INFORME TAREA DISTRIBUCIONES

Martín Palacios Undurraga

La función de este informe es presentar los resultados obtenidos dada la realización de la tarea “RANDOM VARIABLE SIMULATION” del curso de Derivados Avanzados. Por razones prácticas no pegare imágenes de cada pregunta sino únicamente de los códigos y outputs.

Por temas de orden de proyecto, al programar trabajo con un archivo principal (“main” o “script”) que llama los scripts individuales para cada pregunta. Para una correcta funcionalidad del código se recomienda correr los códigos con la función “step” en vez de “run”. De esta manera los scripts corren de manera individual y entregaran los outputs deseados sin importar la versión de Matlab que se esté usando.

Preguntas:

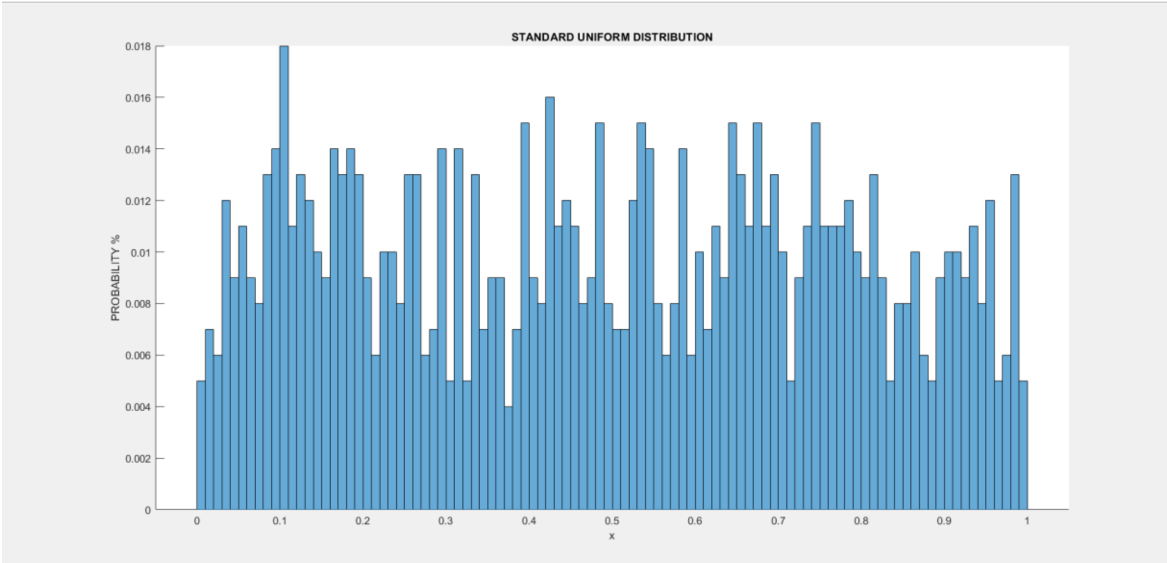
La pregunta 1 solicita generar 1000 variables aleatorias con distribución uniforme estándar entre los valores 0 y 1 y luego graficarlas en un histograma de distribución de probabilidad empírica. El código utilizado para MATLAB fue el siguiente:

Editor - C:\Users\marti\OneDrive - Universidad Adolfo Ibanez\Desktop\Universidad\Derivados Avanzados\Tarea Distribuciones\Tarea aparte\QUESTION1.m

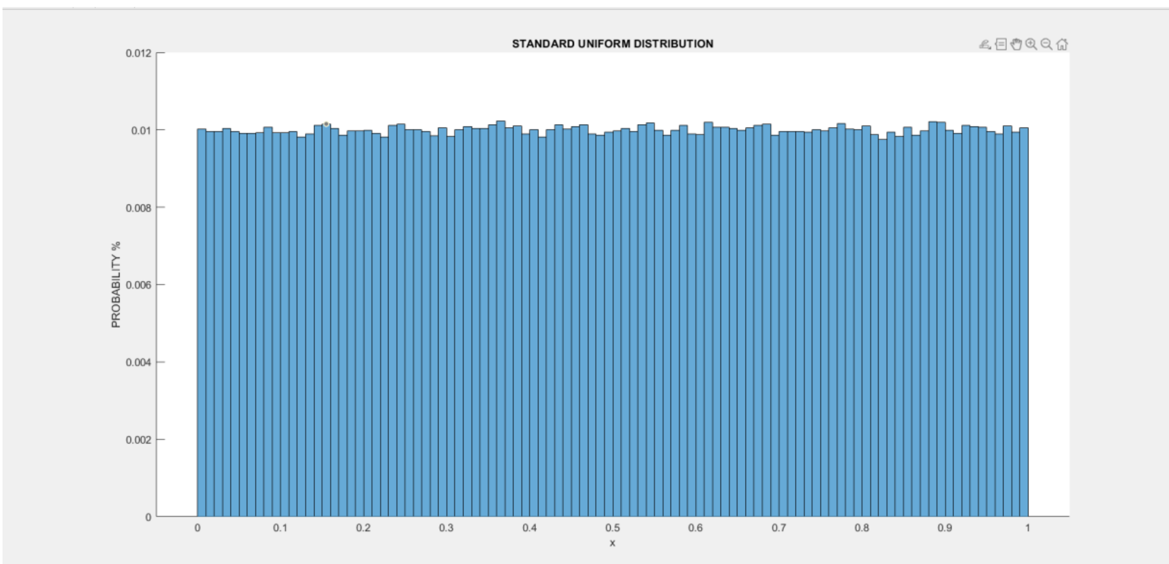
script.m | QUESTION1.m | QUESTION2.m | QUESTION3.m | QUESTION4.m | Q5FUNCTION_q4.m

```
% QUESTION 1 :

clc
clear

% DATA :
N = 1000;
Q = 100;

%  SEED :
rng(0);

% QUESTION 1 :
x_q1 = rand(N,1);

hold on;
h_q1 = histogram(x_q1, Q, "Normalization","probability");
title('STANDARD UNIFORM DISTRIBUTION');
xlabel('x');
ylabel('PROBABILITY %');
hold off;
```

El output generado por este script se presenta a continuación:



Como se puede ver la probabilidad se distribuye de manera bastante uniforme, pero por si quedan dudas uno puede aumentar N a 100000 por ejemplo, esto devolverá el siguiente output:

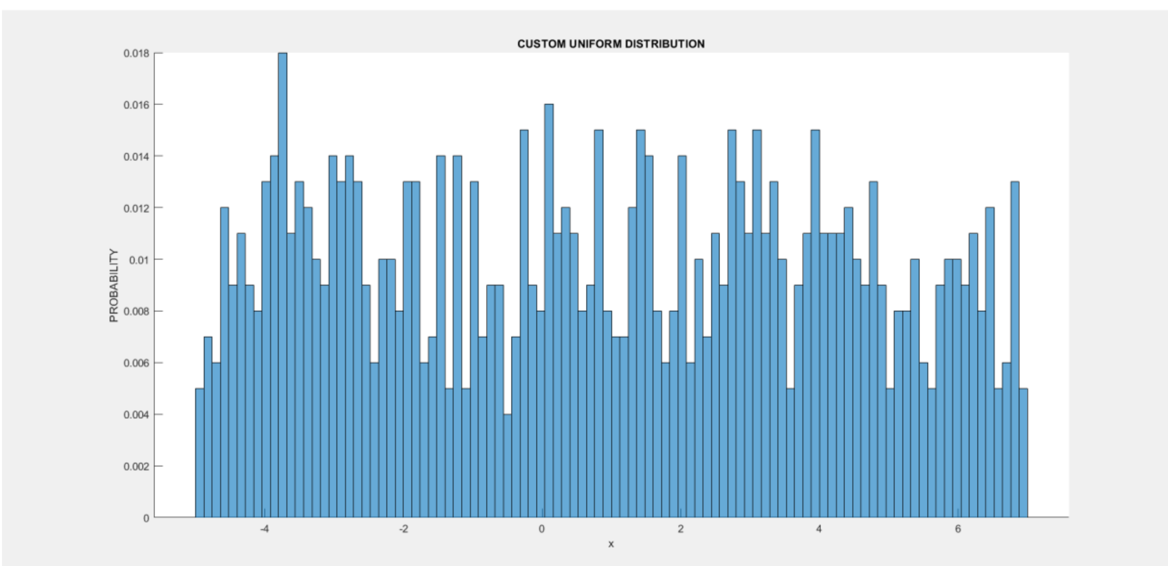


Como se puede observar, al aumentar N se hace mucho mas notorio que hay una distribución uniforme entre los valores 0 y 1 y que cada punto tiende a tener la misma probabilidad individual dentro de la distribución.

La pregunta 2 es bastante similar a la 1 con la diferencia que se nos pide que los valores estén entre -5 y 7. Para esto generé el siguiente script:

```
% QUESTION 2 :

clc
clear

% DATA :
N = 1000;
Q = 100;

%  SEED :
rng(0);

min = -5;
max = 7;

x_q1 = rand(N,1);
x_q2 = min + x_q1 * (max-min);

hold on;
h_q2 = histogram(x_q2, Q,"Normalization","probability");
title('CUSTOM UNIFORM DISTRIBUTION');
xlabel('x');
ylabel('PROBABILITY');
hold off;
```

Para contestar esta pregunta editamos levemente la variable x_q1 generada en la pregunta anterior para que cumpla con los requisitos. Este código devuelve el siguiente output:



Al igual que en la pregunta anterior (y esto aplica para todas las preguntas que siguen), si es que uno aumenta el valor de N (numero de observaciones de la variable aleatoria) , la distribución de probabilidad empírica se acerca cada vez más a la distribución teórica, esto se debe a la aplicación de la ley de los grandes números.

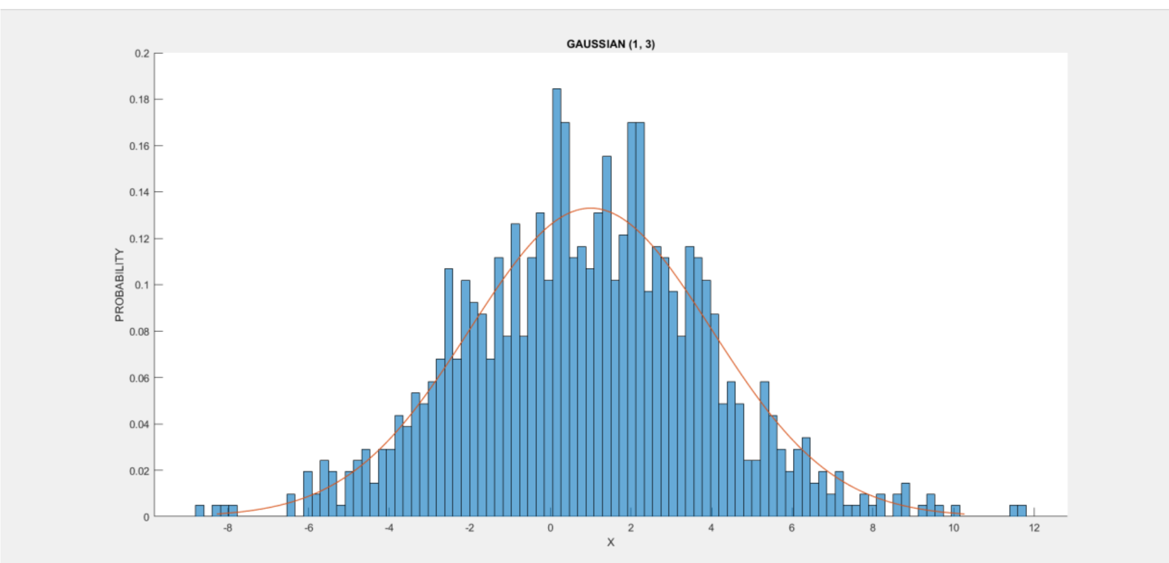
Para la pregunta 3 se pide generar una distribucion gaussiana con media 1 y desviación estándar 3 para luego graficar la distribución de probabilidad empirica junto con la teórica. El código usado para esta tarea es el siguiente:

Editor - C:\Users\marti\OneDrive - Universidad Adolfo Ibanez\Desktop\Universidad\Derivados Avanzados\Tarea Distribuciones\Tarea aparte\QUESTION3.m

script.m | QUESTION1.m | QUESTION2.m | QUESTION3.m | QUESTION4.m | Q5FUNCTION_q4.m

```
% QUESTION 3 :

clc
clear

% DATA :
N = 1000;
Q = 100;

%  SEED :
rng(0);

MU = 1;
SIGMA = 3;

% RANDOM NUMBERS (1, 3)
x_q3 = SIGMA * randn(N, 1) + MU;

% PLOTTING :
    % EMPYRICAL :
hold on;
h_q3 = histogram(x_q3, Q, "Normalization","pdf");

    % THEORETICAL :
y_q3 = makedist("Normal", "mu", MU, "sigma", SIGMA);
plot_q3 = plot(y_q3, "PlotType","pdf");
title('GAUSSIAN (1, 3)');
xlabel('X');
ylabel('PROBABILITY');
hold off;
```

Con esto se generan los gráficos solicitados por la pregunta y los outputs son los mostrados a continuación:



Como podemos ver, la distribución empírica se asemeja bastante a la teórica.

Para la pregunta 4 se solicita realizar el mismo ejercicio que en la 3 pero ahora con una distribucion de Cauchy que sigue una PDF bastante particular.

$$\Gamma^{Cauchy}(x) = \frac{1}{\pi\sigma\left[1+\left(\frac{x-\mu}{\sigma}\right)^2\right]}$$

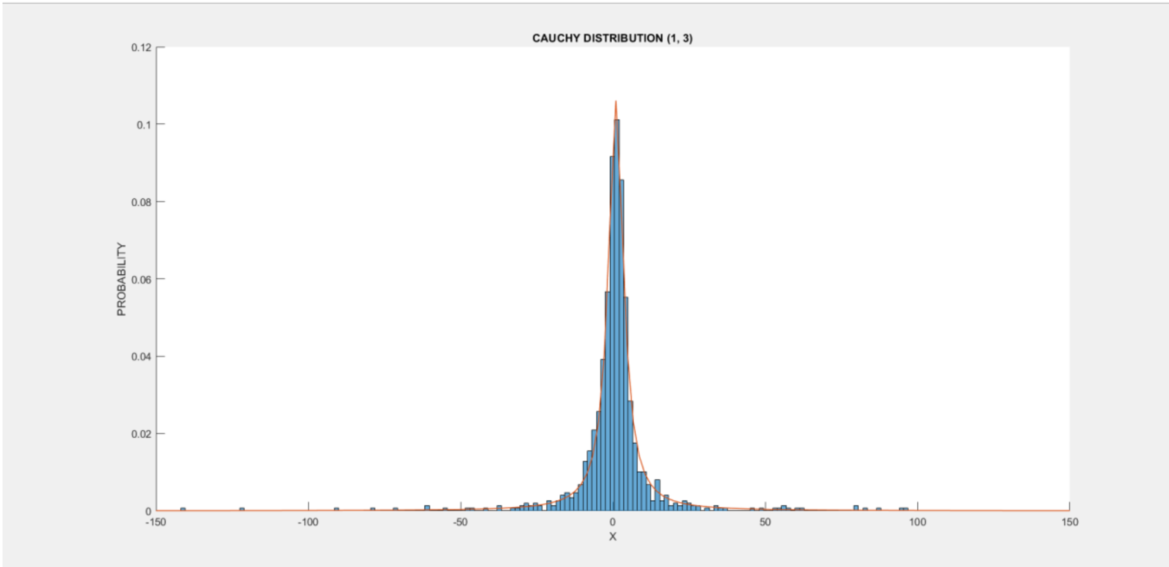
Para realizar esta tarea utilicé el siguiente código:

Editor - C:\Users\marti\OneDrive - Universidad Adolfo Ibanez\Desktop\Universidad\Derivados Avanzados\Tarea Distribuciones\Tarea aparte\QUESTION4.m

script.m | QUESTION1.m | QUESTION2.m | QUESTION3.m | QUESTION4.m | Q5FUNCTION_q4.m

```
% QUESTION 4 :
clc
clear
% DATA : Edite el numero de Q para que el grafico fuera mas estetico
       % Para tener 100 bins deberiamos poner Q = 100
N = 1000;
Q = 1000;
MU = 1;
SIGMA = 3;

% SEED :
rng(0);

% GENERATE RANDOM VARIABLE
x_q1 = rand(N,1);

x_q4 = MU + SIGMA * tan(pi * (x_q1 - 0.5));

% PLOT :

hold on;
histogram(x_q4,'Normalization', 'pdf', 'NumBins', Q);

% CAUCHY
y_q4 = 1./(pi * SIGMA * (1 + ((x_q4 - MU) ./ SIGMA) .^ 2));
pd_q4 = makedist('tLocationScale','mu', MU,'sigma', SIGMA,'nu', 1);
plot(pd_q4 ,"PlotType","pdf")
xlim([-150, 150]);
title("CAUCHY DISTRIBUTION (1, 3)");
xlabel('X');
ylabel('PROBABILITY');
hold off;
```

Para lograr generar la variable aleatoria con distribución de Cauchy fue necesario previamente despejar la función para encontrar la forma de programarlo. En este caso la variable aleatoria quedo denominada como x_q4. Mientras que la distribución de probabilidad se muestra en el siguiente grafico (tanto la empírica como la teórica).



La distribución de Cauchy y la distribución normal son similares en que ambas son distribuciones continuas. Sin embargo, también tienen algunas diferencias importantes. La distribución de Cauchy tiene colas más pesadas que la distribución normal, lo que significa que es más probable que genere valores extremos. La distribución de Cauchy también tiene una media y una varianza indefinidas, mientras que la distribución normal tiene una media y una varianza bien definidas, esto queda ejemplificado más adelante.

En la pregunta 5 se pide evaluar como evolucionan el valor de los primeros 4 momentos a medida que aumenta N para distintas distribuciones, para esto generé distintas funciones que quedan ejemplificadas más adelante:

En primer lugar, veremos lo que sucede con los momento de una distribución uniforme, para esto utilizaremos la variable aleatoria generada en la pregunta 2 y siempre con una seed incluida para que se pueda replicar por el lector en cualquier caso.

```
% QUESTION 5:

clear
clc

% COLOR LEGEND :

    % MEDIA = BLUE
    % DESVIACION STD = RED
    % SKEWNESS = GREEN
    % KURTOSIS = black

% GENERATE x_q2 :
rng(0);
N = 1000;
min = -5;
max = 7;
x_q1 = rand(N,1);
x_q2 = min + x_q1 * (max - min);

Q5FUNCTION_q2(N, x_q2);
```

En esta parte del código se generan los inputs para la función generada.

script.m | QUESTION1.m | QUESTION2.m | QUESTION3.m | QUESTION4.m | Q5FUNCTION_q4.m | Q5FUNCTION_q2.m

```
function Q5FUNCTION_q2 (N, x_q2)

sample = 1:N;

% INICIALIZE VECTORS :
mu = zeros(N, 1);
sigma = zeros(N,1);
skew = zeros(N,1);
kurt = zeros(N,1);

% GENERATE DATA
for i = sample

    S_1N = sum(x_q2(1:i));
    S_2N = sum(x_q2(1:i) .^ 2);
    S_3N = sum(x_q2(1:i) .^ 3);
    S_4N = sum(x_q2(1:i) .^ 4);

    % MEDIA :
    mu(i) = S_1N / i;

    % DESVIACION STD :
    sigma(i) = sqrt ((S_2N / i) - mu(i) .^ 2);

    % SKEWNESS :
    skew(i) = ((S_3N / i) - 3 * (mu(i)) * (S_2N / i) + 2 * (mu(i)) .^ 3) / (sigma(i)) .^ 3;

    % KURTOSIS :
    kurt(i) = ((S_4N / i) - 4 * (mu(i)) * (S_3N / i) + 6 * (mu(i)) .^ 2 * (S_2N / i) - 3 * (mu(i)) .^ 4) / (sigma(i)) .^ 4;


end

% PLOTTING :
hold on;

plot (sample, mu, 'Color', 'blue', 'DisplayName', 'MEDIA');
```

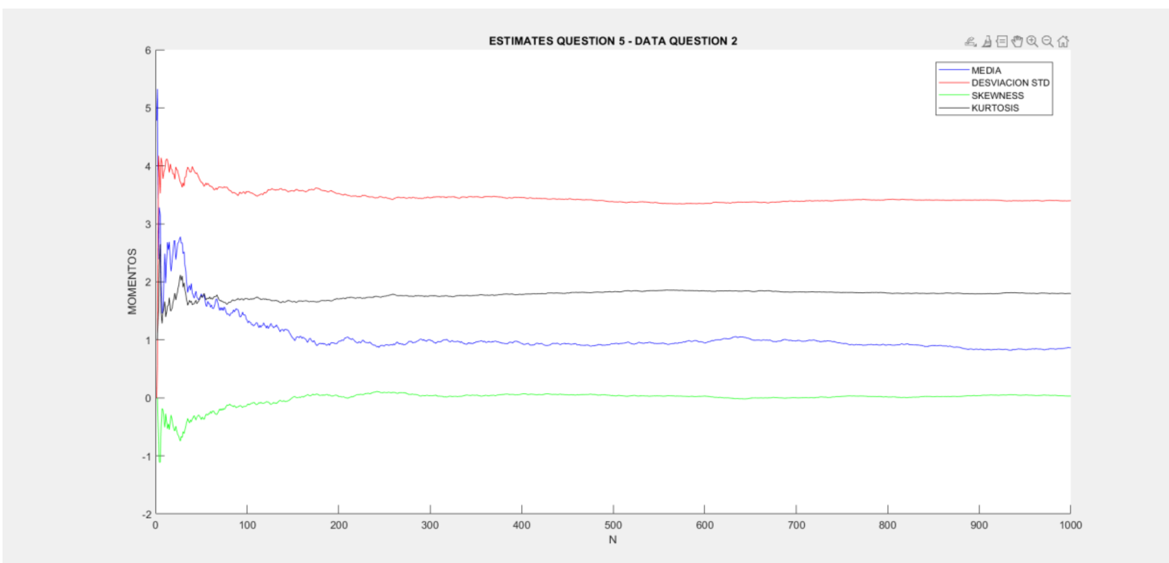
La parte del código que queda fuera de la imagen no es relevante, simplemente son las especificaciones para el funcionamiento del grafico y como la leyenda se encuentra en la primera imagen podemos proceder a analizar los outputs.

Los primeros 4 momentos de una variable aleatoria que se distribuye uniformemente entre dos valores son la media, la varianza, la asimetría y la curtosis. La media es el promedio de los valores, la varianza es la medida de cuánto se dispersan los valores alrededor de la media, la asimetría es la medida de cómo los valores están sesgados a la izquierda o a la derecha de la media y la curtosis es la medida de cómo los valores están distribuidos alrededor de la media.

A medida que aumenta el número de observaciones de la muestra, los primeros 4 momentos de la distribución de probabilidad se acercan a los valores de la distribución teórica. Esto se debe a que la distribución teórica es el límite de la distribución empírica a medida que el número de observaciones de la muestra aumenta sin límite.

En particular, la media de la distribución empírica se aproxima a la media de la distribución teórica, la varianza de la distribución empírica se aproxima a la varianza de la distribución teórica, la asimetría de la distribución empírica se aproxima a la asimetría de la distribución teórica y la curtosis de la distribución empírica se aproxima a la curtosis de la distribución teórica.

Estas aproximaciones son útiles para realizar inferencias sobre las distribuciones de probabilidad a partir de muestras de datos.



Tal y como esperábamos, a medida que aumenta N los valores de los momentos de la variable convergen a cierto valor.

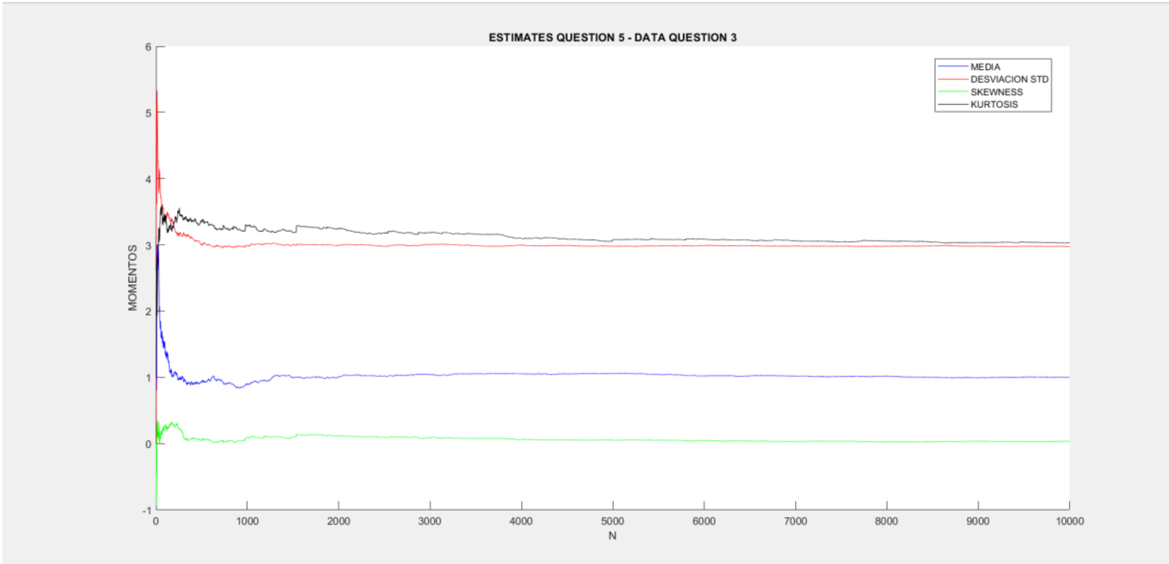
A continuación, analizamos lo sucedido en la distribución Gaussiana. El código para esto es el siguiente:

```
% CLEAN WORKSPACE :
clear
clc

% GENERATE x_q3 :
rng(0);
MU = 1;
SIGMA = 3;
N = 1000;
x_q3 = SIGMA * randn(N, 1) + MU;

% CALL FUNCTION :
Q5FUNCTION_q3(N, x_q3)
```

Esto llama a la siguiente función:

```
function Q5FUNCTION_q3 (N, x_q3)

sample = 1:N;

% INICIALIZE VECTORS :
mu = zeros(N, 1);
sigma = zeros(N,1);
skew = zeros(N,1);
kurt = zeros(N,1);

% GENERATE DATA
for i = sample

    S_1N = sum(x_q3(1:i));
    S_2N = sum(x_q3(1:i) .^ 2);
    S_3N = sum(x_q3(1:i) .^ 3);
    S_4N = sum(x_q3(1:i) .^ 4);

    % MEDIA :
    mu(i) = S_1N / i;

    % DESVIACION STD :
    sigma(i) = sqrt ((S_2N / i) - mu(i) .^ 2);

    % SKEWNESS :
    skew(i) = ((S_3N / i) - 3 * (mu(i)) * (S_2N / i) + 2 * (mu(i)) .^ 3) / (sigma(i)) .^ 3;

    % KURTOSIS :
    kurt(i) = ((S_4N / i) - 4 * (mu(i)) * (S_3N / i) + 6 * (mu(i)) .^ 2 * (S_2N / i) - 3 * (mu(i)) .^ 4) / (sigma(i)) .^ 4;


end
```

La función anterior hace lo mismo que la explicada en la parte anterior de la pregunta, pero lo hace para la variable aleatoria con distribución gaussiana generada para la pregunta 3. Esto devuelve los siguientes outputs.



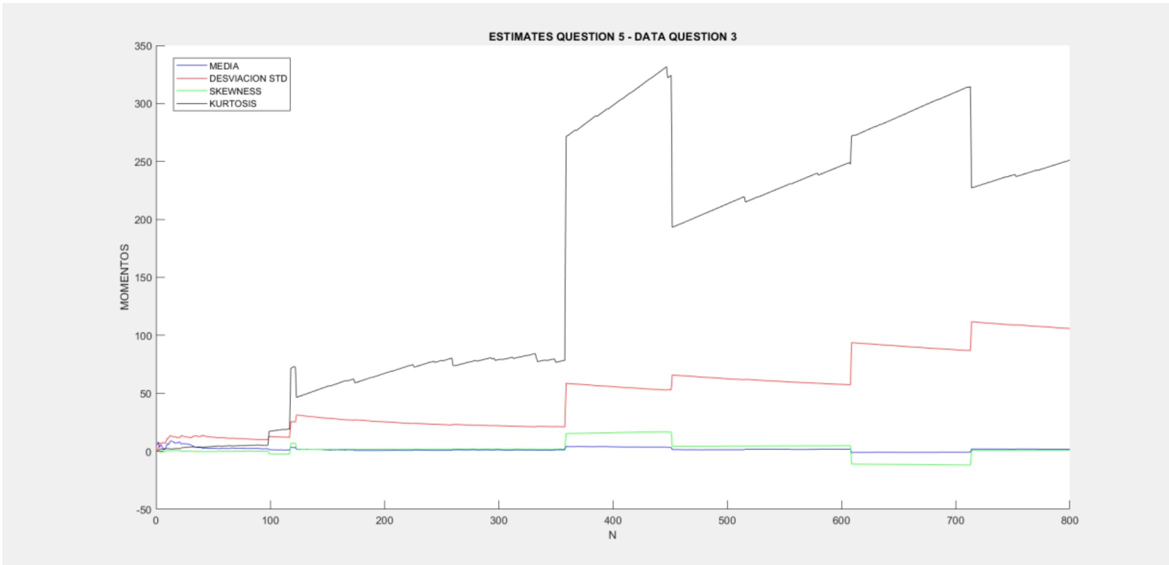
Como podemos ver, el análisis realizado para la parte anterior aplica también para este, a medida que aumenta N los momentos de la variable gaussiana convergen a un valor fijo.

Para la última parte genere una función que hace lo mismo que la función Q4FUNCTION_q3 pero en vez de usar a x_q3 usa a x_q4 como variable aleatoria, la cual tiene una distribución de cauchy.

Los primeros cuatro momentos de una variable aleatoria que distribuye según Cauchy no convergen a un valor definido cuando uno aumenta las observaciones de la muestra. Esto se debe a que la distribución de Cauchy no tiene media ni varianza finitas. La media es la tendencia central de los datos, mientras que la varianza es una medida de la dispersión de los datos alrededor de la media. Si la distribución de Cauchy no tiene media ni varianza finitas, entonces los primeros cuatro momentos no pueden converger a un valor definido.

La distribución de Cauchy es una distribución de cola pesada, lo que significa que tiene una mayor probabilidad de producir valores extremos que una distribución normal. Esto se debe a que la distribución de Cauchy no tiene un pico definido. En su lugar, la distribución de Cauchy se extiende uniformemente alrededor de la media.

La distribución de Cauchy no se utiliza a menudo en estadística porque es difícil de trabajar con. Sin embargo, se utiliza a veces en física y finanzas, donde es importante modelar datos que tienen colas pesadas. Teniendo esto en cuenta podemos ver que se ve confirmado por le siguiente output donde los valores de los momentos no convergen.


ESTIMATES QUESTION 5 - DATA QUESTION 3
MEDIA
DESVIACION STD
SKEWNESS
KURTOSIS
MOMENTOS
N
350
300
250
200
150
100
50
0
-50
0
100
200
300
400
500
600
700
800