M2 ENS Lyon: MDP and RL.

The Intern Problem

Let us consider the following problem. A firm wants to hire one intern. There are N candidates for the job, all of different levels. The firm wants to hire the candidate with the best level. It uses the following interview procedure:
- The N candidates are ordered randomly and get interviewed one by one.
- The interview of a candidate reveals its level.
- At the end of an interview the firm must take one decision:

1. hire the current candidate and stop the process (H);
2. not hire the current candidate and continue with the next one (C).

In both cases the choice over each candidate is final and the firm cannot go back and hire a previously interviewed candidate. If the interviews go all the way to the last candidate, the firm may or may not hire the last candidate.

The goal here is to determine the optimal decision policy for the firm to maximize its expected value. The value is 1 if the hired candidate (if any) is the best one and 0 otherwise. Here, even the second best candidate has no value.

0.1

Explain in a few words why this problem is a MDP over a finite horizon (the construction of the MDP is done in the next questions).

To construct the Bellman equations (see next questions), we consider that the state space is $\{0, 1, 2, 3\}$. At step $1 \leq n \leq N$, State 1 means that the current candidate (number n) is the best among the first n candidates. State 0 means that the current candidate is not the best among the first n candidates. State 2 means that the hiring process has finished at an earlier stage and the best candidate was hired. State 3 means that the hiring process has finished at an earlier stage and the hired candidate is not the best. The initial state is $s_1 = 1$ (why?).

The actions of the firm are: action H means hire the current candidate and action C means do not hire and continue.

0.2

Compute the following quantities as functions of n and N:
$p_g(n)$ is the probability that the current candidate is the best among all the candidates.
$p_\ell(n)$ is the probability that the current candidate is the best among the first n.
$p_c(n)$ is the probability that the current candidate is the best given that it is the best among the first n.

0.3

To construct the Bell optimality equation, we define the expected optimal value for the last steps (from n to N) as $V_n^*(s)$ for $s \in \{0, 1, 2, 3\}$, the current state at step n.

We set $V_{N+1}^*(0) = 0$, $V_{N+1}^*(1) = 0$ and $V_{N+1}^*(2) = 1$ and $V_{N+1}^*(3) = 0$. Explain this choice.

Explain why the Bellman equations for all $n \leq N$ are

$$\begin{aligned}
V_n^*(0) &= \max\{0 + V_{n+1}^*(3), 0 + p_\ell(n+1)V_{n+1}^*(1) + (1 - p_\ell(n+1))V_{n+1}^*(0)\} \\
V_n^*(2) &= \max\{0 + V_{n+1}^*(2), 0 + V_{n+1}^*(2)\} \\
V_n^*(3) &= \max\{0 + V_{n+1}^*(3), 0 + V_{n+1}^*(3)\},
\end{aligned}$$

and give the Bellman equation for state 1 using $p_c(n)$.

In all the equations, does the first term in the max. correspond to action H or to action C?

0.4

Show that the states 2 and 3 can be removed and the equations become for all $n \leq N$:

$$V_n^*(0) = \frac{1}{n+1} V_{n+1}^*(1) + \frac{n}{n+1} V_{n+1}^*(0) \tag{1}$$

$$V_n^*(1) = \max\{\frac{n}{N}, \frac{1}{n+1} V_{n+1}^*(1) + \frac{n}{n+1} V_{n+1}^*(0)\}. \tag{2}$$

0.5

Let π^* be an optimal policy. Assume that at step n, $\pi_n^*(1) = C$, then show that $\pi_{n-1}^*(1) = C$.

Explain why this implies that the optimal policy is of the following form:

There exist an optimal threshold k^* such that if $n < k^*$ then the optimal policy always continues (in states 0 or 1). If $n \geq k^*$ then the optimal policy hires in state 1 and continues in state 0.

0.6

This question is dedicated to the computation of the optimal threshold k^*.

Show by backward induction on n that for all $n \geq k^*$, $V_n^*(1) = \frac{n}{N}$ and $V_n^*(0) = \frac{n}{N}(\frac{1}{n} + \cdots + \frac{1}{N-1})$.

Show that this implies that k^* is the largest integer k such that $\frac{1}{k} + \cdots + \frac{1}{N-1} \geq 1$.

Recall that if m is large, $1 + \frac{1}{2} + \cdots + \frac{1}{m} \approx \log(m)$. Show that when N and k^* are large, the optimal threshold is $k^* \approx N/e$.

0.7

Compute an approximation of the expected value of the optimal policy, $V_1^*(1)$, when N is large.