86.09

Procesos Estocásticos

Aplicaciones del Método Montecarlo

Trabajo Práctico N°1

Grupo N°4

2C 2024

Autor:	**Padrón:**	**Correo:**
Gonzalo Antahuara	109965	gantahuara@gmail.com
Marco Brischetto	110008	mbrischetto@fi.uba.ar
Ignacio Cavicchioli	109428	icavicchioli@fi.uba.ar
Tiago Sandoval	104169	tsandoval@fi.uba.ar

Índice

1. Introducción

El presente informe tiene por objeto la utilizar del método de *Montecarlo* para la simulación de variables aleatorias, cálculo de integrales definidas y modelado de experimentos.

2. Simulación de muestras de una distribución normal

2.1. Inciso (a)

La simulación pedida se realizó en MATLAB y el código se puede ver en la entrega correspondiente. En pocas palabras, la secuencia seguida es la siguiente:

- Se crearon 2 vectores de longitud N que contienen los resultados de simular dos variables aleatorias uniformemente distribuidas en el intervalo (0,1). Para permitir recrear el experimento se fijó la semilla.
- Usando las fórmulas propuestas en las consignas, se armaron 2 vectores que contienen las simulaciones de las variables Z_1 y Z_2.
- Las medias y varianzas correspondientes a cada variable se calcularon con las funciones integradas en el software.
- Se generaron los histogramas correspondientes a las dos simulaciones hechas. Los parámetros de la función se eligieron tal que se pudiese observar correctamente la semejanza de la curva obtenida a una Normal.
- Por último, se calculó el coeficiente de correlación de Pearson entre las variables Z_1 y Z_2 y se creó un gráfico de dispersión.

A continuación se presentan los gráficos y datos obtenidos del programa de MATLAB.

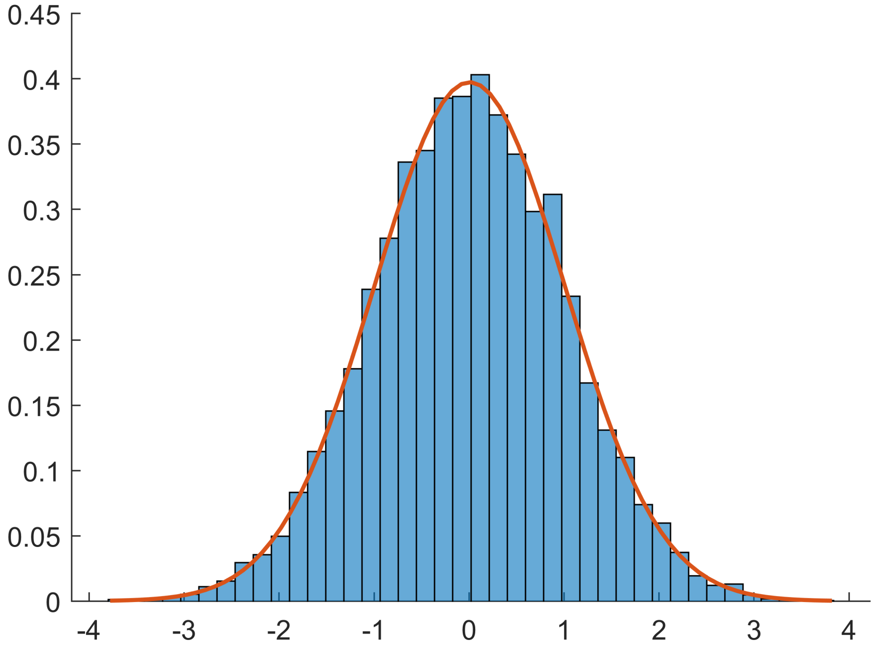

Figura 1: Histograma de Z1 con PDF normal superpuesta

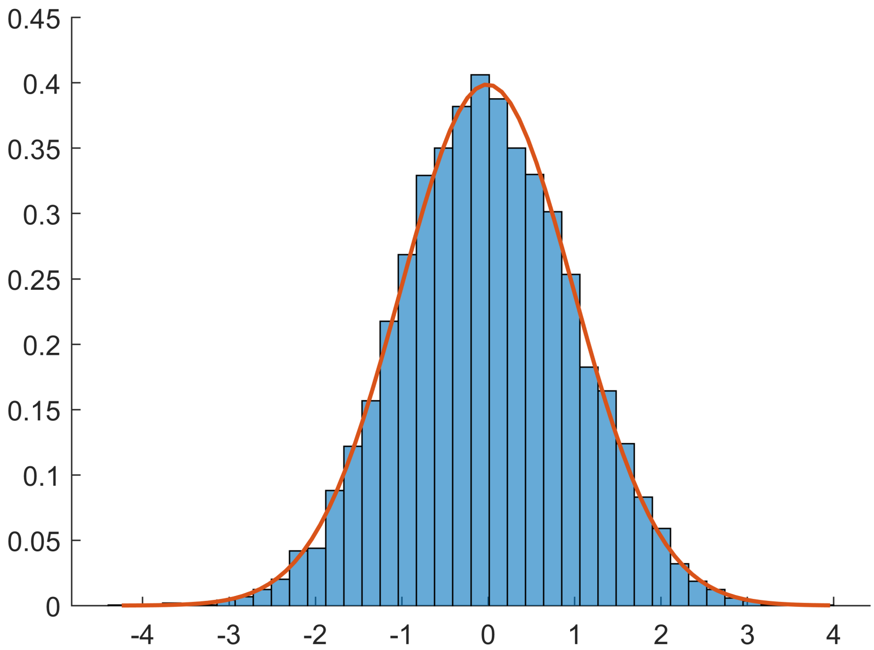

Figura 2: Histograma de Z2 con PDF normal superpuesta

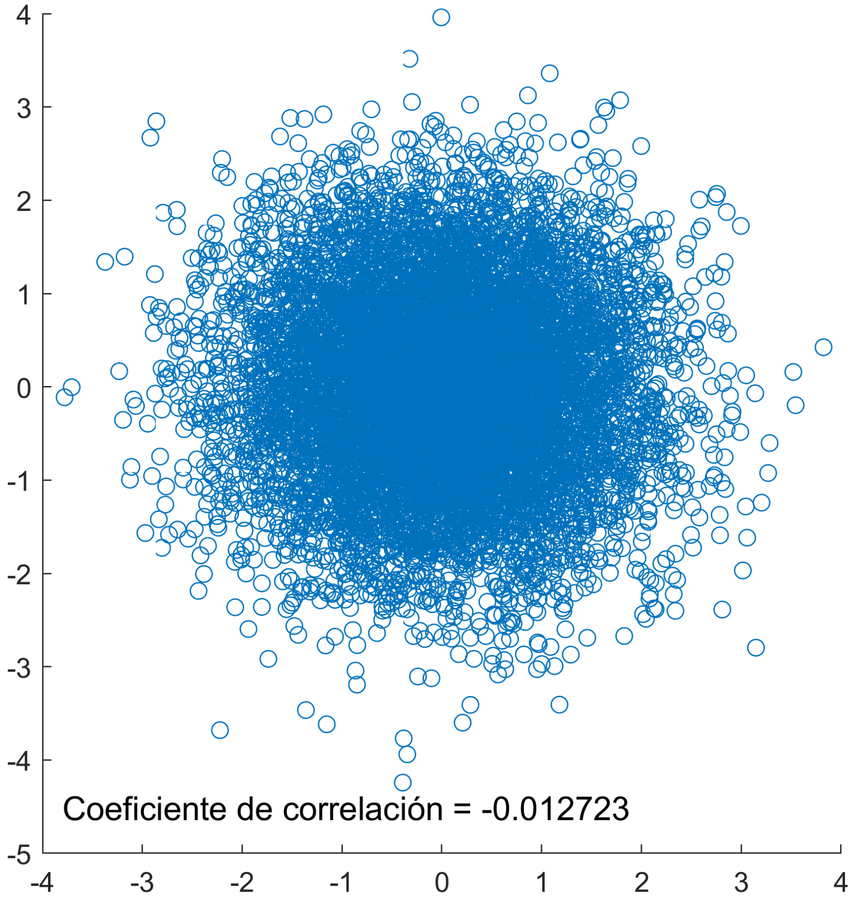

Figura 3: Gráfico de dispersión de Z2 con Z1

Las imágenes 1 y 2 muestran los histogramas de las variables Z_1 y Z_2 junto con la función de densidad normal que mejor aproxima el comportamiento de cada variable. Los parámetros estimados de cada variable son:

Variable	Media	Varianza
Z_1	0.0064	1.0075
Z_2	-0.0126	1.0007

Tabla 1: Estimaciones de media y varianza de Z_1 y Z_2

El gráfico 3 muestra una dispersión de puntos entre Z_2 y Z_1. El coeficiente de correlación de Pearson obtenido es de -0.012723. Dada la cercanía del coeficiente de correlación a cero y la forma de la dispersión, se puede afirmar que las variables en cuestión son independientes.

2.2. Inciso (b)

Utilizando el método *Variables equivalentes* se demuestra que $X = \sigma Z + \mu$ se distribuye como $X \sim N(\mu; \sigma^2)$, siendo Z una variable aleatoria con distribución $N(0;1)$. Para ello se definió la función de distribución acumulativa (CMF) de X de la siguiente forma:

$$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\sigma Z + \mu \leq x) = \mathbb{P}\left(Z \leq \frac{x-\mu}{\sigma}\right) = F_Z\left(\frac{x-\mu}{\sigma}\right)$$

Mediante la función de densidad de probabilidad (PDF) de Z se obtiene la expresión equivalente para $F_Z\left(\frac{x-\mu}{\sigma}\right)$:

$$F_X(x) = \int_{-\infty}^{\frac{x-\mu}{\sigma}} f_Z(z)\mathrm{d}z = \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}\mathrm{d}z$$

Utilizando el cambio de variable $z = \frac{x-\mu}{\sigma}$ donde $dz = \frac{dx}{\sigma}$, se obtiene la siguiente expresión:

$$F_X(x) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}\mathrm{d}x \implies f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

De este modo, debido a que la PDF encontrada para X corresponde a la de una variable aleatoria con distribución Normal de parámetros μ y σ^2, queda demostrado que $X \sim N(\mu; \sigma^2)$.

Para encontrar el valor de $\mathbb{E}[Z]$, se utilizó la propiedad de linealidad de la esperanza tal como se muestra a continuación:

$$\mathbb{E}[Z] = \mathbb{E}\left[\frac{X-\mu}{\sigma}\right] = \frac{\mathbb{E}[X]-\mu}{\sigma} = 0$$

Despejando $\mathbb{E}[X]$ se obtiene que $\mathbb{E}[X] = \mu$.

Del mismo modo, la $\mathbb{V}[X]$ se puede calcular aplicando propiedades conocidas de la siguiente forma:

$$\mathbb{V}[Z] = \mathbb{V}\left(\frac{X-\mu}{\sigma}\right) = \frac{\mathbb{V}[X]}{\sigma^2} = 1$$

Despejando $\mathbb{V}[X]$ se demuestra que $\mathbb{V}[X] = \sigma^2$.

Propiedades usadas:

Sean a, b constantes y X una V.A. :

$$\mathbb{E}[a \cdot X + b] = a \cdot \mathbb{E}[X] + b$$

$$\mathbb{V}[a \cdot X + b] = a^2 \cdot \mathbb{V}[X]$$

2.3. Inciso (c)

Utilizando el método del inciso (a) para generar una distribución normal a partir de una uniforme y teniendo en cuenta la transformación de Box Müller del inciso (b), se realizaron los histogramas de las normales $X_1 \sim N(0,2)$, $X_2 \sim N(1,2)$ y $X_3 \sim N(1,4)$. Los pasos a seguir fueron los siguientes.

- Utilizando el vector que contiene la simulación de la variable Z_1 se aplicaron las transformaciones:
  - $X_1 = \sqrt{2}Z_1$
  - $X_2 = \sqrt{2}Z_1 + 1$
  - $X_3 = 2Z_1 + 1$
- Se calcularon las medias y varianzas de las variables simuladas.
- Se realizaron los histogramas de cada uno de los casos con la función de densidad superpuesta correspondiente en cada caso

A continuación se puede observar una tabla que contiene los parámetros teóricos y experimentales. La fórmula de error % usada es la siguiente:

$$Error\,\% = 100 \cdot \frac{|V_{ref} - V_{med}|}{|V_{ref}|}$$

Variable	$\mathbb{E}[X]$ teórica	$\mathbb{E}[X]$ estimada	Error %	$\mathbb{V}[X]$ teórica	$\mathbb{V}[X]$ estimada	Error %
X_1	0	0,0091	-	2	2,0150	0,75
X_2	1	1,0091	0,91	2	2,0150	0,75
X_3	1	1,0128	1,28	4	4,0299	0,7475

Tabla 2: Tabla comparativa de esperanzas y varianzas

Tanto las esperanzas como varianzas calculadas experimentalmente son porcentualmente cercanas a sus contrapartes teóricas, con la excepción de la esperanza estimada de X_1, que en magnitud se aproxima a cero pero la definición de error relativo porcentual no considera el caso de cero como referencia.

En las siguientes hojas se muestran los histogramas generados.

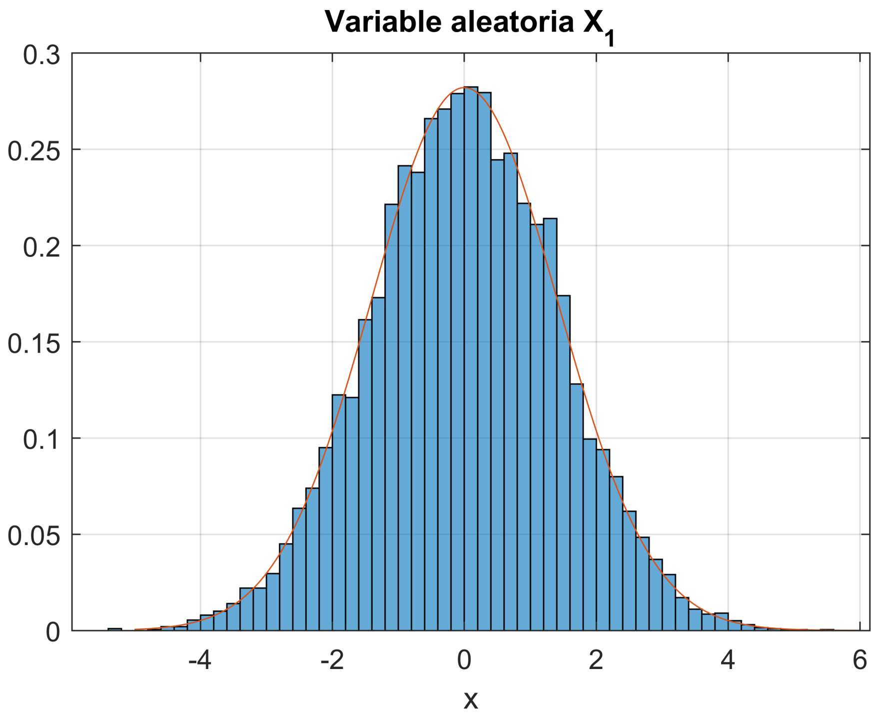

Figura 4: Gráfico de $X_1 = \sqrt{2}Z_1$

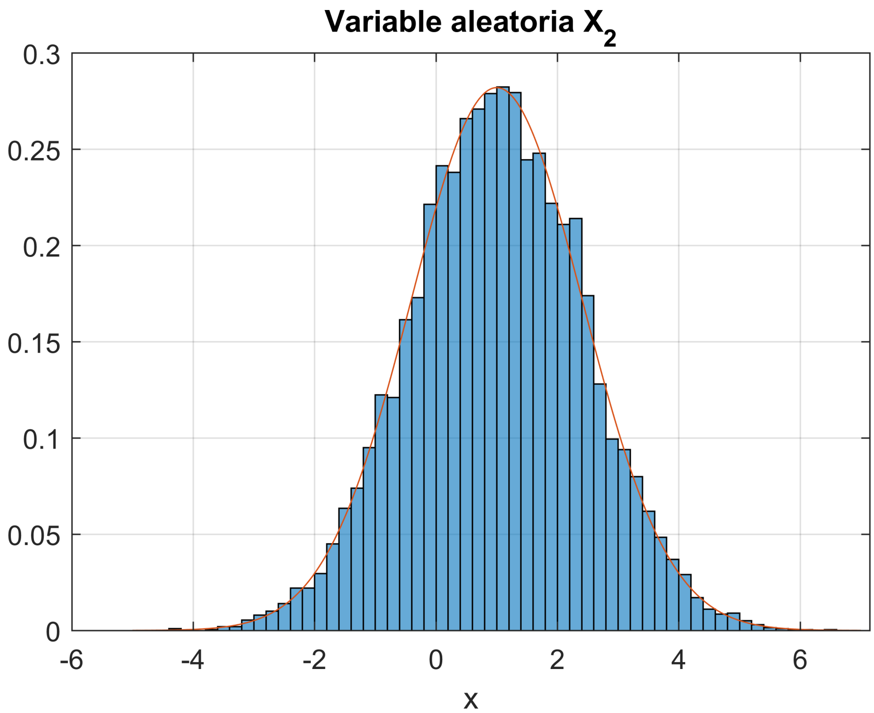

Figura 5: Gráfico de $X_2 = \sqrt{2}Z_1 + 1$

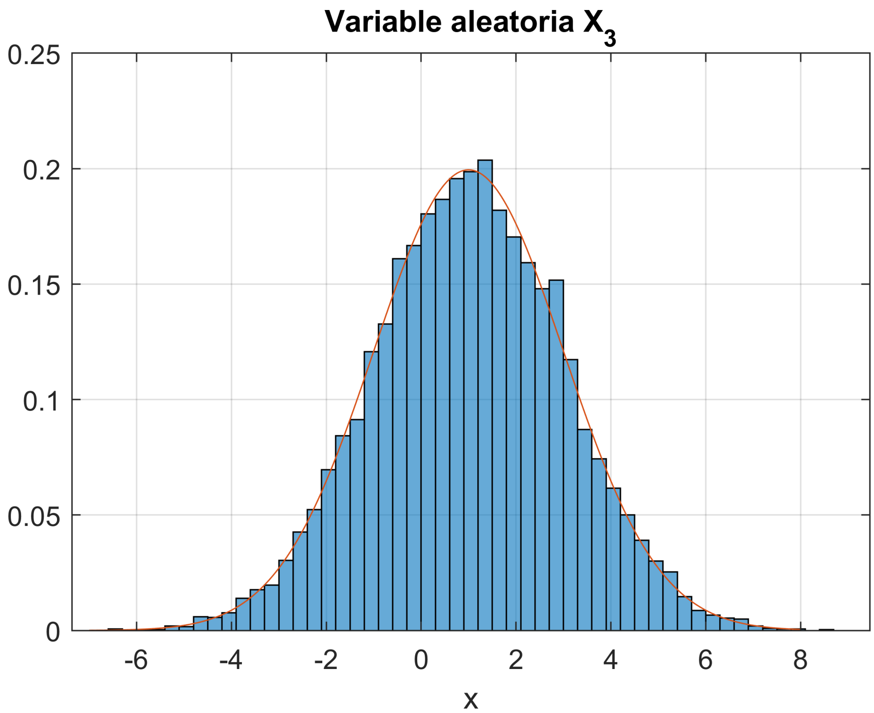

Figura 6: Gráfico de $X_3 = 2Z_1 + 1$

Los histogramas obtenidos coinciden visualmente con las distribuciones esperadas de cada variable simulada, notándose la media y varianza en cada caso.

3. Resolución de integrales

El método de Montecarlo nos permite computar integrales definidas, haciendo realizaciones de una variable aleatoria uniforme evaluada en la función a integrar y tomando su media. En esta sección utilizaremos este método para aproximar distintas probabilidades de una variable aleatoria de distribución normal. Por definición, la función de densidad de una variable aleatoria normal $X \sim N(\mu, \sigma^2)$ se define según la ecuación (1).

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{1}$$

Para hallar la probabilidad de un conjunto comprendido entre dos valores a y b es necesario calcular la integral de la ecuación (2).

$$\mathbb{P}(a < X < b) = \int_a^b f_X(x)\, dx \tag{2}$$

Haciendo uso del método de Montecarlo, es posible aproximar la integral según la ecuación (3).

$$\mathbb{P}(a < X < b) \simeq \frac{b-a}{N} \sum_{i=1}^{N} f_X(u_i) \tag{3}$$

Donde u_i son las realizaciones de una variable aleatoria U distribuida uniformemente sobre el intervalo de integración y N la cantidad de realizaciones de dicha variable.

3.1. Cálculo de probabilidades

Consideraremos una variable aleatoria normal $X \sim N(2, 3)$. En base al marco teórico previo, los pasos para calcular la probabilidad se explican a continuación.

- Se generan N realizaciones de una variable aleatoria U uniforme en el intervalo de integración.
- Se evalúan las realizaciones de U en la función de densidad de la variable aleatoria normal $f_X(x)$.
- Se computa la media de $f_X(u_i)$ y se multiplica por la longitud del intervalo de integración, como se indica en la ecuación (3).

La tabla 4 muestra los resultados del calculo de probabilidades mediante este método para distintos intervalos. En este caso se consideró $N = 10000$.

Inciso	Intervalo	Probabilidad
(a)	$(\mu - \sigma; \mu + \sigma)$	0,683389
(b)	$(\mu - 2\sigma; \mu + 2\sigma)$	0,952131
(c)	$(\mu - 3\sigma; \mu + 3\sigma)$	0,995456

Tabla 3: Resultados del calculo de probabilidad aproximado de una variable aleatoria normal en distintos intervalos

3.2. Comparativa de resultados

Es de interés conocer el error cometido en el cálculo de la integral en función de la cantidad de realizaciones N de la variable aleatoria uniforme, ya que este se relaciona con el tiempo de computo. Para observar esto, se estimó nuevamente la probabilidad del inciso (a) para distintos valores de N y se compararararon los resultados calculando el *Error Cuadrático Medio* con un valor de referencia I. El MSE se define según la ecuación (4), donde $\hat{I}_N$ es el resultado obtenido con el método numérico para un valor de N determinado.

$$MSE = \mathbb{E}[(\hat{I}_N - I)^2] \tag{4}$$

Nuevamente, haciendo uso del la Ley de los Grandes Numeros, estimaremos el MSE segun la ecuacion (5), donde M indica la cantidad de realizaciones del calculo de la integral I_N.

$$MSE \simeq \frac{1}{M} \sum_{k=1}^{M} (\hat{I}_{k,N} - I)^2 \tag{5}$$

La secuencia a realizar para el computo aproximado del MSE se explica a continuación.

- Se generan M realizaciones de la integral I_N, siguiendo el metodo explicado en las secciones previas.
- Se evalúan las realizaciones de la integral en la formula $(\hat{I}_N - I)^2$
- Se calcula la media de los resultado de la evaluación.

La tabla 4 muestra los resultados obtenidos de las integrales y el MSE estimado respecto del valor de referencia para distintos valores de N, con M = 50.

N	I_N	MSE
10	0,682821	$7,64^{-4}$
10^2	0,691315	$1,26^{-4}$
10^3	0,681592	$8,96^{-6}$
10^4	0,681151	$7,07^{-7}$
10^5	0,682304	$1,14^{-7}$
10^6	0,682826	$7,49^{-9}$

Tabla 4: Resultados de las estimaciones de las integrales y el Error Cuadrático Medio

La figura 7 muestra el Error Cuadrático Medio en función de la cantidad de realizaciones. Ambos ejes están en escala logarítmica de base 10 para mejorar la interpretación de los datos.

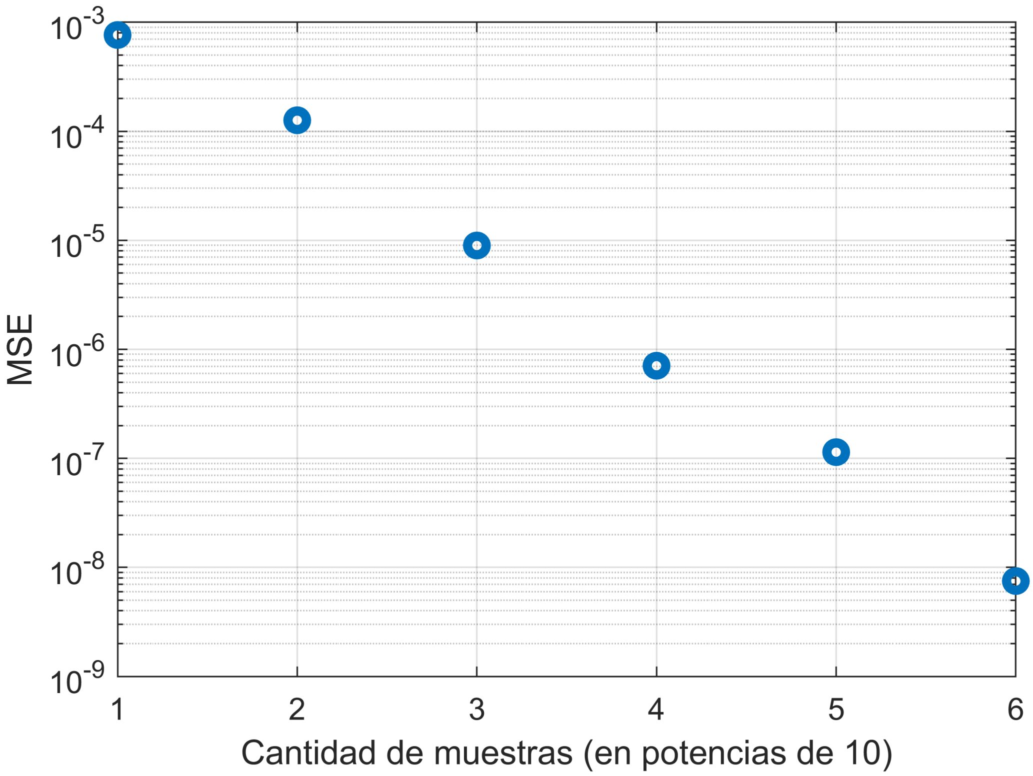

Figura 7: Error Cuadrático Medio en función del numero de realizaciones

Por la forma del gráfico se concluye que, a medida que aumenta la cantidad de muestras, la integral converge más a su valor real y el error se acerca cada vez más a cero. Este comportamiento es lógico y explicable por la Ley de los grandes números.

4. Simulación de lanzamiento de dados

4.1. Inciso (a)

El experimento aleatorio planteado en las consignas consiste en arrojar 2 dados no pesados y sumar los números obtenidos. El experimento resulta favorable cuando la suma de ambas tiradas resulta igual a 7 u 11. Este proceso se puede modelar definiendo las siguientes 3 variables aleatorias:

- X_1 es es resultado de la primer tirada. Su dominio es $1 \leq X_1 \leq 6$.
- X_2 es es resultado de la segunda tirada. Su dominio es $1 \leq X_2 \leq 6$.
- $A = X_1 + X_2$ es la suma de las tiradas. Su dominio es $2 \leq Y \leq 12$.

Se busca la $\mathbb{P}(A = 7 \cup A = 11)$, que se puede calcular como el cociente entre la cantidad de eventos favorables y eventos totales. La variable A es graficable en función de X_1 y X_2 por lo que los eventos favorables se pueden extraer visualmente. Abajo se muestra la variable calculada para cada par de X_1 y X_2.

A	-	X_1					
-	-	1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
X_2	3	4	5	6	7	8	9
	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Tabla 5: Valores de la variable $A = X_1 + X_2$

Usando la tabla 5 se puede plantear que:

$$\mathbb{P}(A = 7 \cup A = 11) = \frac{\text{Eventos favorables}}{\text{Eventos totales}} = \frac{8}{36} = \frac{2}{9} \tag{6}$$

La distribución que mejor modela el experimento es una Bernoulli de probabilidad $\frac{2}{9}$ de éxito. Se define la variable aleatoria Y a partir del evento de éxito “sale 7 u 11 de la suma entre los dados”del experimento de arrojar 2 dados. Así se da que:

$$Y \sim \text{Bernoulli}\left(\frac{2}{9}\right) \tag{7}$$

La distribución de Bernoulli es una distribución discreta que modela un experimento que tiene una probabilidad p de éxito y probabilidad $1 - p$ de fracaso. La variable aleatoria vale 0 en el fracaso y 1 en el éxito.

4.2. Inciso (b)

Se realizó una simulación del juego de dados ya explicado y se generó un histograma en base a los datos obtenidos. Para hacer esto se desarrolló un script en MATLAB siguiendo el siguiente pseudocódigo:

- Se crearon 2 vectores de longitud N que contengan números enteros aleatorios entre 1 y 6. Estos vectores representan las tiradas de dados utilizados en el juego. Seguidamente se sumaron ambos vectores obteniendo así un vector de N sumas que simula el resultado de cada par de tiradas.
- Con el objetivo de evaluar el evento de éxito en cada lanzamiento se recorrió el vector de sumas, se asignó 1 a los componentes iguales a 7 u 11 y 0 a los demás. De este modo se obtuvo un vector de unos y ceros de dimensión N que representa los valores de la variable aleatoria Y ya definida en el inciso previo.
- Finalmente se creó un histograma en base a este vector de éxitos y fracasos.

En la siguiente figura se muestra el histograma con la función de probabilidad teórica superpuesta. Se observa que el histograma coincide con la distribución supuesta en el inciso anterior, verificando la probabilidad del evento de éxito calculada.

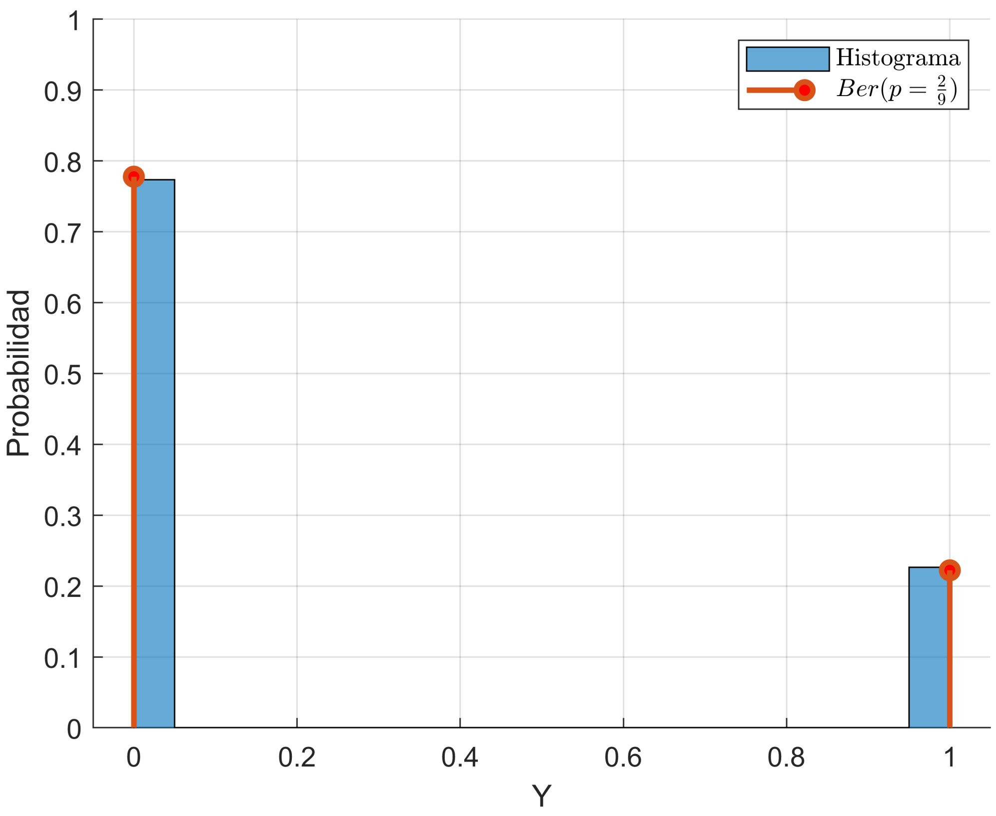

Gráfico 8: Histograma obtenido en la simulación y función de probabilidad teórica.

4.3. Inciso (c)

Recordando, la variable aleatoria Y fue definida como:

Y: "sale 7 u 11 de la suma entre los dados"

Dicha variable aleatoria puede tomar unicamente dos valores:

- $Y = 1$ cuando la suma entre los dos dados es 7 u 11.
- $Y = 0$ cuando la suma entre los dos dados no es 7 ni 11.

La secuencia de experimentos basados en la V.A. Y se puede pensar como un proceso de Bernoulli. De este modo, se da que la cantidad de éxitos en cierta cantidad de experimentos se modela con una distribución binomial. Esta se define como:

Z : "puntaje obtenido luego de repetir el experimento Y una cantidad de n veces"

$$Z \sim \text{Binomial}\left(n, \frac{2}{9}\right) \tag{8}$$

Por lo tanto, la funcion de probabilidad que describe dicha distribución es

$$f_Z(z) = \binom{n}{z} p^z (1-p)^{n-z} = \binom{n}{z} \left(\frac{2}{9}\right)^z \left(\frac{7}{9}\right)^{n-z}$$

$$\Longrightarrow f_Z(z) = \binom{n}{z} \left(\frac{7}{9}\right)^n \left(\frac{2}{7}\right)^z \; z \in \{0, 1, ..., n\} \tag{9}$$

4.4. Inciso (d)

De forma análoga al inciso (B), se buscó corroborar la elección de la distribución binomial por medio de un histograma basado en una simulación. Para esto se desarrolló otro script en MATLAB que sigue los siguientes pasos:

- Se crearon 2 vectores de longitud n que contienen números enteros aleatorios entre 1 y 6.
- Se obtuvo el vector de sumas tal como se hizo en el inciso b. Sin embargo ahora este vector representa las n sumas de los lanzamientos de un solo turno.
- Con el objetivo de evaluar los eventos de éxito por cada turno se recorrió el vector de sumas y se definió una variable auxiliar cuyo valor aumenta en 1 solo cuando los componentes del vector son iguales a 7 u 11.
- El valor obtenido de recorrer los vectores se guarda en un nuevo vector que representa los posibles valores de la variable aleatoria Z.
- Los pasos anteriores se repiten N cantidad de veces, obteniendo así un vector de dimensión N que contiene los puntajes obtenido en cada turno.
- Se genera el histograma con los datos del último vector mencionado.

Cabe destacar que, tal como se hizo en el grafico 8, se superpuso la distribución teórica con el histograma.

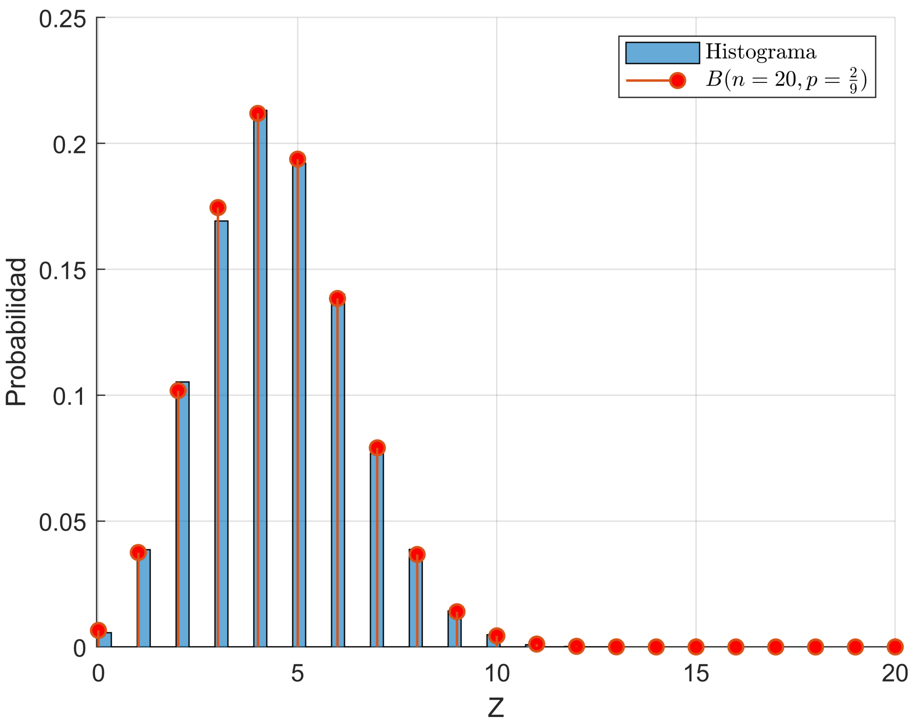

Gráfico 9: Histograma obtenido en la simulación y función de probabilidad teórica.

En el gráfico 9 se nota que el histograma creado a partir de 10000 turnos de 20 tiradas se asemeja a la distribución binomial planteada. Esto verifica la elección de la distribución.

5. Conclusión

Las consignas propuestas se pudieron llevar a cabo de forma satisfactoria, habiendo ahondado en el uso de la herramienta de Montecarlo para la resolución de problemas de diferentes tipos. También se pudo practicar la simulación de variables aleatorias tanto continuas como discretas en MATLAB teniendo en cuenta las discrepancias entre los modelos teóricos exactos y los simulados.