^{-1} \left(\sum_{s=1}^t A_s A_s^\top \underbrace{\theta_x}_{V_t} + \sum_{s=1}^t A_s \eta_s \right)$$

$$= V_t(\gamma)^{-1} V_t \theta_x + V_t(\gamma)^{-1} \underbrace{S_t}_{\text{noise part}}$$

Understanding role of γ in Vanilla bandits

$$V_t^{(\gamma)} = \begin{bmatrix} \eta_1(t) + \gamma & 0 & 0 & \cdots & \cdots \\ 0 & \eta_2(t) + \gamma & 0 & \cdots & \cdots \\ \vdots & \ddots & \ddots & \ddots & \eta_d(t) + \gamma \end{bmatrix}$$

$$\begin{aligned}
 \|\hat{\theta}_t - \theta_*\|_{V_t^{(\gamma)}} &= \left\| V_t^{(\gamma)^{-1}} S_t + V_t^{(\gamma)^{-1}} V_t \theta_* - \theta_* \right\|_{V_t^{(\gamma)}} \\
 &\leq \|V_t^{(\gamma)^{-1}} S_t\|_{V_t^{(\gamma)}} + \|(V_t^{(\gamma)^{-1}} V_t - I) \theta_*\|_{V_t^{(\gamma)}} \\
 &= \left(S_t^T V_t^{(\gamma)^{-1}} V_t^{(\gamma)} V_t^{(\gamma)^{-1}} S_t \right)^{1/2} + (\theta_*^T (V_t^{(\gamma)^{-1}} V_t - I) V_t^{(\gamma)}) (V_t^{(\gamma)^{-1}} V_t - I) \theta_*^{1/2} \\
 &= (S_t^T V_t^{(\gamma)^{-1}} S_t)^{1/2} + (\theta_*^T (V_t^{(\gamma)^{-1}} V_t - I) (V_t - V_t^{(\gamma)}) \theta_*)^{1/2} \\
 &= \|S_t\|_{V_t^{(\gamma)^{-1}}} + (\theta_*^T (V_t^{(\gamma)^{-1}} V_t - I) (-\gamma I) \theta_*)^{1/2} \\
 &= \|S_t\|_{V_t^{(\gamma)^{-1}}}
 \end{aligned}$$

$V_t^{(\gamma)} = \gamma I + V_t$

$$= \|s_t\|_{V_t(\gamma)^{-1}} + \gamma^{1/2} (\theta_*^T (I - V_t(\gamma)^{-1} V_t) \theta_*)$$

$V_t(\gamma)^{-1} V_t$ is positive semi definite

$$\leq \|s_t\|_{V_t(\gamma)^{-1}} + \underbrace{\gamma^{1/2} \|\theta_*\|}$$

Bias due to γ

be invoking maximal inequality
we can bound $\|s_t\|_{V_t(\gamma)^{-1}}$

$$M_t = \sqrt{\frac{\det(\gamma I)}{\det(\gamma I + V_t)}} e^{\frac{1}{2} \|s_t\|_{V_t(\gamma)^{-1}}^2}$$

key : Save this for last.

Invoke maximal inequality

$$\begin{aligned} \text{Prob} \left(\sup_{t \in \mathbb{N}} M_t > \frac{1}{\delta} \right) &\leq E[M_0] \delta \\ &\leq \delta \end{aligned}$$

$$\text{Prob} \left(\sup_{t \in \mathbb{N}} \ln M_t \geq \ln \left(\frac{1}{\delta} \right) \right) \leq \delta$$

$$\ln M_t = \frac{1}{2} \|S_t\|^2 V_t^{(\gamma)}^{-1} + \ln \left(\sqrt{\frac{\det(\gamma I)}{\det(V_t^{(\gamma)})}} \right)$$

$$\text{Prob} \left[\sup_{t \in \mathbb{N}} \frac{1}{2} \|S_t\|^2 V_t^{(\gamma)}^{-1} + \ln \left(\sqrt{\frac{\det(\gamma I)}{\det(V_t^{(\gamma)})}} \right) \geq \ln \left(\frac{1}{\delta} \right) \right] \leq \delta$$

$$\text{Prob} \left[\sup_{t \in \mathbb{N}} \|S_t\|^2 V_t^{(\gamma)}^{-1} \geq 2 \ln \left(\frac{1}{\delta} \right) + 2 \ln \left(\sqrt{\frac{\det(V_t^{(\gamma)})}{\det(\gamma I)}} \right) \right] \leq \delta$$

we need to simplify $\frac{\det(V_t^{(\gamma)})}{\det(\gamma I)}$

$$\circ \det(\gamma I) = \gamma^d \det(I) = \gamma^d$$

$$\circ V_t^{(\gamma)} = \gamma I + \sum_{s=1}^t A_s A_s^T$$

$$\det(V_t(\gamma)) = \prod_{i=1}^d \alpha_i \leq \left(\frac{\sum_{i=1}^d \alpha_i}{d} \right)^d \quad (6M \leq AM)$$

$$\text{trace}(V_t(\gamma)) = \sum_{i=1}^d \alpha_i = \text{sum of the diagonal entries}$$

$$V_t(\gamma) = \gamma I + A_1 A_1^\top$$

A, A_1^\top

$$\begin{bmatrix} A_1(1) \\ \vdots \\ A_1(d) \end{bmatrix} \quad \begin{bmatrix} A_1(1) & \dots & A_1(d) \end{bmatrix}$$

$$= \begin{bmatrix} \gamma + A_1(1)^2 & \text{blah} & \text{blah} \\ \text{blah} & \gamma + A_2(2)^2 & \text{blah} \\ \text{blah} & \text{blah} & \gamma + A_1(d)^2 \end{bmatrix}$$

$$V_2(\gamma) = \begin{bmatrix} \gamma + A_1(1)^2 + A_2(1)^2 & \text{blah} & & \\ \text{blah} & \gamma + A_1(2)^2 + A_2(2)^2 & \text{blah} & \\ & \ddots & \ddots & \\ \text{blah} & & & \gamma + A_1(d)^2 + A_2(d)^2 \end{bmatrix}$$

$$\text{trace}(V_t(\gamma)) = d\gamma + \|A_1\|_2^2 + \|A_2\|_2^2 + \dots + \|A_d\|_2^2$$

Assume that $\|A_t\|_2^2 \leq L^2$

$$\text{trace}(V_t(\gamma)) \leq d\gamma + tL^2$$

$$\frac{\det(V_t(\gamma))}{\det(\gamma I)} \leq \left(\frac{d\gamma + tL^2}{d} \right)^d \cdot \frac{1}{\gamma^d}$$

$$= \left(1 + \frac{tL^2}{d\gamma} \right)^d$$

$$\text{Prob} \left[\sup_{t \in \mathbb{N}} \frac{\|S_t\|^2}{V_t(\gamma)^{-1}} \right] \geq 2 \ln \left(\frac{1}{\delta} \right) + \ln \left[\left(1 + \frac{t L^2}{d\gamma} \right)^d \right] \leq \delta$$

$$\text{Prob} \left[\sup_{t \in \mathbb{N}} \frac{\|S_t\|}{V_t(\gamma)^{1/2}} \right] \leq \sqrt{2 \ln \left(\frac{1}{\delta} \right) + d \ln \left(1 + \frac{t L^2}{d\gamma} \right)} \geq 1 - \delta$$

Compare this with 1 dim case

$$\text{Prob} \left[|\hat{\mu}_t - \mu| \leq \sqrt{\frac{2 \ln(2/\delta)}{t}} \right] \geq 1 - \delta$$

$$\text{Prob} \left[\frac{\|S_t\|}{t} \leq \sqrt{\frac{2 \ln(2/\delta)}{t}} \right] \geq 1 - \delta$$

$$\text{Prob} \left[S_t t^{-1} S_t^\top \leq 2 \ln \left(\frac{1}{\delta} \right) \right] \geq 1 - \delta$$

We know $\|\hat{\theta}_t - \theta_*\|_{V_t(\gamma)^{-1}} \leq \|S_t\|_{V_t(\gamma)^{-1}} + \gamma^{1/2} \|\theta_t\|$

Confidence interval

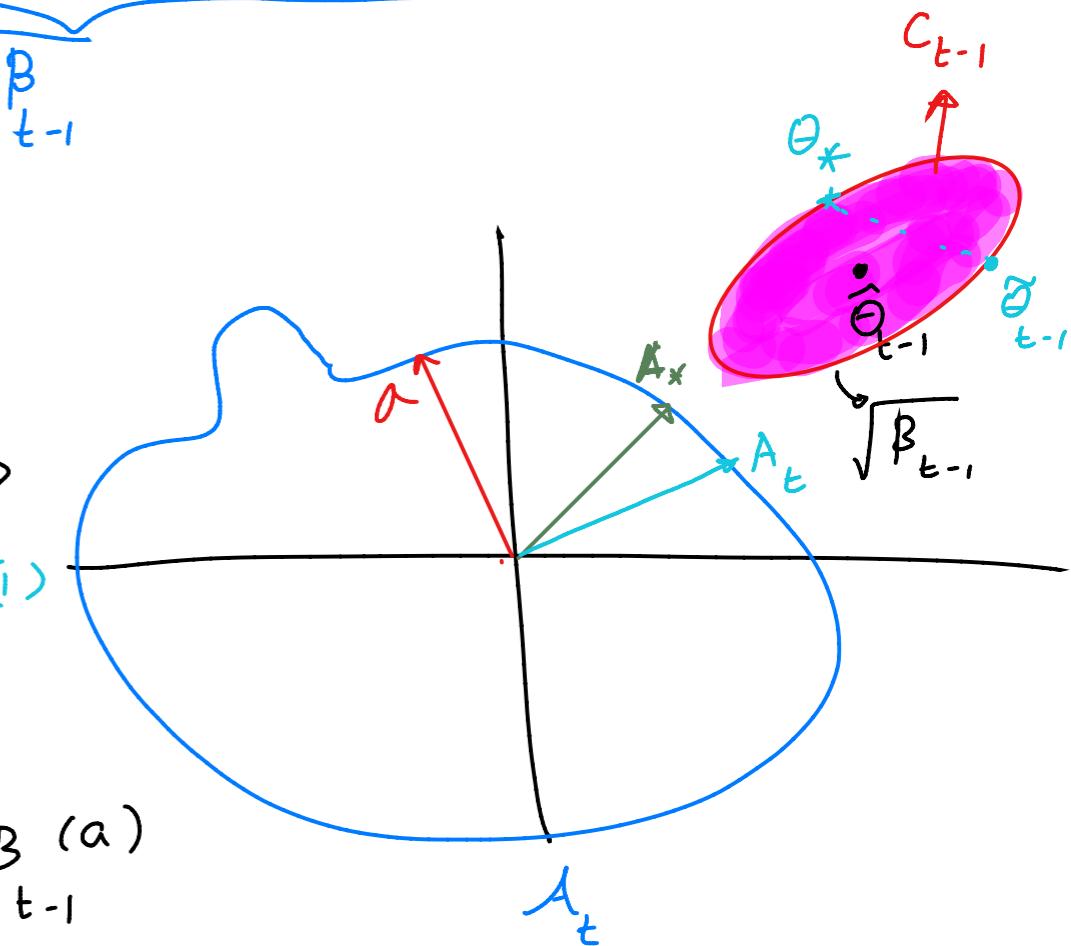
$$C_{t-1} = \left\{ \theta : \frac{\|\hat{\theta}_{t-1} - \theta\|}{\sqrt{V_{t-1}(\theta)}} \leq \sqrt{\gamma^{1/2} \|\theta^*\| + \sqrt{2 \ln(\frac{1}{\delta}) + d \ln\left(1 + \frac{t L^2}{\gamma d}\right)}} \right\}$$

$\sqrt{\gamma}$

B_{t-1}

Linear VCB

- $\forall a \in A_t$, $VCB_{t-1}(a) = \max_{\theta \in C_{t-1}} \langle a, \theta \rangle$ - (1)



- Play : $A_t = \operatorname{argmax}_{a \in A_t} VCB_{t-1}(a)$ - (2)

Best Arm : $A_t^* = \operatorname{argmax}_{a \in A_t} \langle a, \theta^* \rangle$

Regret Analysis: Cannot count as with independent arms case

$$r_t = \langle A_t^*, \theta_* \rangle - \langle A_t, \theta_* \rangle$$

↑
 Best

↑
 Picked by algo

$$\leq UCB_{t-1}(A_t^*) - \langle A_t, \theta_* \rangle$$

max calculation
 in (1) includes $\langle A_t^*, \theta_* \rangle$

$$\leq UCB_{t-1}(A_t) - \langle A_t, \theta_* \rangle$$

max calculation
 in (2) includes $UCB_{t-1}(A_t^*)$

$$= \langle A_t, \tilde{\theta}_{t-1} \rangle - \langle A_t, \theta_* \rangle$$

$$= \langle A_t, \tilde{\theta}_{t-1} - \theta_* \rangle$$

$$= A^\top (\tilde{\theta}_{t-1} - \theta_*)$$

Denote $\tilde{\theta}_{t-1} = \arg \max_{\theta \in C_{t-1}} \langle A_t, \theta \rangle \Rightarrow UCB_{t-1}(A_t) = \langle A_t, \tilde{\theta}_{t-1} \rangle$

$$\begin{aligned}
A_t^\top (\tilde{\theta}_{t-1} - \theta_*) &= A_t^\top V_{t-1}^{-1/2} V_{t-1}^{1/2} (\tilde{\theta}_{t-1} - \theta_*) \\
&\leq \|A_t^\top V_{t-1}^{-1/2}\| \|V_{t-1}^{1/2} (\tilde{\theta}_{t-1} - \theta_*)\| \\
&= \left(A_t^\top V_{t-1}^{-1/2} V_{t-1}^{-1/2} A_t \right)^{1/2} \left((\tilde{\theta}_{t-1} - \theta_*)^\top V_{t-1}^{1/2} V_{t-1}^{1/2} (\tilde{\theta}_{t-1} - \theta_*) \right)^{1/2} \\
&= \|A_t\|_{V_{t-1}^{-1}} \cdot \frac{\|\tilde{\theta}_{t-1} - \theta_*\|}{V_{t-1}^{1/2}}
\end{aligned}$$

$$r_t \leq \|A_t\|_{V_{t-1}^{-1}} \cdot 2 \sqrt{\beta_{t-1}}$$

Fact: β_t is an increasing sequence, w.l.o.g $1 \leq \beta_1 \leq \beta_2 \dots$

Assume:

- $\max_{t \in \mathbb{N}} \sup_{a, b \in A_t} |\langle a - b, \theta_* \rangle| \leq 1$ (Bounded rewards)

- $a \in \cup A_t$, $\|a\|_2 \leq L$

$$r_t \leq \|A_t\|_{V_{t-1}^{-1}} \cdot 2\sqrt{\beta_{t-1}}$$

(regret cannot exceed 2)

$$r_t \leq \min \left\{ 2\sqrt{\beta_{t-1}}, 2\sqrt{\beta_{t-1}} \|A_t\|_{V_{t-1}^{-1}} \right\}$$

$$= 2\sqrt{\beta_{t-1}} \min \left\{ 1, \|A_t\|_{V_{t-1}^{-1}} \right\}$$

$$\text{Regret}(n) = \sum_{t=1}^n r_t, \quad \bar{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \bar{r} = \begin{bmatrix} r_1 \\ \vdots \\ r_t \\ \vdots \\ r_n \end{bmatrix}$$

$$= \langle \bar{1}, \bar{r} \rangle$$

$$\leq \|\bar{1}\| \|\bar{r}\|$$

$$= \sqrt{n} \sqrt{\sum_{t=1}^n r_t^2}$$

$$= \sqrt{n \sum_{t=1}^n 4 \beta_{t-1} \min \left\{ 1, \frac{\|A_t\|^2}{V_{t-1}(x)} \right\}}$$

$$\leq \sqrt{4n \beta_n \sum_{t=1}^n \min \left\{ 1, \frac{\|A_t\|^2}{V_{t-1}(x)} \right\}}$$

We need to bound $\min \left\{ 1, \frac{\|A_t\|^2}{V_{t-1}(x)} \right\}$.

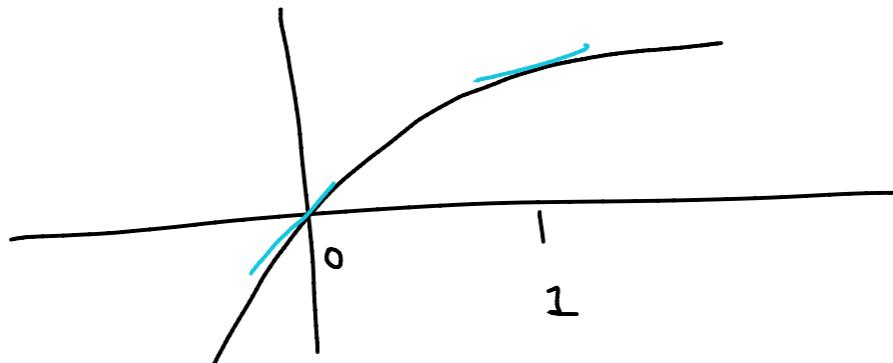
For any $0 \leq x \leq 1$

$$\min \{ 1, x \} \leq x$$

for any $x > 1$

$$\min \{ 1, x \} \leq x$$

Look at $\ln(1+x)$



slope at $x=0$

$$\frac{d}{dx} \ln(1+x) = \frac{1}{1+x} = 1$$

slope at $x=1$ is $\frac{1}{2}$

$$\begin{aligned}
 x &= \int_0^x du = 2 \int_0^x \frac{1}{2} du \\
 &\leq 2 \int_0^x d(\ln(1+u)) \\
 &= 2 [\ln(1+u)]_0^x \\
 &= 2 \ln(1+x)
 \end{aligned}$$

$$\min \{1, x\} \leq 2 \ln(1+x)$$

$$\sum_{t=1}^n \min \left\{ 1, \frac{\|A_t\|^2}{V_{t-1}(\gamma)^{-1}} \right\} \leq 2 \sum_{t=1}^n \ln \left(1 + \frac{\|A_t\|^2}{V_{t-1}(\gamma)^{-1}} \right)$$

$$\begin{aligned}
 V(\gamma) &= \gamma I + V_t = \gamma I + \sum_{s=1}^t A_s A_s^T \\
 &= \gamma I + \sum_{s=1}^{t-1} A_s A_s^T + A_t A_t^T \\
 &= V_{t-1}(\gamma) + A_t A_t^T
 \end{aligned}$$

$$V_t(\gamma) = V_{t-1}(\gamma) + A_t A_t^T$$

$$= V_{t-1}^{1/2} \left[I + V_{t-1}^{-1/2} A_t A_t^T V_{t-1}^{-1/2} \right] V_{t-1}^{1/2}$$

we use $\det(AB) = \det(A) \det(B)$

$$\det(V_t(\gamma)) = \det(V_{t-1}(\gamma)) \det(I + V_{t-1}^{-1/2} A_t A_t^T V_{t-1}^{-1/2})$$

$\underbrace{\qquad\qquad\qquad}_{M}$

$$M = I + yy^T, \quad y = V_{t-1}^{-1/2} A_t$$

\det is product of eigenvalues

• one of eigenvector of M is y itself

$$\begin{aligned} My &= (I + yy^T)y = y + y \|y\|_2^2 \\ &= (1 + \|y\|_2^2) y \end{aligned}$$

- pick any $x \in \mathbb{R}^d$

$$Mx = (I + yy^T)x = x$$

$M = I + yy^T$ has

- one eigenvalue $= \|y\|_2^2$
- other $d-1$ eigenvalues $= 1$

$$\det(M) = \|y\|_2^2$$

For our case $y = V_{t-1}(x) A_t^{-1/2}$

$$y^T y = A_t^T V_{t-1}(x)^{-1} V_{t-1}(x) A_t$$

$$= \|A_t\|_F^2 - 1$$

$$\det(V_n(\gamma)) = \det(V_{n-1}(\gamma)) \left(1 + \frac{\|A_n\|^2}{V_{n-1}(\gamma)^{-1}}\right)$$

⋮

$$= \det(\gamma I) \prod_{t=1}^n \left(1 + \frac{\|A_t\|^2}{V_{t-1}(\gamma)^{-1}}\right)$$

$$\sum_{t=1}^n \ln \left(1 + \frac{\|A_t\|^2}{V_{t-1}(\gamma)^{-1}}\right) = \ln \prod_{t=1}^n \left(1 + \frac{\|A_t\|^2}{V_{t-1}(\gamma)^{-1}}\right)$$

$$= \ln \left(\frac{\det(V_n(\gamma))}{\det(\gamma I)} \right)$$

$$\leq d \ln \left(1 + \frac{nL^2}{\gamma d}\right)$$

(same as
yesterdays
calculation)

Overall regret expression

$$\text{Regret}(n) \leq \sqrt{4n\beta_n \sum_{t=1}^n \min \left\{ 1, \|A_t\|^2 V_{t-1}^{-1} \right\}}$$

$$\leq \sqrt{4n\beta_n d \ln \left(1 + \frac{nL^2}{\delta d} \right)}$$

$$= \sqrt{8n\beta_n d \ln \left(1 + \frac{nL^2}{\delta d} \right)}$$

$$= \sqrt{8n} \sqrt{\beta_n} \sqrt{d \ln \left(1 + \frac{nL^2}{\delta d} \right)}$$

$$\sqrt{\beta_n} = \gamma^n \|A_*\| + \sqrt{2 \ln(1/\delta) + d \ln \left(1 + \frac{nL^2}{\delta d} \right)}$$

writing down dominant terms

$$\text{Regret}(n) \leq C \sqrt{n} d \ln(hL) \quad \text{as } n \rightarrow \infty, \frac{\text{Regret}(n)}{n} \rightarrow 0$$

