Problem 1: Definitions

Simple Stochastic Games

A *simple stochastic game* (SSG) $G = \langle S_{max}, S_{min}, S_{ran}, l, r\rangle$ consists of the following components:

- A set S_{max} of *max states*; a set S_{min} of *min states*; and a set S_{ran} of *random states*. The three sets S_{max}, S_{min}, and S_{ran} are pairwise disjoint. We write $S = S_{max} \cup S_{min} \cup S_{ran}$ and $S^+ = S \cup \{t_{max}, t_{min}\}$, where $t_{max}, t_{min} \notin S$ are special *target states*.
- Two successor functions l: $S \to S^+$ and r: $S \to S^+$.

A state $s \in S$ is *deterministic* if $l(s) = r(s)$. The SSG G is a *min MDP* if all states in S_{max} are deterministic; a *max MDP*, if all states in S_{min} are deterministic; and a *Markov chain*, if all states in $S_{max} \cup S_{min}$ are deterministic.

Plays

The game is played between two players, a max player and a min player. Initially a token is placed on one of the states. If the token is on a max state, then the max player moves it to one of the two successor states; if the token is on a min state, then the min player moves it to one of the two successor states; and if the token is on a random state, then it is moved with probability 1/2 to each of the two successor states. This process is repeated until the token visits a target state. If the token visits the target t_{max}, then the max player wins; if the token visits t_{min}, then the min player wins; and if the game is played forever without the token visiting a target, then neither player wins. Formally, a *winning play* of the SSG G is a finite sequence $\omega = \langle s_0, s_1, \ldots, s_k\rangle$ of states ($k \geq 0$) such that (1) for all $0 \leq i < k$, we have $s_i \in S$ and $s_{i+1} \in \{l(s_i), r(s_i)\}$, and (2) $s_k \in \{t_{max}, t_{min}\}$. The play ω is *max winning* if $s_k = t_{max}$; otherwise it is *min winning*. The *probability* of the play is $\Pr(\omega) = 1/2^m$, where $m = |\{0 \leq i < k \mid s_i \in S_{ran}\}|$ is the number of random states that occur in ω.

Strategies

If a state is visited twice when playing the game, a player may choose different successor states. It can be shown, however, that this is never to the player's benefit. Hence it suffices to consider positional strategies for both players, which prescribe, for each state, whether the player chooses the left (l) or right (r) successor state. Consider an SSG $G = \langle S_{max}, S_{min}, S_{ran}, l, r\rangle$. A *positional max strategy* for G is a function σ: $S_{max} \to \{L, R\}$, and a *positional min strategy* is a function π: $S_{min} \to \{L, R\}$. We write Σ for the set of positional max strategies, and Π for the set of positional min strategies. Given a positional max strategy $\sigma \in \Sigma$, let $G^\sigma = \langle S_{max}, S_{min}, S_{ran}, l', r'\rangle$ be the min MDP such that

- for all states $s \in S_{max}$, if $\sigma(s) = L$ then $l'(s) = l(s)$ and $r'(s) = l(s)$, and otherwise $l'(s) = r(s)$ and $r'(s) = r(s)$;

- for all states $s \in S_{min} \cup S_{ran}$, we have $l'(s) = l(s)$ and $r'(s) = r(s)$.

For a positional min strategy $\pi \in \Pi$, the max MDP G^π is defined analogously. Given both $\sigma \in \Sigma$ and $\pi \in \Pi$, let $G^{\sigma,\pi} = \langle S_{max}, S_{min}, S_{ran}, l', r' \rangle$ be the Markov chain such that

- for all states $s \in S_{max}$, if $\sigma(s) = L$ then $l'(s) = l(s)$ and $r'(s) = l(s)$, and otherwise $l'(s) = r(s)$ and $r'(s) = r(s)$;
- for all states $s \in S_{min}$, if $\pi(s) = L$ then $l'(s) = l(s)$ and $r'(s) = l(s)$, and otherwise $l'(s) = r(s)$ and $r'(s) = r(s)$;
- for all states $s \in S_{ran}$, we have $l'(s) = l(s)$ and $r'(s) = r(s)$.

The Markov chain $G^{\sigma,\pi}$ characterizes the result of the game if the two players follow the strategies σ and π, respectively. The SSG G is *stopping* if for all strategies $\sigma \in \Sigma$ and $\pi \in \Pi$, and all states $s \in S$, the Markov chain $G^{\sigma,\pi}$ has a winning play that starts in s. We will see that the stopping criterion ensures that the probability of the game being played forever without one of the players winning is 0.

Values

Consider a state $s \in S^+$, a positional max strategy $\sigma \in \Sigma$, and a positional min strategy $\pi \in \Pi$. We write $\Omega_s^{\sigma,\pi}$ for the set of winning plays of the Markov chain $G^{\sigma,\pi}$ which start in s, and we write $W_s^{\sigma,\pi} \subseteq \Omega_s^{\sigma,\pi}$ for the set of max winning plays that start in s. If G is stopping, then $\langle \Omega_s^{\sigma,\pi}, \Pr \rangle$ is a discrete probability space (why?). For every state $s \in S^+$ of a stopping SSG, we define the following probabilities:

- Let $x_s^{\sigma,\pi} = \Pr[W_s^{\sigma,\pi}]$ be the probability that a play which starts in s is max winning if the two players follow the strategies σ and π.
- Let $x_s^\sigma = \min_{\pi \in \Pi} x_s^{\sigma,\pi}$ be the least probability of max winning from s that the min player can ensure against the max strategy σ. The positional min strategies $\pi \in \Pi$ at which the minimum is realized are called *optimal* min strategies from s against the given max strategy σ.
- Let $x_s = \max_{\sigma \in \Sigma} x_s^\sigma$ be the greatest probability of max winning from s that the max player can ensure against any min strategy. The positional max strategies $\sigma \in \Sigma$ at which the maximum is realized are called *optimal* max strategies from s against *any* min strategy.

The probability

$$x_s = \max_{\sigma \in \Sigma} \min_{\pi \in \Pi} \Pr[W_s^{\sigma,\pi}]$$

is called the *max value* of the state s in the SSG G. The SSG decision problem asks, given a stopping SSG G and a state s of G, whether $x_s > 1/2$.

Determinacy

Stopping SSGs are *zero-sum* games, which means that if the max player wins the game, then the min player loses, and vice versa. This is because each winning play is either max winning or min winning, but not both. An important property of two-player zero-sum games is *determinacy*, which means that if both players play optimally, then their respective probabilities of winning add up to 1. Stopping SSGs are indeed determined for positional strategies. Formally, for every state $s \in S^+$ of a stopping SSG, define the *min value*

$$y_s = \min_{\pi \in \Pi} \max_{\sigma \in \Sigma} \Pr[\Omega_s^{\sigma,\pi} \backslash W_s^{\sigma,\pi}]$$

to be the greatest probability of min winning from s that the min player can ensure against any max strategy. Then it can be shown that $y_s = 1 - x_s$ for all states $s \in S^+$.

Value Improvement

The max values of a stopping SSG satisfy the following *transition equations*, one for each state:

- for every state $s \in S_{max}$, we have $x_s = \max\{x_{l(s)}, x_{r(s)}\}$;
- for every state $s \in S_{min}$, we have $x_s = \min\{x_{l(s)}, x_{r(s)}\}$;
- for every state $s \in S_{ran}$, we have $x_s = 0.5 \cdot x_{l(s)} + 0.5 \cdot x_{r(s)}$;
- $x_{t_{max}} = 1$ and $x_{t_{min}} = 0$.

The transition equations can be used to compute all max values:

algorithm ValueImprovement
Input: stopping SSG $G = \langle S_{max}, S_{min}, S_{ran}, l, r \rangle$.
Output: max values x_s for all states $s \in S^+$.
$x_{t_{max}} := 1$; $x_{t_{min}} := 0$;
$x_s := 1$ for all $s \in S_{max}$;
$x_s := 0$ for all $s \in S_{min}$;
$x_s := 0.5$ for all $s \in S_{ran}$;
while some x_s changes **do**
  $x_s := \max\{x_{l(s)}, x_{r(s)}\}$ for all $s \in S_{max}$;
  $x_s := \min\{x_{l(s)}, x_{r(s)}\}$ for all $s \in S_{min}$;
  $x_s := 0.5 \cdot x_{l(s)} + 0.5 \cdot x_{r(s)}$ for all $s \in S_{ran}$;
  end while.

The algorithm ValueImprovement is guaranteed to converge to the max values, but it may require an exponential number of iterations of the while loop, even in the special case that G is a Markov chain (can you find an example?). However, if G is a Markov chain or an MDP, then the transition equations can be solved much more efficiently. For Markov chains, we can directly solve the resulting system of linear equations:

$x_s = x_{l(s)}$ for all $s \in S_{max} \cup S_{min}$;
$x_s = 0.5 \cdot x_{l(s)} + 0.5 \cdot x_{r(s)}$ for all $s \in S_{ran}$;
$x_{t_{max}} = 1$ and $x_{t_{min}} = 0$.

For min MDPs, we can solve a linear optimization problem:

$x_s = x_{l(s)}$ for all $s \in S_{max}$;
$x_s \leq x_{l(s)}$ and $x_s \leq x_{r(s)}$ for all $s \in S_{min}$;
$x_s = 0.5 \cdot x_{l(s)} + 0.5 \cdot x_{r(s)}$ for all $s \in S_{ran}$;
$x_{t_{max}} = 1$ and $x_{t_{min}} = 0$;
maximize $\sum_{s \in S_{min}} x_s$.

It follows that, given a positional max strategy $\sigma \in \Sigma$, we can compute on the min MDP G^σ an optimal min strategy against σ by linear programming, in polynomial time (how?). Moreover, we can check if σ is optimal against *any* min strategy in polynomial time (how?). As a corollary, we conclude that the SSG decision problem is in NP $\cap$ coNP (why?). No polynomial-time agorithm is known for solving SSG games.

Random Permutations

A *random permutation* for an SSG is a permutation of the random states. We associate with every random permutation $p = \langle s_1, \ldots, s_n \rangle$ a positional max strategy $\sigma_p \in \Sigma$, which intuitively behaves as follows:

- For all $0 \leq i \leq n$, if s is a state from which the max player can ensure that a state in $\{s_{i+1}, \ldots, s_n, t_{max}\}$ is visited without visiting a state in $\{s_1, \ldots, s_i\}$, then $\sigma_p(s)$ follows such a strategy.
- If s is a state from which the max player cannot ensure that a state in $\{s_1, \ldots, s_n, t_{max}\}$ is visited, then $\sigma_p(s)$ is chosen arbitrarily.

Given a random permutation p, we can compute the strategy $\sigma(p)$ in linear time (how?). Gimbert and Horn (2008) have proved that for every stopping SSG, there exists a random permutation p such that $\sigma(p)$ is optimal against *any* min strategy. Your task is to define and evaluate heuristics for constructing "good" random permutations.