



# **Centro de Investigación en Computación**

Series de Tiempo (Predicción y Caos)

## **Proyecto 1:**

*Caminante Aleatorio*

Autor:

Rodrigo Gerardo Trejo Arriaga

Abril 2024

# Proyecto 1:

## *Caminante Aleatorio*

Un caminante aleatorio es un proceso estocástico que modela una secuencia de pasos tomados al azar. En el contexto de las series de tiempo, este fenómeno describe cómo una variable de interés evoluciona a lo largo del tiempo de manera que cada valor sucesivo depende del valor anterior más alguna variación aleatoria. Formalmente, se puede representar como:

$$P_t = P_{t-1} + \epsilon_t$$

donde  $P_t$  es el estado del sistema en el tiempo  $t$ , y  $\epsilon_t$  es un término aleatorio que representa la innovación o perturbación en el tiempo  $t$ , usualmente asumiendo una distribución normal con media cero.

Aunque el concepto de caminante aleatorio es simple, sus aplicaciones son vastas y variadas. En economía, se utiliza para modelar la hipótesis de mercados eficientes, sugiriendo que los precios de los activos financieros siguen un camino aleatorio y, por lo tanto, son fundamentalmente impredecibles. Esta hipótesis implica que la información pasada no tiene poder predictivo sobre los movimientos futuros de los precios.

En la física, el movimiento browniano es un ejemplo clásico de un caminante aleatorio, describiendo el movimiento aparentemente aleatorio de partículas suspendidas en un fluido. Además, en ecología, los patrones de movimiento de diferentes animales pueden ser modelados como caminantes aleatorios, ayudando a entender los procesos de búsqueda de alimento y la dispersión de poblaciones.

En este proyecto modelaremos dicho proceso estocástico mediante la animación del movimiento de partículas en una cuadrícula; tomando como elementos aleatorios los siguientes parámetros:

- Posición inicial  $(x_0, y_0)$ .
- Dirección de movimiento (del 1 al 8 en cuanto a pixeles).
- Distancia a recorrer en la dirección seleccionada

Dichos elementos serán generados mediante números random, los cuales seguirán distribuciones *normal*, *gamma* y *uniforme*, con el fin de observar el comportamiento del sistema ante diferentes distribuciones de probabilidad en su entrada.

Este proyecto consta de cuatro fases:

## I. Generación de Números Random con Distribución Uniforme

En primer lugar, generamos un total de 100,000 números aleatorios con distribución *uniforme* que se guardaron en un archivo de texto llamado `randomsUniforme.txt`, los cuales generaron un histograma como el siguiente:

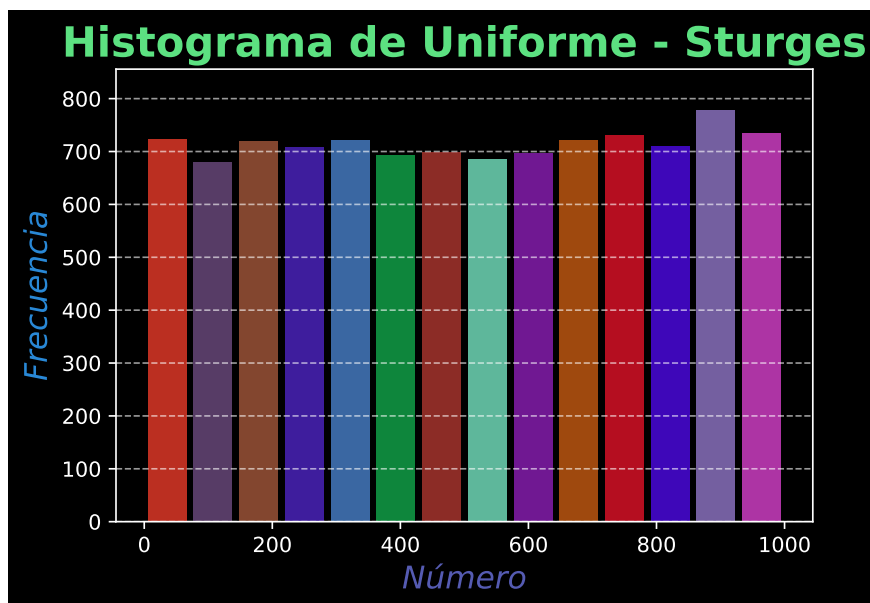


Figura 1: Histograma de números con distribución uniforme.

## II. Transformación de distribución

En segundo lugar, transformamos los números generados con una distribución uniforme mediante una técnica específica. Esta transformación nos permite obtener un nuevo conjunto de números aleatorios, pero esta vez siguiendo distribuciones de probabilidad distintas, concretamente la distribución normal y la gamma.

### II.1. Transformación a Distribución Normal

La transformación de variables aleatorias con distribución uniforme a una distribución normal se realiza mediante el uso de la función de error inversa (*inverse error function*), la cual juega un papel crucial en estadística para la transformación de variables aleatorias.

Dado un valor  $q$  de una distribución uniforme, el objetivo es encontrar el correspondiente valor en una distribución normal con media  $\mu$  y desviación estándar  $\sigma$ . La transformación se lleva a cabo en dos pasos principales:

La función de error inversa se calcula utilizando una aproximación basada en una fórmula específica. Para un valor  $x$ , la función se define como:

$$\text{inverse\_erf}(x) = \text{sign}(x) \cdot \sqrt{\sqrt{\left(\frac{2}{\pi a} + \frac{\log(1-x^2)}{2}\right)^2 - \frac{\log(1-x^2)}{a}} - \left(\frac{2}{\pi a} + \frac{\log(1-x^2)}{2}\right)}$$

donde  $a$  es un coeficiente que en esta aproximación toma el valor de 0.147. Esta función permite mapear los valores de  $x$  de la distribución uniforme a valores que siguen la función de error, la cual está estrechamente relacionada con la distribución normal.

Con la función de error inversa calculada, el siguiente paso es utilizarla para transformar el cuantil  $q$  de la distribución uniforme a un valor que siga una distribución normal estándar, y luego ajustarlo a cualquier media  $\mu$  y desviación estándar  $\sigma$  deseada. Esto se logra mediante la expresión:

$$\text{norm\_ppf}(q, \mu, \sigma) = \mu + \sigma \cdot \sqrt{2} \cdot \text{inverse\_erf}(2q - 1)$$

Este método asegura que los valores generados a partir de una distribución uniforme puedan ser transformados efectivamente a valores que siguen una distribución normal con los parámetros especificados.

Dicha transformación se aplicó a los 100,000 números distribuidos uniformemente, se guardó el resultado en un archivo `randomsNormal.txt` teniendo a la salida números normalmente distribuidos, como se muestra en el histograma:

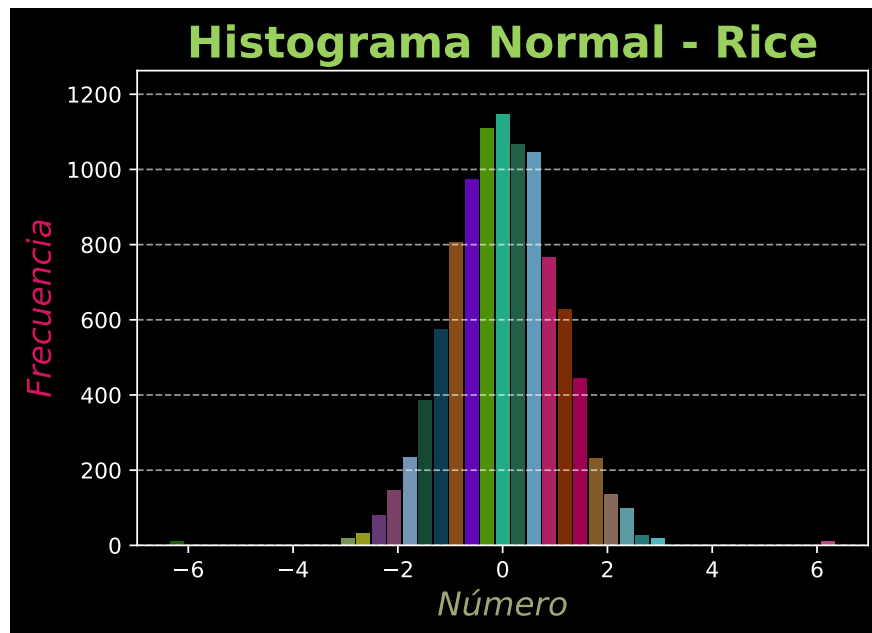


Figura 2: Histograma de números con distribución normal.

## II.2. Transformación a Distribución Gamma

Para generar variables aleatorias que sigan una distribución Gamma con parámetros de forma  $a > 1$  y escala  $\theta$ , utilizamos el siguiente algoritmo:

Definimos  $\lambda = a - 1$  y calculamos un valor  $c$  usando la expresión:

$$c = \frac{(a-1)^{a-1} e^{1-a}}{\Gamma(a)},$$

donde  $\Gamma(a)$  es la función Gamma para el parámetro  $a$ .

Repetimos los siguientes pasos hasta que se acepte un candidato:

1. Generamos una variable aleatoria  $y$  de una distribución exponencial con parámetro  $\lambda$ , utilizando la transformación  $y = -\frac{\log(1-U)}{\lambda}$ , donde  $U$  es un número aleatorio uniforme entre 0 y 1.
2. Generamos un número aleatorio uniforme  $u$ .
3. Aceptamos  $y$  si  $u \leq \frac{y^{a-1} e^{-y}}{c \lambda^{a-1} e^{-\lambda y}}$ . En caso contrario, rechazamos  $y$  y repetimos el proceso.

Una vez aceptado, el valor de salida es  $y \cdot \theta$ , asegurando que la variable aleatoria generada siga una distribución Gamma con los parámetros de forma  $a$  y escala  $\theta$ .

De esta manera, obtenemos números con distribución gamma que se almacenaron en `randomsGamma.txt`. Su transformación la observamos en el siguiente histograma:

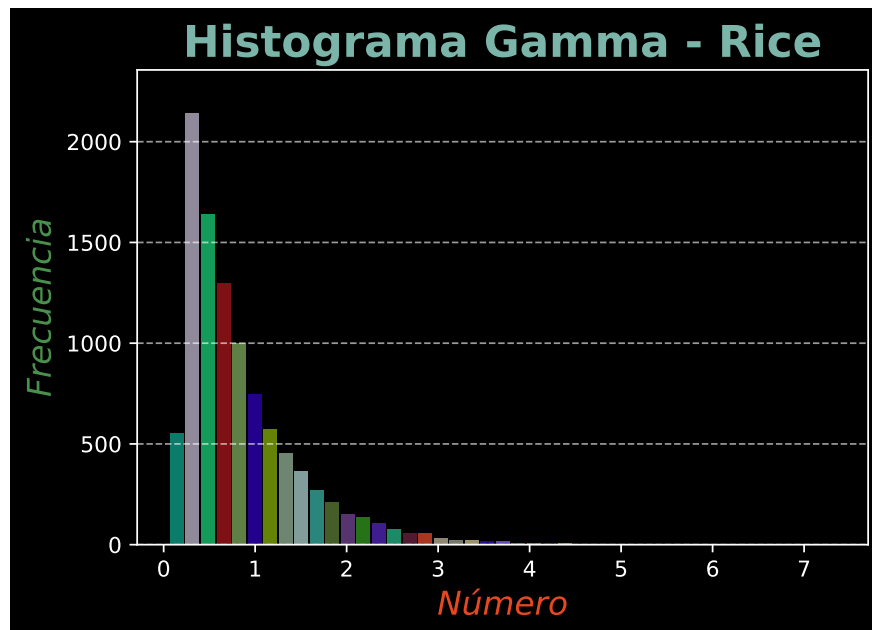


Figura 3: Histograma de números con distribución gamma.

### III. Métricas de las Distribuciones de Probabilidad

Se realizó una investigación respecto a las expresiones matemáticas que definen a las distribuciones de probabilidad *uniforme*, *gamma* y *normal* (las cuales se detallan en el documento **Métricas de Probabilidad**), que se codificaron como sigue para poder aplicarlas en nuestros set de datos anteriores:

#### III.1. Métricas para un set de datos genérico

```
1 class DistribucionProbabilidad:
2     def __init__(self, datos):
3         self.datos = np.array(datos)
4         self.n = len(datos)
5
6     def media(self):
7         return np.mean(self.datos)
8
9     def mediana(self):
10        return np.median(self.datos)
11
12    def moda(self):
13        frecuencias = {}
14        for dato in self.datos:
15            if dato in frecuencias:
16                frecuencias[dato] += 1
17            else:
18                frecuencias[dato] = 1
19
20        modas = [valor for valor, frecuencia in frecuencias.items()
21                  if frecuencia == max(frecuencias.values())]
22
23        if len(modas) == 1:
24            return modas[0]
25        else:
26            return modas
27
28    def media_geometrica(self):
29        return gmean(self.datos)
30
31    def asimetria(self):
32        media = self.media()
33        desv_std = self.desviacion_estandar()
34        suma_tercero_momento = np.sum((self.datos - media) ** 3)
35        skewness = (self.n / ((self.n - 1) * (self.n - 2))) * (
36            suma_tercero_momento / (desv_std ** 3))
37        return skewness
38
39    def rango(self):
40        return self.datos.max() - self.datos.min()
41
42    def desviacion_estandar(self):
43        return np.std(self.datos, ddof=1)
```

```

43     def varianza(self):
44         return np.var(self.datos, ddof=1)
45
46     def coeficiente_variacion(self):
47         return self.desviacion_estandar() / self.media()
48
49     def percentil(self, p):
50         if 0 <= p <= 100:
51             return np.percentile(self.datos, p)
52         else:
53             raise ValueError("El percentil debe estar entre 0 y
54                               100.")
55
56     def cuartil(self, q):
57         if q in [1, 2, 3]:
58             return self.percentil(q * 25)
59         else:
60             return np.nan
61
62     def curtosis(self):
63         media = self.media()
64         s = self.desviacion_estandar()
65         suma_cuarto_momento = np.sum((self.datos - media) ** 4)
66         curtosis = (self.n * (self.n+1) * suma_cuarto_momento) / ((
67             self.n-1) * (self.n-2) * (self.n-3) * s**4) - (3 * (self
68             .n-1)**2) / ((self.n-2) * (self.n-3))
69         return curtosis
70
71     def entropia(self):
72         p, counts = np.unique(self.datos, return_counts=True)
73         p = counts / len(self.datos)
74         entropia = -np.sum(p * np.log(p))
75         return entropia
76
77     def calcular_metricas(self, percentil, cuartil):
78         resultados = {
79             'Media': self.media(),
80             'Mediana': self.mediana(),
81             'Moda': self.moda(),
82             'Media Geometrica': self.media_geometrica(),
83             'Rango': self.rango(),
84             'Desviacion Estandar': self.desviacion_estandar(),
85             'Varianza': self.varianza(),
86             f'Percentil {percentil}': self.percentil(percentil),
87             f'Cuartil {cuartil}': self.cuartil(cuartil),
88             'Asimetria': self.asimetria(),
89             'Coeficiente de Variacion': self.coeficiente_variacion
90             (),
91             'Curtosis': self.curtosis(),
92             'Entropia': self.entropia()
93         }
94
95         for metrica, valor in resultados.items():
96             print(f"{metrica}: {valor}")

```

### III.2. Métricas para una Distribución Uniforme

```
1      class DistribucionUniforme(DistribucionProbabilidad):
2
3          def __init__(self, datos):
4              super().__init__(datos)
5              self.a = min(datos)
6              self.b = max(datos)
7
8          def media(self):
9              return (self.a + self.b)/2
10
11         def mediana(self):
12             return (self.a + self.b)/2
13
14         def media_geometrica(self):
15             return super().media_geometrica()
16
17         def moda(self):
18             return super().moda()
19
20         def rango(self):
21             return self.b - self.a
22
23         def varianza(self):
24             return ((self.b - self.a)**2)/12
25
26         def desviacion_estandar(self):
27             return (((self.b - self.a)**2)/12) ** (1/2)
28
29         def coeficiente_variacion(self):
30             return (((self.b - self.a)**2)/12) ** (1/2) / ((self.
31                 a + self.b) / 2)
32
33         def percentil(self, num_perc):
34             p = num_perc/100
35             return self.a + p*(self.b - self.a)
36
37         def cuartil(self, num_cuartil):
38             num_perc = num_cuartil * 25
39             return self.percentil(num_perc)
40
41         def asimetria(self):
42             return super().asimetria()
43
44         def curtosis(self):
45             return super().curtosis()
46
47         def entropia(self):
48             return math.log(self.b - self.a)
```



```
48
49         def calcular_metricas(self, percentil, cuartil):
50             return super().calcular_metricas(percentil, cuartil)
```

### III.3. Métricas para una Distribución Normal

```
1         class DistribucionNormal(DistribucionProbabilidad):
2
3             def __init__(self, datos):
4                 super().__init__(datos)
5                 self.mu = self.media()
6                 self.sigma = self.desviacion_estandar()
7
8             def media(self):
9                 return super().media()
10
11             def mediana(self):
12                 return self.mu
13
14             def media_geometrica(self):
15                 return "No aplica"
16
17             def moda(self):
18                 return self.mu
19
20             def rango(self):
21                 return super().rango()
22
23             def percentil(self, num_perc):
24                 p = num_perc/100
25                 if 0 <= p <= 1:
26                     valor_z = norm.ppf(p)
27                     return self.mu + valor_z * self.sigma
28                 else:
29                     return np.nan
30
31             def cuartil(self, num_cuartil):
32                 if num_cuartil in [1,2,3,4]:
33                     c = num_cuartil * 25
34                     return self.percentil(c)
35
36                 else:
37                     return np.nan
38
39             def varianza(self):
40                 return self.sigma**2
41
42             def desviacion_estandar(self):
43                 return super().desviacion_estandar()
44
45             def coeficiente_variacion(self):
46                 if self.mu != 0:
```

```

47         return self.sigma/self.mu
48     else:
49         return np.nan
50
51     def asimetria(self):
52         return super().asimetria()
53
54     def curtosis(self):
55         return super().curtosis()
56
57     def entropia(self):
58         return 0.5 * np.log(2 * np.pi * np.e * self.sigma ** 2)
59
60     def calcular_metricas(self, percentil, cuartil):
61         return super().calcular_metricas(percentil, cuartil)

```

### III.4. Métricas para una Distribución Gamma

```

1     class DistribucionGamma(DistribucionProbabilidad):
2
3         def __init__(self, datos):
4             super().__init__(datos)
5             self.k, self.theta = self.estimated_params(self.datos)
6
7         @staticmethod
8         def estimated_params(data):
9             def neg_log_likelihood(params, data):
10                 return -np.sum(gamma.logpdf(data, a=params[0],
11                                     scale=params[1]))
12
13             pesos_iniciales = [1.0, 1.0]
14             result = minimize(neg_log_likelihood, pesos_iniciales,
15                               args=(data), method='L-BFGS-B', bounds=((0, None),
16                               (0, None)))
17
18             k_est, theta_est = result.x
19             return k_est, theta_est
20
21     def media(self):
22         return self.k * self.theta
23
24     def mediana(self):
25         return super().mediana()
26
27     def media_geometrica(self):
28         return super().media_geometrica()
29
30     def moda(self):
31         if self.k > 1:
32             return (self.k - 1) * self.theta
33         else:

```

```
32         return np.nan
33
34     def rango(self):
35         return super().rango()
36
37     def varianza(self):
38         return self.k * self.theta**2
39
40     def desviacion_estandar(self):
41         return (self.k)**(1/2)*self.theta
42
43     def coeficiente_variacion(self):
44         return 1/(self.k)**(1/2)
45
46     def percentil(self, num_perc):
47         p = num_perc/100
48         return gamma.ppf(p, self.k, scale=self.theta)
49
50     def cuartil(self, num_cuartil):
51         c = num_cuartil * 25
52         return self.percentil(c)
53
54     def asimetria(self):
55         return super().asimetria()
56
57     def curtosis(self):
58         return 6/self.k
59
60     def entropia(self):
61         return self.k + np.log(self.theta) + gammaln(self.k) +
62             (1 - self.k) * psi(self.k)
63
64     def calcular_metricas(self, percentil, cuartil):
65         return super().calcular_metricas(percentil, cuartil)
```

Así, para los sets de datos `randomsUniforme.txt`, `randomsNormal.txt` y `randomsGamma.txt` se obtuvieron los siguientes resultados:

```

Distribución Uniforme
Media: 500.0
Mediana: 500.0
Moda: 994.0
Media Geométrica: 0.0
Rango: 1000.0
Desviación Estándar: 288.67513459481285
Varianza: 83333.33333333333
Percentil 20: 200.0
Cuartil 3: 750.0
Asimetría: 0.031077114339680464
Coeficiente de Variación: 0.5773502691896257
Curtosis: -1.1631198834242384
Entropía: 6.907755278982137

```

Figura 4: Métricas para el set `randomsUniforme.txt`.

```

Distribución Normal
Media: 0.01619842591887649
Mediana: 0.01619842591887649
Moda: 0.01619842591887649
Media Geométrica: No aplica
Rango: 12.727364117383477
Desviación Estándar: 1.0437856087734922
Varianza: 1.0894883970826499
Percentil 20: -0.8622737057227252
Cuartil 3: 0.7202211204387743
Asimetría: 0.02366738300979561
Coeficiente de Variación: 64.43747151734904
Curtosis: 2.4611763333062124
Entropía: 1.461792645995829

```

```

Distribución Gamma
Media: 0.8564571806387282
Mediana: 0.650667238385453
Moda: 0.4486894230367014
Media Geométrica: 0.6626515448038219
Rango: 7.288659290248604
Desviación Estándar: 0.5909616096086177
Varianza: 0.3492356240312083
Percentil 20: 0.3638674608659149
Cuartil 3: 1.1492325372283594
Asimetría: 3.111463611204522
Coeficiente de Variación: 0.6900071865447968
Curtosis: 2.856659504900796
Entropía: 0.714628803071913

```

Figura 5: Métricas para el set `randomsNormal.txt`.  
Figura 6: Métricas para el set `randomsGamma.txt`.

A continuación se analizan los resultados obtenidos al aplicar las métricas a los conjuntos de datos.

### III.5. Análisis de Resultados

#### III.5.1. Para la distribución uniforme

Tras aplicar las métricas estadísticas a una muestra de datos generados con una distribución uniforme, hemos obtenido los siguientes resultados:

- La **media** y la **mediana** tienen el mismo valor de 500.0, lo cual es esperable en una distribución uniforme donde todos los valores tienen la misma probabilidad.
- La **moda** se presenta en 994.0. En una distribución perfectamente uniforme, la moda no está bien definida ya que cada valor es igualmente probable. Sin embargo, debido a variaciones en la muestra, se puede presentar un pico en la frecuencia.
- La **media geométrica** es 0. Este resultado sugiere que la muestra puede contener valores en o cercanos a cero, lo que arrastra el promedio geométrico hacia abajo.

- El **rango** es 1000.0, lo que indica que los datos están esparcidos uniformemente a través de un intervalo amplio.
- La **desviación estándar** y la **varianza** son coherentes con la fórmula de distribución uniforme  $(b - a)^2/12$ , reflejando la dispersión de los datos.
- El **coeficiente de variación** es 0.577, lo que puede considerarse bajo, indicando una baja dispersión relativa respecto a la media.
- Los valores de **percentil 20** y **cuartil 3** son consistentes con lo que se esperaría en una distribución uniforme.
- La **asimetría** se presenta con un valor de 0.031. Esto indica que la distribución de los datos es casi simétrica, con una ligera inclinación hacia la derecha.
- La **curtosis** negativa sugiere una distribución más achatada que la distribución normal.
- La **entropía** es positiva y está relacionada con el logaritmo del rango de los datos, indicando un grado de incertidumbre asociado a la variabilidad de los datos dentro del rango.

### III.5.2. Para la distribución normal

- La **media**, la **mediana** y la **moda** tienen valores muy cercanos entre sí, lo que es característico de una distribución normal, que es simétrica alrededor de su media.
- El **rango** es aproximadamente 12.73, lo cual nos da una idea de la dispersión total de los datos desde el valor mínimo al máximo.
- La **desviación estándar** es de aproximadamente 1.04, indicando que la mayoría de los datos se encuentran dentro de este rango alrededor de la media en una distribución normal.
- La **varianza**, siendo el cuadrado de la desviación estándar, refleja la dispersión de los datos y en este caso es aproximadamente 1.09.
- El **percentil 20** es aproximadamente -0.86, lo que indica que el 20 % de los valores son menores que este número.
- El **cuartil 3** es aproximadamente 0.70, mostrando que el 75 % de los datos están por debajo de este valor.
- La **asimetría** tiene un valor cercano a cero (aproximadamente 0.024), lo que confirma la simetría de la distribución. Una distribución perfectamente normal tendría una asimetría de cero.
- El **coeficiente de variación** es bastante alto, aproximadamente 64.44, lo que sugiere una variabilidad relativa alta en comparación con la media.
- La **curtosis** es aproximadamente 2.46, lo cual es ligeramente superior a 3, que es la curtosis de una distribución normal estándar. Esto indica una distribución un poco más puntiaguda que la normal, con colas más pesadas.
- La **entropía**, de aproximadamente 1.46, aunque no es comúnmente discutida en el contexto de la distribución normal, proporciona una medida de incertidumbre o aleatoriedad en la distribución de los datos.

### III.5.3. Para la distribución Gamma

- La **media** es aproximadamente 0.856, lo cual representa el promedio de los valores de la muestra.
- La **mediana** de 0.6506 es menor que la media, lo que indica una asimetría hacia la derecha en la distribución.
- La **moda** es aproximadamente 0.4486, lo que es esperable en una distribución Gamma asimétrica, donde la moda es menor que la media y la mediana.
- La **media geométrica** es aproximadamente 0.6626, siendo relevante en distribuciones sesgadas y generalmente menor que la media aritmética para distribuciones con asimetría positiva.
- El **rango**, de aproximadamente 7.29, muestra la diferencia entre los valores máximo y mínimo en la muestra.
- La **desviación estándar**, cercana a 0.59, indica la cantidad de variación o dispersión de los valores en la muestra.
- La **varianza** es aproximadamente 0.3492, proporcionando otra medida de la dispersión de los datos.
- El **percentil 20** es aproximadamente 0.3638, indicando que el 20 % de los valores en la distribución son inferiores a este valor.
- El **cuartil 3** es de 1.1492, lo que sugiere que el 75 % de los datos son menores que este valor.
- La **asimetría** tiene un valor alto de 3.111, confirmando el sesgo positivo de la distribución y la concentración de la mayoría de los datos hacia el lado izquierdo de la media.
- El **coeficiente de variación**, aproximadamente 0.69, muestra una variabilidad relativa en relación con la media de la muestra.
- La **curtosis**, de alrededor de 2.856, indica un pico más alto y colas más gruesas en comparación con una distribución normal (curtosis normal = 3).
- La **entropía**, de aproximadamente 0.7146, es una medida de la incertidumbre inherente en los resultados de la distribución Gamma.

### III.6. Transformaciones al sistema

Con la intención de seguir caracterizando otros sets de datos y observar como se conservan los sistemas dinámicos cuando a su entrada reciben un conjunto de números aleatorios y se procesan a través de transformaciones **lineales y no lineales** se les aplicó a cada set de datos (`randomsUniforme.txt`, `randomsNormal.txt` y `randomsGamma.txt`) las siguientes funciones:

### III.6.1. Transformación lineal: $x^2 - 5x + 7$

Para la transformación cuadrática con origen en datos uniformemente distribuidos:

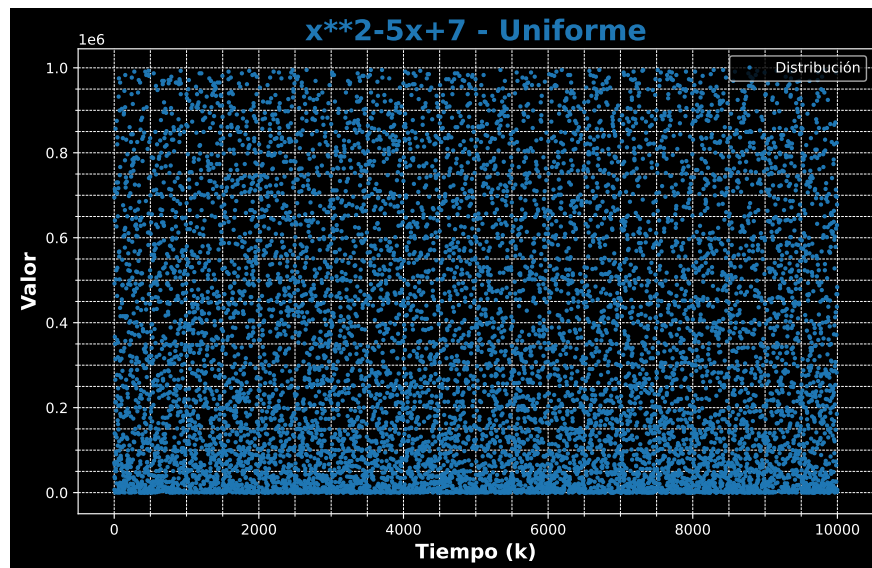


Figura 7: Gráfica de la serie de tiempo para  $x^2 - 5x + 7$  con datos de distribución normal.

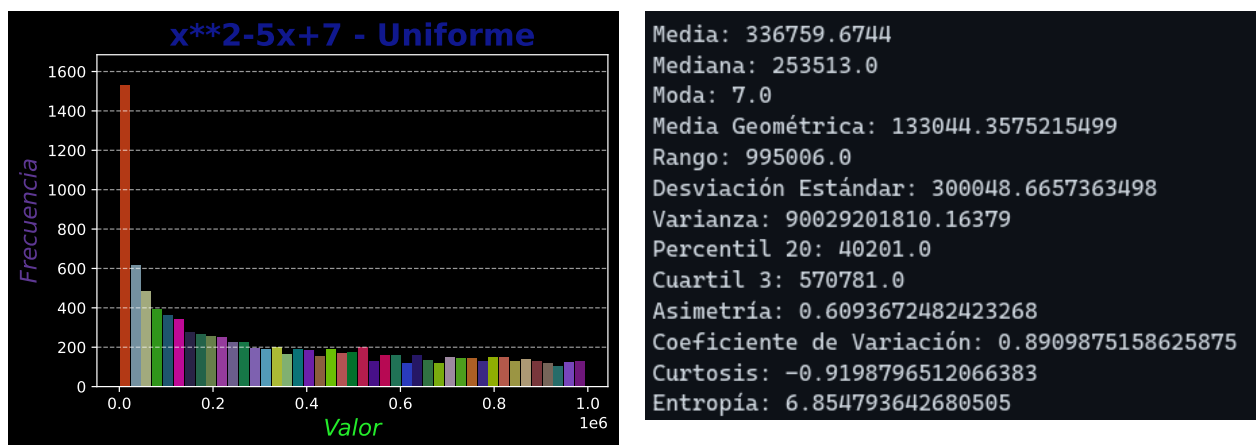


Figura 8: Histograma.

Figura 9: Métricas

La serie de tiempo dada presenta una fluctuación que parece uniforme a lo largo de todos sus valores. El histograma asociado y las métricas calculadas revelan las siguientes características:

- La **media** de 336759, siendo el promedio de los datos, sugiere que la suma total de la serie de tiempo es considerablemente alta. En una distribución uniforme ideal, la media y la mediana son cercanas, lo que no se observa en este conjunto de datos.
- La **mediana** de 253513, al ser menor que la media, puede indicar que una cantidad significativa de valores de la serie de tiempo son relativamente bajos, lo que podría desplazar la media hacia arriba.

- La **moda** de 7 es atípicamente baja y específica en comparación con el rango de los datos. La presencia de una moda tan baja podría ser indicativa de un valor atípico o una acumulación de frecuencias en un punto muy específico de la distribución.
- La **media geométrica**, aproximadamente 133041, es relevante en la descripción de los datos cuando la distribución no es simétrica. Su valor inferior a la media aritmética indica una distribución sesgada positivamente.
- El **rango** de aproximadamente 995006 muestra que los datos se extienden sobre un intervalo muy amplio.
- La **desviación estándar** de 300018 es alta y sugiere que los datos están muy dispersos en relación con la media.
- El **percentil 20** de 40201 señala que hay una gran cantidad de datos que son muy bajos en comparación con el valor medio, lo cual apoya la existencia de asimetría positiva.
- El **cuartil 3** de 570781 indica que hay una amplia dispersión en los valores superiores de la serie de tiempo. En una distribución uniforme, los cuartiles estarían más equitativamente espaciados.
- La **asimetría** positiva de 0.69 sugiere una concentración de datos en el extremo inferior y una cola más larga hacia el lado derecho de la distribución, lo que es indicativo de que hay valores extremadamente altos que están alejando la media de la mediana.
- El **coeficiente de variación** de aproximadamente 0.89 indica una alta variabilidad relativa de los datos. Este valor alto sugiere una distribución con mayor dispersión en relación con la media.
- La **curtosis** de aproximadamente -0.19 es inferior a la de una distribución normal (con una curtosis de 0), lo que indica que la distribución es más plana que la normal y que tiene colas más ligeras.
- La **entropía** de 6.85 es una medida de la incertidumbre y, aunque no proporciona una indicación directa del tipo de distribución, un valor alto sugiere una mayor aleatoriedad en los datos.

Las métricas obtenidas sugieren una distribución con una asimetría positiva y un pico más plano que una distribución normal. La presencia de valores extremos es notable y aleja las características observadas de una distribución uniforme estándar. Con base a estas métricas, los datos parecen alinearse más estrechamente con una distribución que permite valores atípicos y una variabilidad considerable, tal como podría ser una distribución gamma o log-normal, dependiendo de la forma de la distribución observada en los datos.



Para la transformación cuadrática con origen en datos normalmente distribuidos:

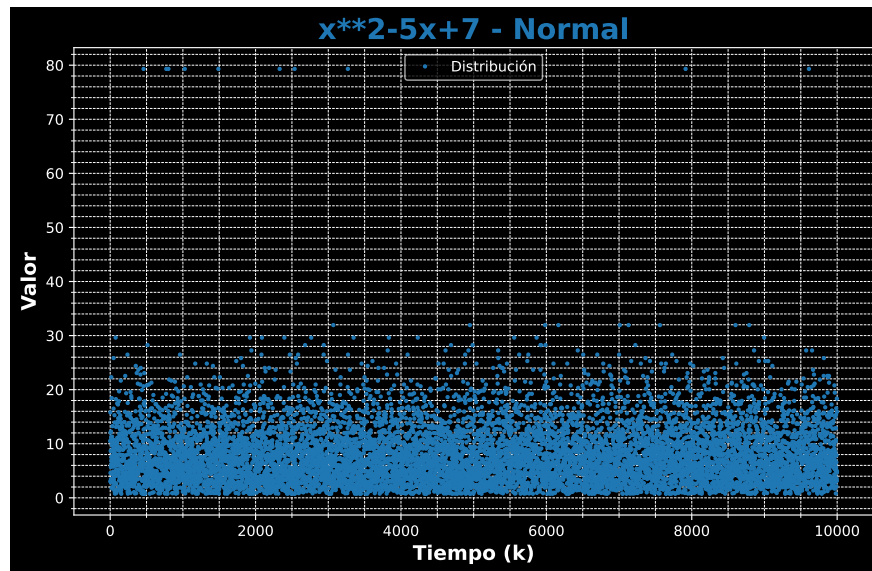


Figura 10: Gráfica de la serie de tiempo para  $x^2 - 5x + 7$  con datos de distribución uniforme.

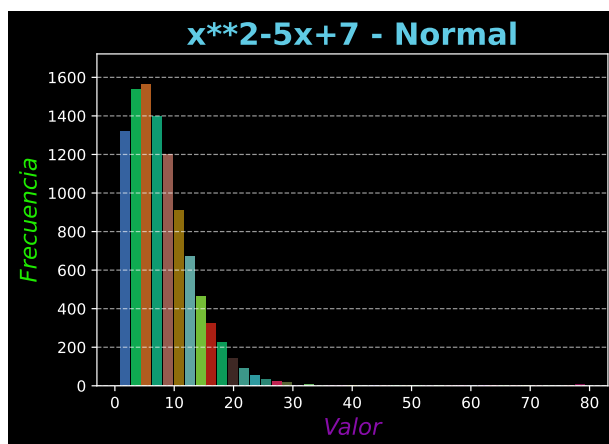


Figura 11: Histograma.

```

Media: 8.008649707650807
Mediana: 6.9499671269505185
Moda: 0.7500822667630311
Media Geométrica: 6.258708843603558
Rango: 78.56477737081079
Desviación Estándar: 5.648955167825946
Varianza: 31.91069448810746
Percentil 20: 3.4165709163423283
Cuartil 3: 10.807966950540425
Asimetría: 2.803216672360859
Coeficiente de Variación: 0.7053567547634649
Curtosis: 25.37734960554213
Entropía: 6.859100114400464
  
```

Figura 12: Métricas

- La **media** de aproximadamente 8.0 indica el valor promedio de la serie de tiempo. Un valor medio bajo en relación con la amplitud del rango sugiere una concentración de valores bajos.
- La **mediana** de 6.9, al estar por debajo de la media, refuerza la indicación de una asimetría positiva, donde más de la mitad de los valores están por debajo del promedio.
- La **moda** de 0.75 sugiere que el valor más frecuente es mucho más bajo que la media y la mediana, lo que podría ser indicativo de una acumulación de frecuencias cerca del límite inferior de los datos.
- La **media geométrica**, con un valor de aproximadamente 6.25, es menor que la media aritmética, lo que se alinea con una distribución que tiene una asimetría hacia la derecha.
- El **rango** es de aproximadamente 78.5, lo que indica la extensión total desde el valor más bajo hasta el más alto en la serie de tiempo.

- La **desviación estándar** es de aproximadamente 6.5, lo que muestra una variabilidad considerable alrededor de la media.
- La **varianza**, que es el cuadrado de la desviación estándar, nos dice cómo se distribuyen los valores alrededor de la media, y con un valor de aproximadamente 31.9, confirma la dispersión observada.
- El **percentil 20** de alrededor de 3.4 indica que el 20 % de los valores son inferiores a este número, lo que nos ayuda a entender la distribución de los valores más bajos.
- El **cuartil 3** de aproximadamente 10.8 indica que el 75 % de los datos se encuentran por debajo de este valor, proporcionando una medida de la distribución en el extremo superior.
- La **asimetría** con un valor de aproximadamente 2.8 sugiere una fuerte inclinación hacia la derecha, lo que significa que hay una cola más larga de valores hacia el extremo superior de la serie de tiempo.
- El **coeficiente de variación** de alrededor de 0.7 indica una variabilidad relativa moderada en relación con el valor medio.
- La **curtosis**, de aproximadamente 25.4, es mucho más alta que 0, lo que indica una distribución muy puntiaguda con colas pesadas, o la presencia de valores atípicos significativos.
- La **entropía**, con un valor de aproximadamente 6.85, refleja la aleatoriedad en la serie de tiempo. Una entropía alta sugiere una mayor incertidumbre y variabilidad en la distribución de los datos.

La combinación de una alta curtosis y una asimetría significativa sugiere que la distribución de la serie de tiempo no sigue un patrón uniforme ni normal. La presencia de una moda baja y una asimetría positiva sugiere una concentración de valores más bajos y la posibilidad de valores atípicos extremos. Estas características pueden estar asociadas con distribuciones que presentan asimetría positiva y colas pesadas, como ciertas variantes de la distribución gamma o distribuciones log-normales, pero siempre dependiendo de la naturaleza específica de los datos en estudio.

Para la transformación cuadrática con origen en datos de distribución gamma:

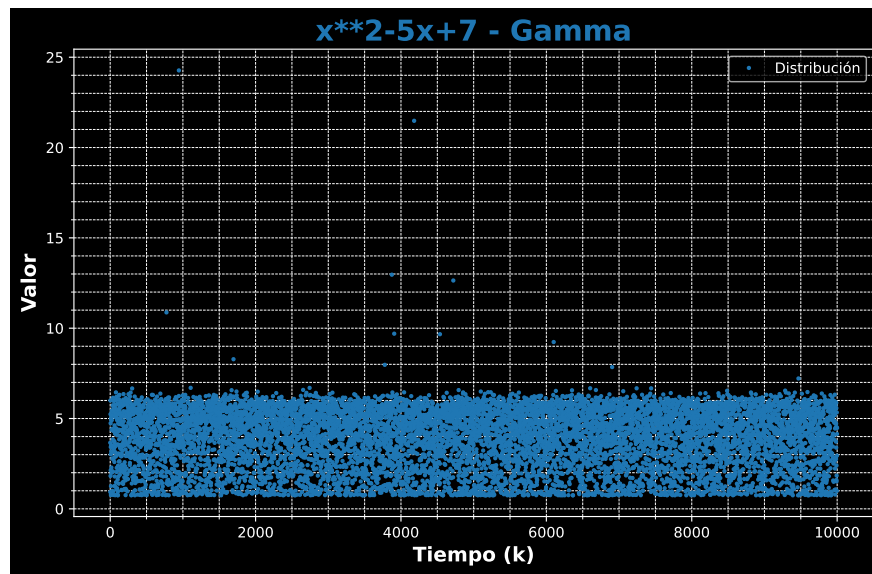


Figura 13: Gráfica de la serie de tiempo para  $x^2 - 5x + 7$  con datos de distribución gamma.

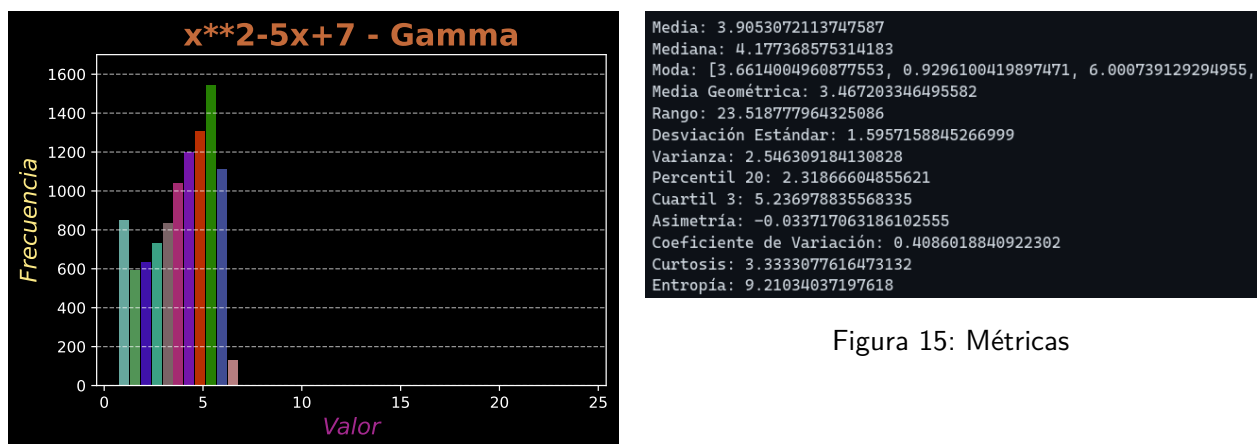


Figura 14: Histograma.

Figura 15: Métricas

- La **media** de aproximadamente 3.91 indica el promedio de los valores observados en la serie de tiempo, lo que puede sugerir una acumulación de eventos alrededor de este punto.
- La **mediana** de 4.12, siendo ligeramente mayor que la media, apunta a una distribución con una cantidad equilibrada de valores por encima y por debajo de este punto central, lo que indica cierta simetría en la distribución de los datos.
- La **moda** reporta múltiples valores (3.66, aproximadamente 0.93 y 6.00), indicando que hay varios picos de frecuencia en la distribución de los datos. Esto podría sugerir una distribución multimodal o la presencia de grupos distintos dentro de los datos.
- La **media geométrica** de 3.47 es una medida de tendencia central que es típicamente más representativa para datos que no siguen una distribución normal, especialmente cuando la distribución es sesgada.

- El **rango** de 23.58 muestra la extensión total de los datos desde el valor más bajo hasta el más alto, lo que indica una variabilidad considerable en la serie de tiempo.
- La **desviación estándar** de aproximadamente 1.60 refleja una variabilidad moderada de los datos alrededor de la media, lo que sugiere que los eventos de la serie de tiempo presentan una dispersión consistente.
- La **varianza** de aproximadamente 2.55, siendo el cuadrado de la desviación estándar, confirma la dispersión de los datos alrededor de la media.
- El **percentil 20** de 2.32 sugiere que el 20 % de los valores son menores a este número, lo que puede ser utilizado para comprender la concentración de los valores más bajos.
- El **cuartil 3** de aproximadamente 5.24 indica que el 75 % de los datos son menores que este valor, ofreciendo una visión de la distribución acumulativa hacia el extremo superior.
- Una **asimetría** ligeramente negativa de -0.03 sugiere que la serie de tiempo es casi simétrica, con una leve inclinación hacia la izquierda, lo que implica que hay una cantidad ligeramente mayor de eventos por encima de la media.
- El **coeficiente de variación** de 0.41 indica que la variabilidad de los datos es proporcionalmente moderada en relación con la media.
- Una **curtosis** de aproximadamente 3.33 sugiere una distribución con un pico similar a una distribución normal (curtosis normal = 3), lo que implica que la serie de tiempo no tiene colas inusualmente pesadas ni ligeras.
- La **entropía** de 9.21 es una medida de la incertidumbre y aleatoriedad en la serie de tiempo, con un valor más alto sugiriendo una mayor diversidad o aleatoriedad en la distribución de los datos.

En conjunto, estas métricas describen una serie de tiempo con una distribución de valores relativamente simétrica y moderadamente dispersa. La presencia de múltiples modas puede indicar la presencia de subgrupos o comportamientos distintos dentro de la serie de tiempo. La distribución de los datos no parece ser excesivamente sesgada ni puntiaguda, lo que podría ser consistente con una variedad de distribuciones de probabilidad no especificadas sin características extremas o atípicas.

### III.6.2. Transformación no lineal: $5 \cdot \cos\left(\frac{x}{2}\right) + 7 \cdot \cos\left(\frac{x^2}{4}\right)$

Para la transformación no lineal con origen en datos de distribución uniforme:

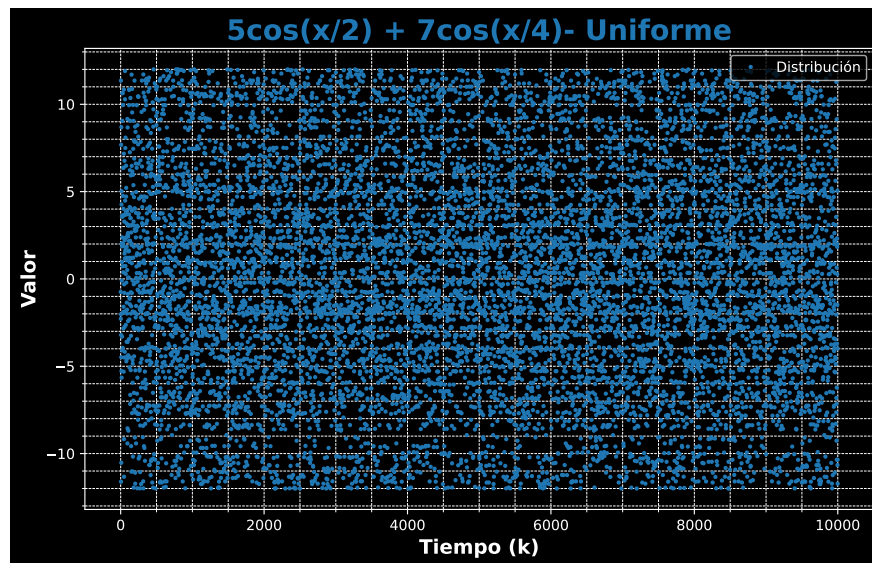


Figura 16: Gráfica de la serie de tiempo para  $5 \cdot \cos\left(\frac{x}{2}\right) + 7 \cdot \cos\left(\frac{x^2}{4}\right)$  con datos de distribución uniforme.

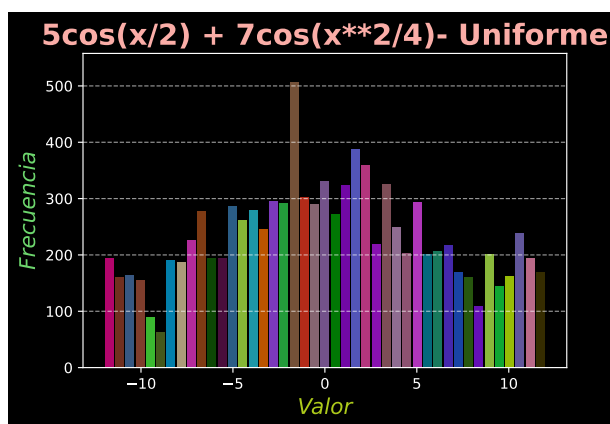


Figura 17: Histograma.

```

Media: 0.07500094721304476
Mediana: -0.1414125866542264
Moda: 2.0219426827710847
Media Geométrica: nan
Rango: 23.999599193264014
Desviación Estándar: 6.142673705819402
Varianza: 37.73244025616507
Percentil 20: -5.65388656326405
Cuartil 3: 4.714552886269873
Asimetría: 0.02803509669008664
Coeficiente de Variación: 81.90128170475985
Curtosis: -0.8046105401758425
Entropía: 6.859100114400464
  
```

Figura 18: Métricas

- La **media** cercana a 0.075 sugiere que el promedio de los valores oscila alrededor de cero, lo cual podría indicar una distribución de datos centrada.
- La **mediana** de -0.14, siendo ligeramente negativa, muestra que más de la mitad de los valores están por debajo de cero, lo que podría señalar una leve inclinación de los datos hacia valores negativos.
- La **moda** de aproximadamente 2.02 indica el valor que ocurre con mayor frecuencia en la serie de tiempo, sugiriendo un pico en la distribución en ese punto.
- La **media geométrica** no es aplicable (NaN), típicamente debido a la presencia de valores negativos en los datos que hacen imposible calcular esta métrica.

- El **rango** de aproximadamente 23.99 refleja una variabilidad significativa en los datos desde el valor más bajo al más alto.
- La **desviación estándar** de alrededor de 6.14 indica que los valores de la serie de tiempo están dispersos de manera considerable respecto a la media.
- La **varianza** de aproximadamente 37.73 cuantifica la dispersión observada y es consistente con una desviación estándar elevada.
- El **percentil 20** de -5.54 muestra que el 20 % de los valores están concentrados hacia el extremo inferior del rango de la serie de tiempo.
- El **cuartil 3** de aproximadamente 4.71 indica que el 75 % de los valores son inferiores a este número, proporcionando una perspectiva de la distribución hacia el extremo superior.
- Una **asimetría** de aproximadamente 0.02 sugiere que la serie de tiempo es casi simétrica con una muy ligera inclinación hacia la derecha.
- El **coeficiente de variación** de alrededor de 81.90 es extremadamente alto y destaca una variabilidad relativa excepcionalmente grande en relación con la media, lo que puede indicar una amplia gama de comportamientos o estados dentro de la serie de tiempo.
- Una **curtosis** de aproximadamente -0.84 es menor que cero, lo que indica una distribución más aplanada que la normal y con colas más ligeras.
- La **entropía** de 6.85 muestra la incertidumbre y la aleatoriedad en la distribución de la serie de tiempo, siendo un valor alto que sugiere una gran diversidad en los datos.

Este análisis muestra una serie de tiempo con una distribución de valores que tiene una gran variabilidad y una distribución aplanada. La simetría casi perfecta y la alta variabilidad relativa podrían ser características de una serie de tiempo con influencias aleatorias significativas y sin un patrón o tendencia dominante.

Para la transformación no lineal con origen en datos de distribución normal:

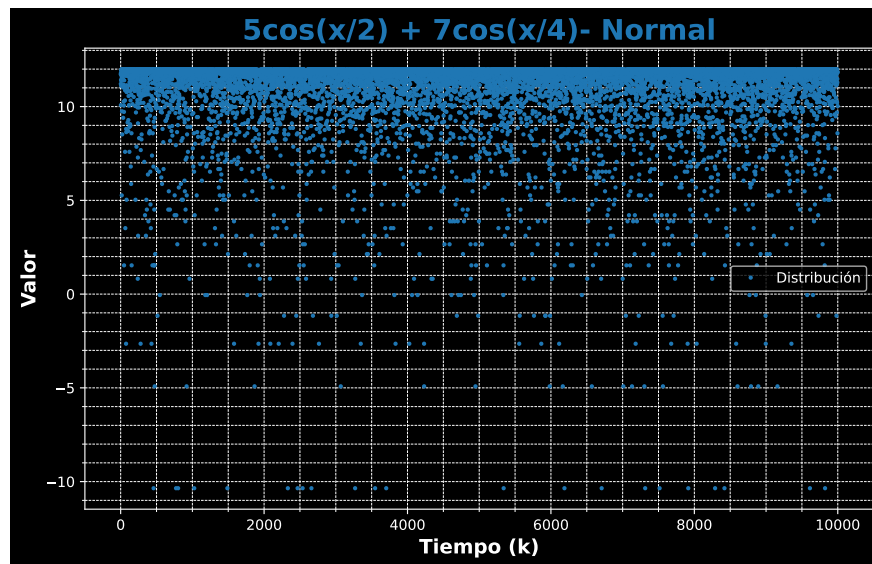


Figura 19: Gráfica de la serie de tiempo para  $5 \cdot \cos\left(\frac{x}{2}\right) + 7 \cdot \cos\left(\frac{x^2}{4}\right)$  con datos de distribución normal.

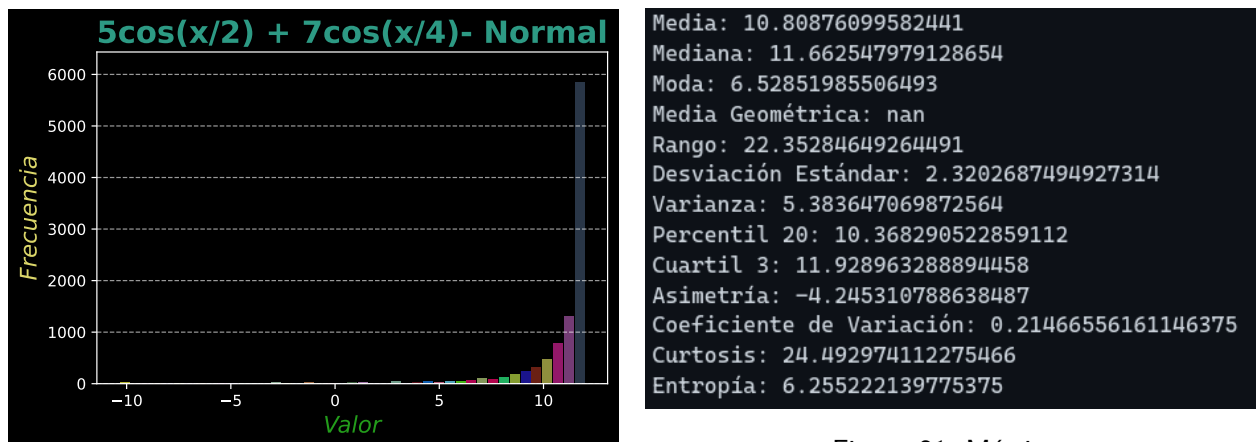


Figura 21: Métricas

Figura 20: Histograma.

- La **media** de aproximadamente 10.81 indica el valor promedio de la serie de tiempo, lo cual sugiere una tendencia central hacia el extremo superior del rango de los datos.
- La **mediana** de 11.66, ligeramente superior a la media, puede señalar una ligera asimetría en los datos, con una concentración de valores más altos.
- La **moda** de aproximadamente 5.63 representa el valor más frecuente dentro de la serie de tiempo y parece estar significativamente desplazada de la media y la mediana, lo que indica una distribución de los valores potencialmente bimodal o multimodal.
- La **media geométrica** no se puede calcular (NaN) debido a la presencia de valores negativos en los datos.



- El **rango** de alrededor de 22.35 muestra una variabilidad significativa entre los valores máximos y mínimos observados en la serie de tiempo.
- La **desviación estándar** de aproximadamente 2.32 refleja la cantidad de dispersión que existe alrededor de la media de los datos.
- La **varianza** de alrededor de 5.38 cuantifica la dispersión de los datos y, junto con la desviación estándar, confirma la presencia de variabilidad en la serie de tiempo.
- El **percentil 20** de 10.36 indica que el 20 % de los datos son menores o iguales a este valor, proporcionando información sobre la distribución acumulativa de los valores inferiores.
- El **cuartil 3** de 11.93 refleja el valor por debajo del cual se encuentra el 75 % de los datos, lo que ayuda a entender la concentración de los valores superiores.
- Una **asimetría** de -4.24 muestra un sesgo considerable hacia la izquierda, indicando una concentración de valores más hacia el lado derecho de la media.
- El **coeficiente de variación** de 0.21 sugiere que la variabilidad de los datos es baja en comparación con la media, lo que indica una relativa uniformidad en la dispersión de los datos.
- La **curtosis** extremadamente alta de 24.49 sugiere una distribución muy puntiaguda con colas pesadas, lo que puede implicar la presencia de valores atípicos o un agrupamiento muy denso alrededor de la media.
- La **entropía** de 6.25 refleja la incertidumbre y la aleatoriedad de los datos. Un valor más alto indica una mayor diversidad o imprevisibilidad en la distribución de la serie de tiempo.

Estas métricas indican una serie de tiempo con un marcado pico central y una notable simetría hacia la derecha, aunque la asimetría negativa y la curtosis alta sugieren una complejidad subyacente. Estos hallazgos son típicos de series que contienen eventos o comportamientos extremos y valores que se aglomeran en torno a un punto central.



Para la transformación no lineal con origen en datos de distribución gamma:

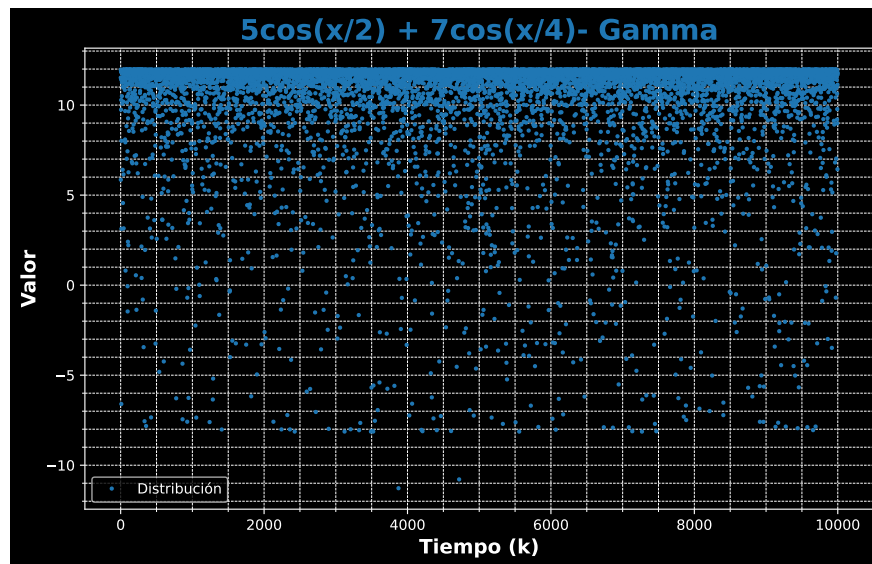


Figura 22: Gráfica de la serie de tiempo para  $5 \cdot \cos\left(\frac{x}{2}\right) + 7 \cdot \cos\left(\frac{x^2}{4}\right)$  con datos de distribución gamma.

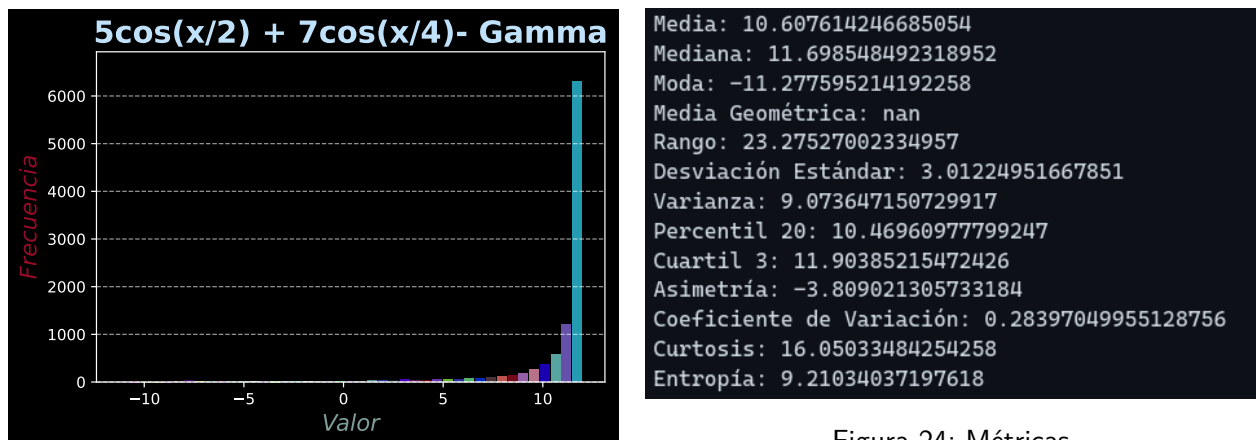


Figura 23: Histograma.

Figura 24: Métricas

- La **media** de aproximadamente 10.61 puede sugerir que la distribución de los valores tiende a ser mayor en comparación con el punto de equilibrio de cero, indicando una posible acumulación de eventos en el lado positivo.
- La **mediana** de 11.69, mayor que la media, indica que más de la mitad de los valores están por encima del promedio, reforzando la idea de que hay una tendencia hacia valores mayores en la serie de tiempo.
- La **moda** de -11.28 es inusualmente baja y distinta de la media y la mediana, lo cual es característico de una distribución que podría ser bimodal o tener un comportamiento atípico en su distribución de valores.

- La **media geométrica** no se calcula (NaN), lo que comúnmente ocurre debido a la presencia de valores negativos en los datos, lo que impide el cálculo de esta métrica.
- El **rango** de 23.25 muestra la diferencia entre los valores máximo y mínimo en la serie de tiempo, lo que indica una extensa variabilidad de los datos.
- La **desviación estándar** de aproximadamente 3.01 sugiere una dispersión considerable de los datos alrededor de la media, lo que indica una amplia gama de comportamientos o valores en la serie de tiempo.
- La **varianza** de aproximadamente 9.07, como el cuadrado de la desviación estándar, cuantifica la dispersión de los datos y es consistente con una desviación estándar alta.
- El **percentil 20** de 10.47 muestra que el 20 % de los datos son menores o iguales a este valor, proporcionando una indicación de la distribución acumulativa de los valores inferiores.
- El **cuartil 3** de 11.90 indica que el 75 % de los valores están por debajo de este número, lo que ayuda a entender la concentración de los valores superiores en la serie de tiempo.
- Una **asimetría** de -3.89 sugiere un sesgo significativo hacia la izquierda, lo que indica una cola larga de eventos hacia el lado izquierdo de la distribución.
- El **coeficiente de variación** de 0.28 sugiere una baja variabilidad relativa de los datos en comparación con la media, lo que podría indicar una concentración de valores alrededor de la media.
- La **curtosis** extremadamente alta de 16.05 indica una distribución con un pico muy agudo y colas pesadas, lo que puede sugerir la presencia de valores extremos o comportamientos muy concentrados alrededor de la media.
- La **entropía** de 2.91 refleja la incertidumbre y la diversidad en la distribución de la serie de tiempo, con un valor más alto indicando una mayor imprevisibilidad o variedad en los datos.

Estos resultados nos llevan a concluir que estamos observando una serie de tiempo con una distribución de datos que tiene una asimetría significativa hacia la izquierda y una curtosis muy elevada. Estas características sugieren una serie con una gran cantidad de valores concentrados alrededor de un punto central y una posible presencia de valores atípicos que influyen en la forma de la distribución.

## IV. Caminante Aleatorio

Con base en los datos generados con una distribución uniforme (`randomsUniforme.txt`), distribución normal (`randomsNormal.txt`) y distribución gamma `randomsGamma.txt` se plantea crear la animación del caminante aleatorio con el fin de observar su comportamiento con distintas distribuciones de probabilidad, y analizar las series de tiempo (con las métricas) generadas por:

- Dirección tomada.
- Distancia recorrida en cada paso.
- Choques con pared
- Posición (x,y) en el plano a lo largo del tiempo.

*Nota:* Por simplicidad, aunque se observen más caminantes aleatorios en cada animación, se analizará solamente al primero.

#### IV.1. Caminante Aleatorio con Distribución Uniforme

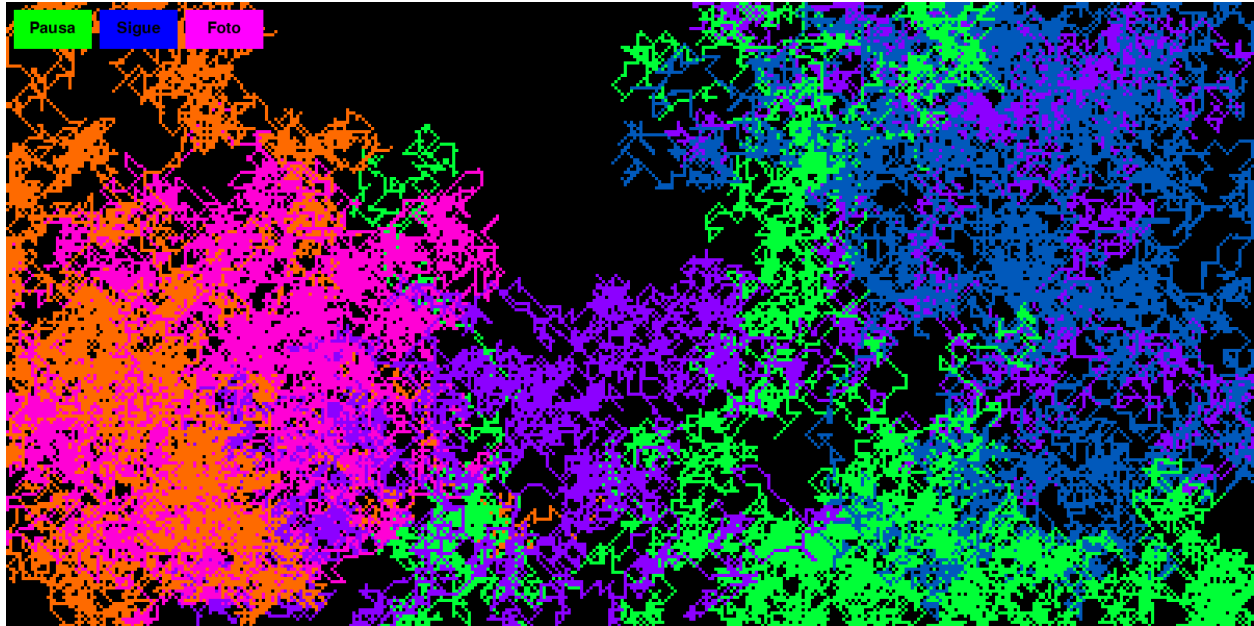


Figura 25: Caminante Aleatorio - Distribución Uniforme