Reinforcement Learning

Taller #1

En este taller se han dispuesto 3 multi armed bandits. Estos se diseñaron con base en el test-bed propuesto en [1] (Figura 1) y contemplan:

- Un escenario no estático.
- Un escenario con alta varianza.
- Un escenario con baja varianza.

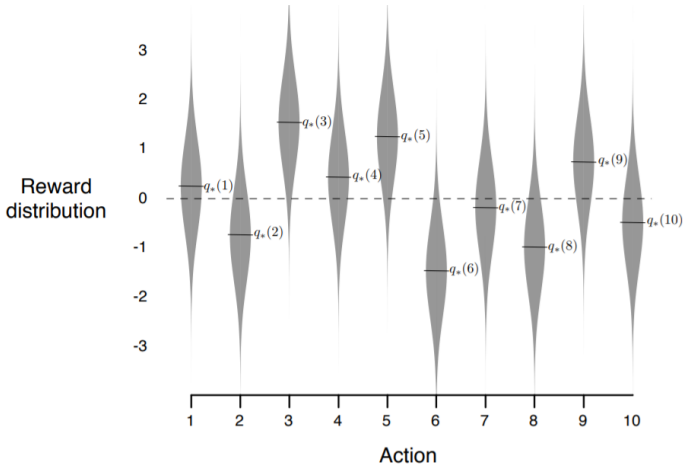

Figura 1: *Test-Bed* para el problema de *Multi-Armed Bandits* [1]

Para interactuar con estos 3 escenarios, adjunto a este documento, encontrará el archivo de Python compilado `envs.pyc`. De este archivo deberá importar la clase `BanditEnvs` y utilizar las funciones que esta provee para desarrollar el taller. Los escenarios no necesariamente están dispuestos en el orden mencionado.

El *Jupyter Notebook* (`HW1_MAB_Example.ipynb`) muestra cómo interactuar con la clase mencionada y los tres escenarios propuestos. Este *notebook* es solo de ejemplo y no es necesario que lo use como base para desarrollar su solución. Para poder ejecutarlo puede cargar el archivo a **Google Colaboratory**, el servicio *cloud* que ofrece *Google* para editar y utilizar *Jupyter Notebooks*.

Usted deberá resolver el problema de *Multi-Armed Bandits*, es decir, encontrar la acción que maximiza la recompensa, para los 3 escenarios, utilizando los siguientes algoritmos:

- Action valued con selección $\epsilon-$**Greedy**
- **Upper Confidence Bound (UCB)**
- **Gradient Bandit Algorithm**

Aunque se puede apoyar en librerías para el manejo de arreglos y matrices como `numpy`, las implementaciones de estos algoritmos deben ser completamente propias.

A partir de sus implementaciones usted debe poder **concluir** cuál de los algoritmos es el mejor para cada escenario, discutiendo sus ventajas y desventajas, ajustando los parámetros necesarios y definiendo las métricas de evaluación que utilice. **Como mínimo debe mostrar la evolución de la recompensa promedio (sobre un horizonte de tiempo) para los tres escenarios con las tres implementaciones**. No obstante, se espera que haga uso de tablas y gráficas adicionales para apoyar sus explicaciones y conclusiones.

Todo su desarrollo, tanto de código como de discusión, debe realizarse en un único *Jupyter notebook* en el entorno de *Google Colab*, por lo que no es necesario realizar un informe adicional. Sin embargo, verifique que todo su *notebook* se ejecuta de forma correcta (**no se va a revisar código que no funcione**) y que todos sus resultados, análisis y conclusiones se encuentran de forma ordenada y sucinta en este documento.

Referencias

[1] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: A Bradford Book, 2018.