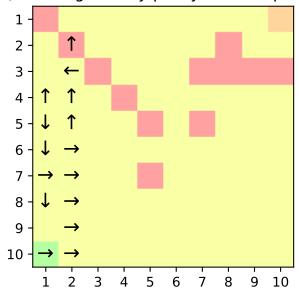
Q-learning Greedy policy at 0 th episode



Q-learning Greedy policy at 10000 th episode

Q-learning Greedy policy at 20000 th episode

1 -	T	\rightarrow	T							
2 -	1	1	1	\rightarrow	1	1	→	1	1	1
3 -	\rightarrow	←	1	\rightarrow	1	1	1	1	1	1
4 -	1	1	\rightarrow	1	\rightarrow	1	←	←	←	←
5 -	↓	1	\downarrow	1	1	1	1	1	←	←
6 -	1	\rightarrow	\rightarrow	\rightarrow	\rightarrow	1	←	←	←	←
7 -	\rightarrow	\rightarrow	\rightarrow	1	1	1	1	←	1	1
							1			
							1			
10 -	\rightarrow	\rightarrow	\rightarrow	\rightarrow			←	←	1	←
	1	2	3	4	5	6	7	8	9	10