Simulación Estocástica: Teoría y Laboratorio
Equipo Docente: Joaquín Fontbona; Camilo Carvajal, Arie Wortsman, Pablo Zúñiga

Integrantes: Javier Maass
Juan Pablo Sepúlveda

Resumen Proyecto Final
Reinforcement Learning en el juego de *Cacho/Dudo*

El Reinforcement Learning (RL) es una de las técnicas que ha revolucionado el área de *ML* en el último tiempo. Ha sido usado por las grandes compañías de tecnología, en proyectos como *Deep Mind* (con *AlphaZero*), automóviles autónomos o para procesamiento de lenguaje natural. Más en general, ha revolucionado la forma en que enseñamos a las máquinas a tomar *decisiones en tiempo real.*

El setting básico de RL es el de un **Markov Decision Process** (MDP), en que un agente toma (en cada instante de tiempo t discreto) **acciones** (A_t) que, según el **medio ambiente**, llevan al agente a un nuevo **estado** (S_t) en el cual se recibe cierta **recompensa** (R_t). Esto nos lleva a un proceso estocástico de la forma: $(S_0, A_0, R_1, S_1, A_1, \ldots, R_T, S_T)$ (donde T es un tiempo *terminal* en que termina un *episodio*). El agente elige su siguiente acción (conociendo el estado en que se encuentra), siguiendo una *política* (distribución de probabilidad condicional $\pi(a|s)$), y obtiene un *retorno* (en el tiempo t) dado por: $G_t = \sum_{k=t+1}^{T} \gamma^{k-(t+1)} R_k$.

El objetivo del RL es lograr que el agente **aprenda la política** π^* que maximice su retorno esperado ($\pi^* \in \arg\max_\pi \mathbb{E}_\pi[G_t]$). Hay resultados teóricos que garantizan que esto puede hacerse en el caso de conocer la *matriz de transición* asociada al ambiente (mediante la llamada *Generalized Policy Iteration*). Cuando esto no es así, se pueden usar métodos de Monte Carlo (en técnicas conocidas como *SARSA* o *Q-Learning*) para obtener un aproximado de la política óptima.

Dado que ya se ha usado el RL en el área de los juegos (como el blackjack, o el Go) pretendemos implementar una IA capaz de aprender la estrategia óptima para jugar Cacho/Dudo (ver este link), un juego típico chileno de complejidad moderada. A nuestro conocimiento no hay grandes resultados teóricos para este juego, por lo que este approach con RL nos podría permitir obtener hacer una primera aproximación de su *política óptima.*

Dado que este juego tiene componentes aleatorias, buscaremos entrenar a nuestro agente usando métodos de *Monte Carlo / SARSA / Q-Learning.* En primera instancia, se buscaría entrenar un agente capaz de jugar contra un **único adversario**, de **política determinista** (para simplificar la implementación); para luego extender la implementación a **varios jugadores**, y un agente capaz de jugar **contra otros agentes también entrenados**, lo cual sería un contexto un poco más realista.

Referencias Bibliográficas:

- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction.* Bradford Books.
- De Granville, C. (2005). *Applying Reinforcement Learning to Blackjack Using Q-Learning*
- Rich, D. (2022). Reinforcement Learning Fundamentals. Youtube: Link. GitHub: Link.
- Geiser, J. & Hasseler T. (2020). *Beating Blackjack - A Reinforcement Learning Approach.* Standford University.