Universidad Adolfo Ibañez

Derivados Avanzados

RANDOM VARIABLE SIMULATION

Alumno:
Felipe Ignacio Durán Aranda

Profesor:
Jacques Burrus

Ayudantes:
- Alejandro Olguín
- Guillermo Sepúlveda

February 25, 2021

Contenidos

1 Prefacio

Mediante este informe se busca presentar los resultados de la simulación de distintos tipos de variables aleatorias, obtenidas a través del programa *Matlab*, cuya dinámica se encuentra planteada en la actividad *Random Variable Simulation*, a través de diversos modelos, metodologías, requerimientos y procedimientos, para finalmente comparar su comportamiento teórico con el obtenido empíricamente.

2 Pregunta Uno

Para esta pregunta, se pide simular 1000 variables aleatorias con una distribución uniforme dentro del intervalo 0 y 1, para posteriormente realizar un histograma con la distribución de probabilidades obtenidas empíricamente. Los resultados se presentan a continuación:

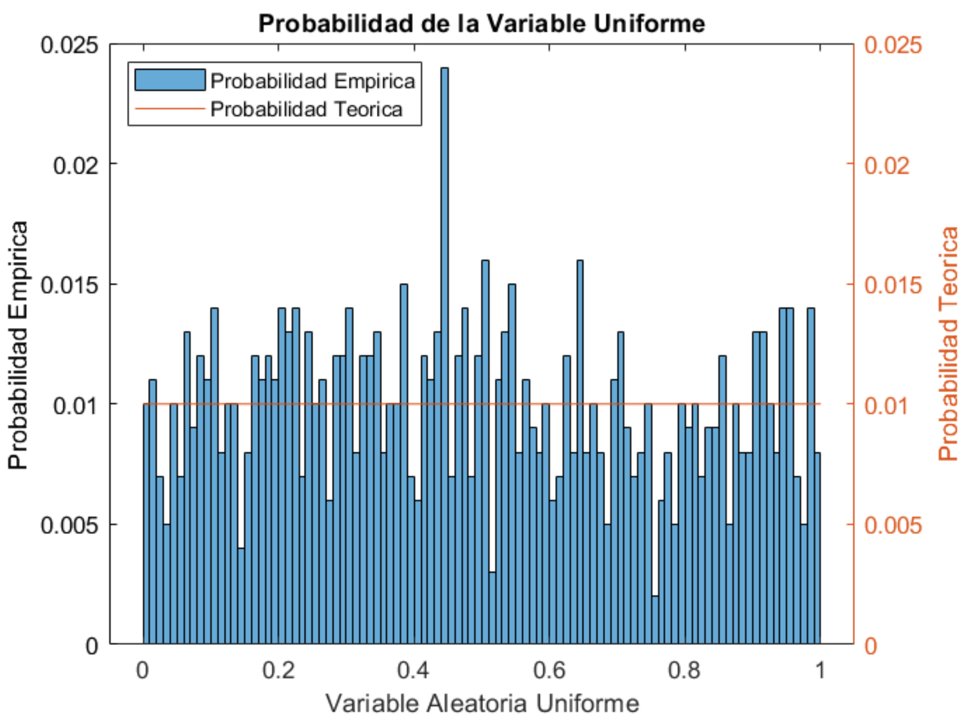

Figura 1: Histograma de Probabilidades Distribución Uniforme

Tal como se puede observar en la figura anteriormente presentada, para una distribución uniforme estándar espaciada de manera ecuánime entre 100 *Bins*, se esperaría que cada espaciado acumulara una probabilidad teórica del (0.01), o como es equivalente del 1%, de tal forma que se cumpliese la siguiente ecuación:

$$\begin{gathered} N°Bins * ProbabilidaddecadaBin = 1 \\ 100 * 0.01 = 1 \end{gathered} \tag{1}$$

En contraste, con respecto a la distribución empírica (representada por el histograma de las 1000 variables aleatorias uniformes estándar simuladas), podemos observar que, a pesar de ciertas excepciones, la probabilidad de cada *Bin* oscila alrededor del 0.01, o como es equivalente, del 1%.

Lo anterior, implica que la distribución empírica, a pesar de no ajustarse perfectamente a la distribución teórica esperada, logra asemejarse a esta, indicando que efectivamente las variables se generaron siguiendo una distribución uniforme estándar.

Las variables utilizadas para generar la distribución se crearon a partir del comando *randn* en *Matlab*, función capaz de generar variables aleatorias para una distribución normal estándar, a través del siguiente código:

```
rng(2);
Z1=rand(N,1);
```

Finalmente, cabe mencionar que se esperaría que el ajuste de la distribución empírica a la distribución teórica mejorase, a medida que aumente el número de observaciones, planteamiento que en parte será explorado parcialmente en la pregunta cinco.

3 Pregunta Dos

De manera similar a la pregunta anterior, se pide simular 1000 observaciones de una variable aleatoria dentro de una distribución uniforme, esta vez delimitada en el espacio comprendido entre (-5) y 7. Los resultados obtenidos se presentan a continuación:

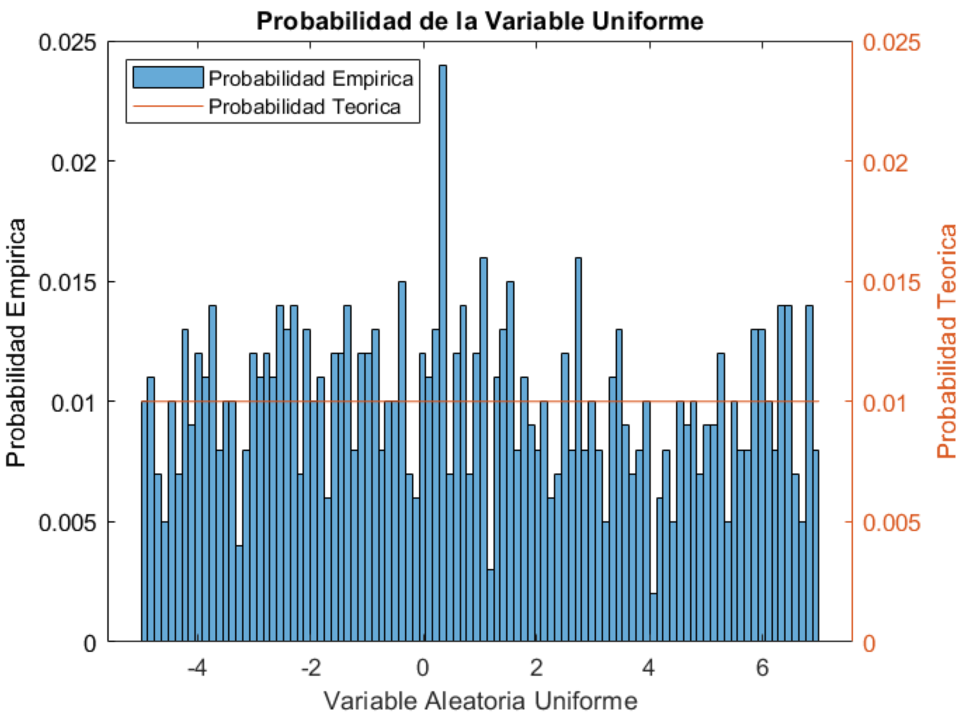

Figura 2: Histograma de Probabilidades Distribución Uniforme

Como se puede observar, comparando la Figura 1 y la Figura 2, ambas presentan la misma distribución, esto se debe en parte a que ambas se generaron usando la misma semilla (2 en este caso), y, por otra parte, debido a la forma utilizada para generar las observaciones. Su representación como código es la siguiente:

```
rng(2);
Z2=(-5)+(12).*rand(N,1);
```

Como se puede apreciar, la simulación de la variable consiste en aplicar operaciones con valores escalares sobre una variable aleatoria uniforme estándar, lo que, sumado a utilizar la misma semilla, provoca que se de la misma forma de la distribución.

No obstante, de igual manera podemos observar que, bajo ciertas excepciones, la distribución empírica obtenida se ajusta moderadamente a la distribución teórica indicada en la figura anteriormente presentada. Lo anterior, indicaría que, dado que de igual manera cambiaron los límites dentro de los cuales se da la distribución uniforme, las variables aleatorias fueron generadas de manera correcta.

De la misma forma que la pregunta anterior, se esperaría que, a medida que aumente el número de observaciones, aumente el "*ajuste*" que se da entre la distribución empírica y la distribución teórica.

4 Pregunta Tres

El objetivo de esta pregunta consiste en simular nuevamente 1000 variables aleatorias, esta vez con una distribución *Gaussiana*, también conocida como distribución normal, con media (μ) igual a 1 y desviación estándar (σ) igual a 3. Los resultados obtenidos son presentados a continuación:

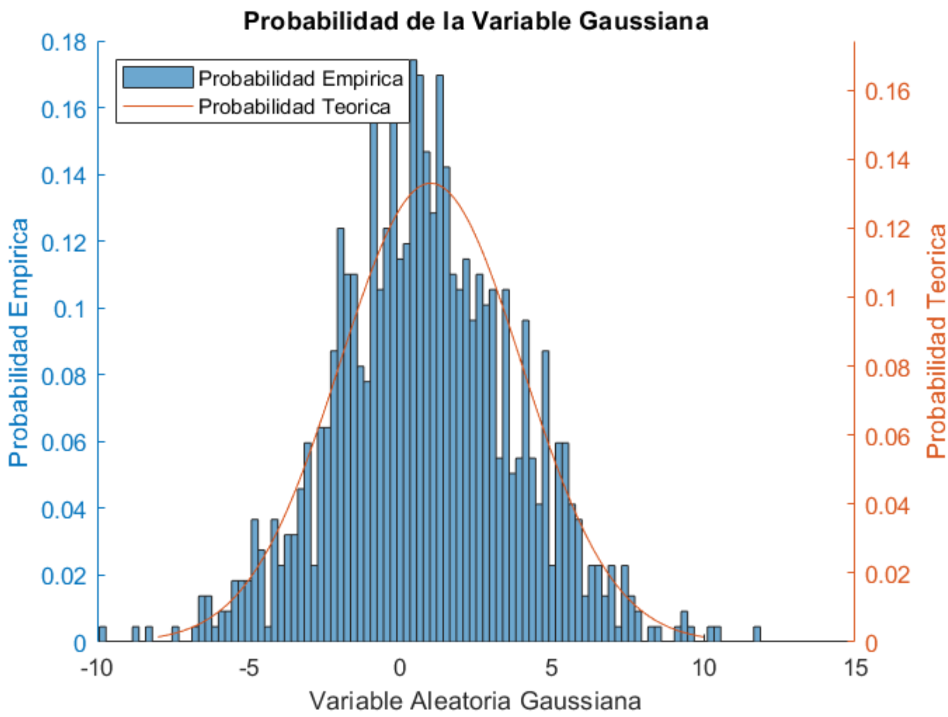

Figura 3: Histograma de Probabilidades Distribución Gaussiana

Las variables aleatorias utilizadas para generar la distribución fueron obtenidas utilizando el comando *normrnd*, función de *Matlab*, capaz de generar números aleatorios a partir de una distribución normal con parámetros de media (μ) y desviación estándar (σ). El código utilizado es presentado a continuación:

```
mu=1;
sigma=3;
rng(2);
Z3=normrnd(mu,sigma,N,1);
%N representa el numero de observaciones que queremos
```

Para obtener la probabilidad teórica con la cuál comparar la distribución empírica, se utilizó la siguiente fórmula:

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

Dentro del código se puede ver de la siguiente manera, para un x definido en un espacio limitado:

```
x = linspace(mu-3*sigma,mu+3*sigma,N);
y = exp(-(x-mu).^2./(2*sigma^2))./(sigma*sqrt(2*pi));
```

Podemos observar, según como es presentado en la Figura 3, que la distribución empírica tiene un buen ajuste a la distribución teórica, siendo ligeramente mayor en los valores cercanos a la media de la distribución.

5 Pregunta Cuatro

Para esta pregunta se pide simular 1000 variables aleatorias según una distribución de *Cauchy*, con parámetros de *location* (μ) igual a 1 y *scale* (σ) igual a 3. Los resultados obtenidos se presentan a continuación:

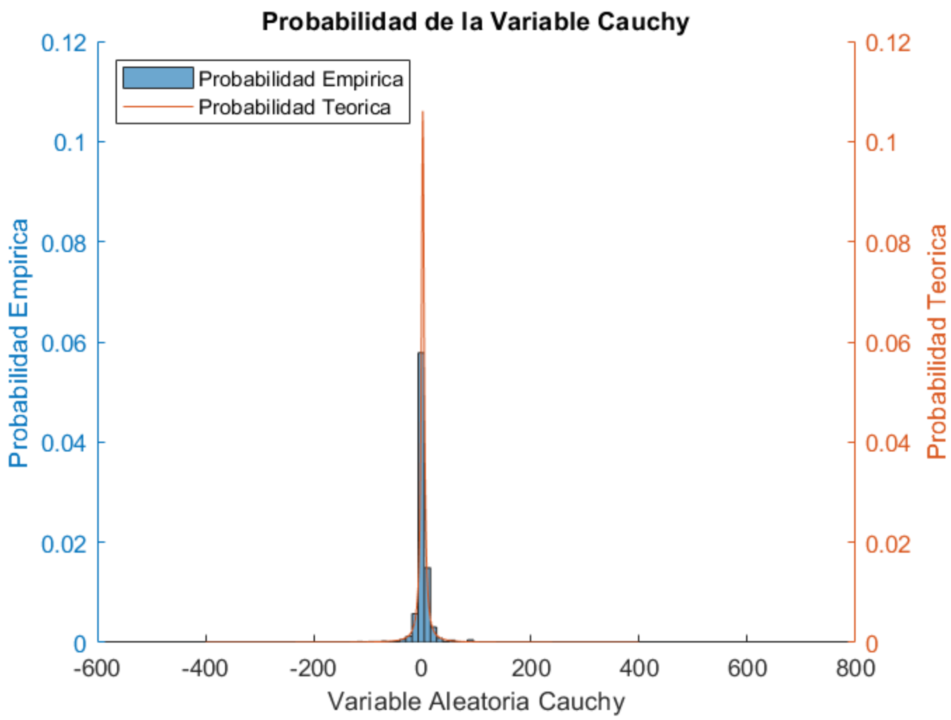

Figura 4: Histograma de Probabilidades Distribución Cauchy

Tal como se puede apreciar en la imagen anterior, la simulación empírica de la variable tendió a ser más pequeña alrededor de μ respecto a la distribución teórica esperada. Sin desmedro de lo anterior, podemos observar que efectivamente la variable aleatoria siguió una distribución de *Cauchy* (según se puede observar la concordancia en el ajuste de la distribución empírica a la distribución teórica presentada, presentada en la figura anterior).

Las variables utilizadas para generar la distribución empírica se generaron de la siguiente manera:

```
mu=1;
sigma=3;
rng(2);
Z4=mu+sigma*tan(pi*(rand(N,1)-1/2));
```

Código que obedece a la siguiente función:

$$F^{-1}(p;\mu,\sigma)=\mu+\sigma\tan\left[\pi\left(p-\frac{1}{2}\right)\right] \tag{3}$$

Con p siendo una variable aleatoria con una distribución uniforme estándar.

Así mismo, para generar la distribución teórica que debería seguir la simulación de variables aleatorias *Cauchy*, se utilizó la función *Makedist* en *Matlab*, mediante el siguiente código:

```
x=(-400):0.01:400;
pd=makedist('tLocationScale','mu',1,'sigma',3,'nu',1);
y=pdf(pd,x);
```

Finalmente, de la misma manera que las distribuciones presentadas con anterioridad, podemos decir que, a pesar de contadas excepciones, la distribución empírica logra ajustarse al modelo teórico, así como el que se esperaría que el *ajuste* mejorase a medida que aumente el número de observaciones de la variable aleatoria.

6 Pregunta Cinco

Para esta pregunta, se pide investigar el efecto que tiene el número de observaciones (N), en la estimación de los primeros cuatro momentos de las distribuciones obtenidas para las preguntas dos, tres, y cuatro.

Al final de esta sección se presentan conjuntamente los gráficos de los momentos agrupados por distribución, como *Subplots* de *Matlab*.

A continuación se presentan los resultados de la media en función del número de observaciones (N) para las distribuciones mencionadas anteriormente:

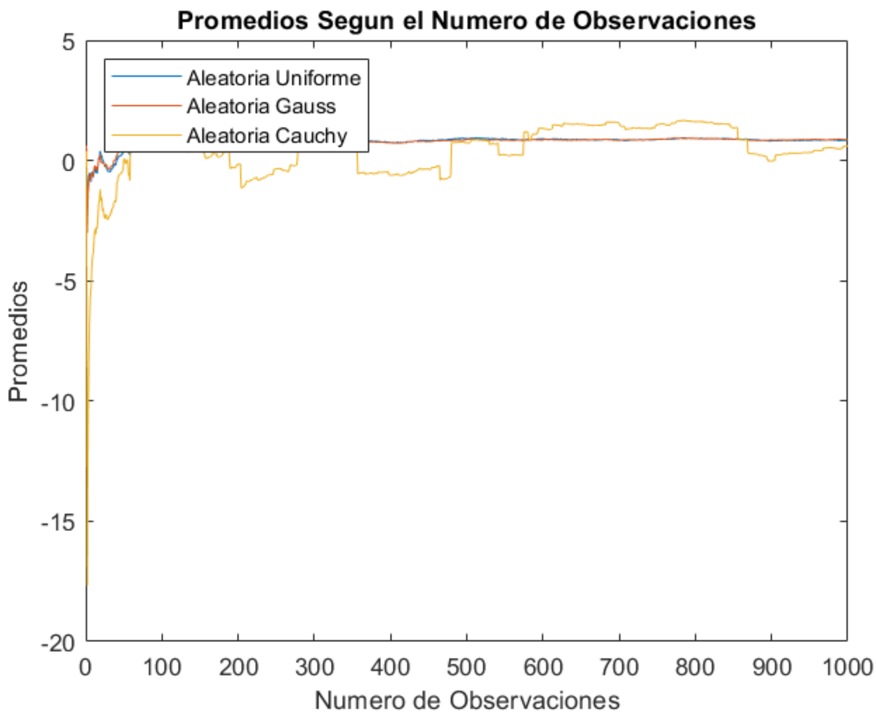

Figura 5: Media según el número de observaciones

Como podemos observar, la distribución aleatoria de *Cauchy* no presenta un valor asintótico para su media, situación concordante con la teoría, en contraste a las distribuciones aleatorias uniforme y *Gaussiana*, las cuales tienden a los valores 0.84 y 0.88 respectivamente (de manera aproximada).

Los valores teóricos de los promedios para las distribuciones uniforme y *Gaussiana* obedecen a las siguientes fórmulas, en función de sus momentos respectivos:

$$\begin{array}{c} Momento\ k\ de\ la\ distribucin\ uniforme: \\ m_k = \frac{1}{k+1}\sum_{i=0}^{k} a^i b^{k-i} \end{array} \tag{4}$$

Finalmente, podemos observar que la media teórica de la distribución uniforme, respondería a la siguiente ecuación:

$$m_1 = \frac{a+b}{2} \tag{5}$$

De la cuál el valor teórico de la media corresponde a 1. Visto lo anterior, podemos concluir que la distribución se acercó al valor teórico de la media, sin embargo, no de manera perfecta, siendo el resultado asintótico empírico igual a 0.84.

Lo anterior, puede deberse a que el número de observaciones (N) igual a 1000, todavía puede no ser suficiente para comprobar la *Ley de los grandes números*, en la cuál esperaríamos que al aumentar el número de observaciones, la distribución tuviera un valor asintótico de la media igual al valor teórico, en este caso a 1.

Otra posible explicación, que, aunque poco probable, vale la pena mencionar, es la posibilidad de que el sistema de generación de números aleatorios de *Matlab* estuviese *sesgado* o no fuese realmente *aleatorio*, lo que ha pasado con anterioridad en otros programas, sin embargo, esta explicación es tan poco probable que no debería dársele mucha importancia.

Respecto al valor de la media de la distribución *Gaussiana*, podemos nuevamente observar que, a pesar de acercarse al valor asintótico teórico de μ igual a 1, siendo el respectivo valor empírico 0.88, este no logra igualarlo perfectamente o al menos aproximadamente. Nuevamente, esto puede deberse a que un N igual a 1000 puede no ser un número lo suficientemente grande para comprobar la *Ley de los grandes números.*

Para la distribución de *Cauchy*, tal como era esperable de la teoría y como se puede apreciar en la figura 5, no se puede apreciar que la media tenga convergencia a ningún valor.

A continuación se presentan los resultados obtenidos para la desviación estándar:

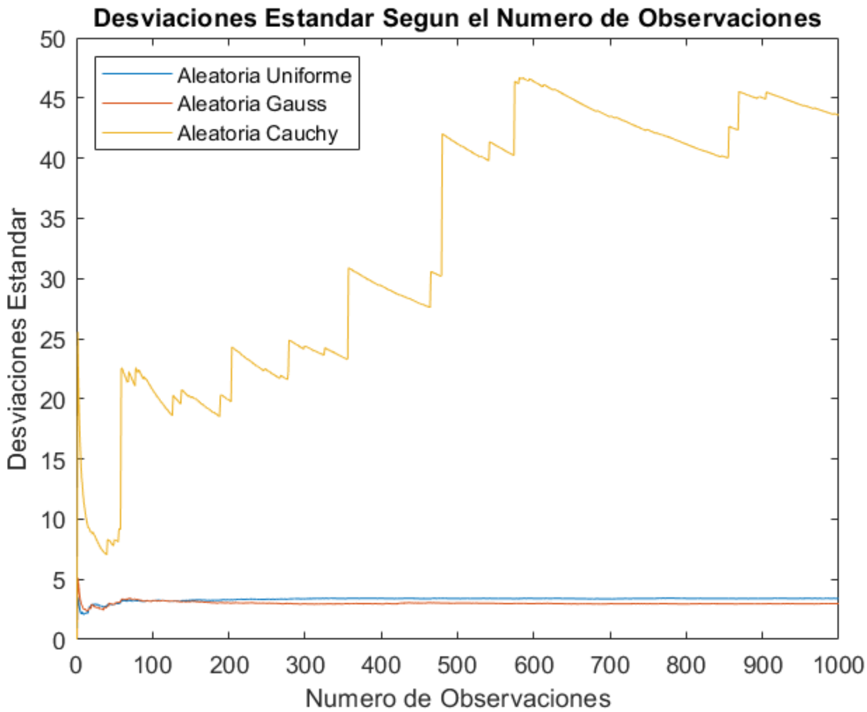

Figura 6: Desviaciones estándar según el número de observaciones

Sobre las distribuciones uniforme y *Gaussiana*, podemos observar en la figura 6 que sus desviaciones estándar tendieron a 3.40 y 2.98 respectivamente. Respecto a la distribución uniforme, el valor teórico esperado de la desviación estándar obedece a las siguientes fórmulas:

$$\begin{gathered}\sigma = \sqrt{m_2 - m_1^2} = \sqrt{\tfrac{(b-a)^2}{12}} \\ \sigma = \sqrt{\tfrac{(7-(-5))^2}{12}} = \sqrt{12} = 3.4641\end{gathered} \tag{6}$$

Siguiendo el resultado de la fórmula anteriormente presentada, siendo de 3.4641 para la desviación estándar teórica de la distribución uniforme, podemos observar que el valor obtenido empíricamente, de 3.40, se acerca bastante a lo esperado, y, debido a la *Ley de los grandes números*, podríamos esperar que este valor se asemejare todavía más a medida que aumentase el número de observaciones (N).

Así mismo, para la distribución *Gaussiana*, podemos observar como el valor obtenido empíricamente, de 2.98, es muy cercano al valor esperado con el que se construyó la distribución en primer lugar, siendo este de un σ igual a 3. En esta ocasión, podemos percibir como el número de observaciones (N) igual a 1000 se ajustó satisfactoriamente al valor esperado.

Por otra parte, podemos observar que para la distribución *Cauchy*, no existe convergencia hacia ningún valor, lo que concuerda con la teoría.

A continuación se presentan los valores obtenidos para la asimetría o *skewness*:

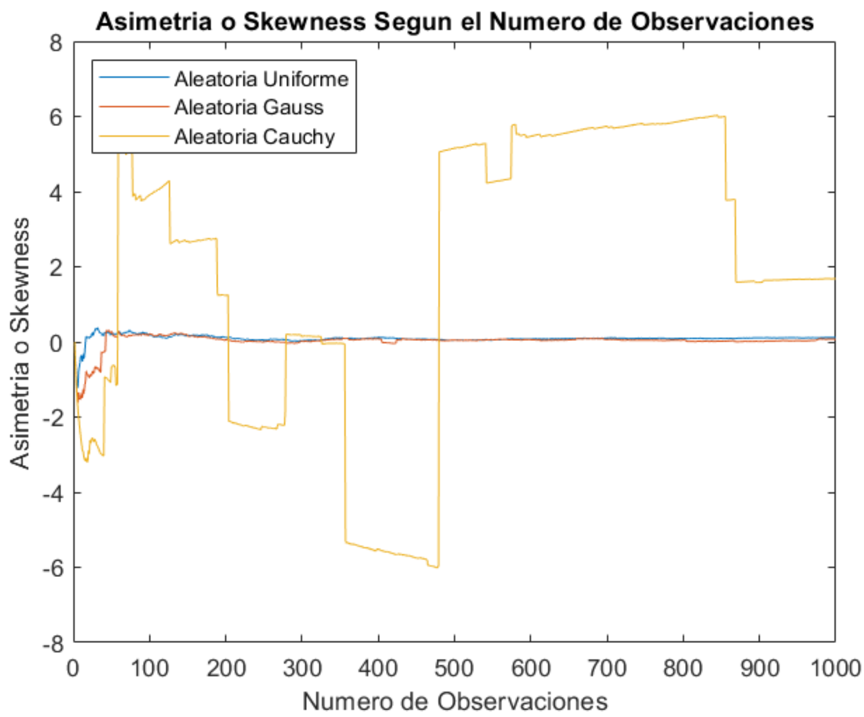

Figura 7: Asimetría o skewness según el número de observaciones

De la figura 7, se puede concluir que efectivamente existe una convergencia del valor de la asimetría para las distribuciones uniforme y *Gaussiana*, siendo este de 0.11 y 0.07 respectivamente. Ambos valores se ajustan a la teoría, que dice que el valor debería tender a 0, teniendo la distribución *Gaussiana* una mayor convergencia que la distribución uniforme.

Por otra parte, podemos observar que, nuevamente, la distribución *Cauchy* no presenta convergencia hacia ningún valor en específico, nuevamente concordando con lo esperado según la teoría.

A continuación se presentan los valores obtenidos para la *Extra-Curtosis*:

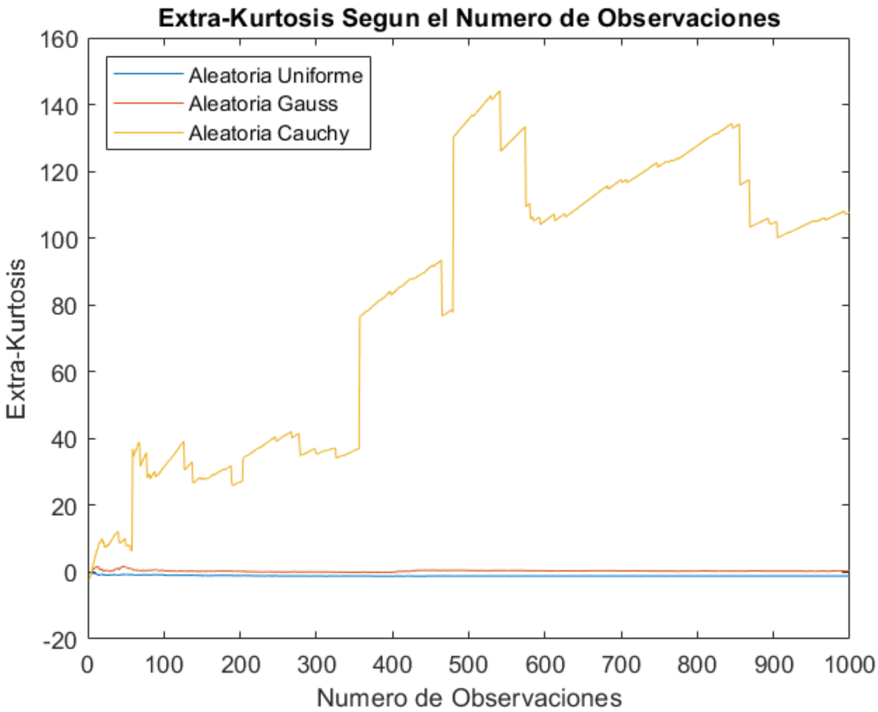

Figura 8: Extra-Curtosis según el número de observaciones

Cuya forma está dada por:

$$\text{Extra-Curtosis} = Curtosis - 3 \tag{7}$$

Donde *Curtosis* corresponde a la formula de *Kurtosis* en *Matlab*, al menos para el ejercicio de esta actividad.

Nuevamente, podemos observar que, para las distribuciones uniforme y *Gaussiana*, la *Extra-Curtosis* si presenta convergencia hacia valores específicos, siendo estos de (-1.13) y 0.35 respectivamente. Los valores teóricos esperados para la *Extra-Curtosis* son de (-1.2) y 0 para la distribución uniforme y *Gaussiana* respectivamente. Bajo estos valores, podemos decir que la distribución uniforme presentó un buen ajuste en relación a la *Extra-Curtosis*, sin embargo, no podemos hacer la misma observación de la distribución *Gaussiana*. Lo anterior, puede deberse a que un número de observaciones (N) de 1000 puede no ser suficientemente grande para encontrarnos con la *Ley de los grandes números*, por lo cual, se esperaría que a medida que aumente el número de observaciones (N) aumentase el ajuste que presenta la distribución *Gaussiana* a la *Extra-Curtosis* teórica.

Finalmente, en las páginas siguientes del documento, se adjuntan los momentos de cada distribución agrupados según el tipo de variable aleatoria utilizada.

A modo de conclusión de las imágenes presentadas a continuación, podemos decir que, exceptuando pequeños desajustes, los momentos de las distribuciones uniforme y *Gaussiana* concordaron con la teoría, convergiendo a valores concretos semejantes a los teóricos. Por otra parte, para el caso de la distribución *Cauchy*, podemos observar que ningún momento convergió a un valor concreto, lo cuál es concordante con el resultado teórico esperado, en el cuál, tanto la media, varianza, asimetría y *Extra-Curtosis*, se encuentran indefinidas, dado que no son asintóticas a ningún valor.

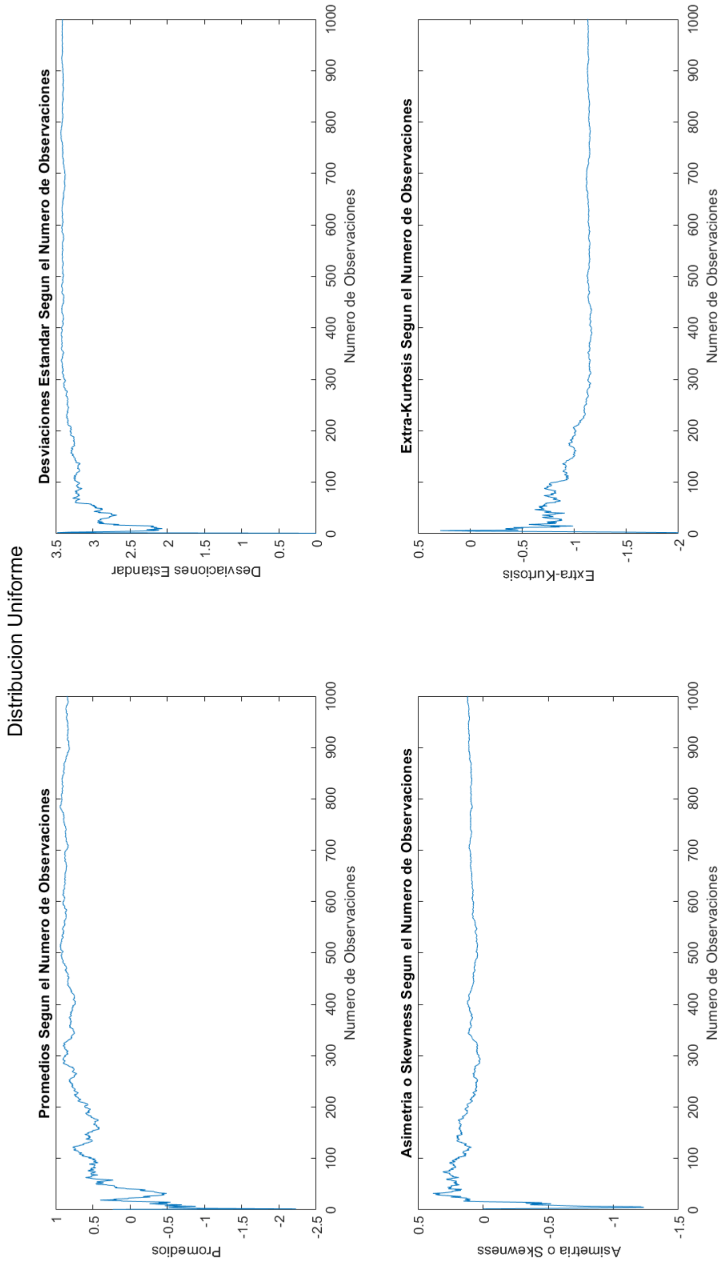

Figura 9: Momentos de la distribución uniforme según el número de observaciones

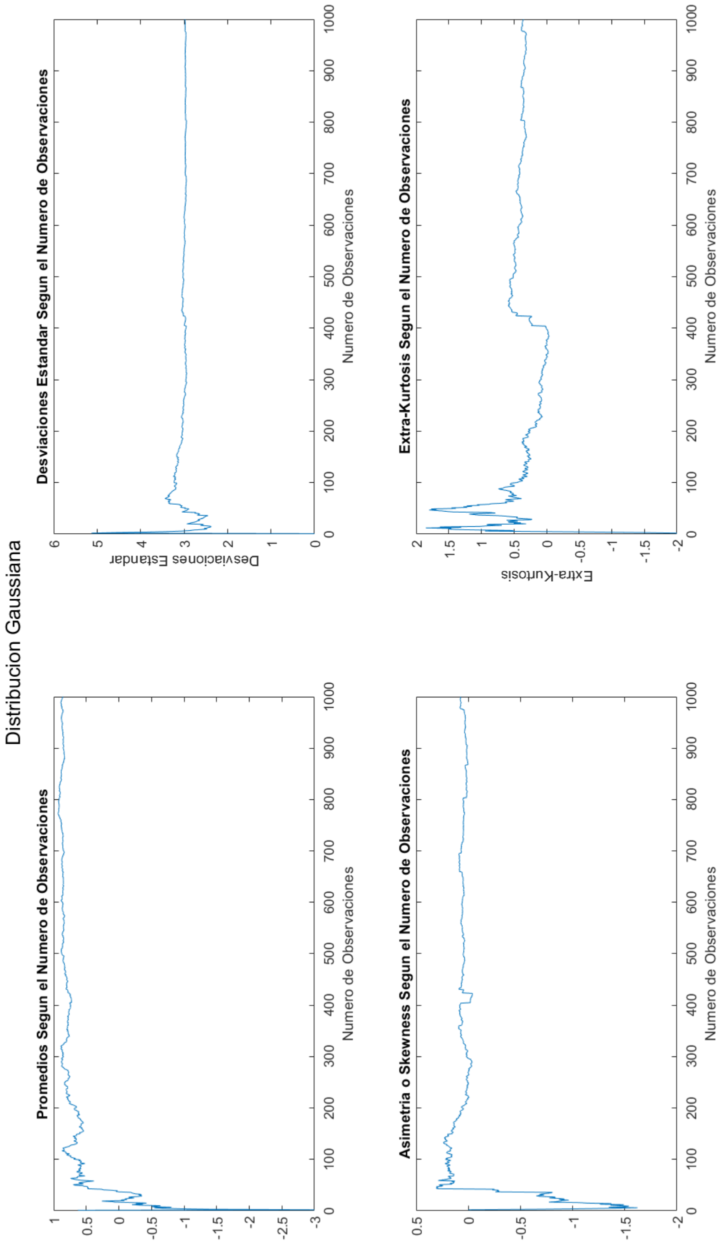

Figura 10: Momentos de la distribución Gaussiana según el número de observaciones

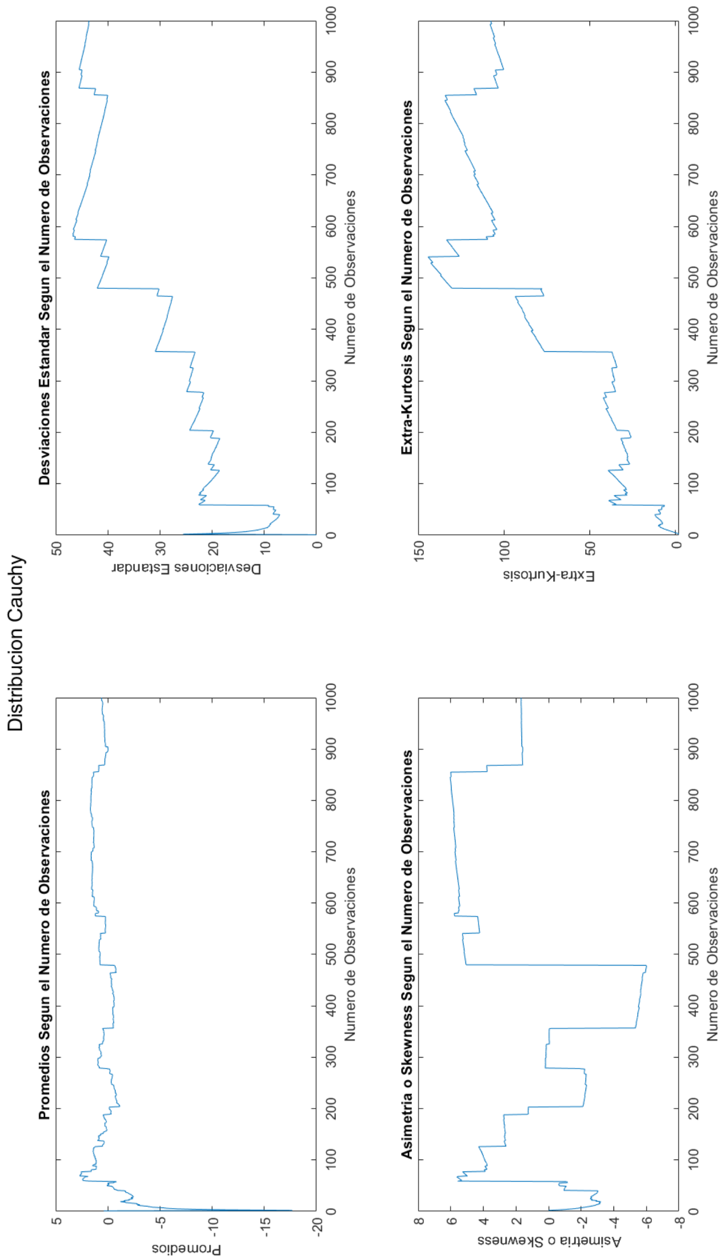

Figura 11: Momentos de la distribución Cauchy según el número de observaciones