AI Theory

Advanced Search Problems :

Intuition :

in real world, there are many uncertainties.

state based models tend to ignore this uncertainties.

how do you account for uncertainties?

This leads to next type of models.

Markov Models:

Applications: Robotics, Agriculture, Resource management etc.

notice one thing: all have real-world $\underbrace{\text{uncertainties}}$ pegged to it.

short form of "probability"

Let's proceed with Markov Models..

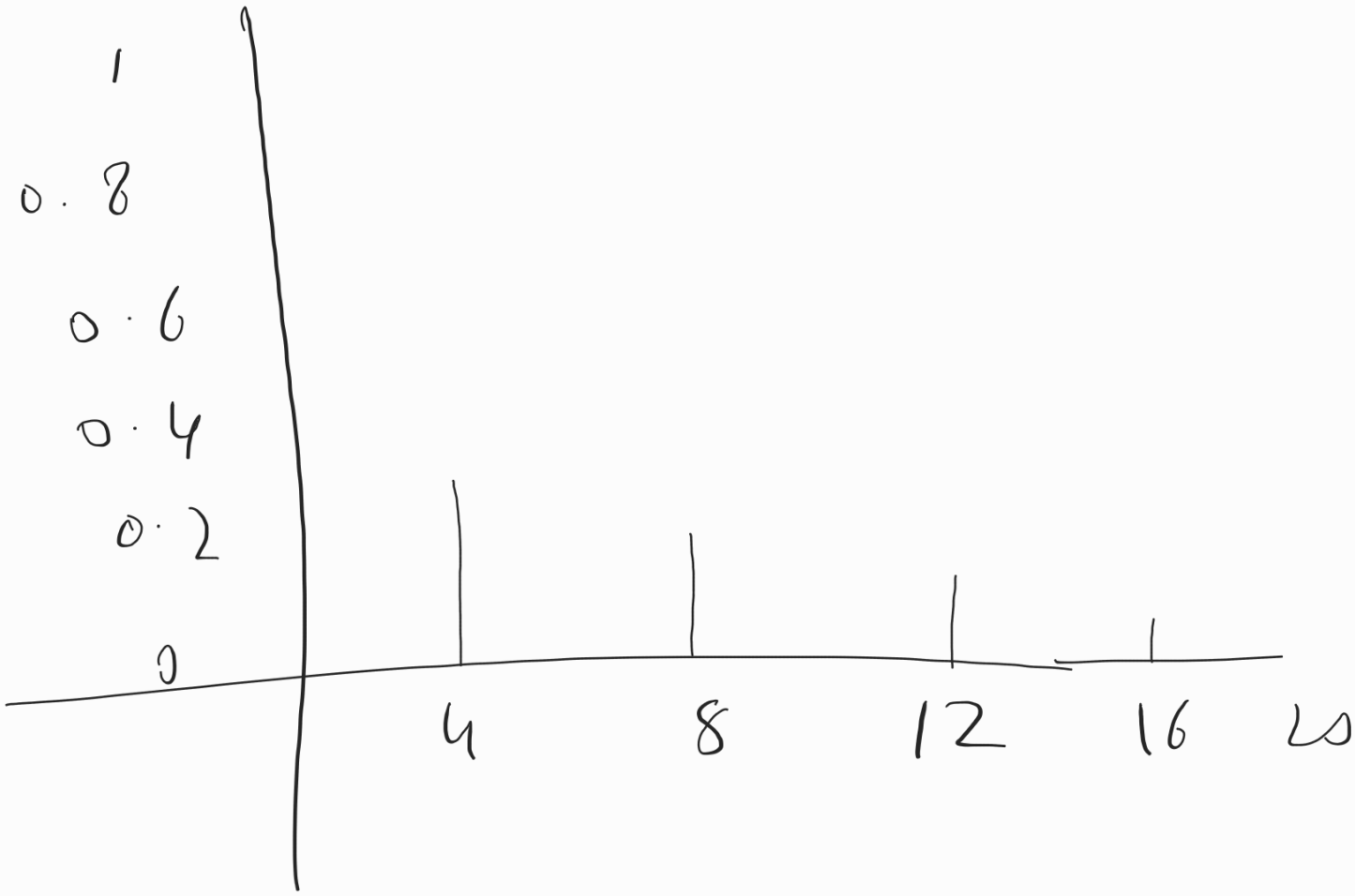
Ex: Dice game

for each round 1, 2, 3, ...

- You choose Stay / Quit.
  - Quit = 10 \$
  - Stay = 4 \$, roll a dice
    - ↳ if 1 or 2, end
    - ↳ continue

What is the probability of getting 4 $ by staying?

Reward Chart:



Expected Utility

$$E[\text{util}] = \frac{1}{3}(4) + \frac{1}{3}\times\frac{2}{3}(8) + \frac{2}{3}\times\frac{2}{3}\times\frac{1}{3}\ldots(12)$$

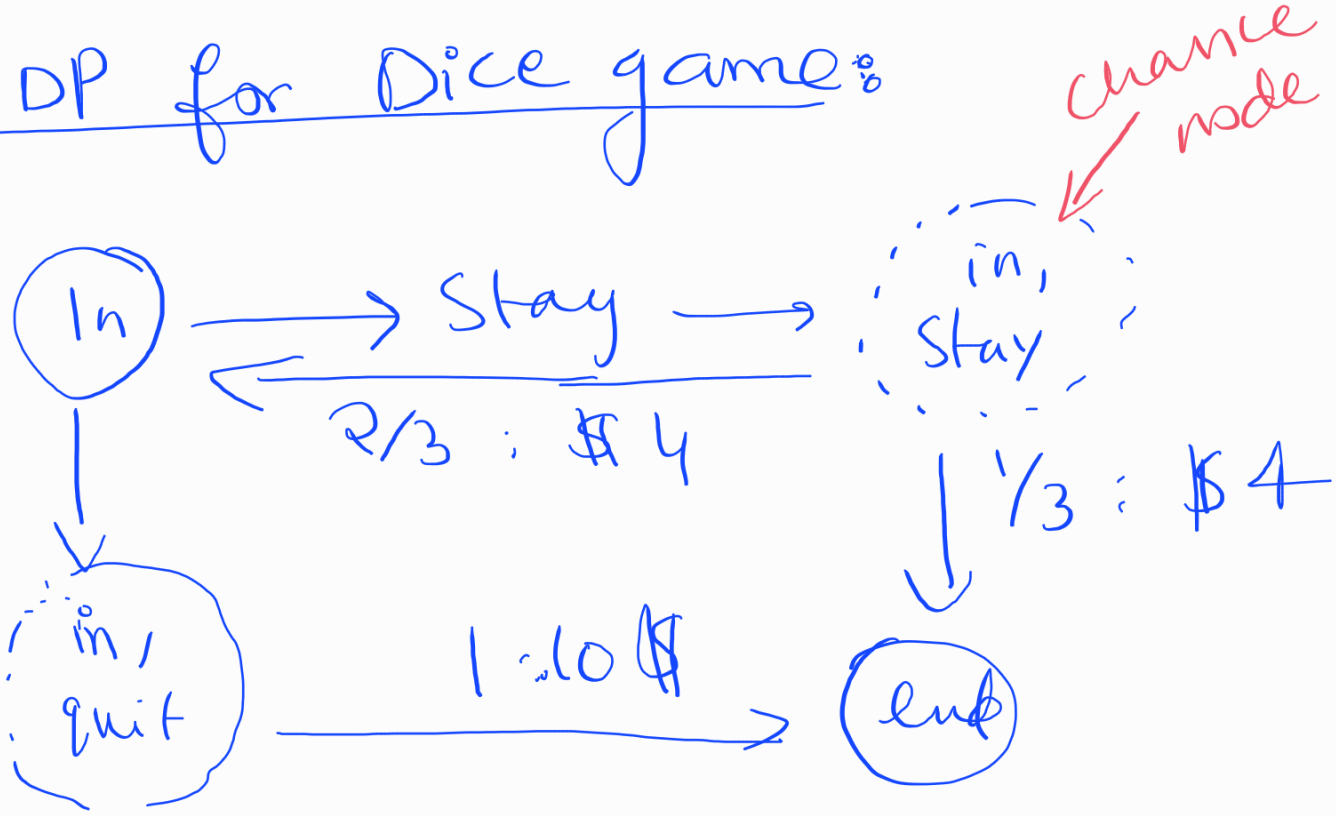
$$= 12$$

Lets say we have a random variable:

random variable	weightage / probability
5	1/4
6	1/4
9	1/2

$$\boxed{\text{Expectation of random variable} = \sum_{i=1}^{n} \text{Value}_i + P(\text{val})_i}$$

MDP for Dice game:



Definition of Markov Processes :

States : set of states

$S_{start} \in$ states : start

Actions (s)

$T(s, a, s')$: $P(s')$ taken s

$R(s, a, s')$: Reward of transition

isGoal(s) : . . .

Discount factor $0 \le \gamma \le 1$

Transition Probability :

Probability of s' provided that s occurs.

s	a	s'	$T(s,a,s')$
in	quit	end	1
in	stay	in	2/3
in	stay	end	1/3

Probabilities sum to one

↳ for each state s and act a:

$$\sum_{s' \in \text{states}} T(s,a,s') = 1$$

in, quit
in, stay etc

Successors:

s' such that $T(s,a,s') > 0$

↓

let's say we went from in, quit,

in.

You can't go from quit

to in. So $P(\text{in}) = 0$.

this one

Transportation Example:

Same example.

walk s+1, 1 min

Tram 2s, 2 min

Tram fails with Probability 0.5.

Solution in MDP:

Policy:

A policy π is a mapping from each state $s \in$ states to an action $a \in$ Actions(s).

Basically: π is the action at that state.

Policy Evaluation:

Following a policy yields <u>random path</u>.

The utility of a policy is the (discounted) sum of rewards on the path (this is the random path).

Path	(s, a, s')	Utility
[in, 4, end]	for each epoch	4
[in, 4, in, 4, in, 4, end]		12
[in, 4, in, 4, end]		8

The value of the policy is the expected utility.

How do we calculate value for

the above policy?

first, calculate probability of these utilities over all epochs. $\rightarrow \mathbb{P}(u) = \frac{\text{COUNT}(u)}{\text{Total vals}}$

Then find E(utility).

Discounting:

Path: $s_0, a_1, r_1, s_1, a_2, r_2, s_2, \ldots$

(action, reward, new state)

Utility with discount γ:

$$u_1 = r_1 + \gamma r_2 + \gamma^2 r_3 + \gamma^4 r_4 \ldots$$

Discount $\gamma = 1$

(save for future) [stay, stay, stay, stay]

$= 16$

<u>Discount $\gamma = 0$</u>

(live in the moment) "

$= 4 + 0 \cdot (4 + \ldots)$

$= 4$

Discount $\gamma = 0.5$ (balanced life)

" $= 4 + 1 \cdot 4 + \ldots 4 + \ldots 4 = 7.5$

<u>Example</u>

	time	val	
10	t_0	10	$(0.5)^0$
5	t_1	10	$(0.5)^1$
2.5	t_2	10	$(0.5)^2$

1.25 $\quad t_3 \quad 10 \quad (0.5)^3$

As we are going into the future, the reward worth is being lessened.

Utility given above with discount factor $\gamma = 0.8$.

$$4 \times (0.8)^0 = 4$$

$$4(0.8)^0 + 4(0.8)^1 + 4(0.8)^2$$

$$= 11.808$$

we thought of getting 12, but we got 11.808.

The purpose of this discount is to partially compensate for future ups/downs in the states.

<u>Value</u>:

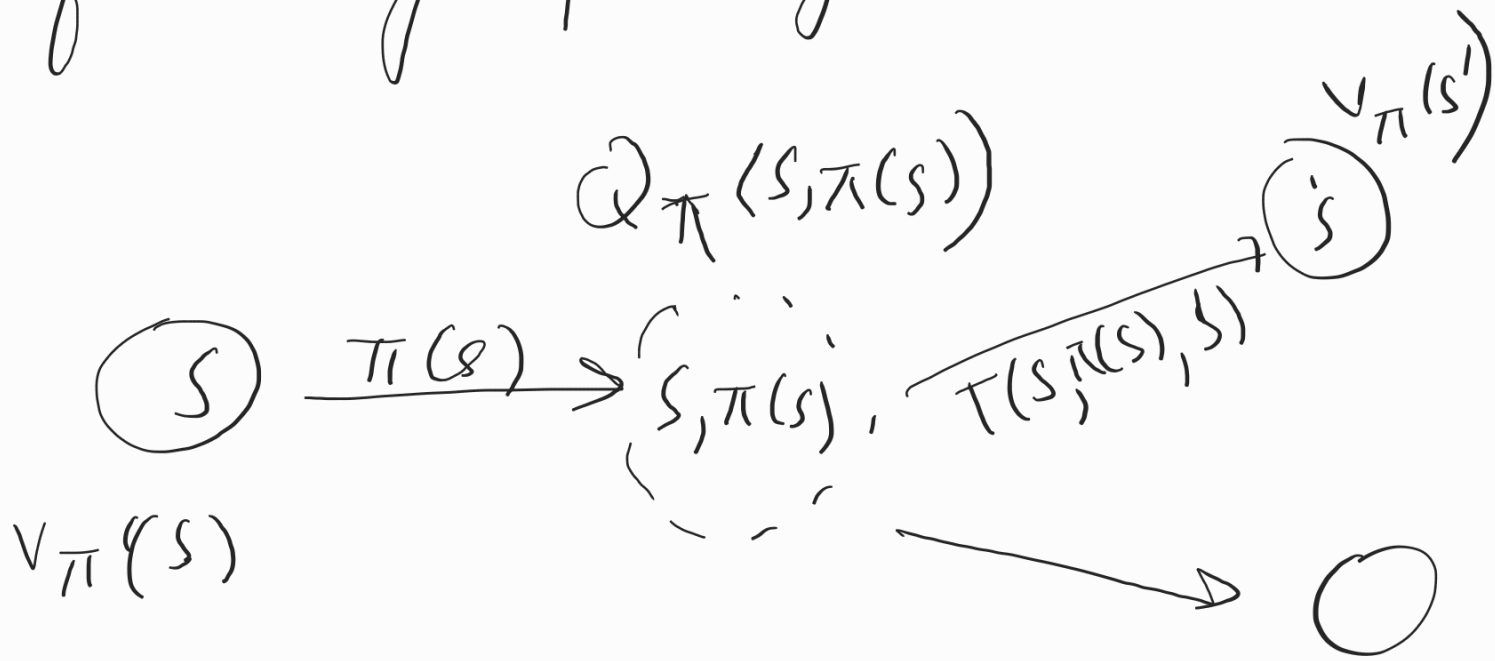
↳ expected utility

Let $V_\pi(s)$ be the expected utility taken from policy π by state s.

<u>Q - value of the policy</u>:

Let $Q_\pi(s,a)$ be the expected

utility of taking action "a" from state s, and then following policy π.



Q-value is staying on that node and getting expected utility of the future.

$$V_\pi(s) = \begin{cases} 0, & \text{Goal} \\ Q_\pi(s, \pi(s)) & \text{otherwise} \end{cases}$$

$$Q_{\pi}(s, a) = \sum_{s'} T(s, a, s') \left[R(s, a, s') + \gamma V_{\pi}(s') \right]$$

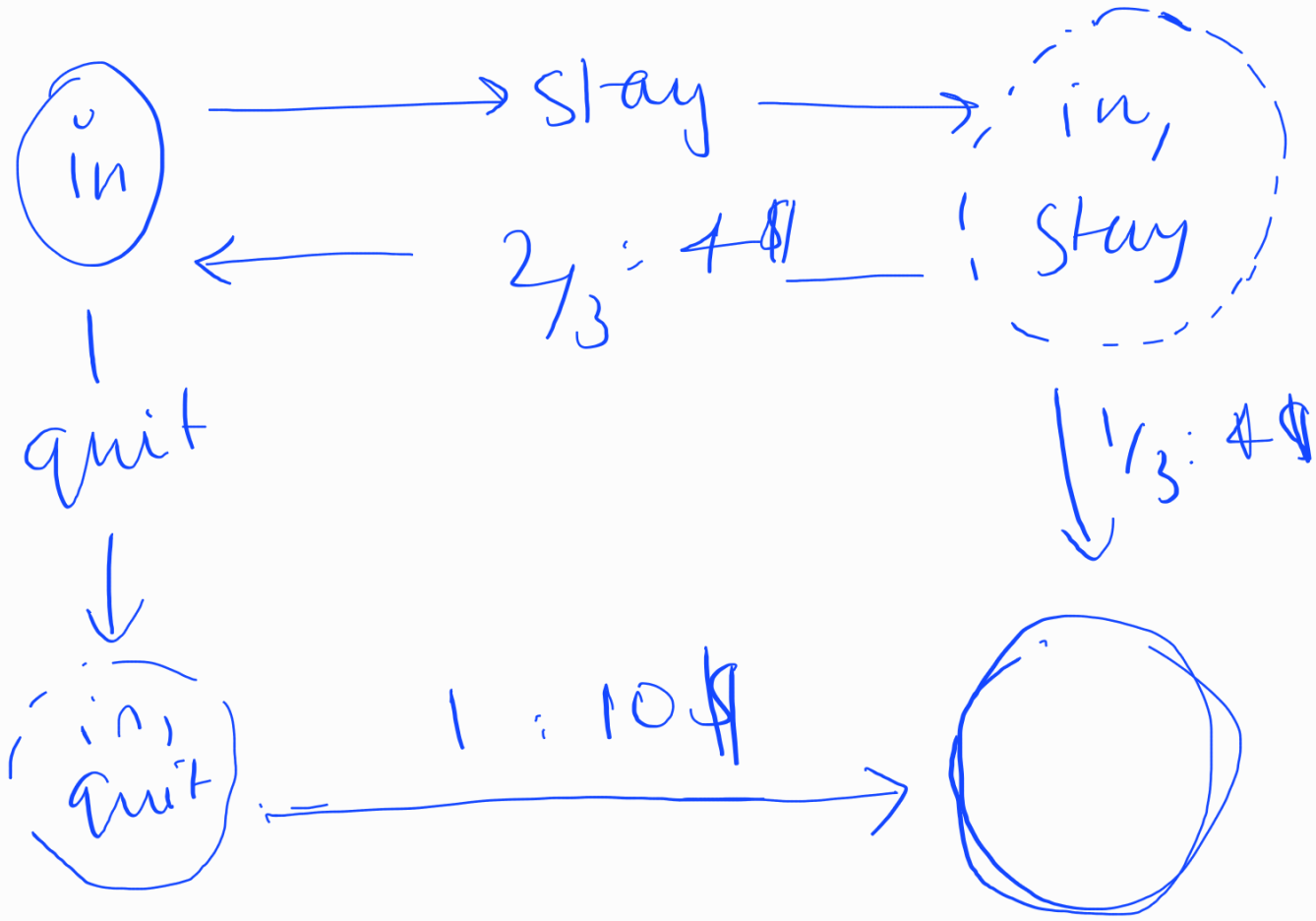
$$= 0.25(4 + 0.9 \times 15)$$

$$+ 0.75(10 + 0.9 \times 20)$$

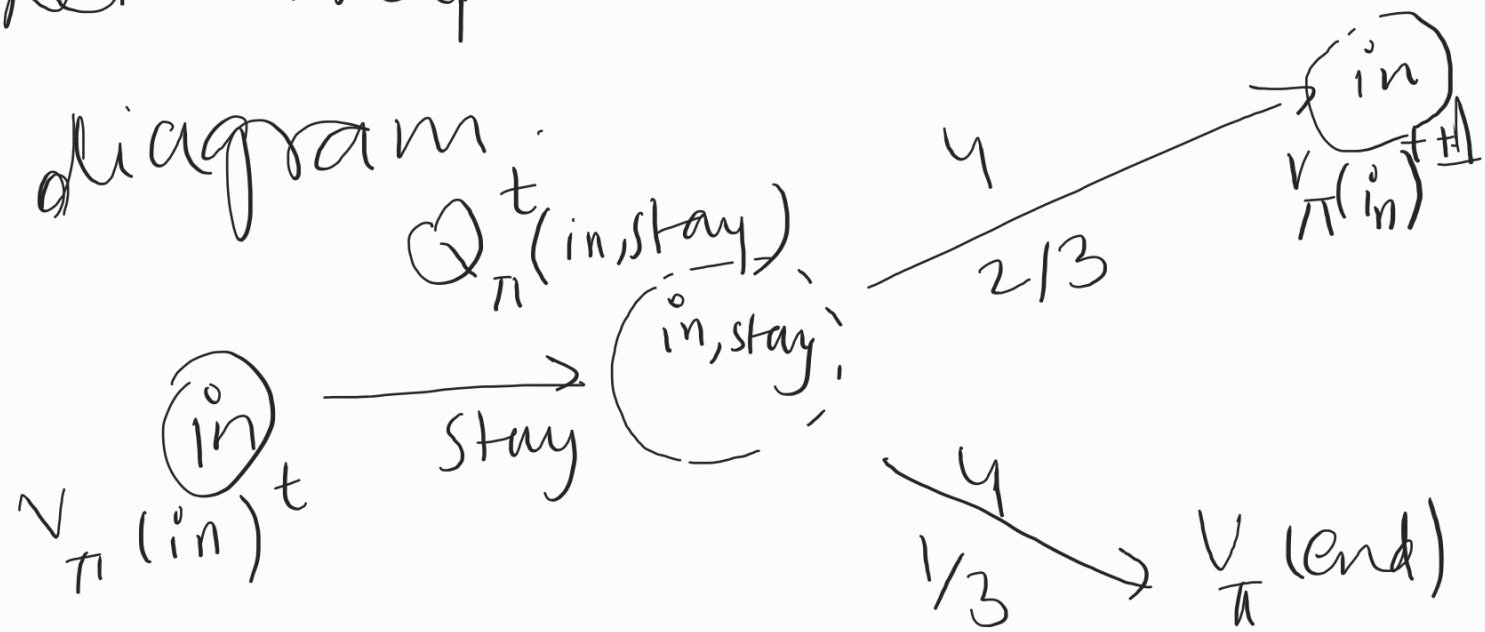
$$\boxed{Q_{\pi}(s, a) = 25.375}$$

we would give more value to this because tran. prob. is greater.

MDP of Dice game:



Let's map it to a node-diagram:



$\gamma = 1$

$$V_\pi(in)^t = \frac{2}{3}\left[4 + 1 \times V_\pi(in)^{t+1}\right]$$

$$+$$

$$\frac{1}{3}\left[4 + 1 \times \underbrace{V_\pi(end)}_{0}\right]$$

$$V_\pi(in) = \frac{2}{3} \times 4 + \frac{2}{3} \times V_\pi(in)$$

$$+ \frac{1}{3}(4) + 0$$

$$= \frac{8}{3} + \frac{2}{3} V_\pi(in) + \frac{4}{3}$$

$$= \frac{8+4}{3} + \frac{2}{3} V_\pi(in)$$

$$\boxed{V_\pi(in) = 4 + \frac{2}{3} V_\pi(in)}$$

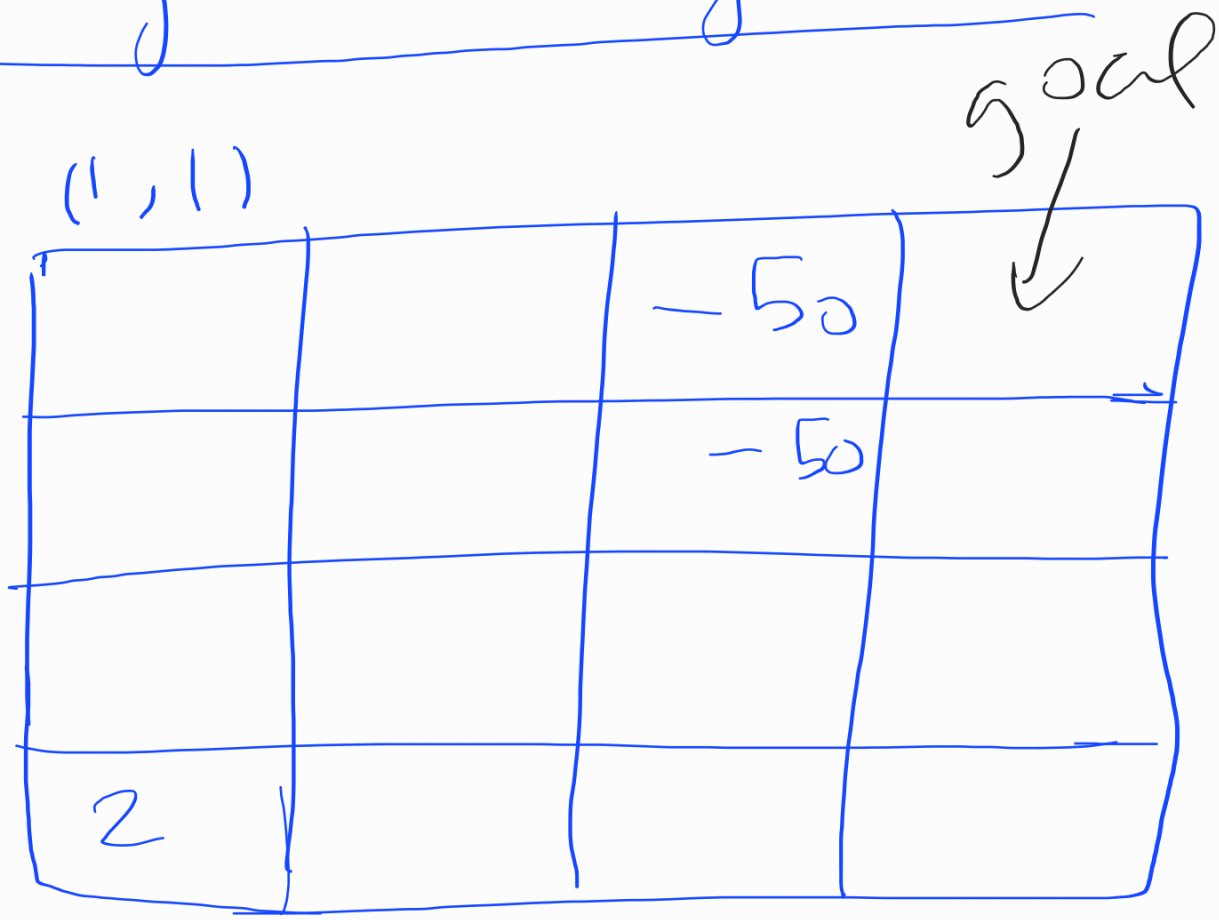
This is the closed form solution directly solving the problem.

$$\frac{1}{3} V_{\pi}(in) = 4$$

$$\boxed{V_{\pi}(in) = 12}$$

for a single action, $V_{\pi} = Q_{\pi}$.

Bridge Crossing Model



all others have reward -0.1.

functions:

states

actions(state)

TransitionProb(s, a, s')

Transition (s, a)

reward (s, a, s')

start_state

IsGoal (state)

Task: Take slip probability as input, and near the squares of -50, you will end the model. Else take deterministic actions.