Bike MDP

* State :- State at time 't' denoted via 'S_t'

$$S_t = [b_t^1 \; b_t^2]$$

S_t → Denotes No. of bikes available at the end of the day at particular time 't'

b_t^i → No. of bycycle at station 'i' at time 't'

* Action state $A([b^1, b^2])$

Considering the Notation -
- Transferring bicycle from stations $1 \to 2$ = +ve
- Transferring bicycle from stations $2 \to 1$ = -ve

eg 1 Consider the state = [2, 6]
- Here we can transfer 5 bicycles from station 1 to 2
- Also we can transfer 2 bicycles from stations 2 to 1

So Action state = [-5..., 2]

eg 2 Consider the state [10, 18]
- Here we can transfer a maximum of 2 bicycles from station $1 \to 2$ as for any station maximum capacity is limited to 20.
- Also we can transfer 5 (max transfer limit) bicycles from $2 \to 1$

So Action state = [-5, ..., 2]

from $2 \to 1$ ← → from $1 \to 2$

So Generally,
Action state $A([b^1, b^2]) =$

$\{-\min(\min(5, b_t^2), 20 - b_t^2), \min(\min(5, b_t^1), 20 - b_t^1)\}$

* Value function :- It maps states to real values.

$$V_\pi : S \longrightarrow R$$

$$|S| = 21 \times 21 \quad (\text{Size of State})$$

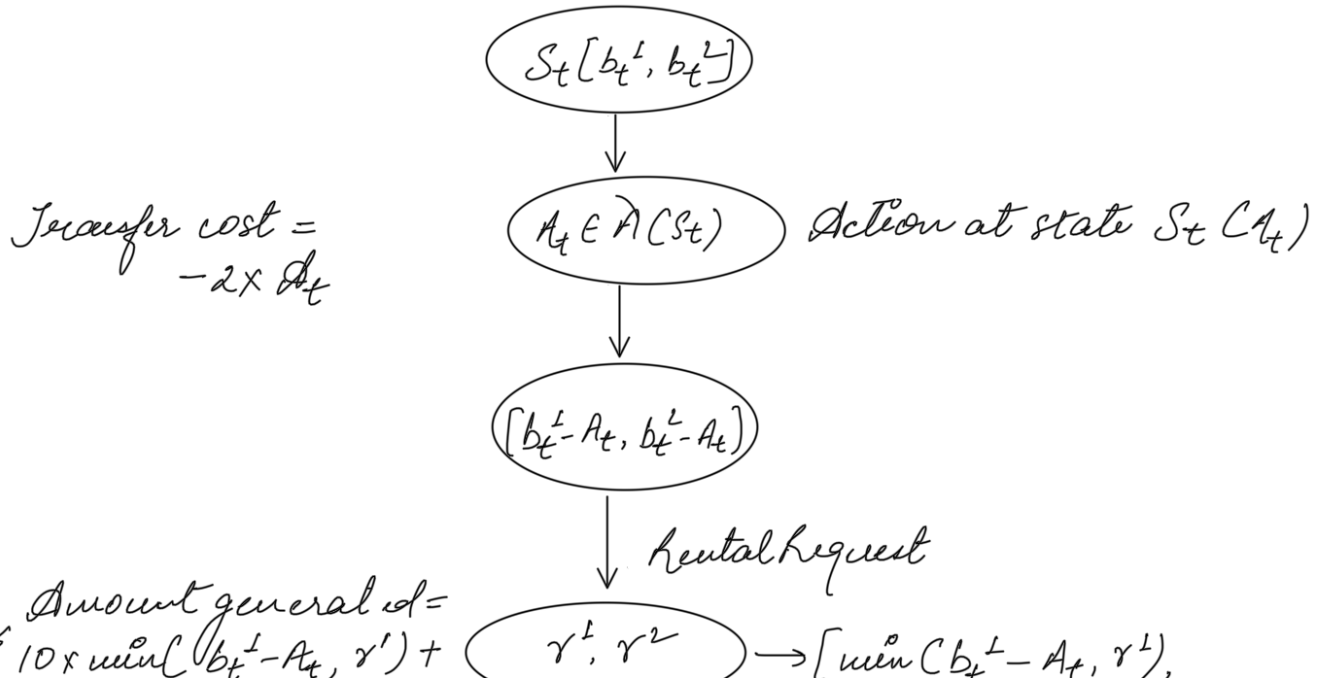
* Action Value Function -

$$Q_\pi : S \times A \longrightarrow R$$

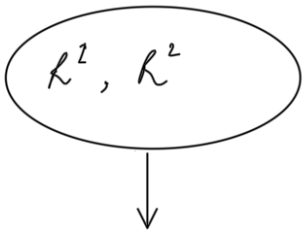
* Transition Probability -

$$P(S_{t+1}, R_{t+1} \mid S_t, A_t)$$

It means that at the state S_t taking action A_t we reach a new state S_{t+1} and received reward R_{t+1}



$10 \times \min(b_t^2 + A_t, r^2)\}$ $\min(b_t^2 + A_t, r^2)]$



$$S_{t+1} = [\, b_{t+1}^1 = \min(b_t^1 - A_t - \min(b_t^1 - A_t, r^1) + R^1, 20)$$
$$b_{t+1}^2 = \min(b_t^2 + A_t - \min(b_t^2 + A_t, r^2) + R^2, 20)\,]$$

$$P(S_{t+1} \mid S_t, A_t) = P(r^1)\,P(r^2)\,P(R^1)\,P(R^2)$$

→ Probability with which we reach from state S_t to S_{t+1} taking action A_t

Probability of receiving rental request r^1 & r^2

$$P(r^1, r^2) = P(r^1)\,P(r^2) \rightarrow \text{Independent.}$$

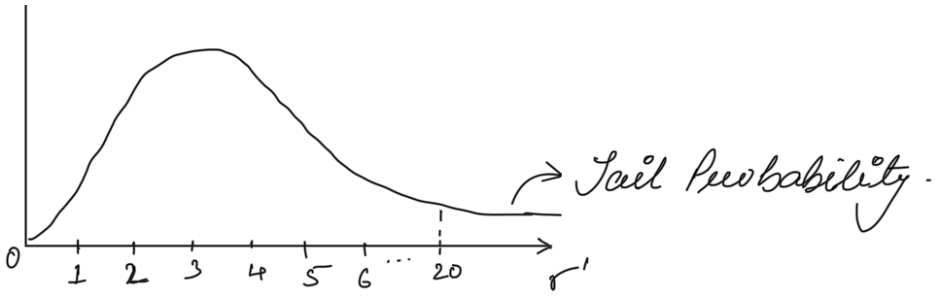
$$= \frac{\lambda_1^{r^1} \cdot e^{-\lambda_1}}{r^1!} \cdot \frac{\lambda_2^{r^2} \cdot e^{-\lambda_2}}{r^2!}$$

Similarly probability of return request will be same as that of rental request with different λ values.

* Reward at S_{t+1}

$$R_{t+1} = \{-2 \times A_t + 10 \times \min(b_t^1 - A_t, r^1) + 10 \times \min(b_t^2 + A_t, r^2)\}$$

$P(r^1)$ ↑



To get a meaningful distribution we divide the tail probability to each values from 0 ... 20.