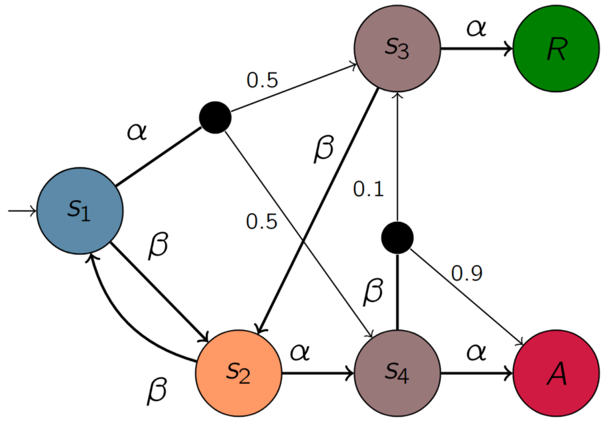
Model Checking: exercise set 10 - Partially Observable MDPs

Due date: May 15

1. Policies

Consider the following POMDP:



The goal is to reach state R with maximal probability. To do that, give observation-based policies of the following types:

1. Memoryless deterministic, i.e. $\pi\colon Z \to A$.
2. Memoryless randomized, i.e. $\pi\colon Z \to \mathcal{D}(A)$.
3. Finite-memory deterministic, i.e. $\pi\colon (Z \times A)^{k-1} \times Z \to A$. You may use as much memory as you want.

Which of these has the highest reachability probability? What is the key problem in this POMDP?

2. Belief updates

Consider the same POMDP as in the previous question. Suppose we have a belief $b_0 = \{s_1 \mapsto 1\}$ (that is, we know we are in the blue state). Perform the following belief updates:

1. $b_1 = \mathsf{BU}(b_0, \alpha, brown)$.
2. $b_2 = \mathsf{BU}(b_1, \beta, orange)$.
3. $b_3 = \mathsf{BU}(b_2, \alpha, brown)$.

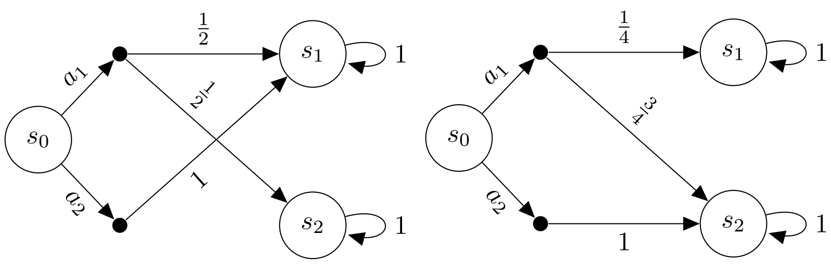
Explain how belief b_3 relates to the history of observations $(blue, brown, orange, brown)$.

3. POMDPs and Parametric Markov chains

Give the (memoryless) parametric Markov chain for the POMDP of question 1.

4. MEMDPs

Consider the following MEMDP with two environments:



1. Construct the disjoint union POMDP for this MEMDP.
2. Give the memoryless parametric Markov chain for this disjoint union POMDP.
3. Start with a uniform distribution over the s_0 states, that is

$$b_0 = \{(s_0, 1) \mapsto 0.5, (s_0, 2) \mapsto 0.5\}$$

   Now use the belief update to compute the two following beliefs:

   - $b_1 = \mathsf{BU}(b_0, a_2, s_1)$,
   - $b_2 = \mathsf{BU}(b_0, a_2, s_2)$.

   What conclusions can we draw from the beliefs b_1 and b_2 regarding the identification problem for MEMDPs?