TRABAJO PRÁCTICO 1.1 SIMULACIÓN DE UNA RULETA

Ornella Colazo
Universidad Tecnológica Nacional
Ingeniería en Sistemas
Legajo 47864
ornecolazo@gmail.com

Diego Navarro
Universidad Tecnológica Nacional
Ingeniería en Sistemas
Legajo 48029
navarrodiego201513@gmail.com

Matias Petrich
Universidad Tecnológica Nacional
Ingeniería en Sistemas
Legajo 46852
matias.petrich@gmail.com

Ramiro Cordoba
Universidad Tecnológica Nacional
Ingeniería en Sistemas
Legajo 46824
ramirocordobautn@gmail.com

Matias Ferullo
Universidad Tecnológica Nacional
Ingeniería en Sistemas
Legajo 48039
matias.ferullo1@gmail.com

25 de Marzo, 2023

ABSTRACT

El siguiente trabajo de investigación incluye el análisis de una simulación de una ruleta desarrollada en el lenguaje Python, calculando asi valores de parámetros estadísticos tales como la probabilidad, la frecuencia relativa, el desvio estandar y la varianza que resultaron de las tiradas realizadas en la simulación. Se encuentran graficados los cuatro parámetros estadísticos mencionados anteriormente en 10000 tiradas de las cuales de corrieron cuatro veces. Se plantea un desarrollo y una conclusión en base a los resultados obtenidos.

1 Introducción

La ruleta es un juego de azar que se puede encontrar en los casinos. Existen variaciones de este juego pero para realizar nuestro análisis utilizamos la ruleta europea, la cual consiste de 37 números. El conjunto de 37 números es del número 0 al 36.

El objetivo del juego de la ruleta es apostar al número o números que predecimos que va a salir en la tirada de la bola sobre el cilindro, cuando este pare de girar. Si apostamos a un solo número, tenemos una probabilidad de acertar de 1/37.

La ruleta es un ejemplo de una experiencia que tiene la particularidad de generar resultados que no son predecibles, existen factores no controlables que generan variación. Es decir que es una experiencia llamada aleatoria. La teoria de la Probabilidad proporciona bases matemáticas para describir estas experiencias aleatorias.

La simulación es una técnica que puede utilizarse en computadoras para analizar modelos matemáticos complejos. Actualmente, gracias a los avances tecnológicos que nos permiten tener computadoras rápidas e inteligentes en nuestras casas, es factible usar una computadora para escribir un código de programación y este ser usado como herramienta para realizar estudios de simulación. Por lo tanto, para estudiar el caso del juego de la ruleta vamos a simularlo con el lenguaje de programación Python. El objetivo de nuestro estudio es observar cómo se comporta una ruleta.

1.1 Conceptos de Probabilidad.

Para empezar, definiremos algunos conceptos a tener en cuenta sobre experiencias aleatorias.

Los resultados individuales de las experiencias aleatorias parecen ocurrir de forma arbitraria. Sin embargo, cuando la experiencia se repite un gran número de veces, aparece un modelo definido de regularidad. Esa regularidad hace posible la construcción de un modelo matemático con el cual se analiza la experiencia.

Para construir nuestro modelo definimos un espacio muestral asociado a la experiencia de la ruleta. El espacio muestral es el conjunto formado por los posibles resultados de la experiencia bajo estudio. Este espacio lo simbolizamos con la letra S.

S = { 0, 1, 2, 3,..., 36 }

Las experiencias aleatorias dan como resultado sucesos elementales. Un suceso elemental es un subconjunto del espacio muestral que contiene un solo resultado posible. Una variable aleatoria se define al asignar un valor numérico a cada suceso elemental de una experiencia aleatoria. Este estudio es de variables aleatorias discretas, es decir que los valores que puede tomar van a ser numerables. Si Ai es un suceso del espacio muestral S de la experiencia de la ruleta y lo definimos como: "sale el número i en una tirada", i siendo un número perteneciente al espacio muestral S que puede tomar los valores entre 0 y 36, entonces la probabilidad del suceso Ai se define como las veces que sale el número i en una tirada sobre la cantidad de posibles resultados, es decir 1/37 (en una tirada puede salir un solo número).

A partir de este concepto sabemos que todos los números tienen la misma probabilidad de salir en una tirada de la ruleta.

2 Desarrollo

Tomamos de la Probabilidad conceptos y fórmulas para construir nuestro análisis [1, 2, 3].

2.0.1 Fórmulas utilizadas.

Frecuencias Las frecuencias se incluyen en una tabla de distribución de frecuencias, la cuál describiremos su utilidad más adelante.

La frecuencia absoluta (fa) corresponde con el número de veces que un dato se repite dentro del conjunto.

La frecuencia relativa (fr) es el cociente entre la frecuencia absoluta y un número n, el cual es el total de observaciones del conjunto. Por lo tanto, corresponde con la proporción de veces que aparece ese dato con respecto al total n.

Tomamos la frecuencia absoluta para luego obtener la frecuencia relativa de un número.

$$f_r = \frac{f_a}{n} \tag{1}$$

Parámetros Estadísticos

Promedio El promedio o media aritmética es el cociente entre la sumatoria de las observaciones y un valor n, el cuál corresponde con el total de las observaciones.

$$\mu = \frac{\sum_{i=1}^{b} x_i}{n} \tag{2}$$

La μ es el símbolo de la media poblacional. $\sum$ nos indica que hay que añadir una lista especifica de números, la i en la fórmula nos indica cuál de las observaciones hay que sumar y n representa el número total de observaciones o de datos utilizados para el cálculo del promedio.

Varianza La varianza σ^2es una medida que representa la variabilidad de los datos respecto de la media.

$$\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 . p_i \tag{3}$$

La n representa el número total de observaciones o de datos utilizados para el cálculo de la varianza. $\sum$ nos indica que hay que añadir una lista especifica de numeros, xi es el valor del dato i y μ es el símbolo de la media poblacional. Por ultimo, pi representa a la función de probabilidad en el dato i.

Desvío Estandar El desvío estandar σ es la raiz cuadrada positiva de la varianza e indica cuánto se desvía del promedio

$$\sigma = \sqrt{\sigma^2} \tag{4}$$

2.1 Planteo y ejecución de la Simulación.

Una vez planteado nuestro espacio muestral, el suceso elemental, las frecuencias y los parámetros, nos hicimos la siguiente pregunta: Si elijo apostar al número 7 y este sale m veces en n tiradas, ¿qué sucede con su frecuencia relativa?

En un principio lo plasmamos en una distribución de frecuencias representando la frecuencia relativa de la frecuencia absoluta del numero 7.

Nº Tirada	Numero	Fa del numero 7	Fr del numero 7
1	7	1	1
2	4	1	0,5
3	25	1	0,333...
4	31	1	0,25
5	0	1	0,2
6	13	1	0,1666...
7	4	1	0,142...
8	7	2	0,25
9	22	2	0,222...
...	...	...	...
25	7	3	0,12
...	...	...	...
60	7	4	0,066...
...	...	...	...

Table 1: Ejemplo de Fr y Fa del numero 7 en una simulación manual

.

Hipótesis. A partir de estas observaciones obtenidas de las tablas, planteamos la siguiente hipótesis: La frecuencia relativa tiende a 0 a medida que se realizan las tiradas. Para demostrar o refutar esta hipótesis, realizamos una simulación, la cual fue planteada con un algoritmo en el lenguaje de programación Python.

Cantidad de tiradas elegidas. Primero debíamos ajustar la cantidad de tiradas que iban a simularse. La frecuencia relativa de 7 varía en cada muestra de n observaciones, pero a la larga surge cierta regularidad, la frecuencia relativa de 7 tiende a estabilizarse alrededor de un valor constante. Es por esto que para el experimento realizamos 10000 tiradas. Este número no fue elegido de una manera arbitraria, si no que fue un número elegido fundamentado por expertos en estadísticas y juegos de azar generalmente recomiendan hacer al menos 10000 tiradas para obtener una simulación razonablemente precisa de la ruleta. Otro fundamento para utilizar dicho número de tiradas es por la ley de los grandes números la cual establece que a medida que aumenta el número de muestras en una distribución, la media de esas muestras se acerca cada vez más a la media real de la población. En el caso de la ruleta, esto significa que a medida que se realizan más tiradas, la distribución de probabilidad de los resultados se acerca cada vez más a la distribución de probabilidad teórica de la ruleta. Teniendo en cuenta los recursos de una computadora y el tiempo de ejecución al realizar una simulación de la ruleta, es importante encontrar un equilibrio entre el número de tiradas que se realizan y la cantidad de recursos y tiempo disponibles[4].

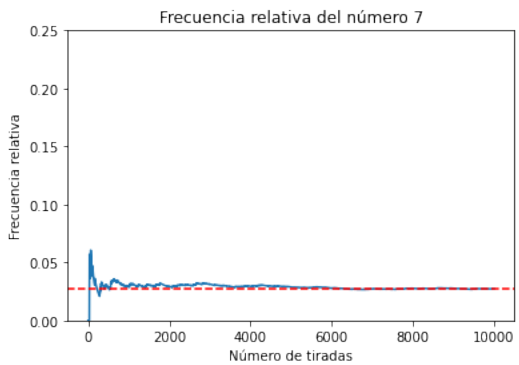
Descripción del código. Comenzamos definiendo una función llamada "ruleta", la cual retorna un número aleatorio entre 0 y 36, representando los números de una ruleta. Previamente realizando una inicialización de variables que vamos a utilizar para cálculos, creamos un bucle, cuyo rango es igual a la cantidad de tiradas que se requieran y registra los resultados en una lista llamada "resultados". Dentro de este bucle, sumamos los resultados a una de las variables para luego calcular el promedio y también aumentamos un contador cada vez que uno de estos resultados es un siete, para calcular su frecuencia relativa, desvío estandar y varianza, realizado con funciones de la librería numpy. Mediante la libreria de Python matplotlib.pyplot, graficamos los cálculos efectuados en función del número de tiradas. Además, en cada gráfica se incluyeron los valores estimados para cada parámetro estadístico [5].

2.2 Gráficas y análisis de los resultados.

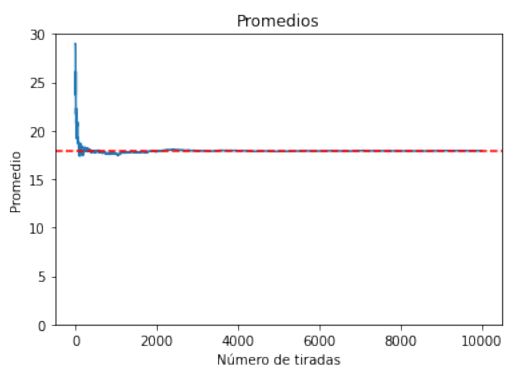
En las cuatro gráficas generadas, se muestran los promedios de los resultados, la frecuencia relativa del número 7, la varianza y el desvío en función del número de tiradas. Cada gráfica también incluye una línea punteada roja que representa el valor esperado teórico.

Realizamos 4 corridas de la simulación para poder obtener una mejor conclusión de los resultados. Queríamos observar si la frecuencia relativa variaba en las diferentes corrida de la simulación y si realmente tendía a 0.

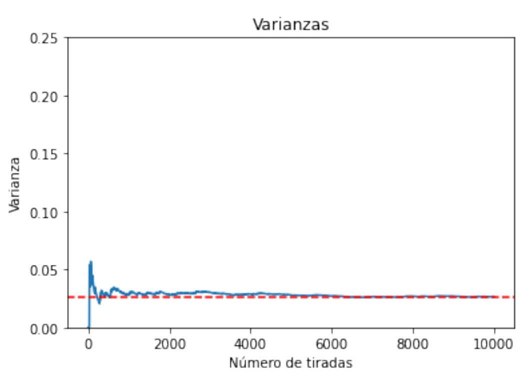
2.2.1 Corrida 1.



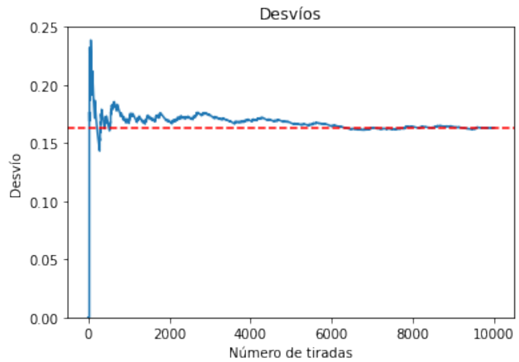
Se puede observar en la grafica que la frecuencia relativa del numero 7 (linea azul) comienza, en las primeras tiradas con valores bastante alejados de la frecuencia relativa esperada, pero luego de varias tiradas vemos que se estabiliza en el valor de la frecuencia relativa esperada, vemos que despues de la tirada 6000 es practicamente el mismo valor de la frecuencia relativa de 7 que el valor de la frecuencia relativa esperada del 7 (linea punteada roja)



Podemos observar que en los primeros valores toma valores promedios anomalos (linea azul) pero a partir de la tirada 2000 los valores promedios y el valor promedio esperado (linea punteada roja) es practicamente el mismo



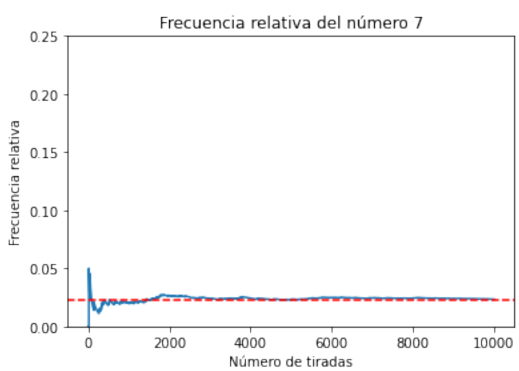
La varianza de 7(Linea azul), como ya vimos anteriormente toma valores, y este caso no es la excepcion, tiene bastante variabilidad, pero al pasar las tiradas, aproximidamente en la 6000, la varianza y varianza esperada (Linea roja punteada) son casi las mismas

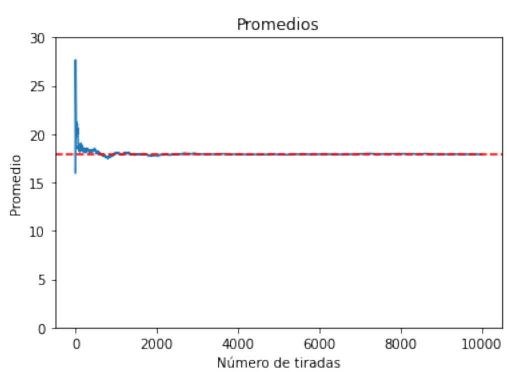


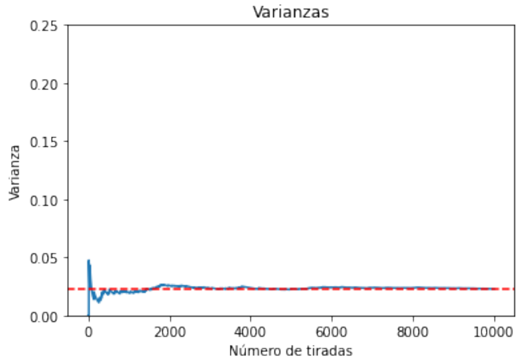
El caso des desvio tampoco es la expecion a la regla, con un numero de tiradas pequeño, el desvio(Linea azul) y el desvio esperado(linea roja punteada) tienen valores dispares, los valores del desvio son muy variables, incluso mas que en los casos anteriores, luego, al aumentar el numero de las tiradas se va estabilizando en torno a las 6000 tiradas.

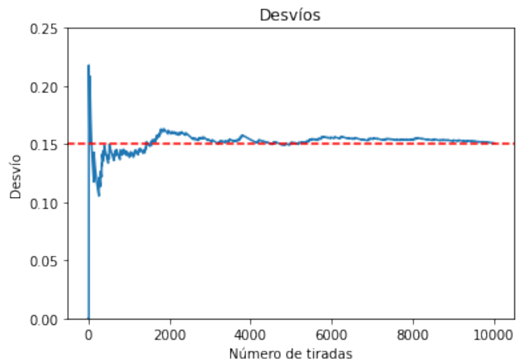
2.2.2 Corrida 2.

El caso de la corrida numero 2, podemos ver que el comportamiento de las funciones a lo largo de las tiradas es muy similar a los de la tirada 1, con exepcion del desvio.



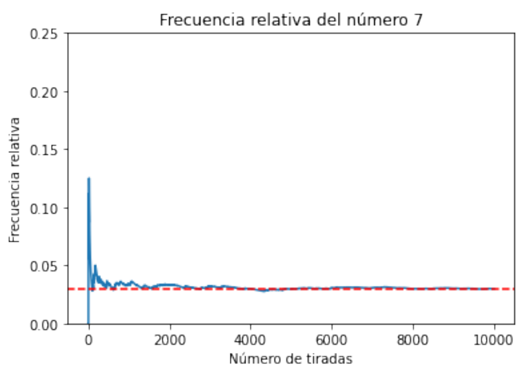


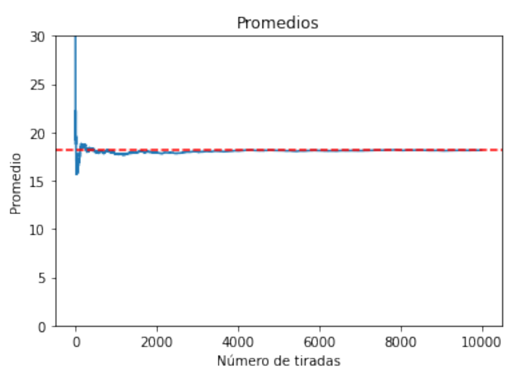


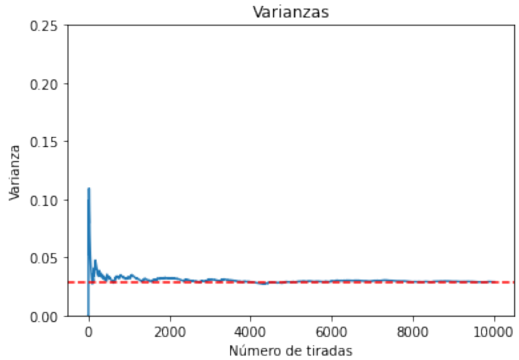


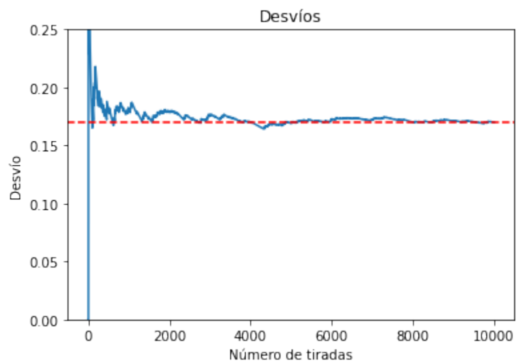
El desvio estandar(Linea azul) es aun mas variable en comparacion a lo esperado (Linea roja punteada) que en la corrida 1, se puede observar que no se puede estabilizar hasta casi la tirada 10000.

2.2.3 Corrida 3.









2.3 Análisis de las gráficas.

En las primeras tiradas es visible que los valores en todos los parámetros varían enormemente, pero que a medida que avanza el número de tiradas, esos valores se van estabilizando hasta practicamente coincidir con el valor estimado. Esta regularidad hallada es consecuencia de haber tomado un número de tiradas adecuado para llegar a los valores que pueden ser, aproximadamente, certeros.

A partir de lo observado en las gráficas, podemos ver que las frecuencias, promedios, varianzas, desvíos entre las diferentes corridas, practicamente no varían. Los valores estimados que se obtuvieron son los siguientes:

$$\mu_e \approx 18 \tag{5}$$
$$\sigma_e^2 \approx 0,02 \tag{6}$$
$$\sigma_e \approx 0,16 \tag{7}$$
$$fr_e \approx 0,02 \tag{8}$$

El promedio μ representa el valor típico central, sin embargo es alterado por números atípicos o extremos, por lo que esta medida resumen puede no ser una representación completamente genuina. Para medir cuánto se desvían los datos de las observaciones de la media tenemos las medidas de dispersión como la varianza σ^2 y el desvío σ. Por lo que estas medidas nos ayudan a leer mejor el resultado obtenido en la media. Interpretando, el valor típico central de los números salidos en la ruleta es 18, con una desviacíon de la media de 0,16.

Con respecto de la frecuencia relativa del número 7, pudimos estudiar nuestra hipótesis y entendemos que la frecuencia no tiende a 0, si no que se acerca aproximadamente a 0,02.

3 Conclusión

Para resumir, realizamos la simulacion de 10000 tiradas en el programa desarrollado en Python y concluimos que la frecuencia relativa del número 7 no tiende a 0, sino que tiende a 0,02. Sabiendo que son 37 los numeros posibles que pueden salir, podemos decir que, probabilisticamente, la frecuencia relativa a la que debería tender cada número sería 1/37, que nos da como resultado aproximadamente 0,0270, lo cual es muy cercano a 0,02, que fue el resultado de la simulacion realizada.

Por lo que podemos concluir que nuestra hipótesis "La frecuencia relativa tiende a 0 a medida que se realizan las tiradas" es falsa ya que tiende a 0,02, es decir, 1/37. Entonces todos los números de la ruleta tienen la probabilidad de salir de 1/37, lo cual es una probabilidad muy baja si se busca utilizar este estudio para justificar apostar o no en el juego de la ruleta, por lo que no se recomendaría apostar cuando las probabilidades son tan cercanas a 0.

References

[1] Tratamiento de datos catedra de probabilidad y estadística UTN facultad regional de Rosario

[2] Introducción a la probabilidad de Raúl David Katz

[3] Variables aleatorias discretas de Raúl David Katz y Pablo Sabatinelli

[4] Libro "Probability Guide to Gambling" de Catalin Barboianu

[5] https://python-para-impacientes.blogspot.com/2014/08/graficos-en-ipython.html