

# Escaleras mágicas:

Un arquitecto se dispone a diseñar su casa de verano y toma una idea de su proyecto anterior, la escuela de magia y hechicería Hogwarts. En este proyecto las escaleras tienen un comportamiento no determinista, llevando a sus usuarios a distintos lugares siguiendo una cierta función de probabilidad. Dado que el proyecto actual es de menor tamaño que el anterior, se anima a hacer los desarrollos a mano, para ello contrata un ayudante, usted, a fin de asistirlo con las cuentas. Le entrega una nota que dice lo siguiente:

Diseño secreto de las escaleras mágicas de la casa de verano:

- Considere los pisos numerados de **1 a 4**.
- Subir del piso **s** a **s+1** toma **1 minuto** y se hace determinísticamente. Salvo para el primer piso que se demora **3 minutos** (por razones mágicas).
- Esperar en el descanso del piso **1** toma **2 minutos** y puede dejarlo en el piso **1** o llevarlo al **4** con la misma probabilidad.
- Esperar en el descanso del piso **2** puede llevarlo al **3** o al **4** con la misma probabilidad. En cualquier caso demora **2 minutos**.
- Esperar en el descanso del piso **3** toma puede llevarlo al **1** o al **4** con la misma probabilidad, si va al **1**, pierde al hacerlo **2.5 minutos**, si va al **4**, **1.5**.
- La única manera de bajar es un tobogán mágico ubicado en el cuarto piso. Este tobogán no forma parte del modelo matemático a desarrollar.

- 1) Modele el problema con un grafo con transiciones de probabilidad.
- 2) Cálculos analíticos:
  - a) Calcule la utilidad esperada de la política “siempre esperar”.
  - b) Calcule la utilidad esperada de la política “siempre subir”.
  - c) Calcule la utilidad esperada de la política “siempre esperar” si la probabilidad de llegar mágicamente al piso 4 por esperar en los descansos es de **0.3**.
- 3) Asumiendo como política inicial “siempre subir” y el MDP descrito originalmente.
  - a) Evalúe iterativamente la policy. ¿Coincide este resultado con el punto 2) b)?
  - b) Calcule un paso de **policy improvement** sobre dicha policy. ¿Es la mejor policy?
  - c) Calcule un paso más de **policy improvement** sobre la nueva policy (**policy iteration**). Comente los resultados obtenidos
- 4) Dadas las mismas condiciones que en ‘3)’ calcule **value iteration** sobre el MDP planteado, ¿qué comentarios tiene sobre este resultado y el de ‘3)c)’?

Nota: Se considera (gamma)  $\gamma = 1$  para todos los ejercicios