Probabilistic Aspects of Computer Science: TD Bonus
MDP – Modelling and Finite-horizon objectives

Chargé de TD: Adrien Koutsos

October 25, 2017

Exercise 1. We study the spinner game presented in the course. In this game, the player has to compose a five-digit number whose digits are randomly chosen by a spinner during five rounds. After every round (except the last one), the player chooses in which position he inserts the current digit. The goal of the player is to obtain the largest number as possible.

Model this problem with a Markov Decision Process. Show that an optimal policy depends only on the position of the unoccupied digit locations; that is, it does not depend on the numbers previously placed in the occupied digits. Using this information, build a smaller equivalent MDP for this problem.

Exercise 2. We consider the secretary problem. An employer seeks to hire an individual to fill a vacancy for a secretarial position. There are N candidates or applicants for this job, with N fixed and known by the employer. Candidates are interviewed sequentially. Upon completion of each interview, the employer decides whether or not to offer the job to the *current* candidate. If he does not offer the job to this candidate, that individual seeks employment elsewhere and is no longer eligible to receive an offer. In this exercise, we assume that the employer wishes to maximize the probability of giving an offer to the best candidate.

More formally, consider a collection of N objects ranked from 1 to N, with that ranked 1 being the most desirable. The true rankings are unknown to the decision maker. He observes the objects one at a time in random order. He can either select the current object and terminate the search, or discard it and choose the next object. His objective is to maximize the probability of choosing the object ranked number 1. We assume that the decision maker's relative rankings are consistent with the absolute rankings. That is, if object A has a lower numerical ranking than object B, the decision maker will prefer object A to object B.

1. Model this problem with a Markov Decision Process with a finite-horizon objective.
2. Draw the graph of the MDP when $N = 4$. In this case, find the optimal policy by applying the algorithm studied in the course.
3. We consider the general case now. By finding a recurrence relation verified by the coefficients of $\mathbf{u}_t^*$ ($0 \leq t \leq N$), show that an optimal policy has the form "Observe the first τ candidates; after that, select the first candidate who is better than all the previous ones" (here τ may depend of N, hence we will denote it $\tau(N)$ in the following).
4. We suppose now that $N > 2$. Show that $\tau(N) \geq 1$. Describe $\tau(N)$ with respect to N. Find again the result for the case $N = 4$. Describe the asymptotical behavior of $\tau(N)$, when N goes to ∞.