Informe trabajo final
Probabilidad

Profesor: Mario E. Arrieta Prieto
Grupo 8

Natalia Monroy Rosas namonroyr@unal.edu.co
Carlos Enrique Nosa Guzman cnosa@unal.edu.co
Oscar Julian Rodriguez Cardenas osrodriguezc@unal.edu.co
Daniel Mauricio Rodriguez Obregon drodriguezob@unal.edu.co

Item A

La función utilizada para generar los números aleatorios es **numpy.random.rand()** del paquete Numpy, que es uno de los paquetes más conocidos para computación científica en Python. Según su documentación, ésta función crea un arreglo de un tamaño dado n y almacena muestras aleatorias de una distribución uniforme en el intervalo $[0, 1)$.
Los números generados por esta función son pseudo-aleatorios, pues sólo un computador cuántico podría generar auténticos números aleatorios. Éstos números pseudo-aleatorios comienzan con una ”semilla” (número base), la cual después se multiplica por un número grande, se le agrega un desplazamiento y luego se toma el módulo de esa suma. El número resultante se utiliza como semilla para generar el siguiente número aleatorio, y de esta manera se llena el arreglo de números aleatorios.

Histograma

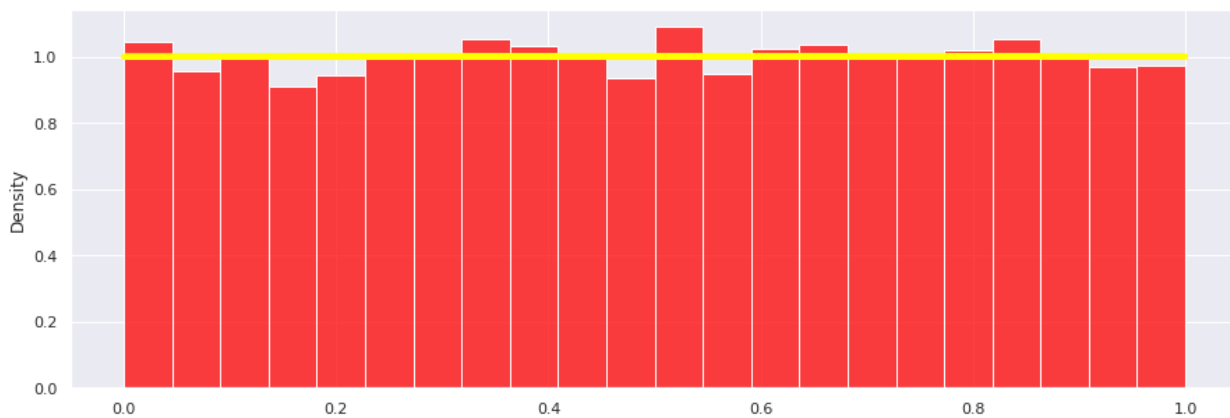


Figure 1: Histograma de los 10000 números generados

Lo ideal de una función que genere números aleatorios dentro de un rango determinado $[a, b]$ sería que cada número tuviera igual probabilidad de salir, pues son ”aleatorios”. Sabemos que una distribución uniforme es usada para describir variables continuas que tienen una probabilidad constante, es decir, modela un rango de valores con igual probabilidad. En la gráfica podemos ver la densidad de una distribución uniforme en el intervalo [0,1] en amarillo y en el histograma se puede verificar que los 10000 números generados tienen una probabilidad de salir muy semejante. Podemos concluir entonces que está función generadora de números aleatorios es bastante buena, pues se aproxima mucho a una distribución uniforme de parámetros $[0, 1]$. Si generáramos ”infinitos” números aleatorios veriamos en el histograma que todos tendrían exactamente la misma probabilidad de salir.

Item B

Función 2

Para calcular k utilizaremos que la propiedad de la función de masa de probabilidad que establece lo siguiente:

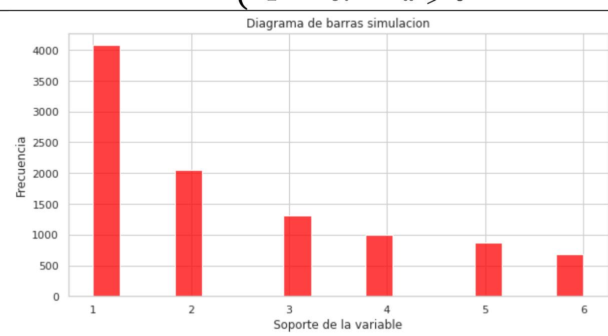
$$\sum_{x \in D_x} p(X = x) = 1$$

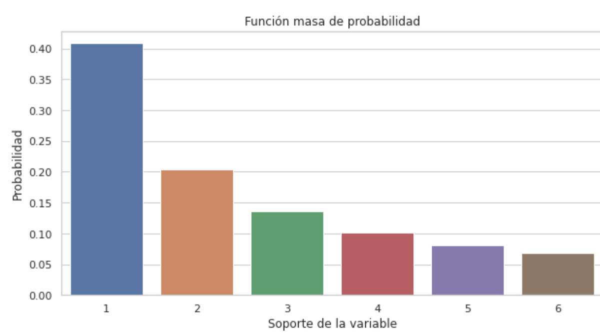
$$\frac{k}{1} + \frac{k}{2} + \frac{k}{3} + \frac{k}{4} + \frac{k}{5} + \frac{k}{6} = 1$$

$$k\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6}\right) = 1$$

$$\frac{49}{20} \cdot k = 1$$

$$k = \frac{20}{49}$$

<table>
<tr><td>Función asignada</td><td>2</td></tr>
<tr><td>Función de masa de probabilidad</td><td>$$P_X(x) = \begin{cases} \frac{20}{49}x & si \quad x = 1, 2, 3, 4, 5, 6 \\ 0 & e.o.c \end{cases}$$</td></tr>
<tr><td>Función de distribución</td><td>$$F_X(x) = \begin{cases} 0 & si \quad x < 1 \\ \frac{20}{49} & si \quad 1 \leqslant x < 2 \\ \frac{30}{49} & si \quad 2 \leqslant x < 3 \\ \frac{110}{147} & si \quad 3 \leqslant x < 4 \\ \frac{125}{147} & si \quad 4 \leqslant x < 5 \\ \frac{137}{147} & si \quad 5 \leqslant x < 6 \\ 1 & si \quad x \geqslant 6 \end{cases}$$</td></tr>
<tr><td>Resultados (10000 simulaciones)</td><td>


</td></tr>
</table>

Función 4

Para calcular k utilizaremos que la propiedad de la función de masa de probabilidad que establece lo siguiente:

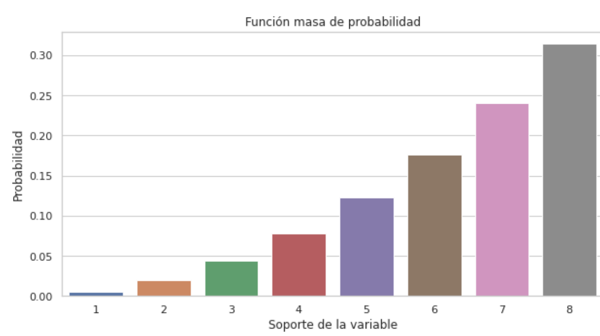
$$\sum_{x \in D_x} p(X = x) = 1$$

$$k(1)^2 + k(2)^2 + k(3)^2 + k(4)^2 + k(5)^2 + k(6)^2 + k(7)^2 + k(8)^2 = 1$$

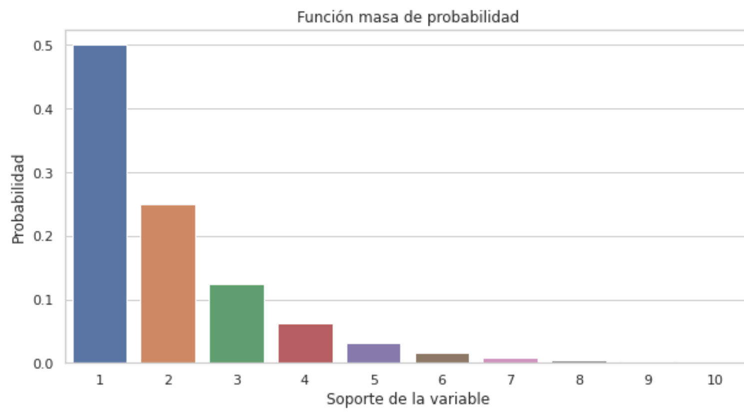
$$k((1)^2 + (2)^2 + (3)^2 + (4)^2 + (5)^2 + (6)^2 + (7)^2 + (8)^2) = 1$$

$$k(204) = 1$$

$$k = \frac{1}{204}$$

Función asignada	4
Función de masa de probabilidad	$P_X\,(x) = \begin{cases} \frac{1}{204}x^2 & si \quad x = 1,2,3,4,5,6,7,8 \\ 0 & e.o.c \end{cases}$
Función de distribución	$F_X\,(x) = \begin{cases} 0 & si \quad x < 1 \\ \frac{1}{204} & si \quad 1 \leqslant x < 2 \\ \frac{5}{204} & si \quad 2 \leqslant x < 3 \\ \frac{14}{204} & si \quad 3 \leqslant x < 4 \\ \frac{30}{204} & si \quad 4 \leqslant x < 5 \\ \frac{55}{204} & si \quad 5 \leqslant x < 6 \\ \frac{91}{204} & si \quad 6 \leqslant x < 7 \\ \frac{140}{204} & si \quad 7 \leqslant x < 8 \\ 1 & si \quad x \geqslant 8 \end{cases}$
Resultados (10000 simulaciones)	

Función 6

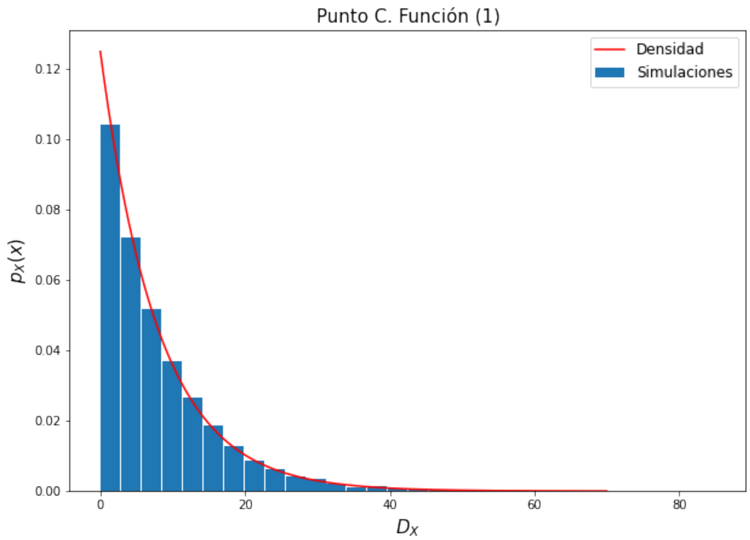
Función asignada	6
Función de masa de probabilidad	$P_X(x) = \begin{cases} (0.5)^{x-1}0.5 & si \quad x = 1, 2, 3, \ldots \\ 0 & e.o.c \end{cases}$
Función de distribución	$F_X(x) = \begin{cases} 1 - (0.5)^{\lfloor x \rfloor} & si \quad x \geq 0 \\ 0 & e.o.c \end{cases}$
Resultados (10000 simulaciones)	 

Función masa de probabilidad teórica vs Simulación

Después de realizar las tres simulaciones y comparando con sus respectivas funciones de masa de probabilidad, es posible concluir que las simulaciones de las variables fueron realizadas correctamente, pues la frecuencia con que salieron los valores del soporte de las variables corresponden con sus probabilidades dadas por su función de masa de probabilidad respectiva. Es posible hacer una comparación con la simulación realizada en el item A, pues también se realizó una simulación de una variable que seguía una distribución uniforme y al sobreponer la función de densidad de la variable, que en este caso al ser variables discretas se trataría de una función de masa de probabilidad, se observó que coincidían notablemente.

Item C

Función 1

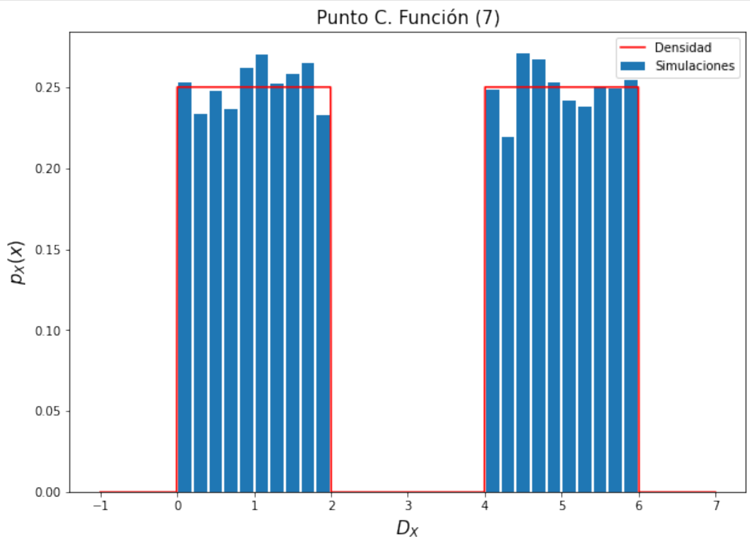
Función asignada	(1) $X \sim Exp(\lambda = \frac{1}{\theta} = \frac{1}{8})$
Función de densidad de probabilidad	$f_X(x) = \begin{cases} \frac{1}{8}e^{-\frac{1}{8}x} & \text{si } x \geq 0, \\ 0 & \text{e.o.c.} \end{cases}$
Función de distribución	$F_X(x) = \begin{cases} 1 - e^{-\frac{1}{8}x} & \text{si } x \geq 0, \\ 0 & \text{si } x < 0 \end{cases}$
Resultados (10000 simulaciones)	

Función 7

Para calcular la constante $k \in \mathbb{R}$ tengamos en cuenta que $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$, de esta forma,

$$1 = \int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = \int_0^2 k\mathrm{d}x + \int_4^6 k\mathrm{d}x = k(x|_0^2 + x|_4^6) = k(2+2) = k \cdot 4$$

por lo tanto, $k = \frac{1}{4}$.

Función asignada	(7) $f_X(x) = \begin{cases} k & 0 \leq x \leq 2, \\ k & 4 \leq x \leq 6, \\ 0 & \text{e.o.c.} \end{cases}$
Función de densidad de probabilidad	$f_X(x) = \begin{cases} \frac{1}{4} & 0 \leq x \leq 2, \\ \frac{1}{4} & 4 \leq x \leq 6, \\ 0 & \text{e.o.c.} \end{cases}$
Función de distribución	$F_X(x) = \begin{cases} 0 & x < 0, \\ \frac{x}{4} & 0 \leq x \leq 2, \\ \frac{1}{2} & 2 \leq x \leq 4, \\ \frac{1}{2} + \frac{x-4}{4} & 4 \leq x \leq 6, \\ 1 & 6 < x \end{cases}$
Resultados (10000 simulaciones)	

Función 8

En primer lugar, calculemos las constantes $k_1, k_2, k_3 \in \mathbb{R}$, basados en los hechos

a. $\int_{-\infty}^{\infty} f_X(x)\mathrm{d}x = 1$

b. $f_X(10^-) = f_X(10)$

c. $f_X(30) = 0$

así,

- De (a) tenemos que

$$\begin{aligned} 1 = \int_{-\infty}^{\infty} f_X(x)\mathrm{d}x &= \int_0^{10} (k_1 x)\ \mathrm{d}x + \int_{10}^{30} (k_3 + k_2 x)\ \mathrm{d}x \\ &= \frac{k_1}{2}(x^2)|_{x=0}^{x=10} + (k_3 x + \frac{k_2}{2}x^2)|_{x=10}^{x=30} \\ &= 50k_1 + 30k_3 + 450k_2 - 10k_3 - 50k_2 \\ &= 50k_1 + 400k_2 + 20k_3 \end{aligned}$$

  es decir, $50k_1 + 400k_2 + 20k_3 = 1$.

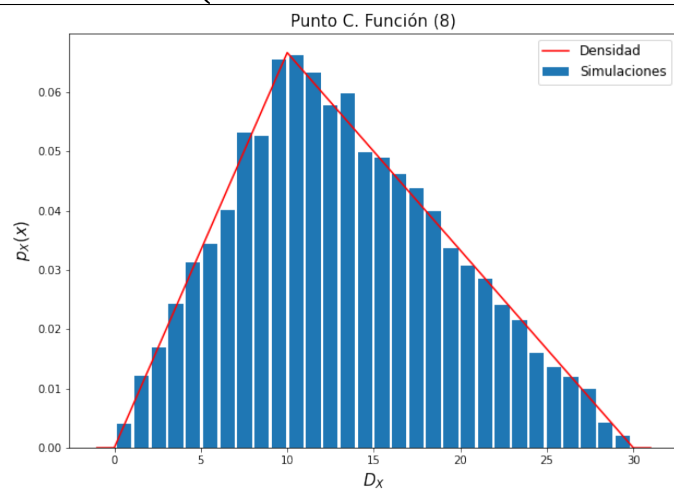
- De (b) podemos inferir que $k_3 + 10k_2 = 10k_1$, es decir, $-10k_1 + 10k_2 + k_3 = 0$
- De (c) se infiere $30k_2 + k_3 = 0$.

Por ende tenemos el sistema de ecuaciones,

$$\begin{cases} 50k_1 + 400k_2 + 20k_3 & = 1 \\ -10k_1 + 10k_2 + k_3 & = 0 \\ 0k_1 + 30k_2 + k_3 & = 0 \end{cases}$$

el cual tiene solución única

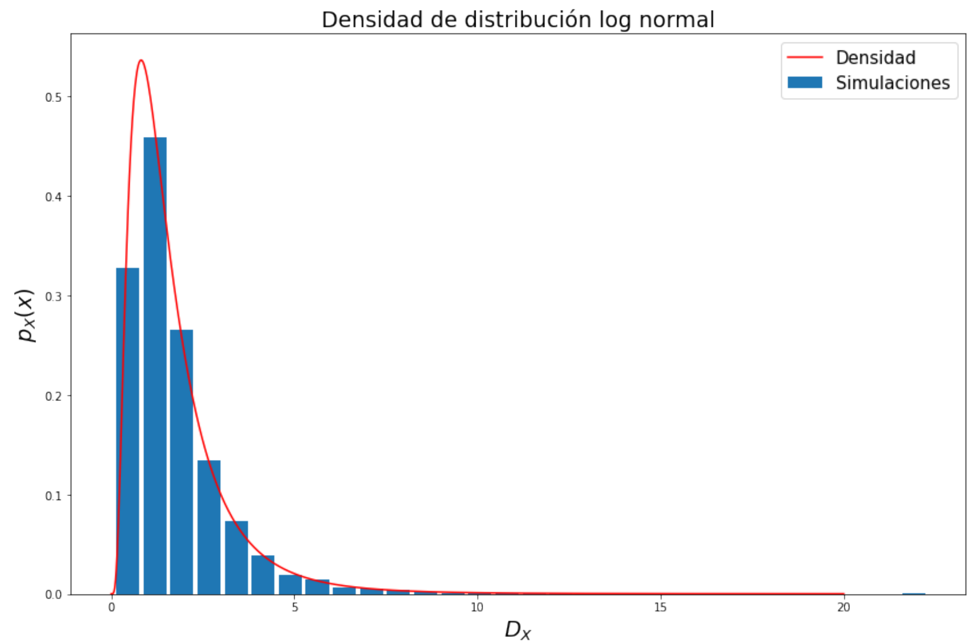
$$k_1 = \frac{1}{150} \qquad k_2 = \frac{-1}{300} \qquad k_3 = \frac{1}{10}$$

Función asignada	(8) $f_X(x) = \begin{cases} k_1 x & 0 \leq x \leq 10, \\ k_3 + k_2 x & 10 \leq x \leq 30, \quad f_X(10^-) = f_X(10) \text{ y } f_X(30) = 0 \\ 0 & \text{e.o.c.} \end{cases}$
Función de densidad de probabilidad	$f_X(x) = \begin{cases} \frac{1}{150}x & 0 \leq x \leq 10, \\ \frac{1}{10} - \frac{1}{300}x & 10 \leq x \leq 30, \\ 0 & \text{e.o.c.} \end{cases}$
Función de distribución	$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{300}x^2 & 0 \leq x \leq 10, \\ \frac{1}{10}x - \frac{1}{600}x^2 - \frac{1}{2} & 10 \leq x \leq 30, \\ 1 & 30 < x \end{cases}$
Resultados (10000 simulaciones)	

Item D

Ejercicio 1

a) Teniendo una variable aleatoria $X \sim N(\mu = 0.3, \sigma^2 = 0.5)$ se simuló una variable aleatoria $Y = e^X$ a partir de la inversa de la función de distribución de la variable X con 10000 simulaciones.



b) La función generadora de momentos de X equivale a la sucesión de momentos ordinarios de Y para todo t en un intervalo $-\alpha < t < \alpha$, dónde $\alpha > 0$:

$$\mu_t := E[Y^t] := E[e^{tX}] := m_X(t), \forall t \in (-\alpha, \alpha)$$

Sabemos que X corresponde a una Distribución Normal cuya f.g.m. es:

$$m_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Utilizaremos esta ecuación para hallar los momentos teóricos de Y, con $\mu = 0.3$ y $\sigma^2 = 0.5$:

- $E[Y] := m_X(1) = e^{0.3+\frac{1}{2}(0.5)} = e^{\ 0.3+\frac{1}{4}} = e^{0.55} \approx 1.733$
- $E[Y^2] := m_X(2) = e^{(0.3)(2)+\frac{1}{2}(0.5)(2^2)} = e^{0.4+1} = e^{1.4} \approx 4.055$
- $E[Y^5] := m_X(5) = e^{(0.3)(5)+\frac{1}{2}(0.5)(5^2)} = e^{1.5+\frac{25}{4}} = e^{7.75} \approx 2321.572$

c) Calculando los momentos $E[Y]$, $E[Y^2]$ y $E[Y^5]$ a través de aproximarlos por polinómios de Taylor, dónde para cualquier variable aleatoria Z que esté definida como $Z = g(X)$ se tiene que su momento es:

$$E[Z] = E[g(X)] \approx g(E[X]) + \frac{g"(E[x])}{2} Var(X)$$

En particular nos queda que $Y = g(X) = e^x$, $Y^2 = g(X)^2 = e^{2x}$ y $Y^5 = g(X)^5 = e^{5x}$, derivando dos veces estas ecuaciones obtenemos: $g_1"(X) = e^x$, $g_2"(X) = 4e^{2x}$ y $g_5"(X) = 25e^{5x}$.

Reemplazando en el polinomio de Taylor, con $Var(X) = \sigma^2 = 0.5$, se obtiene:

$$E[Y] = e^{0.3} + \frac{e^{0.3}}{2} \cdot 0.5 = 1.687$$
$$E[Y^2] = e^{2\cdot(0.3)} + \frac{4 \cdot e^{2\cdot(0.3)}}{2} \cdot 0.5 = 3.644$$
$$E[Y^5] = e^{5\cdot(0.3)} + \frac{25 \cdot e^{5\cdot(0.3)}}{2} \cdot 0.5 = 32.492$$

d) Esta vez para calcular los momentos se va a tomar la media obtenida a partir de la simulación con 10, 100, 1000 y 10000 términos. Para calcularlos mediante este método se tiene:

$$E[Y] \approx \frac{1}{r}\sum_{i=1}^{r} Y = \frac{1}{r}\sum_{i=1}^{r} g(x_i)$$

$$E[Y^2] \approx \frac{1}{r}\sum_{i=1}^{r} Y^2 = \frac{1}{r}\sum_{i=1}^{r} g(x_i)^2$$

$$E[Y^5] \approx \frac{1}{r}\sum_{i=1}^{r} Y^5 = \frac{1}{r}\sum_{i=1}^{r} g(x_i)^5$$

Donde x_i son las simuilaciones de la Variable $X \sim N(\mu = 0.3, \sigma^2 = 0.5)$.

Los resultados obtenidos para los distintos valores de r son:

	$E[Y]$	$E[Y^2]$	$E[Y^5]$
r=10	2.429	11.783	6065.931
r=100	1.782	4.974	581.493
r=1000	1.708	4.918	1169.209
r=10000	1.699	4.669	2173.725

e) Comparando los valores teóricos obtenidos en el punto, con los valores en los puntos c y d, se tienen los siguientes errores porcentuales:

- Método de polinomios de Taylor

$E[Y]$	$E[Y^2]$	$E[Y^5]$
2.64%	26.42%	98.6%

- Método de la media

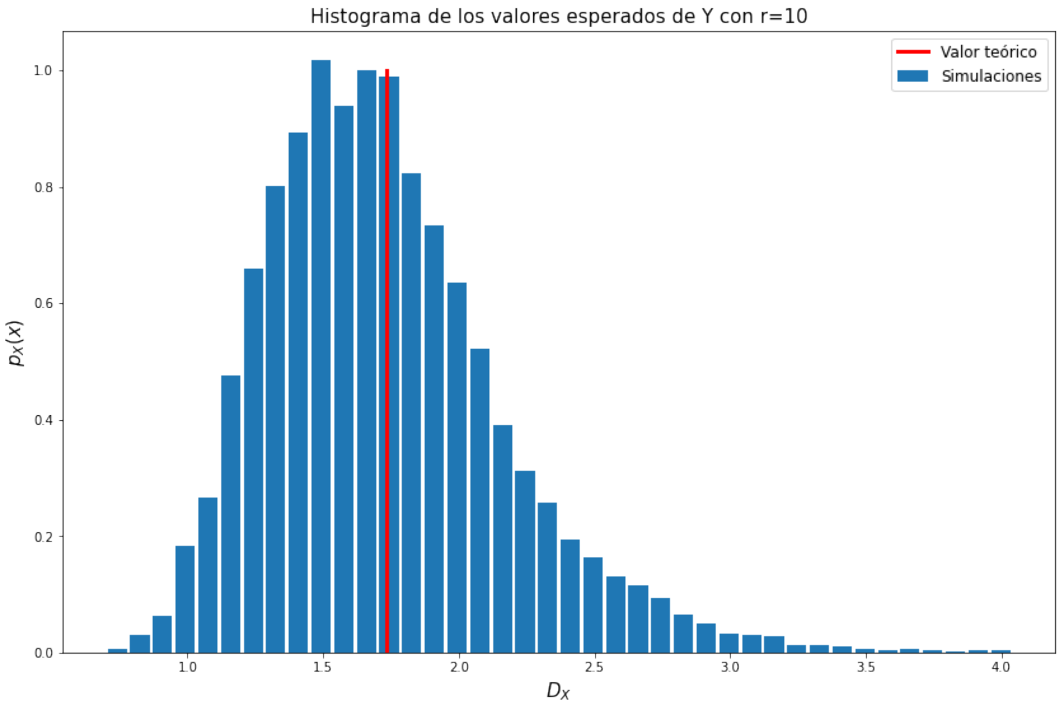
Error % con	$E[Y]$	$E[Y^2]$	$E[Y^5]$
r=10	40.19%	137.9%	161.28%
r=100	2.86%	0.43%	74.952%
r=1000	1.43%	0.68%	49.63%
r=10000	1.96%	5.71%	6.36%

Se puede observar que el método de Taylor no es muy adecuado para la aproximación del valor esperado porque disminuye notablemente su precisión a medida que aumenta el orden de los momentos. Por otra parte, el método 3 permite obtener valores cercanos a los teóricos, aunque su exactitud también disminuye a medida que aumenta el orden de las potencias. En cuánto al impacto de r, observamos que aunque no siempre hay una relación inversamente proporcional estricta entre el valor de r y el error, a medida que aumenta r el error tiende a disminuir y por lo tanto, nos acercamos más al valor esperado. Ésto ocurre por la *Ley Fuerte de los Grandes Números*, la cual afirma que si se repite un experimento de forma independiente una cantidad considerablemente grande de veces, al promediar los resultados se obtiene un valor muy cercano al valor esperdo.

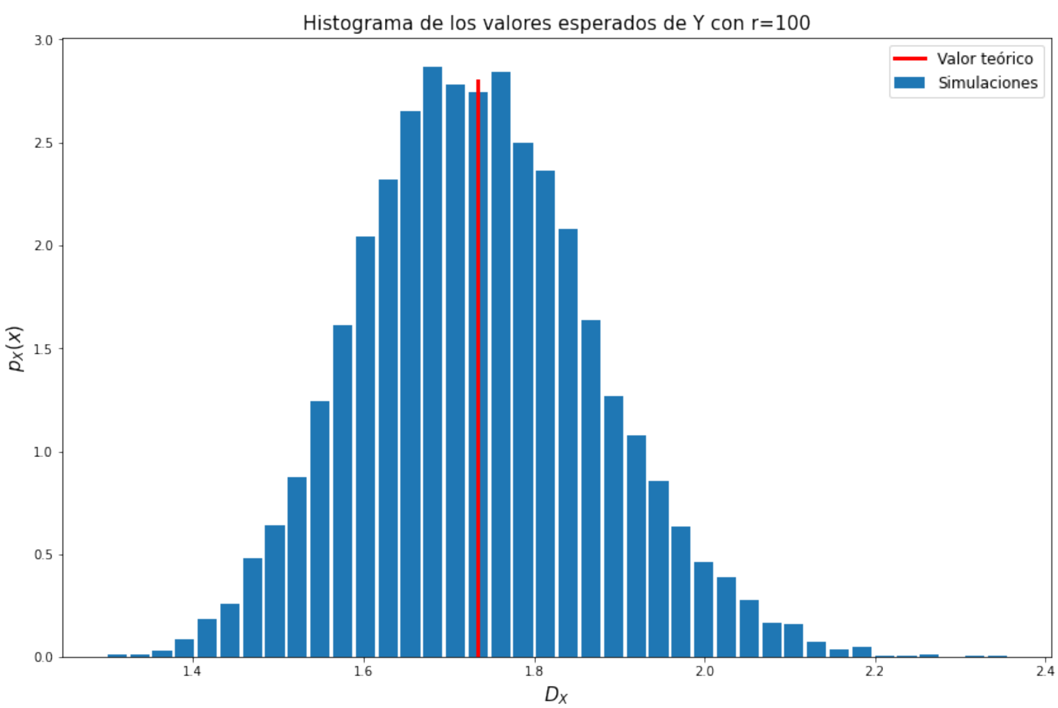
Ejercicio 2

A partir de la Variable aleatoria $X \sim N(\mu = 0.3, \sigma^2 = 0.5)$ se repite el método del valor medio para calcular el valor esperado de la variable aleatoria $Y = e^X$ pero esta vez repitiendo el cálculo 10000 veces y con ello generar un histograma de la frecuencia de los valores esperados.

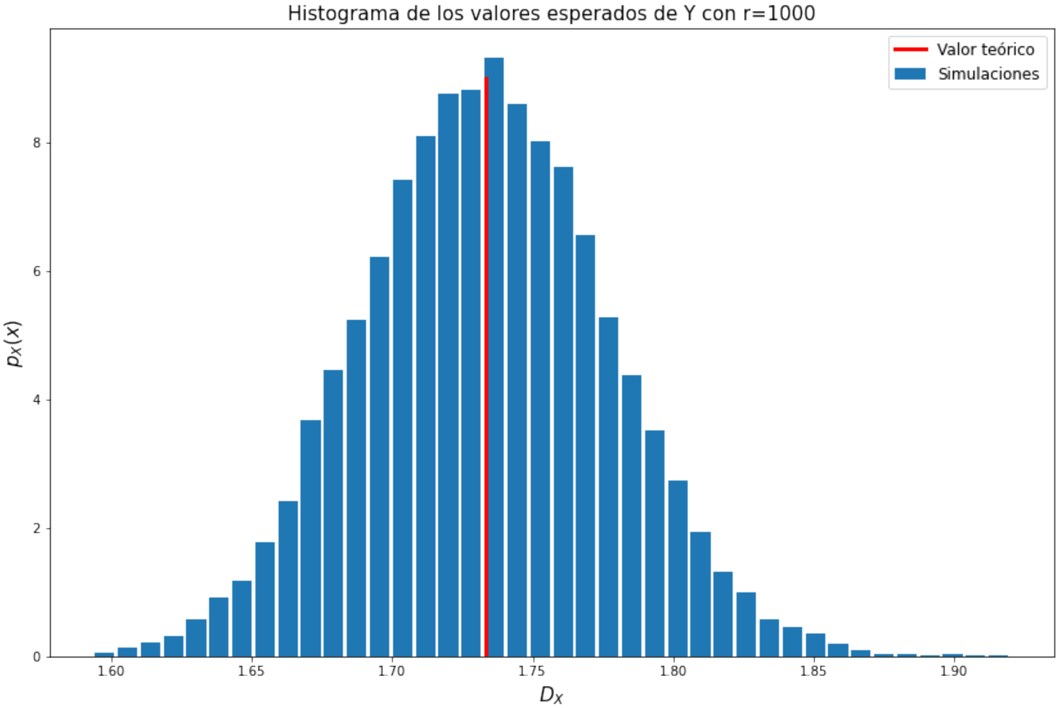
a) Tomando un r=10 para realizar la aproximación del valor esperado de Y se obtiene el siguiente histograma:



b) Incrementando el valor de r a 100 se obtiene el siguiente histograma:



c) Por último, aumentando el valor de r hasta 1000 se obtiene el siguiente histograma:

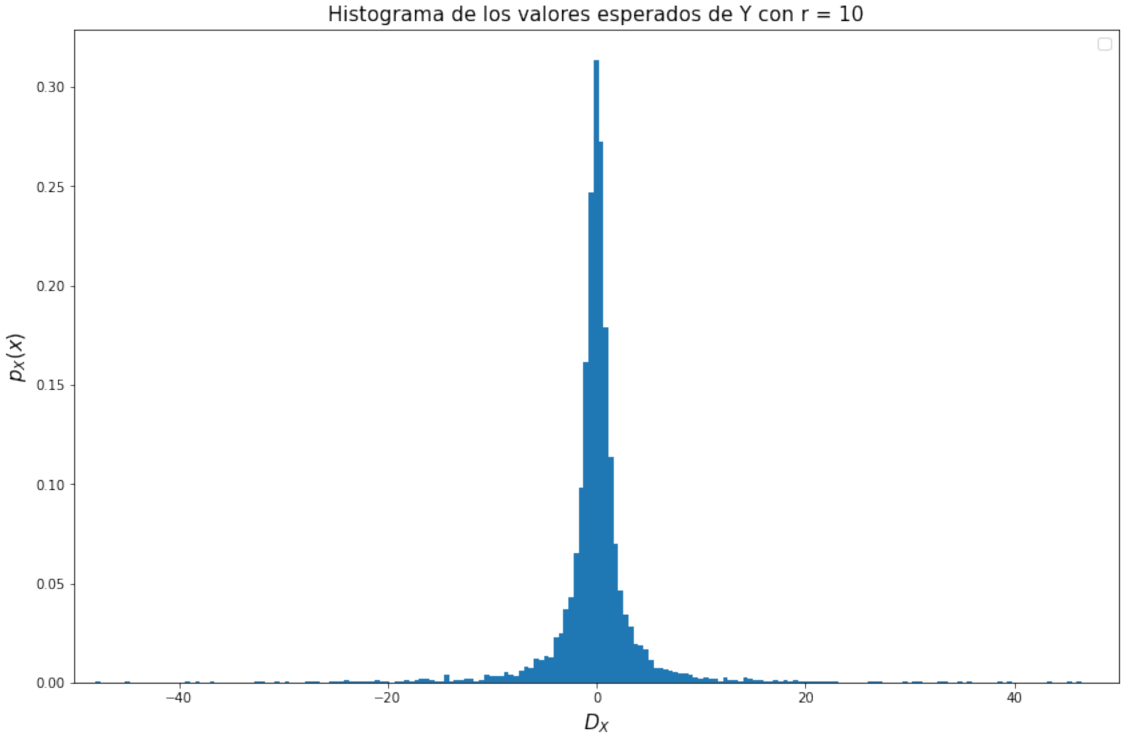


d) Se observa que a medida que se aumenta el valor de r, el histograma tiende a acercarce cada vez más al comportamiento de una variable aleatoria $X \sim N(\mu, \sigma^)$, donde μ es el valor esperado de la variable aleatoria $Y = e^X$. El resultado teórico que justifica esta conclusión es el Teorema del Límite Central. El TLC establece que la distribución muestral de las medias de los resultados se aproxima a una distribución normal a medida que aumenta el tamaño de la muestra, independientemente de la forma de la distribución de la población. Este hecho es especialmente cierto para tamaños de muestra superiores a 30.

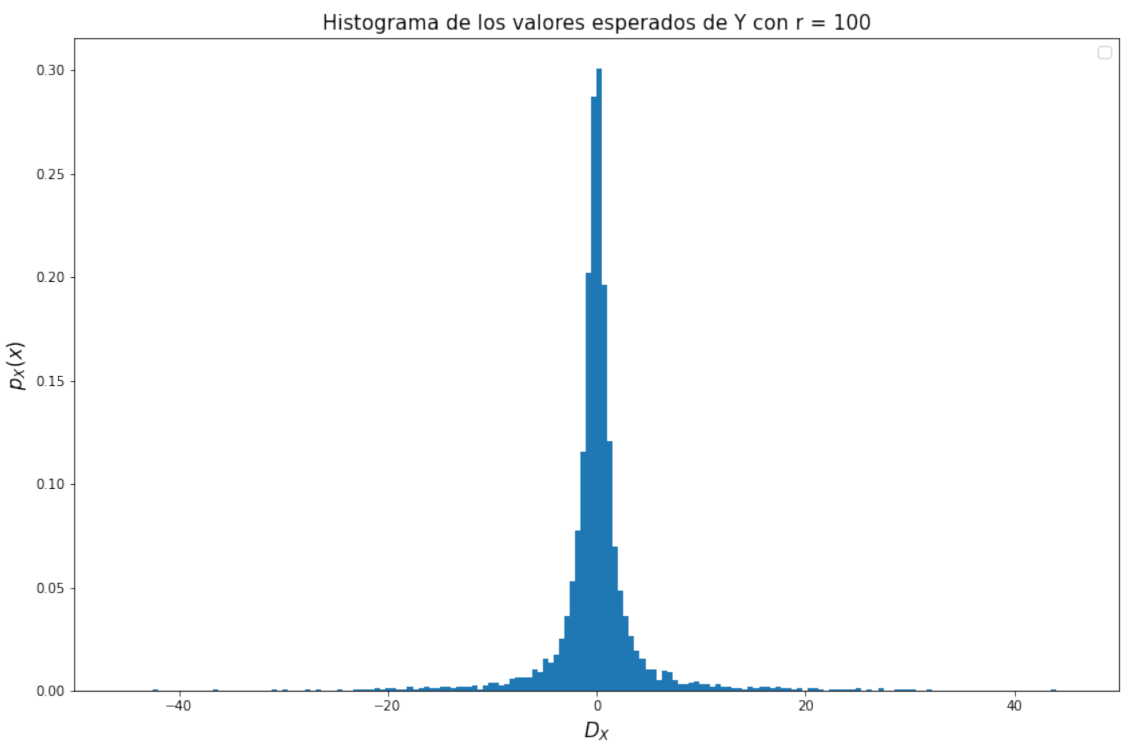
Bonus

Simulación de la Variable aleatoria $Y = \frac{1}{X}$ con X Cauchy ($\alpha = 0$, $\beta = 1$). Las realizaciones de X se realizaron con la función $ss.cauchy.rvs(size = r)$ de la libreria Scipy, cuya función $cauchy.rvs$ genera una lista de tamaño r de números aleatorios que pertenecen a la variable aleatoria de Cauchy. A continuación, se presentan los histogramas obtenidos para r=10, 100 y 1000.

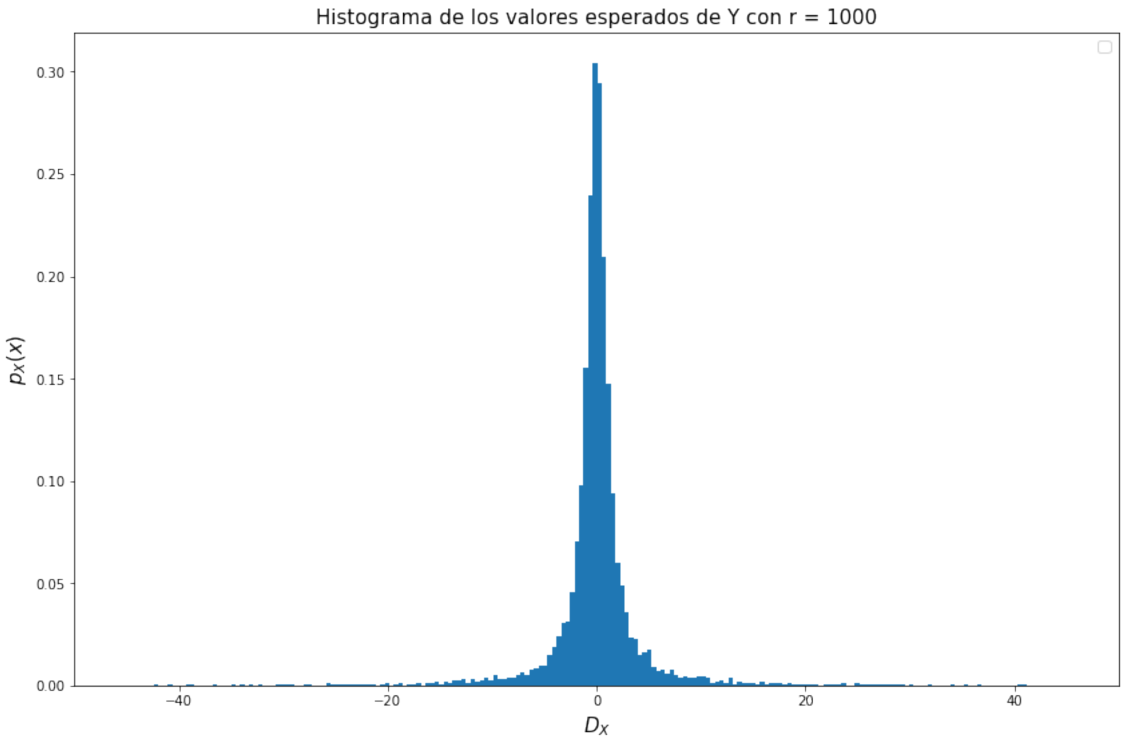
a) r=10



b) r=100



c) r=1000



d) La forma del histograma se mantiene constante a medida que aumenta r. Se observa también que la dispersión, teniendo el cuenta el rango de los resultados, es bastante alta. Éstas son características propias de la distribución de Cauchy, que presenta colas muy largas, lo cual se conoce como "Heavy Tail Distribution". Ésto hace que al añadir más valores a la suma, un valor realmente extremo sea lo suficientemente probable como para compensar la división por un denominador mayor al tomar la media. Además, la distribución de Cauchy NO cumple con la *Ley de los Grandes Números*, pues su integral no converge y por lo tanto, su media o valor esperado no existen.

Item E

Simulación de Montecarlo
Para generar las realizaciones de X en los ejercicios de esta sección se utilizó la función *random.uniform(a,b)* de la librería random de python que implementa generadores de números pseudo-aleatorios para varias distribuciones. La aproximación del valor de las integrales se realizó con el método 3 de la sección anterior teniendo en cuenta que:

$$\int_a^b h(x)\,dx = (b-a)\int_a^b h(x)\frac{1}{b-a}\,dx = (b-a)\cdot E[h(X)]$$

Así el valor de la integral definida equivale a el valor esperado $E[X]$ por la diferencia entre los límites b y a. Luego, la aproximación númerica del valor esperado se realiza mediante la sumatoria:

$$E[h(X)] \approx \frac{1}{r}\sum_{i=1}^{r} h(x_i)$$

Dónde r son los r números aleatorios generados en el intervalo [a,b] por la funcion *random.uniform(a,b)*. A continuación se presenta la tabla con los resultados obtenidos para cada integral con los tres valores sugeridos para r.

- **Aproximaciones numéricas**:

orden	$\int_8^{13} e^{-x^2}\,dx$	$\int_5^{10} \frac{\sqrt{x}\sin x}{(x+1)^3}\,dx$	$\int_0^{15} \frac{x^3}{e^{-\cos x}+\sqrt{x}}\,dx$
r=10	$6.588 \cdot 10^{72}$	0.00448	3156.412
r=100	$2.275 \cdot 10^{71}$	0.00197	2807.411
r=1000	$8.706 \cdot 10^{71}$	0.00104	2830.127
r=10000	$1.010 \cdot 10^{72}$	0.00153	2842.953
Valor Software	$9.596 \cdot 10^{71}$	0.0016	2876.633

- **Errores porcentuales**:

orden	$\int_8^{13} e^{-x^2}\,dx$	$\int_5^{10} \frac{\sqrt{x}\sin x}{(x+1)^3}\,dx$	$\int_0^{15} \frac{x^3}{e^{-\cos x}+\sqrt{x}}\,dx$
r=10	100%	100%	9.726%
r=100	76.295%	23.377%	2.406%
r=1000	9.278%	35.256%	1.617%
r=10000	5.257%	4.387%	1.171%

Se observa que a medida que aumenta el valor de r se incrementa la precisión de las aproximaciones de los valores de cada integral. Lo cual ocurre, de igual forma que en el item D Ej. 1, por la *Ley Fuerte de los grandes números*.

Nota: Las aproximaciones númericas se realizaron con Wolfram Alpha, Mathemathica, e integral-calculator.com. Los respectivos links y soportes se encuentran en el Notebook.

References

[1] Docs.python.org. 2021. *random — Generate pseudo-random numbers*. Python 3.9.6 documentation.[online] Available at: http://docs.python.org/3/library/random.html [Consultado en agosto 11 de 2021].

[2] Probabilitycourse.com. 2021. *Law of Large Numbers*. [online] Available at: https://www.probabilitycourse.com/chapter7/7_1_1_law_of_large_numbers.php [Consultado en agosto 10 de 2021].

[3] Sphweb.bumc.bu.edu. 2021. *Central Limit Theorem*. [online] Available at: https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/BS704_Probability12.html [Consultado en agosto 11 de 2021].

[4] W., pig, F., Zare, D. and Sarwate, D., 2021. *Why does the Cauchy distribution have no mean?*. [online] Cross Validated. Available at: https://stats.stackexchange.com/questions/36027/why-does-the-cauchy-distribution-have-no-mean [Consultado en agosto 13 de 2021].