TP 2.2: GENERADORES DE NÚMEROS PSEUDOALEATORIOS DE DISTINTAS DISTRIBUCIONES DE PROBABILIDAD.

Tomás Ponce
Legajo: 44954
Mail: tomasponce@outlook.com.ar
UTN - FRRO
Zeballos 1341, S2000

Sofía Gasparini
Legajo:44762.
Mail: sofigasparini15@gmail.com
UTN - FRRO
Zeballos 1341, S2000

ABSTRACT

Nuestro trabajo tiene como objetivo mostrar el proceso, resultados y testeo de las distribuciones formadas a partir de la generación de números aleatorios. La programación del trabajo se realizó con el lenguaje Python en su versión 3.7.

1. Introducción

Se quiere simular y poder observar la generación de los distintos tipos de distribuciones a través de la utilización de números pseudoaleatorios. Los tipos de distribuciones que desarrollaremos son: binomial, empírica, exponencial, gamma, hipergeométrica, normal, pascal, poisson y uniforme. Luego de realizar las simulaciones, procederemos a aplicar el test de bondad Chi-Cuadrado a las distribuciones uniforme,exponencial,normal,binomial,poisson y empirica discreta.

2. Números Pseudoaleatorios

Antes de las computadoras, existieron diferentes métodos para generar secuencias de números aleatorios, tales como los procedimientos físicos(monedas, dados, bolilleros, etc), tablas de 40000 dígitos aleatorios por Tipett en 1927(no resultaron con distribución uniforme),tabla de 100.000 números aleatorios hecha por Kendall en 1939, tabla de 1.000.000 numeros aleatorios a partir de ruido electrónico hecha por Rand Corporation en 1955, etc.

Las principales desventajas de todos estos intentos de realizar una secuencia de números aleatorios de manera física son que: no pueden repetirse una misma secuencia, no hay velocidad computacional e incorporar una tabla a la computadora implica gran costo de almacenamiento en relación a la cantidad de números (si se quiere realizar una tabla infinitamente aleatoria)

En 1946 Jon Von Neuman sugirió usar las operaciones aritméticas de una computadora para generar secuencias de número pseudoaleatorios, mediante el "método del valor medio". Este generador cae rápidamente en ciclos cortos, por ejemplo, si aparece un cero se propagará por siempre.

A inicios de 1950s se exploró el método y se propusieron mejoras, por ejemplo para evitar caer en cero. Metrópolis logró obtener una secuencia de 750,000 números distintos al usar semillas de 38 bits (usaba sistema binario), además la secuencia de Metrópolis mostraba propiedades deseables. No obstante, el método del valor medio no es considerado un método bueno por lo común de los ciclos cortos.

A partir de estos primeros métodos de generación de números pseudoaleatorios empezaron a surgir nuevos y mejores métodos para su utilización.

3. Distribuciones de probabilidad

En teoría de la probabilidad y estadística, la distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los sucesos y cada uno de los sucesos es el rango de valores de la variable aleatoria. También puede decirse que tiene una relación estrecha con las distribuciones de frecuencia. De hecho, una distribución de probabilidades puede comprenderse como una frecuencia teórica, ya que describe cómo se espera que varíen los resultados.

La distribución de probabilidad está completamente especificada por la función de distribución, cuyo valor en cada x real es la probabilidad de que la variable aleatoria sea menor o igual que x.

Los tipos de variables existentes son:

• Variable aleatoria: Es aquella cuyo valor es el resultado de un evento aleatorio. Lo que quiere decir que son los resultados que se presentan al azar en cualquier evento o experimento.

• Variable aleatoria discreta: Es aquella que solo toma ciertos valores (frecuentemente enteros) y que resulta principalmente del conteo realizado.

• Variable aleatoria continua: Es aquella que resulta generalmente de la medición y puede tomar cualquier valor dentro de un intervalo dado

3.1. Distribuciones discretas

3.1.1. Distribución de probabilidad Pascal o Binomial Negativa

En estadística la distribución binomial negativa es una distribución de probabilidad discreta que incluye a la distribución de Pascal. Es una ampliación de las distribuciones geométricas, utilizada en procesos en los cuales se ve necesaria la repetición de ensayos hasta conseguir un número de casos favorables (primer éxito).

La variable aleatoria es el número de ensayos Bernoulli necesarios para obtener el primer éxito. Si deseamos conocer el número de estos para conseguir n éxitos, la variable aleatoria es binomial negativa.

El número de experimentos de Bernoulli de parámetro p independientes realizados hasta la consecución del k-ésimo éxito es una variable aleatoria que tiene una distribución binomial negativa con parámetros k y p.

Su función de probabilidad es

$$f_b(x;k,p) = \binom{x-1}{x-k} p^k (1-p)^{x-k}$$

para enteros x mayores o iguales que k, donde

$$\binom{x-1}{k-1} = \binom{x-1}{x-k} = \frac{(x-1)!}{(k-1)!(x-k)!}$$

Su media si se piensa en el número de fracasos únicamente es

$$\mu = \frac{k(1-p)}{p}$$

Su media si se piensa en el número de fracasos únicamente es

$$\mu = \frac{k}{p}$$

Su varianza en ambos casos es

$$\sigma^2 = \frac{k(1-p)}{p^2}$$

3.1.2. Distribución de probabilidad Binomial

En estadística, la distribución binomial es una distribución de probabilidad discreta que cuenta el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí, con una probabilidad fija p de ocurrencia del éxito entre los ensayos.

Un experimento de Bernoulli se caracteriza por ser dicotómico, esto es, solo dos resultados son posibles. A uno de estos se denomina éxito y tiene una probabilidad de ocurrencia p y al otro, fracaso, con una probabilidad q = 1 - p.

En la distribución binomial el anterior experimento se repite n veces, de forma independiente, y se trata de calcular la probabilidad de un determinado número de éxitos. Para n = 1, la binomial se convierte, de hecho, en una distribución de Bernoulli.

Para representar que una variable aleatoria X sigue una distribución binomial de parámetros n y p, se escribe:

$$X \sim B(n, p)$$

La distribución binomial es la base del test binomial de significación estadística.

Su función de probabilidad es:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad 0 \leq p \leq 1$$

donde $x = \{0, 1, 2, \ldots, n\}$,

siendo $\binom{n}{x} = \dfrac{n!}{x!(n-x)!}$

Las combinaciones de n sobre x (n elementos tomados de x en x)

donde:

- n = número de ensayos
- p = probabilidad de éxito
- x = variable aleatoria binomial

3.1.3. Distribución de probabilidad Hipergeométrica

En teoría de la probabilidad la distribución hipergeométrica es una distribución discreta relacionada con muestreos aleatorios y sin reemplazo. Suponga que se tiene una población de N elementos de los cuales, d pertenecen a la categoría A y N-d a la B. La distribución hipergeométrica mide la probabilidad de obtener x elementos de la categoría A en una muestra sin reemplazo de n elementos de la población original.

La función de probabilidad de una variable aleatoria con distribución hipergeométrica puede deducirse a través de razonamientos combinatorios y es igual a

$$P(X = x) = \frac{\binom{d}{x}\binom{N-d}{n-x}}{Nn}$$

donde:

- N es el tamaño de población
- n es el tamaño de la muestra extraída
- d es el número de elementos en la población original que pertenecen a la categoría deseada
- x es el número de elementos en la muestra que pertenecen a dicha categoría.

La notación:

$$\binom{a}{x}$$

hace referencia al coeficiente binomial, es decir, el número de combinaciones posibles al seleccionar x elementos de un total a.

El valor esperado de una variable aleatoria X que sigue la distribución hipergeométrica es:

$$E[X] = \frac{nd}{N}$$

La varianza es:

$$Var[X] = npq\frac{N-n}{N-1}.$$

La distribución hipergeométrica es aplicable a muestreos sin reemplazo.

3.1.4. Distribución de probabilidad Poisson

En teoría de probabilidad y estadística, la distribución de Poisson es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad de que ocurra un determinado número de eventos durante cierto período de tiempo. Concretamente, se especializa en la probabilidad de ocurrencia de sucesos con probabilidades muy pequeñas, o sucesos raros.

La función de densidad de probabilidad de la distribución de Poisson es:

$$f(k, \lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$$

donde

- k es el número de ocurrencias del evento o fenómeno (la función nos da la probabilidad de que el evento suceda precisamente k veces).
- es un parámetro positivo que representa el número de veces que se espera que ocurra el fenómeno durante un intervalo dado.

Tanto el valor esperado como la varianza de una variable aleatoria con distribución de Poisson son iguales a .

Sumas de variables aleatorias de Poisson La suma de variables aleatorias de Poisson independientes es otra variable aleatoria de Poisson cuyo parámetro es la suma de los parámetros de las originales. Dicho de otra manera, si

$$X_i \sim \mathrm{Poi}(\lambda_i)\,, i = 1, \ldots, N$$

son N variables aleatorias de Poisson independientes, entonces

$$Y = \sum_{i=1}^{N} X_i \sim \mathrm{Poi}\left(\sum_{i=1}^{N} \lambda_i\right)$$

Aproximación normal: Como consecuencia del teorema central del límite, para valores grandes de λ , una variable aleatoria de Poisson X puede aproximarse por otra normal dado que el cociente

$$Y = \frac{X - \lambda}{\sqrt{\lambda}}$$

converge a una distribución normal de media 0 y varianza 1.

3.1.5. Distribución de probabilidad Empírica discreta

En esta sección trataremos un método un tanto más general que puede emplearse para simular cualquiera de las siguientes distribuciones: empírica, discreta, continua.

Sin embargo, nos abstendremos de utilizar este método para generar valores de variable aleatoria a partir de distribuciones de probabilidad estándares, debido a que cualquiera de los métodos descritos previamente se espera que proporcionan resultados más satisfactorios, desde el punto de vista de la velocidad de computación, facilidad de programación y requisitos de almacén o memoria. En otros términos, el método que se propondrá en esta sección es una que se utiliza siempre y cuando no se disponga de otra alternativa.

Sea X una variable aleatoria discreta con P (X = b1) = pi.

En consecuencia, resulta evidente que un método para generar X en una computadora es aquel que genera un valor de baile ahora aleatoria y sujeto a una instrucción el informe, el intervalo (0, 1) y un conjunto de valores X= bi siempre que se satisfaga:

$$p_1 + ... + p_{i-1} < r < p_1 + ... + p_i$$

Pese a que se han desarrollado un buen número de técnicas de búsqueda basadas en este método, en su gran mayoría requieren un programa relativamente complejos que a su vez emplean un tiempo de computación excesivo.

Uno de los procedimientos más rápidos para generar valores de variable aleatoria discreta es el desarrollado por G. Marsaglia, quien presupone la disponibilidad de una computadora decimal cuyos bloques o palabras de memoria pueden referirse mediante números. Esta última característica en realidad constituye una propiedad de la gran mayoría de las computadoras actuales. Conviene hacer notar que si bien este método es extremadamente rápido, también requiere por lo menos una memoria de 1000 palabras. Existe otro método desarrollado también por Marsaglia que en forma alternativa utiliza mucho menos capacidad de memoria aunque incrementa ligeramente el tiempo de computación.

3.2. Distribuciones continuas

3.2.1. Distribución de probabilidad Uniforme

En teoría de probabilidad y estadística, la distribución uniforme continua es una familia de distribuciones de probabilidad para variables aleatorias continuas, tales que para cada miembro de la familia, todos los intervalos de igual longitud en la distribución en su rango son igualmente probables. El dominio está definido por dos parámetros, a y b, que son sus valores mínimo y máximo.

La distribución es escrita como $U(a, b)$

La función de densidad de probabilidad de la distribución uniforme continua es:

f(x)= $\begin{cases} \frac{1}{(b-a)} & si \quad a < x < b \\ 0 & si \quad x > b \quad o \quad x < a \end{cases}$

Los valores en los dos extremos a y b no son por lo general importantes porque no afectan el valor de las integrales de f(x)dx sobre el intervalo, ni de xf(x)dx o expresiones similares. A veces se elige que sean cero, y a veces se los elige con el valor $\dfrac{1}{b-a}$.

Su esperanza se calcula como:

$$\frac{a+b}{2},$$

Su varianza se calcula como:

$$\frac{(b-a)^2}{12}.$$

3.2.2. Distribución de probabilidad Exponencial

En estadística la distribución exponencial es una distribución de probabilidad continua con un parámetro $\lambda > 0$

Su función de densidad es:

f(x)= $\begin{cases} \lambda e^{-\lambda x} & si \quad x \quad > \quad 0 \\ 0 & en \quad caso \quad contrario \end{cases}$

El valor esperado y la varianza de una variable aleatoria X con distribución exponencial son:

$$E(X) = \frac{1}{\lambda}, \qquad V(X) = \frac{1}{\lambda^2}$$

La distribución exponencial es un caso particular de distribución gamma con k = 1. Además la suma de variables aleatorias que siguen una misma distribución exponencial es una variable aleatoria expresable en términos de la distribución gamma.

3.2.3. Distribución de probabilidad Gamma

En estadística la distribución gamma es una distribución de probabilidad continua con dos parámetros k>0 y $\lambda > 0$.

La función de densidad para valores x>0 es:

$$f(x) = \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{\Gamma(k)}$$

El valor esperado y la varianza de una variable aleatoria X de distribución gamma son:

$$E[X] = k/\lambda = k\theta$$
$$V[X] = k/\lambda^2 = k\theta^2$$

3.2.4. Distribución de probabilidad Normal

En estadística y probabilidad se llama distribución normal, distribución de Gauss, distribución gaussiana o distribución de Laplace-Gauss, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en estadística y en la teoría de probabilidades.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro estadístico. Esta curva se conoce como campana de Gauss y es el gráfico de una función gaussiana.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos.Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.

De hecho, la estadística descriptiva sólo permite describir un fenómeno, sin explicación alguna. Para la explicación causal es preciso el diseño experimental, de ahí que al uso de la estadística en psicología y sociología sea conocido como método correlacional.

La distribución normal también es importante por su relación con la estimación por mínimos cuadrados, uno de los métodos de estimación más simples y antiguos

En probabilidad, la distribución normal aparece como el límite de varias distribuciones de probabilidad continuas y discretas.

La función de distribución de la distribución normal está definida como sigue:

$$\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma^2}(u)\, du = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}}\, du, \quad x \in \mathbb{R}.$$

donde:

- μ es la media (también puede ser la mediana, la moda o el valor esperado, según aplique)
- σ es la desviación típica [estándar es un anglicismo]
- σ^2 es la varianza
- φ representa la función de densidad de probabilidad

También podemos definir la normal a través de la función de densidad:

$$\phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

La función de distribución normal estándar es un caso especial de la función donde μ =0 y σ =1:

$$\Phi(x) = \Phi_{0,1}(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{u^2}{2}}\, du, \quad x \in \mathbb{R}.$$

4. Métodos de generación de distribuciones aleatorias

Cuando se establecen las bases relacionales subyacentes al empleo de los métodos existentes para generar valores de variables estocasticas en una computadora digital, se parte de dos problemas un tanto divergentes. Estos dos problemas de tipo distinto se pueden clasificar convenientemente como deterministicos, es decir, no probabilístico os o bien, como estocasticos.

En un principio, los métodos de simulación estocastica fueron aplicados por los matemáticos y los científicos relacionados con las áreas de la física, para resolver ciertos problemas deterministicos que se podían expresar mediante ecuaciones matemáticas para las cuales sus soluciones no resultaban fáciles de obtener, utilizando los criterios convencionales de los métodos numéricos o analíticos.

4.1. Método de la transformación inversa

Es un método para la generación de números aleatorios de cualquier distribución de probabilidad cuando se conoce las inversa de su función de distribución acumulada.

Si deseamos generar los valores x de la variable aleatoria a partir de cierta estadística de población cuya función de densidad esté dada por f(x) debemos en primer lugar obtener la función de distribución acumulativa F(X). Puesto que F(x) se define sobre el rango de 0 a 1, podemos generar números aleatorios distribuidos uniformemente y además hacer F(x)=r. Resulta claro, entonces, como queda f determinada únicamente por r= F(x) Sigue, por lo tanto, que para cualquier valor particular de r, que generemos, por ejemplo r0, siempre es posible encontrar el valor de x, en este caso x0, que corresponde a r0 debido a la función inversa de F, si es conocida. Esto es:

$$x_0 = F^{-1}(r_0)$$

Dónde:

$$F^{-1}(x)$$

es la transformación inversa (o mapeo) de r sobre el intervalo unitario en el dominio de x. Si generamos números uniformes aleatorios correspondiente a una F(x) dada, podemos resumir, matemáticamente, este método como sigue:

$$r = F(x) = \int_{-\infty}^{x} f(t)dt \qquad (0)$$

$$P(X \leq x) = F(x) = P[r \leq F(x)] = P[F^{-1}(r) \leq x]$$

$$F^{-1}(r)$$

es una variable que tiene a F(x) como función de densidad de probabilidad.

4.1.1. Método de la transformación inversa para la distribución de probabilidad Uniforme

Para simular una distribución uniforme sobre cierto intervalo conocido (a,b) deberemos, en primer lugar, obtener la transformación inversa para la distribución uniforme acumulativa:

$$F(x) = \int_{0}^{a} \frac{1}{(b-a)}dt = \frac{x-a}{b-a} \qquad (0)$$

0$\leq F(x) \leq 1$

Entonces la transformación inversa tiene la forma:

$$x = a + (b-a)r$$

$$0 \leq r \leq 1$$

En seguida, generamos un conjunto de números aleatorios correspondientes al rango de las probabilidades acumulativas, es decir, los valores de variables aleatorias uniformes definidas sobre el rango 0 a 1. Cada número aleatorio r determina, de manera única, un valor de la variable x uniformemente distribuida.

4.1.2. Método de la transformación inversa para la distribución de probabilidad Exponencial

Existen muchas maneras para lograr la generación de valores de variables aleatorias exponenciales. Puesto que la distribución acumulativa existe explícitamente:

$$F(x) = \int_{0}^{x} \alpha e^{-\alpha t}dt = 1 - e^{-\alpha x} \qquad (0)$$

la técnica de la transformación inversa nos permite desarrollar métodos directos para dicha generación. Debido a la simetría que existe entre la distribución uniforme sigue que la intercambiabilidad de F(x) y 1- F(x) . Por lo tanto:

$$r = e^{-\alpha x}$$

y consecuentemente:

$$x = -\frac{1}{\alpha} log(r) = -EXlog(r).$$

Por consiguiente, para cada valor del número pseudoaleatorio r, se determina un único valor para x. Los valores de x toman tan sólo magnitudes no negativas, debido a que el log (r) es ≤ 0 para ser $0 \leq$ r ≤ 1 y además se ajustan a la función de densidad exponencial con un valor esperado x. Es importante notar que pese a que esta técnica parece en principio muy simple, es preciso recordar que en una computadora digital el cálculo logarítmico natural involucra una expansión en serie de potencias para cada valor de la variable aleatoria un informe que se debe generar.

4.1.3. Método de la transformación inversa para la distribución de probabilidad Normal

El procedimiento para simular valores normales utilizando computadoras requiere el uso de la suma de K valores de variables aleatorias distribuidos uniformemente; esto es la suma de r1, r2... rk con cada ri definida en el intervalo 0 <ri <1. Aplicando la convención notacional de la forma matemática del teorema central del límite, encontramos que:

$$\sigma = \frac{b-a}{(\sqrt{12})} = \frac{1}{(\sqrt{12})}$$

$$z = \frac{\sum_{i=1}^{k} ri - K/2}{\sqrt{K/12}}$$

Pero por definición, z es un valor de variable aleatoria con distribución normal estándar.

Por lo tanto:

$$x = \sigma_x (\frac{12}{k})^{1/2} (\sum_{i=1}^{k} ri - K/2) + \mu_x$$

Por lo tanto, con la ecuación anterior, podemos proporcionar una formulación muy simple para generar valores de variable aleatoria normalmente distribuidos. Para generar un solo valor x bastará con sumar K números aleatorios definidos en el intervalo de 0 a 1. Este procedimiento se puede repetir tantas veces como valores de variables aleatorias normalmente distribuidos se quieran.

El valor de K que debe aplicarse las fórmulas usualmente se determina al establecer las condiciones de balance entre eficiencia de cómputo y la precisión. Al considerar la convergencia asintomática implicada por el procedimiento de límite central, es deseable que K corresponda a un número muy grande .Considerando el tiempo de comprende la generación de K valores uniformes por cada valor de variable aleatoria normal, sería preferible que K estuviera asociada a un número muy chico.

4.2. Método del rechazo

Si f(x) es una función acotada y x tiene además un rango finito, con a<x<b entonces se puede utilizar la técnicas de rechazos para generar los valores de variables aleatorias. La aplicación de esta técnica requiere que se proceda de acuerdo con las siguientes etapas:

1- Normalizar el rango de f mediante un factor de escala c tal que :

$$c.f(x) \leq 1$$
$$a \leq b$$

2-Definir a x como una función lineal de r, es decir:

$$x = a + (b - a)r$$

3- Generar parejas de números aleatorios(r_1, r_2)

4- Siempre que se encuentren una parejas de números aleatorios satisfagan la relación:

$$r_2 \leq c.f[a + (b - a)r_1]$$

dicho par sera aceptado y se utilizara a x=a+(b-a)r_1 como el valor generado de la variable aleatoria.

La teoría sobre la que se apoya este método se basa en el hecho ya conocido relativo a la probabilidad de que r sea menor o igual a c.f(x) es:

$$P[r \leq c.f(x)] = f(x)$$

Se ha demostrado que la esperanza matemática del numero de intentos que se realizan ,antes de encontrar una pareja exitosa es igual a 1/c. Esto implica que para ciertas funciones de densidad de probabilidad este método puede resultar sumamente ineficaz

4.3. Método de composición

Otro método para generar valores de variables estocásticas utilizando computadoras es el llamado método de composición. En este método se expresa a f(x) como una mezcla probabilística de las funciones de densidad g(x) seleccionadas adecuadamente. En términos matemáticos tenemos:

$$f(x) = \sum g_n(x)p_n$$

Para seleccionar las g(x) está dada sobre las consideraciones relativas a la bondad del ajuste y al objetivo de minimizar:

$$f(x) = \sum T_n p_n$$

donde T_n es el tiempo esperado de computación para generar valores de variables aleatorias a partir de $g_n(x)$.

5. Metodología

Utilizando el lenguaje de programación Python implementamos un programa con el propósito de generar números pseudoaleatorios para la formación de las distintas distribuciones. Utilizaremos las librerías Random, Numpy, Scipy, Matplotlib y math para estas tareas.

En el presente trabajo, realizamos nueve distribuciones diferentes, cuatro de ellas de variables aleatorias continuas y las cinco restantes de variables aleatorias discretas. Mediante el método de la transformación inversa pudimos generar determinados números que permitieron, en su conjunto, formar las distintas distribuciones aleatorias.

Luego de la generación de las distintas distribuciones, procedimos a realizar el test chi cuadrado correspondiente a cada distribución. En nuestro trabajo, aplicamos un test chi- cuadrado de distinta manera para cada distribución.

Con los conocimientos obtenidos de la cátedra Probabilidades y estadística e investigaciones propias, pudimos obtener a partir de la función de distribución acumulada de cada una de las distribuciones, las frecuencias esperadas para cada intervalo de los test Chi-cuadrado realizados,

Para los generadores de números pseudoaleatorios de distintas distribuciones,generamos un total de 1000 números , ya que con dicha cantidad, creemos que para la temática del trabajo y los fines que se piensan realizar, podemos considerar dicho valor como aceptable.

Realizamos mas de 30 simulaciones para la generación de números pseudoaleatorios de las distintas distribuciones realizadas. Además se aplicara un test chi-cuadradado que nos indicaran si efectivamente los resultados son pseudoaleatorios para cada distribución.A partir de dichas simulaciones sacamos las respectivas conclusiones.

6. Exposición del código realizado en Python y explicación del mismo

En esta sección mostraremos el código realizado en Python para la obtención de las distintas distribuciones a partir de la generación de números pseudoaleatorios

6.1. Código de distribución Uniforme

```
def uniforme (a,b):
    x=[]
    for i in range(1000):
        r = round(random.random(), 4)
        x.append(a+(b-a)*r)
    return x

uni=(uniforme(1,3))
```

Listing 1: Algoritmo de generación de numeros pseudoaleatorios de distribución Uniforme

En el código se puede observar como se ingresa como parámetro dos valores a y b que expresan los limites de la distribución uniforme.Se generaran mil valores, y cada uno se generara a través del método de la transformación inversa con una variable r cuyo valor es generado por el generador Mersenne twister propio de Python

6.2. Código de distribución Exponencial

```
def exponencial(ex):
    x = []
    for i in range(1000):
        r = random.random()
        x += [-ex*(np.log(r))]
    return x


expo=(exponencial(5))
```

Listing 2: Algoritmo de generación de numeros pseudoaleatorios de distribución Exponencial

En el caso de la generación de valores para la distribución exponencial, también se realizaran mil tiradas y se generaran los valores a partir del método de la transformada inversa. El parámetro ex es la media de la distribución exponencial y la variable r cuyo valor es generado por el generador Mersenne twister propio de Python.

6.3. Código de distribución Gamma

```
def gamma(k,a):
    x=[]
    for i in range(1, 1000):
        tr=1.0
        for j in range(1,k):
            r = random.random()
            tr=tr*r
        x.append(-(math.log10(tr))/a)
    return x

gam=(gamma(5,20))
```

Listing 3: Algoritmo de generación de numeros pseudoaleatorios de distribución Gamma

Para la distribución Gamma, se realizaran 1000 tiradas y se generaran los valores a partir del método de la transformada inversa.El parámetro k representa la cantidad de eventos sucesivos y a representa el parámetro lambda a utilizar.El valor de la variable r es generado por el generador Mersenne twister propio de python.

6.4. Código de distribución Normal

```
def normal(ex,stdx):
    x=[]
    for i in range(1000):
        sum = 0.0
        for j in range (12):
            r=random.random()
            sum += r
        x+=[stdx*(sum-6.0)+ex]
    return x
nor=normal(2.35,30)
```

Listing 4: Algoritmo de generación de numeros pseudoaleatorios de distribución Normal

En el código se puede observar como se ingresa como parámetro la media(ex) y la desviación estándar (stdx).Se generaran mil valores, y cada uno se generara a través del método de la transformación inversa con una variable r cuyo valor es generado por el generador Mersenne twister propio de Python.

6.5. Código de distribución Pascal

```
def pascal(k,q):
    nx = []
    for i in range(1000):
        tr = 1
        qr = math.log10(q)
        u = generadorRandu()
        for j in range(k):
            r = random.random()
            tr *= u[j]
        x = int(math.log10(tr)//qr)
        nx.append(x)
    return nx

pas=pascal(5,0.4)
```

Listing 5: Algoritmo de generación de numeros pseudoaleatorios de distribución Pascal

En el caso de la generación de valores para la distribución Pascal, también se realizaran mil tiradas y se generaran los valores a partir del método de la transformada inversa. El parámetro k es la cantidad de repeticiones hasta el éxito y el parámetro es la probabilidad de no tener éxito. El valor de la variable r es generado por el generador Mersenne twister propio de Python.

6.6. Código de distribución Binomial

```python
def binomial (n,p):
    x=[]
    for i in range(1000):
        y=0.0
        for j in range(1,n):
            r = random.random()
            if (r-p) <0:
                y+=1.0
        x.append(y)
    return x


bino=binomial(1000,0.4)
```

Listing 6: Algoritmo de generación de numeros pseudoaleatorios de distribución Binomial

Para la distribución Binomial, se realizaran 1000 tiradas y se generaran los valores a partir del método de la transformada inversa.El parámetro n representa la cantidad de ensayos independientes de Bernoulli y p representa la probabilidad de éxito .El valor de la variable r es generado por el generador Mersenne twister propio de Python.

6.7. Código de distribución Hipergeométrica

```python
def hipergeometrica(tn,ns,p):
    x=[]
    for i in range(1000):
        tn1=tn
        ns1=ns
        p1=p
        y=0.0
        for j in range(1, ns1):
            r = random.random()
            if(r-p1) > 0:
                s=0.0
            else:
                s=1.0
                y+=1.0
            p1 = (tn1*p1-s)/(tn1 -1.0)
            tn1 -= 1.0
        x.append(y)
    return x

hipergeo=hipergeometrica(5000000,500,0.4)
```

Listing 7: Algoritmo de generación de numeros pseudoaleatorios de distribución Hipergeometrica

En el código se puede observar como se ingresa como parámetro la población total tn, la muestra seleccionada ns y un índice de proporción de 0.4.Se generaran mil valores, y cada uno se generara a través del método de la transformación inversa con una variable r cuyo valor es generado por el generador Mersenne twister propio de Python.

6.8. Código de distribución de Poisson

```
def poisson (lamb):
    x = []
    for i in range(1000):
        cont=0
        tr=1
        b=0
        while(tr-b >= 0):
            b = math.exp(-lamb)
            r = random.random()
            tr=tr*r
            if(tr-b >= 0):
                cont+=1
        x.append(cont)
    return x

poi=poisson(50)
```

Listing 8: Algoritmo de generación de numeros pseudoaleatorios de distribución poisson

En el caso de la generación de valores para la distribución Poisson, también se realizaran mil tiradas y se generaran los valores a partir del método de la transformada inversa. Se ingresara un parámetro cuyo valor sera 50, representando a lambda. El valor de la variable r es generado por el generador Mersenne twister propio de Python.

6.9. Código de distribución Empírica

```
def empirica():
    x=[]
    p=[0.273,0.037,0.195,0.009,0.124,0.058,0.062,0.151,0.047,0.044]
    for i in range(1000):
        r=random.random()
        a=0
        z=1
        for j in p:
            a+=j
            if (r<=a):
                break
            else:
                z+=1
        x.append(z)
    return x

empi=empirica()
```

Listing 9: Algoritmo de generación de numeros pseudoaleatorios de distribución Empirica discreta

Por ultimo, para la distribución Empírica discreta, se realizaran 1000 tiradas y se generaran los valores a partir del método de la transformada inversa.No se ingresara ningún parámetro, definiendo nosotros una serie de probabilidades cuya sumatoria da como resultado el valor de uno .El valor de la variable r es generado por el generador Mersenne twister propio de Python.

7. Exposición de las gráficas y análisis de las mismas

En la presente sección mostraremos las gráficas de las distintas distribuciones realizadas junto a un análisis de los resultados. Las presentes gráficas serán realizadas a partir de las variables aleatorias continuas y discretas propuestas.Las gráficas serán histogramas de frecuencias absolutas.

7.1. Gráficas de las distribuciones de probabilidad continuas

7.1.1. Distribución de probabilidad Uniforme

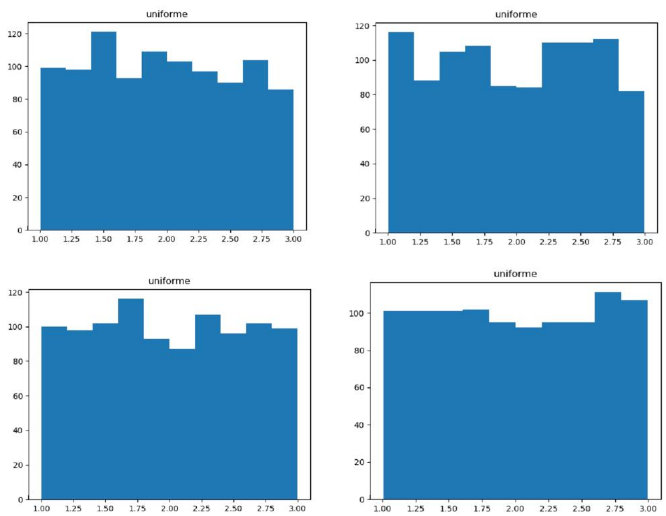

Figura 1: Distintas gráficas obtenidas a partir de la distribución uniforme

En estas gráficas podemos observar las figuras que se forman a partir de la distribución uniforme obtenida a partir de la generación de números pseudoaleatorios y del método de la transformación inversa. En el eje de las abscisas se encuentran los valores valores que rondan el intervalo (a,b) y en el eje de las ordenadas la frecuencia absoluta en ese intervalo.Los valores utilizados para a y b son 1 y 3 respectivamente.

7.1.2. Distribución de probabilidad Exponencial

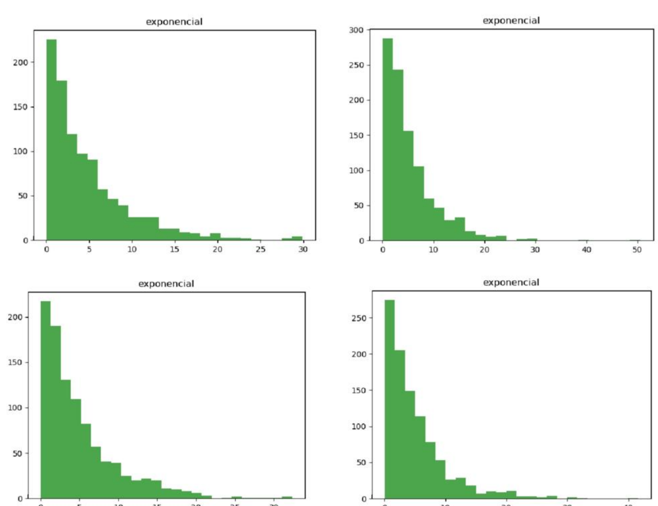

Figura 2: Distintas gráficas obtenidas a partir de la distribución exponencial

En estas gráficas podemos observar las figuras que se forman a partir de la distribución exponencial obtenida a partir de la generación de números pseudoaleatorios y del método de la transformación inversa. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones a partir de un lambda de valor 1/5 ,con una media de 5 y en el eje de las ordenadas la frecuencia absoluta en intervalos de valor 1.

7.1.3. Distribución de probabilidad Gamma

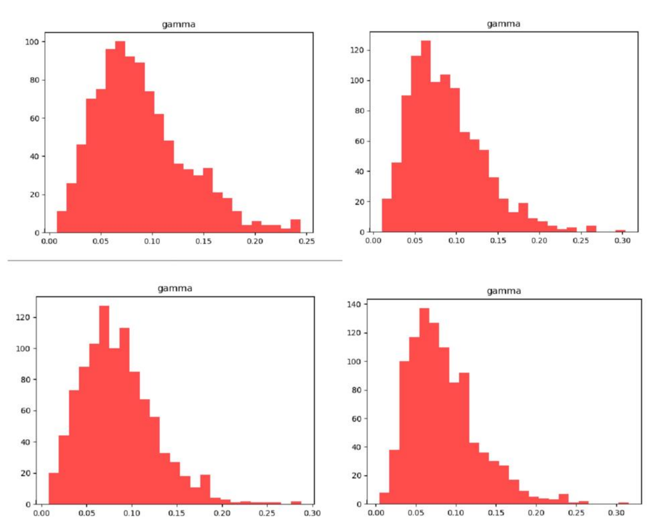

Figura 3: Distintas gráficas obtenidas a partir de la distribución uniforme

En estas gráficas podemos observar las figuras que se forman a partir de la distribución gamma obtenida a partir de la generación de números pseudoaleatorios y del método de la transformación inversa. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones a partir de un lambda de valor 5 y con un total de 20 eventos sucesivos y en el eje de las ordenadas la frecuencia absoluta en intervalos de valor 0.01.

7.1.4. Distribución de probabilidad Normal

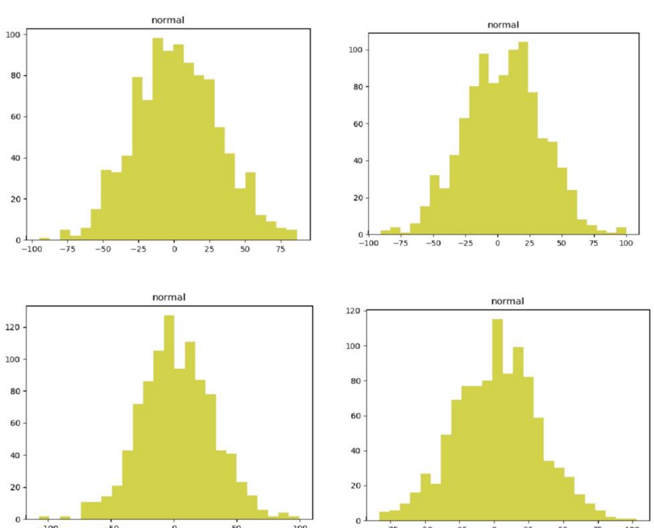

Figura 4: Distintas gráficas obtenidas a partir de la distribución normal

En estas gráficas podemos observar las figuras que se forman a partir de la distribución normal obtenida a partir de la generación de números pseudoaleatorios y del método de la transformación inversa. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones a partir de una media de valor 2.35 y y una desviación de estándar de valor 30 y en el eje de las ordenadas la frecuencia absoluta de los diferentes intervalos del histograma.

7.2. Distribuciones de probabilidad discretas

7.2.1. Distribución de probabilidad Pascal

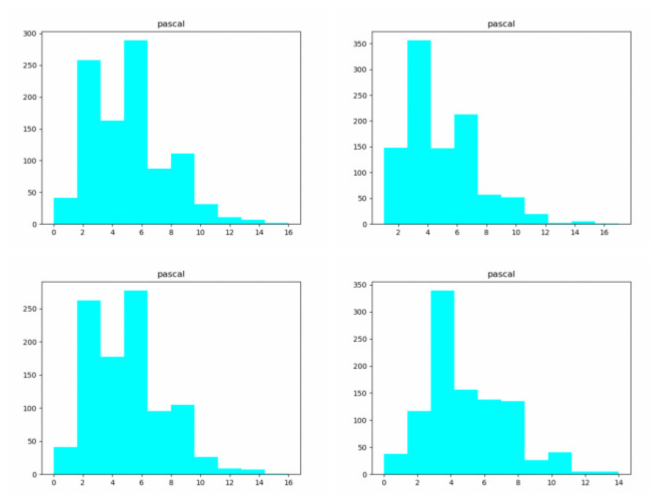

Figura 5: Distintas gráficas obtenidas a partir de la distribución pascal

En estas gráficas podemos observar las figuras que se forman a partir de la distribución Pascal obtenida a partir de la generación de números pseudoaleatorios y del método de la transformación inversa. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones a partir de repeticiones de k éxitos con valor igual a 5 y con una probabilidad de exito de 0.6 y en el eje de las ordenadas la frecuencia absoluta de los diferentes intervalos del histograma.

7.2.2. Distribución de probabilidad Binomial

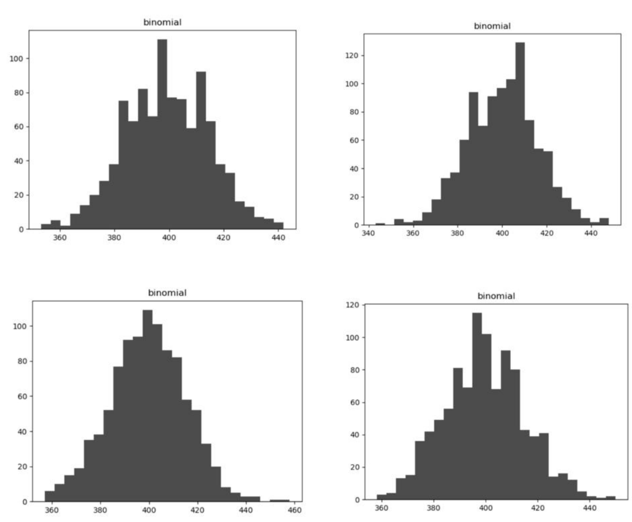

Figura 6: Distintas gráficas obtenidas a partir de la distribución binomial

En estas gráficas podemos observar las figuras que se forman a partir de la distribución Binomial obtenida a partir de la generación de números pseudoaleatorios y del método de la transformación inversa. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones a partir de "n.ensayos independientes de Bernoully para los cuales la probabilidad de éxito es de 0.4 y en el eje de las ordenadas la frecuencia absoluta de los diferentes intervalos del histograma.

7.2.3. Distribución de probabilidad Hipergeometrica

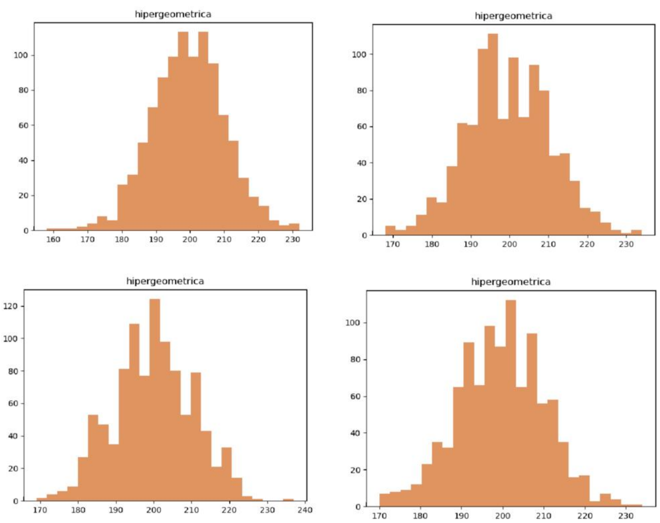

Figura 7: Distintas gráficas obtenidas a partir de la distribución hipergeometrica

En estas gráficas podemos observar las figuras que se forman a partir de la distribución Hipergeometrica obtenida a partir de la generación de números pseudoaleatorios y del método de la transformación inversa. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones a partir de una población N (en nuestro caso seleccionamos un N muy grande de 5000000), una muestra de dicha población de 500 y un índice de proporción de 0.4 y en el eje de las ordenadas la frecuencia absoluta de los diferentes intervalos del histograma.

7.2.4. Distribución de probabilidad Poisson

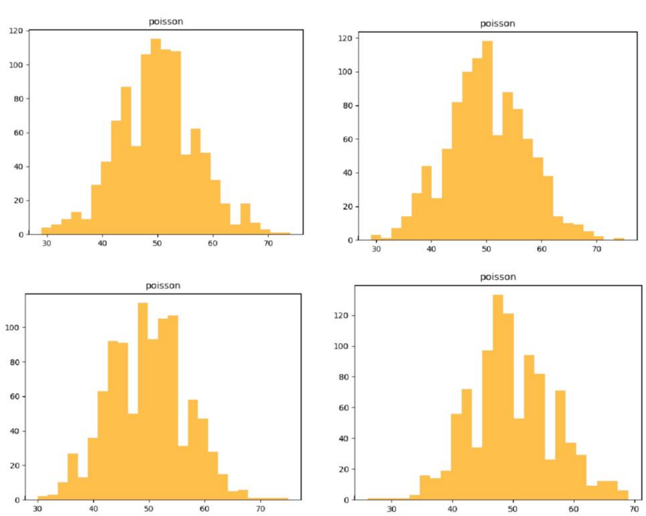

Figura 8: Distintas gráficas obtenidas a partir de la distribución poisson

En estas gráficas podemos observar las figuras que se forman a partir de la distribución Poisson obtenida a partir de la generación de números pseudoaleatorios y del método de la transformación inversa. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones a partir de un lambda de valor 50 y en el eje de las ordenadas la frecuencia absoluta de los diferentes intervalos del histograma.

7.2.5. Distribución de probabilidad Empírica discreta

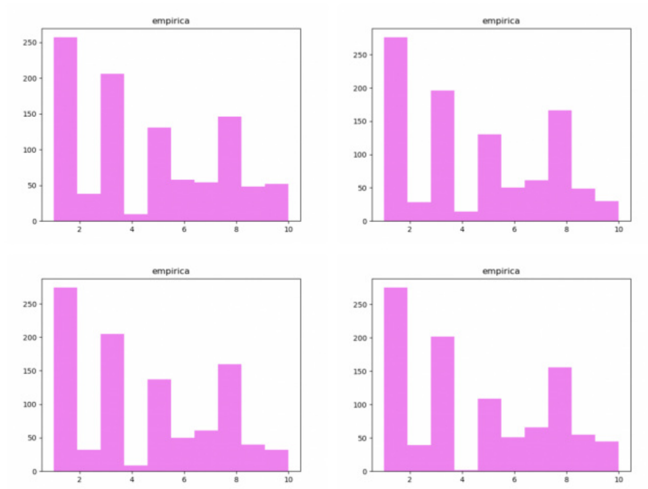

Figura 9: Distintas gráficas obtenidas a partir de la distribución empírica discreta

En estas gráficas podemos observar las figuras que se forman a partir de la distribución Empírica discreta obtenida a partir de la generación de números pseudoaleatorios y del método de la transformación inversa. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones a partir de diez probabilidades (en nuestro caso seleccionamos las siguientes probabilidades:0.273,0.037,0.195,0.009,0.124,0.058,0.062,0.151,0.047,0.044) cuya da un total de 1, y en el eje de las ordenadas la frecuencia absoluta de los diferentes intervalos del histograma.

8. Realización del test de bondad Chi-Cuadrado para las distintas distribuciones

La prueba chi-cuadrado es una de las más conocidas y utilizadas para analizar variables nominales o cualitativas, es decir, para determinar la existencia o no de independencia entre dos variables. Que dos variables sean independientes significa que no tienen relación, y que por lo tanto una no depende de la otra, ni viceversa. Dividiremos en 10 intervalos, agrupando los valores y determinado las frecuencias observadas y esperadas. Aplicaremos este test a las distribuciones Uniforme, exponencial, normal,binomial, poisson y empírica discreta. Como manera de verificación se realizo la sumatoria de todos los valores observados y la sumatoria de todos los valores esperados, ya que cada uno de estos valores debería de poseer un valor de 1000, ya que generaremos 1000 números pseudoaleatorios. Para la visualizacion del valor de chi-cuadrado esperado utilizaremos la tabla con 9 grados de libertad y un 95 por ciento de confianza y la formula para obtener el valor de chi-cuadrado observado sera:

$$x^{(2)} = \sum_{i=1}^{c} \frac{((o_i - e_i)^2}{e_i}$$

Donde:

- e es el valor esperado
- o es el valor observado en cada intervalo

8.1. Test para la distribución Uniforme

Valor observado	Valor esperado	Valor obtenido de chi-cuadrado
95	100	0.25
98	100	0.04
99	100	0.01
106	100	0.36
113	100	1.69
104	100	0.16
101	100	0.01
104	100	0.16
80	100	4.0
100	100	0.0
Suma de todo	—	6.68

Cuadro 1: Tabla con los valores observados,esperados y chi-cuadradado de una distribución Uniforme

Se poseen 10 intervalos para la distribución uniforme, por lo tanto tendremos 9 grados de libertad, es decir el valor esperado de chi cuadrado sera de 16.92.Los intervalos estarán divididos de a cien números, debido a que al poseer mil números generados y al estos estar distribuidos uniformemente, la probabilidad de que un numero pertenezca a alguno de los diez intervalos es de 0.1. Con 9 grados de libertad y con un 95 por ciento de confianza se obtiene un valor de Chi cuadrado de 6.68 en este caso. El valor de chi cuadrado obtenido 6.68 es menor que 16.92, por lo tanto paso el test. Se realizaron múltiples simulaciones y en todas ellas se logro pasar el test. Los datos en la tabla corresponden a una simulación en especifico para mostrar los resultados como ejemplo.

8.2. Test para la distribución Exponencial

Valor observado	Valor esperado	Valor obtenido de chi-cuadrado
60	58.2354	0.05346
42	54.8440	2.1441
62	51.65022	2.0739
46	48.6423	0.1435
25	45.8096	0.01430
27	43.1418	6.0396
46	40.6295	0.7098
46	38.2634	1.5642
40	36.0351	0.43620.4362
584	582.7482	0.0026
Suma de todo	—	13.1820

Cuadro 2: Tabla con los valores observados,esperados y chi-cuadradado de una distribución Exponencial

Se poseen 10 intervalos para la distribución exponencial, por lo tanto tendremos 9 grados de libertad,es decir el valor esperado de chi cuadrado sera de 16.92.Los intervalos estarán divididos de a 0.3 números y el valor esperado para cada intervalo se calculara a través de la multiplicación entre la diferencia entre la función acumulada evaluada en el limite superior e inferior y el valor mil. Esto es debido a que poseemos mil números generados y estos están distribuidos exponencialmente. Con 9 grados de libertad y con un 95 por ciento de confianza se obtiene un valor de Chi cuadrado de 13.182013.1820 en este caso. El valor de chi cuadrado obtenido 13.1820 es menor que 16.92, por lo tanto paso el test. Se realizaron múltiples simulaciones y en todas ellas se logro pasar el test.

8.3. Test para la distribución Normal

Valor observado	Valor esperado	Valor obtenido de chi-cuadrado
6	3.24023	2.3505
8	17.52	0.01308
58	62.8253	0.3706
152	149.48	0.0423
217	236.13	1.5503
258	247.71	0.4269
195	172.5812	2.9122
72	79.83	0.7684
21	24.50 24.50	0.5021
3	4.99	0.7936
Suma de todo	—	9.7304

Cuadro 3: Tabla con los valores observados,esperados y chi-cuadradado de una distribución Normal

Se poseen 10 intervalos para la distribución normal, por lo tanto tendremos 9 grados de libertad,es decir el valor esperado de chi cuadrado sera de 16.92.Los intervalos estarán divididos de a 20 numeros y el valor esperado para cada intervalo se calculara a través de la la diferencia de las probabilidades que se encuentre el valor generado en cada intervalo , buscando los valores en la tabla Z con la utilizacion de una media y desviación calculadas y multiplicado por el valor mil. Esto es debido a que poseemos mil números generados y estos estan distribuidos de manera normal,la probabilidad de que un valor se encuentre en cada intervalo es por lo tanto distinta. Con 9 grados de libertad y con un 95 por ciento de confianza se obtiene un valor de Chi cuadrado de 9.7304 en este caso. El valor de chi cuadrado obtenido 9.7304 es menor que 16.92, por lo tanto paso el test. Se realizaron múltiples simulaciones y en todas ellas se logro pasar el test. En nuestro trabajo utilizamos las siguientes formulas para poder encontrar la media y desviación estándar, necesarias para poder encontrar las probabilidades de los distintos intervalos:

Formula de la media:

$$\mu = \lambda_1 = \sum_{i=1}^{n} \frac{x_i O_i}{n}$$

Donde:

- n es la cantidad de números generados, en nuestro caso 1000.
- μ es la media.
- Oi es la cantidad de valores observados pertenecientes a un intervalo.
- x es la media del intervalo.
- λ_1 es un parámetro necesario para calcular la desviación estándar

Formula de desviación estandar:

$$\lambda_2 = \sum_{i=1}^{n} \frac{x_i^2 O_i}{n}$$

$$\sigma = \lambda_2 - \lambda_1^2$$

Donde:

- λ_2 es el segundo parámetro necesario para calcular la desviación estándar
- Oi es la cantidad de valores observados pertenecientes a un intervalo.
- x es la media del intervalo.
- σ es la desviación estándar

8.4. Test para la distribución Binomial

Valor observado	Valor esperado	Valor obtenido de chi-cuadrado
0	1.2183	1.2183
20	16.3477	0.8159
86	98.4087	1.5646
280	270.4744	0.3354
346	344.0787	0.0107
202	204.4580	0.02955
56	57.02444	0.01840
10	7.4751	0.8527
0	0.4601	0.4601
0	0.01327	0.01327
Suma de todo	—	5.3193

Cuadro 4: Tabla con los valores observados,esperados y chi-cuadradado de una distribución Binomial

Se poseen 10 intervalos para la distribución exponencial, por lo tanto tendremos 9 grados de libertad,es decir el valor esperado de chi cuadrado sera de 16.92.Los intervalos estarán divididos de a 14 números y el valor esperado para cada intervalo se calculara a través de la multiplicación entre la diferencia entre la probabilidad puntual evaluada hasta el limite superior e inferior y el valor mil. Esto es debido a que poseemos mil números generados y estos están distribuidos de manera binomial. Con 9 grados de libertad y con un 95 por ciento de confianza se obtiene un valor de Chi cuadrado de 5.3193 en este caso. El valor de chi cuadrado obtenido 5.3193 es menor que 16.92, por lo tanto paso el test. Se realizaron múltiples simulaciones y en todas ellas se logro pasar el test.

8.5. Test para la distribución Poisson

Valor observado	Valor esperado	Valor obtenido de chi-cuadrado
0	0.0711	0.07112
4	2.6146	0.7339
28	31.2686	0.3416
150	145.8417	0.1185
307	301.3950	0.1042
299	303.2787	0.06036
149	159.8488	0.7363
58	46.8017	2.6793
3	7.9925	3.1186
2	0.8297	1.6503
Suma de todo	—	9.6145

Cuadro 5: Tabla con los valores observados,esperados y chi-cuadradado de una distribución Poisson

Se poseen 10 intervalos para la distribución Poisson, por lo tanto tendremos 9 grados de libertad,es decir el valor esperado de chi cuadrado sera de 16.92.Los intervalos estarán divididos de a 26 numeros y el valor esperado para cada intervalo se calculara a través de la multiplicación entre la diferencia entre la probabilidad puntual evaluada hasta el limite superior e inferior y el valor mil. Esto es debido a que poseemos mil números generados y estos poseen una distribución Poisson. Con 9 grados de libertad y con un 95 por ciento de confianza se obtiene un valor de Chi cuadrado de 9.6145 en este caso. El valor de chi cuadrado obtenido 9.6145 es menor que 16.92, por lo tanto paso el test. Se realizaron múltiples simulaciones y en todas ellas se logro pasar el test.

8.6. Test para la distribución Empírica discreta

Valor observado	Valor esperado	Valor obtenido de chi-cuadrado
268	273	0.091575
32	37	0.6756
194	195	0.00512
8	9	0.1111
135	124	0.9758
58	58	0
52	62	1.6129
158	151	0.3245
42	47	0.53190.5319
53	44	1.8409
Suma de todo	—	6.1695

Cuadro 6: Tabla con los valores observados,esperados y chi-cuadradado de una distribución Empirica Discreta

Se poseen diez números para la distribución Empirica discreta, por lo tanto tendremos 9 grados de libertad,es decir el valor esperado de chi cuadrado sera de 16.92.En este caso especial donde solo poseemos diez valores del 1 al 10,no precisamos la realización de intervalos ,ya que cada valor tiene asignado una determinada probabilidad de ocurrir. La frecuencia esperada para cada valor se calcula a partir de la probabilidad de que ocurra dicho valor multiplicada por mil. Esto es debido a que poseemos mil números generados y estos poseen una distribución Empírica discreta. La suma de todas las probabilidades deber dar el valor de 1. Con 9 grados de libertad y con un 95 por ciento de confianza se obtiene un valor de Chi cuadrado de 6.1695 en este caso. El valor de chi cuadrado obtenido 6.1695 es menor que 16.92, por lo tanto paso el test. Se realizaron múltiples simulaciones y en todas ellas se logro pasar el test.

9. Conclusiones

Luego de realizar múltiples simulaciones y al visualizar todo lo realizado, llegamos a la conclusión de que los números generados por medio del método de la transformación inversa, se distribuyen de manera pseudoaleatoria correspondiendo a cada una de las distribuciones planteadas.

En cuanto a las distribuciones aleatorias continuas y discretas realizadas, pudimos observar como se comportan a partir de los histogramas realizados, pudiendo mostrar los intervalos distintos de cada una y la forma gráfica que se representan, además mostrando las partes mas importantes de cada una.

En cuanto al test chi cuadrado realizado a la distribución Uniforme, Exponencial, Normal, Binomial, Poisson y Empírica discreta, si bien los intervalos a considerar y la forma de calcularlos fue variando, todas las distribuciones nombradas lograron superar los test con resultados mas que satisfactorios.

10. Referencias

https://es.wikipedia.org/wiki/DistribuciC3B3n_binomial#:~:text=En20estadC3Atica2C20la20distribuciC3B3n2C
del20C3A9xito20entre20los20ensayos

https://blog.adrianistan.eu/estadistica-python-distribucion-binomial-normal-poisson-parte-vi

http://www.estadistica.net/test-chi-cuadrado.pdf

Capitulo 4- Técnicas de simulación en computadoras- Naylor

https://mathworld.wolfram.com/NormalDistribution.html

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_uniforme_continua

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_exponencial

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_gamma

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_hipergeom%C3%A9trica

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_de_Poisson

https://es.wikipedia.org/wiki/Distribuci%C3%B3n_binomial_negativa