Resumen

Los juegos estocásticos han servido para modelar diversos sistemas del mundo real donde se aprecia un comportamiento estocástico y adversarial. Ejemplos de esto pueden ser „, „,.

En 2025, Castro y D'Argenio publican el paper *Polytopal Stochastic Games*, en donde por primera vez se presenta el concepto de juego estocástico politópico, respondiendo a la necesidad de modelar juegos estocásticos que puedan capturar mayor incertidumbre sobre las distribuciones de probabilidad que determinan las acciones que toman los distintos jugadores. En el paper estos son estudiados en relación a funciones de recompensa y objetivos de alcanzabilidad.

Este trabajo se concentrará en extender el estudio de juegos estocásticos politópicos con objetivos de Rabin. Los objetivos de Rabin permiten describir especificaciones sobre conjuntos de estados que deben ser visitados infinitas veces y conjuntos de estados que deben ser visitados una cantidad finita de veces. A lo largo del trabajo nos adentraremos sobre por qué resulta interesante estudiar juegos estocásticos politópicos, por qué resulta de importancia estudiar objetivos de Rabin y veremos cómo dar respuestas a las preguntas de quién gana y de cómo se gana un juego estocástico politópico con objetivo de Rabin.

Índice general

Capítulo 1

Introducción

A lo largo de la historia, la teoría de juegos ha cumplido un rol central en muchas áreas de las ciencias de la computación. En particular, la aparición del concepto de *juego estocástico* ha tenido un impacto significativo en la modelización y estudio de sistemas reactivos, sistemas con comportamientos no deterministas, y sistemas con incertidumbre.

El concepto de juego estocástico se contrapone al de juego determinista. Mientras que en el último cada acción de un jugador resulta en un único próximo estado del juego, en el primero cada acción de un jugador determina una distribución de probabilidad sobre los posibles próximos estados del juego. Es por esto que los juegos estocásticos han resultado una herramienta útil para modelar sistemas en diversas áreas como finanzas, inteligencia artificial y telecomunicaciones.

Existen sistemas donde las posibles transiciones de un estado a otro no son discretas, sino que existen una cantidad continua de opciones, y estas se suelen poder describir a partir de un conjunto de restricciones. Ejemplos de estos sistemas son redes de comunicaciones, modelos financieros complejos y planificaciones de movimiento en robótica. Para modelar este tipo de situaciones surgió el concepto de juegos estocásticos politópicos. En ellos, los posibles siguientes estados de un juego forman diferentes politopos en el espacio de estados, y la decisión de un jugador lo que implicará la elección de un politopo y una distribución de probabilidad sobre ese politopo, a partir de los cuales se verá cuál es el próximo estado del juego.

Para su estudio, y en general para el estudio de juegos, la pregunta central a responder es si un jugador tiene una estrategia ganadora. Esto dependerá de cómo se define el ganar para el jugador. En juegos estocásticos, una manera de definir el ganar es a través de *objetivos* dados en forma de una *propiedad ω-regular ϕ*. Como nos concentraremos en juegos de suma cero, el ganar para un jugador será que se cumpla ϕ, mientras que su contrincante ganará si se cumple $\neg\phi$.

Dentro de las propiedades ω-regulares se destacan *las condiciones de Rabin* y su dual, *las condiciones de Street*. La relevancia práctica de ellas viene del hecho de que su forma se corresponde con aquella de las condiciones de equidad en sistemas de transición, además de que cualquier propiedad ω-regular puede ser reescrita como una propiedad de Rabin [1].

Una vez fijado un jugador (generalmente el jugador "maximizador"/ el jugador que busca que se cumpla la propiedad ϕ), al problema de la computación de los estados estados ganadores de un juego se lo suele llamar el análisis cualitativo del juego, mientras que el análisis cuantitativo del juego refiere al problema de computar, para cada estado, cuál es la probabilidad máxima que tiene este de ganar, aún cuando esta sea menor que 1.

En el trabajo nos concentraremos en el análisis cualitativo de juegos estocásticos politópicos con objetivos de Rabin. Mostraremos que el conjunto de sus estados ganadores son iguales a los estados ganadores de un particular juego determinista construido a partir del juego estocástico, y presentaremos un algoritmo para su cómputo y analizaremos su complejidad.

1.1. Estado del arte

Los juegos estocásticos fueron introducidos por L. Shapley en 1953 [2] con el fin de modelar interacciones dinámicas donde el entorno cambia en respuesta a los comportamientos de los jugadores. Estos extienden el modelo de los procesos de decisión de Markov, desarrollado por varios investigadores en la RAND Corporation en el período de 1949 a 1952, a situaciones competitivas con más de un tomador de decisiones.

Gracias a su aporte para la modelización en varias áreas, resultó lógica la división de juegos estocásticos en base a distintos tipos de funciones de pago. El estudio de la existencia, tipo y computabilidad de las estrategias óptimas suele considerarse en juegos con funciones de recompensa cuantitativas (eg., funciones de recompensa media, descontada, pesada o límite) y cualitativas (eg., funciones de recompensas asociadas a objetivos de alcanzabilidad, de Büchi, de Rabin o de Müller) [3, 4, 5].

Resultados para juegos estocásticos con funciones de recompensa media y descontada fueron reportados desde 1958 por Gillete [6], mientras que los primeros resultados para juegos con objetivos de alcanzabilidad fueron publicados por Condon en 1992 [7] y los juegos con objetivos de Rabin comenzaron a ser estudidados en profundidad por Chaterjee desde 2005 [3, 5, 8].

Recientemente, en 2024, Castro y D'Argenio introdujeron el concepto de juegos estocásticos politópicos [9], los cuales expanden la noción de juegos estocásticos. En este mismo paper, se presentan resultados para juegos estocásticos politópicos con objetivos de alcanzabilidad, funciones de pago descontadas y funciones de pago promedio. Nuestro objetivo aquí será extender el estudio de los juegos estocásticos politópicos con objetivos de Rabin.

1.2. Organización del trabajo

- En el **Capítulo 2**, nos adentraremos en el campo de la verificación de modelos con juegos, presentando definiciones básicas de modelos y juegos acompañadas con ejemplos que servirán de guía para el entendimiento de los mismos;

- En el **Capítulo 3**, presentaremos en profundidad los objetivos ω-regulares haciendo especial hincapié en los objetivos de Rabin;

- En el **Capítulo 4**, veremos definiciones formales referidas a los juegos estocásticos politópicos y los resultados obtenidos en el paper *Polytopal Stochastic Games* [9];

- En el **Capítulo 5**, desarrollaremos los resultados originales de esta tesina, presentando en la sección 5.1 a los procesos de decisión de Markov politópicos con sus teoremas asociados y mostrando en la sección 5.2 cómo calcular el conjunto de estados ganadores con probabilidad 1 para el jugador maximizador en un juego estocástico politópico con objetivo de Rabin, y

- Finalmente, en el **Capítulo 6**, concluiremos con un resumen de los aportes realizados en esta tesina y nos expandiremos sobre potenciales caminos para futuras investigaciones en el tema.

Capítulo 2

Verificación de modelos con juegos

En este capítulo presentaremos brevemente la verificación de modelos con juegos. Para ello comenzaremos dando un contexto sobre el campo de estudio de la verificación de modelos y luego presentaremos unos modelos matemáticos que sirven para la representación de distintos tipos de sistemas (las cadenas de Markov, los procesos de decisión de Markov, los juegos estocásticos y los juegos -de grafo- deterministas).

2.1. Sobre la verificación de modelos

La verificación de modelos (*model checking* en inglés) es una técnica prominente dentro de la verificación formal que sirve para evaluar las propiedades funcionales de los sistemas de información y comunicación de manera automatizada. Esta técnica requiere de un modelo del sistema en cuestión y una propiedad deseada, y verifica sistemáticamente si el modelo dado satisface dicha propiedad.

Existen numerosas propiedades de interés para comprobar (a destacar podrían ser la ausencia de interbloqueos o propiedades de solicitud-respuesta) y, a su vez, existen muchos modelos matemáticos que permiten distintas representaciones de sistemas reales, cada uno con sus particularidades. La idea de las próximas secciones será presentar varios de estos modelos, ver ciertas propiedades de los mismos y ver, como motivación, un ejemplo de sistema representable con ellos.

2.2. Cadenas de Markov

El primer modelo matemático que veremos son las cadenas de Markov. Es importante notar que trabajaremos con el tratamiento de las cadenas de Markov como sistemas de transición anotados con probabilidades. Esto en contraposición al enfoque que plantea a las cadenas de markov como una familia de variables aleatorias. Pasemos entonces a su definición.

Definición 2.2.1 (Cadena de Markov)**.** *Una cadena de Markov es una tupla* $M = (\mathcal{S}, P)$ *donde* $\mathcal{S}$ *es un conjunto finito no vacío de estados y* $P : \mathcal{S} \times \mathcal{S} \to [0, 1]$ *es una función tal que para todos los estados* s *vale que*

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

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 representar directamente 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. Es decir, son definidos como secuencias infinitas de estados $\omega = (s_0, s_1, s_2, \dots) \in \mathcal{S}^\omega$ tales que $P(s_i, s_{i+1}) > 0$ para todo $i \geq 0$. A su vez, dado un camino infinito $\omega = (s_0, s_1, \dots)$, podemos pensar en los prefijos finitos de ese camino, $\text{pref}(\omega) = \{(s_0, \dots, s_n) | n \in \mathbb{N}\}$. Notaremos con $\text{Paths}(M)$ al conjunto de todos los caminos en M y con $\text{Paths}_{\text{fin}}(M)$ al conjunto de todos los prefijos finitos de caminos en M.

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

Randomito, el peuqeño robot probabilístico Supongamos que queremos observar el comportamiento de un pequeño robot llamado *Randomito*.

Randomito se encuentra en una grilla 2x2 y se comporta totalmente probabilísticamente siguiendo la siguiente descripción: desde su posición inicial de quietud en la esquina superior izquierda tiene una probabilidad de 0.5 de seguir allí y 0.25 de moverse tanto a la izquierda como hacia abajo. Si sucede que en algún momento se encuentra en la esquina superior o inferior derecha, como a *Randomito* no le gusta quedarse del lado de la derecha, tiene probabilidad 0 de quedarse allí o de ir hacia arriba o abajo, respectivamente. Desde esas esquinas, *Randomito* se mueve con probabilidad 1 hacia la izquierda. Desde la esquina inferior izquierda tendrá una probabilidad de 0.3 de quedarse allí, 0.2 de ir a la derecha y 0.5 de ir hacia arriba.

El comportamiento de *Randomito* se puede modelar con la cadena de Markov que representa el siguiente grafo:

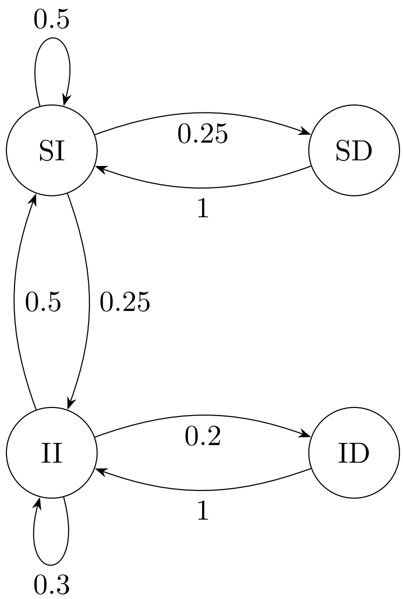

Figura 2.1: Cadena de Markov representando el comportamiento probabilístico de *Randomito*.

Caminos posibles de esta cadena de Markov podrían ser:

- el que mantiene a *Randomito* siempre en la esquina superior izquierda, $(SI, SI, SI, \dots)$, o
- el describe un primer movimiento en “C”de *Randomito* y luego se comporta de cualquier manera, $(SI, SD, SI, II, ID, II, SI, \dots)$.

Una duda natural entonces que podría surgir es ¿qué probabilidad hay de que *Randomito* siga alguno de esos caminos?

Para poder asociar probabilidades a eventos en cadenas de Markov (o, tal vez pensar en preguntas sobre la probabilidad de que *Randomito* siga algún tipo particular de camino), la noción intuitiva de probabilidades en M debe ser formalizada. Lograremos esto asociándole un espacio de probabilidad a M. Los caminos infinitos de M juegarán el rol de resultados de la σ-álgebra asociada. Esto es $Outc^M = \text{Paths}(M)$. Y la σ-álgebra asociada a M será la generada por los conjuntos cilindro formados por los fragmentos de caminos finitos en M.

Definición 2.2.2 (Conjunto cilindro). *El conjunto cilindro de* $\hat{\omega} = (s_0, \ldots, s_n) \in \text{Paths}_{\text{fin}}(M)$ *está definido como*

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

Definición 2.2.3 (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* $\text{Cyl}(\hat{\omega})$ *donde* $\hat{\omega} \in \text{Paths}_{\text{fin}}(M)$.

De conceptos clásicos de teoría de probabilidad se sigue entonces que para 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:

$$\mathbb{P}_s^M(Cyl(s_0, \ldots, s_n)) = \iota(s, s_0) \cdot \prod_{0 \leq i < n} P(si, s_{i+1}), \text{ donde } \iota(s, s_0) = \begin{cases} 1 & \text{si } s = s_0 \\ 0 & \text{en otro caso} \end{cases} \tag{2.1}$$

¿Qué probabilidad hay de que Randomito siga este camino?

Ahora, con las definiciones vistas, podemos pensar en la probabilidad que tiene *Randomito* de tomar algunos caminos en específico. Por ejemplo,

- la probabilidad de que *Randomito* se quede siempre en la esquina superior izquierda, aún empezando desde allí, resulta 0 porque $\mathbb{P}_{SI}^M((SI, SI, SI, \ldots)) = \text{lím}_{n \to \infty}(0{,}5)^n = 0$.

- la probabilidad de que *Randomito* haga un primer movimiento en "C" y luego se comporte de cualquier manera será $\mathbb{P}^{M}_{SI}(\mathrm{Cyl}(SI, SD, SI, II, ID, II, SI)) = 0{,}25 \cdot 1 \cdot 0{,}25 \cdot 0{,}2 \cdot 1 \cdot 0{,}5 = 0{,}00625$.

Notación: en lo que sigue, usaremos notación LTL para describir ciertos eventos en cadenas de Markov. Por ejemplo, para un conjunto $B \subseteq \mathcal{S}$ 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

Decir algo mas de notacion LTL? Algo más de cadenas de markov?

2.3. Procesos de Decisión de Markov

Las cadenas de Markov resultan útiles cuando queremos modelar sistemas en donde existen decisiones probabilistas, pero existen situaciones en donde además de decisiones probabilistas, necesitamos modelar elecciones, es decir, decisiones no-deterministas. Para ello, existen los llamados 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 $\mathcal{A}$ y a continuación presentaremos la definición de un proceso de decisión de Markov:

Definición 2.3.1 (Proceso de Decisión de Markov)**.** *Un proceso de decisión de Markov $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \theta)$ consiste de un conjunto finito de estados $\mathcal{S}$ y de dos componentes, $\mathcal{A}$ y θ, que especifican la estructura de transición:*

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

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

El concepto de caminos en MDPs que mostraremos en este trabajo difiere del concepto de camino que se puede encontrar en alguna literatura y el que presentamos para cadenas de Markov, donde estos eran solo secuencias de estados. Un camino 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 $a \in \mathcal{A}(s)$ 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 2.3.2 (Camino en un MDP)**.** *Un camino en un MDP* $\mathcal{M}$ *es una secuencia infinita* $\omega = (s_0, a_0, s_1, a_1, \dots)$ *tal que* $s \in \mathcal{S}$*,* $a_i \in \mathcal{A}(s_i)$ *y* $a_i(s_{i+1}) > 0$ *para todo* $i \geq 0$*.*

Dado un estado s*, indicaremos con* Paths_s *el conjunto de todos los caminos que se originan en* s*, con* Paths *el conjunto de todos los caminos en* $\mathcal{M}$ *y con* $\mathrm{Paths}_{\mathrm{fin}}$ *el conjunto de todos los prefijos finitos de caminos en* $\mathcal{M}$ *que terminan en un estado* $s \in \mathcal{S}$*.*

Veamos un ejemplo de un proceso de decisión de Markov.

Roborto, el robot controlable

Ahora supongamos que tenemos un robot en una grilla 2x2, *Roborto*, al que podemos manejar a través de un control remoto. Con este control, podemos decidir si se mueve a la izquierda, derecha, arriba o abajo (dependiendo de lo que permita su posición en la grilla). Sin embargo, por irregularidades que puede haber en la grilla cuando seleccionamos una acción habrá una probabilidad de que la acción no tenga el resultado deseado. Es decir, si estando en la esquina superior izquierda, elijo que *Roborto* se mueva a la derecha, esto sucederá con una probabilidad p, pero por las irregularidades del terreno, también puede suceder que se mueva hacia abajo con una probabilidad $1-p$.

Situaremos a *Roborto* en una grilla particular con las siguientes características:

- desde la esquina superior izquierda, si se elije ir a la derecha, habrá una probabilidad de 0,9 de efectivamente ir a la esquina superior derecha y habrá una probabilidad de 0,1 de ir a la esquina inferior izquierda; mientras que si se elije ir hacia abajo, con una probabilidad

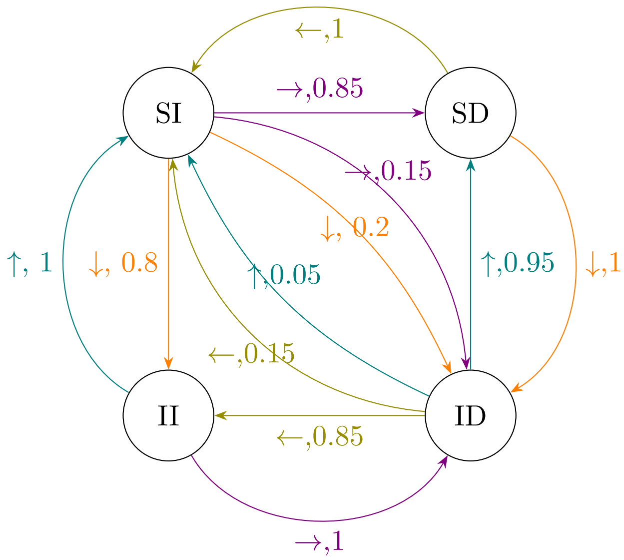

Figura 2.2: Proceso de decisión de Markov representando el comportamiento de *Roborto*.

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 de caminos. Para los MDPs, esto es levemente distinto. Como no hay restricciones en la resolución de las elecciones no deterministas, no hay una única medida de probabilidad asociada a cada estado.

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

Definición 2.3.3 (Estrategia en un MDP)**.** *Sea $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \theta)$ un MDP. Una estrategia para $\mathcal{M}$ es una función $\pi : \mathrm{Paths}_{\mathrm{fin}} \to \mathsf{Dist}(\mathcal{A})$ que asigna una distribución de probabilidad a cada prefijo finito de camino tal que $\pi(\hat{\omega})(a) > 0$ solo si $a \in \mathcal{A}(s)$.*

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 caminos en $\mathcal{M}$.

Definición 2.3.4 (Cadena de Markov de un MDP inducida por una estrategia)**.** *Sea $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \theta)$ un MDP y π una estrategia en $\mathcal{M}$. La cadena de Markov $\mathcal{M}_\pi$ está dada por*

$$\mathcal{M}_\pi = (\mathrm{Paths}_{\mathrm{fin}}, P_\pi)$$

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 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 2.1).

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 2.3.5 (Estrategias sin memoria)**.** *Sea $\mathcal{M}$ un MDP con espacio de estados $\mathcal{S}$. Una estrategia π en $\mathcal{M}$ es sin memoria sii para cada par de caminos $(s_0, a_0, \ldots, s_n)$ 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 $\pi : \mathcal{S} \to \mathsf{Dist}(\mathcal{A})$.

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 camino. 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 2.3.6 (Estrategias con memoria finita)**.** *Sea $\mathcal{M}$ un MDP con espacio de estados $\mathcal{S}$ y conjunto de acciones $\mathcal{A}$. 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,*
- *$\Delta : Q \times \mathcal{A} \times \mathcal{S} \to Q$ es la función de transición del autómata,*
- *$start : \mathcal{S} \to Q$ es la función que determina el modo en el que empieza el automáta para un estado inicial s,*
- *$f_\pi : Q \times \mathcal{S} \to \mathsf{Dist}(\mathcal{A})$ 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 $\hat{\omega} \in \mathrm{Paths}_{\mathrm{fin}}$*, 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 penuitima 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 $\hat{\omega} \in \mathrm{Paths}_{\mathrm{fin}}$ 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 [10, 11]) 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 $\pi : \mathrm{Paths}_{\mathrm{fin}} \to \mathcal{A}$.

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

2.4. Juegos estocásticos

Además de elecciones podemos querer también modelar comportamiento adversarial. Este comportamiento adversarial, donde surgen distintos agentes con distintos objetivos, se ha modelado históricamente a través de juegos. Si estos juegos exhiben también comportamiento probabilístico son llamados juegos estocásticos.

Definir cómo queda con lo de acciones habilitadas esto o MDP ¿qué?

Definición 2.4.1 (Juego estocástico)**.** *Un juego estocástico (SG) es una tupla* $\mathcal{G} = (\mathcal{S}, (\mathcal{S}_\Box, \mathcal{S}_\Diamond), \mathcal{A}, \theta)$ *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*

- $\theta : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0,1]$ *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 $\mathcal{S}_\Box = \emptyset$ o $\mathcal{S}_\Diamond = \emptyset$, entonces $\mathcal{G}$ es un proceso de decisión de Markov, y, como vimos antes, si a la vez $\big|\mathcal{A}(s)\big| = 1$ 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 camino en un juego estocástico, el cual seguirá la misma idea del definido para MDPs: además de estados, posee acciones.

Definición 2.4.2 (Camino en un SG)**.** *Un camino 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* Paths_s *el conjunto de todos los caminos que se originan en* s*, con* Paths *el conjunto de todos los caminos en* $\mathcal{G}$ *y con* $\mathrm{Paths}^i_{\mathrm{fin}}$ *el conjunto de todos los prefijos finitos de caminos en* $\mathcal{G}$ *que terminan en un estado* $s \in \mathcal{S}_i$*, con* $i \in \{\Box, \Diamond\}$.

Ahora sí, podemos definir las dos clases de estrategias que tendremos en un juego estocástico.

Definición 2.4.3 (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* $\pi : \mathrm{Paths}^i_{\mathrm{fin}} \to \mathsf{Dist}(\mathcal{A})$ *que asigna una distribución de probabilidad a cada prefijo finito de camino que termina en un estado del jugador* i *tal que* $\pi(\hat{\omega})(a) > 0$ *solo si* $a \in \mathcal{A}(s)$.

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. 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 $\mathbb{P}_{\mathcal{G}_\mathcal{K},s}^{\pi_\Box,\pi_\Diamond}$ en la σ-álgebra de Borel del conjunto de caminos en $\mathcal{G}$. Si ε es un conjunto medible en la σ-álgebra de Borel, $\mathbb{P}_{\mathcal{G}_\mathcal{K},s}^{\pi_\Box,\pi_\Diamond}(\varepsilon)$ será la probabilidad de que las estrategias $\pi_\Box$ y $\pi_\Diamond$ sigan un comportamiento en ε empezando desde el estado s.

2.5. Juegos deterministas

Ahora bien, dijimos que cuando queremos modelar tanto comportamiento adversarial como probabilístico es cuando necesitamos de juegos estocásticos, pero resulta que también es muy común querer modelar simplemente comportamiento adversarial. Para esto, existen muchas modelizaciones matemáticas del concepto de juego (al que podemos considerar "determinista" por no ser estocástico). Para las definiciones que presentaremos a continuación nos basaremos en lo que se suele conocer en la literatura como juegos de grafo de dos jugadores. Dentro del estudio de juegos de grafos se presentan varias categorías que se pueden paralelizar con las definiciones que vimos: un juego de grafo de un jugador sería simplemente un sistema de transición, un juego de grafo de 1 jugador y medio sería lo que se entiende por proceso de decisión de Markov y un juego de grafo de 2 jugadores y medio sería lo que presentamos como juegos estocásticos.

Procedemos entonces con la definición de juego de grafo de dos jugadores:

Definición 2.5.1 (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* $V = V_\Box \biguplus V_\Diamond$ *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.*

Puede ser de interés notar que en el contexto de juegos de grafo hablamos de vértices y no de estados como en los juegos estocásticos. Además, para los juegos deterministas volveremos a la noción de camino como secuencia de vertices solamente, a los cuales notaremos generalmente con ρ en contraposición a los caminos siendo notados con π para juegos estocásticos.

Definición 2.5.2 (jugada sobre un 2G)**.** *Una jugada en un juego de grafo de dos jugadores es una secuencia infinita de vértices* $\rho = v_0 v_1 v_2 \cdots \in V^\omega$*, donde para todo* $i \in \mathbb{N}_0$ *tenemos que* $v^i \in V$ *y* $(v^i, v^{i+1}) \in E$*.*

También para el análisis de juegos deterministas introduciremos el concepto de estrategia. Esta vez no para poder definir una medida de probabilidad ya que no necesitaremos una, sino para poder estudiar clases de comportamiento adversarial y poder formular y responder preguntas de investigación que se hacen en el contexto del estudio de juegos, las que precisaremos con más detalle en el capítulo siguiente.

Definición 2.5.3 (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* $wv \in V^* V_i$*.*

Otra vez, podemos presentar dos clases distintas de estrategias: estrategias sin memoria y con memoria finita, cuyas definiciones serán como las esperaríamos.

Definición 2.5.4 (estrategia sin memoria en un 2G)**.**

Definición 2.5.5 (estrategia con memoria finita en un 2G)**.**

Agg ejemplo

Capítulo 3

Objetivos ω-regulares

Resulta practico estudiar a los juegos en relacion a algun tipo de objetivo en particular (¿por que? agg algo aca - capaz citas van) Como mencionamos antes nosotros nos enfocamos en el estudio de juegos estoc polito con objetivos de rabin, pero la pregunta es por que? ... kinda that mejor redactado

En este capítulo presentaremos la formalización de los objetivos en juegos, hablaremos un poco sobre las preguntas que nos planteamos cuando estudiamos juegos, veremos una clasificación típica de los objetivos ω-regulares (viendo ejemplos de cada uno) y desarrollaremos sobre la importancia de los objetivos de Rabin y nuestro interés en su estudio.

3.1. Objetivos en juegos y preguntas de investigación

Un objetivo ϕ para un juego es un conjunto de caminos. Un objetivo para el jugador $\Box$ especifica un conjunto de caminos que son ganadores para el jugador $\Box$, y un objetivo para el jugador $\Diamond$ especifica un conjunto de caminos ganadores para el jugador $\Diamond$. En el caso de juegos de suma cero (que son con los que trabajaremos), los objetivos de los dos jugadores son estrictamente competitivos, es decir, uno es el complemento del otro, por eso nos basta con asociarle un objetivo al juego, que será el asociado al jugador $\Box$ y se puede deducir que el objetivo para el jugador $\Diamond$ será su complemento.

Definir la noción de objetivos nos permitirá hacernos la pregunta que uno primero asocia con juegos: "¿quién gana". Para que esta pregunta pueda ser respondida en un contexto académico la deberíamos formalizar, y cuando lo hacemos, en realidad son muchas las preguntas que se pueden pensar sobre "¿quién gana?".

Podemos, por ejemplo, preguntarnos ¿desde qué estados podemos asgurarnos que gana un jugador? Es decir, fijado un estado s nos podemos preguntar si existe una estrategia para algún jugador tal que frente a cualquier estrategia del otro jugador puede asegurar que se produce un camino dentro del conjunto ϕ. Esta pregunta en la literatura de juegos estocásticos es conocida como la pregunta cualitativa (y es esta pregunta sobre la cual enfocaremos nuestro estudio). Se la llama así porque en los juegos estocásticos también nos podemos hacer una pregunta cuantitativa: para un estado s, ¿cuál es la probabilidad de que desde s gane algún jugador?. Esta pregunta puede ser respondida para todos los estados, aún cuando no existe una estrategia que asegure un comportamiento ganador frente a todas las estrtategias del otro jugador. La respuesta a esta pregunta es conocida como el valor del estado.

El conjunto de estados desde donde el jugador $\Box$ gana para algún objetivo ϕ será llamado la región ganadora del juego. El aporte más importante de este trabajo presentará cómo calcular la región ganadora de un juego estocástico politópico con un tipo particular de objetivo ω-regular: un objetivo de Rabin.

Una importante subclase de objetivos son los objetivos ω-regulares. Los objetivos ω-regulares nacen de la extensión de la teoría de lenguajes regulares al dominio de las palabras infinitas. Mientras que los lenguajes regulares clásicos describen conjuntos de cadenas finitas reconocibles por autómatas finitos, los lenguajes ω-regulares operan sobre secuencias infinitas, capturando comportamientos infinitos. Esta generalización fue formalizada en los años sesenta con los autómatas de Büchi, que introdujeron condiciones de aceptación sobre corridas infinitas de autómatas de estado finito.

Esta subclase de objetivos resulta de importancia en el contexto de la verificación y síntesis de sistemas reactivos, donde es necesario especificar y razonar sobre comportamientos que pueden durar indefinidamente. Estos objetivos permiten expresar propiedades como "algo bueno ocurre infinitamente a menudo.º "algo malo ocurre sólo una cantidad finita de veces".

3.2. Clasificación de objetivos ω-regulares

Existen varios tipos de objetivos ω-regulares, cada uno definido por diferentes condiciones de aceptación en los autómatas sobre palabras infinitas. En particular, las siguientes especificaciones de condiciones de aceptación definen objetivos ω-regulares y son las más estudiadas:

- **Objetivos de alcanzabilidad y seguridad.** Una especificación de alcanzabilidad para un juego G es un conjunto $T \subseteq S$ de estados. La especificación de alcanzabilidad requiere que algún estado en T sea visitado. Así, la especificación de alcanzabilidad T define el conjunto $\text{Reach}(T) = \{(s_0, s_1, s_2, \dots) \in \text{Paths}(G) | \exists k \geq 0, s_k \in T\}$ de caminos ganadores; este conjunto es llamado un objetivo de alcanzabilidad.

  Una especificación de seguridad para G es también un conjunto $U \subseteq S$ de estados, llamados estados seguros. La especificación de seguridad U requiere que solo estados en U sean visitados. Formalmente, el objetivo de seguridad definido por U es el conjunto $\text{Safe}(U) = \{(s_0, s_1, \dots) \in \text{Paths}(G) | \forall k \geq 0, s_k \in U\}$ de caminos ganadores. Nótese que alcanzabilidad y seguridad son objetivos duales: $\text{Safe}(U) = \text{Paths}(G) \setminus (\text{Reach}(S \setminus U))$.

- **Objetivos de Büchi y co-Büchi**. Una especificación de Büchi para G es un conjunto $B \subseteq S$ de estados, que son llamados estados de Büchi. La especificación de Büchi requiere que algún estado en B sea visitado infinitamente a menudo. Para un camino $\omega = (s_0, s_1, s_2, \dots)$, escribimos $\text{Inf}(\omega) = \{s \in S | s_k = s \text{ para infinitos } k \geq 0\}$ para el conjunto de estados que ocurren infinitamente a menudo en ω. Entonces, el objetivo de Büchi definido por B es el conjunto $\text{Büchi}(B) = \{\omega \in \text{Paths}(G) | \text{Inf}(\omega) \cap B \neq \emptyset\}$ de caminos ganadores.

  El dual de una especificación de Büchi es una especificación de co-Büchi $C \subseteq S$, que especifica un conjunto de estados llamados co-Büchi. Esta especificación requiere que los estados fuera de C sean visitados solo una cantidad finita de veces. Formalmente, el objetivo de co-Büchi definido por C es el conjunto $\text{co-Büchi}(C) = \{\omega \in \text{Paths}(G) | \text{Inf}(\omega) \subseteq C\}$ de caminos ganadores.

  Cabe destacar también que los objetivos de alcanzabilidad y seguridad pueden ser transformados a objetivos de Büchi y co-Büchi, respectivamente, modificando un poco el juego G. Por ejemplo, si el juego G' resulta de G al transformar cada estado

$s \in T$ en un estado absorvente, entonces un juego jugado en G con el objetivo de alcanzabilidad $\mathrm{Reach}(T)$ es equivalente a un juego jugado en G' con el objetivo de Büchi, $\mathrm{Büchi}(T)$.

- **Objetivos de Rabin y Streett**. Ahora pasamos a combinaciones booleanas de objetivos de Büchi y co-Büchi. Una especificación de Rabin para el juego G es un conjunto finito $R = \{(E_1, F_1) \dots (E_d, F_d)\}$ de pares de conjuntos de estados, esto es, $E_j \subseteq S$ y $F_j \subseteq S$ para todo $1 \leq j \leq d$. Los pares en R son llamados pares de Rabin. Asumimos, sin pérdida de generalidad, que $\bigcup_{1 \leq j \leq d}(E_j \cup F_j) = S$. La especificación de Rabin R requiere que para algun par de Rabin $1 \leq j \leq d$, todos los estados en el conjunto de la izquierda E_j sean visitados solo una cantidad finita de veces, y que algún estado en el conjunto de la derecha F_j sea visitado infinitamente a menudo. Con eso, el objetivo de Rabin definido por R es el conjunto $\mathrm{Rabin}(R) = \{\omega \in \mathrm{Paths}(G) | \exists j \in [1, d], \mathrm{Inf}(\omega) \cap E_j = \emptyset \wedge \mathrm{Inf}(\omega) \cap F_j \neq \emptyset\}$ de conjuntos ganadores. Nótese que el objetivo co-Büchi $\text{co-Büchi}(C)$ es igual al objetivo de Rabin con un único par $\mathrm{Rabin}(\{(C, S)\})$ y que el objetivo de Büchi $\mathrm{Büchi}(B)$ es igual al objetivo de Rabin con dos pares $\mathrm{Rabin}(\{(\emptyset, B), (S, S)\})$ (por la asumpción que hicimos sobre la unión de los conjuntos de los pares).

  Los complementos de los objetivos de Rabin son los objetivos de Streett. Una especificación de Streett para G es también un conjunto $W = \{(E_1, F_1), \dots, (E_d, F_d)\}$ de pares de conjunto de estados $E_j \subseteq S$ y $F_j \subseteq S$ tal que $\bigcup_{1 \leq j \leq d}(E_j \cup F_j) = S$. Los pares en W se llaman pares de Streett. La especificación de Streett W requiere que para cada par de Streett $1 \leq j \leq d$, si algún estado en el conjunto de la derecha F_j es visitado infinitamente a menudo, entonces algún estado en el conjunto de la izquierda E_j es visitado infinitamente a menudo. Formalmente, el objetivo de Streett definido por W es el conjunto $\mathrm{Streett}(W) = \{\omega \in \mathrm{Paths}(G) | \forall j \in [1, d], \mathrm{Inf}(\omega) \cap E_j \neq \emptyset \vee \mathrm{Inf}(\omega) \cap F_j = \emptyset\}$ de caminos ganadores. Nótese que $\mathrm{Streett}(W) = \mathrm{Paths}(G) \setminus \mathrm{Rabin}(W)$.

- **Objetivos de paridad**.Una especificación de paridad para G consiste en un entero no negativo d y una función $p \colon S \to \{0, 1, 2, \dots, 2d\}$, que asigna a cada estado de G un entero entre 0 y $2d$. Para un estado $s \in S$, el valor $p(s)$ se denomina *prioridad* de s. Sin pérdida de generalidad asumimos que $p^{-1}(j) \neq \emptyset$ para todo $0 < j \leq 2d$; esto implica que la especificación queda totalmente determinada por la función p (no es necesario indicar d explícitamente). El número $2d+1$ es el número de prioridades de p. La especificación exige que la prioridad mínima de todos los estados visitados infinitamente a menudo sea par. Formalmente, el objetivo de paridad definido por p es el conjunto $\text{Parity}(p) = \{\omega \in \text{Paths}(G) | \min\{p(s) \mid s \in \text{Inf}(\omega)\} \text{ es par}\}$ de caminos ganadores. Nótese que su complemento es también un objetivo de paridad, pues $\text{Paths}(G) \setminus \text{Parity}(p) = Parity(p+1)$, donde $(p+1)(s) = p(s)+1$ para todo $s \in S$ (si $p^{-1}(0) = \emptyset$ se usa $p-1$ en lugar de $p+1$). Esta autodualidad de los objetivos de paridad resulta muy conveniente al resolver juegos. Es también interesante notar que los objetivos de Büchi son objetivos de paridad con dos prioridades (siendo $p^{-1}(0) = B$ y $p^{-1}(1) = S \setminus B$) y los objetivos de co-Büchi son objetivos de paridad con tres prioridades (siendo $p^{-1}(0) = \emptyset$, $p^{-1}(1) = S \setminus C$ y $p^{-1}(2) = C$).

Los objetivos de paridad también se llaman objetivos Rabin-chain, pues son un caso especial de objetivos de Rabin: si los conjuntos de una especificación Rabin $R = \{(E_1, F_1), \ldots, (E_d, F_d)\}$ forman una cadena $E_1 \subsetneq F_1 \subsetneq E_2 \subsetneq F_2 \subsetneq \cdots \subsetneq E_d \subsetneq F_d$, entonces $\text{Rabin}(R) = \text{Parity}(p)$ para la función de prioridades $p : S \to \{0, 1, \ldots, 2d\}$ que asigna a cada estado en $E_j \setminus F_{j-1}$ la prioridad $2j-1$, y a cada estado en $F_j \setminus E_j$ la prioridad $2j$, donde $F_0 = \emptyset$. Recíprocamente, dada una función de prioridades $p : S \to \{0, 1, \ldots, 2d\}$, podemos construir una cadena $E_1 \subsetneq F_1 \subsetneq \cdots \subsetneq E_{d+1} \subsetneq F_{d+1}$ de $d+1$ pares de Rabin tal que $\text{Parity}(p) = \text{Rabin}(\{(E_1, F_1), \ldots, (E_{d+1}, F_{d+1})\})$ de la siguiente manera: sea $E_1 = \emptyset$ y $F_1 = p^{-1}(0)$, y para todo $1 \leq j \leq d+1$ sea $E_j = F_{j-1} \cup p^{-1}(2j-3)$ y $F_j = E_j \cup p^{-1}(2j-2)$. Por tanto, los objetivos de paridad son una subclase de los objetivos de Rabin que está cerrada bajo complementación; de ello se sigue que todo objetivo de paridad es a la vez un objetivo de Rabin y un objetivo de Streett.

- **Objetivos de Müller**. Una especificación de Müller para G es un conjunto $M \subseteq 2^S$ cuyos elementos se llaman *conjuntos de Müller*. La especificación exige que el conjunto de estados visitados infinitamente a lo largo de una trayectoria pertenezca a M; formalmente, el objetivo de Müller definido por M es $\text{Müller}(M) = \{\omega \in \text{Paths}(G) \mid \text{Inf}(\omega) \in M\}$. Es fácil notar que un objetivo de Rabin es un caso especial de un objetivo de Müller, pero es también cierto que un objetivo de Müller puede ser transformado en un objetivo de Rabin. Se puede encontrar una prueba de esta afirmación que sigue la técnica de *latest appearence record* en la sección 1.4.2 de [1].

  Capaz agg tal ejemplo en tal juego es un objetivo de X

3.3. La importancia de los objetivos de Rabin

Con lo presentado en la sección anterior, poedemos deducir algo muy interesante de los objetivos de Rabin. Mencionamos primero que los objetivos de alcanzabilidad y seguridad pueden transformarse en objetivos de Büchi y co-Büchi. Y luego mencionamos que tanto los objetivos de Büchi como de co-Büchi pueden trasnformarse en objetivos de Rabin (con lo cual también podemos hacer lo mismo con los objetivos de alcanzabilidad y seguridad). A su vez, también mencionamos que los objetivos de paridad y de Müller pueden ser transformados a objetivos de Rabin. Con esto, podemos empezar a deducir lo que se prueba formalmente con la prueba de Safra en [**Safra**]: cada objetivo ω-regular puede ser transformado a un objetivo de Rabin. (Safra lo que prueba en realidad es que cada automata no determinista de Büchi -de los cuales se sabe que pueden reconocer todos los lenguajes ω-regular- se puede trasnformar en un autómata determinista de Rabin).

Esta es la razón principal por la cual decidimos enfocarnos en objetivos de Rabin en esta tesina. Estudiar juegos con objetivos de Rabin es tan general como estudiar juegos con cualquier tipo de objetivo ω-regular; cada resultado que obtengamos vale para juegos con cualquier objetivo ω-regular.

Por otro lado, también resulta de particular interés el estudio de juegos con objetivos de Rabin y su dual, objetivos de Streett, porque su forma coincide con la de las condiciones de equidad (*fairness* en inlglés). Estas propiedades, que son un tipo un particular de propiedades de vitalidad (*liveness* en inglés), son de las más clásicas que se solicitan que tengan los sistemas reactivos, por lo que su estudio tiene implicancias prácticas concretas directas.

Entonces, con lo desarrollado se puede entender bien el porqué de la segunda parte del título de esta tesina: objetivos de Rabin, pero nos falta todavía adentrarnos en el porqué de la primera parte: juegos estocásticos politópicos. Eso es lo que haremos en el próximo capítulo.

Capítulo 4

Juegos Estocásticos Politópicos - PSGs

El propósito de este capítulo es presentar las definiciones y los hallazgos que nos son más relevantes del paper *Polytopal Stochastic Games* [9] haciéndoles ciertos cambios que fueron necesarios para los resultados que mostraremos más adelante. Este trabajo es el que sirvió de inspiración para esta tesina y toda la producción aquí es una extensión del mismo.

El paper, publicado en 2025, es el primero en presentar el concepto de *juego estocástico politópico*, respondiendo a la necesidad de juegos estocásticos que puedan capturar mayor incertidumbre sobre las distribuciones de probabilidad que determinan las acciones, haciendo esto a través de inecuaciones lineales cuyas soluciones forman un politopo.

Dividiremos el capítulo en dos secciones: definiciones y resultados.

4.1. Definiciones

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 $s \in S$.

PSGs

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* $\Theta : \mathcal{S} \to \mathcal{P}_f(\mathrm{DPoly}(\mathcal{S}))$.

Un juego estocástico politópico se diferencia de un juego estocástico tradicional por el hecho de que desde 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 $\mu \in K$, en vez de elegir directamente una acción de entre un conjunto finito. Al igual que en un juego estocástico, el siguiente estado s' se selecciona de acuerdo con la distribución μ correspondiente a la acción seleccionada, y el juego continuará desde s' repitiendo el mismo proceso.

Viendo esto, es natural pensar que desarrollo de un juego estocástico politópico se puede interpretar 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 $p \in [0,1]$).

Como lo hicimos para MDPs y juegos estocásticos a la definición de camino la daremos como una secuencia de estados y acciones. Esto difiere de la formulación original hecha para PSGs en [9], pero cambiaremos acordemente las definiciones que se asocian a ello para mantener la formalidad de los resultados.

Definición 4.1.3 (Camino en un PSG)**.** *Un camino en un PSG es una secuencia infinita que alterna entre estados y conjuntos resultado. Formalmente un camino 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 $i \geq 0$.*

Al igual que para juegos estocásticos, notaremos al conjunto de todos los caminos de un PSG como Paths*. Dado un estado s, indicaremos con* Paths_s *el conjunto de los* *caminos que se originan en s, y con* Paths_s^n *al conjunto de los caminos 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,

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

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 $\pi_i(\omega s)(\mathcal{A}(s)) = 1$.*

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 $\mathbb{P}^{\pi_\Box,\pi_\Diamond,n}_{\mathcal{G}_\mathcal{K},s} : \text{Paths}^{n+1} \to [0,1]$ para todo $s' \in \mathcal{S}$ y $\hat{\omega} \in \text{Paths}^{n+1}_s$ inductivamente como sigue:

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

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

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

(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 $\mathbb{P}^{\pi_\Box,\pi_\Diamond,n}_{\mathcal{G}_\mathcal{K},s} : \mathcal{P}(\text{Paths}^{n+1}) \to [0,1]$ a conjuntos como la suma de la medida sobre los elementos del conjunto.
Sea Σ_{Paths} la σ-álgebra discreta sobre Paths (pues tanto $\mathcal{S}$ como los conjuntos soportes serán conjuntos finitos) y $\Sigma_{\text{Paths}^\omega}$ la σ-álgebra producto usual sobre Paths^ω. Por el teorema de extensión de Carathéodory, $\mathbb{P}^{\pi_\Box,\pi_\Diamond}_{\mathcal{G}_\mathcal{K},s} : \Sigma_{\text{Paths}^\omega} \to [0,1]$ se define como la única medida de probabilidad tal que para todo $n \geq 0$, y $SV_i \in \Sigma_{\text{Paths}}$, $0 \leq i \leq n$,

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

Así se logra interpretar un PSG como un SG pero con un conjunto infinito de acciones. En [9] se demuestra cómo ciertos preguntas que nos podemos hacer sobre esta interpretación infinita se pueden responder analizando en la interpretación con una cantidad finita de acciones 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 [9].

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 [7, 12]. Por lo tanto, este teorema también proporciona inmediatamente una solución algorítmica para los PSGs.

El segundo resultado, se deriva del primero y presenta la complejidad del estudio cuantitativo de los juegos estocásticos politópicos. Enunciado sobre lo que nos interesa y notando como $Val_{\mathcal{G}_{\mathcal{K}},s}(\Diamond G)$ al valor del estado s (es decir, la probabilidad máxima de que $\Box$ gane desde s) en el juego $\mathcal{G}_{\mathcal{K}}$ con un objetivo de alcanzabilidad sobre un conjunto G, el teorema sería:

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* $NP \cap coNP$.

Con estos resultados como contexto surge el interés de estudiar los juegos estocásticos politópicos con otros objetivos. Por lo expuesto en el capítulo anterior, decidimos que sería una buena idea estudiar los objetivos de Rabin. En el próximo capítulo nos explayamos sobre los hallazgos que pudimos hacer durante el desarrollo de esta tesina.

Capítulo 5

Objetivos de Rabin en PSGs

Este capítulo refleja el aporte original principal de la tesina. En primer lugar, se presentará el concepto de Proceso de Decisión de Markov Politópico (una suerte de juego estocástico politópico con un solo jugador o una extensión de los MDPs) acompañados de unos teoremas que nos serán útiles. En segundo lugar, introduciremos los juegos deterministas de adversario justo para así, finalmente, llegar a nuestro resultado principal que establece que la región ganadora para el jugador maximizador en un juego estocástico politópico con objetivo de Rabin es igual a la región ganadora para el jugador maximizador en un juego determinista de adversario justo particular que construiremos y nombraremos como la desrandomización del juego estocástico politópico. Este resultado, a su vez, nos proveerá con una manera de sintetizar una estrategia ganadora, lo que también nos permitirá estudiar la complejidad de responder a la pregunta cualitativa en nuestro contexto.

5.1. Procesos de Decisión de Markov Politópicos - PMDPs

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* $\mathcal{S}_\Diamond = \emptyset$*.*

Los caminos, las estrategias, y la medida de probabilidad en los PMDPs se definen tomando las mismas definiciones de los juegos estocásticos politópicos.

A continuación presentaremos las definiciones de conjuntos estado-resultado, sub-PMDPs y componentes finales que en cierta forma extienden a los conceptos de conjuntos estado-acción, sub-MDPs y componentes finales que se suelen encontrar en la literatura de procesos de decisión de Markov [10, 11].

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* $\mathrm{sub}(\chi) = (C, D)$ *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* $\mathrm{er}(C, D) = \{(s, V') \mid s \in C \wedge V' \in D(s)\}$ *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 $\mathrm{EC}(\mathcal{M})$ *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 para siempre 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 camino dado.

Definición 5.1.5 ($\inf_{\mathrm{er}}$)**.** *Dado un camino* $\omega = (s_0, \alpha_0, s_1, \alpha_1, \dots)$ *indicamos por*

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

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:*

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

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 camino entra a C, nunca sale de C ni se elige una acción que no esté en D. Es más, una vez en C un camino visitará todos los estados de C infinitamente a menudo con probabilidad 1. □

Este teorema nos permite presentar un corolario que nos será útil al momento de probar el teorema principal de esta tesina.

Corolario 5.1.1.1. *Sea $\mathcal{M}$ un PMDP y sea s un estado ganador en él para una condición de Rabin $R = \{(E_1, F_1), \dots, (E_d, F_d)\}$.Entonces, para toda componente final (C, D) alcanzable desde s, vale que existe algún $j \in [1, d]$ tal que $C \cap E_j = \emptyset$ y $C \cap F_j \neq \emptyset$.*

Demostración. Supongamos que existe una componente final (C', D') alcanzable desde s tal que esta no cumple que existe algún $j \in [1, d]$ tal que $C' \cap E_j = \emptyset$ y $C' \cap F_j \neq \emptyset$.

Por teorema 5.1, sabemos que existe una estrategia válida π' tal que permite llegar a C' desde s y quedarse allí para siempre. Podemos plantear un camino que sigue esa estrategia π', llamemóslo ω'. Ahora bien, entonces también por como presentamos que será esa estrategia π' en el teorema 5.1, sabemos que $\text{Inf}(\omega') = C$. Pero, a su vez sabemos que ese camino debe cumplir con la condición de Rabin, es decir, sabemos que existe algún $j \in [1, d]$ tal que $\text{Inf}(\omega') \cap E_j = \emptyset$ y $\text{Inf}(\omega') \cap F_j \neq \emptyset$, lo que quiere decir que existe algún $j \in [1, d]$ tal que $C' \cap E_j = \emptyset$ y $C' \cap F_j \neq \emptyset$ y tendríamos una contradicción.

Esta contradicción viene de suponer que existe tal comopnente final. Con lo que probamos lo que queríamos. $\square$

El próximo resultado establece que, para cualquier estado inicial y cualquier estrategia (de memoria finita), un camino 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,*

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

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

Si podemos mostrar que

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

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

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

. 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 $(t, V') \in \text{er}(C, D)$ tal que $V' \nsubseteq C$.

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

- 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 camino en $\text{Paths}_s^{(C,D)}$ 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 camino en $\text{Paths}_s^{(C,D)}$ 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 camino en $\text{Paths}_s^{(C,D)}$ 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 $\mathbb{P}_{\mathcal{M},s}^{\pi}(\{\omega \in \text{Paths}(s) \mid \omega \in \text{Paths}_s^{(C,D)}\}) = 0$

$\square$

Esto nos permite definir el siguiente corolario útil y clásico en los procesos de decisión de Markov 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 de memoria finita en él. Una condición de Rabin $R = \{(E_1, F_1), \ldots, (E_d, F_d)\}$ 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 $j \in \{1, \ldots, d\}$ tal que $U \cap E_j = \emptyset$ y $U \cap F_j \neq \emptyset$.*

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

Claro, en un principio la idea fue poder probar ambos teoremas en su caso general, sin tener que limitarnos en la clase de las estrategias utilizadas. La idea para probar esto esencialmente fue intentar seguir como modelo las pruebas realizadas para MDPs en [10, 11]. Esto en el caso del teorema 5.1.1 vino sin complicaciones, pero en el caso del teorema 5.1.2 devino en problemas, puesto que la finitud de acciones salientes de un estado probó ser una hipótesis fundamental.

La idea de prueba consiste en proponer un sub-PMDP (C, D) arbitrario que no cumple la definición 5.1.4 de componente final, dividir el análisis por casos de cómo no se cumple la definición y mostrar que, en cada caso, la probabilidad de que un camino genere un sub-PMDP como el propuesto es 0.

Remitiéndonos a nuestra prueba, vemos que en el primer ítem el argumento está explicítamente 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 es finito, por lo que 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 si tengo una cantidad infinita de r_i's distintos, aún cuando cada uno de ellos es menor a 1, la productoria bien podría no converger a 0, como se puede ver en el siguiente caso.

Consideremos un PMDP en donde tenemos dos estados t y 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 camino 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 caminos de la siguiente manera:

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

siendo ω un camino, $\emptyset$ el camino vacío y $s \neq t$.

Si 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 $\frac{1}{2}$.

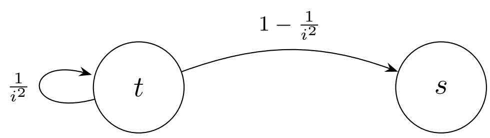

Figura 5.1: Interpretación del PDMDP con la estrategia π fijada.

Ahora bien, es cierto que esta premisa se basa en el hecho de que s no esté en C, que no es algo que podamos afirmar a priori, y que esto que mostramos solo implica que este método de prueba no es útil para probar el teorema 5.1.2 para cualquier tipo de estrategias; no significa que no sea válido.

Esta limitación a la que nos tuvimos que atener aquí, sin embargo, no constituyó una limitación a la prueba del teorema principal sobre regiones ganadoras que mostramos más adelante, porque pudimos reducir lo que necesitabamos para su prueba al corolario 5.1.1.1.

5.2. Juegos justos y desrandomización de un PSG

A fin de responder la pregunta cualitativa en el estudio de los juegos estocásticos politópicos con objetivos de Rabin, presentaremos un tipo nuevo de juegos deterministas: los juegos justos. Probaremos que el conjunto de vértices ganadores con probabilidad 1 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 el cálculo de los estados ganadores y la síntesis de estrategias ganadoras para un PSG con un objetivo de Rabin.

Juegos de adversario justo

Sea G un juego de grafo de dos jugadores y sea $E^l \subseteq (V_\Diamond \times V) \cap E$ un conjunto dado de aristas que deben ser tomadas infinitamente a menudo si el estado del que parten es visitado infinitamente a menudo. Nombraremos $V^l := \mathrm{dom}(E^l)$ al conjunto de vértices del jugador $\Diamond$ que está en el dominio de E^l. Las aristas en E^l representarán suposiciones de equidad 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 infinitas veces, se espera también que toda arista en E^l saliente de v sea tomada una cantidad infinita de veces.

Denotamos por $G^l = \langle G, E^l \rangle$ a un juego de adversario justo, y extendemos nociones como jugadas, estrategias, condiciones ganadoras, regiones ganadoras, etc. de juegos deterministas de manera natural. Una jugada ρ sobre G^l se dice fuertemente justa si satisface la siguiente fórmula en lógica temporal lineal (LTL):

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

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 $\alpha \to \varphi$ desde v_0.

Hay dos observaciones interesantes para hacer sobre los juegos de adversario justo:

Primero, las aristas en E^l 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 en E^l (representada con línea discontinua). Supongamos que la especificación para el jugador $\Box$ es $\varphi = \Box\Diamond q$. Si la arista (p, q) no estuviera en E^l, 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 equidad sobre la arista (p, q) fuerza al jugador $\Diamond$ a elegir infinitamente a menudo la transición hacia q.

Segundo, las suposiciones de equidad modeladas por aristas en E^l 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 con dos aristas en E^l 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.

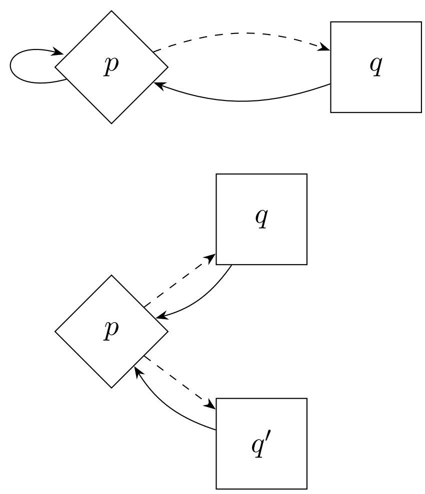

Figura 5.2: Dos juegos de adversario justo.

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:

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

Esta misma condición de equidad se puede expresar como

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

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

Relación entre un PSG y su desrandomización

Nos será útil pensar en transformaciones entre los distintos juegos para la prueba 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:

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

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

Transformación de caminos

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

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

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 caminos en $\mathcal{G}_\mathcal{K}$ que se corresponderían con él. Llamaremos a este conjunto de caminos $\mathrm{rand}(\rho)$. Es decir,

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

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

Randomización 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 camino ω en $\mathcal{G}_\mathcal{K}$. La idea, entonces, será definir π_i de igual manera para todos los elementos del conjunto $\mathrm{rand}(\rho)$.

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 $\omega \in \mathrm{derand}(\rho)$ definir $\pi_i(\omega)(\hat{\alpha}) = 1$ para algún $\hat{\alpha}$ particular tal que $\mathrm{supp}(\hat{\alpha}) = \widehat{V}$.

Llamaremos a esta nueva estrategia π_i obtenida a partir de σ_i, $\mathrm{rand}(\sigma_i)$.

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

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

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

Desrandomización de estrategias

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 camino que podrían cumplir con la condición de ω, la definición de derand(π_i) dependerá de la elección de un prefijo de camino 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 derand(π_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).

¿Cómo se podrían tomar esas elecciones?

Suponiendo que tenemos una estrategia π_i en $\mathcal{G}_\mathcal{K}$, es probable que a su vez tengamos caminos $\omega'_1, \omega'_2, \ldots$ en $\mathcal{G}_\mathcal{K}$ específicos en que estemos interesados que σ_i replique. En ese caso, una manera de tomar las decisiones antes planteadas es usando los prefijos finitos de los distintos ω'_k para las decisiones de σ_i. Esto sería: si tenemos un prefijo de camino $\hat{\omega} = (s_1, \alpha_1, \ldots, s_m, \alpha_m)$ (que respeta π_i), entonces haremos que para $\hat{\rho} = (s_1, v_{\text{supp}(\alpha_1)}, \ldots, s_m)$, $\sigma_i(\hat{\rho}) = v_{\text{supp}(\alpha_m)}$.

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 $\sigma_\Diamond(s_0, v_{V_0}, \dots, s_k, v_{V_k}) = s_{k+1}$ 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 $s_0^{V_k}, s_1^{V_k}, \dots$. 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 camino $\hat{\omega}$ 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).

Respetar una estrategia

Al hablar de caminos específicos es natural pensar que estos fueron producto de una partida en la que los jugadores siguieron determinadas estrategias. Para la prueba que plantearemos a continuación querremos formalizar esta relación entre caminos y las estrategias que se siguieron en ellos y lo haremos a través de la idea de que un camino "respeta"determinadas estrategias.

Sea $\omega = (s_0, \alpha_0, s_1, \alpha_1, \dots)$ un camino en un juego estocástico poliópico y sean $\pi_\Box$ y $\pi_\Diamond$ dos estrategias en el mismo juego. Decimos que ω respeta las estrategias $\pi_\Box$ y $\pi_\Diamond$ si

$\forall i \geq 0$ vale

$$\big(s_i \in \mathcal{S}_\Box \wedge \pi_\Box(s_0, \alpha_0, \ldots, s_i)(\alpha_i) > 0 \wedge s_{i+1} \in \mathrm{supp}(\alpha_i)\big) \bigvee$$
$$\big(s_i \in \mathcal{S}_\Diamond \wedge \pi_\Diamond(s_0, \alpha_0, \ldots, s_i)(\alpha_i) > 0 \wedge s_{i+1} \in \mathrm{supp}(\alpha_i)\big)$$

Sea $\rho = (s_0, v_{V_0}, s_1, v_{V_1}, \ldots)$ un camino en la desrandomización de un juego estocástico politópico y sean $\sigma_\Box$ y $\sigma_\Diamond$ dos estrategias en el mismo juego. Decimos que ρ respeta las estrategias $\sigma_\Box$ y $\sigma_\Diamond$ si $\forall i \geq 0$ vale

$$\sigma_\Diamond(s_0, v_{V_0}, \ldots, s_i, v_{V_i}) = s_{i+1} \bigwedge$$
$$\big(\big(s_i \in \mathcal{S}_\Box \wedge \sigma_\Box(s_0, v_{V_0}, \ldots, s_i) = v_{V_i}\big) \bigvee$$
$$\big(s_i \in \mathcal{S}_\Diamond \wedge \sigma_\Diamond(s_0, v_{V_0}, \ldots, s_i) = v_{V_i}\big)\big)$$

5.3. Prueba de igualdad sobre los conjuntos ganadores

Teorema 5.3.1. *Sea $\mathcal{G}_\mathcal{K} = (\mathcal{S}, (\mathcal{S}_\Box, \mathcal{S}_\Diamond), \mathcal{A}, \theta)$ la interpretación de un juego estocástico politópico, $R = \{(E_1, F_1), \ldots, (E_d, F_d)\}$ una condición de Rabin sobre $\mathcal{S}$, con su especificación LTL φ,*

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

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 con probabilidad 1 en $\mathcal{G}_\mathcal{K}$. 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:

$$\inf_{\pi_\Diamond \in \Pi_\Diamond} \mathbb{P}^{\hat{\pi}^*_{\Box}, \pi_\Diamond}_{\mathcal{G}_{\mathcal{K}}, s}(\varphi) = 1 \tag{5.3}$$

Proponemos $\hat{\pi}^*_{\Box} = \mathrm{rand}(\sigma^*_{\Box})$ y veremos que vale 5.3 por reducción al absurdo.

Supongamos que no vale 5.3, es decir,

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

Esto significa que existe una estrategia $\pi^*_\Diamond$ que hace que $\mathbb{P}^{\hat{\pi}^*_{\Box}, \pi_\Diamond}_{\mathcal{G}_{\mathcal{K}}, s}(\varphi) < 1$. A su vez, esto significa que existe un camino ω^* que empieza desde s, respeta las estrategias $\hat{\pi}^*_{\Box}$ y $\pi^*_\Diamond$, pero no cumple φ.

Pensemos, entonces, en $\sigma^*_\Diamond$ una desrandomización válida de $\pi^*_\Diamond$ que sigue las elecciones de acciones que toman los prefijos de ω^*.

Luego, $\rho^* = \mathrm{derand}(\omega^*)$ respeta las estrategias $\sigma^*_{\Box}$ y $\sigma^*_\Diamond$. Como $\rho^* \cap \mathcal{S} = \omega^* \cap \mathcal{S}$ y ω^* no cumple φ, ρ^* no cumple φ. Es decir, existe un camino que empieza desde s, respeta $\sigma^*_{\Box}$, pero no es ganador para φ. Esto contradice que la estrategia $\sigma^*_{\Box}$ sea ganadora desde s.

Llegamos a esta contradicción por suponer que no vale 5.3. Entonces la ecuación sí vale y, consecuentemente, tenemos que $s \in \mathcal{W}^{as}$, 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 $\sigma^*_\square = \text{derand}(\pi^*_\square)$

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

Una manera de ver esto es probar que para un camino cualquiera ρ generado por $\sigma^*_\square$ y una estrategia válida $\sigma_\diamondsuit$ arbitraria en $Derand(\mathcal{G}_\mathcal{K})$, $\text{Inf}(\rho)_C = \text{Inf}(\rho) \cap \mathcal{S}$ cumple con la especificación φ. Es decir, que vale la siguiente fórmula:

$$\varphi' := \bigvee_{j=1}^{k} \big((\text{Inf}(\rho)_C \cap R_j = \emptyset) \wedge (\text{Inf}(\rho)_C \cap G_j \neq \emptyset)\big) \tag{5.5}$$

Como $\pi^*_\square$ es una estrategia ganadora frente a cualquier estrategia $\pi_\diamondsuit$, siempre se cumple la condición de Rabin con probabilidad 1 en el PMDP $\mathcal{M}'$ que se forma al fijar la estrategia $\pi^*_\square$ en $\mathcal{G}_\mathcal{K}$. Por teorema 5.1.1.1, entonces vale que para cada componente final (C, D) alcanzable desde s en el PMDP $\mathcal{M}^*$ vale que existe algún $j \in [1, d]$ tal que $C \cap E_j = \emptyset$ y $C \cap F_j \neq \emptyset$.

Si denotamos como $\text{Inf}(\rho)_D$ a la función que le asocia a cada s_i en $\text{Inf}(\rho)_C$ el conjunto formado solo por el V_i correspondiente al vértice v_{V_i} en ρ, esto que presentamos anteriormente quiere decir que si podemos probar que $(\text{Inf}(\rho)_C, \text{Inf}(\rho)_D)$ es una componente final alcanzable desde s en el PMDP $\mathcal{M}^*$, entonces vale la ecuación 5.5.

Veamos entonces primero que $(\text{Inf}(\rho)_C, \text{Inf}(\rho)_D)$ cumple las dos condiciones para ser componente final en $\mathcal{M}^*$:

- Para cada V^i que aparece como subíndice de vértices *especiales* en $\text{Inf}(\rho)$, 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 $\pi_\diamondsuit$ 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 $\text{Inf}(\rho)$ es fuertemente conexo. Si fuese de otra manera habría dos vértices $u, v \in \text{Inf}(\rho)$ tales que v no sería alcanzable desde u, contradiciendo así que u y v son visitados infinitas veces por σ.

Luego, podemos ver que $\mathrm{Inf}(\rho)_C$ 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 $\mathrm{Inf}(\rho)_C$, puesto que tanto $\pi^*_\Diamond$ como $\pi^*_\Box$ les dan probabilidad positiva a los mismos vértices en $\mathcal{S}$ que $\sigma_\Diamond$ asegura visitar infinitamente.

Con lo cual hemos probado que vale l ecuación 5.5 y, por lo tanto, que $\sigma_\Box*$ es ganadora en $Derand(\mathcal{G}_{\mathcal{K}})$ desde s, con lo que hemos probado que $s \in \mathcal{W}$.

$\square$

Complejidad del cálculo de estados ganadores en un PSG con objetivo de Rabin

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 [13] 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 $O(n^2 d!)$ donde n es la cantidad de vértices en G y d 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 $O((nl)^2 d!)$ donde n es la cantidad de estados en $\mathcal{K}$, d 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 $\mathcal{K}$.

Sintésis de estrategias ganadoras en un PSG con objetivo de Rabin

El cálculo de desde qué estados gana el jugador $\Box$ responde parcialmente a la pregunta de "¿quién gana?", pero resulta no de menor importancia el plantearse "¿cómo se gana?". Esta pregunta abarca tanto con qué clase de estrategias se puede ganar como también podría enfocarse en describir paso a paso cómo se deberían ver esas estrategias y se asocia a lo que es conocido como el problema de *sintésis de estrategias*, remarcado como de interés en la literatura. En cierto punto, con lo demostrado, también respondemos parcialmente a esta pregunta por lo siguiente.

El algoritmo simbólico planteado en la subsección anterior y propuesto en [13] permite extraer una estrategia sin memoria ganadora para el jugador $\Box$ en el juego justo, por lo que tendríamos así una estrategia sin memoria ganadora en la desrandomización de un juego estocástico politópico con objetivo de Rabin. Como nuestra prueba nos indica cómo, a partir de una estrategia en la desrandomización, obtener una estrategia en el juego original, con esto estamos respondiendo, en cierta manera, a la pregunta de "¿cómo se gana?".

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

Capítulo 6

Conclusiones

6.1. Trabajo Futuro

Referencias

[1] E. Graedel, W. Thomas y T. Wilke. *Automata, Logics, and Infinite Games: A Guide to Current Research*. Springer Berlin Heidelberg, ene. de 2002. ISBN: 978-3-540-00388-5. DOI: 10.1007/3-540-36387-4.

[2] L. S. Shapley. «Stochastic Games». En: *Proceedings of the National Academy of Sciences* 39.10 (1953), págs. 1095-1100. DOI: 10.1073/pnas.39.10.1095. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.39.10.1095. URL: https://www.pnas.org/doi/abs/10.1073/pnas.39.10.1095.

[3] K. Chatterjee. «Stochastic ω-Regular Games». Tesis doct. University of California at Berkeley, 2007.

[4] A. Kučera. «Turn-Based Stochastic Games». En: *Lectures in Game Theory for Computer Scientists*. Ed. por K. R. Apt y E. Grädel. Cambridge University Press, 2011, págs. 146-184.

[5] K. Chatterjee y T. A. Henzinger. «A survey of stochastic ω-regular games». En: *Journal of Computer and System Sciences* 78.2 (2012). Games in Verification, págs. 394-413. ISSN: 0022-0000. DOI: https://doi.org/10.1016/j.jcss.2011.05.002. URL: https://www.sciencedirect.com/science/article/pii/S0022000011000511.

[6] D. Gillette. «Stochastic Games with Zero Stop Probabilities». En: *Contributions to the Theory of Games, Volume III*. Princeton: Princeton University Press, 1958, págs. 179-188. ISBN: 9781400882151. DOI: doi:10.1515/9781400882151-011. URL: https://doi.org/10.1515/9781400882151-011.

[7] 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.

[8] K. Chatterjee, L. de Alfaro y T. A. Henzinger. «The Complexity of Stochastic Rabin and Streett Games». En: *Automata, Languages and Programming*. Ed. por L. Caires, G. F. Italiano, L. Monteiro, C. Palamidessi y M. Yung. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, págs. 878-890. ISBN: 978-3-540-31691-6.

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

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

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

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

[13] 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.