Universidad Nacional de Rosario

Tesina de grado

para la obtención del grado de

Licenciado en Ciencias de la Computación

Vale por un título

Autor:

Tu persona

Directores:

XX

YY

Departamento de Ciencias de la Computación

Facultad de Ciencias Exactas, Ingeniería y Agrimensura

Av. Pellegrini 250, Rosario, Santa Fe, Argentina

4 de junio de 2025

May the force be with you.

YODA (Y PARTICULARMENTE SEBA EN ESTE MOMENTO)

Agradecimientos

Quiero agradecer a quien quiera agradecer. (Por ejemplo, a la universidad pública).

Resumen

Versión cortita de tu tesina.

Índice general

Capítulo 1

Introducción

1.1. Contribuciones

...

1.2. Organización del trabajo

HOLA

- En el **Capítulo ??**, presentaré ...;
- En el **Capítulo 3**, extenderé y explicaré en detalle los métodos de esta tesina ...;
- En el **Capítulo 5**, presentaré mis resultados y conseguidos ...;
- Finalmente, en el **Capítulo 6**, concluyo con un resumen de los aportes realizados en esta tesina, menciono las implicancias de esta investigación y hallazgos realizados, y sugiero potenciales caminos para futuras investigaciones.

Capítulo 2

Preliminares

2.1. sigma algebras

2.1.1. Baier y Katoen

Una σ-**álgebra** es un par $(Outc, \mathfrak{E})$ donde $Outc$ es un conjunto no vacío y $\mathfrak{E} \subseteq 2^{Outc}$ es un conjunto que consiste en subconjuntos de $Outc$ que contiene el conjunto vacío y es cerrado bajo complementación y uniones contables, es decir:

- $\emptyset \in \mathfrak{E}$,
- si $E \in \mathfrak{E}$, entonces $Outc \setminus E \in \mathfrak{E}$,
- si $E_1, E_2, \ldots \in \mathfrak{E}$, entonces $\bigcup_{n=1}^{\infty} E_n \in \mathfrak{E}$.

Nótese que las condiciones de las σ-álgebras implican que $Outc \in \mathfrak{E}$, ya que $Outc \neq \emptyset$, y que $\mathfrak{E}$ es cerrado bajo intersecciones contables, ya que $\bigcap_{n=1}^{\infty} E_n = \overline{\bigcap_{n=1}^{\infty} \overline{E_n}}$.

A veces se supone que el conjunto $Outc$ está fijado y $\mathfrak{E}$ se llama una σ-álgebra. Los elementos de $Outc$ a menudo se llaman resultados (*outcomes*), mientras que los elementos de $\mathfrak{E}$ se llaman eventos.

Para cualquier conjunto $Outc$, el conjunto potencia $\mathfrak{E} = 2^{Outc}$ produce una σ-álgebra sobre $Outc$. En esta σ-álgebra, todos los subconjuntos de $Outc$ son eventos. El otro extremo es la σ-álgebra que consiste en el conjunto vacío y $Outc$, es decir, $\mathfrak{E} = \{\emptyset, Outc\}$. Aquí, ningún subconjunto propio no vacío de $Outc$ es un evento.

Una **medida de probabilidad** sobre $(Outc, \mathfrak{E})$ es una función $Pr : \mathfrak{E} \to [0, 1]$ tal que $Pr(Outc) = 1$, y si $(E_n)_{n=1}^{\infty}$ es una familia de eventos disjuntos dos a dos con $E_n \in \mathfrak{E}$, entonces:

$$Pr\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} Pr(E_n).$$

Un **espacio de probabilidad** es una σ-álgebra equipada con una medida de probabilidad, es decir, es un trupla $(Outc, \mathfrak{E}, Pr)$ donde $(Outc, \mathfrak{E})$ es una σ-álgebra y Pr es una medida de probabilidad sobre $(Outc, \mathfrak{E})$. El valor $Pr(E)$ se llama la medida de probabilidad de E, o simplemente la probabilidad de E.

En el contexto de las medidas de probabilidad, los eventos (es decir, los elementos de $\mathfrak{E}$) a menudo se dice que son medibles. Es decir, la medibilidad de un conjunto $E \subseteq Outc$ significa que $E \in \mathfrak{E}$, y por lo tanto, tiene sentido hablar de la medida de probabilidad de E.

En general, siempre que $Outc$ sea contable, se puede obtener una medida de probabilidad sobre el conjunto potencia de $Outc$ fijando una función $\mu : Outc \to [0, 1]$ tal que

$$\sum_{e \in Outc} \mu(e) = 1.$$

Tales funciones μ se llaman **distribuciones** sobre $Outc$. Cualquier distribución μ induce una medida de probabilidad sobre la σ-álgebra $E = 2^{Outc}$ de la siguiente manera. Para un subconjunto E de $Outc$, $Pr_{\mu}(E)$ se define como

$$Pr_\mu(E) = \sum_{e \in E} \mu(e).$$

De hecho, es fácil verificar que μ satisface las condiciones de una medida de probabilidad. En lo sucesivo, $Pr_\mu(E)$ a menudo se abrevia como $\mu(E)$ y $Distr(Outc)$ se usa para denotar el conjunto de distribuciones sobre $Outc$.

Resumamos algunas propiedades fundamentales de las medidas de probabilidad. Dado que $E \cup \overline{E} = Outc$ y E y $\overline{E}$ son disjuntos, las condiciones anteriores implican que

$$Pr(E) = 1 - Pr(\overline{E}).$$

En particular, $Pr(\emptyset) = Pr(Outc) = 1 - Pr(Outc) = 1 - 1 = 0$.

Las medidas de probabilidad son monótonas, es decir, para eventos E y E' tales que $E \subseteq E'$, se cumple que

$$Pr(E') = Pr(E') + Pr(E' \setminus E) \geq Pr(E).$$

Además, si $(E_n)_{n=1}^{\infty}$ es una familia de eventos, posiblemente no disjuntos entre sí, entonces:

$$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} E'_n,$$

donde $E'_1 = E_1$ y $E'_n = E_n \setminus (E_1 \cup \ldots \cup E_{n-1})$ para $n \geq 2$. Dado que $E'_n \cap E'_m = \emptyset$ para $n \neq m$, tenemos que:

$$Pr\left(\bigcup_{n=1}^{\infty} E_n\right) = Pr\left(\bigcup_{n=1}^{\infty} E'_n\right) = \sum_{n=1}^{\infty} Pr(E'_n).$$

Si $E_1 \subseteq E_2 \subseteq E_3 \subseteq \cdots$ y E'_n es como se indicó antes, entonces tenemos $E'_n = E_n \setminus E_{n-1}$ para $n \geq 2$, lo que resulta en:

$$Pr\left(\bigcup_{n=1}^{\infty} E_n\right) = Pr(E_1) + \sum_{n=2}^{\infty}(Pr(E_n) - Pr(E_{n-1})) =$$
$$= \lim_{N\to\infty}\left(Pr(E_1) + \sum_{n=2} NPr(E_n) - Pr(E_{n-1})\right) = \lim_{N\to\infty} Pr(E_N).$$

Este límite existe y concuerda con el supremo de $\{Pr(E_1), Pr(E_2), \ldots\}$, ya que la monotonía de Pr implica que $Pr(E_1) \leq Pr(E_2) \leq \cdots \leq 1$. Para intersecciones numerables se aplican resultados análogos. Es decir, si $E_1 \supseteq E_2 \supseteq \ldots$, entonces

$$Pr\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} Pr(E_n) = \inf_{n\geq 1} Pr(E_n).$$

Esto se sigue del hecho de que la secuencia $(\overline{E_n})_{n\geq 1}$ de los complementos $\overline{E_n} = \text{Outc} \setminus E_n$ es decreciente. Por lo tanto, obtenemos:

$$Pr\left(\bigcap_{n=1}^{\infty} E_n\right) = 1 - Pr(\overline{\bigcup_{n\geq 1} \overline{E_n}}) = 1 - \lim_{n\to\infty} Pr(\overline{E_n}) = 1 - \lim_{n\to\infty} Pr(\overline{E_n}) = \lim_{n\to\infty} Pr(E_n).$$

Cualquier evento E con $Pr(E) = 1$ se dice que ocurre **casi seguramente**. Nótese que si E ocurre casi seguramente, entonces $Pr(D) = Pr(E \cap D)$ para todos los eventos D, ya que $Pr(D \setminus E) = 0$, dado que $D \setminus E$ es un subconjunto de $\overline{E}$ y $Pr(\overline{E}) = 1 - Pr(E) = 1 - 1 = 0$. Esto implica que:

$$Pr(D) = Pr(E \cap D) + Pr(D \setminus E) = Pr(E \cap D).$$

En particular, el evento $E_1 \cap E_2$, de eventos E_1 y E_2 que ocurren casi seguramente, también ocurre casi seguramente. Como se puede demostrar por inducción, esto se extiende a cualquier evento que pueda escribirse como una intersección finita $\bigcap_{1\leq i\leq n} E_i$

de eventos $E_1, \dots, E_n$ que ocurren casi seguramente. Al tomar el límite de tales intersecciones finitas, obtenemos que $\bigcap_{i \geq 1} E_i$ ocurre casi seguramente si $Pr(E_i) = 1$ para todo $i \geq 0$.

Para cada conjunto Outc y cada subconjunto Π de 2^{Outc}, existe la σ-álgebra más pequeña que contiene a Π. Esto se debe a las siguientes observaciones:

- El conjunto potencia 2^{Outc} de Outc es una σ-álgebra, y
- La intersección de σ-álgebras es una σ-álgebra.

La intersección $\mathfrak{E}_\Pi = \bigcap_{\mathfrak{E}} \mathfrak{E}$, donde $\mathfrak{E}$ varía sobre todas las σ-álgebras sobre Outc que contienen a Π, es una σ-álgebra y está contenida en cualquier σ-álgebra $\mathfrak{E}$ tal que $\Pi \subseteq \mathfrak{E}$. A esta σ-álgebra $\mathfrak{E}_\Pi$ se la llama la **σ-álgebra generada por** Π, y Π es la **base** para $\mathfrak{E}_\Pi$.

2.1.2. Kucera

Espacios de probabilidad

Sea A un conjunto finito o numerablemente infinito. Una **distribución de probabilidad** sobre A es una función $\mu : A \to [0, 1]$ tal que $\sum_{a \in A} \mu(a) = 1$.

Una distribución μ es **racional** si $\mu(a)$ es racional para cada $a \in A$, positiva si $\mu(a) > 0$ para cada $a \in A$, y **de Dirac** si $\mu(a) = 1$ para algún $a \in A$. Una distribución de Dirac μ donde $\mu(a) = 1$ también se denota por μ_a o simplemente a.

Sea Ω un conjunto de eventos elementales. Una **σ-álgebra** sobre Ω es un conjunto $\mathcal{F} \subseteq 2^\Omega$ que incluye Ω y está cerrado bajo complementos y uniones numerables. Un **espacio medible** es un par $(\Omega, \mathcal{F})$ donde $\mathcal{F}$ es una σ-álgebra sobre Ω. Una función **$\mathcal{F}$-medible** sobre $(\Omega, \mathcal{F})$ es una función $f : \Omega \to \mathbb{R}$ tal que $f^{-1}(I) \in \mathcal{F}$ para cada intervalo I en $\mathbb{R}$.

Una **medida de probabilidad** sobre un espacio medible $(\Omega, \mathcal{F})$ es una función $\mathcal{P} : \mathcal{F} \to [0, 1]$ tal que, para cada colección numerable $\{A_i\}_{i \in I}$ de elementos disjuntos de a pares de $\mathcal{F}$, tenemos que $\mathcal{P}\left(\bigcup_{i \in I} A_i\right) = \sum_{i \in I} \mathcal{P}(A_i)$, y además $\mathcal{P}(\Omega) = 1$.

Un **espacio de probabilidad** es una terna donde $(\Omega, \mathcal{F})$ es un espacio medible y $\mathcal{P}$ es una medida de probabilidad sobre $(\Omega, \mathcal{F})$.

Una **variable aleatoria** sobre un espacio de probabilidad $(\Omega, \mathcal{F}, \mathcal{P})$ es una función $\mathcal{F}$-medible $X : \Omega \to \mathbb{R}$. El **valor esperado** de X, denotado por $E(X)$, se define como la integral de Lebesgue $\int_\Omega X \, d\mathcal{P}$.

Una **variable aleatoria** X **es discreta** si existe un subconjunto finito o numerablemente infinito N de $\mathbb{R}$ tal que $P(X^{-1}(N)) = 1$. El valor esperado de una variable aleatoria discreta X es igual a $\sum_{n \in N} n \cdot \mathcal{P}(X = n)$, donde $X = n$ denota el conjunto $X^{-1}(\{n\})$.

σ-álgebra de Borel

Sea $T = (S, \to)$ un sistema de transiciones. Sea $\mathcal{B}$ la menor σ-álgebra sobre *Run* que contiene todos los cilindros básicos $\mathit{Run}(w)$ donde $w \in \mathit{Fpath}$ (es decir, $\mathcal{B}$ es la σ-álgebra de Borel generada por conjuntos abiertos en la topología de Cantor sobre *Run*). Entonces, $(\mathit{Run}, \mathcal{B})$ es un espacio medible, y los elementos de $\mathcal{B}$ se llaman **conjuntos de Borel de ejecuciones**.

La **σ-álgebra de Borel** $\mathcal{B}$ contiene muchos elementos interesantes. Por ejemplo, sea $s, t \in S$ y sea $\mathit{Reach}(s, t)$ el conjunto de todas las ejecuciones iniciadas en s que visitan t. Obviamente, $\mathit{Reach}(s, t)$ es la unión de todos los cilindros básicos $\mathcal{R}\text{un}(w)$ donde $w \in \text{Fpath}(s)$ y w visita t, y por lo tanto $\text{Reach}(s, t) \in \mathcal{B}$.

De manera similar, se puede demostrar que el conjunto de todas las ejecuciones iniciadas en s que visitan t infinitamente a menudo es Borel. En realidad, la mayoría de los conjuntos 'interesantes' de ejecuciones son de Borel, aunque también existen subconjuntos de *Run* que no están en $\mathcal{B}$.

Sea $A \in \mathcal{B}$, y sea $f : \mathit{Run} \to \{0, 1\}$ una función que asigna a un dado $w \in \mathcal{R}\text{un}$ 1 o 0, dependiendo de si $w \in A$ o no, respectivamente. Entonces, f es $\mathcal{B}$-medible, porque para cada intervalo I en $\mathbb{R}$ tenemos que $f^{-1}(I)$ es igual a $\mathcal{R}\text{un}$, A, $\mathcal{R}\text{un} \setminus A$ o $\emptyset$, dependiendo de si $I \cap \{0, 1\}$ es igual a $\{0,\ 1\}$, $\{1\}$, $\{0\}$, o $\emptyset$, respectivamente.

2.2. Rabin

Capítulo 3

Verificación de modelos con juegos

3.1. Sobre la verificación de modelos

blabla

problema de sintesis de church

idea de la necesidad de juegos y nuestro entendimiento de ellxs

¿Que preguntas nos hacemos cuando trabajamos con juegos? Mas que nada con juegos estocasticos tenemos las preguntas que se hace Kucera, las que presenta Chatterjee y se puede investigar mas Capaz esto iría mas en la introducción para platear la idea de trabajo

3.2. Cadenas de Markov

Definición 3.2.1 (Cadena de Markov)**.** *Una cadena de Markov es una tupla* $M = (S, P)$ *donde* S *es un conjunto (finito ?) no vacío de estados y* $P : S \times S \to [0, 1]$ *es una función* *tal que para todos los estados s vale que*

$$\sum_{s' \in S} P(s, s') = 1$$

La función de probabilidad de transición P especifica para cada estado s la probabilidad $P(s, s')$ de moverse de s a s' en un solo paso. La restricción impuesta en P asegura que la función sea una distribución.

Una cadena de Markov induce un grafo subyacente, donde los estados actúan como vértices y hay una arista entre s y s' si y solo si $P(s, s') > 0$. Las cadenas de Markov se suelen representa por su grafo subyacente donde sus aristas estarán anotadas con las probabilidades en el intervalo $(0, 1]$.

Los caminos en una cadena de Markov son los caminos en el grafo subyacente. Son definidos como secuencias infinitas de estados $\omega = (s_0, s_1, s_2, \dots) \in S^\omega$ tales que $P(s_i, s_{i+1}) > 0$ para todo $i \geq 0$.

Veamos un pequeño ejemplo de cómo sería una cadena de Markov.

Ejemplo 3.2.1. *Supongamos que queremos observar el comportamiento de un pequeño robot llamado Roberto. Roberto se comporta probabilísticamente de la siguiente manera: desde su posición inicial de quietud tiene una probabilidad de …*

Para poder asociar probabilidades a eventos en cadenas de Markov, la noción intuitiva de probabilidades en M es formalizada al asociarle un espacio de probabilidad (2.1). Los caminos infinitos de M juegan el rol de resultados. Esto es $Outc^M = Paths(M)$. La σ-álgebra asociada con M es generada pot los conjuntos cilindro formados por los fragmentos de caminos finitos en M.

Me falta la noción de prefijo.

Definición 3.2.2 (Conjunto cilindro)**.** *El conjunto cilindro de $\hat{\omega} = (s_0, \dots, s_n) \in Paths_{fin}(M)$ está definido como*

$$Cyl(\hat{\omega}) = \{\omega \in Paths(M) \mid \hat{\omega} \in pref(\omega)\}$$

Definición 3.2.3 (σ-álegra de una cadena de Markov). *La σ-álgebra $\mathfrak{E}^M$ asociada a la cadena de Markov M es la σ-álgebra más pequeña que contiene todos los conjuntos cilindro $Cyl(\hat{\omega})$ donde $\hat{\omega} \in Paths_{fin}(M)$.*

De conceptos clásicos de teoría de probabilidad se sigue que por cada estado s existe una única medida de probabilidad $\mathbb{P}_s^M$ en la σ-álgebra $\mathfrak{E}^M$ asociada a M, donde las probabilidades para los conjunto cilindros (es decir, los eventos) están dadas por:

$$\mathbb{P}_s^M(Cyl(s_0, \ldots, s_n)) = \iota(s, s_0) \cdot \prod_{0 \leq i < n} P(si, s_{i+1}), \text{ donde } \iota(s, s_0) = \begin{cases} 1 & \text{si } s = s_0 \\ 0 & \text{en otro caso} \end{cases} \tag{3.1}$$

No estoy muy segura de cómo manejar lo de estado inicial.

Notación: en lo que sigue, usaremos notación LTl para describir ciertos eventos en cadenas de Markov. Por ejemplo, para un conjunto $B \subseteq S$ de estados, $\Diamond B$ denota el evento de llegar eventualmente a (algún estado en) B, mientras que *siempeventB* describe el evento en el que B es visitado infinitamente a menudo. Algo más? sí

EJEMPLO DE ROBOT y la pregunta de alcanzabilidad.

3.3. Procesos de Decisión de Markov

Un proceso de decisión de Markov (MDP, por sus siglas en inglés) es una generalización de una cadena de Markov donde un conjunto de acciones posibles es asociado a cada estado. A cada par estado-acción le corresponde una distribución de probabilidad en los estados, que es usada para seleccionar el próximo estado. A su vez, una cadena de Markov se corresponde a un MDP donde hay exactamente una acción asociada a cada estado. Asumiremos la existencia de un conjunto fijo de acciones *Act*. La definición de un proceso de decisión de Markov es como sigue:

Definición 3.3.1 (Proceso de Decisión de Markov). *Un proceso de decisión de Markov $\mathcal{M} = (S, A, \theta)$ consiste de un conjunto finito de estados S y de dos componentes A y θ que especifican la estructura de transición:*

- *A es un conjunto de acciones. Para cada $s \in S$, $A(s) \subseteq Acts$ es el conjunto finito no vacío de acciones disponibles en s. Para cada estado $s \in S$ se requiere que $A(s) \neq \emptyset$.*

- *$\theta : S \times A \times S \to [0,1]$ es una función de transición probabilística. Para todo estado $s \in S$, si $a \in A(s)$ tenemos que $\sum_{s' \in S} \theta(s,a,s') = 1$, mientras que si $a \notin A(s)$, $\sum_{s' \in S} \theta(s,a,s') = 0$. Para cada $s, t \in S$ y $a \in A(s)$, $\theta(s,a,t)$ es la probabilidad de transicionar de s a t cuando la acción a es seleccionada.*

Agregar ejemplo de robot

Un comportamiento en un proceso de decisión de Markov es una secuencia alternante infinita de estados y acciones, construida iterativamente por un proceso de dos pasos. Primero, dado un estado s, una acción $a \in A(s)$ es seleccionada no-determinísticamente. Luego, el sucesor t de s es seleccionado de acuerdo a la distribución asociada a la acción a. La definición formal es como sigue:

Definición 3.3.2 (Comportamiento en un MDP)**.** *Un comportamiento en un MDP $\mathcal{M}$ es una secuencia infinita $\omega = (s_0, a_0, s_1, a_1, \dots)$ tal que $s \in S$, $a_i \in A(s_i)$ y $a_i(s_{i+1}) > 0$ para todo $i \geq 0$.*

Dado un estado s, indicaremos con Ω_s el conjunto de todos los comportamientos que se originan en s, con $\Omega_{\mathcal{M}}$ el conjunto de todos los comportamientos en $\mathcal{M}$ y con $\Omega_{\mathcal{M}}^{fin}$ el conjunto de todos los prefijos finitos de comportamientos en $\mathcal{M}$ que terminan en un estado $s \in S$.

Para cadenas de Markov, el conjunto de caminos está equipado con una σ-álgebra y una medida de probabilidad que refleja la noción intuitiva de probabilidad para conjuntos (medibles) de caminos. Para los MDPs, esto es levemente distinto. Como no hay restricciones en la resolución de las elecciones no determinista, no hay una única medida de probabilidad asociada. (capaz acá podría ir algo tipo un ejemplo como el de la página 841)

Para poder razonar sobre probabilidades de conjuntos de comportamientos en un MDP necesitamos resolver de alguna manera el no determinismo, y para ello introduciremos el concepto de estrategia.

Definición 3.3.3 (Estrategia en un MDP)**.** *Sea $\mathcal{M} = (S, A, \theta)$ un MDP. Una estrategia para $\mathcal{M}$ es una función $\pi : \Omega^{fin} \to \mathsf{Dist}(A)$ que asigna una distribución de probabilidad a cada prefijo finito de comportamiento tal que $\pi(\hat{\omega})(a) > 0$ solo si $a \in A(s)$.*

Comentar algo de que en la literatura se suelen bajar las acciones de la definición de estrategia? O capaz la idea sería dropear las acciones de la definición de estrategia acá?

Lo que está a continuación me hace un poco de ruido por la parte de la definición de la cadena de Markov. Es como lo presentan en el libro de Baier y Katoen y tiene esta idea de ir armando una continuidad entre MCs y MDPs en la explicación. La otra opción es ir poco más con el enfoque de Luca de Alfaro y presentar la σ-álgebra sin mencionar nada de MDPs.

Como una estrategia resuelve todas las elecciones no deterministas en un MDP, induce una cadena de Markov. Esto es, el funcionamiento de un MDP $\mathcal{M}$ siguiendo las decisiones de una estrategia π puede ser formalizado por una cadena de Markov $\mathcal{M}_\pi$, donde los estados son los prefijos finitos de comportamientos en $\mathcal{M}$.

Definición 3.3.4 (Cadena de Markov de un MDP inducida por una estrategia)**.** *Sea $\mathcal{M} = (S, A, \theta)$ un MDP y π una estrategia en $\mathcal{M}$. La cadena de Markov $\mathcal{M}_\pi$ está dada por*

$$\mathcal{M}_\pi = (\Omega^{fin}_{\mathcal{M}}, P_\pi)$$

donde para $\hat{\omega} = (s_0, a_0, s_1, a_1, \ldots, s_n)$:

$$P_\pi(\hat{\omega}, \hat{\omega} a s_{n+1}) = \pi(\hat{\omega})(a) \cdot \theta(s_n, a, s_{n+1})$$

Nótese que $\mathcal{M}_\pi$ cuenta con un espacio de estados infinito aun, aun cuando el MDP $\mathcal{M}$ es finito. Entonces cuenta como una MC?.

Como $\mathcal{M}_\pi$ es una cadena de Markov, uno ahora puede razonar sobre las probabilidades de los conjuntos medibles de caminos que siguen la estrategia π, simplemente usando las distintas medidas de probabilidad $\mathbb{P}_s^{\mathcal{M}_\pi}$ asociadas a la cadena de Markov $\mathcal{M}_\pi$ (véase 3.1).

Responder preguntas de probabilidad de cosas en MDP robot

Intuitivamente, el estado $(s_0, a_0, \ldots, s_n)$ de $\mathcal{M}_\pi$ representa la configuración donde el MDP $\mathcal{M}$ está en el estado s_n y cuenta con la historia $(s_0, a_0, \cdots, s_{n-1}, a_{n-1})$. Según la definición que vimos las estrategias pueden depender de la historia en su totalidad, produciendo resultados distintos si al menos una acción o estado en su historia cambia, pero es cierto que este caso no es lo usual. Veamos algunos tipos particulares de estrategias donde esto no sucede.

Definición 3.3.5 (Estrategias sin memoria)**.** *Sea $\mathcal{M}$ un MDP con espacio de estados S. Una estrategia π en $\mathcal{M}$ es sin memoria sii para cada par de comportamientos $(s_0, a_0, \ldots, s_n)$ y $(t_0, a'_0, \ldots, t_m)$ con $s_n = t_m$ vale que:*

$$\pi(s_0, a_0, \ldots, s_n) = \pi(t_0, a'_0, \ldots, t_m)$$

En este caso, π puede ser vista como una función $\pi : S \to \mathsf{Dist}(A)$.

Coloquialmente, una estrategia es sin memoria si no recuerda nada de la historia y solo elige probabilidades para las acciones basándose en el estado actual. Esto puede ser bastante extremo en ciertos casos, por eso existe una variante que busca reflejar la idea de finitud sin ser tan restrictiva: las estrategias de memoria finita. Una estrategia de memoria finita puede ser pensada intuitivamente como que solo puede guardar hasta una cantidad finita fija de información de la historia, por lo que no podrá ser distinta para **todo** prefijo finito de comportamiento. Formalmente, la definiremos a través de una autómata determinista finito (DFA). La distribución de probabilidad de las acciones será seleecionada a partir del estado actual en $\mathcal{M}$ y el estado actual del autómata (al que llamaremos modo). Veamos su definición:

Definición 3.3.6 (Estrategias con memoria finita)**.** *Sea $\mathcal{M}$ un MDP con espacio de estados S y conjunto de acciones A. Una estrategia de memoria finita para $\mathcal{M}$ es una tupla $\pi = (Q, f_\pi, \Gamma, start)$ donde*

- *Q es un conjunto finito de modos,*
- *$\Delta : Q \times A \times S \to Q$ es la función de transición del autómata,*
- *$start : S \to Q$ es la función que determina el modo en el que empieza el automáta para un estado inicial s,*

- $f_\pi : Q \times S \to \mathsf{Dist}(A)$ *es la función que asigna la distribución de probabilidad en las acciones desde un estado s, es decir, lo que veníamos entendiendo como estrategia en sí.*

El funcionamiento del MDP bajo la estrategia de memoria finita sería como sigue. En principio, se inicializa el modo del DFA a $q_0 = start(s_0)$. Luego, desde cada estado s_i posterior el proceso será iterativo. Primero, se seleccionará la distribución de probabilidad en las acciones a partir del modo actual q_i del autómata con $f_\pi((q_i, s_i))$. Una vez tomada la decisión, se determina probabilísticamente la siguiente acción a_{i+1}, y, a partir de ella, se determina también probabilísticamente el siguiente estado s_{i+1}. Con la nueva acción y estado se selecionará el próximo modo del DFA $q_{i+1} = \Delta(q_i, a_{i+1}, s_{i+1})$ y se repetirá el proceso.

Para $\hat{\omega} \in \Omega^{fin}_{\mathcal{M}}$, notaremos con $\pi(\hat{\omega})$ a la distribución obtenida al realizar el proceso explicado anteriormente con $\hat{\omega}$ pero qué pasa con la elección probabilística de las acciones? No habría que tenerlo? Y que hay de sin memoria? Ahi en realidad entonces no tendria que ser ultimo estado y penultima accion?.

Además de su categorización en base a qué tanto dependen de su historia, existe otro tipo notable de estrategias: las estrategias deterministas. Una estrategia es determinista cuando para cada $\hat{\omega} \in \Omega^{fin}_{\mathcal{M}}$ la distribución $\pi(\hat{\omega})$ es una distribución de Dirac (es decir, una distribución δ_a tal que $\delta_a(a) = 1$ y $\delta_a(b) = 0$ para todo $b \neq a$). En la literatura (véase [1, 2]) es usual encontrarse con el estudio de estrategias solo en su variante determinista y, en ese caso, se puede pensar a las estrategias como una función $\pi : \Omega^{fin}_{\mathcal{M}} \to A$.

Mencionar de que tipos son las estrategias vistas en el ejemplo de robot

3.4. Juegos estocásticos

¿textito de conexión con MDPs?

Definir cómo queda con lo de acciones habilitadas esto o MDP

Definición 3.4.1 (Juego estocástico)**.** *Un juego estocástico (SG) es una tupla $\mathcal{G} = (\mathcal{S}, (\mathcal{S}_\square, \mathcal{S}_\Diamond), \mathcal{A}, \theta)$ donde:*

- *$\mathcal{S}$, un conjunto finito de estados con $\mathcal{S}_\Box, \mathcal{S}_\Diamond \subseteq \mathcal{S}$ siendo una partición de él,*
- *$\mathcal{A}$ es un conjuno de acciones, y*
- *$\theta : \mathcal{S} \times \mathcal{A} \times S \to [0,1]$ es una función de transición probabilística tal que para cada $s \in \mathcal{S}$, $\theta(s, a, \cdot) \in \mathsf{Dist}(\mathcal{S})$ o $\theta(s, a, \mathcal{S}) = 0$.*

Notaremos con $\mathcal{A}(s) = \{a \in \mathcal{A} \mid \theta(s, a, \mathcal{S}) = 1\}$ al conjunto de acciones habilitadas en s, y

Se puede notar que si $S_\Box = \emptyset$ o $\S_\Diamond = \emptyset$, entonces $\mathcal{G}$ es un proceso de decisión de Markov, y, como vimos antes, si a la vez $\left|\mathcal{A}(s)\right| = 1$ entonces $\mathcal{G}$ es una cadena de Markov.

Otra manera de verlo es que podemos pensar a los juegos estocásticos como un procesos de decisión de Markov donde el conjunto de estados está particionado en dos: los estados correspondientes al jugador $\Box$ y los estados correspondientes al jugador $\Diamond$. La idea de esta anotación de a quién pertence los estados es lo que introduce esta idea adversarial: no solo habrá un ente que tome las decisiones no deterministas, sino dos. Esto quiere decir que tendremos dos tipos de estrategias, una por cada tipo de jugador.

Al igual que con los MDPs, antes de definir las estrategias, presentamos el concepto de comportamiento en un juego estocástico.

Definición 3.4.2 (Comportamiento en un SG)**.** *Un comportamiento en un SG $\mathcal{G}$ es una secuencia infinita $\omega = (s_0, a_0, s_1, a_1, \dots)$ tal que $s_i \in \mathcal{S}$, $a_i \in \mathcal{A}(s_i)$ y $a_i(s_{i+1}) > 0$ para todo $i \geq 0$.*

Dado un estado s, indicaremos con Ω_s el conjunto de todos los comportamientos que se originan en s, con $\Omega_\mathcal{G}$ el conjunto de todos los comportamientos en $\mathcal{G}$ y con $\Omega^{fin}_{\mathcal{G},i}$ el conjunto de todos los prefijos finitos de comportamientos en $\mathcal{G}$ que terminan en un estado $s \in \mathcal{S}_i$, con $i \in \{\Box, \Diamond\}$.

Definición 3.4.3 (Estrategia en un juego estocástico)**.** *Sea $\mathcal{G} = (\mathcal{S}, (\mathcal{S}_\Box, \mathcal{S}_\Diamond), \mathcal{A}, \theta)$ un SG. Una estrategia π_i para el jugador i en $\mathcal{G}$ es una función $\pi : \Omega^{fin}_i \to \mathsf{Dist}(A)$ que asigna una distribución de probabilidad a cada prefijo finito de comportamiento que termina en un estado del jugador i tal que $\pi(\hat{\omega})(a) > 0$ solo si $a \in A(s)$.*

Llamaremos $\Pi_\Box$ al conjunto de todas las estrategias del jugador $\Box$ y $\Pi_\Diamond$ al conjunto de todas las estrategias del jugador $\Diamond$.

Podemos ver que las estrategias en un juego estocástico se definen igual a cómo se las definen para procesos de decisión de Markov con la salvedad de que pertenecerán a un jugador específico. Los distintos tipos de estrategias presentadas en la sección anterior se extienden naturalemente a SG. Ver redacción. Llamaremos Π_i^M al conjunto de las estrategias sin memoria del jugador i, Π_i^F al conjunto de las estrategias de memoria finita del jugador i, Π_i^D al conjunto de las estrategias deterministas del jugador i, Π_i^{MD} al conjunto de las estrategias deterministas sin memoria del jugador i y Π_i^{FD} al conjunto de las estrategias deterministas con memoria finita del jugador i.

De manera similar a como lo razonamos para procesos de decisión de Markov, si fijamos dos estrategias $\pi_\Box \in \Pi_\Box$ y $\pi_\Diamond \in \Pi_\Diamond$ en un juego estocástico $\mathcal{G}$ obtenemos una cadena de Markov a la que denotaremos $\mathcal{G}^{\pi_\Box,\pi_\Diamond}$. Esta cadena de Markov, para cada $s \in \mathcal{S}$ definirá una medida de probabilidad $\mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}$ en la σ-álgebra de Borel del conjunto de comportamientos en $\mathcal{G}$. ¿Es comportamientos?. Si ε es un conjunto medible en la σ-álgebra de Borel, $\mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\varepsilon)$ será la probabilidad de que las estrategias $\pi_\Box$ y $\pi_\Diamond$ sigan un comportamiento en ε empezando desde el estado s.

Tal como fue aclarado en las dos secciones anteriores, usaremos notación LTL para representar conjunto específicos de comportamientos. Sin embargo, estando en el análisis de juegos nos interesan hacernos preguntas sobre probabilidades de ganar. Para poder hablar de ellas necesitamos formalizar ciertos conceptos que presentamos a continuación.

Valor en un juego y maneras de ganar

Determinismo

Objetivos en juegos estocásticos

Lo primero que será importante notar es que vamos a trabajar con juegos de suma cero.

Recompensas cualitativas

Sea $C = \{c_1, \ldots, c_n\}$ un conjunto finito de colores, y $\nu : V \to 2^C$ una valuación. Una clase importante y bien estudiada de recompensas cualitativas son las **funciones**

características de subconjuntos ω-regulares de *Run*(G).

La pertenencia a un conjunto ω-regular de ejecuciones se determina por una de las **condiciones de aceptación** listadas a continuación. Estas condiciones corresponden a criterios de aceptación de autómatas de estado finito sobre palabras infinitas (ver Sección 5.2.2 para más detalles).

- **Alcanzabilidad y seguridad (Reachability and safety).** Una ejecución $w \in Run(G)$ satisface la condición de **alcanzabilidad** determinada por un color $c \in C$ si $c \in \nu(w(i))$ para algún $i \geq 0$. La condición de **seguridad** determinada por c es dual, es decir, $c \notin \nu(w(i))$ para todo $i \geq 0$.

- **Büchi y co-Büchi.** Una ejecución $w \in Run(G)$ satisface la condición de **Büchi** determinada por un color $c \in C$ si $c \in \nu(w(i))$ para infinitos $i \geq 0$. La condición de **co-Büchi** es dual, es decir, solo hay un número finito de $i \geq 0$ tal que $c \in \nu(w(i))$.

- **Rabin, Rabin-chain y Street.** Sea *Pairs* un conjunto finito de pares de colores $\{(c_1, d_1), \ldots, (c_m, d_m)\}$. Una ejecución $w \in Run(G)$ satisface la condición de **Rabin** determinada por *Pairs* si existe $(c, d) \in Pairs$ tal que w satisface la condición de Büchi determinada por d y la condición de co-Büchi determinada por c.

  La condición de **Street** determinada por *Pairs* es dual a Rabin, es decir, para cada $(c, d) \in Pairs$ tenemos que w satisface la condición de co-Büchi determinada por d ó la condición de Büchi determinada por c.

  Para un color dado c, sea $V(c)$ el conjunto de todos los $v \in V$ tales que $c \in \nu(v)$. La condición de **Rabin-chain** (o paridad) es un caso especial de la condición de Rabin donde *Pairs* y ν satisfacen $V(c_1) \subset V(d_1) \subset \cdots \subset V(c_m) \subset V(d_m)$.

- **Muller.** Sea $M \subseteq 2^C$ un conjunto de subconjuntos de colores. Una ejecución $w \in Run(G)$ satisface la condición de **Muller** determinada por M si el conjunto de todos los $c \in C$ tales que w satisface la condición de Büchi determinada por c es un elemento de M. (En criollo, “los conjuntos de colores que w visita infinitas veces están en M”)

Nótese que los conjuntos ω-regulares de ejecuciones son relativamente simples en el sentido de que están contenidos en los dos primeros niveles de la jerarquía de Borel (los conjuntos de ejecuciones que satisfacen las condiciones de alcanzabilidad y seguridad están en el primer nivel).

distintos tipos y organizacion / ver de que manera presentarlos

Kucera, omega reg, Chatterjee

3.4.1. Una breve historia sobre los juegos estocásticos

Enfoque mc ->mdp ->sg ->psg

Enfoque juegos deterministas ->juegos estocasticos (Shapley etc)

3.5. Juegos deterministas

Agg transiciones de texto escrito entre definiciones

Definición 3.5.1 (juego de grafo de 2 jugadores)**.** *Definimos a un juego de grafo de dos jugadores (2G) como una tupla* $G = (V, V_\Box, V_\Diamond, E)$ *donde* $V = V_\Box \uplus V_\Diamond$ *es un conjunto de vértices (o estados) particionado en* $V_\Box$ *y* $V_\Diamond$*, y* $E \subseteq (V \times V)$ *es una relación que denota el conjunto de aristas (dirigidas) que representan transiciones de un estado a otro del juego.*

Los 2 jugadores son llamados $\Box$ *y* $\Diamond$ *y controlan los vértices* $V_\Box$ *y* $V_\Diamond$*, respectivamente.*

Definición 3.5.2 (estrategia sobre un 2G)**.** *Una estrategia para un jugador* i *con* $i \in \{\Box, \Diamond\}$ *es una función* $\sigma_i : V^*V_i \to V$ *con la restricción de que* $\sigma_i(wv) \in E(v)$ *para todo* $wv \in V^*V_i$*.*

Definición 3.5.3 (jugada sobre un 2G)**.** *Una jugada en un juego de grafo de dos jugadores es una secuencia infinita de vértices* $\rho = v_0v_1v_2\cdots \in V^\omega$*, donde para todo* $i \in \mathbb{N}_0$ *tenemos que* $v^i \in V$ *y* $(v^i, v^{i+1}) \in E$*.*

Sean $\sigma_\Box$ *y* $\sigma_\Diamond$ *un par de estrategias y sea* v_0 *un vértice inicial, la jugada que cumple con* $\sigma_\Box$ *y* $\sigma_\Diamond$ *es la única jugada* $\rho = v_0v_1v_2\ldots$ *para la cual cada* $i \in \mathbb{N}_0$*, si* $v_i \in V_\Box$ *entonces* $v_{i+1} = \sigma_\Box(v_0 \ldots v_i)$*, y si* $v_i \in V_\Diamond$ *entonces* $v_{i+1} = \sigma_\Diamond(v_0 \ldots v_i)$*.*

Definición 3.5.4 (condición ganadora)**.** *Una condición ganadora* φ *en un juego de grafo de dos jugadores es un conjunto de jugadas sobre el juego, i.e.,* $\varphi \subseteq V^\omega$*. Usaremos notación LTL para describir conjuntos de jugadas específicos.*

Definición 3.5.5 (regiones ganadoras)**.** *El jugador* $\Box$ *gana el juego de grafo de dos jugadores* G *para una condición ganadora* φ *desde un vértice* v_0 *so existe una estrategia* $\pi_\Box$ *tal que para cada* $\pi_\Diamond$*, la jugada* ρ *que sigue* $\pi_\Box$ *y* $\pi_\Diamond$ *satisface* φ*, i.e.,* $\rho \in \varphi$*. La región ganadora* $\mathcal{W} \subseteq V$ *para el jugador* $\Box$ *es el conjunto de vértices desde donde el jugador* $\Box$ *gana el juego.*

Agg ejemplo

3.6. Comparación de los distintos juegos

Incluir cuadros comparativos entre juegos // ejemplos para todas las nociones nuevas explicadas

(Capaz agregar un apéndice con cuadritos de esto ?)

(estaría bueno ver problemas y agregar un apéndice con ejemplos?)

Capítulo 4

Juegos Estocásticos Politópicos - PSGs

4.1. Definiciones

4.1.1. Politopos

Un **politopo convexo** en $\mathbb{R}^n$ es un conjunto acotado $K = \{x \in \mathbb{R}^n | Ax \leq b\}$ con $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$, para algún $m \in \mathbb{N}$.

Por acotado nos referimos a que $\exists\ M \in \mathbb{R}_{\geq 0}$ tal que $\sum_{i=1}^{n} |x_i| \leq M\ \forall x \in K$, y siendo S un conjunto finito, como las funciones en $\mathbb{R}^S$ pueden ser vistas como vectores en $\mathbb{R}^{|S|}$, generalmente, nos referiremos a politopos en $\mathbb{R}^S$.

Sea $\mathbf{Poly}(S)$ el conjunto de todos los politopos convexos en $\mathbb{R}^S$. Nótese que el conjunto de todas las funciones de probabilidad en S forma el politopo convexo

$\mathrm{Dist}(S) = \{\mu \in \mathbb{R}^S \mid \sum_{s \in S} \mu(s) = 1 \text{ y } \forall s \in S : \mu(s) \geq 0\}$

Definimos $\mathbf{DPoly}(S) = \{K \cap \mathrm{Dist}(S) \mid K \in \mathrm{Poly}(S)\}$ (el conjunto de los politopos formados por distribuciones de probabilidad sobre S?). Cada $K \in \mathrm{DPoly}(S)$ es un politopo convexo cuyos elementos son también funciones de probabilidad sobre S, y su conjunto de desigualdades característico $Ax \leq b$ ya codifica las desigualdades $\sum_{s \in S} x_s =$

1 y $x_s \geq 0$ para todo $s \in S$.

4.1.2. Juegos estocásticos

Un juego estocástico es una tupla $\mathcal{G} = (\mathcal{S}, (\mathcal{S}_\Box, \mathcal{S}_\Diamond), \mathcal{A}, \theta)$, donde $\mathcal{S}$ es un conjunto finito de estados, con $\mathcal{S}_\Box, \mathcal{S}_\Diamond \subseteq \mathcal{S}$ siendo una partición de $\mathcal{S}$, y $\theta : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$ es una función de transición probabilística tal que para todo $s \in \mathcal{S}$ y $a \in \mathcal{A}$, $\theta(s, a, \cdot) \in \mathrm{Dist}(\mathcal{S})$ o $\theta(s, a, \mathcal{S}) = 0$ (dados un estado s y una acción a, θ define una distribución en s o no hay estados alcanzables por a desde s).

Sea $\mathcal{A}(s) = \{a \in \mathcal{A} \mid \theta(s, a, \mathcal{S}) = 1\}$ el conjunto de acciones habilitadas en el estado s. Si $\mathcal{S}_\Box = \emptyset$ o $\mathcal{S}_\Diamond = \emptyset$, entonces $\mathcal{G}$ es un proceso de decisión de Markov (o MDP). Si, además, $|\mathcal{A}(s)| = 1$ para todo $s \in \mathcal{S}$, $\mathcal{G}$ es una cadena de Markov (o MC).

Un camino en el juego $\mathcal{G}$ es una secuencia infinita de estados $\rho = s_0 s_1 \dots$ tal que $\theta(s_k, a, s_{k+1}) > 0$ para alguna $a \in \mathcal{A}$ y para todo $k \in \mathbb{N}$. Para $i \geq 0$, ρ_i indica el i-ésimo estado en el camino ρ (nótese que ρ_0 es el primer estado en ρ). Denotamos por $\mathrm{Paths}_\mathcal{G}$ el conjunto de todos los caminos, y por $\mathrm{FPaths}_\mathcal{G}$ el conjunto de los prefijos finitos de caminos. De manera similar, $\mathrm{Paths}_{\mathcal{G},s}$ y $\mathrm{FPaths}_{\mathcal{G},s}$ denotan el conjunto de caminos y el conjunto de caminos finitos que comienzan en el estado s.

Una estrategia para el jugador i (para $i \in \{\Box, \Diamond\}$) en un juego G es una función $\pi_i : \mathcal{S}^* \mathcal{S}_i \to \mathrm{Dist}(\mathcal{A})$ que asigna una distribución probabilística a cada secuencia finita de estados, tal que $\pi_i(\hat{\rho} s)(a) > 0$ solo si $a \in \mathcal{A}(s)$. El conjunto de todas las estrategias para el jugador i se denomina Π_i. Cuando sea conveniente, indicamos que el conjunto de estrategias Π_i pertenece al juego $\mathcal{G}$ escribiendo $\Pi_{\mathcal{G},i}$.

Una estrategia π_i se dice pura o determinista si, para todo $\hat{\rho} s \in \mathcal{S}^* \mathcal{S}_i$, $\pi_i(\hat{\rho} s)$ es una distribución de Dirac (es decir, una distribución δ_a tal que $\delta_a(a) = 1$ y $\delta_a(b) = 0$ para todo $b \neq a$), y se llama sin memoria si $\pi_i(\hat{\rho} s) = \pi_i(s)$, para todo $\hat{\rho} \in \mathcal{S}^*$. Sea Π_i^M el conjunto de todas las estrategias sin memoria para el jugador i, y Π_i^{MD} el conjunto de todas sus estrategias deterministas y sin memoria.

Dadas las estrategias $\pi_\Box \in \Pi_\Box$ y $\pi_\Diamond \in \Pi_\Diamond$, y un estado inicial s, el resultado del juego es una cadena de Markov denotada $\mathcal{G}_s^{\pi_\Box, \pi_\Diamond}$.

La cadena de Markov $\mathcal{G}_s^{\pi_\Box, \pi_\Diamond}$ define una **medida de probabilidad** $\mathbb{P}_{\mathcal{G},s}^{\pi_\Box, \pi_\Diamond}$ sobre la

σ-álgebra de Borel generada por los cilindros de $\text{Paths}_{\mathcal{G},s}$. Si ξ es un conjunto medible en dicha σ-álgebra de Borel, $\mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G},s}(\xi)$ es la probabilidad de que las estrategias $\pi_\Box$ y $\pi_\Diamond$ sigan un camino en ξ partiendo del estado s.

Usamos la notación LTL para representar conjuntos específicos de caminos, en particular: $DU^nC = \{\rho \in \mathcal{S}^\omega \mid \rho_n \in C \text{ y } \forall j < n : \rho_j \in D\} = D^n \times C \times \mathcal{S}^\omega$ es el conjunto de caminos que alcanzan $C \subseteq \mathcal{S}$ en exactamente $n \geq 0$ pasos, recorriendo antes solo estados en $D \subseteq \mathcal{S}$; $\Diamond^n C = \mathcal{S}U^nC$ es el conjunto de todos los caminos que alcanzan los estados en C en exactamente n pasos; $\Diamond C = \bigcup_{n\geq 0}(\mathcal{S} \setminus C)U^nC$ es el conjunto de todos los caminos que alcanzan un estado en C.

4.1.3. PSGs

Los juegos estocásticos politópicos fueron presentados en 2024 por Castro y D'Argenio [3]. La idea de su creción ...

Un juego estocástico politópico se caracteriza a través de una estructura que contiene un conjunto finito de estados divididos en dos conjuntos, cada uno perteneciente a un jugador diferente. Además, a cada estado se le asigna un conjunto finito de politopos convexos de distribuciones de probabilidad sobre los estados. La definición formal es como sigue:

Definición 4.1.1 (PSG). *Un juego estocástico politópico (abreviado PSG, por sus siglas en inglés) es una estructura $\mathcal{K} = (\mathcal{S}, (\mathcal{S}_\Box, \mathcal{S}_\Diamond), \Theta)$ tal que $\mathcal{S}$ es un conjunto finito de estados particionado en $\mathcal{S} = \mathcal{S}_\Box \biguplus \mathcal{S}_\Diamond$ y $\Theta : \mathcal{S} \to \mathcal{P}_f(\text{DPoly}(\mathcal{S}))$.*

La idea de un PSG es la esperada: en un estado $s \in \mathcal{S}_i$ (para $i \in \{\Box, \Diamond\}$), el jugador i elige jugar un politopo $K \in \Theta(s)$ y una distribución $\mu \in K$. El siguiente estado s' se selecciona de acuerdo con la distribución μ, y el juego continúa desde s' repitiendo el mismo proceso.

El desarrollo de un juego estocástico politópico se interpreta en términos de un juego estocástico donde el número de transiciones salientes desde los estados de los jugadores puede ser no numerable. Formalmente, la interpretación de un PSG es la siguiente.

Definición 4.1.2 (Interpretación de un PSG). *La interpretación del juego estocástico politópico $\mathcal{K}$ se define por el juego estocástico $\mathcal{G}_\mathcal{K} = (\mathcal{S}, (\mathcal{S}_\Box, \mathcal{S}_\Diamond), \mathcal{A}, \theta)$, donde $\mathcal{A} =$*

$\bigcup_{s\in\mathcal{S}} \Theta(s) \times \mathrm{Dist}(\mathcal{S})$ *y*

$$\theta(s,(K,\mu),s') = \begin{cases} \mu(s') & \textit{si } K \in \Theta(s) \textit{ y } \mu \in K \\ 0 & \textit{en otro caso.} \end{cases}$$

Nótese que el conjunto de acciones $\mathcal{A}$ puede ser no numerable, al igual que cada conjunto $\mathcal{A}(s) = \bigcup_{K\in\Theta(s)}\{K\} \times K$ de todas las acciones realizables en el estado s, identificado por el politopo elegido y la distribución seleccionada dentro del politopo. Por lo tanto, necesitamos extender las estrategias a este dominio no numerable, que debe estar dotado de una σ-álgebra adecuada.

Para esto, utilizamos una construcción estándar para darle a $\mathrm{Dist}(S)$ una σ-álgebra: $\Sigma_{\mathrm{Dist}(S)}$ se define como la σ-álgebra más pequeña que contiene los conjuntos $\{\mu \in \mathrm{Dist}(S) \mid \mu(S) \geq p\}$ para todo $S \subseteq S$ y $p \in [0,1]$. Ahora, dotamos a $\mathcal{A}$ con la σ-álgebra producto $\Sigma_{\mathcal{A}} = \mathscr{P}\left(\bigcup_{s\in S}\Theta(s)\right) \otimes \Sigma_{\mathrm{Dist}(S)}$, y llamamos $\mathrm{PMeas}(A)$ al conjunto de todas las medidas de probabilidad sobre $\Sigma_{\mathcal{A}}$. No es difícil comprobar que cada conjunto de acciones habilitadas $\mathcal{A}(s)$ es medible (es decir, $\mathcal{A}(s) \in \Sigma_{\mathcal{A}}$) y que la función $\theta(s,\cdot,s')$ es medible (es decir, $\{a \in \mathcal{A} \mid \theta(s,a,s') \leq p\} \in \Sigma_{\mathcal{A}}$ para todo $p \in [0,1]$).

Ahora bien, para definir formalmente las estrategias, introduciremos el concepto de comportamiento en PSGs. Esto difiere de la formulación original hecha para PSGs en [3], pero no afecta a los resultados que se mostrarán en este trabajo.

Definición 4.1.3 (Comportamiento en un PSG)**.** *Un comportamiento en un PSG es una secuencia infinita que alterna entre estados y conjuntos resultado. Formalmente un comportamiento es un* $\omega = (s_0,(K_0,\mu_0),s_1,(K_1,\mu_1),\dots)$ *tal que* $s_i \in \mathcal{S}$*,* $(K_i,\mu_i) \in \mathcal{A}(s_i)$ *y* $\mu_i(s_{i+1}) > 0$ *para todo* $i \geq 0$*.*

Notaremos al conjunto de todos los comportamientos de un PSG como Ω*. Dado un estado s, indicaremos con* Ω_s *el conjunto de los comportamientos que se originan en s, y con* Ω_s^n *al conjunto de los comportamientos que se originan en s y tiene un total de n estados.*

Definición 4.1.4 (Conjuntos soporte y acciones)**.** *Dada una acción* (K,μ) *definiremos su conjunto soporte, supp*$((K,\mu))$ *como el conjunto formado por los estados a los cuales* (K,μ) *les asigna una probabilidad positiva. Es decir,*

$$supp((K,\mu)) = \{s \in \mathcal{S} \mid \mu(s) > 0\}$$

Dado un estado s y un conjunto de estados V, definiremos como $acc(s,V)$ a las acciones que parten desde s y tienen como conjunto soporte V. Es decir,

$$acc(s,V) = \{\alpha \in \mathcal{A}(s) \mid supp(\alpha) = V\} = \{(K,\mu) \in \mathcal{A}(s) \mid \mu(V) = 1 \wedge \forall v \in V,\ \mu(v) > 0\}$$

Por último, dado un estado s, definiremos como V_s al conjunto de todos los soportes que pueden tener las distintas acciones desde s. Es decir,

$$\begin{aligned} V_s &= \{supp(\mu) \mid \exists K \in \Theta(s) : \mu \in K\} = \\ &= \{V' \subseteq \mathcal{S} | \exists \mu \in K, K \in \Theta(s) \text{ tal que } \mu(V') = 1 \text{ y } \forall s' \in V', \mu(s') > 0\} \end{aligned}$$

Entonces, ahora sí podemos extender el concepto de estrategia en un PSG:

Definición 4.1.5 (Estrategia en un PSG)**.** *Una estrategia π_i para el jugador i ($i \in \{\Box, \Diamond\}$) en un PSG será una función $\pi_i : (\mathcal{S} \times \mathcal{A})\mathcal{S}_i \to \mathrm{PMeas}(\mathcal{A})$ que asigna una medida de probabilidad a cada ωs tal que $\pi_i(\omega s)(\mathcal{A}(s)) = 1$.*

Habría que agregar alguna justificación de esta idea de comportamientos?

Medida de probabilidad que no va a ir probablemente

Con la formalización del concepto de estrategia ahora podemos presentar formalmente la definición de $\mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}$, la medida de probabilidad definida por la cadena de Markov dada por $\mathcal{G}_\mathcal{K}$ y las estrategias $\pi_\Box$ y $\pi_\Diamond$.
Para eso, primero para cada $n \geq 0$ y $s \in \mathcal{S}$ definimos $\mathbb{P}^{\pi_\Box,\pi_\Diamond,n}_{\mathcal{G}_\mathcal{K},s} : \Omega^{n+1} \to [0,1]$ para todo $s' \in \mathcal{S}$ y $\hat{\omega} \in \Omega^{n+1}_s$ inductivamente como sigue:

$$\mathbb{P}^{\pi_\Box,\pi_\Diamond,0}_{\mathcal{G}_\mathcal{K},s}(s') = \delta_s(s')$$

$$\mathbb{P}^{\pi_\Box,\pi_\Diamond,n+1}_{\mathcal{G}_\mathcal{K},s}(\hat{\omega} s_n V' s') = \mathbb{P}^{\pi_\Box,\pi_\Diamond,n}_{\mathcal{G}_\mathcal{K},s}(\hat{\omega} s_n) \int_{\{a \in A(s_n) | \mathrm{sop}(a) = V'\}} \theta(s_n, a, s')\, d(\pi_i(\hat{\omega} s_n)(a))$$

$$\text{si } s_n \in \mathcal{S}_i \text{ con } i \in \{\Box, \Diamond\}$$

(sop(a) = V' significa que el soporte de a es V')
(capaz restringirse sobre las acciones con soporte V'? No sé bien cómo va a quedar esto)
y extendemos $\mathbb{P}^{\pi_\Box,\pi_\Diamond,n}_{\mathcal{G}_\mathcal{K},s} : \mathcal{P}(\Omega^{n+1}) \to [0,1]$ a conjuntos como la suma de la medida sobre los elementos del conjunto.
Sea Σ_Ω la σ-álgebra discreta sobre Ω (pues tanto $\mathcal{S}$ como los conjuntos soportes serán conjuntos finitos) y Σ_{Ω^ω} la σ-álgebra producto usual sobre Ω^ω. Por el teorema de extensión de Carathéodory, $\mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s} : \Sigma_{\Omega^\omega} \to [0,1]$ se define como la única medida de probabilidad tal que para todo $n \geq 0$, y $SV_i \in \Sigma_\Omega$, $0 \leq i \leq n$,

$$\mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(SV_0 \times \cdots \times SV_n \times \Omega^\omega) = \mathbb{P}^{\pi_\Box,\pi_\Diamond,n}_{\mathcal{G}_\mathcal{K},s}(SV_0 \times \cdots \times SV_n).$$

Entonces, logramos interpretar un PSG como un SG pero con un conjunto infinito de acciones. La idea de este paper va a ser demostrar cómo se pueden probar ciertos resultados de esta interpretación infinita con resultado en una interpretación finita, que presentamos a continuación.

Definición 4.1.6 (Interpretación extrema de un PSG). *Dada $\mathcal{G}_\mathcal{K} = (\mathcal{S}, (\mathcal{S}_\Box, \mathcal{S}_\Diamond), \mathcal{A}, \theta)$ la interepretación de un PSG $\mathcal{K}$, la interpretación extrema de $\mathcal{K}$ es el juego estocástico $\mathcal{H}_\mathcal{K} = (\mathcal{S}, (\mathcal{S}_\Box, \mathcal{S}_\Diamond), \mathbb{V}(\mathcal{A}), \theta_{\mathcal{H}_\mathcal{K}})$ donde $\theta_{\mathcal{H}_\mathcal{K}}$ es la restricción de θ a las acciones en $\mathbb{V}(\mathcal{A}) = \{(K, \mu) \in \mathcal{A} \mid \mu \in \mathbb{V}(K)\}$. Es decir, para todos $s, s' \in \mathcal{S}$ y $a \in \mathbb{V}(\mathcal{A})$, $\theta_{\mathcal{H}_\mathcal{K}}(s, a, s') = \theta(s, a, s')$.*

Como $\mathbb{V}(\mathcal{A})$ es finito, $\mathcal{H}_{\mathcal{K}}$ es un juego estocástico finito.

4.2. Teoremas

En esta sección presentamos los resultados principales desarrollados en [3].

Si bien el paper aborda 4 tipos distintos de objetivos (a saber, de alcanzabilidad, de recompensa media, de recompensa total acumulada y de recompensa total descontada), solo nos concentraremos en los resultados que el paper presenta para objetivos de alcanzabilidad.

El primer teorema, establece la determinación de los juegos estocásticos a la vez de que establece que los objetivos de alcanzabilidad en PSGs puede ser resueltos de manera equivalente en la interpretación extrema del PSG.

Teorema 4.2.1 (Reducción de PSGs)**.** *Sean $\mathcal{G}_{\mathcal{K}}$ y $\mathcal{H}_{\mathcal{K}}$ la interpretación y la interpretación extrema, respectivamente, de un juego estocástico politópico $\mathcal{K}$. Sea C un subconjunto de estados en $\mathcal{K}$. Entonces vale que:*

$$
\begin{aligned}
&\inf_{\pi_\Diamond \in \Pi_\Diamond} \sup_{\pi_\Box \in \Pi_\Box} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_{\mathcal{K}},s}(\Diamond C) = \\
&\inf_{\pi_\Diamond \in \Pi^{MD}_\Diamond} \sup_{\pi_\Box \in \Pi^{MD}_\Box} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{H}_{\mathcal{K}},s}(\Diamond C) = \\
&\sup_{\pi_\Box \in \Pi^{MD}_\Box} \inf_{\pi_\Diamond \in \Pi^{MD}_\Diamond} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{H}_{\mathcal{K}},s}(\Diamond C) = \\
&\sup_{\pi_\Box \in \Pi_\Box} \inf_{\pi_\Diamond \in \Pi_\Diamond} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_{\mathcal{K}},s}(\Diamond C)
\end{aligned}
$$

Dado que las interpretaciones extremas son finitas, los valores se pueden calcular siguiendo algoritmos conocidos [4, 5]. Por lo tanto, este teorema también proporciona inmediatamente una solución algorítmica para los PSGs.

El segundo resultado, se deriva fácilmente del primero y será el que nos permitirá hablar de la complejidad en juegos estocástico politópico.

Teorema 4.2.2 (Complejidad en un PSG)**.** *Para todo PSG $\mathcal{K}$, $q \in \mathbb{Q}$, $G \subseteq \mathcal{S}$ y $s \in \mathcal{S}$, el problema de decidir si $Val_{\mathcal{G}_{\mathcal{K}},s}(\Diamond G) \geq q$ está en $NP \cap coNP$.*

¿Agregar algo en prelim o apéndice de clases de complejidad?

Capítulo 5

Objetivos de Rabin en PSGs

5.1. Procesos de Decisión de Markov Politópicos - PMDPs

5.1.1. Draft

Acá se está dando directamente la interpretación de un PMDP

Lo podría definir directamente como PSGs con un solo jugador.

Definición 5.1.1. *Un proceso de decisión de Markov politópico (PMDP por sus siglas en inglés) es una tupla $\mathcal{M} = (\mathcal{S}, Act, \theta')$ donde S es un conjunto finito de estados, Act es un conjunto de pares (politopo, distribución) y $\theta' : \mathcal{S} \times Act \times \mathcal{S} \to [0, 1]$ es la función de transición entre estados. También podemos definir a un PDMP como un PSG donde $\mathcal{S}_\Box = \emptyset$ o $\mathcal{S}_\Diamond = \emptyset$.*

También introducimos el concepto de *comportamiento* en un PMDP. Un comportamiento será una secuencia alternante de estados y conjuntos de próximos estados posibles, que refleja un proceso de selección de dos pasos. Desde un estado s, primero, el jugador selecciona un politopo y una distribución cuyo conjunto soporte será un $V \in V_s$, y luego el próximo estado s' será elegido probabilísticamente en base a la distribución seleccionada. Presentamos la definición formal de la siguiente manera:

Los comportamientos, las estrategias, y la medida de probabilidad en los PMDPs se definen tomando las mismas definiciones de los juegos estocásticos politópicos, mientras que las definiciones de conjuntos estado-acción, sub-MDPs y componentes finales de los procesos de decisión de Markov tradicionales se extienden a los PMDPs de la siguiente manera:

Definición 5.1.2 (conjuntos estado-resultado y sub-PMDPs). *Dado un PMDP $\mathcal{M} = (S, Act, \theta')$ un conjunto estado-resultado es un subconjunto $\chi \subseteq \{(s, V') \mid s \in S \wedge V' \in V_s\}$. Un sub-PMDP es un par (C, D), donde $C \subseteq S$ y D es una función que asocia a cada $s \in C$ un conjunto $D(s) \subseteq V_s$ de subconjuntos de estados próximos posibles. Hay una relación uno-a-uno entre sub-PMDPs y conjuntos de estado-acción:*

- *dado un conjunto estado-resultado χ, denotamos $sub(\chi) = (C, D)$ al sub-PMDP definido por:*

$$C = \{s \mid \exists V'.(s, V') \in \chi\} \quad D(s) = \{V' \mid (s, V') \in \chi\}$$

- *dado un sub-PMDP (C, D), denotamos por $er(C, D) = \{(s, V') \mid s \in C \wedge V' \in D(s)\}$ al conjunto estado-resultado correspondiente a (C, D).*

Si definimos para cada V', un vértice único nuevo $v_{V'}$ podemos ver que cada sub-PMDP (C, D) induce una *relación de aristas*: hay una arista $(s, v_{V'})$ de $s \in C$ a $v_{V'}$ para cada $V' \in D(s)$ y hay una arista de $(v_{V'}, t)$ de $v_{V'}$ con $V' \in D(s)$ a $t \in \mathcal{S}$ sii es posible ir de s a t en un paso con probabilidad positiva utilizando una acción cuyo conjunto resultado sea V'. La definición formal es como sigue:

Definición 5.1.3 (relación de aristas ρ). *Para un sub-PMDP, definimos la relación $\rho_{(C,D)}$ como*

$$\rho_{(C,D)} = \{(s, v_{V'}) \mid \exists V' \in D(s)\} \cup \{(v_{V'}, t) \mid t \in V'\}$$

Definición 5.1.4 (componente final). *Un sub-PMDP es una componente final si:*

- *$V' \subseteq C$ para todo V' tal que existe un $s \in C$ donde $V' \in D(s)$*
- *el grafo $(C \cup \{v'_V \mid \exists s \in C : V' \in D(s)\}, \rho_{(C,D)})$ es fuertemente conexo.*

Llamaremos $EC(\mathcal{M})$ al conjunto de todas las componentes finales en un PMDP $\mathcal{M}$.

Intuitivamente, una componente final representa un conjunto de pares estado-resultado que, una vez en ellos, es posible quedase allí si la estrategia escoge las acciones de manera apropiada. Esta intuición se hará precisa con los siguientes teoremas.

Antes de enunciar estos teoremas, introducimos una abreviatura para el conjunto de estados-resultados que ocurren infinitamente a menudo en un comportamiento dado y una notación para el conjunto (infinito) de acciones con el mismo conjunto resultado.

Definición 5.1.5 (*inf*)**.** *Dado un comportamiento $\omega = (s_0, \alpha_0, s_1, \alpha_1, \dots)$ indicamos por*

$$\mathit{inf}(\omega) = \{(s, V') \mid s_k = s \wedge \mathit{supp}(\alpha_k) = V' \text{ para infinitos } k \in \mathbb{N}_0\}$$

al conjunto de pares estado-resultado que ocurren infinitas veces en él.

Ahora sí, podemos pasar a presentar las primeras demostraciones:

Teorema 5.1.1 (estabilidad de componentes finales)**.** *Sea (C, D) una componente final. Entonces, para cada estrategia π existe una estrategia π', que difiere de π solo en C, tal que:*

$$\mathbb{P}_s^{\pi}(\Diamond C) = \mathbb{P}_s^{\pi'}(\Diamond C) = \mathbb{P}_s^{\pi'}(inft(\omega) = er(C, D)) \tag{5.1}$$

para todo $s \in S$.

Demostración. Considérese una estrategia π' definida como sigue para cada secuencia $s_0 \dots s_n$ con $n \geq 0$:

- Si $s_n \in C$, la estrategia asignará probabilidad positiva a una única acción $(K, \mu) \in A(s_n, V')$ para cada conjunto resultado V' (notaremos a esta acción particular con $(K, \mu)_{V'}$), y la probabilidad de elegir cada una de esas acciones se distribuirá de manera uniforme. Es decir,

$$\pi'(s_0 \dots s_n)(K, \mu) = \begin{cases} \frac{1}{|D(s_n)|} & \text{si } (K, \mu) = (K, \mu)_{V'} \text{ para algún } V' \in D(s); \\ 0 & \text{en otro caso} \end{cases}$$

- Si $s_n \notin C$, la estrategia π' coincide con π, i.e.

$$\pi(s_0 \dots s_n)(K, \mu) = \pi'(s_0 \dots s_n)(K, \mu) \quad \forall (K, \mu) \in Act$$

La primera igualdad en 5.1 es una consecuencia del hecho de que π y π' coinciden fuera de C.

Para la segunda igualdad, basta con ver que bajo la estrategia π' una vez que un comportamiento entra a C, nunca sale de C ni se elige una acción que no esté en D. Es más, una vez en C un comportamiento visitará todos los estados de C infinitamente a menudo con probabilidad 1. □

El próximo resultado establece que, para cualquier estado inicial y cualquier estrategia (de memoria finita), un comportamiento terminará con probabilidad 1 en una componente final. Esta es la razón detrás del nombre "componente final".

Teorema 5.1.2 (teorema fundamental de las componentes finales)**.** *Sea $\mathcal{M}$ un PMDP. Para todo $s \in S$, toda estrategia π de memoria finita,*

$$\mathbb{P}^{\pi}_{\mathcal{M},s}(\{\omega \in Paths(s) \mid sub(inft(\omega)) \text{ } es \text{ } una \text{ } componente \text{ } final\}) = 1$$

Demostración. Consideremos un sub-PMDP (C, D) que no sea una componente final y sea $\Omega_s^{(C,D)} = \{\hat{\omega} \in \Omega_s \mid inft(\hat{\omega}) = er(C, D)\}$ el conjunto de comportamientos cuyo conjunto de pares estado-resultado que se repiten infinitas veces en él forman el sub-PMDP (C, D).

Si podemos mostrar que

$$\mathbb{P}^{\pi}_{\mathcal{M},s}(\{\omega \in Paths(s) \mid \omega \in \Omega_s^{(C,D)}\}) = 0 \tag{5.2}$$

como (C, D) es un sub-PMDP cualquiera y como hay una cantidad finita de sub-PMDPs en $\mathcal{M}$, esto es lo mismo que mostrar que

$$\mathbb{P}^{\pi}_{\mathcal{M},s}(\{\omega \in Paths(s) \mid sub(inft(\omega)) \text{ es una componente final}\}) = 1$$

. Veamos que vale 5.2, dividiendo en casos según cuál es la condición de la definición 5.1.4 que no se cumple para (C, D):

- Primero, asumamos que existe un $(t, V') \in er(C, D)$ tal que $V' \nsubseteq C$.

  Sabemos que cada comportamiento en $\Omega_s^{(C,D)}$ toma el par estado-resultado (t, V') infinitas veces. Llamemos I al conjunto de índices infinito que representa los momentos en los que se visita (t, V'). Indiquemos con μ_i a la distribución elegida en el momento $i \in I$ y definamos como $r_i = \sum_{u \in C} \mu_i(u)$ a la probabilidad de quedarnos en C en el momento $i \in I$ (que en cada caso será menor a 1 porque $V' \nsubseteq C$).

  Como π es de memoria finita sucederá que π solo puede elegir una cantidad finita de acciones (y, por lo tanto, distribuciones) distintas desde t. Esto hace que el conjunto $R = \{r_i \mid i \in I\}$ tenga un máximo, llamémoslo r.

  Para que valga que ω esté en $\Omega_s^{(C,D)}$ tiene que valer que en infinitos momentos i nos quedemos en C. Entonces que vale que $\mathbb{P}_s^\pi(\omega \in \Omega_s^{(C,D)}) < r^k$ para todo $k > 0$ natural. Como sabemos que $r < 1$, tenemos que $\mathbb{P}_{\mathcal{M},s}^\pi(\{\omega \in \mathit{Paths}(s) \mid \omega \in \Omega_s^{(C,D)}\}) = 0$.

- Si no, asumamos que existen $t_1, t_2 \in C$ tales que no hay camino de t_1 a t_2 en $(C \cup \{v'_V \mid \exists s \in C : V' \in D(s)\}, \rho(C, D))$.

  La falta de camino de t_1 a t_2 en $(C, \rho_{(C,D)})$ implica que para cada subsecuencia $s_m V_m s_{m+1} ... s_n$ de comportamiento en $\Omega_s^{(C,D)}$ que vaya de $s_m = t_1$ a $s_n = t_2$, hay 2 opciones:

  1. existe un $j \in [m+1, n-1]$ tal que $s_j \notin C$. Como cada comportamiento en $\Omega_s^{(C,D)}$ contiene infinitas subsecuencias de t_1 a t_2 y tenemos una cantidad finita de estados, sabemos que habrá una cantidad infinita de s_j iguales. Pero si infinitas veces se toma un estado s_j entonces, $s_j \in C$, lo que contradice la hipótesis anterior. Absurdo.
  2. existe un $j \in [m, n-1]$ tal que $V_j \notin D(j)$. Como cada comportamiento en $\Omega_s^{(C,D)}$ contiene infinitas subsecuencias de t_1 a t_2 y tenemos una cantidad finita de conjuntos resultado, sabemos que habrá una cantidad infinita de V_j iguales, con lo que $V_j \in D(j)$, lo que contradice la hipótesis anterior. Absurdo

  Con lo que arribamos a que $\mathbb{P}_{\mathcal{M},s}^\pi(\{\omega \in \mathit{Paths}(s) \mid \omega \in \Omega_s^{(C,D)}\}) = 0$

□

Esto nos permite definir el siguiente corolario útil para el análisis de condiciones de Rabin:

Corolario 5.1.2.1. *Sea $\mathcal{M}$ un PMDP, s un estado en él y sea π una estrategia (lo hacemos con estrategias o planificadores?) en él. Una condición de Rabin $\widetilde{R} = \{\langle G_1, R_1\rangle, ..., \langle G_k, R_k\rangle\}$ se satisface desde s, siguiendo la estrategia π con probabilidad 1 si y solo si para cada componente final U alcanzable desde s, existe un $j \in \{1, 2, ..., k\}$ tal que $U \cap R_j = \emptyset$ y $U \cap G_j \neq \emptyset$.*

Ahora bien, una pregunta muy natural que surge es "¿por qué nos restringimos a estrategias de memoria finita en este último teorema?"

Nota sobre la limitación a estrategias finitas del Teorema 5.1.2

La prueba del teorema 5.1.2 está inspirada en las pruebas realizadas para MDPs (véase [1] [2]). En ellas, la prueba consiste en proponer un sub-PMDP (C, D) arbitrario que no cumple la definición 5.1.4, dividir el análisis por casos de cómo no se cumple la definición de componente final y mostrar que en cada caso la probabilidad de que un comportamiento que siga la estrategia del enunciado genere un sub-PMDP como el propuesto es 0.

Si nos remitimos a la prueba que mostramos en 5.1.2, vemos que el método de prueba se ajusta bien a nuestro caso hasta llegar al primer ítem donde el argumento está explícitamente respaldado en el hecho de que π es de memoria finita. De esta manera, se puede decir que el conjunto de las distintas probabilidades de quedarse en C en los momentos i tiene un máximo, r, lo que permite acotar la productoria $\Pi_{i \in I}\ r_i$ por $\lim_{k\to\infty} r^k$, que sabemos que converge a 0, puesto que $r < 1$.

Si no nos restringimos a estrategias de memoria finita no podemos garantizar esto. Sabemos que la productoria va a estar acotada inferiormente por 0 y superiormente por 1, pero bien podría no converger a 0, sino a algún otro número. Por ejemplo, veamos lo siguiente:

Consideremos un juego estocástico politópico donde tenemos un vértice t y un vértice s. Desde t, existen acciones μ de la forma $\mu(s) = \frac{1}{k^2}$, $\mu(t) = 1 - \frac{1}{k^2}$, por lo que se puede definir una estrategia de memoria infinita π a partir de la cantidad de t que se encuentran en el comportamiento de la siguiente manera:

$$\pi(\omega t)(s) = \frac{1}{(\#_t(\omega))^2}$$
$$\pi(\omega t)(t) = 1 - \frac{1}{(\#_t(\omega))^2}$$

donde $\#_t$ se define inductivamente sobre comportamientos de la siguiente manera:

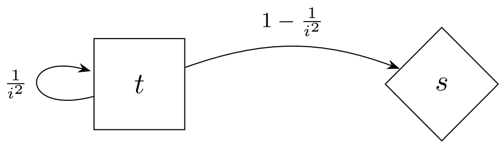

Figura 5.1: Interpretación del juego estocástico con la estrategia π fijada.

$$\#_t(\emptyset) = 0$$
$$\#_t(\omega V' t) = \#_t(\omega) + 1$$
$$\#_t(\omega V' s) = \#_t(\omega)$$

donde ω es un comportamiento, $\emptyset$ representa el comportamiento vacío (no sé qué tan bien estaría esto) y $s \neq t$.

Si, remitiéndonos otra vez a la estrategia de prueba, suponemos que s no forma parte de C, tenemos que la probabilidad de quedarnos en C sería $\Pi_{i \geq 0} 1 - (\frac{1}{i^2})$, que sabemos que converge a $\frac{1}{2}$.

No me gustan mucho estos siguientes párrafos, pero no sé muy bien cómo decir lo que quiero, y siento que vos habías dicho algo más interesante sobre esto.

Ahora bien, es importante aclarar que esto no constituye de ninguna manera un contraejemplo, además de que faltaría formalidad en su presentación, resulta que plantear $s \notin C$ siendo que tengo probabilidad positiva de visitarlo cada vez no tiene mucho sentido. (Esto no sé que tanto es así).

En cualquier caso, podría ser una muestra de que la aplicación de este fórmula de prueba no es útil *directamente* para probar el teorema para estrategias de memoria infinita. Pero resaltamos el directamente porque la productoria constituye una cota a la probabilidad de que $\omega \in \Omega_s^{(C,D)}$, no es directamente el valor de la probabilidad.

Esto de acá abajo capaz se va

A su vez, el teorema anterior nos permite empezar a considerar especialmente un tipo de propiedades, aquellas que solo dependen del comportamiento asintótico.

Definición 5.1.6 (propiedad límite)**.** *Una propiedad de tiempo lineal se llama propiedad límite de tiempo lineal si para todos los comportamientos* ω, ω' $\omega \in P \wedge \mathit{inf}(\omega) = \mathit{inf}(\omega') \implies \omega' \in P$.

Si T *es un subconjunto de estados para un PMDP* $\mathcal{M}$*, diremos que* $T \models P$ *para alguna propiedad límite* P *sii para todo comportamiento tal que* $sub(\mathit{inf}(\omega)) = (T, A)$ *para algún* A*, vale que* $\omega \in P$*.*

Por el teorema 5.1.2, solo los conjuntos T donde T es un conjunto de estados de una componente final son relevantes a la hora de analizar las probabilidades de una propiedad límite P. Denotemos U_P a la unión de todos los conjuntos T de todas las componentes finales (T, A) de $\mathcal{M}$ tales que $T \models P$, y V_P la unión de los conjuntos T de todos los $(T, A) \in EC(\mathcal{M})$ tales que $\neg(T \models P)$. Esto nos permite presentar el siguiente interesante resultado:

Teorema 5.1.3. *Sea* $\mathcal{M}$ *un PMDP y sea* P *una propiedad de límite. Entonces existe una estrategia de memoria finita* π *tal que para todo estado* s *de* $\mathcal{M}$*:*

(a) $\sup_\pi \mathbb{P}^\pi_{\mathcal{M},s}(P) = \sup_\pi \mathbb{P}^\pi_{\mathcal{M},s}(\Diamond U_P)$

(b) $\inf_\pi \mathbb{P}^\pi_{\mathcal{M},s}(P) = 1 - \sup_\pi \mathbb{P}^\pi_{\mathcal{M},s}(\Diamond V_P)$

Demostración. Primero, probemos el item *(a)*.

Para cada estrategia tenemos que $\mathbb{P}^\pi_{\mathcal{M},s}(P) \leq \mathbb{P}^\pi_{\mathcal{M},s}(\Diamond U_P)$, puesto que por teorema 5.1.2 vale que:

$$\mathbb{P}^{\pi}_{\mathcal{M},s}(P) = \mathbb{P}^{\pi}_{\mathcal{M},s}\{\omega \in \mathit{Paths}(s) \mid \mathit{inf}(\omega) \in EC(\mathcal{M}) \wedge \mathit{inf}(\omega) \models P\}$$

y por definición de U_P, $\omega \models \Diamond U_P$ para cada camino ω tal que $\mathit{inf}(\omega)$ es una componente final y $\mathrm{inf}(\omega) \models P$.

Si podemos probar que existe una estrategia π tal que $\mathbb{P}^{\pi}_{\mathcal{M},s}(P) = \sup_{\pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(\Diamond U_P)$, podemos ver que vale la igualdad, por lo expuesto anteriormente.

Consideremos, entonces, una estrategia sin memoria π_0 que maximiza las probabilidades de llegar a U_P para cada s en $\mathcal{M}$ (esta estrategia existe por resultado X de literatura y el teorema 1 de [3]). Además, para cada componente final (C, D), definamos $\pi_{(C,D)}$: una estrategia que asegura quedarse siempre en C mientras visita infinitamente a menudo todos los estados $s \in C$ (esta estrategia existe por el teorema 5.1). Y, por último, para cada estado $u \in U_P$ seleccionamos una componente final $EC(u) = (C, D)$ tal que $u \in C$ y $C \models P$. Todo esto nos servirá para definir la estrategia π que nos asegura que $\mathbb{P}^{\pi}_{\mathcal{M},s}(P) = \sup_{\pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(\Diamond U_P)$.

Definimos π como la estrategia que primero se comporta como π_0 hasta que llega a un estado $u \in U_P$. Desde ese momento, π se comporta como $\pi_{EC(u)}$. En consecuencia, para esta estrategia, los caminos que eventualmente entran a U_P visitarán todos los estados de una componente final (C, D) tal que $C \models P$ infinitamente a menudo con probabilidad 1. En particular, $\mathit{inf}(\pi) \models P$ vale para todo ω que sigue la estrategia π y eventualmente alcanza U_P. Esto lleva a que:

$$\begin{aligned}
\mathbb{P}^{\pi}_{\mathcal{M},s}(P) &= \sum_{u \in U_P} \mathbb{P}^{\pi_0}_{\mathcal{M},s}((\neg U_P)\ \mathcal{U}\ u) \cdot \mathbb{P}^{\pi_{EC(u)}}_{\mathcal{M},u}(P) \\
&= \sum_{u \in U_P} \mathbb{P}^{\pi_0}_{\mathcal{M},s}((\neg U_P)\ \mathcal{U}\ u) \cdot 1 \\
&= \sum_{u \in U_P} \mathbb{P}^{\pi_0}_{\mathcal{M},s}((\neg U_P)\ \mathcal{U}\ u) \\
&= \sup_{\pi'} \mathbb{P}^{\pi'}_{\mathcal{M},s}(\Diamond U_P)
\end{aligned}$$

(Revisar la manera en la que están escritas estas igualdades).

Hemos probado así el ítem *(a)*.

Ahora, probemos el ítem *(b)*. Este deriva del ítem anterior usando el hecho de que las propiedades límite son cerradas bajo negación (es decir, si P es una propiedad límite entonces $\overline{P}$ también lo es) y que $\mathbb{P}^{\pi}_{\mathcal{M},s}(P) = 1 - \mathbb{P}^{\pi}_{\mathcal{M},s}(\overline{P})$ para toda estrategia π. Por lo tanto:

$$\inf_{\pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(P) = 1 - \inf_{\pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(\overline{P}) \tag{5.3}$$

y cualquier estrategia de memoria finita que maximiza la probabilidades para $\overline{P}$, minimiza las probabilidades para P. Para un subconjunto de estados T, tenemos que $T \models \overline{P}$ sii $\neg(T \models P)$. Entonces, el conjunto V_P de todos los estados t que están contenidos en una componente final (T, A) con $\neg(T \models P)$ coincide con el conjunto $U_{\overline{P}}$ que surge de la unión de todos las componentes finales (T, A) donde $T \models \overline{P}$. Y con eso, el ítem *(a)* aplicado a $\overline{P}$ hace que valga:

$$\sup_{\pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(\overline{P}) = \sup_{\pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(\Diamond U_{\overline{P}}) = \sup_{\pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(\Diamond V_P)$$

Y estas igualdades aplicadas a 5,3 nos dan lo que queríamos probar.

□

5.2. Juegos justos: la respuesta a la pregunta cualitativa

A fin de responder la pregunta "¿desde qué estados existe probabilidad uno de ganar?", presentaremos un tipo nuevo de juegos deterministas: los juegos justos. Probaremos que el conjunto de vértices casi seguramente ganadores en un juego estocástico politópico con un objetivo de Rabin es igual al conjunto de vértices ganadores de un juego justo construido a partir del PSG.

Para eso primero presentaremos el concepto de juego justo y la construcción del juego justo a partir del PSG. Luego, explicitaremos un poco más la relación entre el PSG y el juego justo construido, al cual llamaremos su desrandomización. Seguido de eso, expondremos la prueba formal de la igualdad entre los conjuntos. Y, por último, mostraremos un algoritmo para la sintésis de estrategias de memoria finita casi seguramente ganadoras para un PSG con un objetivo de Rabin.

5.2.1. Juegos de adversario justo

Sea G un juego de grafo de dos jugadores y sea $E^l \subseteq (V_1 \times V) \cap E$ un conjunto dado de aristas vivas ver traducción. Sea $V^l := \mathrm{dom}(E^l)$ el conjunto de vértices del jugador $\Diamond$ que están en el dominio de E^l. Intuitivamente, las aristas en E^l representan suposiciones de justicia sobre el jugador $\Diamond$: para cada arista $(v, v') \in E^l$, si v es visitado infinitamente a menudo en una jugada, se espera que la arista (v, v') sea elegida también infinitamente a menudo por el jugador $\Diamond$. Es decir, si un vértice v es visitado infinitamente, se espera también que toda arista viva saliente de v sea tomada infinitamente a menudo.

Denotamos por $G^l = \langle G, E^l \rangle$ a un juego de grafo de dos jugadores con aristas vivas, y extendemos nociones como jugadas, estrategias, condiciones ganadoras, regiones ganadoras, etc., de juegos de grafo a juegos de grafos con aristas vivas. Una jugada π sobre G^l es fuertemente justa si satisface la siguiente fórmula en lógica temporal lineal (LTL):

$$\alpha := \bigwedge_{(v,v') \in E^l} \left(\Box\Diamond v \rightarrow \Box\Diamond(v \wedge \bigcirc v')\right).$$

Dado G^l y una condición de victoria φ, el jugador $\Box$ gana el juego de adversario justo sobre G^l con respecto a la condición de victoria φ desde un vértice $v_0 \in V$ si gana el juego sobre G^l para la condición de victoria $\alpha \rightarrow \varphi$ desde v_0.

Hay dos observaciones interesantes sobre los juegos de adversario justo.

Primero, las aristas vivas permiten descartar ciertas estrategias del jugador $\Diamond$, facilitando que el jugador $\Box$ gane en determinadas situaciones. Por ejemplo, consideremos un grafo de juego (figura X, parte superior) con dos vértices p y q. El vértice p pertenece al jugador $\Diamond$ y el vértice q al jugador $\Box$. La arista (p, q) es una arista viva (representada con línea discontinua). Supongamos que la especificación para el jugador $\Box$ es $\varphi = \Box\Diamond q$. Si la arista (p, q) no fuera viva, el jugador $\Box$ no ganaría desde p, porque el jugador $\Diamond$ podría mantener el juego atrapado en p eligiéndose a sí mismo como sucesor en cada turno. En contraste, el jugador $\Box$ gana desde p en el juego adversarial justo, porque la suposición de liveness sobre la arista (p, q) fuerza al jugador $\Diamond$ a elegir infinitamente a menudo la transición hacia q.

Segundo, las suposiciones de justicia modeladas por aristas vivas restringen las elecciones de estrategias del jugador $\Diamond$ menos que lo que restringirían la suposición de que el jugador $\Diamond$ elige probabilísticamente entre estas aristas. Consideremos, por ejemplo, un juego de adversario justo con un único vértice del jugador $\Diamond$, p (cuadrado) con dos aristas vivas salientes hacia los estados q y q', como se muestra en la Figura 1 (parte inferior). Si el jugador $\Diamond$ elige aleatoriamente entre las aristas (p, q) y (p, q'), toda secuencia finita de visitas a los estados q y q' ocurrirá infinitamente a menudo con probabilidad uno. Esto no es cierto en el juego de adversario justo. Aquí el jugador $\Diamond$ puede elegir una secuencia particular de visitas a q y q' (por ejemplo, simplemente $qq'qq'qq' \dots$), siempre que ambos sean visitados infinitamente a menudo.

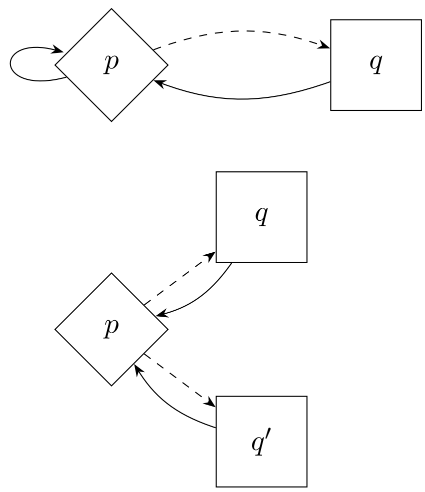

Figura 5.2: Dos juegos de adversario justo.

5.2.2. Desrandomización de PSGs

Dada la interpretación de un juego estocástico politópico $\mathcal{G}_{\mathcal{K}} = (\mathcal{S}, (\mathcal{S}_\Box, \mathcal{S}_\Diamond), \mathcal{A}, \theta)$, se define la *desrandomización de* $\mathcal{G}_{\mathcal{K}}$, $Derand(\mathcal{G}_{\mathcal{K}}) := ((V, V_\Box, V_\Diamond, E), E^l)$, como el (vamos con esta traducción?) juego de grafo de 2 jugadores extremadamente equitativo con:

$$\begin{aligned}
\tilde{V} &= \bigcup_{\substack{s \in \mathcal{S} \\ V' \in V_s}} v_{V'} \\
V &= \mathcal{S} \cup \tilde{V} \\
V_{\Box} &= \mathcal{S}_{\Box} \\
V_{\Diamond} &= \mathcal{S}_{\Diamond} \cup \tilde{V} \\
E &= \{(s, v_V) : V \in V_s\} \\
E^l &= \{(v_V, s') : s' \in V\}
\end{aligned}$$

y para el cual se cumple la siguiente condición de equidad:

$$\varphi^l := \bigwedge_{(v_{V'}, s') \in E^l} (\Box\Diamond v_{V'} \to \Box\Diamond(v_{V'} \wedge \bigcirc s'))$$

Esta misma condición de equidad se puede expresar como

$$\varphi^l := \textstyle\bigwedge_{v_V \in \tilde{V}} \varphi^V, \text{ donde}$$

$$\varphi^V := \textstyle\bigwedge_{s' \in V} (\Box\Diamond v_{V'} \to \Box\Diamond(v_{V'} \wedge \bigcirc s'))$$

5.2.3. Relación entre un PSG y su desrandomización

Nos será útil pensar en transformaciones entre los distintos juegos para las pruebas que tenemos más adelante, así que veremos algunas definiciones. Antes, fijemos la interpretación de un juego estocástico politópico $\mathcal{G}_{\mathcal{K}}$ y su desrandomización $Derand(\mathcal{G}_{\mathcal{K}})$.

Condición de equidad en el juego estocástico politópico

Podemos expresar fácilmente la condición de equidad en el juego estocástico politópico de la siguiente forma:

$$\hat{\varphi}^l := \textstyle\bigwedge_{\alpha \in \mathcal{A}} \hat{\varphi}^{\mathrm{supp}(\alpha)}, \text{ con}$$

$$\hat{\varphi}^{\mathrm{supp}(\alpha)} := \textstyle\bigwedge_{s' \in \mathrm{supp}(\alpha)} (\Box\Diamond\alpha \to \Box\Diamond(\alpha \wedge \bigcirc s'))$$

Comportamientos y caminos

Dado un comportamiento $\omega = (s_0, \alpha_0, s_1, \alpha_1, \dots)$ en $\mathcal{G}_{\mathcal{K}}$, podemos obtener un único camino en $Derand(\mathcal{G}_{\mathcal{K}})$ al que llamaremos derand(ω) y tendrá la siguiente forma:

$$\text{derand}(\omega) = (s_0, v_{\text{supp}(\alpha_0)}, s_1, v_{\text{supp}(\alpha_1),\dots})$$

Por otro lado, si tenemos un camino $\rho = (s_0, v_{V_0}, s_1, v_{V_1}, \dots)$ en $Derand(\mathcal{G}_{\mathcal{K}})$, existen varios comportamientos en $\mathcal{G}_{\mathcal{K}}$ que se corresponderían con él. Llamaremos a este conjunto de comportamientos rand(ρ). Es decir,

$$\text{rand}(\rho) = \{(s_0, \alpha_0, s_1, \alpha_1, \dots) \in \Omega_{\mathcal{G}_{\mathcal{K}},s} \mid \forall i \geq 0,\ \text{supp}(\alpha_i) = V_i\}$$

Estas definiciones también se extienden a prefijos finitos de comportamientos y caminos de manera natural.

Randomización y desrandomización de estrategias

Supongamos que tenemos una estrategia σ_i en $Derand(\mathcal{G}_{\mathcal{K}})$ y queremos construir una estrategia π_i en $\mathcal{G}_{\mathcal{K}}$ que tenga un comportamiento similar. σ_i está definida para todos los prefijos finitos de caminos ρ en $Derand(\mathcal{G}_{\mathcal{K}})$ y nuestra idea es definir π_i para todos los prefijos finitos de comportamiento ω en $\mathcal{G}_{\mathcal{K}}$. La idea, entonces, será definir π_i de igual manera para todos los elementos del conjunto rand(ρ).

Solo nos va a interesar ver cómo σ_i se comporta en los prefijos finitos de caminos ρ que terminen en un estado $s \in \mathcal{S}$ [1]. Para cada uno de estos ρ, tenemos que $\sigma_i(\rho) = v_{\widehat{V}}$ para algún v_V. Lo que haremos para la construcción de π_i es para cada $\omega \in \text{derand}(\rho)$ definir $\pi_i(\omega)(\hat{\alpha}) = 1$ para algún $\hat{\alpha}$ particular tal que $\text{supp}(\hat{\alpha}) = \widehat{V}$.

Llamaremos a esta nueva estrategia π_i obtenida a partir de σ_i, rand(σ_i).

$$\sigma_i(s_0, v_{V_0}, \dots, s_k) = v_{V_k} \implies \text{rand}(\sigma_i)(s_0, \alpha_0, \dots, s_k)(\alpha_k) = 1 \text{ donde } \forall i \ \text{supp}(\alpha_i) = V_i.$$

Figura 5.3: Una forma de ver la randomización de estrategias en $Derand(\mathcal{G}_{\mathcal{K}})$

[1] Desde los prefijos finitos de caminos donde el último elemento es un vértice v_V, las decisiones son tomadas por el jugador $\diamondsuit$, pero solo reflejan lo que sería la decisión probabilística en $\mathcal{G}_{\mathcal{K}}$.

Supongamos ahora que tenemos una estrategia π_i en $\mathcal{G}_{\mathcal{K}}$ y queremos construir una estrategia σ_i en $Derand(\mathcal{G}_{\mathcal{K}})$ que tenga un comportamiento similar.

Para cada prefijo finito de camino ρ debemos definir qué vértice será el que elija $\sigma_i(\rho)$. La idea será que elegiremos un vértice v_V que se corresponde al soporte de una acción a la que $\pi_i(\omega)$ (siendo ω tal que $\text{derand}(\omega) = \rho$) le asigne una probabilidad positiva.

Como existen varios prefijos de comportamiento que podrían cumplir con la condición de ω, la definición de $\text{derand}(\pi_i)$ dependerá de la elección de un prefijo de comportamiento ω para cada prefijo de camino ρ tal que $\text{derand}(\omega) = \rho$.

Además, como cada acción le puede dar probabilidad positiva a varias acciones con distintos soportes, la definición de $\text{derand}(\pi_i)$ también dependerá de la elección de una acción a la que $\pi_i(\omega)$ le asigne una probabilidad positiva (por cada ω seleccionado en el paso anterior).

Con estas decisiones tomadas, podemos definir como se comportará σ_i para cada prefijo de camino que termina en un estado $s \in \mathcal{S}$.

Ahora bien, para el caso de $\sigma_\Diamond$ también hay que definir cómo se comporta la estrategia en los caminos que terminan en los vértices de la forma v_V. Es decir, debemos elegir qué estado s_{k+1} será el que cumpla $\sigma_\Diamond(s_0, v_{V_0}, \ldots, s_k, v_{V_k}) = s_{k+1}$ para cada k. En este caso, lo que nos debemos asegurar es que se cumpla la condición de equidad.

¿Cómo asegurar la condición de equidad?

Presentamos dos maneras fáciles con las cuales se podría asegurar esto, pero podrían existir infinidades de ellas:

1. como cada V_k es finito, podemos numerar cada $s^{V_k} \in V_k$ de manera $s_0^{V_k}, s_1^{V_k}, \ldots$. Esto nos permite seleccionar el próximo estado en base a la cantidad de veces que se visitó v_{V_k}. Si n fueron las veces que se visitó v_{V_k} en el prefijo finito de camino ρ', entonces podemos definir $\sigma_{\Diamond}(\rho' v_{V_k}) = s_n^{V_k}$.

2. si tenemos un comportamiento $\hat{\omega}$ válido que respeta la estrategia $\pi_{\Diamond}$, y a partir de los prefijos de este $\hat{\omega}$ es que estamos definiendo las elecciones de prefijos y acciones de $\sigma_{\Diamond}$, podemos también usar este $\hat{\omega}$ para las decisiones desde los estados v_{V_k}, eligiendo como próximo vértice s_{k+1} al que se eligió probabilísticamente en $\hat{\omega}$.

Con estas decisiones ya tomadas resulta la construcción de la σ_i que queríamos, a la cual llamaremos derand(π_i).

5.2.4. Prueba de igualdad sobre los conjuntos ganadores

Teorema 5.2.1. *Sea $\mathcal{G}_{\mathcal{K}} = (\mathcal{S}, (\mathcal{S}_{\Box}, \mathcal{S}_{\Diamond}), \mathcal{A}, \theta)$ la interpretación de un juego estocástico politópico, $\widetilde{\mathcal{R}} = \{\langle G_1, R_1\rangle, ..., \langle G_k, R_k\rangle\}$ una condición de Rabin sobre $\mathcal{S}$, con su especificación LTL φ,*

$$\varphi := \bigvee_{j \in [1,k]} \left(\Diamond\Box\overline{R_j} \wedge \Diamond\Box G_j \right)$$

y sea Derand$(\mathcal{G}_{\mathcal{K}})$ su desrandomización.

Sea $\mathcal{W} \subseteq \mathcal{S}$ el conjunto de todos los estados desde los cuales el jugador $\Box$ gana en Derand$(\mathcal{G}_{\mathcal{K}})$ y sea $\mathcal{W}^{as}$ el conjunto de vértices desde los cuales el jugador $\Box$ gana casi seguramente con estrategias de memoria finita. Entonces, $\mathcal{W} = \mathcal{W}^{as}$.

Es más, a partir de una estrategia ganadora en $Derand(\mathcal{G}_{\mathcal{K}})$ se puede construir fácilmente una estrategia ganadora en $\mathcal{G}_{\mathcal{K}}$, y viceversa.

Demostración. Probaremos la doble contención:

Primera contención: $\mathcal{W} \subseteq \mathcal{W}^{as}$

Sea $s \in \mathcal{W}$. Entonces, sabemos que existe al menos una estrategia del jugador $\Box$ ganadora desde s en $Derand(\mathcal{G}_{\mathcal{K}})$. Llamemos a esta $\sigma^*_\Box$.

Queremos ver que $s \in \mathcal{W}^{as}$, lo que requiere ver que existe una estrategia $\hat{\pi}^*_\Box$ tal que:

$$\inf_{\pi_\Diamond \in \Pi_\Diamond} \mathbb{P}^{\hat{\pi}^*_\Box, \pi_\Diamond}_{\mathcal{G}_{\mathcal{K}}, s}(\varphi) = 1 \tag{5.4}$$

Proponemos $\hat{\pi}^*_\Box = \text{rand}(\sigma^*_\Box)$.

Ahora bien, ¿cómo podemos probar 5.4?

Por definición de $\sigma^*_\Box$ sabemos que si se está frente a una estrategia $\sigma_\Diamond$ que respeta la condición de equidad φ^l, entonces la jugada determinada por ambas estrategias cumple la especificación φ.

A esto de acá abajo es que le faltaría una justificación más precisa. Capaz tampoco es la manera en la que habría que definir el varphil original.

Versión 2 Si definimos

$$\Psi = \bigwedge_{\substack{s \in pred(v_V) \\ s' \in V}} \big(\Box\Diamond s \to \Box\Diamond(s \wedge \bigcirc s')\big)$$

$$\hat{\Psi} = \bigwedge_{s' \in \cup_{\alpha \in acc(s,V)} \text{supp}(\alpha)} \big(\Box\Diamond s \to \Box\Diamond(s \wedge \bigcirc s')\big)$$

sabemos que $\Psi \implies \varphi^l$ y que $\hat{\Psi}$ es la adaptación de Ψ a un juego estocástico politópico. Con lo que tendríamos que

$$\inf_{\pi_\Diamond \in \Pi_\Diamond} \mathbb{P}^{\hat{\pi}^*_\Box, \pi_\Diamond}_{\mathcal{G}_{\mathcal{K}}, s}(\hat{\Psi} \to \varphi) = 1 \tag{5.5}$$

Acá iría algo similar a la versión 1 ... Vemos que $\neg\hat{\Psi}$ es 0, viendo que para cada $\alpha \in acc(s, V)$ y cada $s' \in \mathrm{supp}(\alpha)$,

$$\mathbb{P}^{\hat{\pi}^*_\Box, \pi_\Diamond}_{\mathcal{G}_\mathcal{K}, s} \left(\Box\Diamond s \wedge \Diamond\Box\neg(s \wedge \bigcirc s')\right) = 0$$

Como $\hat{\pi}^*_\Box$ solo le da probabilidad positiva a una $\alpha \in acc(s, V)$ por definición de $\mathrm{rand}(\sigma_\Box)$, podemos concentrarnos en esa única acción. mmm esto. Mejor vamos con

Sea α una acción y sea

Versión 1

Pedir que valga φ^l en $Derand(\mathcal{G}_\mathcal{K})$ es pedir que valga $\hat{\varphi}^l$ en $\mathcal{G}_\mathcal{K}$ para todo comportamiento en el juego y acción α en el comportamiento, vale que:

$$\hat{\varphi}^l = \bigwedge_{s' \in \mathrm{supp}(\alpha)} \left(\Box\Diamond\alpha \to \Box\Diamond(\alpha \wedge \bigcirc s')\right)$$

Capaz es notable el que por cada soporte V, por cómo está definido rand de estrategias, habrá una única acción α que se elije (por eso se sabe la equivalencia? por eso se sabe que habrá infinitas elecciones de α?).

Entonces, por cómo $\hat{\pi}^*_\Box$ está definida (a partir de $\sigma^*_\Box$) sabemos que

$$\inf_{\pi_\Diamond \in \Pi_\Diamond} \mathbb{P}^{\hat{\pi}^*_\Box, \pi_\Diamond}_{\mathcal{G}_\mathcal{K}, s} (\hat{\varphi}^l \to \varphi) = 1 \tag{5.6}$$

Veamos ahora que la ecuación 5.4 valdrá porque la probabilidad de que no valga $\hat{\varphi}^l$ es 0. Veamos:

Esto no me está quedando bien con el tema de que esto es para cada α que aparece en el comportamiento, pero medio que entendiendo que es algo del camino capaz está bien y ni vale la pena hacerse mucha cabeza

$$
\begin{aligned}
&\mathbb{P}^{\hat{\pi}^*_{\Box},\pi_{\Diamond}}_{\mathcal{G}_{\mathcal{K}},s}(\neg\hat{\varphi}^l) = \\
&\mathbb{P}^{\hat{\pi}^*_{\Box},\pi_{\Diamond}}_{\mathcal{G}_{\mathcal{K}},s}\left(\neg\left(\bigwedge_{s'\in\mathrm{supp}(\alpha)}(\Box\Diamond\alpha \to \Box\Diamond(\alpha\wedge\bigcirc s'))\right)\right) = \\
&\mathbb{P}^{\hat{\pi}^*_{\Box},\pi_{\Diamond}}_{\mathcal{G}_{\mathcal{K}},s}\left(\bigvee_{s'\in\mathrm{supp}(\alpha)}\neg(\Box\Diamond\alpha \to \Box\Diamond(\alpha\wedge\bigcirc s'))\right) = \\
&\mathbb{P}^{\hat{\pi}^*_{\Box},\pi_{\Diamond}}_{\mathcal{G}_{\mathcal{K}},s}\left(\bigvee_{s'\in\mathrm{supp}(\alpha)}(\Box\Diamond\alpha \wedge \Diamond\Box\neg(\alpha\wedge\bigcirc s'))\right) \leq \\
&\sum_{s'\in\mathrm{supp}(\alpha)}\mathbb{P}^{\hat{\pi}^*_{\Box},\pi_{\Diamond}}_{\mathcal{G}_{\mathcal{K}},s}\big(\Box\Diamond\alpha \wedge \Diamond\Box\neg(\alpha\wedge\bigcirc s')\big)
\end{aligned}
$$

Mostramos que este último término es igual a 0, viendo que para cada $s' \in \mathrm{supp}(\alpha)$,

$$\mathbb{P}^{\hat{\pi}^*_{\Box},\pi_{\Diamond}}_{\mathcal{G}_{\mathcal{K}},s}\big(\Box\Diamond\alpha \wedge \Diamond\Box\neg(\alpha\wedge\bigcirc s')\big) = 0$$

Sea ω una jugada. Notemos con μ a la distribución asociada a la acción α. Sea I el conjunto infinito de momentos donde se elije la acción α. La probabilidad de elegir s' como próximo vértice en cualquier momento $i \in I$ estará dada por $\mu(s')$.

Entonces, para cada momento i, la probabilidad de no visitar s' por los próximos k momentos i, está acotada superiormente por $(1-\mu(s'))^k$, que converge a 0 cuando k tiende de a ∞. Por lo tanto, tenemos que $\mathbb{P}^{\hat{\pi}^*_{\Box},\pi_{\Diamond}}_{\mathcal{G}_{\mathcal{K}},s}\big(\Box\Diamond\alpha \wedge \Diamond\Box\neg(\alpha\wedge\bigcirc s')\big) = 0$.

Queda raro esto porque no hay un estado s de por sí.

Consideremos un $s' \in \mathrm{supp}(\alpha)$ arbitrario. Sea ω una jugada y sea I el conjunto infinito de momentos donde se llega a s y se elige una acción $(K,\mu) \in acc(s, \mathrm{supp}(\alpha))$. Llamemos μ_i a la distribución elegida en el momento $i \in I$. La probabilidad de elegir s' como próximo vértice en el momento $i \in I$ estará dada por $\mu_i(s')$.

Como estamos trabajando con estrategias de memoria finita, habrá solo una cantidad finita de distribuciones que $\pi_{\Diamond}$ puede elegir desde s, así que existirá una distribución, llamémosla $\mu_{\text{máx}}$ tal que da la probabilidad máxima de ir a s'.

No tengo porqué hablar de acciones de memoria finita mepa. Voy a elegir infinitas veces la misma acción. Solo elijo una acción por cada soporte por la manera en la que definimos rand.

Entonces, para cada momento i, la probabilidad de no visitar s' por los próximos k momentos i, está acotada superiormente por $(1-\mu_{max}(s'))^l$, que converge a 0 cuando l tiende de a ∞. Por lo tanto, tenemos que $\mathbb{P}^{\hat{\pi}^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\Box\Diamond\alpha \wedge \Diamond\Box\neg(\alpha \wedge \bigcirc s')) = 0$

En consecuencia, se sigue que $\sum_{s'\in\text{supp}(\alpha)} \mathbb{P}^{\hat{\pi}^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}\left(\Box\Diamond\alpha \wedge \Diamond\Box\neg(\alpha \wedge \bigcirc s')\right) = 0$, lo que a su vez establece que $\mathbb{P}^{\hat{\pi}^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\neg\hat{\varphi}^l) = 0$.

Luego vale que:

$$\mathbb{P}^{\hat{\pi}^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\hat{\varphi}^l \rightarrow \varphi) = \mathbb{P}^{\hat{\pi}^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\neg\hat{\varphi}^l \vee \varphi) \leq$$
$$\mathbb{P}^{\hat{\pi}^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\neg\hat{\varphi}^l) + \mathbb{P}^{\hat{\pi}^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\varphi) = 0 + \mathbb{P}^{\hat{\pi}^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\varphi) =$$
$$\mathbb{P}^{\hat{\pi}^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\varphi)$$

Y por 5.6 entonces vale que $\inf_{\pi_\Diamond\in\Pi_\Diamond} \mathbb{P}^{\hat{\pi}^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\varphi) = 1$ como queríamos probar.

Segunda contención: $\mathcal{W}^{as} \subseteq \mathcal{W}$

Sea $s \in \mathcal{W}^{as}$, veamos que $s \in \mathcal{W}$.

Como $s \in \mathcal{W}^{as}$, entonces existe $\pi^*_\Box$ en $\mathcal{G}_\mathcal{K}$ tal que $\inf_{\pi_\Diamond\in\Pi_\Diamond} \mathbb{P}^{\pi^*_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\varphi) = 1$.

Lo que queremos ver para probar que $s \in \mathcal{W}$ es que existe una estrategia $\sigma^*_\Box$ tal que $\Box$ gana con ella desde s en $Derand(\mathcal{G}_\mathcal{K})$.

Definimos $\sigma^*_\Box = \text{derand}(\pi^*_\Box)$

Entonces ahora queremos ver que $\sigma_\Box *$ es ganadora en $Derand(\mathcal{G}_\mathcal{K})$ desde s.

Una manera de ver esto es probar que para un comportamiento cualquiera ρ generado por $\sigma^*_\Box$ y una estrategia válida $\pi_\Diamond \in \Pi_\Diamond$ arbitraria, *inf*(ρ) cumple con la especificación φ. Es decir, que vale la siguiente fórmula:

$$\varphi' := \bigvee_{j=1}^{k} \big((\mathit{inf}(\rho) \cap R_j = \emptyset) \wedge (\mathit{inf}(\rho) \cap G_j \neq \emptyset)\big) \tag{5.7}$$

Si podemos probar que $\mathit{inf}(\rho)$ es un sub-PMDP alcanzable en el PMDP que se forma al fijar la estrategia de memoria finita $\pi^*_\Box$ en $\mathcal{G}_\mathcal{K}$, por el teorema 5.1.2.1 podemos ver que vale la ecuación 5.7.

Primero, podemos ver que $\mathit{inf}(\rho)$ (entendiendo los vértices de la forma v_V como el conjunto resultado V) es una componente final en el polytopal markov decision process que se forma al fijar la estrategia $\pi^*_\Box$ en $\mathcal{G}_\mathcal{K}$, llamémoslo (C, D), viendo que cumple las dos propiedades:

- Para cada V^i que aparece como subíndice de vértices *especiales* en $\mathit{inf}(\rho)$, vale que $V^i \subseteq C$. En el caso de que existiese algún $s_x \in V^i$ tal que $s_x \notin C$, se estaría contradiciendo que $\pi_\Diamond$ sea una estrategia válida en $Derand(\mathcal{G}_\mathcal{K})$, puesto que visitaría infinitas veces v_{V^i} y solo finitas veces s_x, uno de sus sucesores.
- El grafo dirigido inducido por $\mathit{inf}(\rho)$ es fuertemente conexo. Si fuese de otra manera habría dos vértices $u, v \in \mathit{inf}(\rho)$ tales que v no sería alcanzable desde u, contradiciendo así que u y v son visitados infinitas veces por σ.

Luego, podemos ver que $\mathit{inf}(\rho)$ es alcanzable en $\mathcal{G}_\mathcal{K}$ viendo que existe una estrategia $\pi^*_\Diamond$ en $\mathcal{G}_\mathcal{K}$ que lo posibilita.

Definamos $\pi^*_\Diamond = \mathrm{rand}(\sigma_\Diamond)$.

Esta estrategia permite llegar con probabilidad positiva a $\mathit{inf}(\rho)$, puesto que tanto $\pi^*_\Diamond$ como $\pi^*_\Box$ le da probabilidad positiva a los mismos vértices que $\pi_\Diamond$ asegura visitar infinitamente. Capaz esto se podría explicar mejor.

Con lo cual hemos probado que vale 5.7 y, por lo tanto, que $\sigma_\Box *$ es ganadora en $Derand(\mathcal{G}_\mathcal{K})$ desde s, con lo que hemos probado que $s \in \mathcal{W}$.

$\Box$

Haber podido probar esta igualdad nos permite calcular el conjunto de estados casi seguramente ganadores para $\Box$ en $\mathcal{G}_{\mathcal{K}}$ mediante el algoritmo que se propone en [6] para calcular el conjunto de vértices ganadores en un juego de adversario justo G con un objetivo de Rabin R. Este algoritmo tiene una complejidad de $O(n^2 k!)$ donde n es la cantidad de vértices en G y k la cantidad de pares en R.

Esto quiere decir que podemos calcular el conjunto de estados casi seguramente ganadores para $\Box$ en un PSG $\mathcal{K}$ con un objetivo de Rabin R con una complejidad de $O((nl)^2 k!)$ donde n es la cantidad de estados en $\mathcal{K}$, k es la cantidad de pares en R y $l = \text{máx}\{|V_s| \mid s \in \mathcal{S}\}$ es la cantidad máxima de conjuntos soporte para las acciones que puede haber desde un estado s en $\mathcal{K}$.

¿Agregar algo sobre cómo es el algoritmo? ¿Sobre el teorema en sí? ¿Dejar el segundo párrafo como teorema? ¿Qué vamos a hablar de complejidad en PSG Rabin?

¿Cómo represento gráficamente la idea de politopos? ¿Capaz dándoles forma a las acciones?

5.3. Transformar Rabin en alcanzabilidad: la respuesta a varias preguntas

5.3.1. Draft

La idea es intentar hacer algo al estilo

Teorema 5.3.1 (Reducción de Rabin). *Sea $\mathcal{G}_{\mathcal{K}}$ un juego estocástico politópico y sea $\mathcal{H}_{\mathcal{K}}$ su interpretación extrema tal como se presenta en [3]. Si tenemos una propiedad de Rabin R, y su respectivo conjunto de estados almost sure winning W_R entonces valen las siguientes ecuaciones:*

$$
\begin{aligned}
&\inf_{\pi_\Diamond \in \Pi_{\mathcal{G}_\mathcal{K},\Diamond}} \sup_{\pi_\Box \in \Pi_{\mathcal{G}_\mathcal{K},\Box}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(R) = \\
&= \inf_{\pi_\Diamond \in \Pi_{\mathcal{G}_\mathcal{K},\Diamond}} \sup_{\pi_\Box \in \Pi_{\mathcal{G}_\mathcal{K},\Box}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\Diamond W_R) = \\
&= \inf_{\pi_\Diamond \in \Pi^{MD}_{\mathcal{H}_\mathcal{K},\Diamond}} \sup_{\pi_\Box \in \Pi^{MD}_{\mathcal{H}_\mathcal{K},\Box}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{H}_\mathcal{K},s}(\Diamond W_R) = \\
&= \sup_{\pi_\Box \in \Pi^{MD}_{\mathcal{H}_\mathcal{K},\Box}} \inf_{\pi_\Diamond \in \Pi^{MD}_{\mathcal{H}_\mathcal{K},\Diamond}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{H}_\mathcal{K},s}(\Diamond W_R) = \\
&= \sup_{\pi_\Box \in \Pi_{\mathcal{G}_\mathcal{K},\Box}} \inf_{\pi_\Diamond \in \Pi_{\mathcal{G}_\mathcal{K},\Diamond}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\Diamond W_R) = \\
&= \sup_{\pi_\Box \in \Pi_{\mathcal{G}_\mathcal{K},\Box}} \inf_{\pi_\Diamond \in \Pi_{\mathcal{G}_\mathcal{K},\Diamond}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(R)
\end{aligned}
$$

Demostración.

$$
\begin{aligned}
&\inf_{\pi_\Diamond \in \Pi_{\mathcal{G}_\mathcal{K},\Diamond}} \sup_{\pi_\Box \in \Pi_{\mathcal{G}_\mathcal{K},\Box}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(R) \\
&\quad \leq \inf_{\pi_\Diamond \in \Pi_{\mathcal{G}_\mathcal{K},\Diamond}} \sup_{\pi_\Box \in \Pi_{\mathcal{G}_\mathcal{K},\Box}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\Diamond W_R) && \text{(por lema 5.3.2)} \\
&\quad \leq \inf_{\pi_\Diamond \in \Pi^{MD}_{\mathcal{H}_\mathcal{K},\Diamond}} \sup_{\pi_\Box \in \Pi^{MD}_{\mathcal{H}_\mathcal{K},\Box}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{H}_\mathcal{K},s}(\Diamond W_R) && \text{(por Teorema 1)} \\
&\quad \leq \sup_{\pi_\Box \in \Pi^{MD}_{\mathcal{H}_\mathcal{K},\Box}} \inf_{\pi_\Diamond \in \Pi^{MD}_{\mathcal{H}_\mathcal{K},\Diamond}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{H}_\mathcal{K},s}(\Diamond W_R) && \text{(por Teorema 1)} \\
&\quad \leq \sup_{\pi_\Box \in \Pi_{\mathcal{G}_\mathcal{K},\Box}} \inf_{\pi_\Diamond \in \Pi_{\mathcal{G}_\mathcal{K},\Diamond}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\Diamond W_R) && \text{(por Teorema 1)} \\
&\quad \leq \sup_{\pi_\Box \in \Pi_{\mathcal{G}_\mathcal{K},\Box}} \inf_{\pi_\Diamond \in \Pi_{\mathcal{G}_\mathcal{K},\Diamond}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(R) && \text{(por lema 5.3.4)} \\
&\quad \leq \inf_{\pi_\Diamond \in \Pi_{\mathcal{G}_\mathcal{K},\Diamond}} \sup_{\pi_\Box \in \Pi_{\mathcal{G}_\mathcal{K},\Box}} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(R) && \text{(por prop de sup e inf)}
\end{aligned}
$$

□

Lema 5.3.2. *Sea $\mathcal{G}_\mathcal{K}$ la interpretación (?) de un juego estocástico politópico y sea R una propiedad límite y W_R su correspondiente conjunto casi seguramente ganador. Entonces vale que*

$$\inf_{\pi_\Diamond \in \Pi_\Diamond} \sup_{\pi_\Box \in \Pi_\Box} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(R) \leq \inf_{\pi_\Diamond \in \Pi_\Diamond} \sup_{\pi_\Box \in \Pi_\Box} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s}(\Diamond W_R)$$

Demostración. Nombremos $\pi^*_\Diamond$ a una estrategia tal que

$$\sup_{\pi_\Box \in \Pi_\Box} \mathbb{P}^{\pi_\Box, \pi^*_\Diamond}_{\mathcal{G}_\mathcal{K}, s}(\Diamond W_P) = \inf_{\pi_\Diamond \in \Pi_\Diamond} \sup_{\pi_\Box \in \Pi_\Box} \mathbb{P}^{\pi_\Box, \pi_\Diamond}_{\mathcal{G}_\mathcal{K}, s}(\Diamond W_P)$$

(explicar algo más hablado? decir algo sobre su existencia?). Ahora podemos hacer las siguientes deducciones:

$$\begin{aligned}
&\inf_{\pi_\Diamond \in \Pi_\Diamond} \sup_{\pi_\Box \in \Pi_\Box} \mathbb{P}^{\pi_\Box, \pi_\Diamond}_{\mathcal{G}_\mathcal{K}, s}(P) \\
&\quad \leq \sup_{\pi_\Box \in \Pi_\Box} \mathbb{P}^{\pi_\Box, \pi^*_\Diamond}_{\mathcal{G}_\mathcal{K}, s}(P) && \text{(por def. de inf)} \\
&\quad = \sup_{\pi_\Box \in \Pi_\Box} \mathbb{P}^{\pi_\Box, \pi^*_\Diamond}_{\mathcal{G}_\mathcal{K}, s}(\Diamond W_P) && \text{(por lema 5.3.3)} \\
&\quad = \inf_{\pi_\Diamond \in \Pi_\Diamond} \sup_{\pi_\Box \in \Pi_\Box} \mathbb{P}^{\pi_\Box, \pi_\Diamond}_{\mathcal{G}_\mathcal{K}, s}(\Diamond W_P) && \text{(por def. de } \pi^*_\Diamond\text{)}
\end{aligned}$$

y con esto hemos probado el enunciado. □

Lema 5.3.3. *Sea $\mathcal{M}$ un PMDP, π una estrategia en $\mathcal{M}$, s un estado en $\mathcal{M}$, R una propiedad de Rabin y W_R el conjunto casi seguramente ganador de R (en el juego estocástico debería ser, no?), entonces vale que:*

¿Cambiar wr por conjunto desde el que eciste una estrategia que da probabilidad 1 de ganar?

$$\sup_{\pi \in \Pi} \mathbb{P}^{\pi}_{\mathcal{M}, s}(R) = \sup_{\pi \in \Pi} \mathbb{P}^{\pi}_{\mathcal{M}, s}(\Diamond W_R)$$

Demostración. Por definición de conjunto casi seguramente ganador

$$\mathbb{P}^{\pi}_{\mathcal{M}, s}(R) = \mathbb{P}_{\mathcal{M}, s}(\{\omega \in \mathit{Paths}(s) \mid \mathit{inf}(\omega) \models R\})$$

Claramente los caminos que hacen que valga la propiedad también son caminos desde donde llego a algún estado desde donde existen estrategias para ganar con probabilidad 1. Por lo tanto, $\mathbb{P}^{\pi}_{\mathcal{M}, s}(R) \leq \mathbb{P}^{\pi}_{\mathcal{M}, s}(\Diamond W_R)$.

Para ver que vale la igualdad, veremos que existe una estrategia de memoria finita π que hace que $\mathbb{P}^{\pi}_{\mathcal{M},s}(R) = \sup_{\pi\in\Pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(\Diamond W_R)$. Para ello, consideremos la estrategia sin memoria π_0 que maximiza las probabilidades de alcanzar W_R desde todos los estados $s \in \mathcal{M}$ (sabemos que esta estrategia existe por [3, 4]). A su vez, sabemos que para cada estado $s \in W_R$ existe una estrategia π_s que asegura ganar con probabilidad 1 frente al objetivo R.

Sea entonces π la estrategia que primero se comporta como π_0, hasta llegar a un estado t en W_R, y a partir de allí π se comporta como π_t. Con eso tenemos que:

$$\begin{aligned}\mathbb{P}^{\pi}_{\mathcal{M},s}(R) &= \sum_{t\in W_R} \mathbb{P}^{\pi_0}_{\mathcal{M},s}((\neg W_R \,\mathcal{U}\, t)) \cdot \underbrace{\mathbb{P}^{\pi_t}_{\mathcal{M},t}(R)}_{=1} \\ &= \sup_{\pi\in\Pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(W_R)\end{aligned}$$

Como $\sup_{\pi\in\Pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(W_R)$ es una cota superior las probabilidades para R bajo todas las estrategias, con esto podemos concluir la igualdad $\sup_{\pi\in\Pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(R) = \sup_{\pi\in\Pi} \mathbb{P}^{\pi}_{\mathcal{M},s}(\Diamond W_R)$. $\square$

Esto es lo importante que habría que probar y no está probado.

Lema 5.3.4. *Sea $\mathcal{G}_{\mathcal{K}}$ la interpretación (?) de un juego estocástico politópico y sea R una propiedad límite y W_R su correspondiente conjunto casi seguramente ganador. Entonces vale que*

$$\sup_{\pi_\Box\in\Pi_\Box} \inf_{\pi_\Diamond\in\Pi_\Diamond} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_{\mathcal{K}},s}(\Diamond W_R) \leq \sup_{\pi_\Box\in\Pi_\Box} \inf_{\pi_\Diamond\in\Pi_\Diamond} \mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_{\mathcal{K}},s}(R)$$

¿Agregar una sección hablando de algo de complejidad en PSG con objetivos de Rabin?

Capítulo 6

Conclusiones

6.1. Trabajo Futuro

Referencias

[1] C. Baier y J.-P. Katoen. *Principles of Model Checking*. Vol. 26202649. The MIT Press, 2008. ISBN: 978-0-262-02649-9.

[2] L. de Alfaro. *Formal Verification of Probabilistic Systems*. Inf. téc. Stanford, CA, USA, 1998.

[3] P. F. Castro y P. R. D'Argenio. «Polytopal Stochastic Games». En: (2024). Enviado para publicación.

[4] A. Condon. «The complexity of stochastic games». En: *Information and Computation* 96.2 (1992), págs. 203-224. ISSN: 0890-5401. DOI: https://doi.org/10.1016/0890-5401(92)90048-K. URL: https://www.sciencedirect.com/science/article/pii/089054019290048K.

[5] J. Filar y K. Vrieze. *Competitive Markov Decision Processes*. Springer New York, 2012. ISBN: 9781461240549. URL: https://books.google.com.ar/books?id=uXDjBwAAQBAJ.

[6] T. Banerjee, R. Majumdar, K. Mallik, A.-K. Schmuck y S. Soudjani. «A Direct Symbolic Algorithm for Solving Stochastic Rabin Games». En: *Tools and Algorithms for the Construction and Analysis of Systems*. Ed. por D. Fisman y G. Rosu. Cham: Springer International Publishing, 2022, págs. 81-98. ISBN: 978-3-030-99527-0.

Apéndice A

Titulo del Apendice

A.1. Titulo de la seccion

Hola