TRABAJO PRÁCTICO RULETA - SIMULACIÓN 2023

Gardeñes Fernando
Cátedra de Simulación - ISI
UTN - FRRo
Zeballos 1341, S2000 Rosario, Santa Fe
fergardenes7@gmail.com

Biscaldi Ivan
Cátedra de Simulación - ISI
UTN - FRRo
Zeballos 1341, S2000 Rosario, Santa Fe
ibiscaldi@frro.utn.edu.ar

Saludas Fausto
Cátedra de Simulación - ISI
UTN - FRRo
Zeballos 1341, S2000 Rosario, Santa Fe
fausaludas14@gmail.com

Schlieper Tadeo
Cátedra de Simulación - ISI
UTN - FRRo
Zeballos 1341, S2000 Rosario, Santa Fe
tadeo.sch@hotmail.com

Nicolás Fierro
Cátedra de Simulación - ISI
UTN - FRRo
Zeballos 1341, S2000 Rosario, Santa Fe
nicofierro1@gmail.com

29 de mayo de 2023

ABSTRACT

El presente trabajo aborda el estudio y la simulación de varios lanzamientos de una ruleta francesa utilizando el lenguaje de programación Python para su posterior interpretación y obtención de conclusiones.

1. Introducción

La ruleta es un juego de azar típico de los casinos, cuyo nombre viene del término francés *roulette*, que significa "ruedita" o "rueda pequeña". Su uso como elemento de juego de azar, aún en configuraciones distintas de la actual, no está documentado hasta bien entrada la Edad Media. Es de suponer que su referencia más antigua es la llamada Rueda de la Fortuna, de la que hay noticias a lo largo de toda la historia, prácticamente en todos los campos del saber humano.

La "magia" del movimiento de las ruedas tuvo que impactar a todas las generaciones. La aparente quietud del centro, el aumento de velocidad conforme nos alejamos de él, la posibilidad de que se detenga en un punto al azar; todo esto tuvo que influir en el desarrollo de distintos juegos que tienen la rueda como base.

Las ruedas, y por extensión las ruletas, siempre han tenido conexión con el mundo mágico y esotérico. Así, una de ellas forma parte del tarot, más precisamente de los que se conocen como arcanos mayores.

Según los indicios, la creación de una ruleta y sus normas de juego, muy similares a las que conocemos hoy en día, se debe a Blaise Pascal, matemático francés, quien ideó una ruleta con treinta y seis números (sin el cero), en la que se halla un extremado equilibrio en la posición en que está colocado cada número. La elección de 36 números da un alcance aún más vinculado a la magia (la suma de los primeros 36 números da el número mágico por excelencia: seiscientos sesenta y seis).

Esta ruleta podía usarse como entretenimiento en círculos de amistades. Sin embargo, a nivel de empresa que pone los medios y el personal para el entretenimiento de sus clientes, no era rentable, ya que estadísticamente todo lo que se apostaba se repartía en premios (probabilidad de 1/36 de acertar el número y ganar 36 veces lo apostado).

En 1842, los hermanos Blanc modificaron la ruleta añadiéndole un nuevo número, el 0, y la introdujeron inicialmente en el Casino de Montecarlo. Ésta es la ruleta que se conoce hoy en día, con una probabilidad de acertar de 1/37 y ganar 36 veces lo apostado, consiguiendo un margen para la casa del 2,7 % (1/37).

Más adelante, en algunas ruletas (sobre todo las que se usan en países anglosajones) se añadió un nuevo número (el doble cero), con lo cual el beneficio para el casino resultó ser doble (2/38 o 5,26 %), aunque también se cambiaron otras reglas del juego.

La ruleta francesa es la más conocida y la más extendida, debido al trabajo de difusión que los franceses han realizado de este juego, promocionándola en todos los casinos más lujosos de Europa, como en el casino de Montecarlo que se ubica en el Principado de Mónaco.

Para el caso de estudio se seleccionó la ruleta francesa, conteniendo 37 números (del 0 al 36), debido a su mayor popularidad y sencillez.

2. Marco Teórico

2.1. Experiencia aleatoria

La experiencia aleatoria es la experiencia en la cual no se puede saber el resultado particular que ocurrirá, pero si se puede describir el conjunto de todos los resultados posibles. Nosotros definimos la experiencia como:

E: se lanza la bola y se observa el número que sale en la ruleta

2.2. Variables y espacio muestral

Se denomina espacio muestral al conjunto de todos los posibles sucesos de la experiencia E. Dada la experiencia aleatoria E, definida anteriormente, su espacio muestral es:

$$A = \{0, 1, 2, 3, ..., 35, 36\} \quad \textit{Espacio muestral de E} \tag{1}$$

Siendo los casos posibles(resultados posibles realizar el experimento E) la cardinalidad de este conjunto:

$$|A| = 37 \quad \textit{Cantidad de posibles resultados} \tag{2}$$

Y se denomina variable a X, el resultado de realizar la experiencia aleatoria

$$X \quad \textit{Variable aleatoria} \tag{3}$$

2.3. Frecuencia absoluta

La frecuencia absoluta es una medida estadística que nos da información acerca de la cantidad de veces que se repite un suceso al realizar un número determinado de experimentos aleatorios. En otras palabras es la cantidad de veces que X asumió un valor i luego de N experimentos.

$$f_i \quad \textit{Frecuencia absoluta de ocurrencia de del suceso i} \tag{4}$$

Dado el espacio muestral A, la sumatoria de las frecuencias absolutas de sus i elementos es igual a la cantidad de experimentos:

$$N = \sum_{i=0}^{36} f_i \tag{5}$$

2.4. Frecuencia relativa

La frecuencia relativa es una medida estadística que se calcula como el cociente de la frecuencia absoluta de algún valor de la población/muestra entre el total de valores que componen la población/muestra. Las frecuencias relativas son valores que varían entre 0 y 1. La suma de todas las frecuencias relativas es siempre 1.

Para calcular la frecuencia relativa, se necesita calcular anteriormente la frecuencia absoluta.

$$fr_i = \frac{f_i}{N} \tag{6}$$

2.5. Probabilidad de ocurrencia

Al calcular sucesivamente la frecuencia relativa de los elementos de A, realizando n experimentos, cuando n sea lo suficientemente grande, cada fr_i tendrá tendencia a cierto valor, el cual se denominara probabilidad de ocurrencia de i:

$$\lim_{n\to\infty} fr_i = \lim_{n\to\infty} \frac{f_i}{n} = P(X = i) \tag{7}$$

Sucede que en el caso de la ruleta al observar los valores de $P(X = i)$ estos tienden a ser iguales para todo i. Esto indica que los i sucesos son equiprobables. Y según Regla de Laplace, en todo experimento aleatorio en que los sucesos sean equiprobables, se define a la probabilidad del suceso i como el cociente entre el numero de casos afirmativos y el numero de casos posibles.

$$P(X = i) = \frac{\textbf{Casos afirmativos}}{\textbf{Casos posibles}} \tag{8}$$

Como los casos afirmativos de i son unicamente i(es decir solo hay un caso afirmativo), y los casos posibles son $|A| = 37$:

$$P(X = i) = \frac{1}{37}$$

2.6. Media aritmética

La media aritmética es el valor obtenido al sumar todos los datos y dividir el resultado entre el número total de observaciones. Se suele representar con la letra X. Si tenemos una muestra de N valores, X_i, la media aritmética, $\bar{X}$, es la suma de los valores divididos por el número de elementos; en símbolos:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{X_1 + X_2 + ... + X_{N-1} + X_N}{N} \tag{9}$$

Pero en el caso de la ruleta X asumirá f_i veces cada valor posible, asi por 6, 7 y por 9, si se realiza un numero lo suficientemente grande de experimentos:

$$\mu = \frac{1}{N}\sum_{i=0}^{|A|-1} X_i * f_i = \sum_{i=0}^{|A|-1} X_i * fr_i = \sum_{i=0}^{36} \frac{1}{37} X_i = 18$$

2.7. Varianza

La varianza es la medida de dispersión que indica cuanto varia un conjunto de datos respecto a la media. Se calcula encontrando la media de los cuadrados de las desviaciones de cada valor en la muestra con respecto a la media muestral. Si la varianza es baja, significa que los valores están cercanos a la media, mientras que si la varianza es alta, significa que los valores están más dispersos.

$${\sigma_N}^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \bar{X})^2 \tag{10}$$

En la ruleta la varianza es:

$${\sigma_{37}}^2 = \frac{1}{37}\sum_{i=1}^{37}(X_i - 18)^2 = 114$$

2.8. Desviación estándar

La desviación estándar o desvio es una medida de extensión o variabilidad. Se utiliza para calcular la variación o dispersión en la que los puntos de datos individuales difieren de la media. La desviación estándar es igual a la raíz cuadrada de la varianza

$$\sigma_N = \sqrt{{\sigma_N}^2} \tag{11}$$

Siendo el desvio en el caso de la ruleta:

$$\sigma_{37} = \sqrt{114} \approx 10{,}6771$$

2.9. Muestras como estimadores de población

Se denomina población a la totalidad de las observaciones sobre las que se realiza el estudio estadístico.

Se denomina muestra a un subconjunto de la población.

En la mayoría de los estudios estadísticos no es posible observar la totalidad de la población ya sea porque esta es infinita o muy grande (por ej. la población es el conjunto de resultados obtenidos de jugar a la ruleta). Cuando esto sucede es conveniente tomar una o varias muestras.

Para poder inferir a partir de una muestra sobre la población es necesario que la muestra sea representativa. Seleccionar intencionalmente algunos resultados solo generara un sesgo, por ello la muestra debe ser aleatoria.

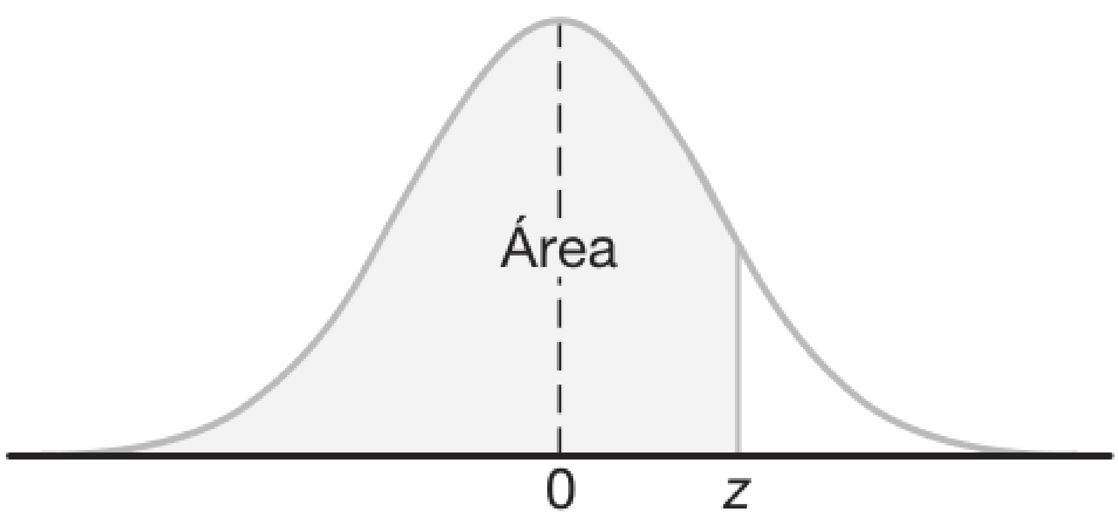
Suponiendo que una muestra es aleatoria, y su media aritmética es $\bar{X}$. Es probable que $\bar{X} \neq \mu$, en cuyo caso el error es de:

$$e = |\mu - \bar{X}| \tag{12}$$

Para inferir sobre la población con $100 \cdot (1 - \alpha)\,\%$ de confianza, la misma debe tener un tamaño n:

$$n = \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{e}\right)^2 \tag{13}$$

Donde $z_{\frac{\alpha}{2}}$ es el valor en que $100 \cdot (1 - \alpha)\,\%$ es igual al porcentaje del área bajo la curva de la distribución normal estandarizada($\mu = 0$ y $\sigma = 1$) desde $-\infty$ hasta z.



3. Metodología

En este estudio, se llevarán a cabo un total de n observaciones que serán repetidas m veces. El objetivo principal es demostrar que los valores de diferentes estadísticos tienden a un valor esperado, que puede ser calculado antes de la realización de los experimentos. Durante las m repeticiones, se analizará la variación de varios estadísticos para obtener una mejor comprensión de los resultados obtenidos.

La ruleta que utilizamos en nuestro estudio es la francesa, que tiene los 36 números (del 1 al 36) y el 0. Dado que todos los números tienen la misma probabilidad de ocurrencia, nuestra frecuencia relativa será de $\frac{1}{37}$, el valor esperado del promedio es 18.

También existen otros parámetros, llamados varianza y desvío, que son medidas de dispersión de los datos en la población, el valor esperado de la varianza es 114 y en consecuencia el valor esperado del desvío es 10,6771.

Decidimos que se admitirá que la media aritmética de cada muestra tenga un error de $\pm 0{,}05$, y se desea tener una confianza del $99{,}98\,\%$. Por lo tanto las observaciones requeridas para que la muestra sea representativa a tal precisión deben de no menos que:

$$n = \left(\frac{3{,}49 \cdot 10{,}6771}{0{,}05}\right)^2 \approx 555$$

3.1. Herramientas utilizadas

Python es un lenguaje de programación interpretado, fácil de aprender y leer. La licencia de Python es administrada por la Python Software Fundation, pero el lenguaje es de código abierto y se desarrolla bajo licencias aprobadas por la OSI.

Python cuenta con un índice de paquetes (PyPI/Python Package Index), que es un repositorio que alberga módulos/librerías de terceros así como los módulos estándar del lenguaje. Estos módulos pueden tener miles de aplicaciones como por ejemplo en la ciencia y los números.

La función numpy.random.randint() de la librería NumPy nos permitió generar muestras aleatorias de números enteros dentro de un rango específico. Por ejemplo, para generar una muestra aleatoria de 100 números enteros entre 0 y 10, se utiliza la siguiente línea de código: muestras = numpy.random.randint(0, 10, size=100). Esto nos permitió generar muestras aleatorias de datos para analizarlos en nuestro estudio.

La función ‘numpy.arange()‘ de NumPy nos permitió generar secuencias numéricas para crear muestras de datos específicas. Por ejemplo, si queríamos generar una muestra de datos con números enteros consecutivos desde 1 hasta 100, podríamos utilizar la siguiente línea de código: ‘muestras = numpy.arange(1, 101, 1)‘. Esto nos permitió generar muestras de datos personalizadas para nuestro análisis.

Tambien utilizamos las librerías statistics y matplotlib de Python para analizar y visualizar nuestros datos. En particular, empleamos las funciones de la librería statistics para calcular medidas estadísticas como el promedio, la desviación estándar, el desvío y la varianza de nuestras muestras de datos. Utilizamos las funciones statistics.mean(), statistics.stdev(), statistics.variance() y statistics.median() para calcular estas medidas estadísticas respectivamente. Estas funciones nos permitieron obtener información valiosa sobre la distribución de nuestros datos.

4. Casos de estudio

Se estudiarán dos casos y luego se hará una comparación entre ellos. El caso consiste en simular, al menos una vez, varios experimentos y medir por cada uno la frecuencia relativa, el valor promedio, el valor del desvío estándar y el valor de la varianza, para estudiar su evolución. Para la frecuencia relativa es necesario establecer un número a partir del cual se mide su frecuencia a lo largo de la simulación.

Entonces, cada caso de estudio tiene tres condiciones: la cantidad de simulaciones (también llamado corridas), la cantidad de experimentos por simulación (también llamado tiradas o repeticiones), y el número a observar (que corresponde al número con el que se mide la frecuencia). Le daremos nombres de variable a estas condiciones con el fin de identificarlas en el código fuente. A la variable de la cantidad de corridas la llamamos ***c***, a la cantidad de tiradas ***t***, y al número a observar ***numObs***.

El Caso 1 consistirá en 1 corrida, mientras que el Caso 2 en 37 corridas. Ambos tendrán 560 tiradas por corrida y el número 7 a observar.

4.1. Caso 1: 1 corrida

En este caso, ***c*** toma el valor de 1 ya que se tiene en cuenta una única simulación. ***t*** = 560 es la cantidad de experimentos, es decir la cantidad de repeticiones de la experiencia aleatoria.

$$\boldsymbol{c} = 1$$

$$t = 560$$

$$numObs = 7$$

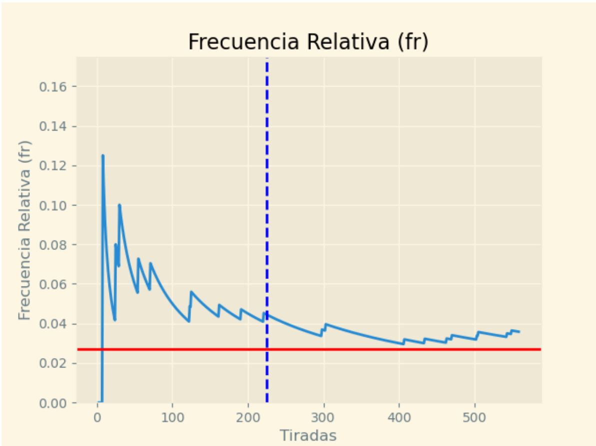


Figura 1: Comparativa de la frecuencia relativa del número 7 a lo largo de 560 tiradas respecto al valor esperado(recta de rojo)

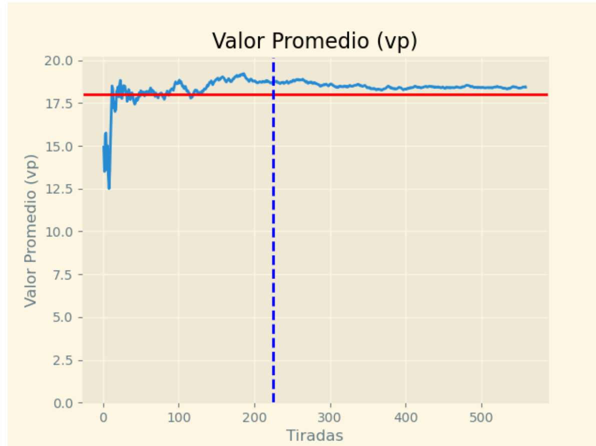


Figura 2: Comparativa del valor promedio a lo largo de 560 tiradas respecto al valor esperado(recta de rojo)

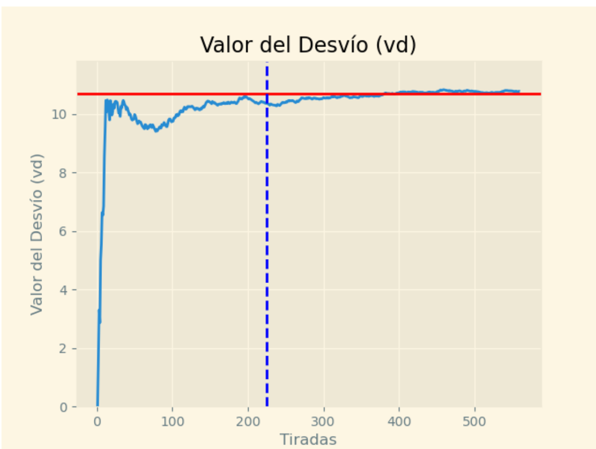


Figura 3: Comparativa del desvío a lo largo de 560 tiradas respecto al valor esperado(recta de rojo)

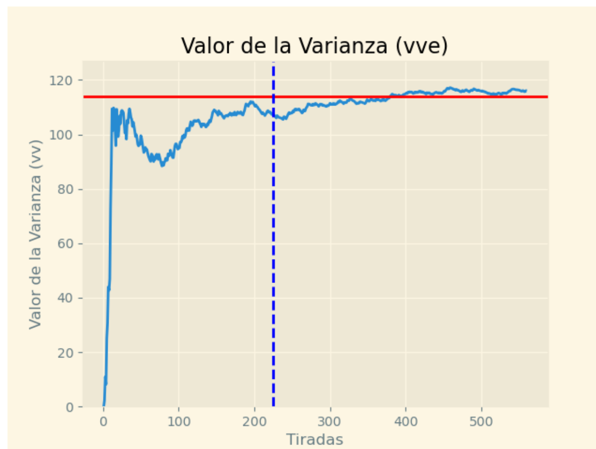


Figura 4: Comparativa de la varianza a lo largo de 560 tiradas respecto al valor esperado(recta de rojo)

4.2. Caso 2: 15 corridas

Para el segundo caso ***c*** toma el valor de 37 considerando que permitirá una buena evaluación de resultados. Luego, ***t*** vuelve a ser 560 y ***numObs*** a ser 7.

$$c = 37$$

$$t = 560$$

$$numObs = 7$$

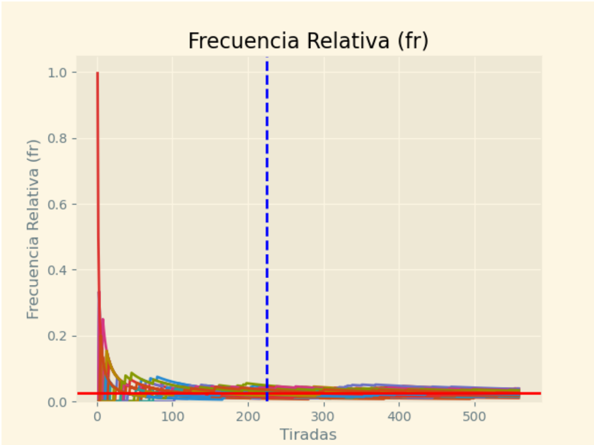


Figura 5: Frecuencia relativa del número 7 de 37 corridas a lo largo de 560 tiradas

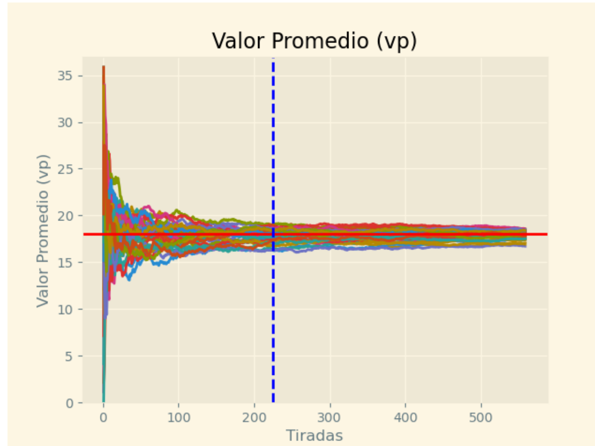


Figura 6: Promedio de 37 corridas a lo largo de 560 tiradas

Figura 7: Desvío de 37 corridas a lo largo de 560 tiradas

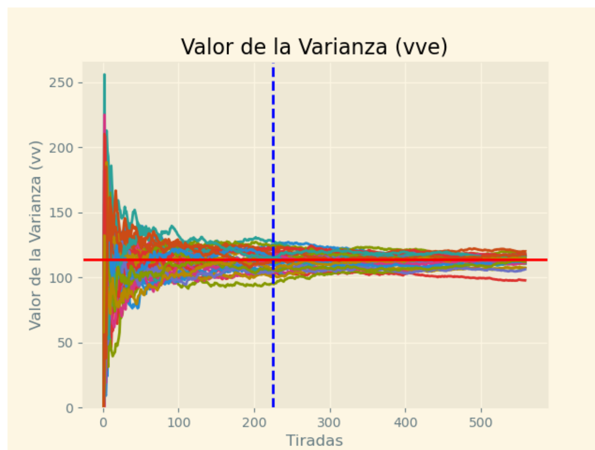


Figura 8: Varianza de 37 corridas a lo largo de 560 tiradas

5. Conclusión

En el presente estudio se llevaron a cabo 560 tiradas en cada uno de los dos casos de estudio. Se observó que hasta cierto valor de n(numero de tiradas), las gráficas presentaron una gran variabilidad y oscilación entre diferentes valores. Sin embargo, al superar un valor conveniente de n, las gráficas comenzaron a suavizarse y a tomar una forma más lineal, llegando en algunos casos a converger con la gráfica esperada.

Esto sucede en torno a las 200 tiradas. Como curiosidad, si se divide al tamaño ideal de la muestra entre la cantidad de posibles resultados, y al resultado de esto de lo eleva al cuadrado se obtiene el valor que aparece como una recta azul en las gráficas:

$$\left(\frac{555}{37}\right)^2 = 225$$

Es importante destacar que, aunque no se presentó en este estudio, existe la posibilidad de que surja un resultado anómalo. Por esta razón, es crucial realizar varias repeticiones del experimento, ya que una única corrida no sería suficiente para determinar resultados significativos y podríamos estar sesgados por un resultado no representativo.

Como se puede observar en todos los gráficos, a partir de las 200 tiradas los valores del promedio, desvío y frecuencia relativa alcanzan y/o se aproximan a los valores esperados. Esto indica que manteniendo la apuesta por un numero a lo largo de las tiradas, durante todas las tiradas previas a la 200, la frecuencia de tener éxito en la apuesta puede ser aun menor que 1/37.

Referencias

[1] Código fuente utilizado para la realización del Trabajo Práctico:

https://github.com/Faus14/Simulacion_2023

[2] Formulas de calculo probabilistico:

Probabilidad y estadística para ingeniería y ciencias 9º edición- Pearson
https://www.pearson.com/en-us/subject-catalog/p/probability--statistics-for-engineers--scientists/P200000007119/9780137273546

[3] Documentacion acerca del lenguaje de programacion Python

https://www.python.org/about/
https://pypi.org/

[4] Documentación utilizada como soporte para la realización de cálculos matemáticos:

https://docs.python.org/3/library/statistics.html

[5] Documentación sobre uso de librerías para conformación de gráficas:

https://matplotlib.org/stable/plot_types/index.html
https://www.w3schools.com/python/matplotlib_pyplot.asp

[6] Documentacion sobre el juego de la ruleta en casinos:

https://www.betsson.es/blog/casino-online/ruletas/guia-ruleta/tipos-de-ruleta/