Modelado de la ruleta - Simulación

Gianfranco Scaldini
Cátedra Simulación
UTN - FRRO
Zeballos 1341, S2000
gianscaldinis6@gmail.com

Luciano Vanelli
Cátedra Simulación
UTN - FRRO
Zeballos 1341, S2000
lucianovannelli@gmail.com

Manuel Ángel Bahamonde
Cátedra Simulación
UTN - FRRO
Zeballos 1341, S2000
manuelbahamonde6@gmail.com

Ramiro Sammataro
Cátedra Simulación
UTN - FRRO
Zeballos 1341, S2000
ramasammataro@gmail.com

30 de marzo de 2021

Abstract

El siguiente documento tiene por objetivo que el lector pueda comprender el comportamiento de un juego de azar muy conocido, la ruleta. Para alcanzar este meta utilizaremos la estadística.

Enunciado

Realizar la construcción de un programa en lenguaje Python 3.x. Dicha construcción debe simular el funcionamiento del plato de una ruleta. Para esto, se debe tener en cuenta lo siguientes temas a investigar:

- Generación de valores aleatorios enteros.
- Uso de listas para el almacenamiento de datos.
- Uso de la estructura de control FOR para iterar las listas.
- Empleo de funciones estadísticas.
- Gráficas de los resultados mediante el paquete Matplotlib.

1. Muestra

Se utilizará para realizar el estudio una ruleta tradicional ideal que posee 37 números, que están distribuidos desde el 0 hasta el 36 sin repetición. Teniendo estos datos, podemos afirmar que la probabilidad de ocurrencia de un número sigue una distribución discreta uniforme[1].
Las variables que utilizaremos en nuestro estudio serán:

- Media.
- Frecuencia relativa.
- Desvio Estandar
- Varianza

2. Metodología

Dividiremos este estudio en dos estudios:

- Primer estudio: estudiaremos una réplica de 5000 tiradas
- Segundo estudio: estudiaremos 5 réplicas de 5000 tiradas cada una

Para ambas partes, podemos calcular las variables de probabilidad analíticamente (ya que la distribución es discreta uniforme). Procedemos a calcularlos con sus fórmulas definidas por su distribución

$$Promedio: \frac{a+b}{2} = \frac{02+36}{2} = 18 \qquad (1)$$

[1] https://es.wikipedia.org/wiki/Distribución_uniforme_discreta

$$Varianza: \frac{(b-a+1)^2-1}{12} = \frac{(36-0+1)^2-1}{12} \quad (2)$$

=114

$$Desvio: \sqrt{\frac{(b-a+1)^2-1}{12}} = \sqrt{\frac{(36-0+1)^2-1}{12}} = \quad (3)$$

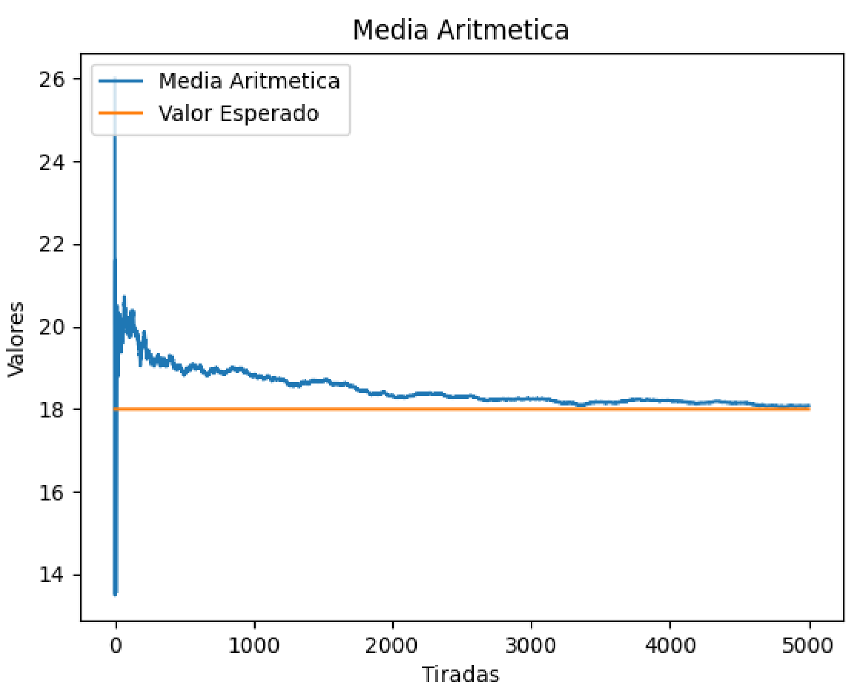
=10,67707...

Considerando "n" como la cantidad de posibles resultados que tiene la ruleta, la expresión que determina la frecuencia relativa queda:

$$\frac{1}{n} = \frac{1}{37} = 0,027027... \quad (4)$$

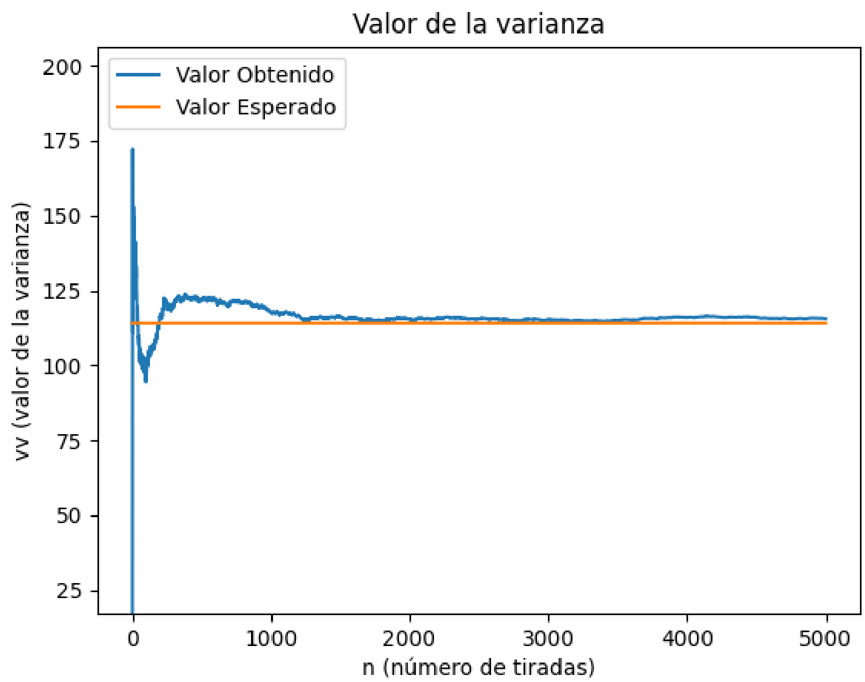
3. Primer estudio

3.1. Media



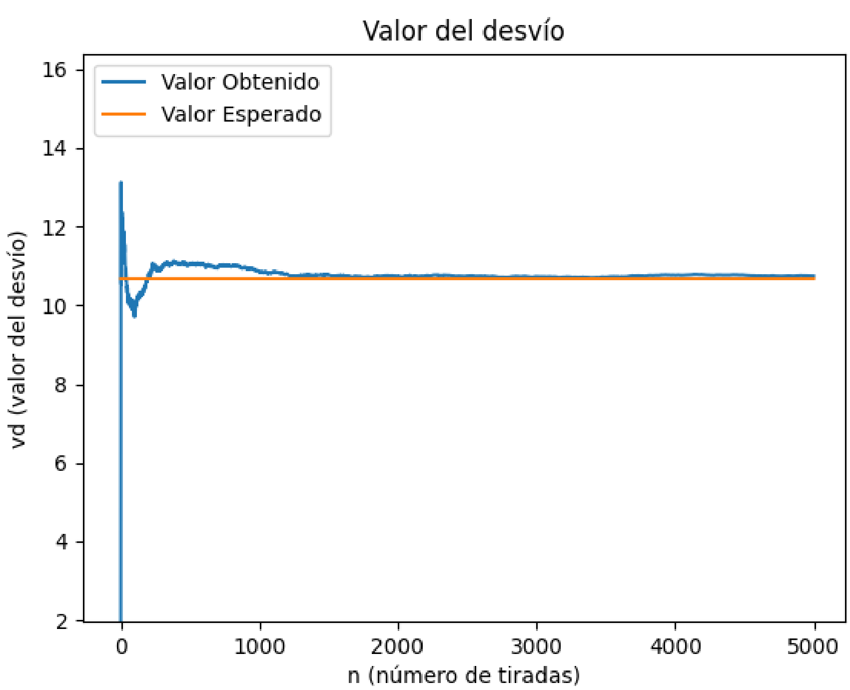
Al comienzo, vemos una dispersión muy fuerte del promedio, lo cual era esperable debido a la aleatoriedad de la ruleta, pero una vez que va aumentando el número de tiradas, el promedio se acerca al valor esperado (18)

3.2. Varianza



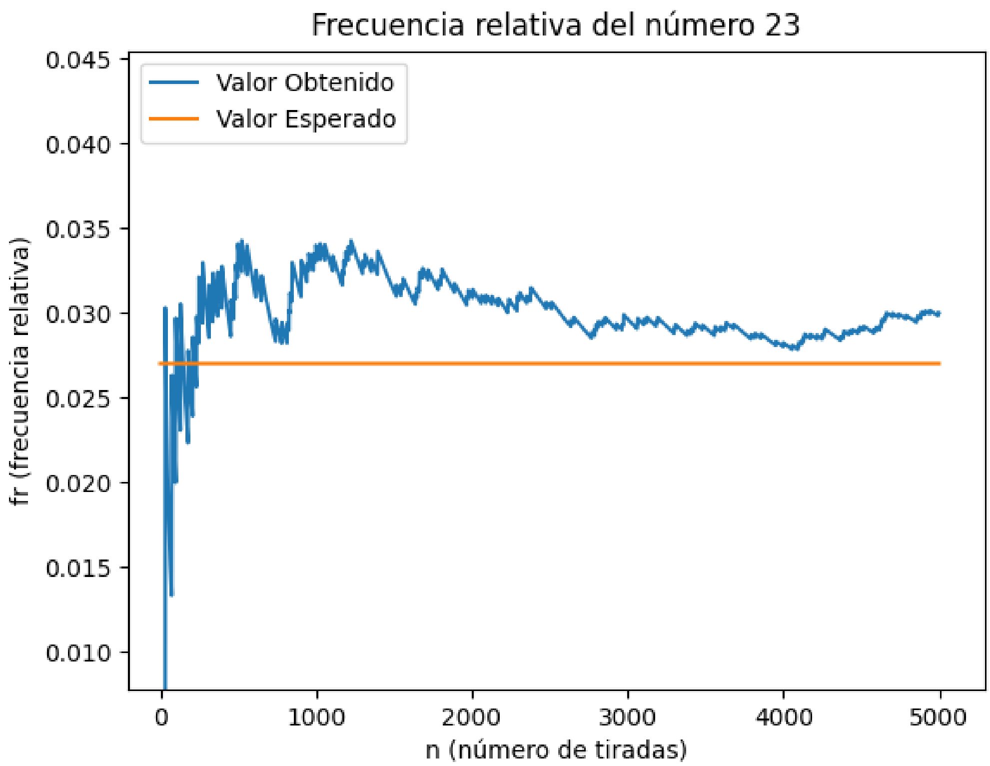
Al principio, vemos que pasa lo mismo que con el promedio: encontramos alta dispersión, pero luego se va acercando al valor esperado.

3.3. Desvío estandar

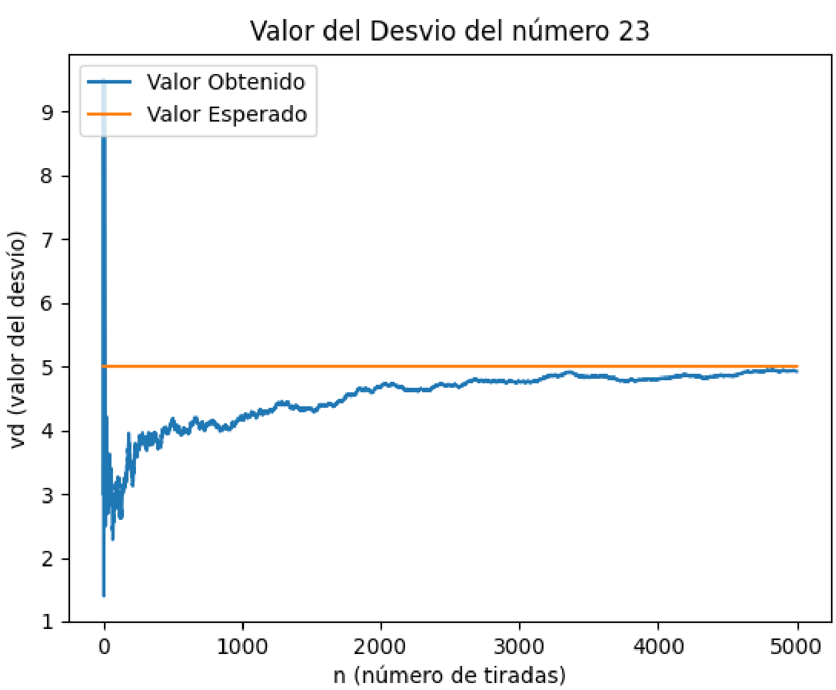


Al ser la raíz cuadrada de la varianza, vemos un comportamiento muy similar al de la misma.

3.4. Frecuencia relativa y desvío con respecto al 23



Observamos que al principio, la frecuencia relativa del 23 es 0, hasta que este sale por primera vez. A partir de ese momento, la frecuencia relativa empieza a aumentar hasta alcanzar el valor esperado. Para comprender un poco mejor este comportamiento, podemos utilizar el desvío o la varianza de la muestra.



Observamos que al tender a infinito la cantidad de tiradas, el desvío tiende a 5 (diferencia entre 18 [media] y 23 [número a estudiar])

4. Segundo estudio

Como nombramos anteriormente, el segundo caso de estudio será de 5 réplicas de 5000 tiradas cada una. Analizaremos distintos parámetros con los 100, 500, 1000 y 5000 primeros individuos. Con esto, buscamos demostrar que es lo que ocurre con cada parámetro al ir aumentando el tamaño de la muestra

4.1. Valor medio

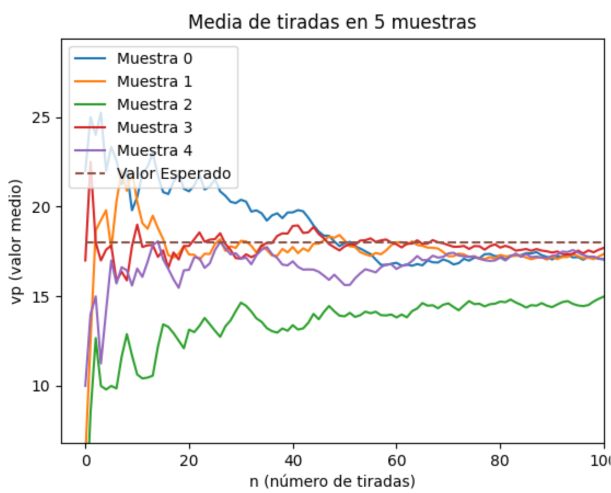
Con respecto al valor medio, como primer paso calcularemos el valor de la esperanza poblacional.

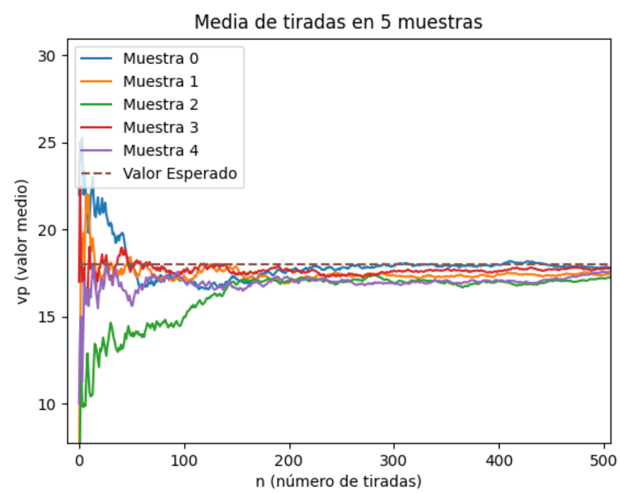
$$E(X) = \frac{\sum_{i=0}^{36} x_i}{37} = 18 \tag{5}$$

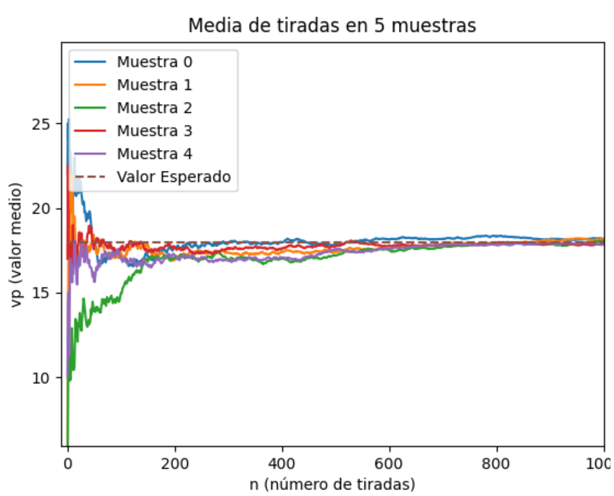
Definimos a el valor medio de una muestra con respecto a su tamaño como: $\bar{x}$

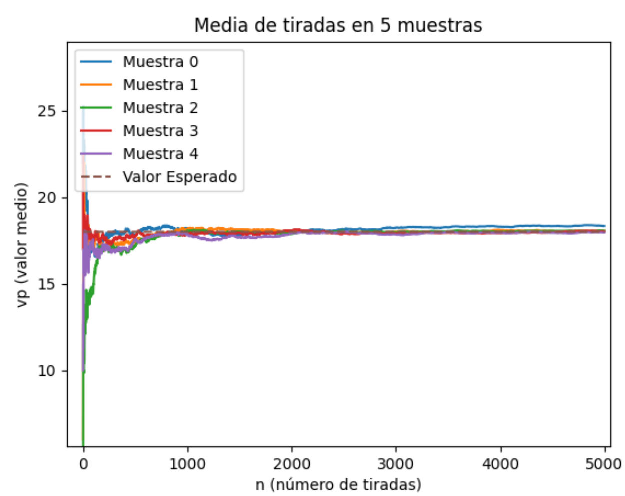
A continuación presentaremos gráficos en los cuales vamos a ver que a medida que la muestra aumenta de tamaño, el valor medio de la muestra va convergiendo hacia un valor puntual, E(X), esto se debe a que para muestras de un tamaño suficiente "n", está demostrado que:

$$\lim_{n \to \infty} f(\bar{x}) = E(X) \tag{6}$$









4.2. Frecuencia Relativa

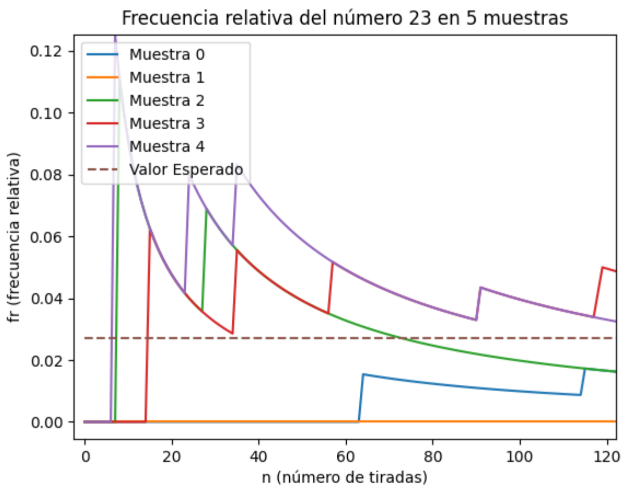
Primero buscaremos el valor de la probabilidad para que X=x, siendo m el número de casos posibles:

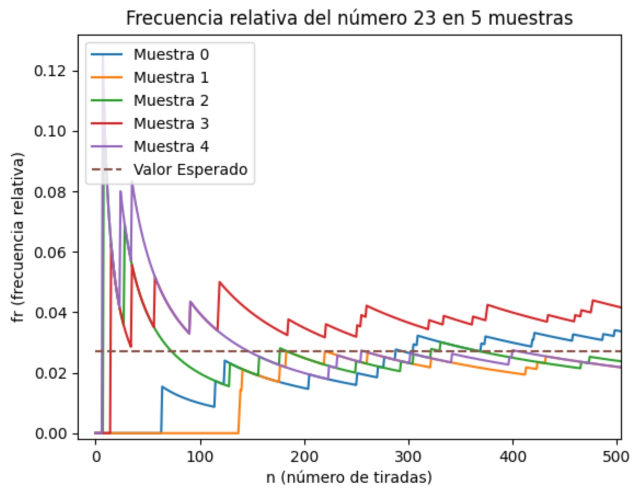
$$P(X = n) = \frac{1}{m} = \frac{1}{37} = 0,27 \tag{7}$$

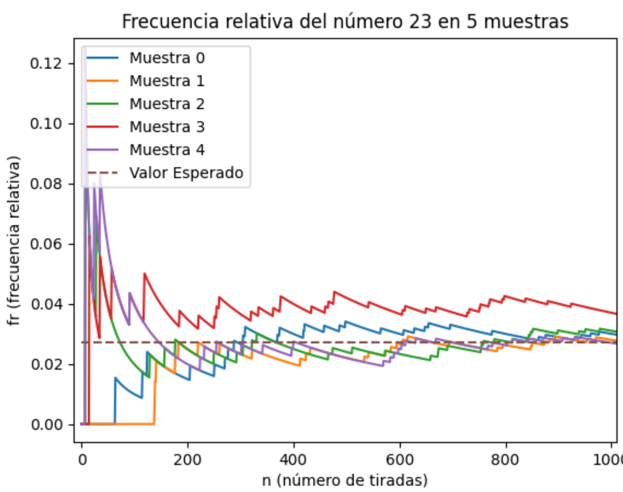
Definimos la frecuencia relativa como: $\mathrm{f}_{ri}(n)$

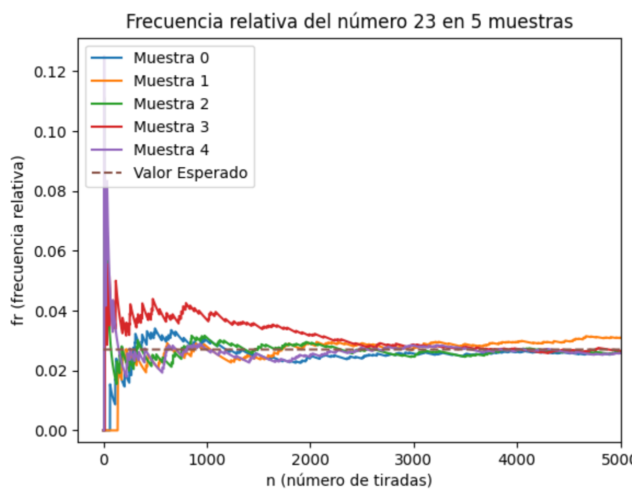
Ahora tomamos un x cualquiera de los valores de la ruleta y calculamos la frecuencia relativa conforme aumenta el tamaño de la muestra. En los gráficos siguientes se muestra la siguiente afirmación:

$$\lim_{n\to\infty} f_{ri}(n) = P(X = n) \tag{8}$$









4.3. Parámetro de Dispersión

Primero buscamos el valor de la Desviación Estándar:

$$D(X) = \sqrt{\frac{\sum_{i=0}^{36}(x_i - E(X))^2}{37}} = 10,68 \tag{9}$$

Luego definiremos al coeficiente de dispersión de una muestra con respecto a su tamaño como: S

Al igual que en el valor medio, podemos observar que el coeficiente de dispersión de la muestra converge hacia el valor de la desviación estándar.

$$\lim_{n\to\infty} S = D(X) \tag{10}$$

4.4. Distribución Muestral

Para continuar el análisis formaremos cuatro distribuciones muestrales con los promedios de 5 muestras de 5000 individuos cada una:
EspacioMuestral $= \sqsubset \bar{x}_1, \bar{x}_2, ..., \bar{x}_{37} \sqsupset$
$||\, EspacioMuestral \,|| = 37$

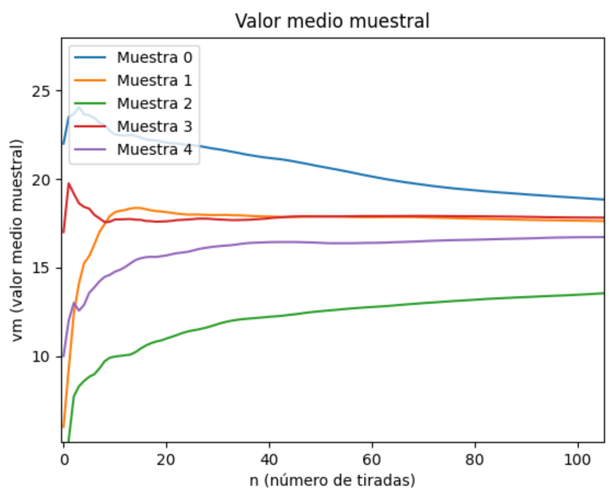
De este modo definimos la variable:
$\bar{X}$: *Promedio de una muestra de 5000*

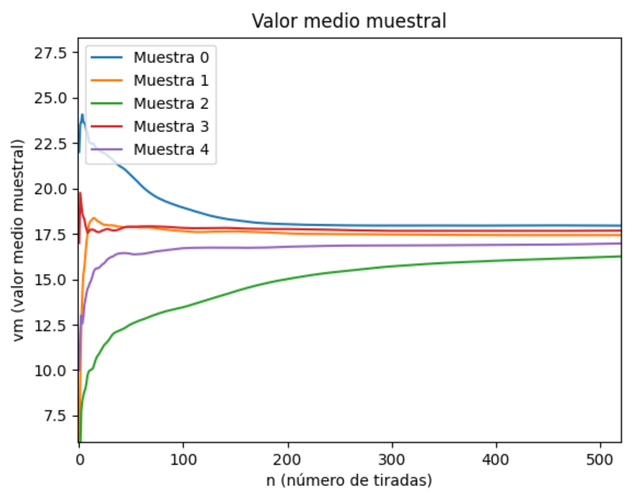
Esta variable es una distribución muestral inferida a partir de las 5 muestras antes mencionada.

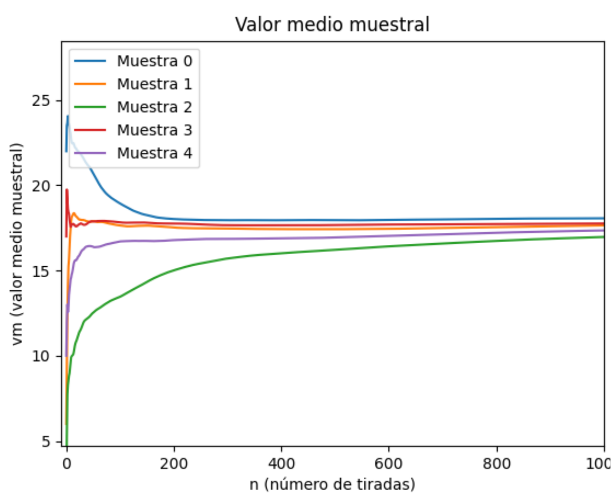
De cada muestra deberemos su valor medio para así poder calcular la Esperanza y la Desviación Estándar de esta variable.

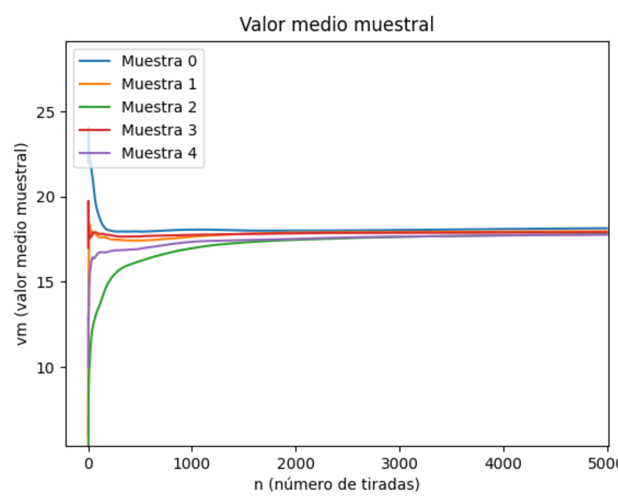
4.5. Valor Medio Muestral

El valor medio muestral lo definimos como el valor medio de los valores medios de las muestras, y vamos a observar que este valor tiende a asemejarse a la Esperanza, al igual que el valor medio común. La esperanza del Promedio Muestral es el promedio la las Esperanzas de cada población, en este caso como se trabaja con la misma población en todos los casos, la Esperanza Muestral y la Esperanza de la población son iguales. Lo característico es que lo realiza de forma mucho más acelerada que este, como se puede apreciar en los siguientes gráficos.







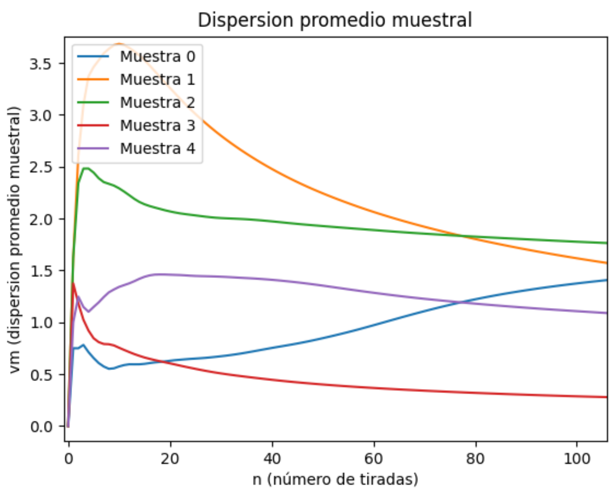


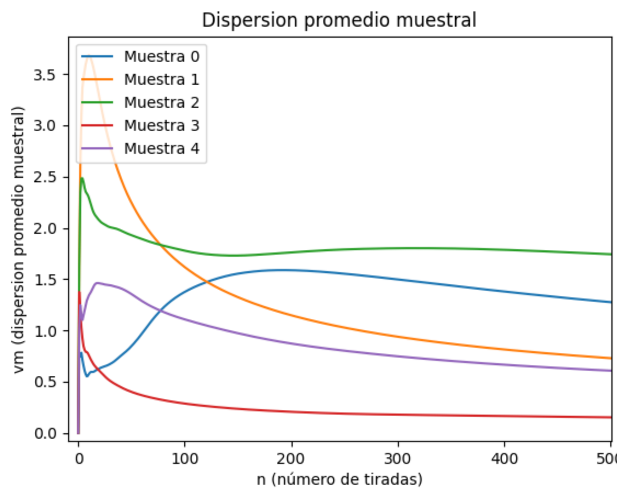
4.6. Dispersión Muestral

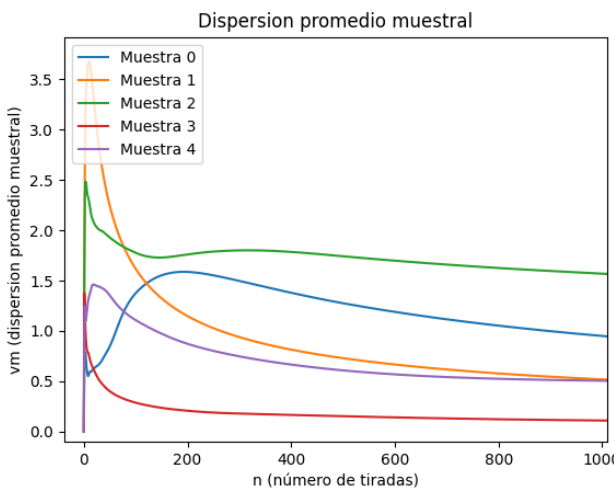
Ahora vamos a ver el motivo por el cual el valor medio muestral se aproxima de forma más repentina a la Esperanza. La razón de esto es que la varianza muestral es la suma de las varianzas dividido la cantidad de muestras que estamos trabajando. Se puede deducir claramente que cuanto más muestras trabajemos, más exacta va a ser nuestra inferencia y más va a reducirse la dispersión. Eventualmente, la dispersión tenderá a 0, conforme aumente el tamaño de las muestras o la cantidad de muestras:

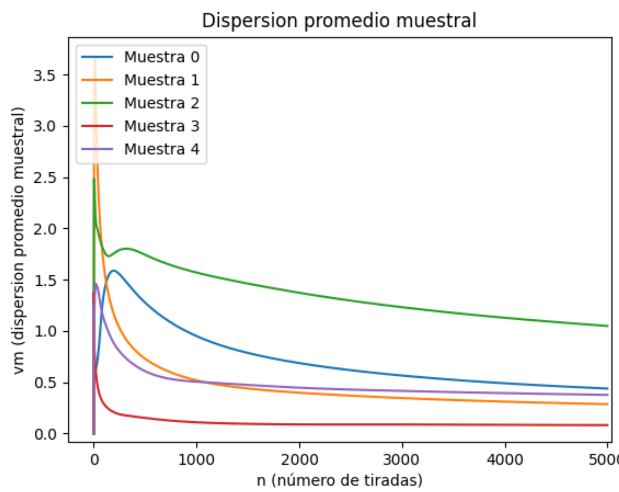
$$\lim_{m \vee n \to \infty} \sqrt{\frac{\sum_{i=1}^{35}(\bar{x}_i - E(\bar{X}))^2}{n}} = 0 \qquad (11)$$

La conclusión es que a medida que se reduce la dispersión, más exacta será nuestra estimación de la Esperanza del promedio muestral.









5. Conclusiones generales

Ahora que presentamos ambos casos de estudio, podemos llegar a una conclusión sobre la cantidad de observaciones del experimento. Si bien lo realizamos con 5000 tiradas, nos damos cuenta de que entre las tiradas 1500 y 3000 ya los resultados empiezan a repetirse. Entonces consideramos que sería ideal realizar los experimentos con esas cantidades antes que con 5000. Esto optimizaría el experimento.

Referencias

[1] Documentación matplotlib: `https://matplotlib.org/stable/tutorials/introductory/pyplot.html`

[2] Documentación Python: `https://docs.python.org/3/`

[3] Repaso de conceptos y fórmulas de probabilidad: `https://es.khanacademy.org/math/statistics-probability`