TP 1.1 SIMULACION DE UNA RULETA

Leilen Avila
Legajo: 41610
Mail: leilenavila@gmail.com

Natalia Fernandez
Legajo: 44758
Mail: nata.fernandez77@gmail.com

April 11, 2020

ABSTRACT

El trabajo consiste en construir una programa en lenguaje Python que simule el funcionamiento del plato de una ruleta. Para esto se tuvo en cuenta la generación de valores aleatorios enteros, el uso de listas para el almacenamiento de datos, el uso de la estructura de control FOR para iterarlas, el empleo de funciones estadísticas y gráficas de los resultados mediante la librería Matplotlib.

1 Introduction

Se quiere simular una ruleta y poder observar su comportamiento. La ruleta se la considera ideal, ya que no se puede determinar el resultado que va a arrojar. Las ruletas, después de un gran número de repeticiones, demuestra una distribución regular de los números que fueron arrojados. En el presente trabajo simularemos el modelo de una ruleta convencional, es decir aquella que posee números desde el cero hasta el treinta y seis. En una experiencia aleatoria, cada resultado posible se conoce con el nombre de suceso, y al conjunto de todos los sucesos posibles se lo denomina espacio muestral S. En nuestro ejemplo S = 0,1,2,. . . .,36.

2 Descripcion del experimento

Partiendo de lo explicado en la introducción, cuándo se juega a tirar la ruleta se supone que las probabilidades de obtener cada uno de los posibles resultados es idéntica e igual a 1/37 (sucesos equiprobables). También comenzaremos diciendo que la variable de estudio es una variable aleatoria discreta ya que sólo puede tomar un número finito de valores dentro de un intervalo. Además sabiendo que los sucesos son equiprobables demostraremos que esta variable aleatoria presenta una distribución uniforme discreta. En nuestro experimento simularemos que giramos una ruleta un número n veces de repeticiones. Por cada repetición se generará un número aleatorio y para ello utilizamos la función randint() de la librería random. Esta función recibe dos números entre los cuales queremos que se genere un número aleatorio entero. Este procedimiento se repetirá entre 5 y 10 veces, generando así diferentes corridas del experimento. En nuestro estudio comenzaremos corriendo el programa para tomar 5 muestras de 1500 repeticiones independientes y evaluar los resultados obtenidos de los diferentes parámetros estadísticos. Los estadísticos permiten visualizar las características de la distribución de los datos ayudando a sintetizar la dimensión de cambio de una variable, y son conceptos básicos de estadística.

3 Distribución de probabilidad.

Una distribución de probabilidades para una variable aleatoria discreta es un listado mutuamente excluyente de todos los resultados numéricos posibles para esa variable aleatoria tal que una probabilidad específica de ocurrencia se asocia con cada resultado. La distribución de este experimento es una distribución uniforme discreta porque el resultado de la experiencia aleatoria puede ser un conjunto finito de 37 posibles resultados, todos ellos igualmente probables. Procederemos a demostrar la afirmación mencionada anteriormente, a través de un diagrama de barras que permite visualizar 5 corridas de una cantidad de repeticiones lo suficientemente grande. De estas corridas se anotó la frecuencia absoluta de cada número de la ruleta. 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. Mediante este experimento podemos observar una curva azul que representa el valor promedio de la frecuencia absoluta de un número r de la ruleta entre todas las corridas realizadas. Se observa que esta curva tiende a estabilizarse entre los valores de 260 y 280, este es un resultado muy exacto que demuestra que mientras más se incremente el número de repeticiones más tenderá esta curva a ser una recta perfecta. Esta recta será una constante que tendrá el siguiente valor:

$$\frac{numero\,repeticiones}{cant\,numeros\,ruleta} = \frac{numero\,repeticiones}{37} \tag{1}$$

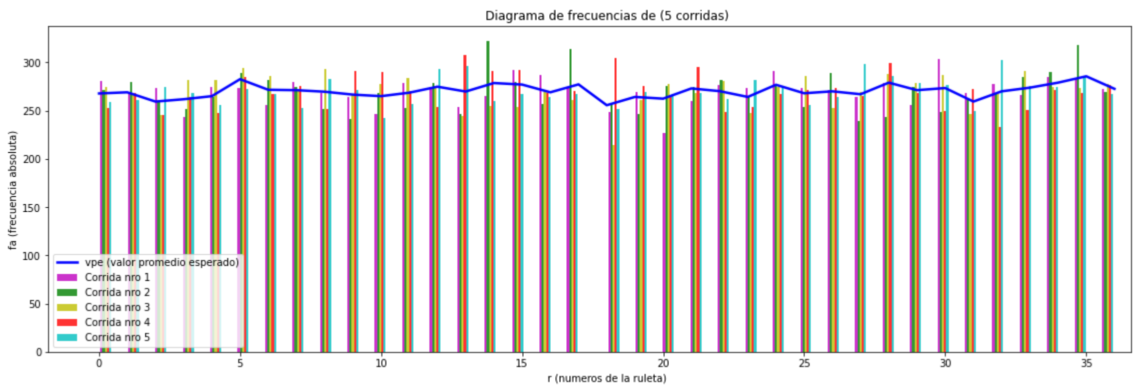

Figure 1: Diagrama de frecuencias de 5 corridas.

En nuestro ejemplo el número de repeticiones es de 10000 y por lo tanto esta constante, al incrementar el número de corridas, tenderá a tener un valor de 270. Para confirmar mejor esto experimentamos con un número de corridas igual a 20 y el resultado fue una curva mucho más suave que tiende al valor esperado.

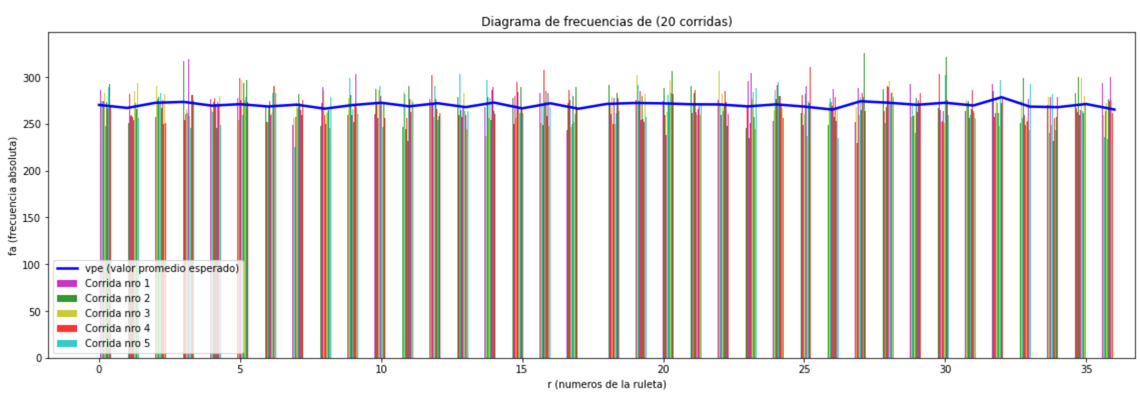

Figure 2: Diagrama de frecuencias de 20 corridas.

Luego dibujamos, para el sistema de ejes coordenados, sobre el eje horizontal cada uno de los casos y levantamos, para cada uno de los valores, una barra cuya altura será igual al valor promedio de todas las corridas de la variable en ese caso. En este diagrama de barras podremos notar la uniformidad de la distribución de una forma muy clara.

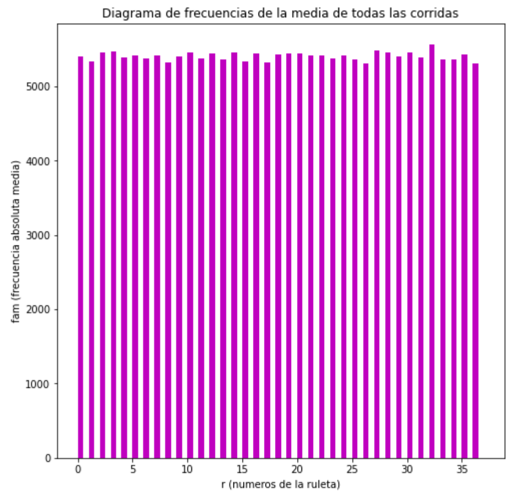

Figure 3: Diagrama de barras de promedio corridas.

Tomamos también una muestra aleatoria correspondiente a una sola corrida y decidimos evaluar su comportamiento mediante un diagrama de barras. Podemos visualizar que si bien no es tan exacta como la anterior, su comportamiento es muy aproximado variando entre las cantidades de 230 y 310.

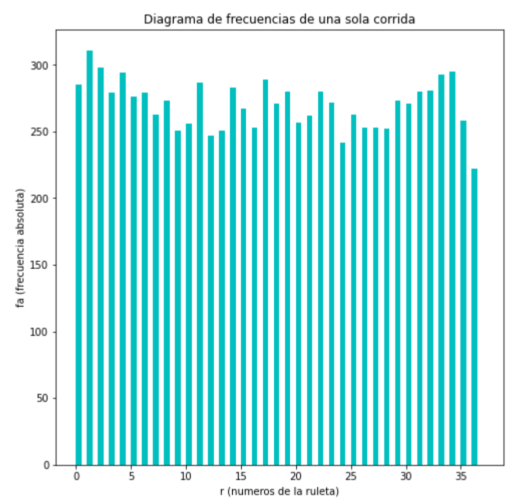

Figure 4: Diagrama de barras de una corrida.

4 Evaluación de la frecuencia relativa.

El concepto de aleatoriedad lleva implícitas dos ideas matemáticas: repetibilidad de un experimento aleatorio e independencia de los ensayos sucesivos. En el estudio de las frecuencias relativas se encuentran implícitas estas dos exigencias, por lo que su comprensión precisa la de un experimento aleatorio. Puede observarse empíricamente una estabilización gradual de las frecuencias relativas en una serie de repeticiones suficientemente largas. En la Figura 5 se puede observar cómo las distintas corridas oscilan entre los valores de 0,01 y 0,05. A partir de estas gráficas procedimos a calcular el promedio del valor de la frecuencia relativa para cada número n de tiradas y obtuvimos una curva de la frecuencia relativa esperada (representada en color azul).

Analiticamente sabemos que por ser un experimento aleatorio con una distribución uniforme discreta, la frecuencia relativa para un número de repeticiones suficientemente grande tiende a ser igual a la probabilidad de que ocurra dicho suceso, entonces la frecuencia relativa será igual a:

$$\textbf{frecuencia relativa} = \frac{nro\,casos\,favorables}{nro\,casos\,totales} = \frac{1}{37} \simeq 0.027 \tag{2}$$

Por lo tanto, el valor esperado de la frecuencia relativa, con un margen de error muy pequeño, coincide con el valor calculado analiticamente.

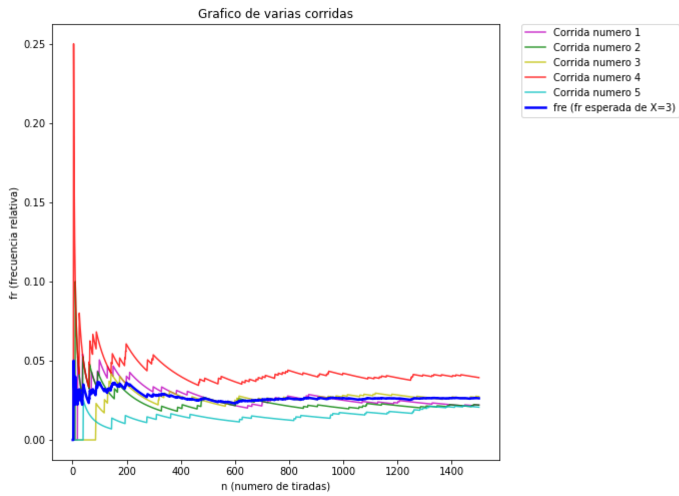

Figure 5: Grafico de varias corridas con respecto a la frecuencia relativa.

Utilizamos el valor de la frecuencia relativa esperada calculada en la gráfica anterior para compararla con una sola muestra(corrida en color rosa) obtenida al azar, para evaluar qué tan confiable es la misma. Como se ve gráficamente, el valor de una sola muestra de 1500 repeticiones es muy precisa, y tiende a estabilizarse alrededor de los valores esperados.

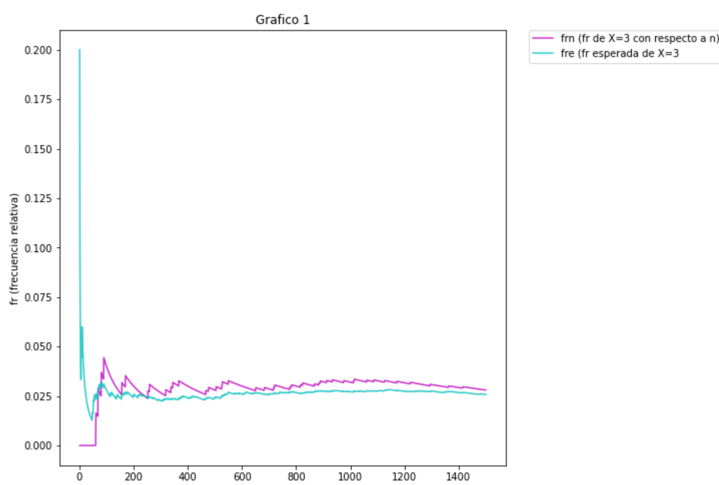

Figure 6: Grafico de una corrida elegida al azar con respecto a la frecuencia relativa.

5 Evaluación del valor promedio.

A continuación realizaremos la misma experimentación que en el apartado anterior pero para estimar el valor de la media aritmética. La media aritmética se define como la suma de todos los valores observados de la distribución, dividida por el número total de observaciones. Como sabemos que el experimento tiene una distribución uniforme discreta podemos calcular su media aritmética a través de la fórmula de la misma.

$$\mu = \frac{1}{n}\sum_{i}^{n} x_i = \frac{1}{37}\sum_{i}^{n} x_i = 18 \tag{3}$$

Para demostrar esto se procedió a tomar 5 muestras de 1500 repeticiones cada una (graficadas en diferentes colores). Luego calculamos el valor promedio de todas las muestras para cada número de tirada y se obtuvo una curva(en color azul) que representa la media de las medias aritméticas de cada muestra. Se puede ver cómo se estabilizan los valores, conforme crece el número de repeticiones, hacia el valor 18. Lo cual coincide exactamente con el valor calculado analiticamente.

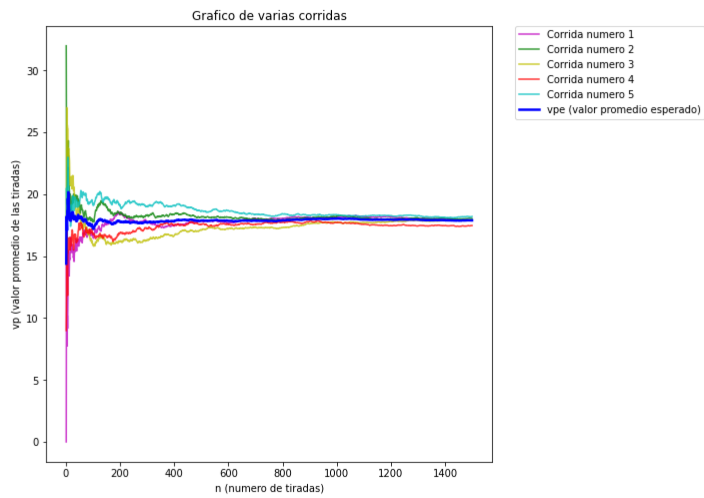

Figure 7: Grafico de varias corridas con respecto al valor de la media.

Utilizamos el valor de la media aritmética esperada calculada en la gráfica anterior para compararla con una sola muestra(corrida en color rosa) obtenida al azar, para evaluar qué tan confiable es la misma. Como se ve gráficamente, el valor de una sola muestra de 1500 repeticiones es muy precisa, y tiende a estabilizarse alrededor de los valores esperados.

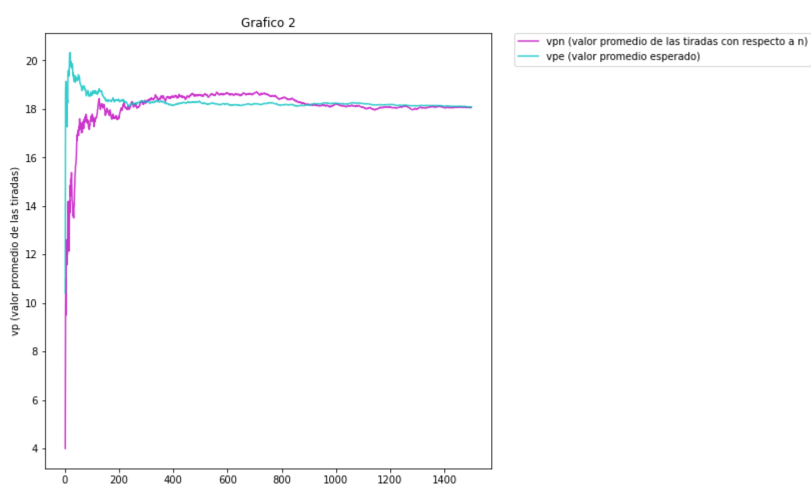

Figure 8: Grafico de una corrida con respecto al valor de la media.

6 Evaluación de la varianza y del desvío estándar.

La desviación estándar es la medida de dispersión más común, que indica qué tan dispersos están los datos con respecto a la media. Mientras mayor sea la desviación estándar, mayor será la dispersión de los datos. Su fórmula, para la distribución uniforme discreta es la siguiente:

$$\sigma = \sqrt{\frac{1}{n}\sum_{i}^{n}(x_i - \mu)^2} = 10,67 \tag{4}$$

Procedimos a hacer el mismo procedimiento que los apartados anteriores, obteniendo una curva del promedio del desvío estándar de 5 corridas y luego a partir de esta evaluar la exactitud del desvío estándar de una muestra aleatoria. Nuevamente puede verse que los resultados empíricos coinciden con los analiticos tendiendo la curva de la muestra(color rosa) al valor 10.67

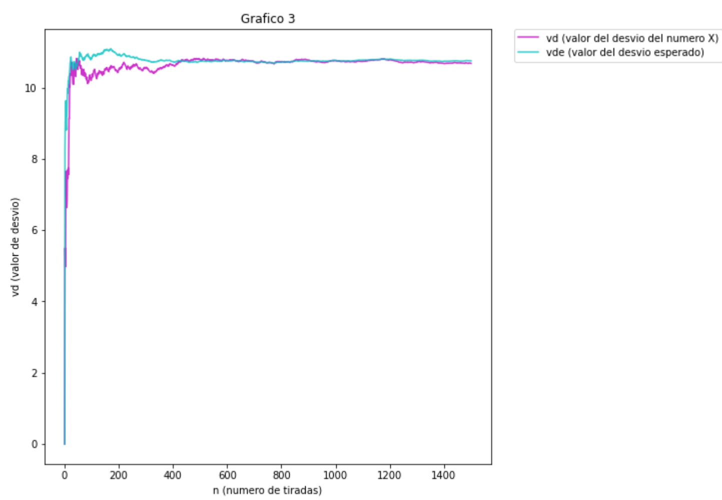

Figure 9: Grafico de una corrida elegida al azar con respecto al desvio estandar.

La unidad de medida de la varianza será siempre la unidad de medida correspondiente a los datos pero elevada al cuadrado. La varianza siempre es mayor o igual que cero. Al elevarse los residuos al cuadrado es matemáticamente imposible que la varianza salga negativa. Y de esa forma no puede ser menor que cero. Su fórmula es la siguiente:

$$\sigma^2 = \frac{1}{n}\sum_{i}^{n}(x_i - \mu)^2 = 114 \tag{5}$$

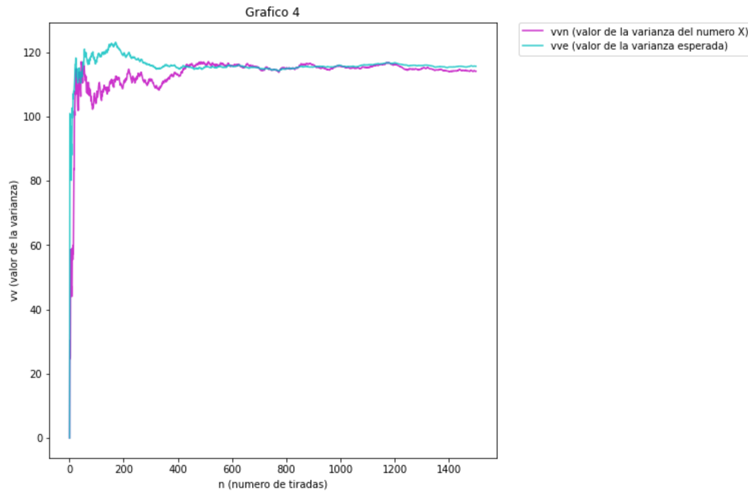

Figure 10: Grafico de varias corridas con respecto a la varianza.

7 Observación importante.

Dado que las muestras que estamos obteniendo presentan resultados contundentes nos comenzamos a preguntar cual tendrá que ser el mínimo valor de la cantidad de repeticiones para que se puedan seguir obteniendo estos resultados exactos. Cuando seleccionamos una muestra, esperamos que sea representativa de la población, y utilizaremos criterios que faciliten la realización del estudio concreto pero que no generen excesivas diferencias con la población objetivo, para poder generalizar los resultados. Por lo tanto para obtener el tamaño ideal de la muestra nos basaremos la fórmula que sirve para calcular el tamaño de muestra cuando se desconoce el tamaño de la población:

$$n = \frac{Z_a^2 * P^2}{e^2} = \frac{1,96^2 * 0,5^2}{0,05^2} = 384 \tag{6}$$

Donde:

- Z = nivel de confianza deseado
- P = probabilidad de éxito, o proporción esperada.
- e = error máximo admisible en términos de proporción

Dando un margen de error del 5% y con un nivel de confianza del 95% se obtiene que el tamaño de la muestra debería ser de 385 repeticiones.

Sin embargo, si se quiere tener un error del 3% y con un intervalo de confianza del 99% entonces se obtiene, según la fórmula, que la muestra debería ser del tamaño de 1844 repeticiones.

Generamos las gráficas de los diferentes parámetros estadísticos de una muestra aleatoria de 385 repeticiones para evaluar sus resultados. Tomando como referencia curva celeste que habíamos calculado en los apartados anteriores.

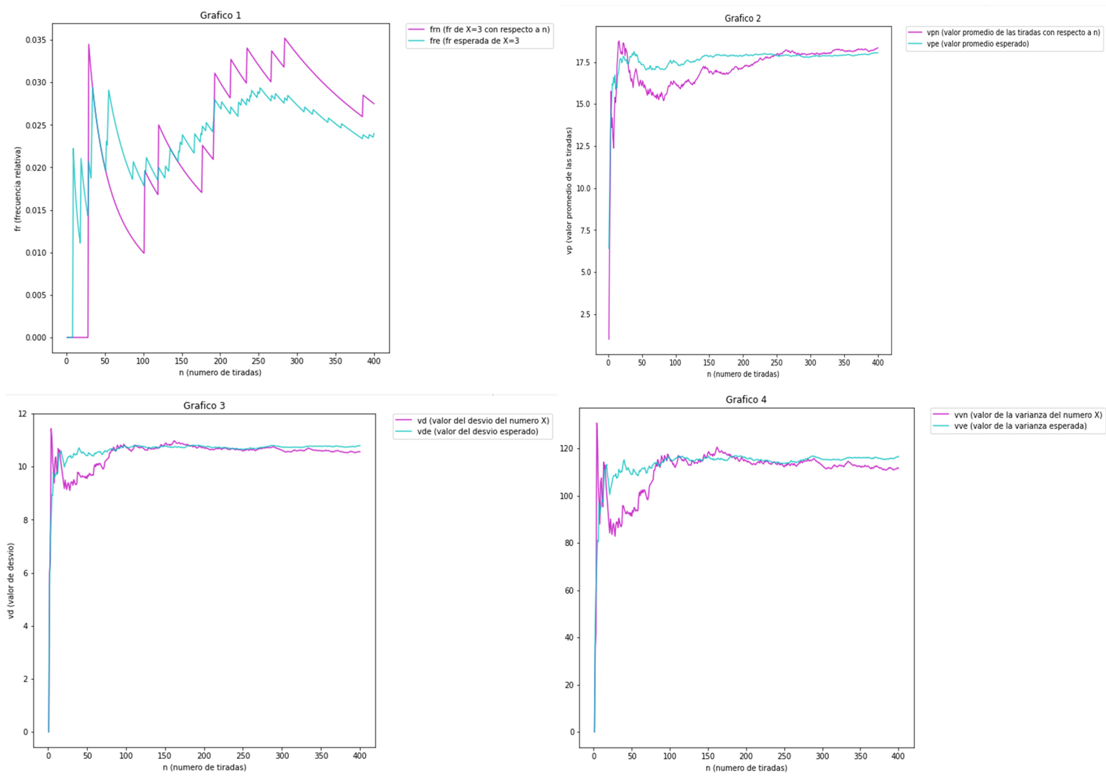

Figure 11: Parametros de una corrida elegida al azar de una muestra pegueña.

La representatividad es la característica más importante de una muestra estadística, y se define como la capacidad de un subconjunto de presentar iguales características a las del conjunto completo, por ende el muestreo adquiere todo su sentido en cuanto se garantice que las características que se quieren observar en la población quedan reflejadas adecuadamente en la muestra. Al respecto, una muestra será representativa de la población de la que fue tomada, si "la suma de sus características se aproxima al conjunto de características de la población". No obstante, las muestras no tienen que ser representativas en todos los aspectos, limitándose a las características importantes para los intereses reales del estudio. Por lo tanto podemos concluir que con un tamaño de la muestra muchísimo más pequeño que el que habíamos tomado inicialmente se pueden obtener resultados muy exactos y que son representativos de la población.

8 Conclusión.

En nuestro estudio hemos identificado un conjunto de elementos de significado diferenciados en el estudio de la frecuencia relativa que permite destacar la complejidad de comprensión del mismo frente a su aparente simplicidad. Asimismo hemos descrito un conjunto de situaciones problemáticas asociadas al concepto de muestreo en el que encontramos la muestra ideal para obtener resultados contundentes. Encontramos formas de analizar los diferentes estadisticos del espacio muestral S para determinar los comportamientos de cada corrida, y mediante estos poder predecir con solo una muestra el comportamiento del juego de la ruleta. Pudimos tambien contranstar la experiencia con la teoria y concluir así, que los resultados eran los adecuados. Todo esto fue gracias a que la simulación nos permite entender el funcionamiento de sistemas de la vida real, aplicando tecnología y esto sumado a nuestro conocimientos de probabilidad y estadistica permiten obtener resultados optimos. Esto sirvió para formular teorías y confirmarlas, prediciendo la forma de comportamiento de un sistema. Como se pudo analizar en este trabajo, a medida que se incrementó el número de observaciones de las muestras, los valores estadísticos fueron acercándose más a los valores reales, calculados previamente.

References

[1] Distribucion Uniforme Discreta https://prezi.com/ebfpilzc2iki/distribucion-uniforme-discreta/

[2] MatPlotLib http://research.iac.es/sieinvens/python-course/source/matplotlib.html

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

[4] Probabilidad y estadistica https://relopezbriega.github.io/blog/2015/06/27/probabilidad-y-estadistica-con-python/