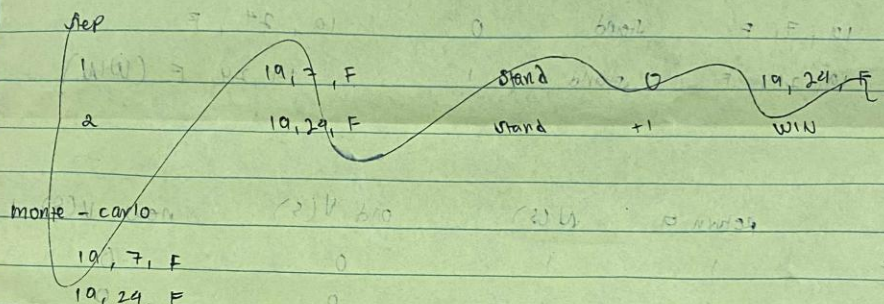
	step	state (P, D, A)	Action	reward	Next State
Ep	1	19, 7, F	stand	0	19, 24, F
		19, 24, F	Stand	+1	win
		20, 20, F	Stand	0	20, 20, F
	2	20, 20, F	Stand	0	TIE
		17, 2, F	Stand	0	17, 23, F
	3	17, 23, F	stand	+1	win



4	13, 9, F	HIT	0	23, 9, F
	21, 19, F	stand	1	21, 19, F (WIN)
5	8, 5, F	HIT	0	18, 5, F
	18, 20, F	stand	-1	18, 20, F (LOSE) BUST
6	15, 10, F	HIT	0	21, 10, F
	21, 26, F	stand	1	21, 26, F (WIN)
7.	16, 3, F	Hit	0	26, 3, F
	26, 3, F	stand	-1	26, 23, F (Bust)
8.	12, 5, F	HIT	0	19, 5, F
	19, 5, F	stand	0 1	19, 18, F (WIN)

9.	21, 10, F	stand	0	21, 10, F
	21, 10, F	stand	1	21, 10, F (WIN)
10.	18, 11, F	stand	0	18, 11, F
	18, 11, F	stand	1	18, 23, F (WIN)

step

1	10, 7, F	stand	0	10, 24, F
2	10, 24, F	stand	1	10, 24, F (WIN)

monte carlo

State	return G	N(s)	old V(s)	new V(s)
10, 7, F	1	1	0	0
10, 24, F	1	1	0	0

step 1

a. state (10, 7, F)

b. Number of visits : N(10, 7 F) = 0

c. V(10, 7, F) = 0

d. increment visit count = 1

e. update value func : V(17, 6, False) = $0 + \frac{1}{1}(1-0) = 1$

step 2

a. State = (10, 24, F)

b. number of visits N(10, 24, F) = 0

c. V(10, 24, F) = 0

d. increment visit count : N(10, 24, F) = 1

e. update V_func V(10, 24, F) = $0 + \frac{1}{1}(1-0) = 1$

State	return G	N(s)	old V(s)	new V(s)
10, 7, F	+1	1	0	1
10, 24, F	+1	1	0	1

sample Temporal Difference (0) update

a. state = (19, 7, F)

b. Reward r = 0,

c. next state = (19, 24, F)

d. update: $V(19,7,F) = 0 + 0.5(0 + V(19,24,F) - 0) = 0$

step 2. (update next state immediately using

a. state = (19, 24, F)

b. Reward = r = +1

c. Next state = WIN

d. update Vfunc: $V(19,24,F) = 0 + 0.5(1 + 0 - 0) = +0.5$

	state	reward	n-state	old V(s)	new V(s)
step 1	(19, 7, F)	0	(19, 24, F)	0	0
2	(19, 24, F)	+1	WIN	0	+0.5

Step 3.

~~update~~ all previously visited state

a. state = (19, 7, F)

b. Reward r = +1

c. next state = (19, 24, F)

d. update Vfunc: V(19,7,F)

$V(19,7,F) = 0 + 0.5(1 + 0.5 - 0) = 0.25$

Step					
1	19, 7, F	0	19, 24, F	0	0.25
2	19, 24, F	+1	WIN	0	0.5
		+1	1	0	1
19, 24, F		+1	1	0	1