Trabajo Práctico 1

Introducción

La vida nos enfrenta constantemente a decisiones que nos obligan a equilibrar entre la seguridad de lo familiar y la promesa de lo desconocido, un dilema conocido como *"explore vs. exploit"*. Esta dicotomía, se manifiesta en una multitud de escenarios cotidianos. Por ejemplo, consideremos la elección de una cafetería para una merienda. ¿Optamos por un lugar al que hemos ido muchas veces, conocido por su calidad constante (*'exploit'*), o probamos uno de los nuevos locales que abrieron en Rosario con el furor del café de especialidad que podría ofrecer una experiencia culinaria increíble o decepcionante (*'explore'*)? Esta elección representa una encrucijada entre lo seguro y lo novedoso, entre el confort de lo familiar y la emoción de la novedad.

Este dilema también se extiende a decisiones más significativas en nuestras vidas, como la elección de una carrera, donde *'exploit'* implicaría seguir en un campo donde ya tenemos habilidades y experiencia, mientras que *'explore'* nos llevaría a aventurarnos en un nuevo dominio, potencialmente más gratificante pero también más arriesgado. Esta tensión entre explorar y explotar no es solo una curiosidad teórica; es un principio fundamental que guía nuestras decisiones diarias. Navegar entre estas dos opciones requiere una comprensión profunda de nuestras metas, recursos y el entorno en el que operamos, y es una habilidad esencial para la adaptación y el éxito en un mundo en constante cambio.

Estos párrafos de *coaching emocional* sirven como la introducción a este trabajo práctico, donde estudiaremos el problema del *multi-armed bandit*, que pone énfasis en el dilema *"explore vs. exploit"*. La traducción de *multi-armed bandit* es bandido multibrazo por lo que, por motivos obvios, nos quedaremos con la expresión en inglés.

El *multi-armed bandit*

El *multi-armed bandit* nos enfrenta a tres máquinas tragamonedas, tragaperras (si es por usar traducciones poco felices) o simplemente maquinitas (como les decimos en Rosario). El juego de las maquinitas consiste en hacer girar sus rodillos (analógicos o digitales) con el objetivo de obtener una combinación de símbolos ganadora y así acceder a un premio monetario (¡a esto sí que se le puede llamar éxito!).

Cada máquina tiene una probabilidad de éxito desconocida y potencialmente diferente, es decir, una probabilidad distinta de entregar un premio. El desafío consiste en decidir a cuál máquina dedicar nuestras tiradas con el objetivo de maximizar las ganancias totales. Aquí es donde entra el dilema: ¿conviene *"explotar"* la máquina que hasta ahora ha dado mejores resultados, o *"explorar"* otras máquinas que podrían tener una tasa de éxito mayor pero aún desconocida?

En la fase inicial, cuando se sabe poco sobre las máquinas, podría ser más prudente *"explorar"*, probando cada máquina varias veces para obtener una estimación aproximada de sus probabilidades de éxito. A medida que se acumulan datos sobre el rendimiento de cada máquina, la estrategia podría cambiar a *"explotar"* la máquina que ha demostrado ser la más rentable. Sin embargo, siempre existe la incertidumbre y la posibilidad de que una de las máquinas menos utilizadas tenga en realidad una tasa de éxito mayor. Este problema se complica aún más por el hecho de que cada elección de máquina proporciona información que podría alterar nuestra comprensión de cuál es la mejor opción. La solución óptima a este problema involucra un equilibrio cuidadoso entre explorar para ganar información y explotar esa información para maximizar las ganancias.

En este trabajo práctico consideraremos la situación simplificada e imaginaria en la que no cuesta dinero jugar con una máquina. Es decir, si obtenemos una combinación ganadora, sumamos una unidad monetaria , pero si no, no perdemos nada. Supondremos, además, un escenario ficticio en que el deseo por descubrir cual es la máquina ganadora nos tendrá jugando los 366 días del año 2024 . Lo que sí, cada día jugaremos con una sola máquina y volveremos al día siguiente…

El objetivo del trabajo consiste en evaluar y comparar diferentes estrategias de juego. Se analizarán mediante simulaciones diferentes estrategias de exploración y explotación de la máquina. ¿Dónde aparece la inferencia bayesiana? Partiremos de una creencia *a priori* para la probabilidad de éxito de cada máquina y la iremos actualizando con cada jugada.

Para el estudio mediante simulaciones, consideraremos que las probabilidades de éxito de las tres máquinas son $\theta_a = 0.30$, $\theta_b = 0.55$ y $\theta_c = 0.45$. Recordemos que estas probabilidades son desconocidas (no podemos basar nuestras estrategias en esos valores, sino en las estimaciones que vamos haciendo de ellos).

1. Simule 1000 repeticiones de una persona que tiene información confidencial y privilegiada y juega 366 días con la mejor máquina. Realice un histograma del dinero acumulado al finalizar el período. ¿Cuánto se espera que gane en promedio?

Estrategias

Utilizando nuestro ingenio e imaginación podríamos inventarnos al menos un par de estrategias que nos ayuden a determinar la máquina mas pagadora. Sin embargo, el género humano ya se ha inventado y debatido un sinfín de alternativas a seguir y nosotros podemos aportar nuestro granito de arena a la discusión mientras aprendemos estadística bayesiana y ejercitamos nuestras habilidades en R, ¡qué ofertón!.

Por lo tanto, para cada una de las estrategias presentadas debajo:

1. Construya una función en R que elija una máquina siguiendo el método indicado, obtenga un resultado (éxito o fracaso) y actualice la credibilidad sobre los posibles valores de la probabilidad de éxito correspondiente.
2. Utilice esa función para simular una secuencia de 366 días de juego. Registre la evolución diaria del dinero acumulado cada día, la cantidad de veces que se juega en cada máquina, y la distribución *a posteriori* de cada probabilidad de éxito. Muestre gráficamente los resultados.
3. Simule 1000 secuencias de 366 días de juego y analice los resultado
4. ¿Podría considerarse bayesiano este método de elección de máquina?

Consideraremos, en todos los escenarios, que la creencia *a priori* para θ_a, θ_b y θ_c se corresponde con una distribución Beta$(2, 2)$

Completamente al azar

Esta es la estrategia (o no-estrategia) más elemental: cada día, jugar con una máquina seleccionada al azar con probabilidad uniforme.

Greedy con tasa observada

Se elige la máquina que tenga la mayor tasa de éxito observada hasta el momento.

Greedy con probabilidad *a posteriori*

Se elige la máquina que tenga, hasta el momento, mayor probabilidad de éxito promedio *a posteriori*.

ϵ-*greedy* (con tasa observada)

Se selecciona la mejor máquina (la de mayor tasa de éxito observada según los datos actuales) con una probabilidad de $1 - \epsilon$ y se elige una máquina al azar con una probabilidad ϵ.

Softmax

Dada la tasa observada para cada máquina i, π_i, se calcula una probabilidad de elegir cada máquina utilizando la función *softmax*:

$$\Pr(i) = \frac{e^{\pi_i/\tau}}{\sum_{j=1}^{3} e^{\pi_i/\tau}}$$

donde τ es un parámetro de "temperatura" que controla el grado de exploración. Luego, se elige la máquina i con probabilidad $\Pr(i)$.

Además:

- **i.** Implemente la función *softmax* que, dadas tres tasas observadas, devuelva la probabilidad de elegir cada máquina. La función debe recibir la "temperatura" como argumento.
- **ii.** Explique qué función cumple el parámetro de "temperatura" en términos de explorar versus explotar.

Upper-bound

Se selecciona la máquina que tenga el mayor extremo derecho de un intervalo de credibilidad (construido a partir de la distribución *a posteriori* de la probabilidad de éxito).

Thompson sampling

Para seleccionar una máquina, se toma una muestra de la distribución *a posteriori* de las probabilidades de éxito de cada máquina y se elige la máquina correspondiente a la muestra más grande.