למידה במערכות דינמיות אביב תש"פ

תרגיל בית 1

מגישים:

יאיר נחום 034462796

דר ערבה 205874951

* 1. All probabilities are greater or equal to 0. So, there can’t be in a row I that is negative. From the total probability formula on all possible next states we get:  
     In other words, this is because, given that we are in state *i*, the next state must be one of the possible states. Thus, when we sum over all the possible values of *j*, we should get one. That is, the rows of any state transition matrix must sum to one.
  2. From a we get that the sum of each row is 1. Therefore, if we multiply the P matrix with a vector of we get that . Thus, is an eigenvector of eigenvalue 1. So, since the probabilities sum up on each row to 1, there is always an eigenvalue 1 that has an eigenvector .
  3. Let’s assume the opposite. Meaning, we have an eigenvalue which has an eigenvector .

We denote the maximum component of vector as   
and the minimum component of vector as

In case and :  
On one hand, for and on the other hand

So, we got a contradiction.  
In case and :   
the and on one hand, for but on the other hand   
So, again we got a contradiction. Thus, it can’t be true that .

We can show the same thing when assuming symetrically when the .  
That’s proves the claim that .

**3.**

1. Let be the probability that after observing candidates the candidate has the highest score (of all candidates).

Let be the score of the candidate.

By observing candidates we know either if-

1. One of them has a higher score than and therefore is not the highest score seen so far .

Or -

1. is higher than all the scores seen so far one .

Therefore -

Another way to think of it, is as follows:

1. Let be the transition probability of the following candidate be the highest scoring candidate seen so far (with knowing of the current state).

In the same way will be the transition probability of the following candidate to **not** be the highest scoring candidate seen so far (with knowing of the current state).

Clearly since the transition probability in independent of the current state. For the candidate selected at time (the next state) to be it must be larger than all scores prior to it. The probability is the same whether was the largest(current state ) or was the largest of the first states(current state - ).

One of the candidates has the highest score. Since they are randomly(uniformly) selected, each has same probability of being the largest in the group.  
One can think of it as selecting the last place in the t+1 places for the largest number in these t+1 places.

1. Let denote the maximal probability of choosing the best candidate from state 𝑠 at time 𝑡 assuming no candidate had been chosen so far.

is actually the maximum probability to select the best candidate given we are in state s.  
we can think of it as computing this value (probability function) in DP and deciding when to stop (optimal stop time) according to it. If we’re in s = 1 for example, we select between the immediate reward of compared to selecting the possible future reward. The future reward is the expected value of . Meaning, .  
The action space has only one 2 actions, stop or continue and we max over it.

If we got to the last candidate (and haven’t chosen one of the candidates beforehand)-

since all candidates have been reviewed -

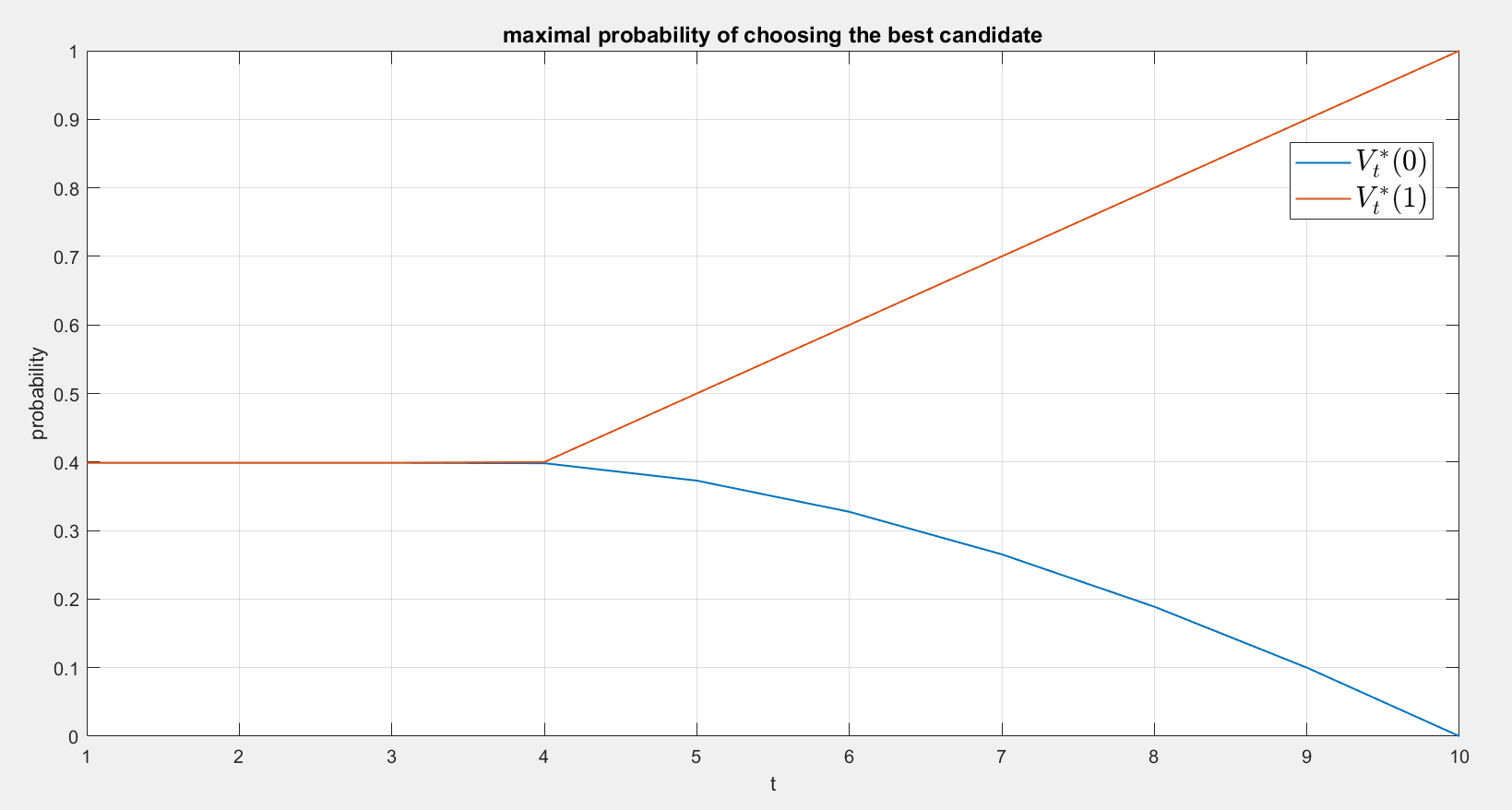
Therefore, the maximal probability is **.**

always,and since we have reached the final candidate, choosing the highest scoring candidate is impossible - **.**

1. From all previous results-

We get-

After solving the induction numerically for N=10 via MATLAB (see code in Appendix A), we got the following result:



1. The optimal strategy for choosing a candidate is the strategy that chooses a candidate at moment when is maximal (highest probability of being the best of all).

As seen from section (4), there in an initial period where , meaning that the maximal probability of choosing the best candidate is the same regardless to whether or not the current candidate is the best seen so far.

After this first initial period, in monotonal increasing while is monotonal decreasing. In order to choose the candidate at the highest we want to choose at a point in time where (the current candidate is best seen). This way and we are located on the increasing probability plot.

After waiting an initial period, we are at risk that the best candidate is chosen in the initial discarded group (and then for all the remaining candidates decreasing ). In this case the strategy will wait for , get to the last candidate and be forced to hire him.

The initial period is as seen since there is a tradeoff between the probability that the best candidate will be chosen in the period will be low (short ) and the probability that the best candidate will remain unseen is low and we get a better chance to hire someone good(long ).  
It turns out that this period converges to .

**4.** Need to prove the following equation:

We will mark k as t+1 and replace the integration start and the integrand expression:

Since the policy is a given time invariant one, and the dynamics are also stationary, the reward doesn’t depend on the time and therefor and are distributed the same(for every time index ). For two random variables with same probability distribution we have equality between their Mean value. Since we start to sum up the same expression from the same start stage to infinity, we get:

**5.**

1. Let be the second moment of the discounted return when starting from state s and following policy π.

Similarly -

We got a linear dependency of on :

1. In order to calculate for all states, we solve the equation above and the equation for :

A total of  **equations** for variables-.

1. Let be the variance of the discounted return when starting from state s and following policy π.

In order to calculate , we first calculate and (as described in equations form previous section). Secondly, we subtract from .

**Appendix A**

Code for question 3 section 4.

clc; %clear all;

N = 10;

t = 1:1:N;

Vt0 = zeros(1,N);

Vt1 = zeros(1,N);

Vt0(N) = 0;

Vt1(N) = 1;

for i = (N-1):-1:1

Vt0(i) = (1/(i+1))\*Vt1(i+1)+(i/(i+1))\*Vt0(i+1);

Vt1(i) = max(i/N,Vt0(i));

end

plot(t,Vt0,t,Vt1,'LineWidth',1); grid on;

title('maximal probability of choosing the best candidate', 'fontsize', 12);

ylabel('probability');

xlabel('t');

legend('$V^\*\_t(0)$','$V^\*\_t(1)$', 'Interpreter','latex', 'fontsize', 16);