



UNIVERSIDAD DE BUENOS AIRES
FACULTAD DE INGENIERÍA
Año 2024 - 2^{do} cuatrimestre

PROCESOS ESTOCÁSTICOS (TA132)

TRABAJO PRÁCTICO 1 - APLICACIONES DEL MÉTODO MONTECARLO

Fecha de entrega: 26/09/2024

ESTUDIANTES: Grupo 5

Giani, Tomás	107629
<code>tgiani@fi.uba.ar</code>	
Lalía, Manuel	107979
<code>mlalia@fi.uba.ar</code>	
Perassi, Felipe	107630
<code>fperassi@fi.uba.ar</code>	
Rodríguez, Maximiliano	107604
<code>masrodriguez@fi.uba.ar</code>	

Índice

1. Introducción	2
2. Simulación de Muestras de una Distribución Normal	2
2.1. Transformación de Box Muller	2
2.2. Simulación de la Transformación de Box Muller	4
2.3. Transformación Lineal de una Normal Estándar	6
2.4. Generación de Normales a partir de una Estándar	8
3. Resolución de Integrales	9
4. Lanzamiento de Dados	13
4.1. Modelo Probabilístico de Lanzamiento de Dados	13
4.2. Simulación de Lanzamiento de Dados	14
4.3. Modelo Probabilístico de n Lanzamientos de Dados	15
4.4. Simulación de n Lanzamientos de Dados	16
5. Conclusiones	17

1. Introducción

Este trabajo práctico, aborda la aplicación del Método Montecarlo en distintos problemas probabilísticos y simulaciones. Para ello, se desarrollaron varias actividades que van desde la simulación de muestras de distribuciones normales mediante la Transformación de Box-Muller, hasta la resolución de integrales y la modelización de experimentos de lanzamiento de dados.

En el primer apartado, se emplea el método de Box-Muller para generar variables aleatorias con distribución normal a partir de variables uniformes, implementándolo y simulándolo en el *Software* de *Matlab*. Luego, se lleva a cabo la resolución de integrales a través del método Montecarlo, empleándolo para calcular probabilidades dentro de una distribución normal y calculando el error cuadrático medio según la cantidad de muestras empleadas. Finalmente, se estudian distintos experimentos de lanzamiento de dados, calculando la probabilidad de éxito y generando realizaciones a través de *Matlab*, para así, obtener la distribución de cada una de las experiencias.

2. Simulación de Muestras de una Distribución Normal

2.1. Transformación de Box Muller

A modo de introducir el tema a estudiar en este inciso, se desarrolló la transformación de Box Muller, tal que, se generaron dos variables aleatorias con distribución normal estándar ($Z_1 \sim \mathcal{N}(0; 1)$ y $Z_2 \sim \mathcal{N}(0; 1)$), a partir de dos variables uniformes ($U_1 \sim \mathcal{U}(0; 1)$ y $U_2 \sim \mathcal{U}(0; 1)$). Para ello, se plantearon las siguientes transformaciones.

$$\begin{cases} Z_1 = \sqrt{-2 \log(U_1)} \cdot \cos(2\pi U_2) \\ Z_2 = \sqrt{-2 \log(U_1)} \cdot \sin(2\pi U_2) \end{cases}$$

Para abordar lo planteado, se propone dividir la transformación en dos. Por ende, en primer lugar, se tiene lo siguiente.

$$\begin{cases} R = \sqrt{-2 \log U_1} \\ \Theta = 2\pi U_2 \end{cases} \quad (1)$$

Teniendo en cuenta que se define $(R, \Theta) = \tilde{g}(U_1, U_2) = (\sqrt{-2 \log U_1}, 2\pi U_2)$, se puede usar el método de la transformación biyectiva y suave. Por lo tanto, se obtiene que la función de densidad conjunta de R y Θ tendrá la forma que se ve a continuación.

$$f_{R,\Theta}(r, \theta) = \frac{f_{U_1, U_2}(u_1, u_2)}{|\det(J_{\tilde{g}(u_1, u_2)})|} \Big|_{\tilde{g}^{-1}(r, \theta)} \quad (2)$$

Partiendo de esta ecuación, se opera. En primera instancia, se sabe que si U_1 y U_2 son independientes, su densidad conjunta es la multiplicación de las marginales.

$$f_{U_1, U_2}(u_1, u_2) = f_{U_1}(u_1) \cdot f_{U_2}(u_2) = \mathbb{1}\{0 < u_1 < 1\} \cdot \mathbb{1}\{0 < u_2 < 1\}$$

Luego, se despeja $(U_1, U_2) = \tilde{g}^{-1}(R, \Theta)$ y se calcula $|\det(J_{\tilde{g}(u_1, u_2)})|$, tal que, se tiene:

$$(U_1, U_2) = \tilde{g}^{-1}(R, \Theta) = \left(e^{\frac{-R^2}{2}}, \frac{\theta}{2\pi} \right)$$

$$\det(J_{\tilde{g}(U_1, U_2)}) = \begin{vmatrix} -\frac{1}{U_1 \cdot \sqrt{-2 \log U_1}} & 0 \\ 0 & 2\pi \end{vmatrix} = -\frac{2\pi}{U_1 \cdot \sqrt{-2 \log U_1}} \Rightarrow |\det(J_{\tilde{g}(U_1, U_2)})| = \frac{2\pi}{U_1 \cdot \sqrt{-2 \log U_1}}$$

De esta manera, se reemplazan los resultados obtenidos en la ecuación 2.

$$\begin{aligned}
f_{R,\Theta}(r, \theta) &= \frac{f_{U_1, U_2}(u_1, u_2)}{|\det(J_{\tilde{g}}(u_1, u_2))|} \Big|_{\tilde{g}^{-1}(r, \theta)} \\
&= \frac{\mathbb{1}\{0 < u_1 < 1\} \cdot \mathbb{1}\{0 < u_2 < 1\}}{\frac{2\pi}{u_1 \cdot \sqrt{-2 \log u_1}}} \Big|_{\tilde{g}^{-1}(r, \theta)} \\
&= \frac{\mathbb{1}\{0 < e^{-\frac{r^2}{2}} < 1\} \cdot \mathbb{1}\{0 < \frac{\theta}{2\pi} < 1\}}{\frac{2\pi}{e^{-\frac{r^2}{2}} \cdot \sqrt{-2 \log e^{-\frac{r^2}{2}}}}} \\
&= \frac{\mathbb{1}\{r > 0\} \cdot \mathbb{1}\{0 < \theta < 2\pi\}}{\frac{2\pi}{e^{-\frac{r^2}{2}} \cdot r}} \\
&= \frac{r e^{-\frac{r^2}{2}}}{2\pi} \cdot \mathbb{1}\{r > 0\} \mathbb{1}\{0 < \theta < 2\pi\}
\end{aligned}$$

Como resultado parcial de lo anteriormente realizado, se tiene lo siguiente.

$$f_{R,\Theta}(r, \theta) = \frac{r e^{-\frac{r^2}{2}}}{2\pi} \cdot \mathbb{1}\{r > 0\} \mathbb{1}\{0 < \theta < 2\pi\}$$

Donde se ve que una parte de la densidad conjunta depende, pura y exclusivamente, de r y otra de θ . De esta forma, se puede ver que $R \sim \text{Ray}(\sigma = 1)$, es decir, sigue una distribución Rayleigh. Mientras que, $\Theta \sim \mathcal{U}(0; 2\pi)$ es una distribución uniforme de 0 a 2π .

$$f_{R,\Theta}(r, \theta) = \underbrace{r e^{-\frac{r^2}{2}} \mathbb{1}\{r > 0\}}_{R \sim \text{Ray}(\sigma=1)} \cdot \underbrace{\frac{1}{2\pi} \mathbb{1}\{0 < \theta < 2\pi\}}_{\Theta \sim \mathcal{U}(0; 2\pi)}$$

Por otro lado, se propone realizar la segunda transformación para así generar las variables aleatorias con distribución normal estándar.

$$\begin{cases} Z_1 = R \cdot \cos(\Theta) \\ Z_2 = R \cdot \sin(\Theta) \end{cases} \quad (3)$$

A partir de este punto se opera análogamente a lo realizado para la primera transformación. Por lo tanto, también se utiliza el método de la transformación biyectiva y suave, pues $(Z_1, Z_2) = \vec{h}(R, \Theta) = (R \cos(\Theta), R \sin(\Theta))$ cumple con las condiciones.

Entonces, la densidad conjunta de Z_1 y Z_2 tiene la siguiente forma.

$$f_{Z_1, Z_2}(z_1, z_2) = \frac{f_{R, \Theta}(r, \theta)}{|\det(J_{\vec{h}}(r, \theta))|} \Big|_{\vec{h}^{-1}(z_1, z_2)} \quad (4)$$

Teniendo en cuenta esta expresión, se despeja $(R, \Theta) = \vec{h}^{-1}(Z_1, Z_2)$ y se calcula el determinante del jacobiano.

$$\begin{aligned}
(R, \Theta) &= \vec{h}^{-1}(Z_1, Z_2) = \left(\sqrt{Z_1^2 + Z_2^2}, \text{atan2}\left(\frac{Z_2}{Z_1}\right) \right) \\
\det(J_{\vec{h}(R, \Theta)}) &= \begin{vmatrix} \cos(\Theta) & -R \sin(\Theta) \\ \sin(\Theta) & R \cos(\Theta) \end{vmatrix} = R \cos^2(\Theta) + R \sin^2(\Theta) = R \Rightarrow \therefore |\det(J_{\vec{h}(R, \Theta)})| = R
\end{aligned}$$

Luego, se reemplaza esto en la ecuación 4. A fin de obtener la densidad conjunta de Z_1 y Z_2 .

$$\begin{aligned}
f_{Z_1, Z_2}(z_1, z_2) &= \frac{f_{R, \Theta}(r, \theta)}{|\det(J_{\vec{h}(r, \theta)})|} \bigg|_{\vec{h}^{-1}(z_1, z_2)} \\
&= \frac{\frac{r e^{-\frac{r^2}{2}}}{2\pi} \cdot \mathbb{1}\{r > 0\} \mathbb{1}\{0 < \theta < 2\pi\}}{r} \bigg|_{\vec{h}^{-1}(z_1, z_2)} \\
&= \frac{e^{-\frac{r^2}{2}}}{2\pi} \cdot \mathbb{1}\{r > 0\} \mathbb{1}\{0 < \theta < 2\pi\} \bigg|_{\vec{h}^{-1}(z_1, z_2)} \\
&= \frac{e^{-\frac{(\sqrt{z_1^2 + z_2^2})^2}{2}}}{2\pi} \cdot \mathbb{1}\{\sqrt{z_1^2 + z_2^2} > 0\} \mathbb{1}\{0 < \text{atan2}\left(\frac{z_2}{z_1}\right) < 2\pi\} \\
&= \frac{e^{-\frac{z_1^2 + z_2^2}{2}}}{2\pi} \quad ; \quad z_1, z_2 \in \mathbb{R}
\end{aligned}$$

De esta manera, la densidad conjunta obtenida como consecuencia de las dos transformaciones desarrolladas es la que se ve a continuación.

$$f_{Z_1, Z_2}(z_1, z_2) = \frac{e^{-\frac{z_1^2 + z_2^2}{2}}}{2\pi} \quad ; \quad z_1, z_2 \in \mathbb{R}$$

Se observa claramente que es posible expresar la función de densidad conjunta como el producto de las densidades marginales de Z_1 y Z_2 , lo que sugiere que estas variables aleatorias son independientes. Además, al continuar con este análisis, se puede concluir que ambas variables siguen una distribución normal estándar, es decir, tienen media cero y varianza 1, $Z_1, Z_2 \sim N(\mu = 0; \sigma^2 = 1)$.

$$f_{Z_1, Z_2}(z_1, z_2) = \underbrace{\frac{e^{-\frac{z_1^2}{2}}}{\sqrt{2\pi}}}_{Z_1 \sim \mathcal{N}(0;1)} \cdot \underbrace{\frac{e^{-\frac{z_2^2}{2}}}{\sqrt{2\pi}}}_{Z_2 \sim \mathcal{N}(0;1)}$$

2.2. Simulación de la Transformación de Box Muller

Utilizando la transformación de Box Muller, genere dos variables aleatorias con distribución normal $Z_1 \sim \mathcal{N}(0; 1)$ y $Z_2 \sim \mathcal{N}(0; 1)$ a partir de dos variables uniformes $U_1 \sim \mathcal{U}(0; 1)$ y $U_2 \sim \mathcal{U}(0; 1)$. Para ello utilice las siguientes transformaciones:

$$\begin{aligned}
Z_1 &= \sqrt{-2 \log(U_1)} \cdot \cos(2\pi U_2) \\
Z_2 &= \sqrt{-2 \log(U_1)} \cdot \sin(2\pi U_2)
\end{aligned}$$

Considere una simulación de $N = 10000$ muestras para ambas variables uniformes. Estime la media y varianza de Z_1 y Z_2 y grafique los histogramas correspondientes. Realice también un gráfico de dispersión (*scatter plot*) de Z_2 versus Z_1 y calcule el coeficiente de correlación de Pearson entre ambas variables simuladas para verificar su independencia.

De acuerdo con lo indicado en la consigna del trabajo práctico, se desarrolló un *script* en *Matlab* para llevar a cabo la simulación correspondiente. A partir de esto, se estimaron la media y la varianza tanto de Z_1 como de Z_2 , lo que permitió que, al generar el histograma, se lo pueda comparar con la función de densidad de cada distribución normal.

Por ende, se presenta el pseudo código que se realizó previamente a desarrollar el programa. Este cumple con lo anteriormente descrito y se puede apreciar en el siguiente cuadro.

Pseudocódigo

1. Se definió un largo de 10000 realizaciones.
2. Se definieron los bins como la raíz cuadrada del largo de las realizaciones.
3. Se crearon dos Uniformes de largo 10000.
4. Se definieron las dos Normales en base a la definición de la Transformación de Box Muller.
5. Se calculó la media, la varianza y el coeficiente de correlación para cada una de las Normales.
6. Se crearon los histogramas de cada variable y el gráfico de dispersión.
7. Se definieron las funciones de densidad de las Normales (0,1) para comparar con los histogramas.

De esta manera, se obtuvieron las imágenes que se observan a continuación.

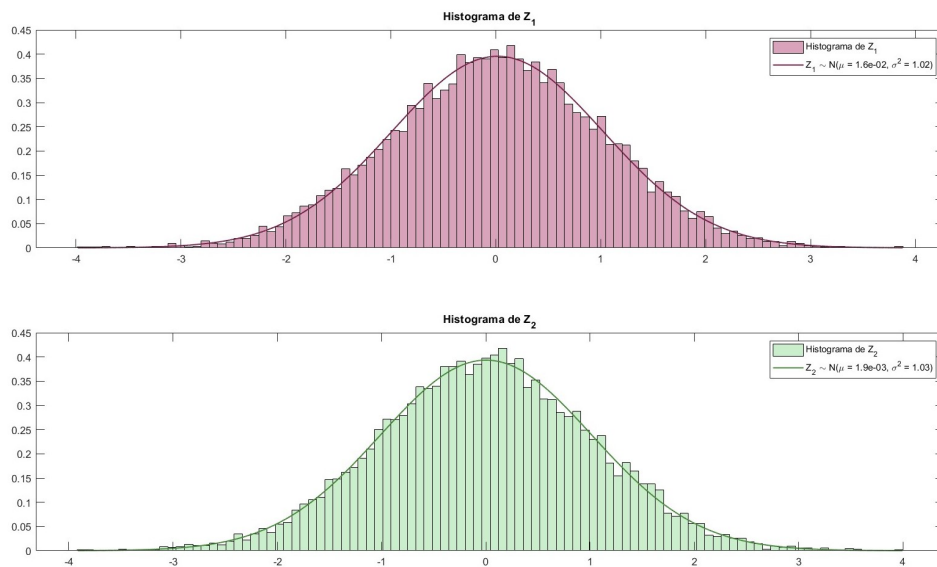


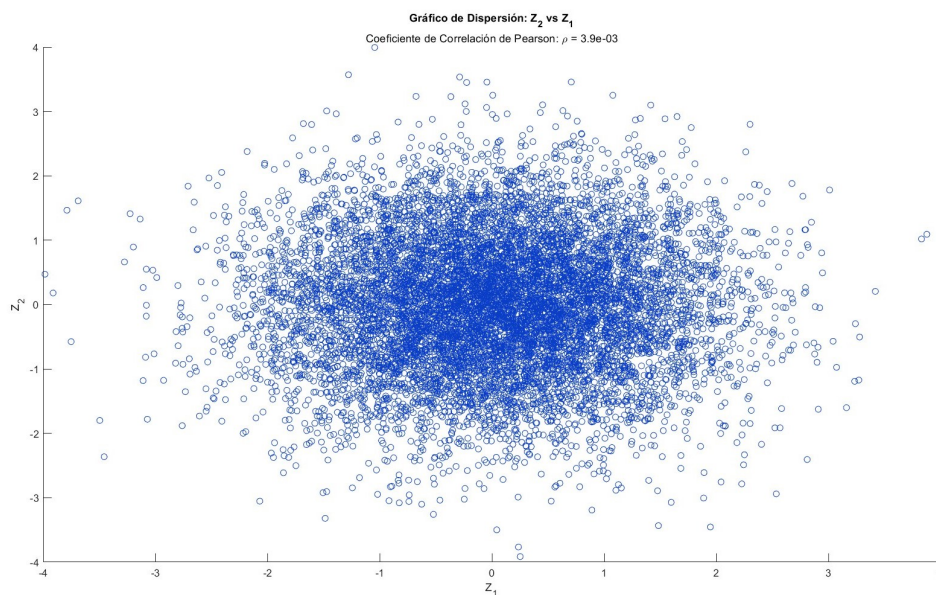
Figura 1: Comparación de histogramas con la función de densidad de Z_1 y Z_2

En las leyendas de ambas figuras, se ven claramente los parámetros de cada una de las normales, lo que corresponde a la estimación de la media y la varianza de Z_1 y Z_2 . De igual forma, a continuación se las especifica en el informe, para mejorar la claridad de lo realizado.

$$\begin{cases} Z_1 \sim \mathcal{N}(\mu = 0,016; \sigma^2 = 1,02) \\ Z_2 \sim \mathcal{N}(\mu = 1,9 \times 10^{-3}; \sigma^2 = 1,03) \end{cases}$$

A partir de estos resultados y viendo que en la Fig. 1 la distribución del histograma sigue la función de densidad de probabilidad teórica de ambas gaussianas, se puede concluir que tanto Z_1 como Z_2 siguen una distribución normal estándar al simular con $N = 10000$ muestras.

Por otro lado, se graficó el diagrama de dispersión o *Scatter Plot* y se calculó el coeficiente de correlación de Pearson entre ambas normales, Z_1 y Z_2 .

Figura 2: Gráfico de dispersión de Z_2 vs Z_1

En base al análisis de la Fig. 2, se obtiene un coeficiente de correlación $\rho = 3,9 \times 10^{-3}$, un valor muy cercano al cero teórico que se espera cuando las variables son independientes, como es el caso aquí. La pequeña diferencia con respecto al valor teórico puede atribuirse a las dispersiones propias de las simulaciones, las cuales impiden que el coeficiente sea exactamente cero. Sin embargo, el resultado sigue siendo lo suficientemente próximo como para considerarse prácticamente nulo.

2.3. Transformación Lineal de una Normal Estándar

Demuestre que dada una variable aleatoria normal estándar $Z \sim \mathcal{N}(0; 1)$, la media de una variable normal definida a partir de la transformación $X = \sigma Z + \mu$ es una variable normal $\sim \mathcal{N}(\mu; \sigma^2)$ de media $E[X] = \mu$ y varianza $V[X] = \sigma^2$.

En este apartado, se pide demostrar que a partir de una transformación lineal de una normal estándar, se obtiene otra gaussiana con media $E[X] = \mu$ y varianza $V[X] = \sigma^2$.

Para abordar lo que se pide, se comienza justificando que una transformación lineal de una distribución normal produce otra distribución normal. Intuitivamente, esto se comprende al observar cómo afecta la transformación la forma de la función de densidad de las distribuciones gaussianas. Al aplicar una transformación lineal, esencialmente se multiplica por un factor, lo que provoca una expansión o contracción, y se suma un valor, lo que genera un desplazamiento en el eje x. Como resultado, la forma de la distribución se conserva, pues sigue manteniéndose como una gaussiana, pero con una media modificada (desplazamiento en el eje x) y una varianza ajustada (expansión o contracción de la campana).

Luego, siguiendo con lo solicitado, se obtiene la media de la variable aleatoria definida como $X = \sigma Z + \mu$. Para esto, se utilizó que el operador esperanza es lineal y que la esperanza de una constante es igual a la constante.

$$\begin{aligned} E[X] &= E[\sigma Z + \mu] \\ &= E[\sigma Z] + E[\mu] \\ &= \underbrace{\sigma E[Z]}_0 + \underbrace{E[\mu]}_{\mu} \implies \therefore E[X] = \mu \end{aligned}$$

Por otra parte, se obtuvo la varianza de X . Para ello, se usaron distintas propiedades correspondientes a la varianza, principalmente que $V[X] = E[X^2] - E[X]^2$, y las anteriormente vistas para la media. De esta manera, operando se tiene el resultado que se buscaba.

$$\begin{aligned}
V[X] &= E[X^2] - E[X]^2 = E[(\sigma Z + \mu)^2] - \underbrace{E[X]^2}_{\mu^2} \\
&= E[\sigma^2 Z^2 + \sigma Z\mu + \mu^2] - \mu^2 \\
&= E[\sigma^2 Z^2] + E[\sigma Z\mu] + \underbrace{E[\mu^2]}_{\mu^2} - \mu^2 \\
&= \sigma^2 E[Z^2] + \sigma \mu \underbrace{E[Z]}_0 \\
&= \sigma^2 E[Z^2] \\
&= \sigma^2 (\underbrace{V[Z]}_1 + \underbrace{E[Z]^2}_0) \Rightarrow \therefore V[X] = \sigma^2
\end{aligned}$$

Habiendo hecho este análisis, se agregó una simulación realizada por computadora que sigue los siguientes lineamientos.

Pseudocódigo

1. Se definió un largo de 10000 realizaciones.
2. Se definieron los bins como la raíz cuadrada del largo de las realizaciones.
3. Se definieron μ y σ^2 para que sean valores aleatorios.
4. Se definió una Normal a través de la Transformación de Box Muller.
5. A esa Normal se le aplica la transformación $X = AZ + b$ con $A = \sigma$ y $b = \mu$.
6. Se creó el histograma.
7. Se definió la función de densidad de la Normal (μ, σ^2) para comparar con el histograma.

Esta simulación, demuestra que la transformación lineal $X = \sigma Z + \mu$ de una normal estándar Z es otra normal X con los parámetros vistos anteriormente. Esto se puede apreciar en la imagen que se ve a continuación.

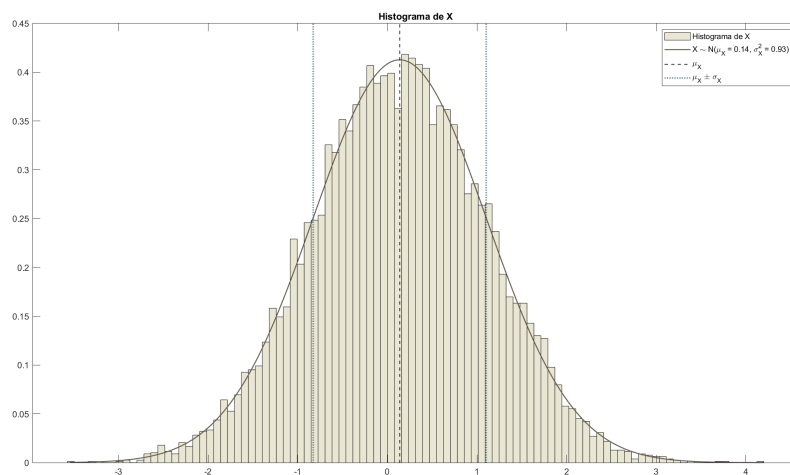


Figura 3: Comparación de histograma con la función de densidad de X

En la Fig. 4, se ve claramente que el histograma simulado se condice con la función de densidad de probabilidad de la normal dada por la transformación estudiada.

2.4. Generación de Normales a partir de una Estándar

Aplicando Box Muller para obtener muestras normales estándar, genere tres nuevas variables aleatorias normales $X_1 \sim \mathcal{N}(0; 2)$, $X_2 \sim \mathcal{N}(1; 2)$ y $X_3 \sim \mathcal{N}(1; 4)$. Estime en cada caso la media y varianza para su comparación con los valores teóricos. Grafique los histogramas de cada variable superpuesto a las curvas de densidad de probabilidad teóricas para verificar la coincidencia.

Con el objetivo de simular lo requerido en esta sección del trabajo práctico, se desarrolló un algoritmo computacional en el programa recomendado por la cátedra. Este, genera las imágenes que serán analizadas posteriormente y sigue la lógica que se ve a continuación.

Pseudocódigo

1. Se definió un largo de 10000 realizaciones.
2. Se definieron los bins como la raíz cuadrada del largo de las realizaciones.
3. Se definió una Normal Z a través de la Transformación de Box Muller.
4. A esa Normal se le aplicaron 3 transformaciones de la forma $X_i = AZ + b, i = \{1, 2, 3\}$, para cumplir con las medias y las varianzas requeridas.
5. Se crearon los histogramas.
6. Se definieron las funciones de densidad de las Normales para comparar con el histograma.
7. Se estimaron las medias y las varianzas con las funciones `-mean-` y `-var-`.

De esta manera, se obtienen los siguientes gráficos. En ellos, se especifica cada una de las gaussianas generadas y se compara su histograma con su función de densidad teórica.

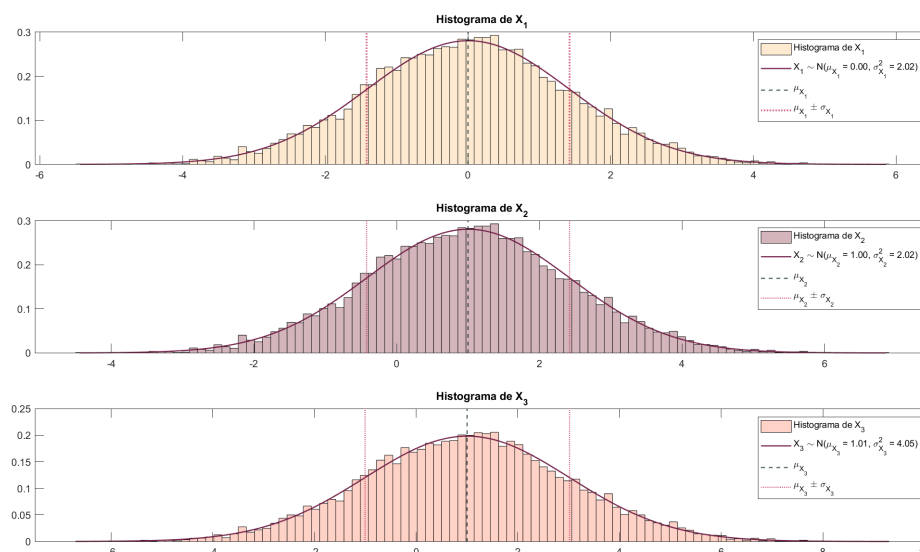


Figura 4: Comparación de histograma con la función de densidad de cada X

En los cuales se puede visualizar que cada una de las transformaciones realizadas, sigue una distribución normal cuya media y varianza cambian según lo requerido en cada caso. De esta manera, se tiene lo siguiente.

$$\begin{cases} X_1 \sim \mathcal{N}(\mu = 0; \sigma^2 = 2,02) \\ X_2 \sim \mathcal{N}(\mu = 1; \sigma^2 = 2,02) \\ X_3 \sim \mathcal{N}(\mu = 1,01; \sigma^2 = 4,05) \end{cases}$$

Con lo cual se puede concluir que las transformaciones realizadas en el *script*, han generado correctamente las normales solicitadas por la cátedra.

3. Resolución de Integrales

En esta sección, se aplica el método **Monte Carlo** para aproximar integrales definidas, ya que, se busca calcular la probabilidad de que una variable aleatoria $X \sim \mathcal{N}(\mu = 2; \sigma^2 = 3)$ se encuentre dentro de un rango $[a, b]$ determinado.

Para esto, antes de comenzar con los cálculos requeridos, se demuestra de donde se obtienen las aproximaciones de las integrales, lo cual se va a utilizar en este apartado del informe.

En primer lugar, siendo $f_X(x)$ la función de densidad de X , se busca obtener la integral que se ve a continuación.

$$I = \int_a^b f_X(x) dx$$

Recordando que, si $U \sim \mathcal{U}(a, b)$, entonces $f_U(u) = \frac{1}{b-a}$. Por lo tanto, se introduce esto operando de la siguiente forma.

$$I = (b-a) \frac{1}{b-a} \int_a^b f_X(x) dx = (b-a) \int_a^b f_X(x) f_U(x) dx = (b-a) E[f_X(x)]$$

Habiendo llegado a este resultado, se define X_i como las N muestras de una variable aleatoria uniforme en el intervalo $[a, b]$, que representan los posibles valores de la variable de integración x . De esta manera, por la **Ley Fuerte de los Grandes Números**, se tiene lo siguiente.

$$\frac{1}{N} \sum_{i=1}^N f_X(x_i) \xrightarrow{a.s.} E[f_X(x)]$$

Esto, refleja que la secuencia definida en base a las medias muestrales converge de forma casi segura (o *almost sure*) a la media de la variable aleatoria.

Por lo tanto, a partir de este resultado, se obtiene la siguiente aproximación de la integral.

$$\hat{I} = (b-a) \frac{1}{N} \sum_{i=1}^N f_X(x_i) \xrightarrow{a.s.} I$$

Donde se puede visualizar claramente que, en función de la *LFGN*, la aproximación \hat{I} tiende en forma casi segura al valor real I cuando la cantidad de muestras N tiende a infinito.

Suponga que X es una variable aleatoria normal de media $\mu = 2$ y varianza $\sigma^2 = 3$. Calcule $\mathbb{P}(\mu - \sigma < X < \mu + \sigma)$ con $N = 10000$ muestras.

En este primer inciso, recordando que $X \sim \mathcal{N}(\mu = 2; \sigma^2 = 3)$ se busca calcular la siguiente probabilidad.

$$\mathbb{P}(\mu - \sigma < X < \mu + \sigma) = \mathbb{P}(2 - \sqrt{3} < X < 2 + \sqrt{3})$$

A fin de resolver esto, se aplica el método de Monte Carlo siguiendo los lineamientos descriptos anteriormente. Para ello, se implementó un código que se basa en el pseudocódigo visto a continuación.

Pseudocódigo

1. Se generaron N realizaciones (X_i) uniformes en el intervalo de integración.
2. Se evaluó la función de densidad $f_X(x)$ en los puntos X_i generados.
3. Se aproximó la integral de acuerdo a la expresión hallada.

De esta forma, se obtuvieron los resultados que se observan en la Fig. 5.

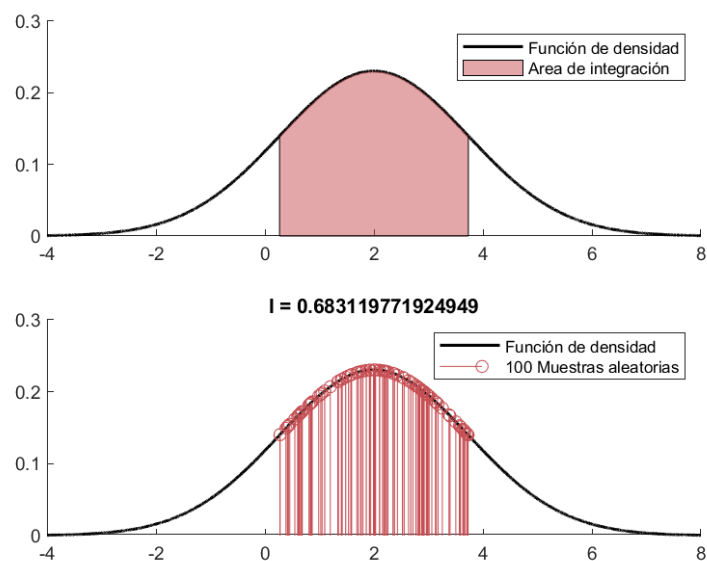


Figura 5: Aproximación de la integral

En la parte superior de la imagen se observa el área a integrar, mientras que en la parte inferior se muestran 100 de las 10000 muestras aleatorias utilizadas para realizar la aproximación.

Comparando ambas figuras, se puede ver a simple vista que la aproximación obtenida se corresponde con el resultado teórico, el cual indica que en la región comprendida en $\mu \pm \sigma$ se encuentra el 68 % de la probabilidad total.

Calcule $\mathbb{P}(\mu - 2\sigma < X < \mu + 2\sigma)$ con $N = 10000$ muestras.

En este punto, se trabajó análogamente a lo ya realizado en el anterior, obteniendo los resultados de la Fig. 6.

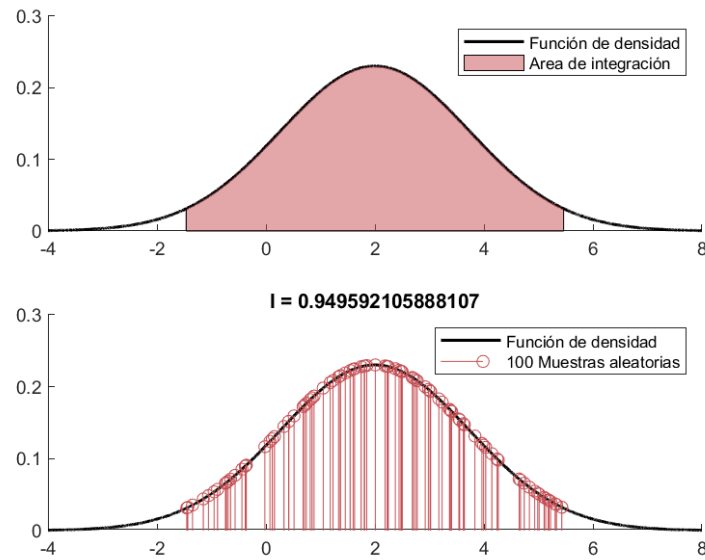


Figura 6: Aproximación de la integral

Nuevamente, la aproximación llevada a cabo se condice con el resultado teórico, el cual establece que el 95 % de la probabilidad total se encuentra en el intervalo $\mu \pm 2\sigma$.

Calcule $\mathbb{P}(\mu - 3\sigma < X < \mu + 3\sigma)$ con $N = 10000$ muestras.

Se procede de la misma forma que se hizo en los dos puntos anteriores, pero cambiando los intervalos de la probabilidad. En este caso, se obtuvo el resultado de la Fig. 7.

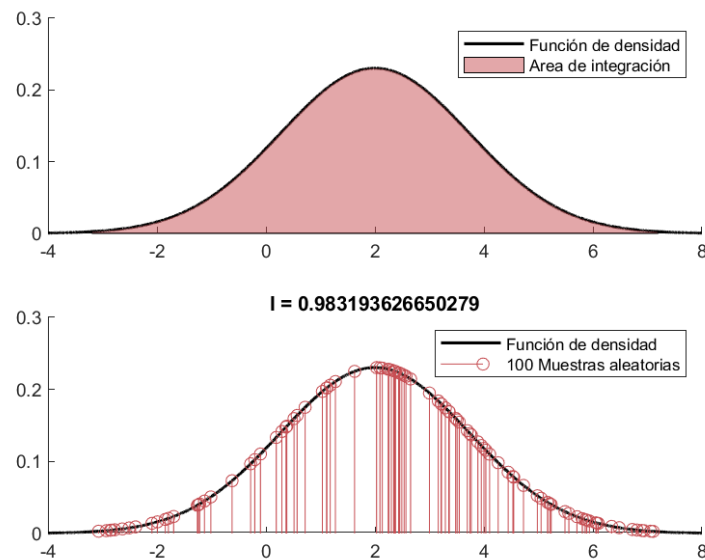


Figura 7: Aproximación de la integral

Donde se ve claramente que lo simulado por computadora coincide con lo teórico. También, se verifica que el 99 % de la probabilidad esta contenida en $\mu \pm 3\sigma$.

Calcule $\mathbb{P}(\mu - \sigma < X < \mu + \sigma)$ para distintas muestras de $N = (10, 10^2, 10^3, 10^4, 10^5, 10^6)$. Para comparar los resultados, grafique el error cuadrático medio $MSE = E[(\hat{I}_N - I)^2]$ en función de N , donde \hat{I}_N es el resultando de la integral para N muestras e I el valor teórico supuesto de la integral (asuma para este caso $I = 0,682687273250961$).

En este punto, se quiere analizar cómo se modifica el error de la aproximación realizada por el **Método Montecarlo** según la cantidad de muestras utilizadas (N). Para esto, se aproximará el error cuadrático medio (o MSE) de la siguiente manera.

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

Tomando $M = 50$ realizaciones, según lo que se dispone en las consignas del trabajo.

Con el fin de cumplir con lo solicitado, se desarrolló un programa en *Matlab* que implementa los pasos necesarios para llevar a cabo los cálculos correspondientes. El programa sigue una estructura clara y eficiente, basada en el pseudocódigo que se muestra a continuación.

Pseudocódigo

1. Se generaron 6 vectores de muestras de uniformes en el intervalo de integración, uno para cada valor de N .
2. Se evaluó la densidad en estas muestras uniformes para cada vector.
3. Se aproximó la integral de la misma manera que en los puntos a, b y c para cada vector.
4. Se calculó el error al cuadrado $(\hat{I} - I)^2$ para los 6 resultados.
5. Se repitió este procedimiento M veces para obtener las 50 realizaciones.
6. Se aproximó el MSE calculando el promedio para cada N .

De esta forma, se obtuvo el gráfico mostrado en la Fig. 8.

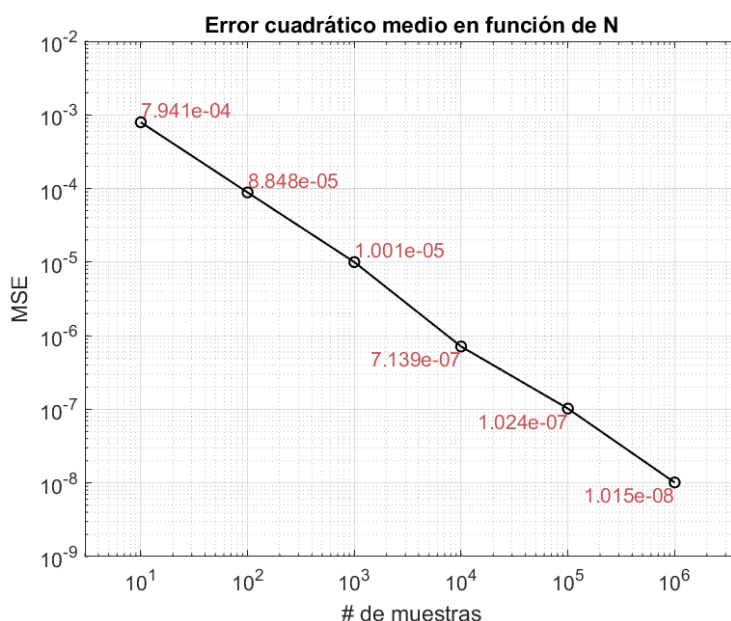


Figura 8: Aproximación de la integral

Se observa que el Error Cuadrático Medio (o MSE) disminuye de manera lineal a medida que aumenta el número de muestras N . Esto se debe a que, a medida que se incrementa el tamaño de la muestra, la estimación realizada a partir de la simulación se vuelve más precisa en comparación a su valor teórico.

4. Lanzamiento de Dados

4.1. Modelo Probabilístico de Lanzamiento de Dados

Suponiendo que tenemos dos dados no pesados (es decir que cada cara tiene la misma probabilidad de salir que las demas) considerando que un evento de éxito será cuando la suma de ambos dados dé 7 u 11. Determine la probabilidad de ocurrencia de ese evento e indique qué tipo de distribución lo modela adecuadamente.

En esta sección, se opera a fin de obtener la probabilidad de ocurrencia de que la suma de dos dados no pesados sea igual a 7 u 11. Además, se determina el modelo probabilístico que mejor se ajusta a este evento.

En primer lugar, se define A como el evento en el cual la suma de los dos dados es igual a 7 o 11. Para calcular la probabilidad ocurrencia de este evento, se utiliza la siguiente expresión.

$$\mathbb{P}(A) = \frac{\text{número de resultados favorables}}{\text{número total de resultados posibles}} \quad (5)$$

De esta manera, se define el espacio muestral correspondiente al evento bajo estudio. Este espacio está formado por todas las combinaciones posibles de caras entre ambos dados. Por ende, utilizando la regla del producto y entendiendo que cada dado tiene 6 caras, el número total de combinaciones posibles al lanzar los dos dados es la que se observa.

$$S = 6 \times 6 = 36$$

Luego, los casos de este espacio que se definen como éxito son los eventos en los cuales la suma de los dos dados es igual a 7 u 11. Por lo tanto, teniendo en cuenta que los dados están numerados (es decir, (4, 3) será una realización distinta a (3, 4)), se tienen las combinaciones que llevan a estos resultados.

- Para que la suma sea 7, las combinaciones posibles son:

$$(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$$

- Para que la suma sea 11, las combinaciones posibles son:

$$(5, 6), (6, 5)$$

A partir de este análisis de casos, se tiene lo que se lee a continuación.

- El número de resultados favorables es $n(A) = 8$, que corresponde a las combinaciones previamente enumeradas.
- El número total de resultados posibles es $S = 36$, ya que, hay 36 combinaciones posibles al lanzar dos dados.

Entonces, volviendo a 5, la probabilidad de que ocurra el evento A es la siguiente.

$$\mathbb{P}(A) = \frac{\text{número de resultados favorables}}{\text{número total de resultados posibles}} = \frac{8}{36} = \frac{2}{9}$$

Siguiendo con el razonamiento realizado anteriormente, este experimento sigue una distribución de Bernoulli, pues en cada lanzamiento de los dados sólo hay dos posibles resultados:

- Éxito, si la suma de los dos dados es 7 o 11. Su probabilidad es la calculada para el evento A , $p = \mathbb{P}(A) = \frac{2}{9}$.
- Fracaso, si la suma no es ni 7 ni 11. Su probabilidad es la complementaria del caso de éxito, por lo tanto, será $q = 1 - p = \frac{7}{9}$.

Por último, se tiene que la función de masa de probabilidad o *PMF* para una variable aleatoria de Bernoulli tiene la forma que se ve a continuación.

$$P(X = k) = p^k(1 - p)^{1-k}, \quad k = \{0, 1\}$$

Donde (X) es la variable aleatoria que representa el éxito ($X = 1$) o el fracaso ($X = 0$) y p es la probabilidad de éxito, mencionada anteriormente.

4.2. Simulación de Lanzamiento de Dados

Genere 10000 realizaciones que simulen un lanzamiento de dos dados con el que se evalúe el evento de éxito. Realice un histograma y compárelo con la función de probabilidad teórica de acuerdo a la distribución propuesta en el ejercicio anterior.

A fin de evaluar experimentalmente el ejercicio anterior, se genera una simulación de 10000 realizaciones del lanzamiento de dos dados. Esto permite aproximar empíricamente la probabilidad del evento, observando la frecuencia relativa de los resultados simulados.

Para ello, se desarrolló un código en *Matlab* que sigue los pasos descritos en el pseudocódigo mostrado más adelante.

Pseudocódigo

1. Se definió un largo de 10000 realizaciones
2. Se definieron los bins como la raíz del largo de las realizaciones
3. Se definieron dos realizaciones aleatorias con distribución uniforme (1,6), que se refieren a la tirada de los dos dados
4. Se definió una variable "SumaDados"
5. Se iteró "SumaDados" para contar la cantidad de éxitos y fracasos
6. Se realizó el histograma y se comparó con la función de masa de probabilidad de la Binomial teórica

Entonces, a continuación se presenta el histograma que muestra la distribución de los valores obtenidos en las 10000 simulaciones.

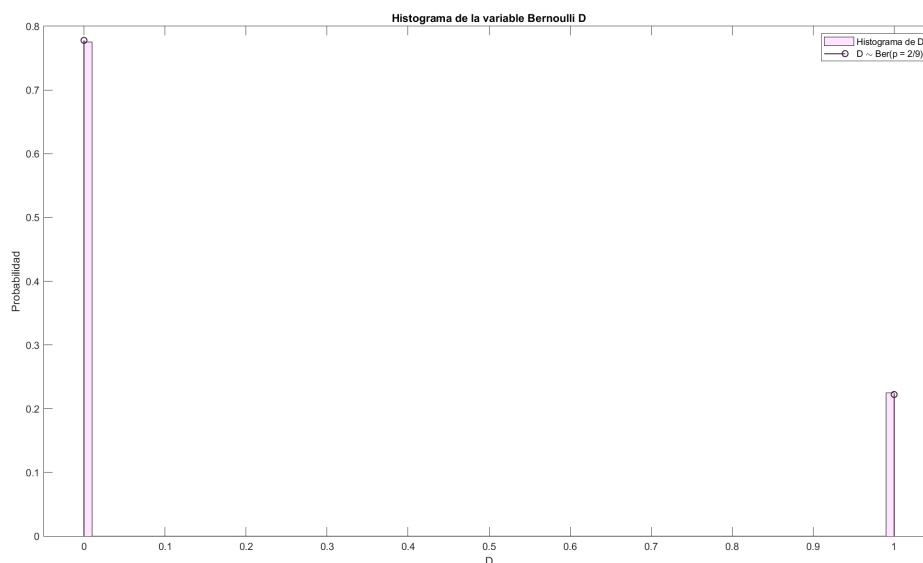


Figura 9: Histograma de la variable Bernoulli

Al igual que lo realizado en los anteriores puntos, se compara este histograma con la función de probabilidad teórica. A partir de esto, se puede observar si lo obtenido en la simulación se corresponde con las probabilidades esperadas teóricamente, que fueron calculadas en el punto anterior.

Como se puede ver en la Fig. 9, la *PMF* de la distribución Bernoulli mencionada anteriormente ($X \sim \text{Ber}(p = \frac{2}{9})$) y el histograma, convergen a valores cercanos. A fin de analizarlos, se calcula el error relativo entre el valor teórico y el obtenido a partir de las simulaciones en el histograma.

$$\% \epsilon_p = \frac{|p_{\text{his}} - p_{\text{teo}}|}{p_{\text{teo}}} \times 100 = \frac{|0,2253 - \frac{2}{9}|}{\frac{2}{9}} \times 100 \approx 1,4\%$$

$$\% \epsilon_q = \frac{|q_{\text{his}} - q_{\text{teo}}|}{q_{\text{teo}}} \times 100 = \frac{|0,7747 - \frac{7}{9}|}{\frac{7}{9}} \times 100 \approx 0,4\%$$

Por lo tanto, al comparar los resultados empíricos con teóricas, se puede decir que la diferencia es casi despreciable, pues se trata de un porcentaje menor al 2 % en ambos casos. De esta forma, se puede decir que para 10000 realizaciones, el resultado converge al valor teórico.

4.3. Modelo Probabilístico de n Lanzamientos de Dados

Suponga que ahora deseamos saber el puntaje total en n lanzamientos (es decir, la cantidad de éxitos por turno). Determine qué distribución es adecuada para modelar el puntaje obtenido por turno y defina la expresión de su función de probabilidad.

En este ejercicio, se modelaron la cantidad de éxitos en n lanzamientos, donde un éxito se define como el evento en el cual la suma de los dos dados es 7 u 11, de acuerdo con el problema anterior.

Como se ha visto anteriormente, cada lanzamiento de los dos dados es un experimento de Bernoulli, dado que tiene dos posibles resultados:

- Éxito, si la suma de los dados es 7 u 11, con probabilidad $p = \frac{2}{9}$.
- Fracaso, si la suma de los dados no es 7 ni 11, con probabilidad $1 - p = \frac{7}{9}$.

Dado que interesan la cantidad de éxitos en n lanzamientos independientes, donde cada lanzamiento es un experimento de Bernoulli con la misma probabilidad de éxito, la variable aleatoria que cuenta el número de éxitos sigue una distribución Binomial, de parámetro $p = \frac{2}{9}$ y n según lo que se requiera.

La función de probabilidad de la distribución binomial sigue la forma que se ve a continuación.

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n \quad (6)$$

Donde:

- X es la variable aleatoria que representa el número de éxitos en n lanzamientos.
- n es el número total de lanzamientos.
- k es la cantidad de éxitos en los n lanzamientos.
- $p = \frac{2}{9}$ es la probabilidad de éxito en cada lanzamiento (Bernoulli).

Luego, teniendo en cuenta que $p = \frac{2}{9}$, según 6 la función de probabilidad que modela este experimento en particular tendrá la siguiente forma.

$$P(X = k) = \binom{n}{k} \left(\frac{2}{9}\right)^k \left(\frac{7}{9}\right)^{n-k}, \quad k = 0, 1, 2, \dots, n$$

4.4. Simulación de n Lanzamientos de Dados

Genere 10000 realizaciones simulando en cada una $n = 20$ lanzamientos para contabilizar el puntaje en cada turno (de acuerdo a la variable aleatoria propuesta en el punto anterior). Realice un histograma y compárelo con la función de probabilidad teórica para dicha variable.

Para evaluar el comportamiento de la variable aleatoria que cuenta la cantidad de éxitos en $n = 20$ lanzamientos, se genera una simulación con 10000 realizaciones. En cada una, se simulan 20 lanzamientos de los dados y se cuentan cuántos éxitos se obtienen.

La variable aleatoria correspondiente sigue una distribución binomial con parámetros $n = 20$ y $p = \frac{2}{9}$, como se indicó en el ejercicio anterior.

Habiendo aclarado esto, a fin de simular este inciso, se llevó a cabo un *script* basado en el siguiente pseudocódigo.

Pseudocódigo

1. Se definió un largo de 10000 realizaciones
2. Se definieron los bins como la raíz cuadrada del largo de las realizaciones
3. Se definió la constante turnos con valor 20
4. Se definieron dos realizaciones aleatorias con distribución uniforme (1,6)
5. Se definió una variable SumaDados
6. Se definió una matriz llamada éxitos de dimensión turnos x 10000
7. Se hicieron dos for anidados para iterar la cantidad de éxitos de cada turno, iterando SumaDados en cada turno para contar la cantidad de éxitos y fracasos
8. Se sumó la cantidad de éxitos de cada columna y se definió un vector llamado exitostotales
9. Se realizó el histograma de exitostotales y se comparó con la función de masa de probabilidad de una Bernoulli

A continuación, se presenta un histograma que muestra la distribución de los puntajes obtenidos en las 10000 simulaciones.

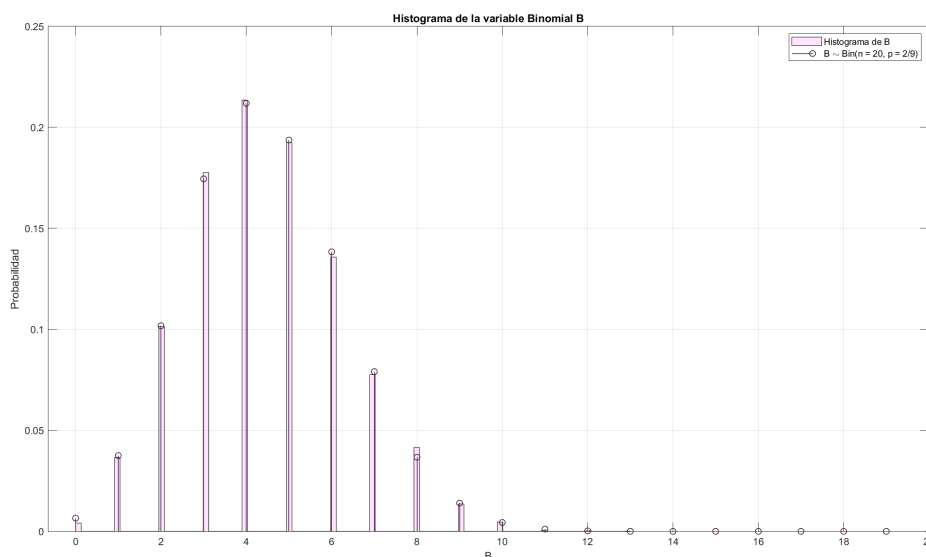


Figura 10: Histograma de la variable Binomial

El histograma muestra la frecuencia relativa de cada cantidad de éxitos obtenidos en los 10000 experimentos. Al compararlo con la función de probabilidad teórica para una variable $X \sim \text{Bin}(n = 20, p = \frac{2}{9})$, se puede observar que las frecuencias obtenidas en la simulación se corresponden con la distribución teórica.

La función de probabilidad teórica de la distribución binomial en este caso es la siguiente, cabe aclarar que sigue la forma dada de la ecuación general 6.

$$P(X = k) = \binom{20}{k} \left(\frac{2}{9}\right)^k \left(\frac{7}{9}\right)^{20-k}, \quad k = 0, 1, 2, \dots, 20$$

5. Conclusiones

El trabajo realizado demuestra cómo el Método Montecarlo es eficaz para la resolución de problemas probabilísticos complejos, permitiendo la aproximación de distribuciones a través de la Transformación de Box-Muller y la simulación de escenarios probabilísticos. Los resultados obtenidos a partir de las simulaciones concuerdan satisfactoriamente con los valores teóricos esperados, tanto en la generación de variables normales como en la resolución de integrales y el modelado de experimentos con dados. Un punto a destacar es que la precisión de los resultados mejora a medida que se incrementa el número de muestras simuladas.

Siguiendo con esta lógica, en cada uno de los ejercicios se observa claramente que lo simulado se ajusta a los valores teóricos obtenidos en el análisis previo. Esto se evidencia en todas las secciones, como en la simulación de normales estándar, la transformación lineal de gaussianas y la resolución de integrales. Asimismo, en las simulaciones de lanzamiento de dados, se obtienen errores porcentuales menores al 2 %. Todo esto, muestra la robustez del método para generar aproximaciones precisas en situaciones donde los cálculos analíticos resultan difíciles.

Por último, cabe destacar que este trabajo práctico ayudó al desarrollo de un pseudocódigo, previo a comenzar a generar el *script* en *Matlab*. De esta manera, se pudieron implementar simulaciones y aproximaciones, lo que demuestra aplicaciones prácticas y contribuye a una mejor comprensión de los conceptos teóricos aprendidos en clase.