Quantitative Verification Session 9

January 12, 2017

Rewards on MDPs and DTMCs

Note. On MDPs, we only talk about maximum/minimum rewards and on DTMCs, we ask what is *the* reward. This is because in MDPs, it makes sense to talk about rewards only when we have fixed a scheduler (Θ). So what does maximum mean? It is the supremeum over all schedulers, the rewards obtained in the Markov chain induced by the scheduler (i.e. $sup_{\Theta}\ R^{\Theta}$). Some of the examples given below correspond to DTMCs, which you may be able to figure out.

1 Instantaneous rewards

Example 1. What is the maximal expected number of pieces in the play area after 50 rounds?

$R_{max} =? [I = 50]$

Example 2. The expected number of messages still to be delivered after 10 time steps is atleast 4.

$R_{\geq 4} [I = 10]$

Example 3. What is the maximum expected size of the queue after 200 steps?

$R_{max} =? [I = 200]$

Instantaneous reward at step k is the state rewards expected at step k. See slide 33+ of Chapter 4 for more details.

2 Step-bounded cumulative rewards

Example 1. What is the maximal expected number of times I kick out a piece of the opponent within the first 100 steps?

$R_{max} =? [C \leq 100]$

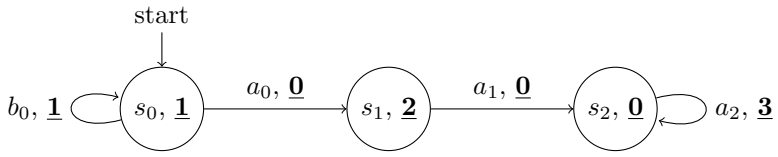

Figure 1: MDP which shows that memoryless schedulers don't exist for instantaneous rewards. (e.g. take b_0 twice and a_0 to maximize instantaneous reward in 3 steps)

Example 2. The expected power consumption within the first 100 time steps of operation is less than or equal to 5.5.

$R_{\leq 5.5}\ [C \leq 100]$

See slide 33 of Chapter 4 for the equation on how to compute the step bounded cumulative rewards.

Exercise 1: Check $R_{\geq 1}\ [C \leq 2]$ on Figure 2.
Exercise 2: Compute $R_{max} =?\ [C \leq 3]$ on Figure 1.

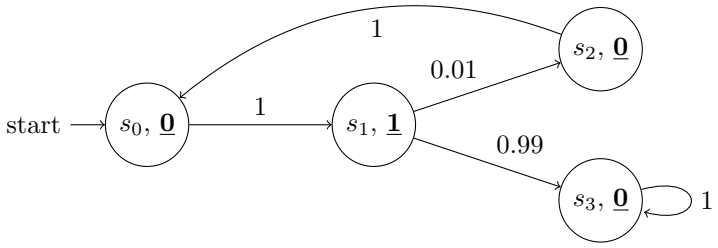

Figure 2: An example DTMC. Note that the values on the edges are probabilities. Only state rewards are present and they are bold-faced and underlined.

3 Cumulative rewards to reach target

Example 1. What is the minimal expected number of steps before the game ends?

$R_{max} =?\ [F\ \text{“}finish\text{”}]$

Example 2. Expected number of correctly delivered messages is atleast 14.

$R_{\geq 14}\ [F\ \text{“}done\text{”}]$

Exercise 1: Compute $R =?\ [F\ s = 3]$ on Figure 2.
Exercise 2: Open `prism-examples/consensus/coinX.nm` in Prism and run `R{"steps"}max=? [F "finished"]` on it.

4 Long run average reward (or Mean-payoff)

Example 1. We have an MDP for planning investments. Different actions represent buying, selling or holding investments. Each of these carry a reward. How do we maximize our profits in the long run?

See Chapter 4, slide 34+ for more details.

Idea for Value Iteration. In the long run, the system would run around within maximal end components.

(Continued in the next session)