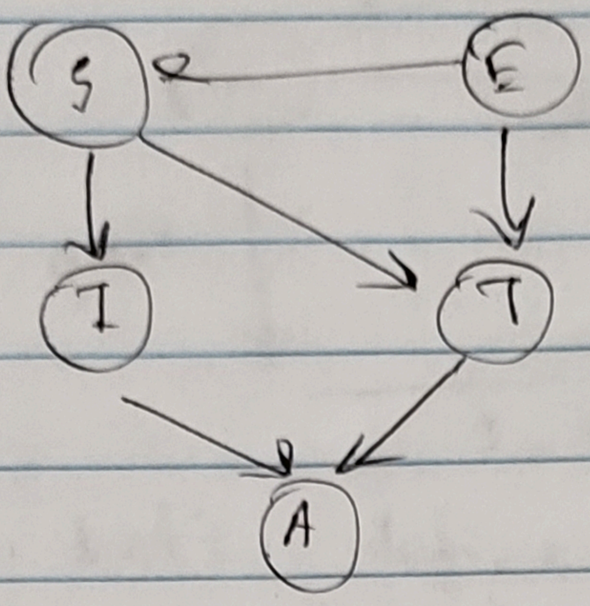
Bayesian Network



Note: All binary variables

We want to compute $P(+A, +T)$.

So, the free variables are E, S, I.

For variable elimination, which of the following orders are easy to compute?

① E → S → I

$$P(+A,+T) = \sum_E P(E) \sum_S P(S|E) P(+T|S,E) \sum_I P(I|S) P(+A|I,+T)$$

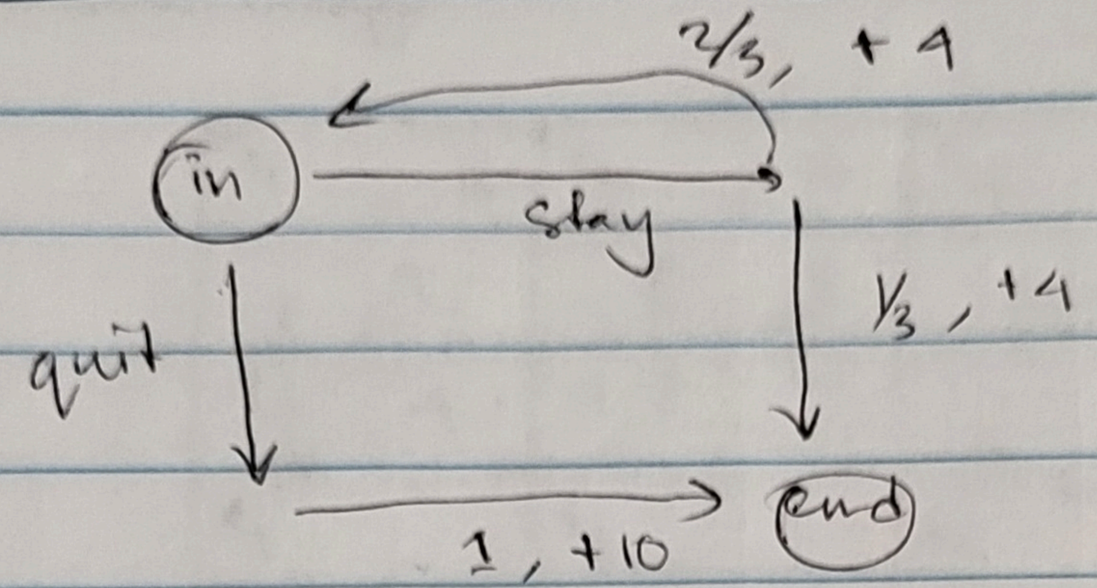
② I → E → S

$$P(+A,+T) = \sum_I P(+A|I,+T) \sum_E P(E) \sum_S P(I|S) P(S|E) P(+T|S,E)$$

Out of these two, ② has three terms and more variables in the last part.

So, ① seems easier.

MDP (Dice Game)



Let's define this MDP in tabular form.

s	a	s'	T(s,a,s')	R(s,a,s')
in	stay	in	2/3	+4
in	stay	end	1/3	+4
in	quit	end	1	+10

* No rows for s = end (why?)

* $V'_{in} = V'_{end} = 0$, $\gamma = 0.5$

Let's do VALUE ITERATION

iteration 1:

$Q'(in, stay) = \frac{1}{3}[+4 + \gamma.0] + \frac{2}{3}[+4 + \gamma.0]$
$= +4$

$Q'(in, end) = \cancel{\frac{1}{3}}\ 1.\ 1.[+10 + \gamma.0] = +10$

$V^2_{\pi}(in) = \max_a Q'(in, a) = +10$

$\pi^2(in) = \arg\max_a Q'(in, a) = \underline{end}$

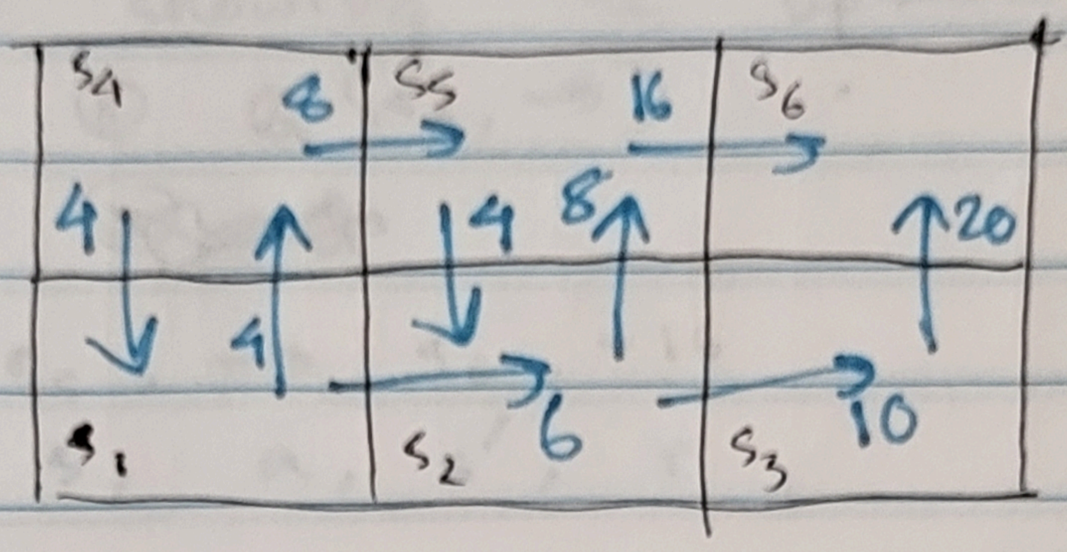
iteration 2:

$Q^2(in, stay) = \frac{1}{3}[+4 + 0.5 \times 0] + \frac{2}{3}[+4 + 0.5 \times 4]$
$= 16/3$

$Q^2(in, end) = 1.(+10 + 0) = +10$

$V^3(in) = 10$, $\pi^3_{\pi}(in) = end$

Q - learning



Arrows indicate Q-values (after a few iterations)

Let's list all of them.

$Q(s_1, \uparrow) = 4$, $Q(s_1, \rightarrow) = 6$

$Q(s_2, \uparrow) = 8$, $Q(s_2, \rightarrow) = 10$

$Q(s_3, \uparrow) = 20$

$Q(s_4, \rightarrow) = 8$, $Q(s_4, \downarrow) = 4$

$Q(s_5, \rightarrow) = 16$, $Q(s_5, \downarrow) = 4$

Now, assume the following episode.

$s_1, \uparrow, s_4, \rightarrow, s_5, \rightarrow, s_6$

$(r=0)$ $(r=0)$ $(r=+16)$

Update Q-values. $(\gamma = 0.3, \alpha = 0.8)$

s_1, a, s', r

① $(s_1, \uparrow, s_4, 0)$

$$\text{sample} = r + \gamma \max_{a'} Q(s', a')$$

$$= 0 + 0.3 \times 8 = 2.4$$

$$Q(s_1, \uparrow) = (1-\alpha) Q(s_1, \uparrow) + \alpha \cdot \text{sample}$$

$$= 0.2 \times 4 + 0.8 \times 2.4$$

$$= 2.72.$$

Not showing the updates for

(ii) $Q(s_4, \rightarrow)$.

~~(iii)~~

(iii) $s_5, \rightarrow, s_6, +16$

s, a, s', r

$$\text{sample} = r + \gamma \max_{a'} Q(s', a')$$
$$= +16 + 0.3 \times 0$$
$$= +16$$

$$Q(s_5, \rightarrow) = 0.2 \times 16 + 0.8 \times 16$$
$$= 16$$