c.u _____

# Examen Parcial 2: Bandits y Cadenas de Markov (Individual)

Este examen es en equipo y consta de dos partes: una escrita y otra oral. Este es el examen individual, solo se permite "papel y lapiz/pluma"

**Instrucciones**:
Escribe tu clave en la parte superior izquierda de esta hoja y en cada una de las páginas siguientes.

## 1. Markov-Bandit

### 1.1 Problema de Decisión (34 puntos)

Formula el problema de decisión más completo posible para el caso del *bandit markoviano* cuando se tiene la opción de consultar la probabilidad de ganar en una máquina específica.

Debes definir el problema de manera **formal y matemática**, incluyendo también la **intuición** sobre su significado y la razón de ser de esta formulación. Debe venir la definicion matematica formal, y que representa en le problema actual. Debe tener absolutamente todos los componentes de un problema de decision.

**Nota**:
Asegúrate de mencionar **todas las partes que componen un problema de decisión**, tanto desde la perspectiva intuitiva como formal.
La formulación matemática es prioritaria: debe ser precisa, explícita y concisa, tal como se planteó en el problema original.

**1.1.1 Hint** Componentes de un problema de decision. 1. **Estados (S)**: Conjunto de posibles estados verdaderos del entorno.

2. **Acciones (A)**: Conjunto de acciones que el agente puede tomar.
3. **Función de Transición (P)**: Probabilidad de transición de estado: $P(s'|s, a)$.
4. **Función de Recompensa (R)**: Recompensa inmediata: $R(s, a, s')$.
5. **Factor de Descuento ($\gamma$)**: Valor entre 0 y 1 que reduce la importancia de recompensas futuras.
6. **Política ($\pi$)**: Estrategia de decisión: $\pi(a|s)$ o $\pi(a|b)$ para políticas basadas en creencias.
7. **Distribución Inicial de Estados ($\rho$)**: Distribución de probabilidad sobre los estados iniciales: $\rho(s)$.
8. **Observaciones / Proxys (O)**: Señales o lecturas observables que el agente puede percibir.
9. **Función de Observación (e)**: Probabilidad de observar $o$ dado el estado resultante y la acción: $e(o|s', a)$
10. **Modelo de Inferencia / Actualización de Creencia ($P(s|o)$)**: Probabilidad posterior sobre los estados dado una observación (usado cuando los estados no son directamente observables).