Generadores de números pseudoaleatorios por medio de distribuciones de probabilidad

Ramiro Di Giacinti
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional - FRRO
Zeballos 1341, S2000, Argentina
dia.digiacinti.ramiro@gmail.com

Bruno Mollo
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional - FRRO
Zeballos 1341, S2000, Argentina
dia.mollo.bruno@gmail.com

Facundo Braida
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional - FRRO
Zeballos 1341, S2000, Argentina
facundobraida98@gmail.com

Lucía Cappellini
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional - FRRO
Zeballos 1341, S2000, Argentina
luciacappll@gmail.com

Adriel Gorosito
Ingeniería en Sistemas de Información
Universidad Tecnológica Nacional - FRRO
Zeballos 1341, S2000, Argentina
adrielgorosito14@gmail.com

March 28, 2024

Abstract

En base al trabajo realizado en el anterior informe, pudimos calcular números aleatorios entre 0 y 1 distribuidos uniformemente por medio de los generadores de números aleatorios. Sin embargo, ahora buscamos de generar números de acuerdo a un tipo de distribución de probabilidad que nosotros deseemos. Para ello nos basaremos en dos métodos de obtención de números aleatorios: el método de la inversa y el método de aceptación y rechazo.

1 Introducción

1.1 Distribuciones de Probabilidad

1.1.1 Uniforme

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 respectivamente.

Función de densidad de probabilidad:

$$f(x; a, b) = \frac{1}{b-a} \quad a \leq x \leq b$$

1.1.2 Exponencial

La distribución exponencial es utilizada comúnmente en estadística y probabilidad para modelar el tiempo entre eventos en un proceso de Poisson. Esta distribución es especialmente útil en situaciones en las que se necesita modelar el tiempo

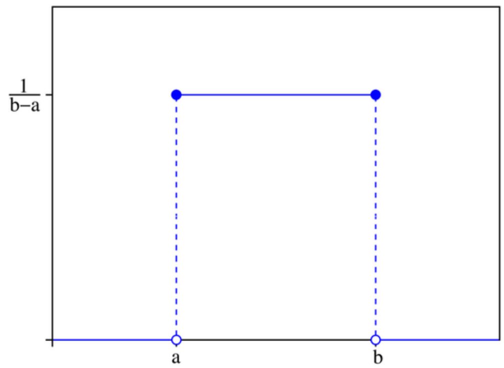

Figure 1: Función de frecuencia de distribución uniforme

entre dos eventos, como el tiempo que transcurre entre llamadas entrantes en un centro de llamadas, el tiempo que tarda un cliente en llegar a una tienda después de ver un anuncio, o el tiempo que tarda un dispositivo en fallar. Su parámetro es lambda,un número positivo que representa la tasa promedio de ocurrencia de eventos por unidad de tiempo

Función de densidad de probabilidad:

$$f(x;\lambda) = \lambda \cdot e^{-\lambda x} \quad x \geq 0, \lambda > 0$$

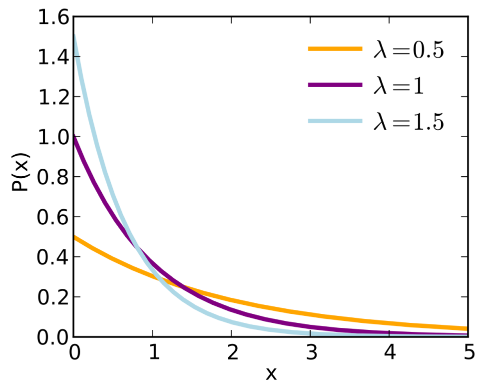

Figure 2: PDF de exponencial

1.1.3 Gamma

La distribución gamma es una distribución con dos parámetros que pertenece a las distribuciones de probabilidad continuas. La distribución exponencial, distribución de Erlang y la distribución chi-cuadrado son casos particulares de la distribución gamma. Hay dos diferentes parametrizaciones que suelen usarse:

- Con parámetro de forma k y parámetro de escala θ .
- Con parámetro de forma $\alpha = k$ y parámetro inverso de escala $\beta = 1/\theta$.

Nosotros usaremos la segunda parametrización en este estudio.

Función de densidad de probabilidad:

$$f(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0, \quad \alpha, \beta > 0$$

1.1.4 Normal

La gráfica de su función de densidad de esta distribución 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

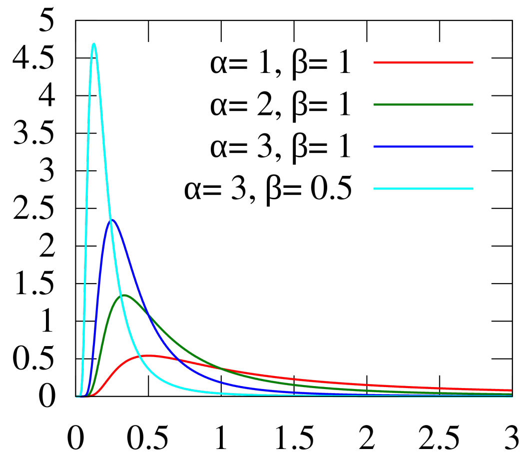

Figure 3: PDF de la distribución Gamma

gaussiana.La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Se puede definir a partir de dos parámetros, al media μ y el desvió estándar σ.

Función de densidad de probabilidad:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad \mu \in \mathbb{R}, \sigma^2 > 0$$

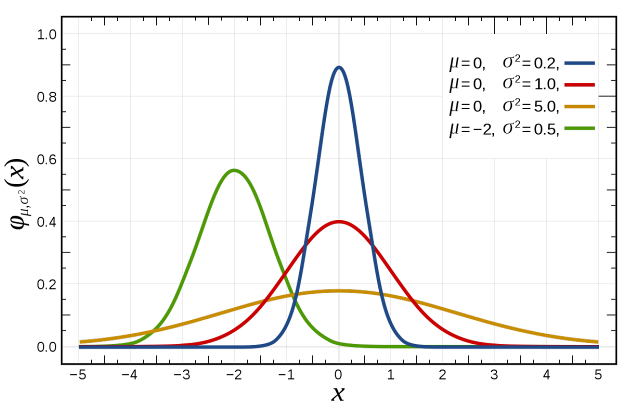

Figure 4: PDF de la distribución

1.1.5 Binomial

Una distribución binomial busca evaluar el numero de éxitos que se tuvieron respecto de un evento con probabilidad p de ocurrencia, en n ensayos de Bernoulli (Se obtiene un éxito o no) que son independientes entre si, es decir, existe reposición.

Función de probabilidad de x:

$$P(X = x) = n\mathbb{C}x \; p^x(1-p)^{n-x}, \; para \; k = 0, \; 1, \; ..., \; n, \; 0 \leq p \leq 1$$

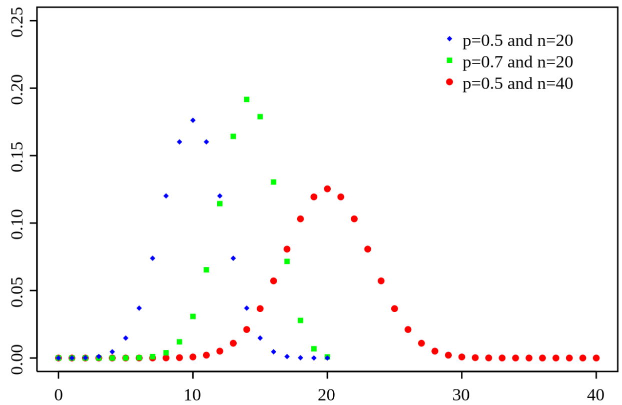

Figure 5: PDF de la distribución

1.1.6 Pascal

Es muy similar a una binomial. Esta distribución representa el numero de intentos necesarios hasta que ocurre el éxito, por lo que solamente toma la probabilidad de la ocurrencia p para realizarlo.

Función de probabilidad de x:

$$P(X = x) = p(1 - p)^{x-1}$$

1.1.7 Hiper-geométrica

Tanto la distribución hiper-geométrica como la distribución binomial describen el número de veces que un evento ocurre en un número fijo de ensayos. Para la distribución binomial la probabilidad es igual para cada ensayo, sin embargo para la distribución hiper-geométrica, cada intento cambia la probabilidad del subsiguiente puesto que no hay reposición. Utiliza 2 parámetros, la cantidad inicial de eventos que pueden ocurrir K, el tamaño de la muestra N

Función de probabilidad de x:

$$P(X = x) = \frac{K\mathbb{C}x \ (N - K)\mathbb{C}(n - x)}{N\mathbb{C}n}$$

1.1.8 Poisson

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.

Función de Probabilidad:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$$

Donde:

- k: es el número de eventos que se observan en el intervalo.
- λ: es el parámetro de la distribución de Poisson que representa la tasa media de ocurrencia de los eventos.

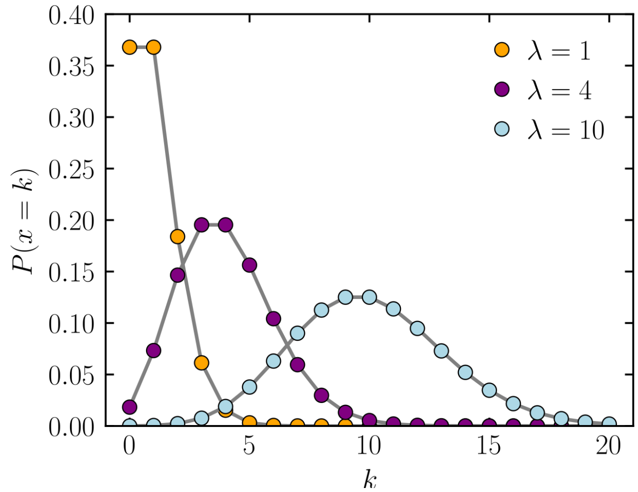

Figure 6: PDF de la distribución

1.2 Distribución empírica discreta

Una distribución empírica discreta es una forma de representar la distribución de probabilidad de una variable discreta utilizando datos empíricos observados. En lugar de especificar la función de masa de probabilidad de forma analítica, la distribución empírica discreta se basa en la frecuencia relativa de los valores observados.

Para construir una distribución empírica discreta, primero recopilas datos discretos y cuentas la frecuencia con la que ocurre cada valor. Luego, calculas la frecuencia relativa dividiendo cada frecuencia por el número total de observaciones. La frecuencia relativa se puede interpretar como una estimación de la probabilidad de que ocurra cada valor.

Para nuestro estudio, utilizaremos la siguiente distribución de frecuencias.

Valor	**Probabilidad**
1	0.3
2	0.4
4	0.3

1.3 Métodos de generación de números

1.3.1 Transformada Inversa

Una de las maneras para poder generar números aleatorios en base a una distribución, requiere del conocimiento de su función de probabilidades acumulada F(x) = $P(X \leq x) =$ r, donde r es una probabilidad $0 \leq r \leq 1$.

Conociendo esto podemos replantear la ecuación de la siguiente manera: $F^{-1}(r) = x$ Es importante tener en cuenta que no todas las distribuciones de probabilidad tienen una transformada inversa analítica debido a su naturaleza, por lo que se deben analizar por otros medios.

1. A partir de la función acumulada F(x), si existe encontrar su función inversa $F^{(} - 1)(y)$ analítica
2. Generar un lote de números aleatorios con distribución u U(0,1)
3. Evaluar cada uno de los valores del lote en la función inversa $F^{(} - 1)(u) = z$
4. Finalmente el conjunto de todos los valores z siguen la distribución buscada.

1.3.2 Método de aceptación y rechazo

Se trata de un método universal alternativo al de inversión para el caso de que no se pueda emplear la función cuantil, pero se dispone de una expresión (preferiblemente sencilla) para una función de densidad objetivo f(x). Este método se puede implementar cuando la función densidad objetivo f(x) se trate de una función acotada, y además x pueda definirse en un intervalo finito [a, b]. La técnica del rechazo consiste en las siguientes etapas:

3. Normalizar la función densidad f(x) multiplicándola por un factor "c" de manera que cumpla

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

2. Definir a x como una función lineal de r, siendo r_1 un numero aleatorio uniforme dentro del rango (0,1)

$$z = a + (b - a) * r_1$$

   Para simplificar este paso, no utilizamos esta función lineal para obtener z sino que aprovechamos la librería de python numpy para generar un valor aleatorio que este entre el valor mínimo y máximo de la distribución, los cuales se obtuvieron a partir de la regla empírica.

3. Para determinar si dicho valor aleatorio generado z es aceptado o rechazado, se evalúa según la siguiente expresión, donde r_2 es otro valor aleatorio generado con la librería numpy de python, el cual pertenece al intervalo (0, 1)

$$r_2 \leq c.f(z)$$

1.4 Test Kolmogorov

El método de Kolmogorov es una prueba de bondad que se emplea para obtener un indicador que le dé una idea al investigador de si dos distribuciones son distintas o si una distribución de probabilidad subyacente difiere de una distribución hipotética. Principalmente, se usa cuando en el análisis de datos se quiere comparar la distribución empírica de una variable con una distribución teórica. El test se basa en la comparación de las funciones de distribución acumulativa (CDF) de los datos observados y la CDF de la distribución teórica. En principio, se plantea una hipótesis nula que asume que los datos generados se distribuyen de acuerdo a la distribución teórica. El test calcula una estadística de prueba que representa el valor máximo absoluto de las diferencias entre ambas funciones de distribución acumuladas. El procedimiento del test consiste en ordenar los datos generados de manera ascendente y calcular las diferencias entre ambas CFD en cada punto. La estadística de prueba queda determinada por la máxima diferencia absoluta y se compara con un valor crítico determinado por el nivel de confianza deseado. Cuando el valor de la estadística de prueba sea menor o igual al valor crítico, no habrá suficiente evidencia para rechazar la hipótesis nula y se concluye que los datos se distribuyen de acuerdo a la función teórica. Por el contrario, si la estadística de prueba supera el valor crítico, se rechaza la hipótesis nula y se concluye que los datos generados tienen una distribución diferente a la buscada.

2 Metodología

Para realizar las distintas operaciones, utilizaremos Python en conjunto con un ciertas librerías (Math, NumPy, SciPy) para poder trabajar con las distintas distribuciones de probabilidades y Matplotlib para realizar los distintos gráficos.

2.1 Método del rechazo

Para realizar el método de aceptación y rechazo utilizaremos el siguiente código, en el cual se define una función genérica que tomará una distribución de probabilidades que se le pase como parámetro y la evaluara en base al resto de parámetros de entrada.

```
def metodo_rechazo(pdf_estudio, techo, min, max, size):
  accepted = []
  for _ in range(size):
      # Generamos un numero aleatorio x-U(min,max)
      x = np.random.uniform(min, max)
      # Calculamos la probabilidad de aceptar x
      prob_pass = pdf_estudio(x) / techo
      # Aleatoriamente, decidimos si es aceptado
      if (np.random.uniform(0, 1) <= prob_pass):
          # Si lo es, lo agregamos a nuestra lista de aceptados
          accepted.append(x)
  return accepted
```

2.2 Método de la transformada inversa

El método de la transformada inversa permite obtener un conjunto de valores con una distribución deseada a partir de un conjunto que presenta una distribución uniforme.

Éste método consiste en determinar la inversa de la función densidad de la distribución deseada y aplicarla a cada uno de los valores del conjunto con distribución uniforme.

2.2.1 Distribución Uniforme

```
def uniforme(a, b, size):
    datos = np.random.uniform(a, b, size)

    # Valores de la funcion de densidad en un rango de valores de x
    x = np.linspace(a, b, 1000)
    y = np.full_like(x, 1 / (b - a))

    plt.plot(x, y, color='red', label='Distribucion Uniforme')
    plt.hist(datos, bins=10, density=True)
    plt.legend(["Funcion de densidad de probabilidad"], loc='upper right')
    plt.title('Distribucion Uniforme')
    plt.xlabel('Valor')
    plt.ylabel('Probabilidad')
    plt.show()
```

2.2.2 Distribución Normal

La distribución normal es una de las distribuciones más importantes en la estadística y se utiliza en muchos campos, como la física, la ingeniería y la economía. Sin embargo, en ocasiones es difícil obtener datos que sigan una distribución normal, por lo que se recurre a métodos de transformación de distribuciones. Existen diversas técnicas para transformar los datos y ajustarlos a una distribución normal, una de ellas es la distribución lambda generalizada (DLG).

En este enfoque, primero se parte de un conjunto de datos con distribución uniforme y se establece una función objetivo para ajustar los datos a una distribución normal. Luego se utilizan técnicas de estimación para ajustar los parámetros de la a los datos uniformes. Finalmente, se generan datos transformados utilizando la DLG ajustada y se transforman en una distribución normal.

$$\lambda_1 = 0,\ \lambda_2 = 0.1975,\ \lambda_3 = \lambda_4 = 0.1349$$

La transformación de una distribución uniforme en una distribución normal no siempre será perfecta, pero la DLG es una herramienta útil para obtener una aproximación a una distribución normal.

```
def newton_raphson(p, mu=0, sigma=1, tol=1e-6, max_iter=100):
    x = mu # initial guess
    for i in range(max_iter):
        pdf = norm.spicy.pdf(x, loc=mu, scale=sigma)
        cdf = norm.spicy.cdf(x, loc=mu, scale=sigma)
        x_new = x - (cdf - p) / pdf
        if abs(x_new - x) < tol:
            break
        x = x_new
    return x
```

```
def normal(media, sigma, size):
    fig, ax = plt.subplots()

    # Distribucion generada a partir de la uniforme
    datos = np.random.uniform(0, 1, size)
    transformZ = np.vectorize(lambda u: (pow(u, 0.1349)-pow(1-u,0.1349))
        /0.1975)
    datosZ = transformZ(datos);
    transformX = np.vectorize(lambda u: u*sigma+media)
    datosX=transformX(datosZ)
    plt.hist(datosX,
            bins=30,
            density=True,
            label="Normal por transformada inversa")

    #Normal teorica
    x=np.linspace(media-3*sigma, media+3*sigma, size)
    y = (1/(sigma*np.sqrt(2*np.pi)))*(np.exp(-pow(x-media,2)/(2*pow(sigma
        ,2))))
    ax.plot(x, y, 'r-', linewidth=2)
    plt.plot(x, y, color='red', label='Normal teorica')
    plt.title('Distribucion Normal')
    plt.xlabel('Valor')
    plt.ylabel('Probabilidad')
    plt.legend(["Funcion de densidad de probabilidad"], loc='upper right')
    plt.show()
```

3 Resultados

En este apartado se exhiben gráficos en los que se ilustra la función teoría de cada distribución junto un histograma o gráfico de barra que toma los valores del conjunto de datos generado, dependiendo la distribución sea discreta o continua. El objetivo es analizar la similitud entre la función teórica y el conjunto de datos. Cuanto mas similares sean ambas gráficas, estaremos ante un conjunto de datos con una distribución muy aproximada a la distribución deseada. Además implementamos el Test Kolmogorov para confirmar que los datos generados verifican la distribución buscada.

3.1 Método de la transformada inversa

3.1.1 Distribución Uniforme

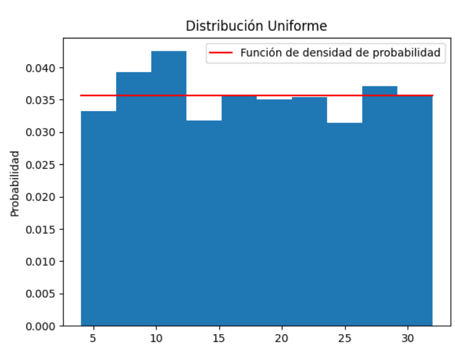

Figure 7: Gráfica distribución Uniforme con transformada inversa

3.1.2 Distribución Normal

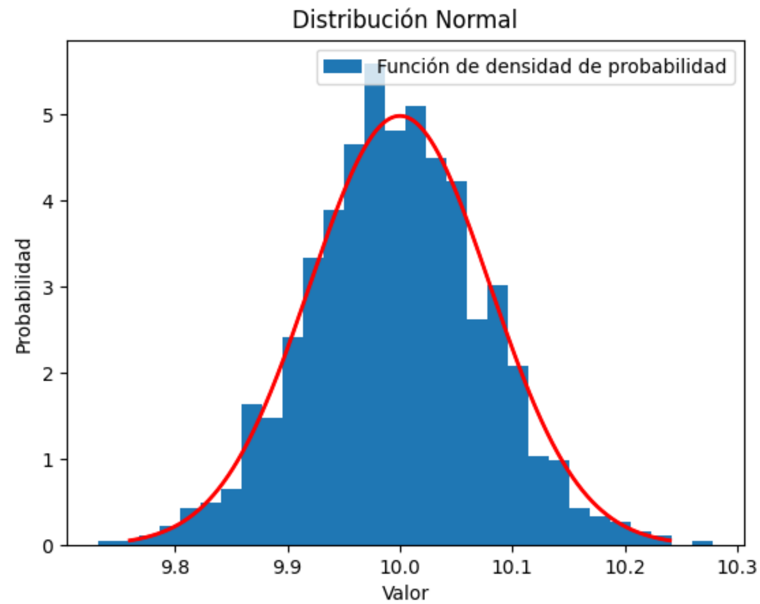

Figure 8: Gráfica distribución Normal con transformada inversa

3.1.3 Distribución Exponencial

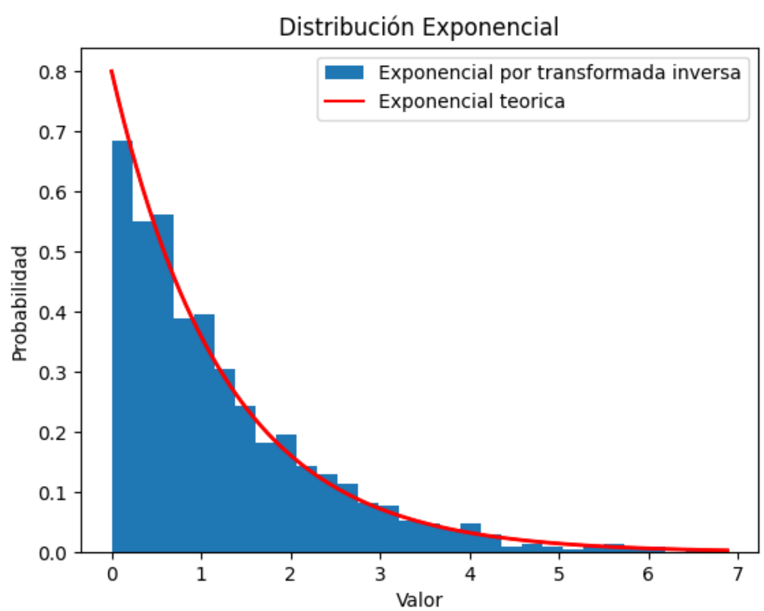

Figure 9: Gráfica distribución Exponencial con transformada inversa

3.2 Método del rechazo

3.2.1 Distribución Uniforme

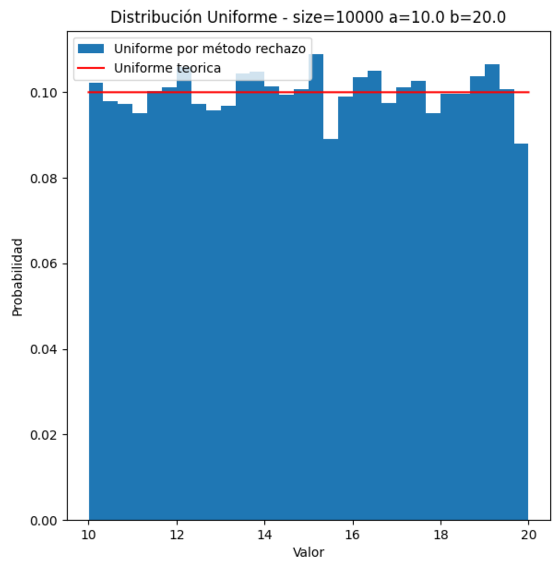

Figure 10: Gráfica distribución Uniforme con método del rechazo

3.2.2 Distribución Exponencial

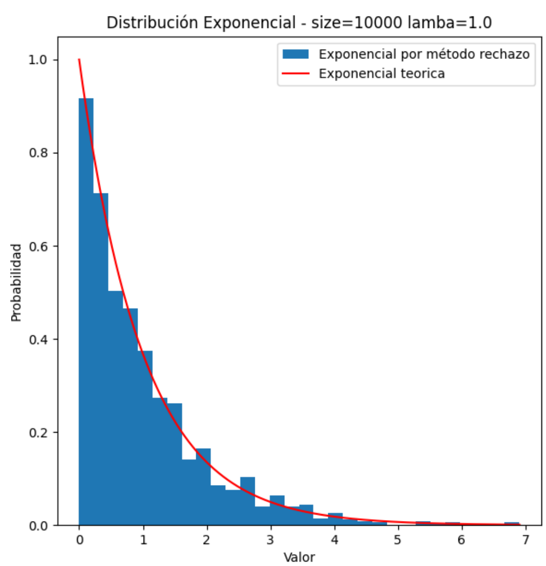

Figure 11: Gráfica distribución Exponencial con método del rechazo

3.2.3 Distribución Gamma

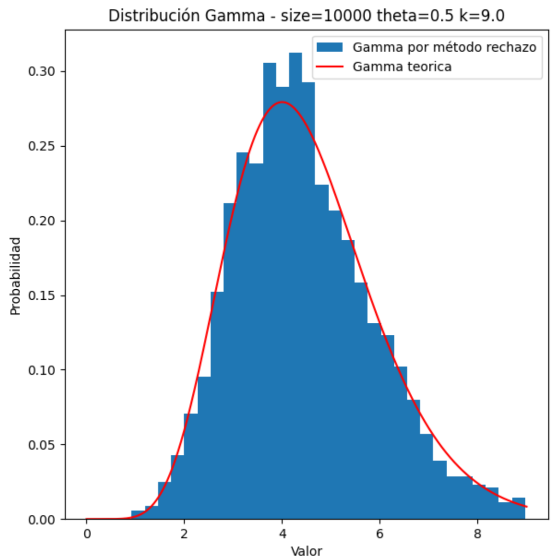

Figure 12: Gráfica distribución Gamma con método del rechazo

3.2.4 Distribución Normal

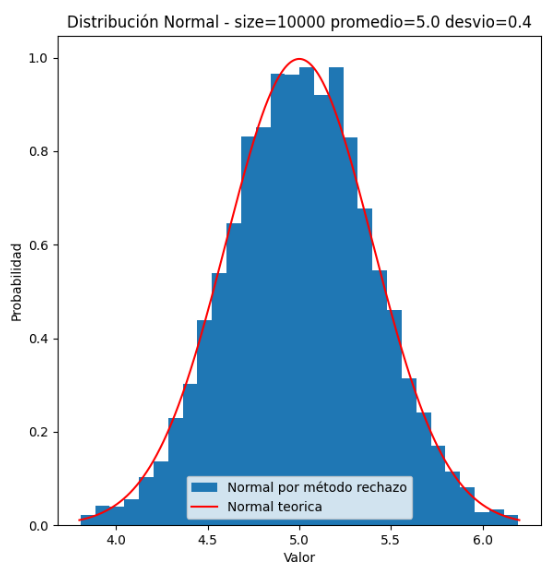

Figure 13: Gráfica distribución Normal con método del rechazo

3.2.5 Distribución Pascal

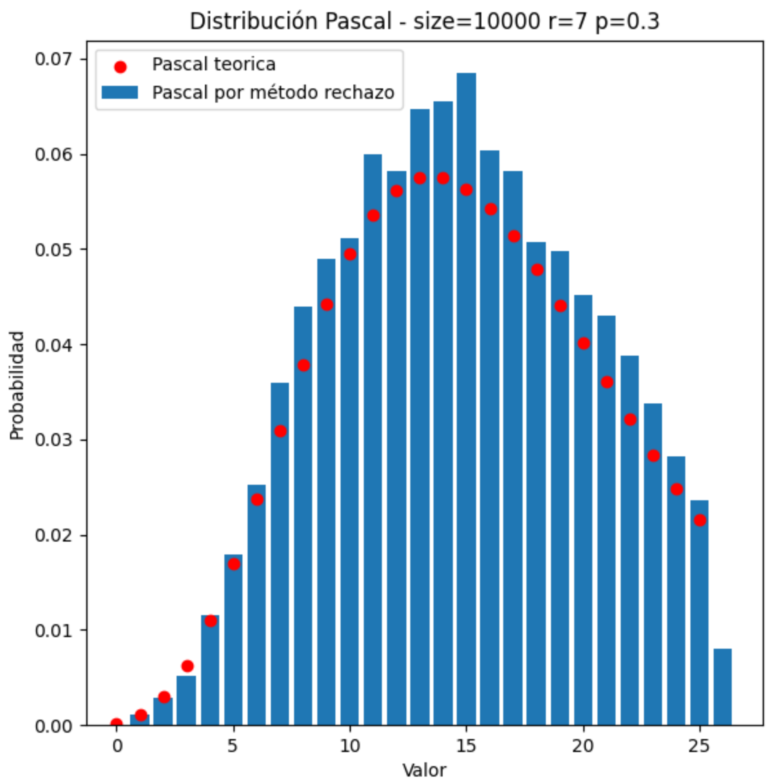

Figure 14: Gráfica distribución Pascal con método del rechazo

3.2.6 Distribución Binomial

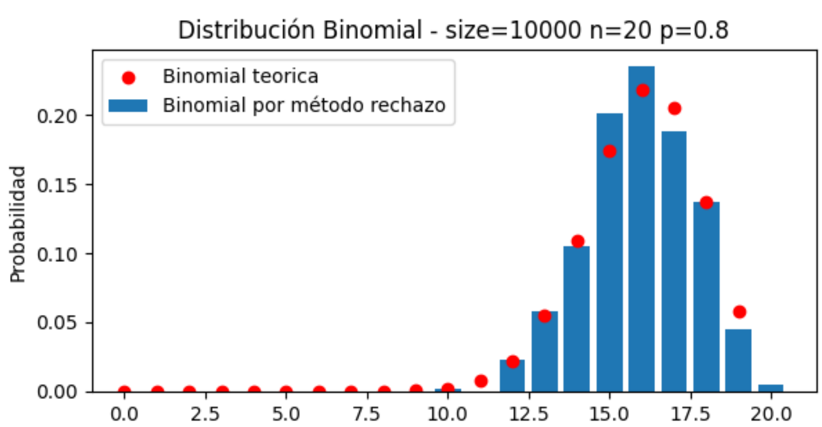

Figure 15: Gráfica distribución Binomial con método del rechazo

3.2.7 Distribución Hipergeometrica

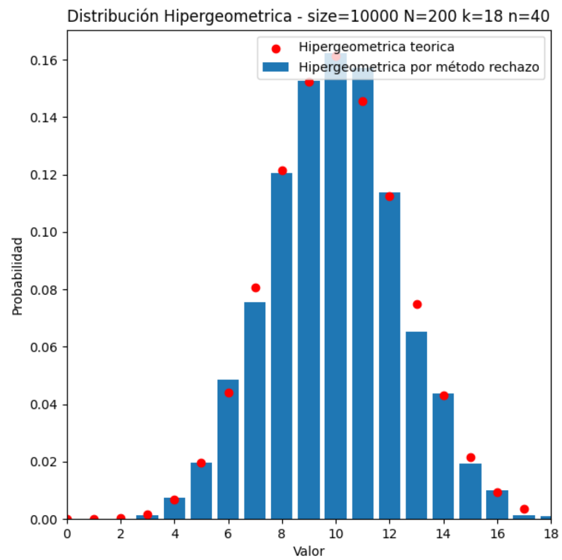

Figure 16: Gráfica distribución Hipergeometrica con método del rechazo

3.2.8 Distribución Poisson

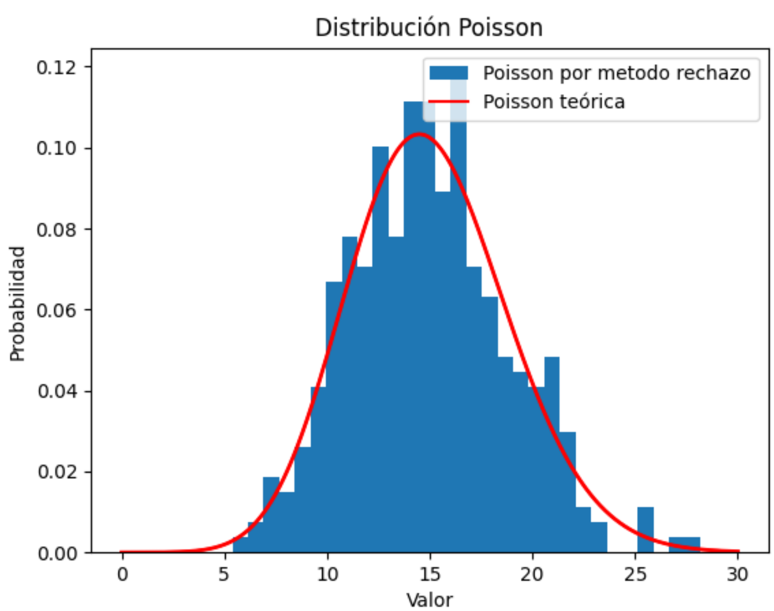

Figure 17: Gráfica distribución Poisson con método del rechazo

3.2.9 Distribución Empírica discreta

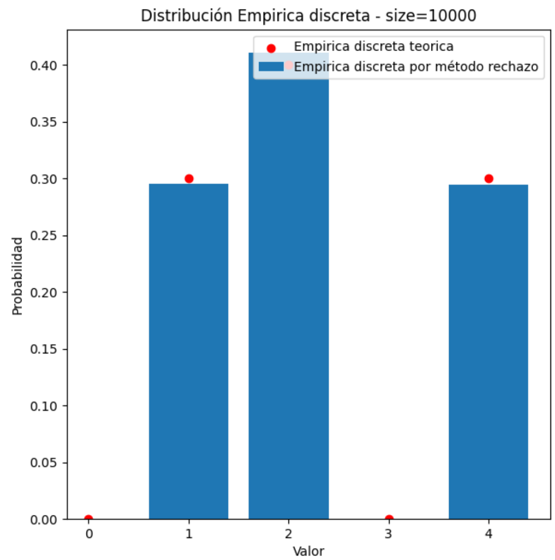

Figure 18: Gráfica distribución Empírica discreta con método del rechazo

3.3 Resultados obtenidos por medio de la prueba de Kolmogórov-Smirnov

Distribución	**Transformada inversa**	**Método del rechazo**
Uniforme	Pasó	Pasó
Normal	Pasó	Pasó
Binomial	-	Pasó
Poisson	-	Pasó
Exponencial	Pasó	Pasó
Gamma	-	Pasó
Hipergeometrica	-	Pasó

4 Conclusión

En este informe implementamos dos métodos para generar datos que presenten una distribución específica deseada a partir de un conjunto de datos que se distribuye uniformemente. Éstos métodos son el método de la transformada inversa y el método de aceptación y rechazo.

En primer lugar, comprobamos que el método de la transformada inversa puede no ser siempre aplicable; algunas distribuciones de probabilidad no cuentan con una función inversa analítica, lo que hace muy difícil la implementación del método. Por su parte, el método de aceptación y rechazo normaliza cada función de estudio para aceptar y rechazar valores aleatorios de una manera más simple, permitiendo así la implementación del método en todas las distribuciones sin importar su ley. Vale aclarar que el método de la transformada inversa es una mejor aproximación que el método del rechazo.

Posteriormente, a partir de gráficas que ilustran la distribución del conjunto de valores generados junto con las funciones de probabilidad de estudio y de la implementación del test Kolmogorov, en todos los casos de estudio apreciamos que el conjunto de datos respeta la distribución esperada, por lo que podemos concluir que los algoritmos desarrollados funcionan correctamente y son confiables.

Para cerrar se puede concluir que se pueden generar una distribución deseada cualquiera a partir de un conjunto de valores que siguen una distribución uniforme, ya sea con el método de trasformada inversa o de aceptación y rechazo, por lo que si en algún campo especifico de trabajo se necesita un tipo de distribución para realizar simulaciones podemos utilizar estos algoritmos para obtener la distribución necesaria en tal campo.

5 Referencias

- Código Simulación: https://github.com/Ramiro-DG/Simulacion2023/tree/main/DDP
- Naylor, T.H. Técnicas de simulación en computadoras, 1982.
- https://en.wikipedia.org/wiki/Binomial_distribution
- https://en.wikipedia.org/wiki/Exponential_distribution
- https://en.wikipedia.org/wiki/Gamma_distribution
- https://en.wikipedia.org/wiki/Binomial_distribution