Problem 1.

Solution:

1. State space: $X = \{-5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5\}$ - your advantage at the moment

   Control space: $U = \{0\ (timid), 1\ (bold)\}$

   Transition probabilities:

$$P\{x_{t+1} = x_t - 1 |\, x_t, 0\} = 0.1,$$
$$P\{x_{t+1} = x_t \,|\, x_t, 0\} = 0.8,$$
$$P\{x_{t+1} = x_t + 1 |\, x_t, 0\} = 0.1$$
$$P\{x_{t+1} = x_t - 1 |\, x_t, 1\} = 0.55,$$
$$P\{x_{t+1} = x_t + 1 |\, x_t, 1\} = 0.45$$

   For all other states $y \in X$, we have $P\{x_{t+1} = y |\, x_t, u_t\} = 0$;

   Reward function for one period: $r_t(x, u) = 0$.

   Dynamic programming equations: The value function $v_t^*(x)$ expresses the probability to win at time t if the state (score) is x.

$$v_t^*(x) = \max\begin{Bmatrix} 0.8v_{t+1}^*(x) + 0.1v_{t+1}^*(x-1) + 0.1v_{t+1}^*(x+1), \\ 0.45v_{t+1}^*(x+1) + 0.55v_{t+1}^*(x-1) \end{Bmatrix},$$
$$t = 1, \dots, 5$$
$$for\ t = 1, x = 0;$$
$$t = 2, x = -1, 0, 1;$$
$$t = 3, x = -2, -1, 0, 1, 2;$$
$$t = 4, x = -3, -2, -1, 0, 1, 2, 3;$$
$$t = 5, x = -4, -3, -2, -1, 0, 1, 2, 3, 4;$$

$$v_6^*(x) = \begin{cases} 1, & if\ x > 0 \\ 0, & if\ x < 0 \\ 0.45, & if\ x = 0, u = 1 \\ 0.50, & if\ x = 0, u = 0 \end{cases}$$

$$v_t^*(x) = \begin{cases} 0.45, & if\ u = 1 \\ 0.50, & if\ u = 0 \end{cases}, \qquad t > 6$$

0											
0	0						any				
0	0	0					any	any			
0	0	0.1013	0.104				any	bold	timid		
0	0.225	0.23	0.3038	0.3089			bold	timid	bold	timid	
0.50	0.50	0.5513	0.555	0.5643	0.5684	timid	timid	bold	timid	bold	timid
1	0.95	0.91	0.8826	0.8603			timid	timid	timid	timid	
1	1	0.995	0.987				any	timid	timid		
1	1	1					any	any			
1	1						any				
1											

Code:

In [14]:

```
v=[[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,
0,0,0,0],
  [0.50,0,0,0,0,0],
  [1,0,0,0,0,0],[1,0,0,0,0,0],[1,0,0,0,0,0],[1,0,0,0,0,0],[1,0,0
,0,0,0],]
policy=[[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],
[0,0,0,0,0,0],
  [0,0,0,0,0,0],
  [0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0
,0,0,0],]


for t in range(1,6):
    for x in range(1,10):
        if x<t or x+t>10:
            v[x][t]='NaN'
            policy[x][t]='NaN'
        else:
            timid=round(0.8*v[x][t-1]+0.1*v[x-1][t-1]+0.1*v[x+1]
[t-1],4)
            bold=round(0.45*v[x+1][t-1]+0.55*v[x-1][t-1],4)
            if timid>bold:
                v[x][t]=timid
                if v[x][t]==0 or v[x][t]==1:
                    policy[x][t]='any'
                else:
                    policy[x][t]='timid'
            else:
                v[x][t]=bold
                if v[x][t]==0 or v[x][t]==1:
                    policy[x][t]='any'
                else:
                    policy[x][t]='bold'
    x=x-1
t=t-1
for x in range(0,11):
    print(v[x][0], v[x][1], v[x][2], v[x][3], v[x][4], v[x][5],)
print("\n")
for x in range(0,11):
    print(policy[x][0], policy[x][1], policy[x][2], policy[x][3]
, policy[x][4], policy[x][5],)
print("\n")
```

```
0 0 0 0 0 0
0 0.0 NaN NaN NaN NaN
0 0.0 0.0 NaN NaN NaN
0 0.0 0.1013 0.104 NaN NaN
0 0.225 0.23 0.3038 0.3089 NaN
0.5 0.5 0.5513 0.555 0.5643 0.5684
1 0.95 0.91 0.8826 0.8603 NaN
1 1.0 0.995 0.987 NaN NaN
1 1.0 1.0 NaN NaN NaN
1 1.0 NaN NaN NaN NaN
1 0 0 0 0 0
```

```
0 0 0 0 0 0
0 any NaN NaN NaN NaN
0 any any NaN NaN NaN
0 any bold timid NaN NaN
0 bold timid bold timid NaN
0 timid bold timid bold timid
0 timid timid timid timid NaN
0 any timid timid NaN NaN
0 any any NaN NaN NaN
0 any NaN NaN NaN NaN
0 0 0 0 0 0
```

2. Let the probability of wining in the case of a timid play be 0, 0.3, 0.6, 0.9, 0.12, 0.15, we can see the higher probability of wining, the more we choose to use timid way.
   Code:

In [9]:

```
v=[[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,
0,0,0,0],
  [0.50,0,0,0,0,0],
  [1,0,0,0,0,0],[1,0,0,0,0,0],[1,0,0,0,0,0],[1,0,0,0,0,0],[1,0,0
,0,0,0],]
policy=[[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],
[0,0,0,0,0,0],
  [0,0,0,0,0,0],
  [0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0,0,0,0],[0,0,0
,0,0,0],]

for a in [0, 0.03, 0.06, 0.09, 0.12, 0.15]:
    for t in range(1,6):
        for x in range(1,10):
            if x<t or x+t>10:
                v[x][t]='NaN'
                policy[x][t]='NaN'
            else:
                timid=round((1-2*a)*v[x][t-1]+a*v[x-1][t-1]+a*v[
x+1][t-1],4)
                bold=round(0.45*v[x+1][t-1]+0.55*v[x-1][t-1],4)
                if timid>bold:
                    v[x][t]=timid
                    if v[x][t]==0 or v[x][t]==1:
                        policy[x][t]='any'
                    else:
                        policy[x][t]='timid'
                else:
                    v[x][t]=bold
                    if v[x][t]==0 or v[x][t]==1:
                        policy[x][t]='any'
                    else:
                        policy[x][t]='bold'
        x=x-1
    t=t-1
    print('a =',a,'\n')
    #for x in range(0,11):
    #    print(v[x][0], v[x][1], v[x][2], v[x][3], v[x][4], v[x]
[5],)
    #print("\n")
    for x in range(0,11):
        print(policy[x][0], policy[x][1], policy[x][2], policy[x
][3], policy[x][4], policy[x][5],)
    print("\n")
```

Output:

```
a = 0
0 0 0 0 0 0
0 any NaN NaN NaN NaN
0 any any NaN NaN NaN
0 any bold bold NaN NaN
0 bold bold bold bold NaN
0 timid bold bold bold bold
0 any any any any NaN
0 any any any NaN NaN
0 any any NaN NaN NaN
0 any NaN NaN NaN NaN
0 0 0 0 0 0
```

```
a = 0.03
0 0 0 0 0 0
0 any NaN NaN NaN NaN
0 any any NaN NaN NaN
0 any bold timid NaN NaN
0 bold timid bold timid NaN
0 timid bold timid bold timid
0 timid timid timid timid NaN
0 any timid timid NaN NaN
0 any any NaN NaN NaN
0 any NaN NaN NaN NaN
0 0 0 0 0 0
```

```
a = 0.06
0 0 0 0 0 0
0 any NaN NaN NaN NaN
0 any any NaN NaN NaN
0 any bold timid NaN NaN
0 bold timid bold timid NaN
0 timid bold timid bold timid
0 timid timid timid timid NaN
0 any timid timid NaN NaN
0 any any NaN NaN NaN
0 any NaN NaN NaN NaN
0 0 0 0 0 0
```

```
a = 0.09
0 0 0 0 0 0
0 any NaN NaN NaN NaN
0 any any NaN NaN NaN
0 any bold timid NaN NaN
0 bold timid bold timid NaN
0 timid bold timid bold timid
0 timid timid timid timid NaN
0 any timid timid NaN NaN
0 any any NaN NaN NaN
0 any NaN NaN NaN NaN
0 0 0 0 0 0
```

```
a = 0.12
0 0 0 0 0 0
0 any NaN NaN NaN NaN
0 any any NaN NaN NaN
0 any bold timid NaN NaN
0 bold timid bold timid NaN
0 timid bold timid timid timid
0 timid timid timid timid NaN
0 any timid timid NaN NaN
0 any any NaN NaN NaN
0 any NaN NaN NaN NaN
0 0 0 0 0 0
```

```
a = 0.15
0 0 0 0 0 0
0 any NaN NaN NaN NaN
0 any any NaN NaN NaN
0 any bold timid NaN NaN
0 bold timid bold timid NaN
0 timid bold timid timid timid
0 timid timid timid timid NaN
0 any timid timid NaN NaN
0 any any NaN NaN NaN
0 any NaN NaN NaN NaN
0 0 0 0 0 0
```

Problem 2.

Solution:

(1) State space: $X = \{0, 1, 2, 3, 4\}$ – the size of tree at that moment

Control space: $U = \{0\ (maintain), 1\ (harvest)\}$

Transition probabilities:

$$P\{x_{t+1} = x_t \mid x_t = 0\} = 1,$$
$$P\{x_{t+1} = 0 \mid x_t = 1\} = 0.05,$$
$$P\{x_{t+1} = 1 \mid x_t = 1\} = 0.15,$$
$$P\{x_{t+1} = 2 \mid x_t = 1\} = 0.7,$$
$$P\{x_{t+1} = 3 \mid x_t = 1\} = 0.1,$$
$$P\{x_{t+1} = 0 \mid x_t = 2\} = 0.05,$$
$$P\{x_{t+1} = 2 \mid x_t = 2\} = 0.2,$$
$$P\{x_{t+1} = 3 \mid x_t = 2\} = 0.7,$$
$$P\{x_{t+1} = 4 \mid x_t = 2\} = 0.05,$$
$$P\{x_{t+1} = 0 \mid x_t = 3\} = 0.05,$$
$$P\{x_{t+1} = 3 \mid x_t = 3\} = 0.5,$$
$$P\{x_{t+1} = 4 \mid x_t = 3\} = 0.45,$$
$$P\{x_{t+1} = 0 \mid x_t = 4\} = 0.05,$$
$$P\{x_{t+1} = 4 \mid x_t = 4\} = 0.95$$

For all other states $x, y \in X$, we have $P\{x_{t+1} = y \mid x_t = x\} = 0$;

Reward function for one period:

$$r(x,u)=\begin{cases}-10 & x=0,u=0\\ -30 & x=0,u=1\\ -11 & x=1,u=0\\ 115 & x=1,u=1\\ -12 & x=2,u=0\\ 140 & x=2,u=1\\ -13 & x=3,u=0\\ 165 & x=3,u=1\\ -14 & x=4,u=0\\ 210 & x=4,u=1\end{cases}$$

Dynamic programming equations: The value function ${v_t}^*(x)$ expresses the biggest profit at time t if the state (score) is x.

$$v_t{}^*(0)=max\{r(0,0)+v_{t+1}{}^*(0),r(0,1)\}$$

$$v_t{}^*(1)=max\{r(1,0)+0.05v_{t+1}{}^*(0)+0.15v_{t+1}{}^*(1)+0.7v_{t+1}{}^*(2)+0.1v_{t+1}{}^*(3),r(1,1)\}$$

$$v_t{}^*(2)=max\{r(2,0)+0.05v_{t+1}{}^*(0)+0.2v_{t+1}{}^*(2)+0.7v_{t+1}{}^*(3)+0.05v_{t+1}{}^*(4),r(2,1)\}$$

$$v_t{}^*(3)=max\{r(3,0)+0.05v_{t+1}{}^*(0)+0.5v_{t+1}{}^*(3)+0.45v_{t+1}{}^*(4),r(3,1)\}$$

$$v_t{}^*(4)=max\{r(4,0)+0.05v_{t+1}{}^*(0)+0.95v_{t+1}{}^*(4),r(4,1)\}$$

(2) I choose value iteration and policy iterations to solve this problem.

Value iterations:

I got $v^*=(-30,115,140,165,210)^T$ in 4 iterations.

MATLAB Code:

```
% [v_lo,n_it] = dne1_value_iteration_revised (0.9,10000);
function [v_lo,n_it] = dne1_value_iteration_revised (alpha,max_it)
i = 0;
n_it = max_it;
v=[0,0,0,0,0];
vv=[0,0,0,0,0];
v_lo=[0,0,0,0,0];
v_up=[0,0,0,0,0];
while (i < n_it)
    vv(1) = max(-10+alpha*v(1),-30);
    vv(2) = max(-11+alpha*(0.05*v(1)+0.15*v(2)+0.7*v(3)+0.1*v(4)),115);
    vv(3) = max(-12+alpha*(0.05*v(1)+0.2*v(3)+0.7*v(4)+0.05*v(5)),140);
    vv(4) = max(-13+alpha*(0.05*v(1)+0.5*v(4)+0.45*v(5)),165);
    vv(5) = max(-14+alpha*(0.05*v(1)+0.95*v(5)),210);
    v_lo = vv + min(vv-v)*alpha/(1-alpha);
```

```
    v_up = vv + max(vv-v)*alpha/(1-alpha);
    if (isequal(v,vv))
        n_it=i;
    end
    i=i+1;
    v(1)=vv(1);
    v(2)=vv(2);
    v(3)=vv(3);
    v(4)=vv(4);
    v(5)=vv(5);
end
end
```

Output:

```
>> dne1_value_iteration_revised(0.9, 10000)

ans =

   -30   115   140   165   210
```

Policy iterations:

$\pi^1 = (0,0,0,0,0)^T$, v^1 is the solution of the system of equations

$$\begin{cases} v^1(0) = -10 + 0.9v^1(0) \\ v^1(1) = -11 + 0.045v^1(0) + 0.135(1) + 0.63v^1(2) + 0.09v^1(3) \\ v^1(2) = -12 + 0.045v^1(0) + 0.18(2) + 0.63v^1(3) + 0.045v^1(4) \\ v^1(3) = -13 + 0.045v^1(0) + 0.45(3) + 0.405v^1(4) \\ v^1(4) = -14 + 0.045v^1(0) + 0.885v^1(4) \end{cases}$$

So

$$\begin{cases} v^1(0) = -100 < -30 \\ v^1(1) = -138.7296 < 115 \\ v^1(2) = -144.4066 < 140 \\ v^1(3) = -150.2767 < 165 \\ v^1(4) = -160.8696 < 210 \end{cases}$$

The policy maximize is $\pi^1 = (1,1,1,1,1)^T$, the v^2 is the solution of the system of equations. So the policy is $\pi^2 = (1,1,1,1,1)^T$, we have

$\pi^* = (1,1,1,1,1)^T$, $v^* = (-30,115,140,164,210)^T$.