Model Checking: exercise set 6
Robust Markov Decision Processes

Due date: March 13

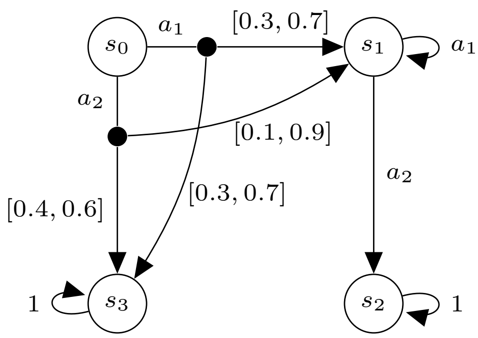
1. Robust Bellman equations

Consider a robust MDP $(S, A, \mathcal{P}, R)$ and a reachability property $\mathbb{P}_{\max}(\Diamond T)$ for some target set $T \subset S$. Give the *robust* Bellman equation that maximizes the value of this objective.

Hint: Recall the Bellman equation for reachability from the MDP lecture, and see how it differes from the (robust) Bellman equations discussed for discounted reward.

2. IMDPs

Consider the IMDP below.



Answer the following questions:

(a) Consider the reachability specification $\mathbb{P}_{\max}(\Diamond s_2)$. What are the best-case and worst-case instances at state-action pairs (s_0, a_1) and (s_0, a_2), respectively.

(b) For (s_0, a_1) and (s_0, a_2), define the convex polytopes that contain all valid probability distributions at each of these two state-action pairs.

(c) Draw both convex polytopes. *Hint*: start by constructing a box from the intervals, then find where the valid probability distributions are within this box.

(d) For each of the convex polytopes, list the extreme points (vertices), check your answers at question 1.

3. Inner problems

Consider the inner minimization problem for IMDPs (slide 25). Describe what changes need to be made to the algorithm for IMDPs, such that it solves the *inner maximization* problem.

4. Learning

Suppose we have a single state-action pair (s, a) with 4 successor states s_1, s_2, s_3, s_4. We take $N = 12$ samples of (s, a), with counts $\#(s, a, s_1) = 2$, $\#(s, a, s_2) = 6$, $\#(s, a, s_3) = 0$, $\#(s, a, s_4) = 4$. This state-action pair is part of a larger MDP with $\sum_{s,a} |\text{Post}(s, a)| = 8$.

(a) Use frequentist learning to learn the probabilities for the transitions from (s, a).

(b) Use PAC-learning to learn probability intervals for the transitions $(s, a, s_i), i = 1, \ldots, 4$. Use an error rate of $\varepsilon = 0.01$.