TOOLKIT DE MUESTREO Y TEOREMA DEL LÍMITE CENTRAL

EQUIPO:

- Alonso Corral Martha Julieta
- Leal Salguero Ari Iván
- Jiménez Cornejo Emanuel Adrián
- Diaz Garduño José Angel
- Franco Guzmán Alberto

INSTITUTO POLITÉCNICO NACIONAL

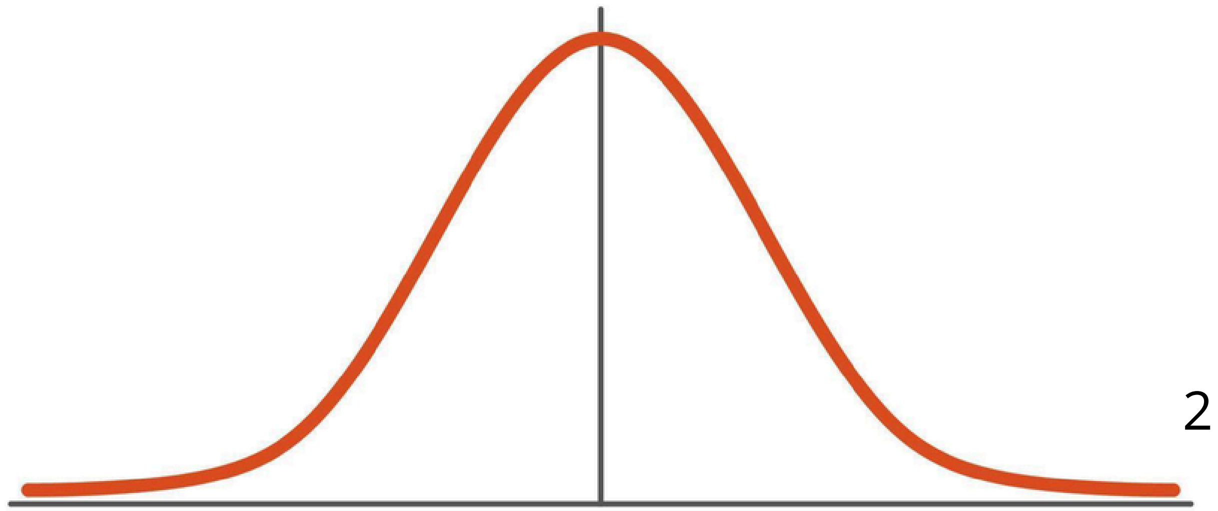
ESCOM

INTRODUCCIÓN

Es un simulador que toma poblaciones "desordenadas" y demuestra cómo, al promediarlas, se vuelven ordenadas y predecibles.

- **Objetivo:** Desarrollar un software interactivo que demuestre la convergencia a la normalidad de diversas distribuciones base.

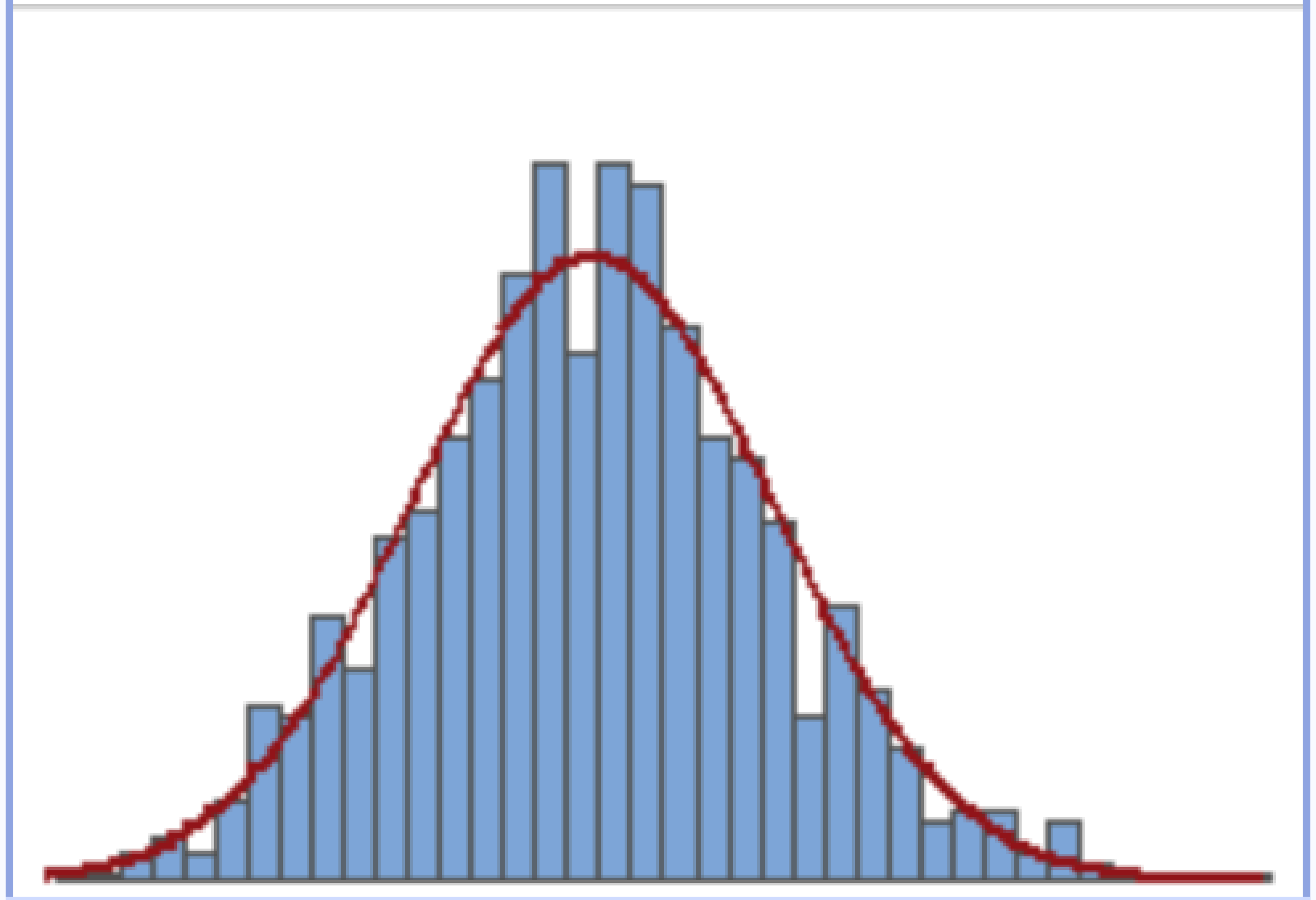
- **Reto:** Demostrar que no importa de dónde vengan los datos, si tomamos suficientes muestras, siempre terminaremos en una campana de Gauss.



TEOREMA DEL LÍMITE CENTRAL

Establece que si se toman muestras suficientemente grandes de una población, las medias de las muestras tendrán una distribución normal , incluso si la población no tiene una distribución normal.

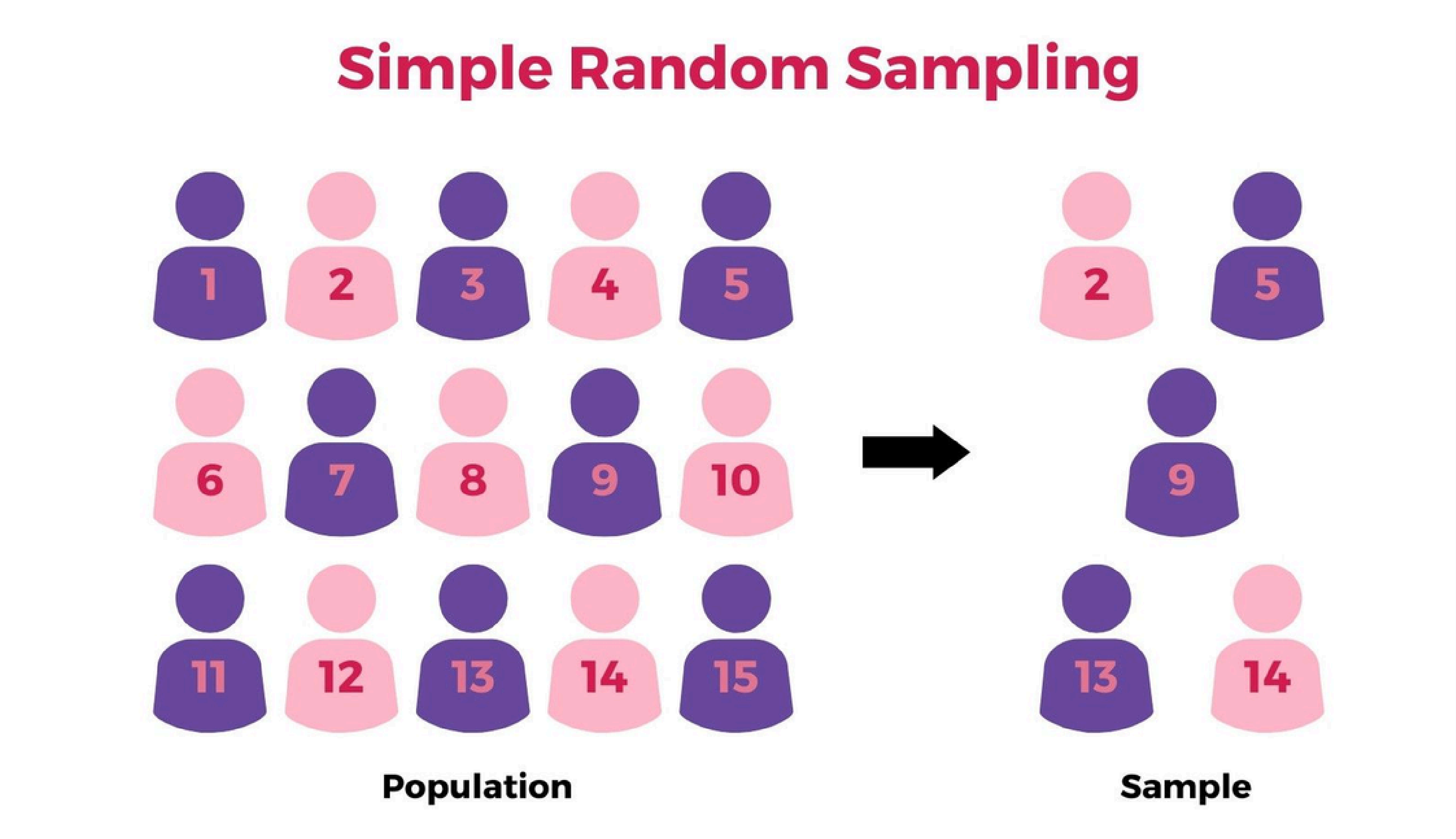
$$\bar{X}_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$



Medias de las muestras

MODELADO ESTADÍSTICO.

El modelado estadístico de nuestra aplicación se divide en tres componentes críticos: la definición de las poblaciones base, el proceso de muestreo aleatorio y la caracterización de la distribución resultante.



DISTRIBUCIONES DE POBLACIÓN (ENTRADA).

Para demostrar que el TLC funciona independientemente de la forma original, modelamos tres escenarios:

- Distribución Uniforme (U[0, 1]): Representa procesos equiprobables. Parámetros teóricos: $\mu = 0.5$ y $\sigma^2 = 1/12 \approx 0.0833$.
- Distribución Exponencial ($\lambda = 1$): Crucial para modelar tiempos de espera. Es una distribución no simétrica con $\mu = 1$ y $\sigma^2 = 1$. Su asimetría representa el reto principal para la validación visual del TLC.
- Distribución Bernoulli ($p = 0.5$): Modela eventos dicotómicos (éxito/fracaso). Permite observar cómo el promedio de valores discretos converge a una variable continua.

EL TEOREMA DEL LÍMITE CENTRAL (PROCESO):

MEDIA MUESTRAL

Es el promedio de un subconjunto de una muestra de datos extraído de una población más grande.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

VARIANZA MUESTRAL

Es una medida de dispersión que representa la variabilidad de una serie de datos con respecto a su media.

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

ERROR ESTÁNDAR

Estima qué tanto se alejaría la media de tu muestra de la verdadera media de toda la población si repitieras el experimento muchas veces.

$$\frac{s}{\sqrt{n}}$$

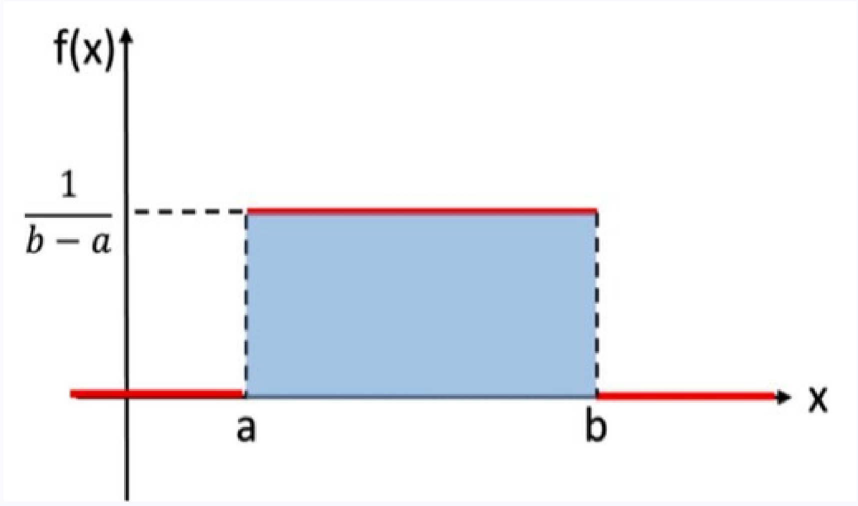
MONTE CARLO

Tipo de algoritmo computacional que utiliza muestreo aleatorio repetido para obtener la probabilidad de que se produzca un rango de resultados.

$$\bar{X}_n \to \mu$$
$$n \to \infty$$

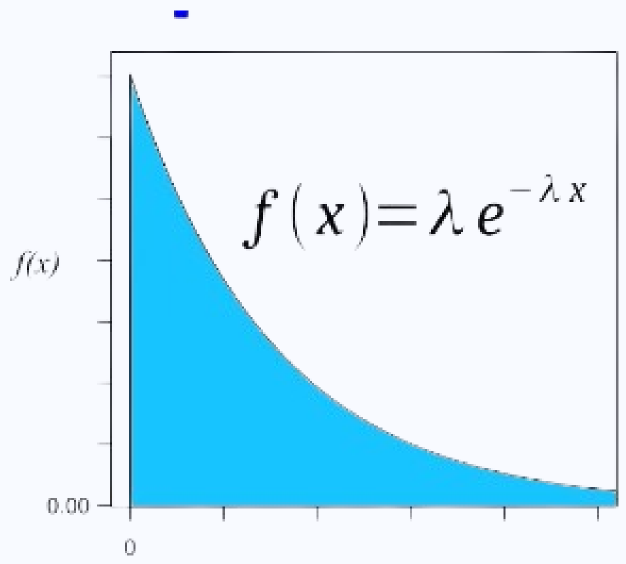
UNIFORME

Probabilidad constante en un rango (Escenario base)



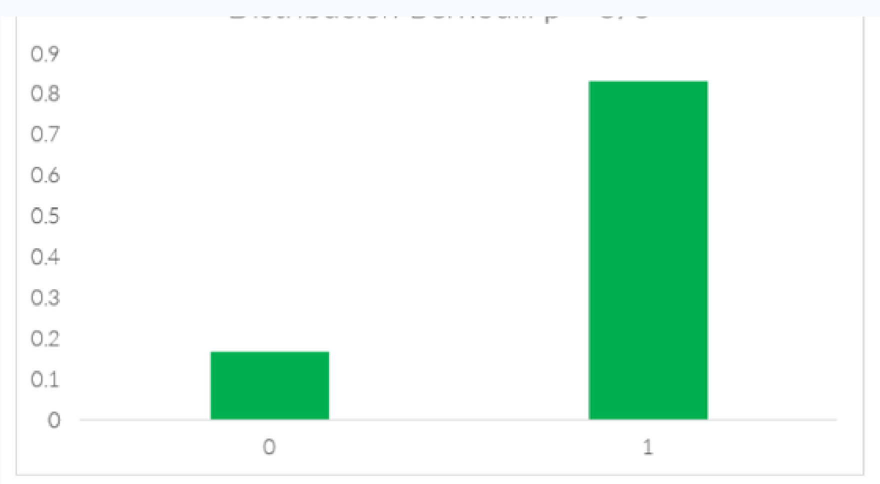
EXPONENCIAL

Modela tiempos de espera (Escenario con asimetría).



BERNOULLI

Eventos dicotómicos de éxito o fracaso (Escenario discreto).



LENGUAJE Y LIBRERÍAS

Lenguaje: Python 3.10.

Framework: Streamlit para la interfaz de usuario reactiva.

Procesamiento: NumPy para cálculos vectorizados de alta velocidad.

Visualización: Matplotlib y SciPy Stats para el renderizado de densidades teóricas.

TUTORIAL DE USO

01 Crear un nuevo archivo.

02 Abrir una terminal

03 Ejecutar python -m venv env en terminal y dar enter

04 Ejecutar pip install streamlit numpy matplotlib scipy en la terminal

05 Esperar a que lo descargue

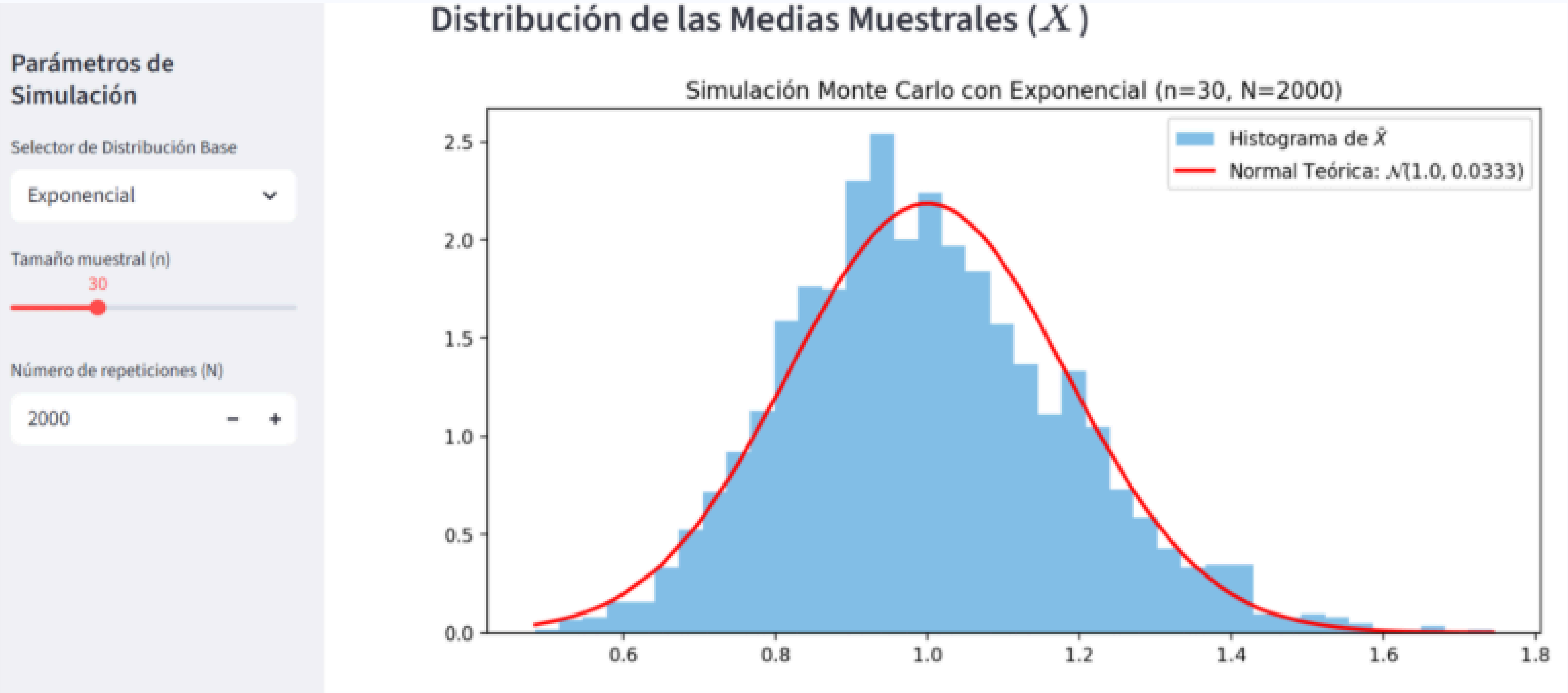
06 Guarda el archivo en una carpeta

07 Para ejecutar el programa ejecuta streamlit run <nombre de archivo>.py

08 De primera vez pedirá correo escribir correo electrónico

FUNCIONAMIENTO

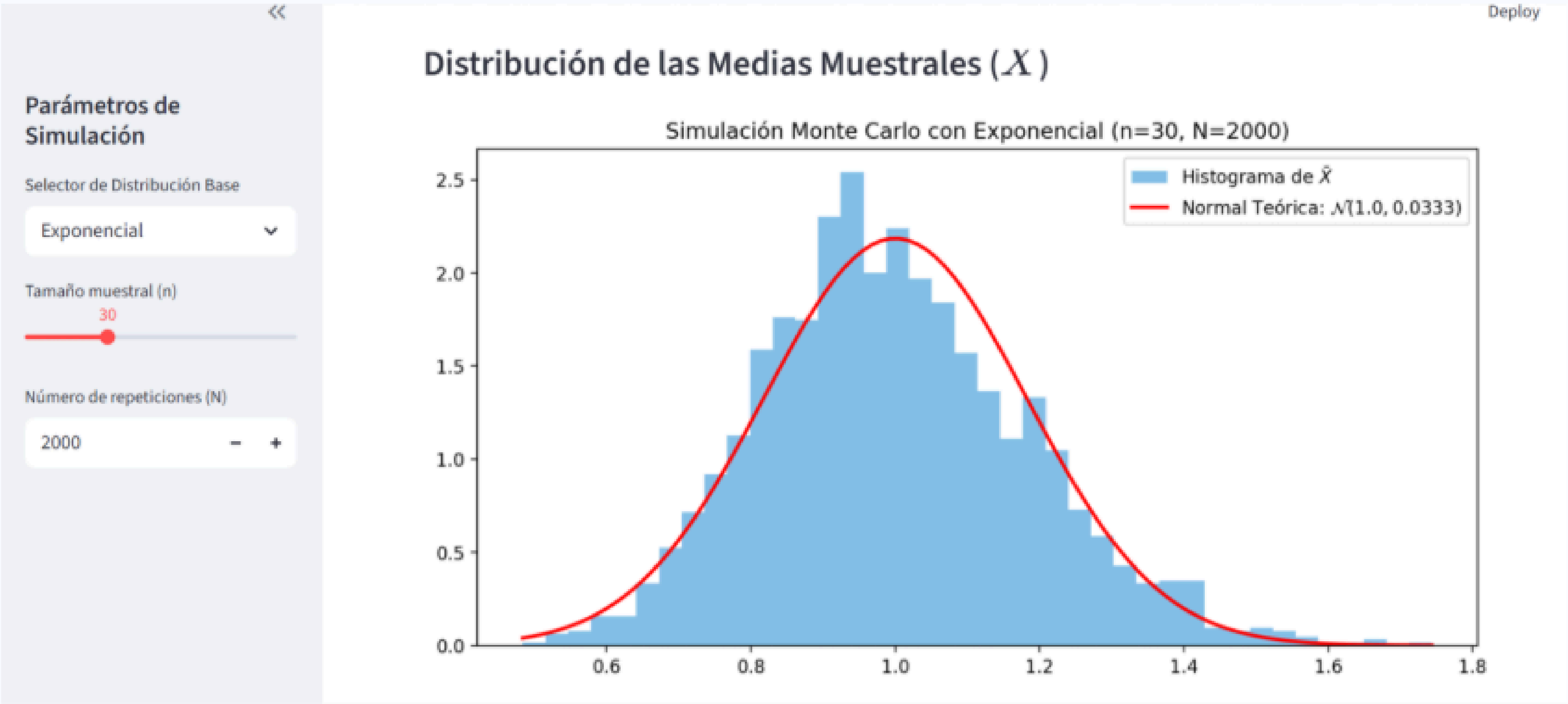
Entrada (Input): Selección de distribución, tamaño de muestra **(n)** y número de experimentos **(N)**.
Procesamiento: Muestreo aleatorio iterativo y cálculo de estadísticos descriptivos.
Salida (Output): Histograma de frecuencias relativas superpuesto con la curva normal teórica.



I. VALIDACIÓN DE CONVERGENCIA (HISTOGRAMAS)

PRUEBA CON N=30

PRUEBA CON N=2000



OBSERVACION

A pesar de que la distribución exponencial original es altamente sesgada, el histograma de las medias muestrales (azul) se ajusta casi perfectamente a la curva Normal teórica

2. VALIDACIÓN NUMÉRICA

El programa calcula la diferencia entre la teoría y la simulación.

Ejemplo de ejecución:

Media de $\bar{X}$: 0.9982 (Teórica: 1.0000).

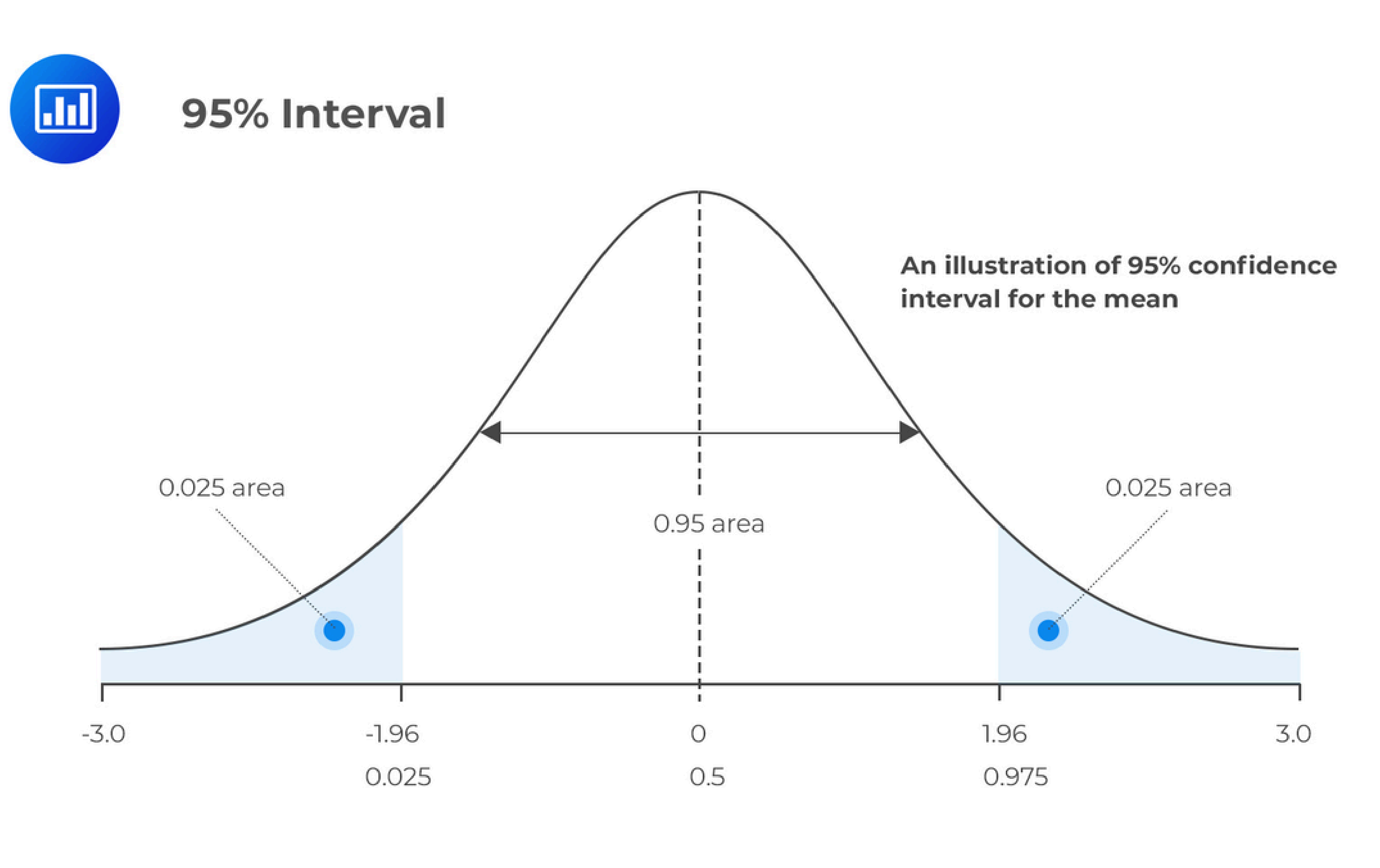
Varianza de $\bar{X}$: 0.0341 (Teórica: 1/30 = 0.0333).

La cercanía entre estos valores confirma que $\bar{X}$ es un estimador insesgado y consistente.

INTERVALO DE CONFIANZA

Es un rango de valores que, con un cierto nivel de probabilidad (confianza), contiene el verdadero valor de un parámetro poblacional desconocido (como la media) estimado a partir de datos de una muestra, ayudando a cuantificar la incertidumbre de la estimación puntual.

Se expresa como un límite inferior y superior, y su nivel (ej., 95%) indica cuántas veces el intervalo capturaría el valor real si se repitiera el muestreo muchas veces, ofreciendo una medida de precisión para la estimación

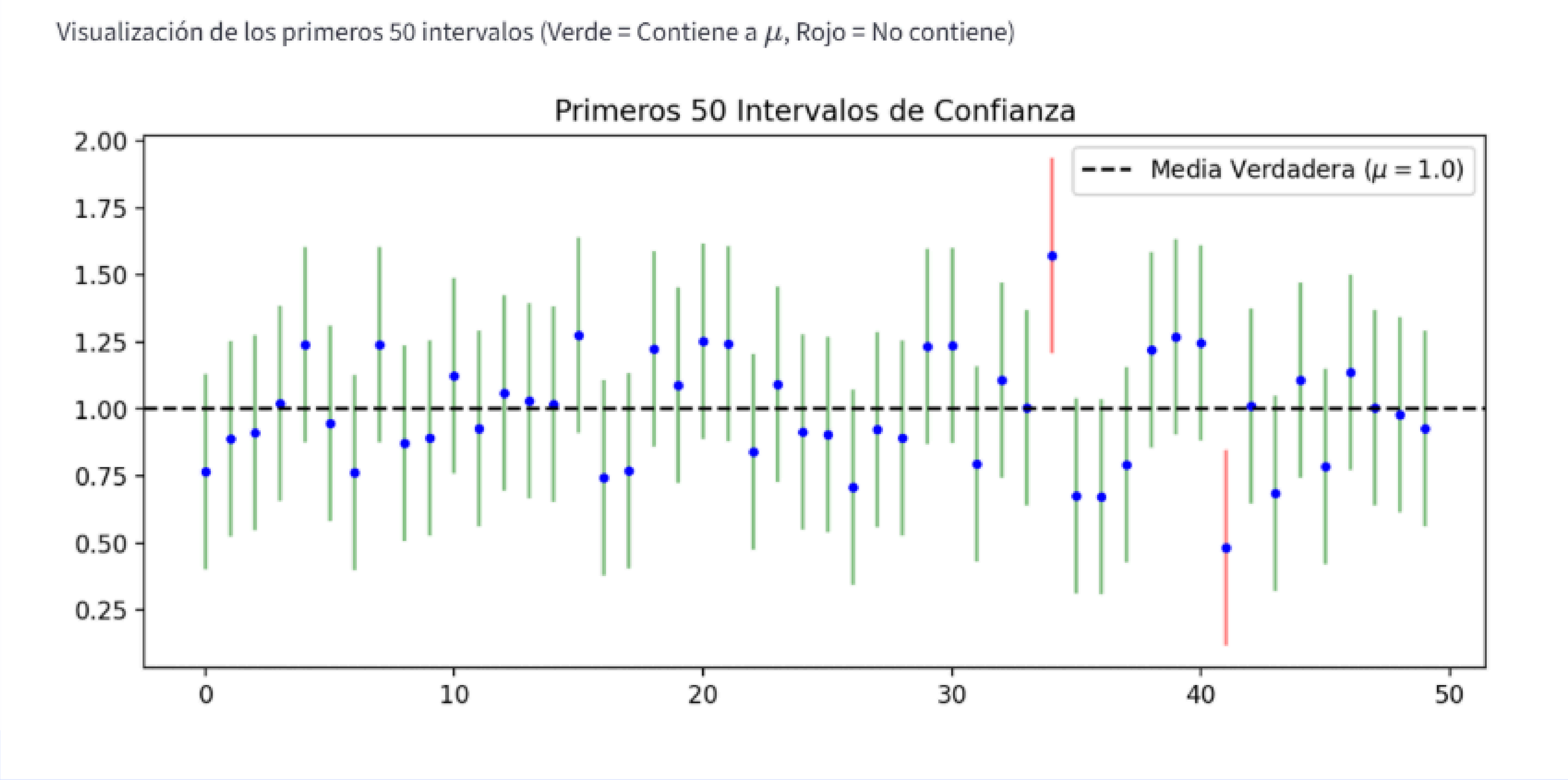


3. COBERTURA DE INTERVALOS DE CONFIANZA

Se implementó una visualización visual para validar la cobertura del 95%.

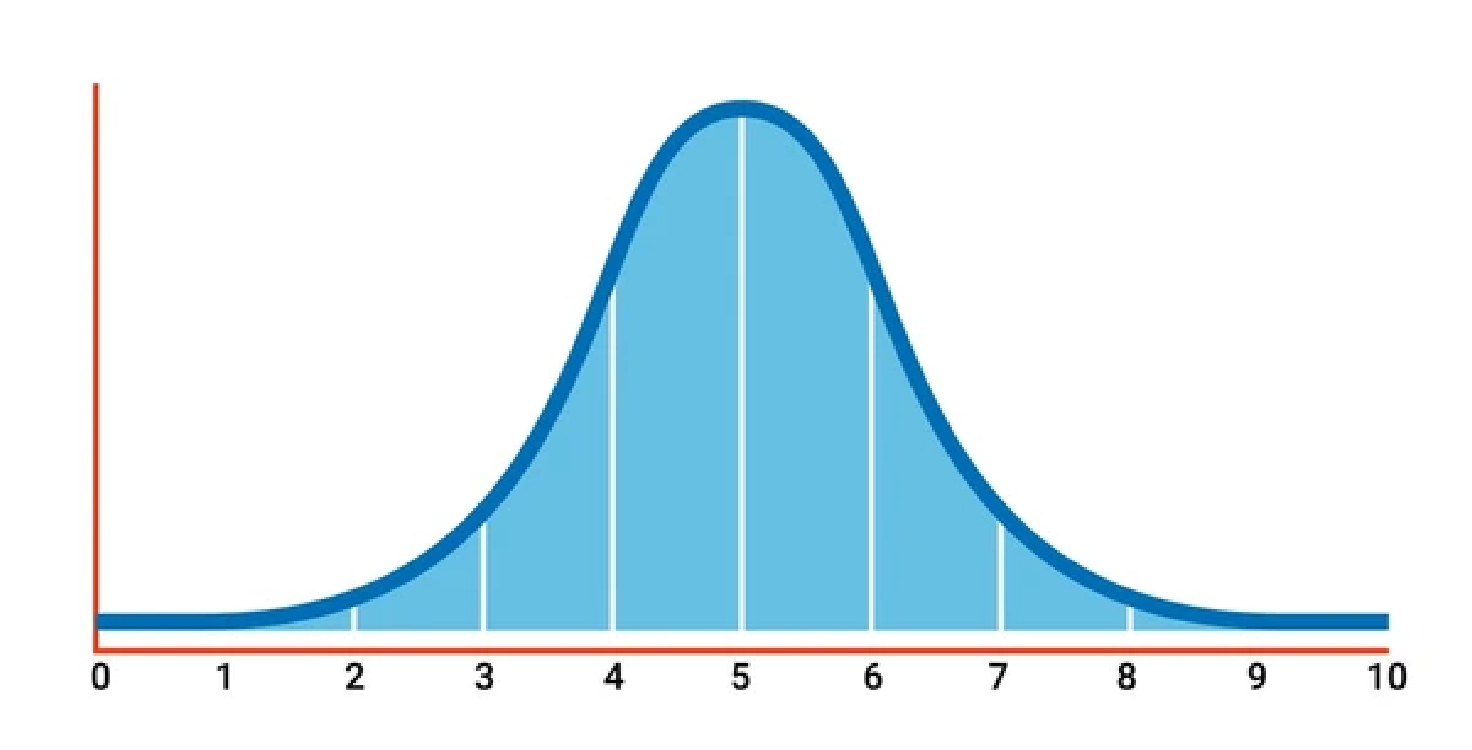
Caso A (n pequeño): Con n=5 en una distribución Exponencial, el sistema arroja una advertencia, mostrando una cobertura real menor al 93% y múltiples líneas rojas en el gráfico, indicando que la aproximación normal falló.

Caso B (n adecuado): Con n=30 o mayor, el indicador de cobertura se torna verde (cercano al 95%), validando el TLC.



CONCLUSIONES

El Teorema del Límite Central es valioso porque se aplica a muchos casos, la simulación computacional hace más fácil aprender estadística y, además, un buen tamaño de muestra permite obtener resultados más confiables.



REFERENCIAS

Walpole, R. E., et al. (2012). Probabilidad y estadística para ingeniería y ciencias.

Documentación de Streamlit y NumPy (2024).

Notas de curso: Probabilidad y Estadística, ESCOM.

Wackerly, D., Mendenhall, W., y Scheaffer, R. (2010) Estadística Matemática con Aplicaciones.