TP 2.2 - Generadores de números pseudoaleatorios de distintas Distribuciones de Probabilidad

Lautaro Teta Musa
Universidad Tecnológica Nacional
lautarotetamusa@gmail.com

Ignacio Roca
Universidad Tecnológica Nacional
ignacioroca@gmail.com

Agustín Luzzini
Universidad Tecnológica Nacional
agusluzzini@gmail.com

Nazareno Necchi
Universidad Tecnológica Nacionala
nazanecchi.cer@gmail.com

13 de junio, 2024

Abstract

Este informe presenta el desarrollo y análisis de generadores de números aleatorios aplicados a distintas distribuciones de probabilidad, utilizando un Generador Lineal Congruencial (GLC) previamente programado. El trabajo abarca la implementación de nueve distribuciones de probabilidad (uniforme, exponencial, gamma, normal, Pascal, binomial, hipergeométrica, Poisson y empírica discreta), cada una introducida de manera teórica, incluyendo sus parámetros de entrada, función de probabilidad y gráficas.

***K*eywords** Simulación · Trabajo práctico · Generadores Pseudoaleatorios · Distribuciones de probabilidad

1 Introducción

El estudio anterior a este, demostró que la generación de números pseudoaleatorios no es una tarea sencilla. Por suerte para nosotros es un tema felizmente resuelto para nuestras necesidades. Ahora que tenemos un generador testeado, es hora de empezar a utilizarlo. El problema surge como hacerlo. Si lo empleamos tal y como está, este solo genera valores uniformes continuos entre 0 y 1. Como nosotros hemos aprendido en la materia Probabilidad y Estadística, hay otras distribuciones de probabilidad tanto continuas como discretas. La pregunta que debemos responder es: ¿cómo hacemos para generar valores de distintas distribuciones?. La buena noticia es que la solución ya ha sido inventada y es nuestra tarea como ingenieros redescubrir el mecanismo para implementarlo en nuestros experimentos. Por lo tanto el objetivo de este trabajo es construir generadores de números pseudoaleatorios de distintas probabilidades por medio de un material antiguo pero sumamente práctico que es el elaborado por el autor Thomas Naylor.

2 Distribuciones

En esta sección, se presentan los fundamentos teóricos de los métodos de generación de números pseudoaleatorios y las pruebas estadísticas utilizadas para evaluar su calidad. Se describen dos métodos específicos de generación: el método de los Cuadrados Medios y el Generador Lineal Congruencial (GLC). Además, se detallan cuatro pruebas estadísticas que se aplicarán para evaluar el rendimiento de estos generadores.

2.1 Uniforme

La distribución uniforme es una distribución de probabilidad en la que todos los intervalos de igual longitud dentro del rango de la distribución son igualmente probables. Esta distribución se define en un intervalo [a,b] y su función de densidad de probabilidad (FDP) es constante entre estos dos límites.

2.1.1 Parametros de entrada

- **a:** Límite inferior.
- **b:** Límite superior.

2.1.2 Función de Densidad de Probabilidad (PDF)

La función de densidad de probabilidad (PDF) para la distribución uniforme está dada por:

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

2.1.3 Función Acumulada (CDF)

La función acumulada (CDF) para la distribución uniforme está dada por:

$$F(x) = \frac{x-a}{b-a} \quad \text{para} \quad a \leq x \leq b$$

Paso a paso:

1. Definición de la PDF:

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

2. Definición de la CDF:

$$F(x) = \int_a^x f(t)\, dt$$

3. Sustitución de la PDF:

$$F(x) = \int_a^x \frac{1}{b-a}\, dt$$

4. Integración:

$$F(x) = \left[\frac{t}{b-a}\right]_a^x = \frac{x-a}{b-a}$$

5. Resultado Final:

$$F(x) = \frac{x-a}{b-a} \quad \text{para} \quad a \leq x \leq b$$

2.1.4 Función Inversa

La función inversa de la CDF está dada por:

$$F^{-1}(y) = a + y(b-a)$$

Paso a paso:

1. Definición de la CDF:

$$F(x) = \frac{x-a}{b-a}$$

2. Igualar la CDF a una variable y:

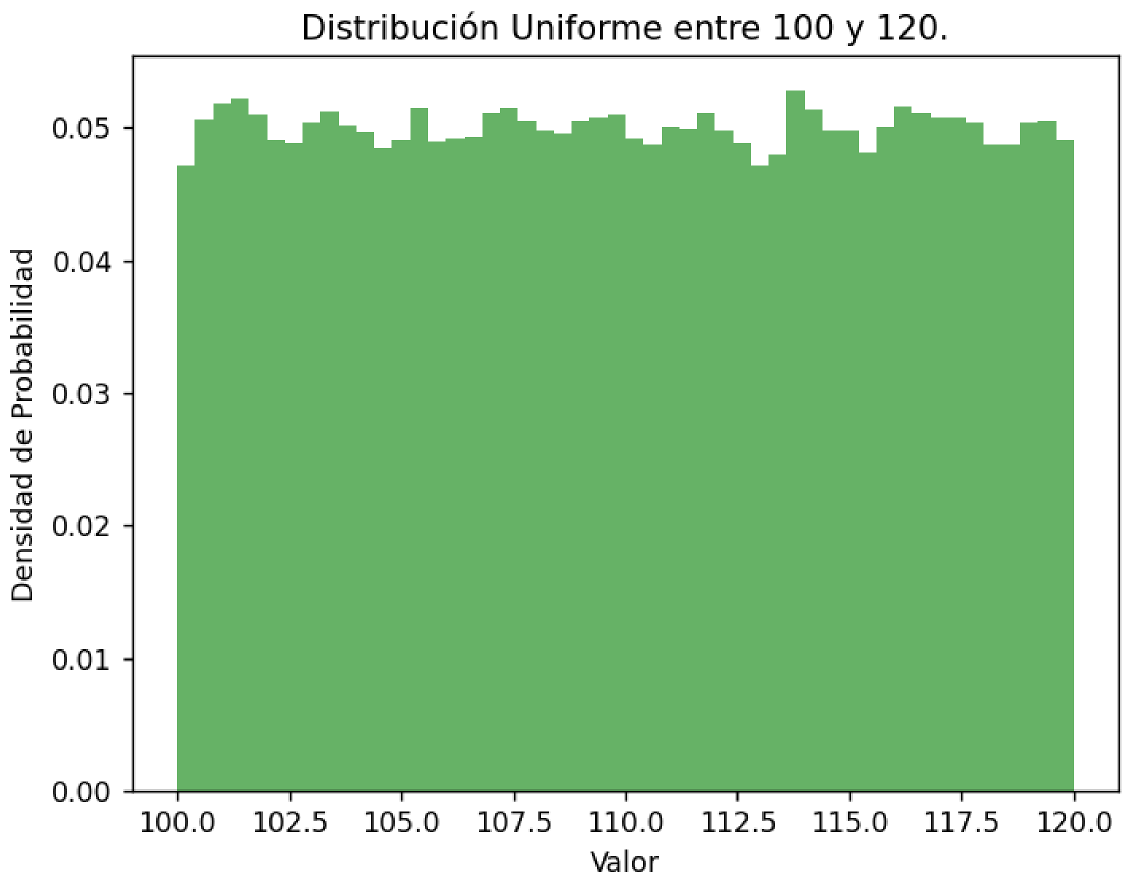
$$y = \frac{x-a}{b-a}$$

3. Despejar x:

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

4. Resultado Final:

$$F^{-1}(y) = a + y(b-a)$$



2.1.5 Codigo

```
import matplotlib.pyplot as plt
import os
import sys
sys.path.append(os.path.abspath(os.path.join(os.path.dirname(__file__), '..')))
from TP2_1.generators import GLCGenerator

glc = GLCGenerator(10, 2**31-1, 12345, 1103515245)

def distribucion_uniforme(a, b, x):
    values = []
    for _ in range(x):
        values.append(glc.next())
    return [a + (b - a) * v for v in values]

uniform_values = distribucion_uniforme(100, 120, 100000)

def grafica_distribucionUniforme():
    plt.hist(uniform_values, bins=50, density=True, alpha=0.6, color='g')
    plt.title('Distribucion Uniforme entre 100 y 120')
    plt.xlabel('Valor')
    plt.ylabel('Densidad de Probabilidad')
    plt.show()

grafica_distribucionUniforme()
```

Listing 1: Generación y Gráfica de la Distribución Uniforme

2.1.6 Método de Rechazo

El método de rechazo para la distribución uniforme no es necesario, ya que se puede generar directamente utilizando las variables aleatorias uniformemente distribuidas.

Para generar una muestra X de una distribución uniforme en el intervalo $[a, b]$, se puede utilizar:

$$X = a + (b - a) \cdot U$$

donde U es una variable aleatoria uniforme en el intervalo $[0, 1]$.

2.1.7 Testeo

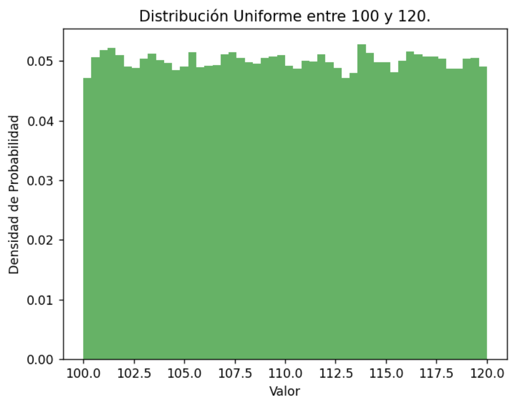


Figure 1: Grafica distribucion uniforme utilizando el generador GLC

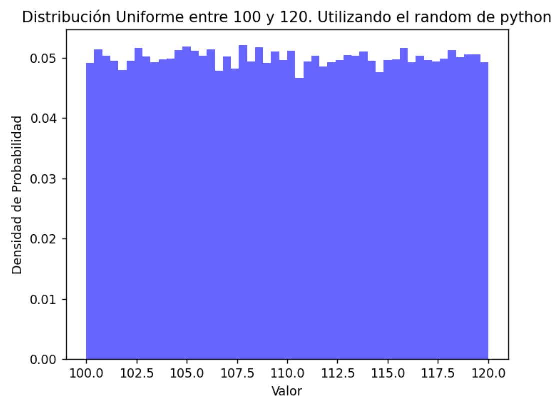


Figure 2: Grafica distribucion uniforme utilizando el generador Random de Python

2.2 Exponencial

La distribución exponencial es una distribución de probabilidad continua que se utiliza para modelar el tiempo entre eventos en un proceso de Poisson. Esta distribución se define mediante un único parámetro λ, que es la tasa de eventos por unidad de tiempo.

2.2.1 Parámetros de entrada

- λ**:** Tasa de eventos por unidad de tiempo.

2.2.2 Función de Densidad de Probabilidad (PDF)

La función de densidad de probabilidad (PDF) para la distribución exponencial está dada por:

$$f(x) = \lambda e^{-\lambda x} \quad \text{para} \quad x \geq 0$$

2.2.3 Función Acumulada (CDF)

La función acumulada (CDF) para la distribución exponencial está dada por:

$$F(x) = 1 - e^{-\lambda x} \quad \text{para} \quad x \geq 0$$

Paso a paso: 1. Definición de la PDF:

$$f(x) = \lambda e^{-\lambda x} \quad \text{para} \quad x \geq 0$$

2. Definición de la CDF:

$$F(x) = \int_0^x f(t)\,dt$$

3. Sustitución de la PDF:

$$F(x) = \int_0^x \lambda e^{-\lambda t}\,dt$$

4. Integración por partes:

$$F(x) = \left[-e^{-\lambda t}\right]_0^x = -e^{-\lambda x} - (-e^0) = 1 - e^{-\lambda x}$$

5. Resultado Final:

$$F(x) = 1 - e^{-\lambda x} \quad \text{para} \quad x \geq 0$$

2.2.4 Funcion Inversa

La función inversa de la CDF está dada por:

$$F^{-1}(y) = -\frac{1}{\lambda}\ln(1-y)$$

Paso a paso:
1. Definición de la CDF:

$$F(x) = 1 - e^{-\lambda x}$$

2. Igualar la CDF a una variable y:

$$y = 1 - e^{-\lambda x}$$

3. Despejar $e^{-\lambda x}$:

$$e^{-\lambda x} = 1 - y$$

4. Aplicar el logaritmo natural a ambos lados:

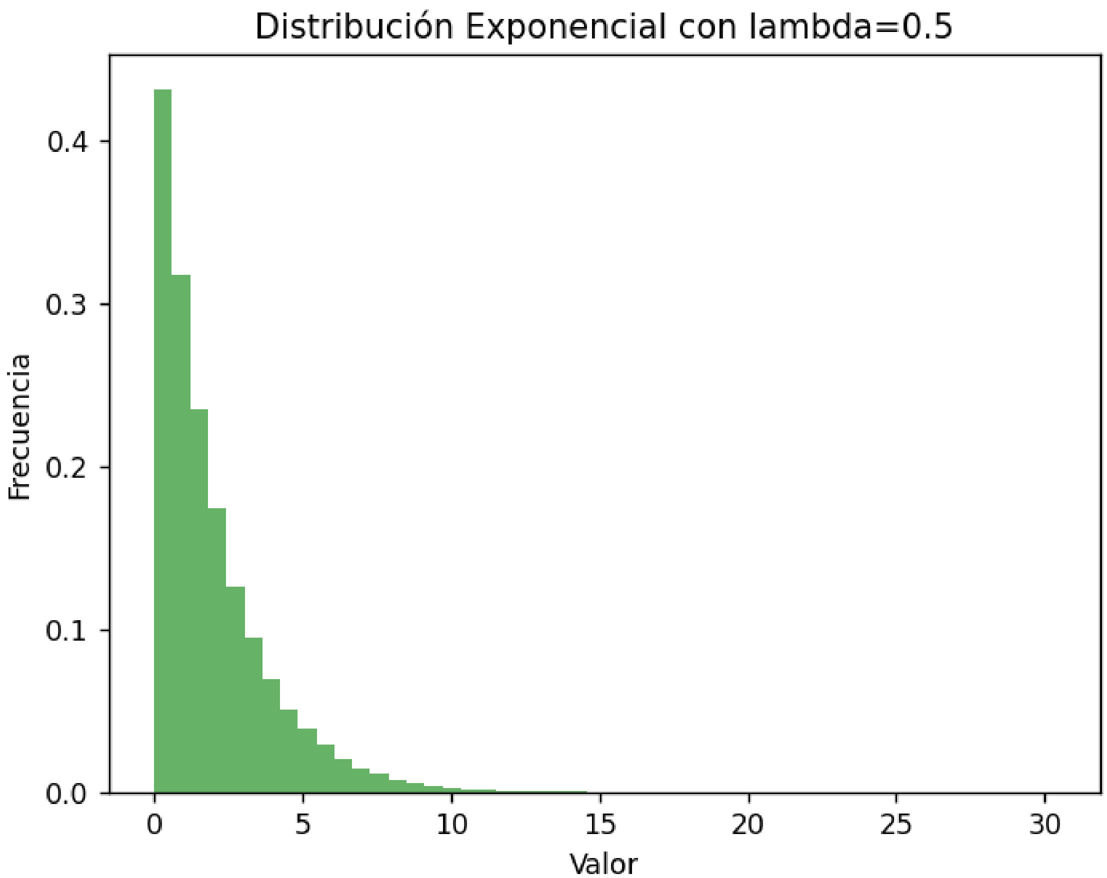
$$-\lambda x = \ln(1-y)$$

5. Despejar x:

$$x = -\frac{1}{\lambda}\ln(1-y)$$

6. Resultado Final:

$$F^{-1}(y) = -\frac{1}{\lambda}\ln(1-y)$$



2.2.5 Método de Rechazo

El método de rechazo para la distribución exponencial se basa en el uso de una distribución uniforme.

1. Generar dos variables aleatorias U_1 y U_2 independientes y uniformemente distribuidas en el intervalo $[0, 1]$.
2. Calcular $Y = -\frac{1}{\lambda}\ln(U_1)$.
3. Calcular $c \cdot f(Y)$, donde c es una constante mayor o igual al máximo valor de $f(x)$.
4. Si $U_2 \leq \frac{f(Y)}{c}$, aceptar Y como una muestra de la distribución exponencial. Si no, repetir desde el paso 1.

2.2.6 Testeo

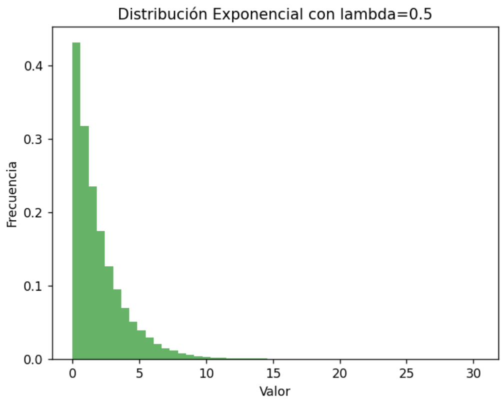


Figure 3: Gráfica de distribución exponencial utilizando GLC

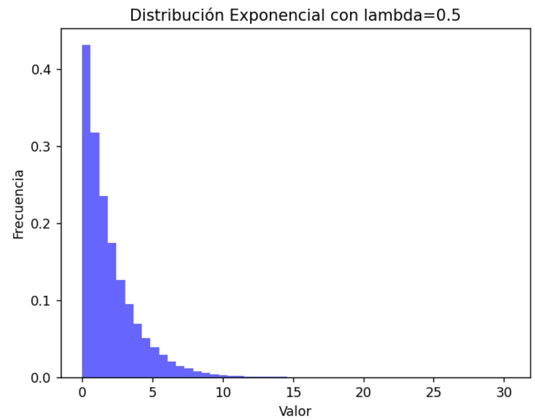


Figure 4: Gráfica de distribución exponencial utilizando el generador random de Python

2.3 Gamma

La distribución gamma es una distribución de probabilidad continua que se utiliza para modelar tiempos de espera y eventos en procesos con tasa variable. Esta distribución se define mediante dos parámetros: la forma α y la escala β.

2.3.1 Parámetros de entrada

- α**:** Parámetro de forma.
- β**:** Parámetro de escala.

2.3.2 Función de Densidad de Probabilidad (PDF)

La función de densidad de probabilidad (PDF) para la distribución gamma está dada por:

$$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} \quad \text{para} \quad x \geq 0$$

donde $\Gamma(\alpha)$ es la función gamma, definida como:

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}\,dt$$

2.3.3 Función Acumulada (CDF)

La función acumulada (CDF) para la distribución gamma está dada por:

$$F(x) = \frac{1}{\Gamma(\alpha)}\int_0^x t^{\alpha-1}e^{-t/\beta}\,dt$$

Paso a paso:

1. Definición de la PDF:

$$f(x) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} \quad \text{para} \quad x \geq 0$$

2. Definición de la CDF:

$$F(x) = \int_0^x f(t)\,dt$$

3. Sustitución de la PDF:

$$F(x) = \int_0^x \frac{t^{\alpha-1} e^{-t/\beta}}{\beta^\alpha \Gamma(\alpha)} \, dt$$

4. Factorización de constantes fuera de la integral:

$$F(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-t/\beta} \, dt$$

5. Transformación de variable:

$$u = \frac{t}{\beta}, \quad du = \frac{dt}{\beta}, \quad t = u\beta$$

6. Cambio de límites:

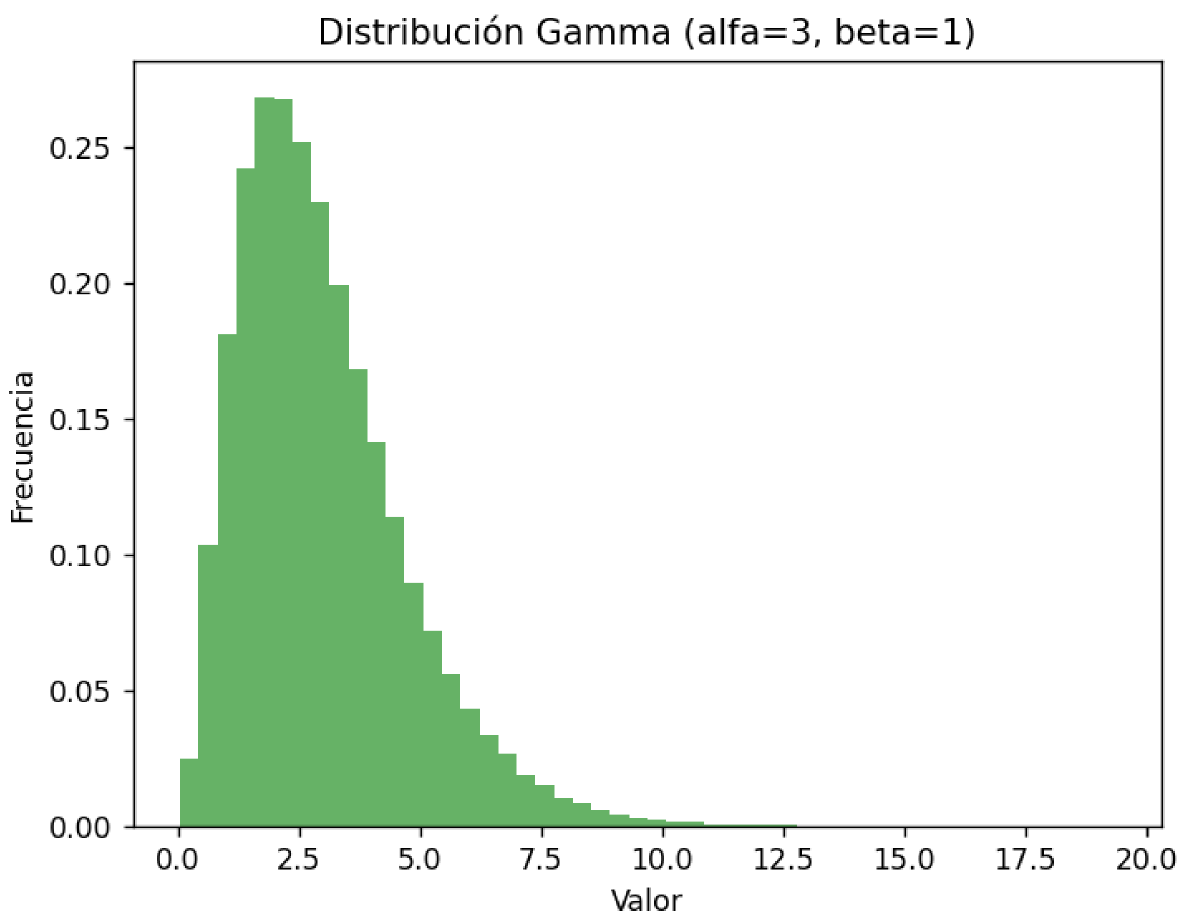
$$F(x) = \frac{1}{\Gamma(\alpha)} \int_0^{x/\beta} u^{\alpha-1} e^{-u} \, du$$

7. Resultado Final:

$$F(x) = \frac{\gamma(\alpha, x/\beta)}{\Gamma(\alpha)}$$

donde $\gamma(\alpha, x/\beta)$ es la función gamma incompleta definida como:

$$\gamma(\alpha, x/\beta) = \int_0^{x/\beta} t^{\alpha-1} e^{-t} \, dt$$



2.3.4 Método de Rechazo

El método de rechazo para la distribución gamma se puede implementar utilizando la distribución uniforme.

1. Generar dos variables aleatorias U_1 y U_2 independientes y uniformemente distribuidas en el intervalo $[0, 1]$.
2. Calcular $Y = -\ln(U_1)$.
3. Calcular $X = \beta Y$.
4. Calcular $c \cdot f(X)$, donde c es una constante mayor o igual al máximo valor de $f(x)$.
5. Si $U_2 \leq \frac{f(X)}{c}$, aceptar X como una muestra de la distribución gamma. Si no, repetir desde el paso 1.

2.4 Normal

La distribución normal es una distribución de probabilidad continua que se utiliza para modelar fenómenos naturales y variables aleatorias. Esta distribución se define mediante dos parámetros: la media μ y la desviación estándar σ.

2.4.1 Parámetros de entrada

- μ**:** Media de la distribución.
- σ**:** Desviación estándar de la distribución.

2.4.2 Función de Densidad de Probabilidad (PDF)

La función de densidad de probabilidad (PDF) para la distribución normal está dada por:

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

2.4.3 Función Acumulada (CDF)

La función acumulada (CDF) para la distribución normal está dada por:

$$F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

donde erf es la función error, definida como:

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt$$

Paso a paso:

1. Definición de la PDF:

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

2. Definición de la CDF:

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

3. Sustitución de la PDF:

$$F(x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt$$

4. Transformación de variable:

$$z = \frac{t-\mu}{\sigma\sqrt{2}}, \quad dz = \frac{dt}{\sigma\sqrt{2}}$$

5. Cambio de límites:

$$F(x) = \int_{-\infty}^{\frac{x-\mu}{\sigma\sqrt{2}}} \frac{1}{\sqrt{\pi}} e^{-z^2}\, dz$$

6. Uso de la función error:

$$F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

7. Resultado Final:

$$F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right]$$

2.4.4 Función Inversa

La función inversa de la CDF no tiene una forma cerrada simple, pero puede ser aproximada numéricamente. Para la distribución normal estándar ($\mu = 0$ y $\sigma = 1$), se usa la función inversa de la CDF estándar, denotada como $\Phi^{-1}(y)$.

Paso a paso:

1. Definición de la CDF para la distribución normal estándar:

$$\Phi(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

2. Igualar la CDF a una variable y:

$$y = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$$

3. Despejar $\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)$:

$$\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) = 2y - 1$$

4. Aplicar la función inversa de la función error:

$$\frac{x}{\sqrt{2}} = \operatorname{erf}^{-1}(2y-1)$$

5. Despejar x:

$$x = \sqrt{2}\operatorname{erf}^{-1}(2y-1)$$

6. Generalización para cualquier μ y σ:

$$F^{-1}(y) = \mu + \sigma\sqrt{2}\operatorname{erf}^{-1}(2y-1)$$

7. Resultado Final:

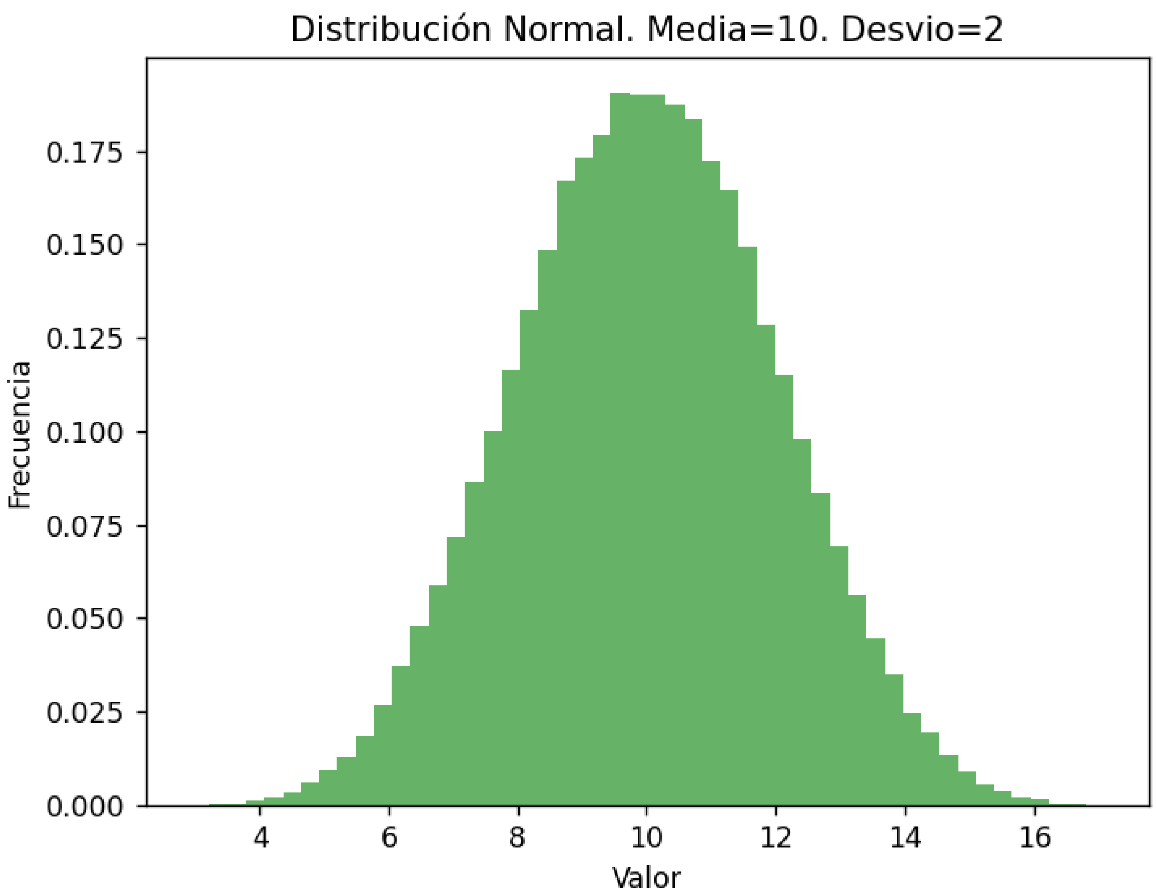
$$F^{-1}(y) = \mu + \sigma\sqrt{2}\operatorname{erf}^{-1}(2y-1)$$

2.4.5 Codigo

```
import matplotlib.pyplot as plt
import numpy as np
import os
import sys
sys.path.append(os.path.abspath(os.path.join(os.path.dirname(__file__), '..')))
from TP2_1.generators import GLCGenerator

glc = GLCGenerator(10, 2**31-1, 12345, 1103515245)

def generador_normal(media, desvio, K, x):
    values = []
    for _ in range(x):
        sum_normal = 0
        for _ in range(K):
            sum_normal += glc.next()
        sum_normal = ( sum_normal - (K/2) )
        multiplicar = desvio * pow((12/K), 0.5)
```

```
        values.append(sum_normal * multiplicar + media)
    return values
K = 5
media = 10
desvio = 2
x = 100000

valores_normales = generador_normal(media, desvio, K, x)

# Graficar la distribucion normal
def grafica_distribucionNormal():
    plt.hist(valores_normales, bins=50, density=True, alpha=0.6, color='b')
    plt.title('Distribucion Normal. Media=10. Desvio=2')
    plt.xlabel('Valor')
    plt.ylabel('Frecuencia')
    plt.show()
```

Listing 2: Generación y Gráfica de la Distribución Normal



2.4.6 Método de Rechazo

El método de rechazo para la distribución normal se puede implementar usando la distribución uniforme.

1. Generar dos variables aleatorias U_1 y U_2 independientes y uniformemente distribuidas en el intervalo $[0, 1]$.
2. Calcular $Y = -\ln(U_1)$.
3. Calcular $X = \mu + \sigma\sqrt{2Y}\cos(2\pi U_2)$.

4. Calcular $c \cdot f(X)$, donde c es una constante mayor o igual al máximo valor de $f(x)$.
5. Si $U_1 \leq \frac{f(X)}{c}$, aceptar X como una muestra de la distribución normal. Si no, repetir desde el paso 1.

2.4.7 Testeo

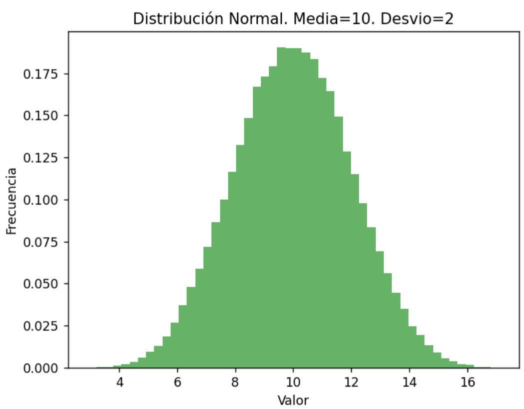


Figure 5: Gráfica de distribución normal utilizando GLC

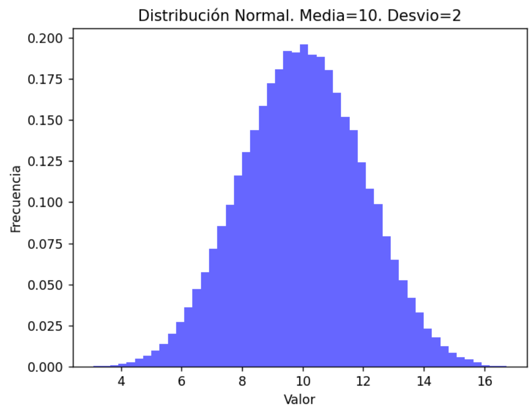


Figure 6: Gráfica de distribución normal utilizando el generador random de Python

2.5 Pascal

La distribución Pascal, también conocida como la distribución binomial negativa, es una distribución de probabilidad discreta que se utiliza para modelar el número de ensayos necesarios para obtener un número fijo de éxitos en experimentos de Bernoulli independientes. Esta distribución se define mediante dos parámetros: el número de éxitos r y la probabilidad de éxito en cada ensayo p.

2.5.1 Parámetros de entrada

- r**:** Número de éxitos.
- p**:** Probabilidad de éxito en cada ensayo.

2.5.2 Función de Masa de Probabilidad (PMF)

La función de masa de probabilidad (PMF) para la distribución Pascal está dada por:

$$P(X = k) = \binom{k+r-1}{k}(1-p)^k p^r \quad \text{para} \quad k = 0, 1, 2, \ldots$$

donde $\binom{k+r-1}{k}$ es el coeficiente binomial, definido como:

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

2.5.3 Función Acumulada (CDF)

La función acumulada (CDF) para la distribución Pascal está dada por:

$$F(k) = \sum_{i=0}^{k} P(X = i)$$

Paso a paso:

1. Definición de la PMF:

$$P(X = k) = \binom{k+r-1}{k}(1-p)^k p^r \quad \text{para} \quad k = 0, 1, 2, \ldots$$

2. Definición de la CDF:

$$F(k) = \sum_{i=0}^{k} P(X = i)$$

3. Sustitución de la PMF:

$$F(k) = \sum_{i=0}^{k} \binom{i+r-1}{i} (1-p)^i p^r$$

4. Sumar los términos:

$$F(k) = p^r \sum_{i=0}^{k} \binom{i+r-1}{i} (1-p)^i$$

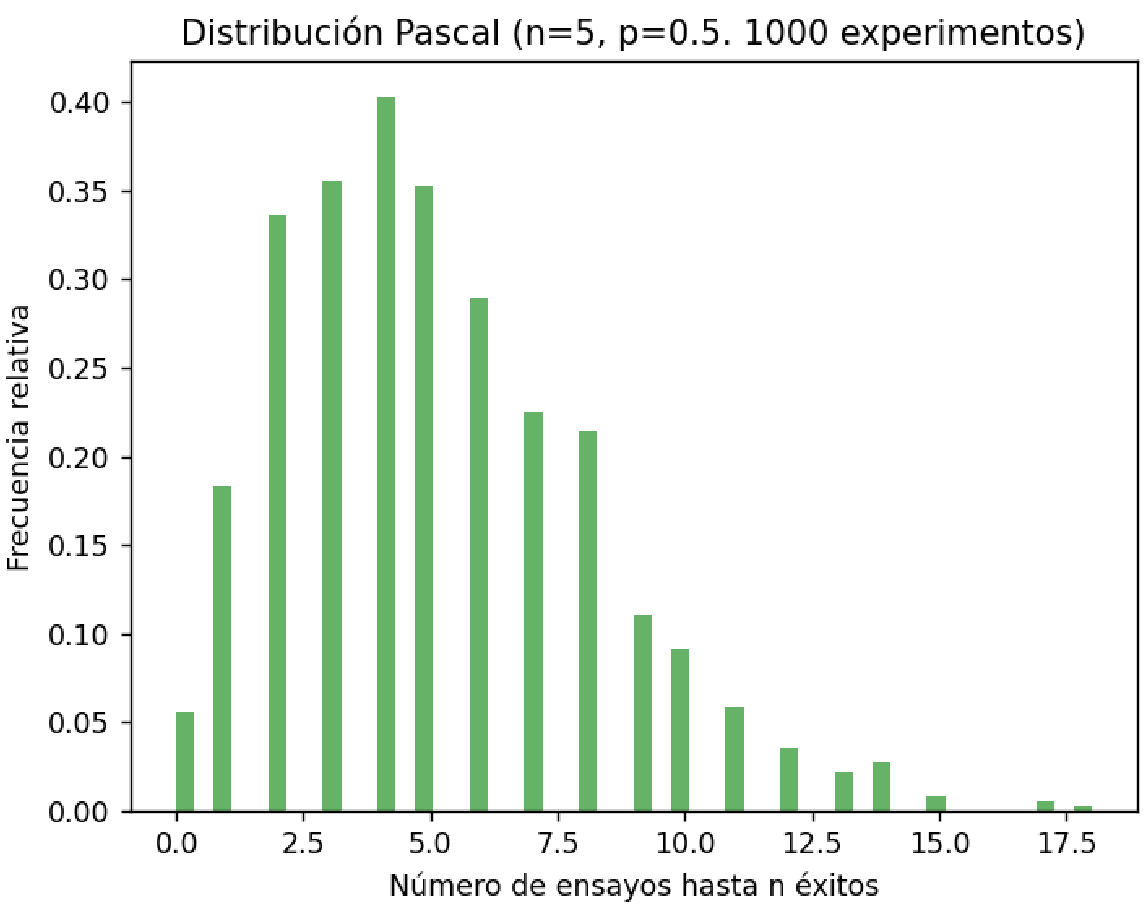
5. Resultado Final:

$$F(k) = 1 - I_{1-p}(r, k+1)$$

donde $I_{1-p}(r, k+1)$ es la función beta incompleta regularizada definida como:

$$I_x(a,b) = \frac{B(x; a, b)}{B(a, b)}$$

donde $B(x; a, b)$ es la función beta incompleta y $B(a, b)$ es la función beta.



2.5.4 Método de Rechazo

El método de rechazo para la distribución Pascal se puede implementar usando la distribución uniforme.

1. Generar dos variables aleatorias U_1 y U_2 independientes y uniformemente distribuidas en el intervalo $[0, 1]$.

2. Calcular $Y = -\ln(U_1)$.
3. Calcular $X = \left\lfloor \frac{Y}{(1-p)} \right\rfloor$.
4. Calcular $c \cdot P(X)$, donde c es una constante mayor o igual al máximo valor de $P(k)$.
5. Si $U_2 \leq \frac{P(X)}{c}$, aceptar X como una muestra de la distribución Pascal. Si no, repetir desde el paso 1.

2.6 Binomial

La distribución binomial es una distribución de probabilidad discreta que se utiliza para modelar el número de éxitos en una secuencia de n ensayos de Bernoulli independientes, cada uno con una probabilidad de éxito p. Esta distribución se define mediante dos parámetros: el número de ensayos n y la probabilidad de éxito en cada ensayo p.

2.6.1 Parámetros de entrada

- n**:** Número de ensayos.
- p**:** Probabilidad de éxito en cada ensayo.

2.6.2 Función de Masa de Probabilidad (PMF)

La función de masa de probabilidad (PMF) para la distribución binomial está dada por:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{para} \quad k = 0, 1, 2, \ldots, n$$

donde $\binom{n}{k}$ es el coeficiente binomial, definido como:

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

2.6.3 Función Acumulada (CDF)

La función acumulada (CDF) para la distribución binomial está dada por:

$$F(k) = \sum_{i=0}^{k} P(X = i)$$

Paso a paso:

1. Definición de la PMF:

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \quad \text{para} \quad k = 0, 1, 2, \ldots, n$$

2. Definición de la CDF:

$$F(k) = \sum_{i=0}^{k} P(X = i)$$

3. Sustitución de la PMF:

$$F(k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$$

4. Sumar los términos:

$$F(k) = \sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i}$$

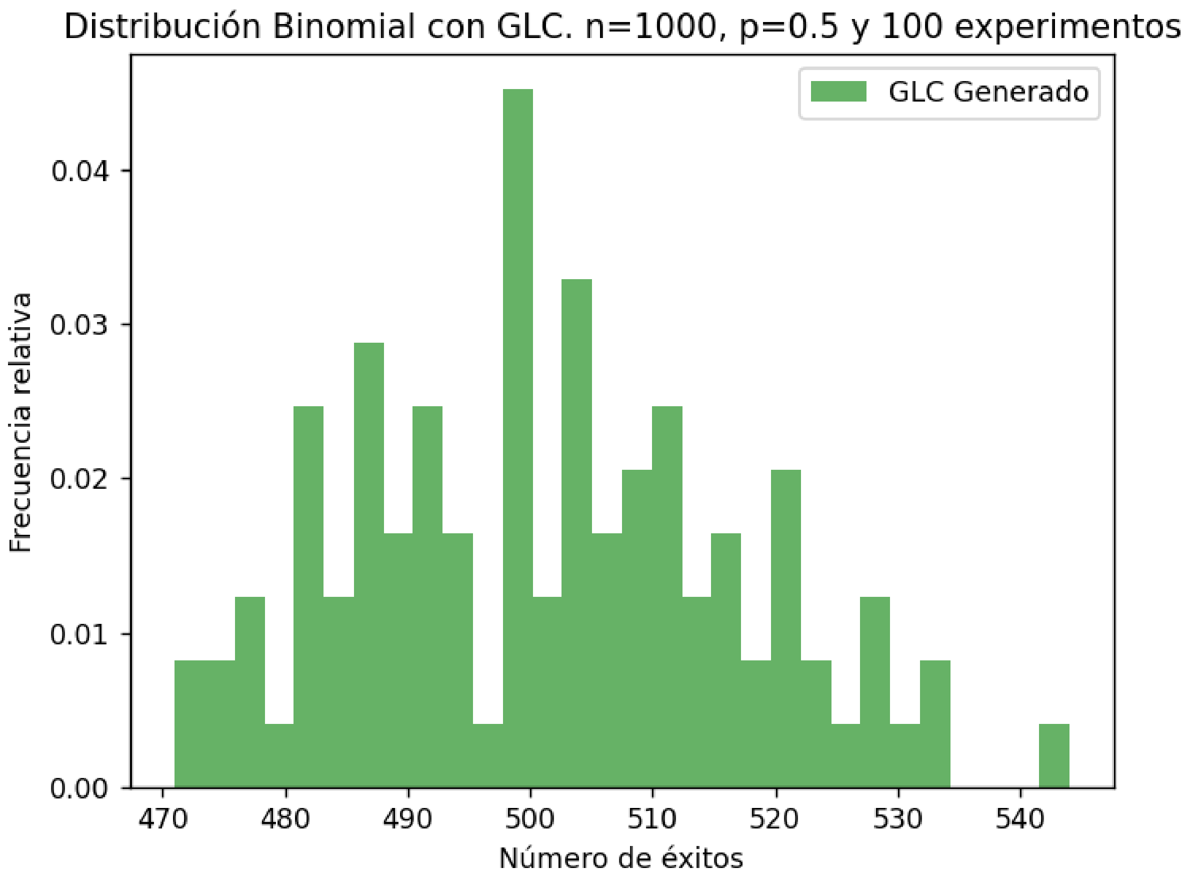
5. Resultado Final:

$$F(k) = I_p(k+1, n-k)$$

donde $I_p(k+1, n-k)$ es la función beta incompleta regularizada definida como:

$$I_x(a,b) = \frac{B(x;a,b)}{B(a,b)}$$

donde $B(x;a,b)$ es la función beta incompleta y $B(a,b)$ es la función beta.



2.6.4 Testeo

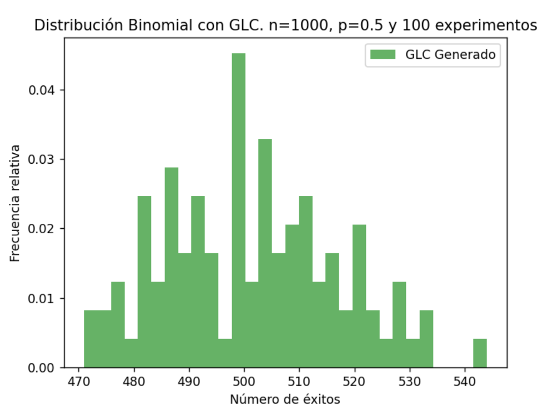


Figure 7: Gráfica de distribución binomial utilizando GLC

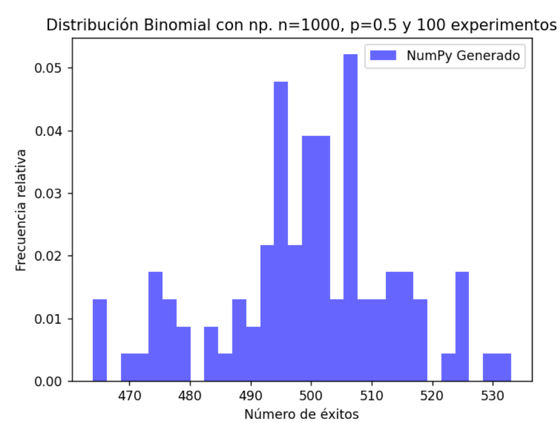


Figure 8: Gráfica de distribución binomial utilizando el generador random de Python

2.6.5 Método de Rechazo

El método de rechazo para la distribución binomial se puede implementar utilizando la distribución uniforme.

1. Generar dos variables aleatorias U_1 y U_2 independientes y uniformemente distribuidas en el intervalo $[0, 1]$.
2. Calcular $Y = -\ln(U_1)$.
3. Calcular $X = \left\lfloor \frac{Y}{(1-p)} \right\rfloor$.
4. Calcular $c \cdot P(X)$, donde c es una constante mayor o igual al máximo valor de $P(k)$.
5. Si $U_2 \leq \frac{P(X)}{}$

2.7 Hipergeometrica

La distribución hipergeométrica es una distribución de probabilidad discreta que se utiliza para modelar el número de éxitos en una muestra extraída sin reemplazo de una población finita. Esta distribución se define mediante tres parámetros: el tamaño de la población N, el número de éxitos en la población K, y el tamaño de la muestra n.

2.7.1 Parámetros de entrada

- N**:** Tamaño de la población.
- K**:** Número de éxitos en la población.
- n**:** Tamaño de la muestra.

2.7.2 Función de Masa de Probabilidad (PMF)

La función de masa de probabilidad (PMF) para la distribución hipergeométrica está dada por:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \quad \text{para} \quad k = \max(0, n-(N-K)), \ldots, \min(K, n)$$

donde $\binom{a}{b}$ es el coeficiente binomial, definido como:

$$\binom{a}{b} = \frac{a!}{b!(a-b)!}$$

2.7.3 Función Acumulada (CDF)

La función acumulada (CDF) para la distribución hipergeométrica está dada por:

$$F(k) = \sum_{i=\max(0,n-(N-K))}^{k} P(X = i)$$

Paso a paso:

1. Definición de la PMF:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \quad \text{para} \quad k = \max(0, n-(N-K)), \ldots, \min(K, n)$$

2. Definición de la CDF:

$$F(k) = \sum_{i=\max(0,n-(N-K))}^{k} P(X = i)$$

3. Sustitución de la PMF:

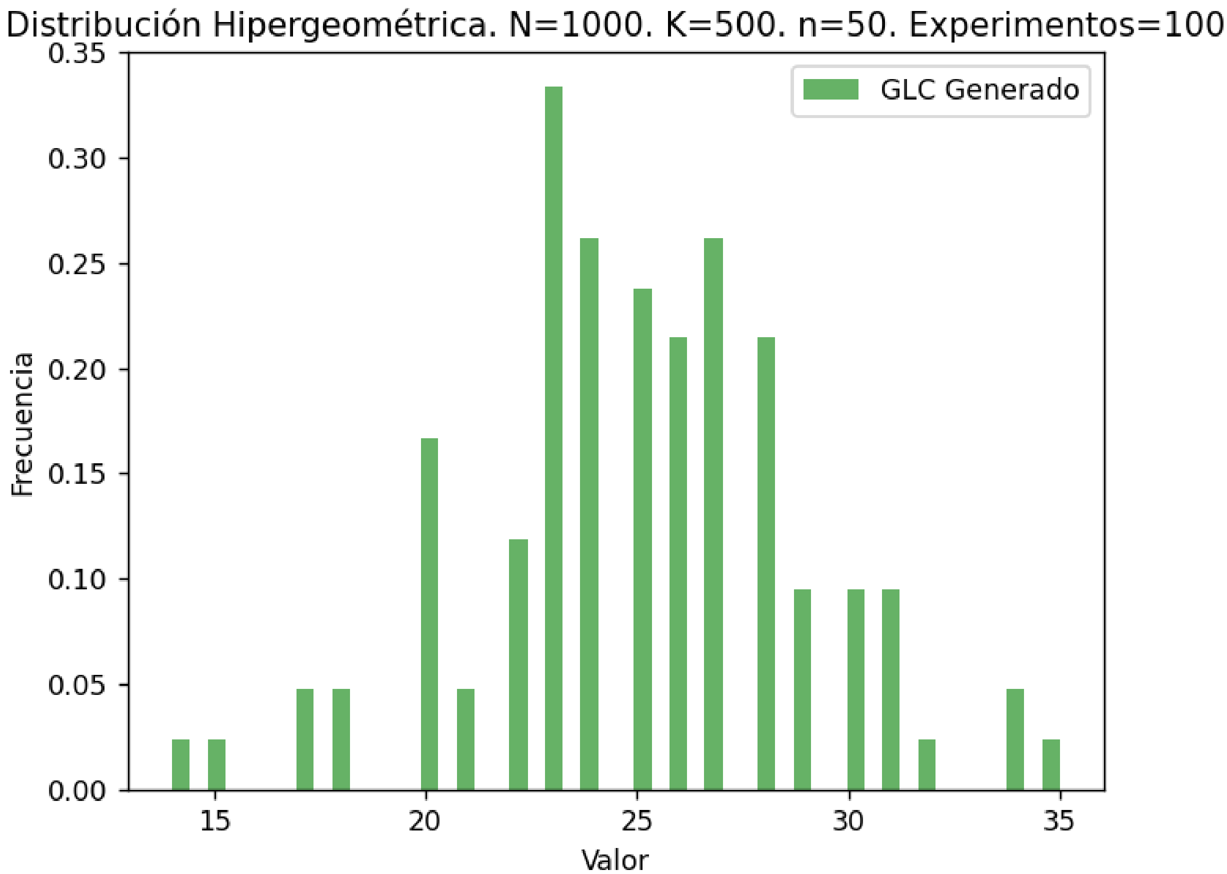
$$F(k) = \sum_{i=\max(0,n-(N-K))}^{k} \frac{\binom{K}{i}\binom{N-K}{n-i}}{\binom{N}{n}}$$

4. Sumar los términos:

$$F(k) = \frac{1}{\binom{N}{n}} \sum_{i=\max(0,n-(N-K))}^{k} \binom{K}{i}\binom{N-K}{n-i}$$

5. Resultado Final:

$$F(k) = \frac{1}{\binom{N}{n}} \sum_{i=\max(0,n-(N-K))}^{k} \binom{K}{i}\binom{N-K}{n-i}$$



2.7.4 Método de Rechazo

El método de rechazo para la distribución hipergeométrica utiliza la distribución uniforme.

1. Generar dos variables aleatorias U_1 y U_2 independientes y uniformemente distribuidas en el intervalo $[0, 1]$.

2. Calcular $Y = U_1 \cdot \binom{N}{n}$.

3. Si $Y \leq \binom{K}{k}\binom{N-K}{n-k}$ para algún k, aceptar k como una muestra de la distribución hipergeométrica. Si no, repetir desde el paso 1.

2.8 Poisson

La distribución de Poisson es una distribución de probabilidad discreta que se utiliza para modelar el número de eventos que ocurren en un intervalo de tiempo fijo o en un área fija. Esta distribución se define mediante un único parámetro λ, que es la tasa media de eventos por intervalo.

2.8.1 Parámetros de entrada

- **λ:** Tasa media de eventos por intervalo.

2.8.2 Función de Masa de Probabilidad (PMF)

La función de masa de probabilidad (PMF) para la distribución de Poisson está dada por:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{para} \quad k = 0, 1, 2, \ldots$$

2.8.3 Función Acumulada (CDF)

La función acumulada (CDF) para la distribución de Poisson está dada por:

$$F(k) = \sum_{i=0}^{k} P(X = i)$$

Paso a paso:

1. Definición de la PMF:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{para} \quad k = 0, 1, 2, \ldots$$

2. Definición de la CDF:

$$F(k) = \sum_{i=0}^{k} P(X = i)$$

3. Sustitución de la PMF:

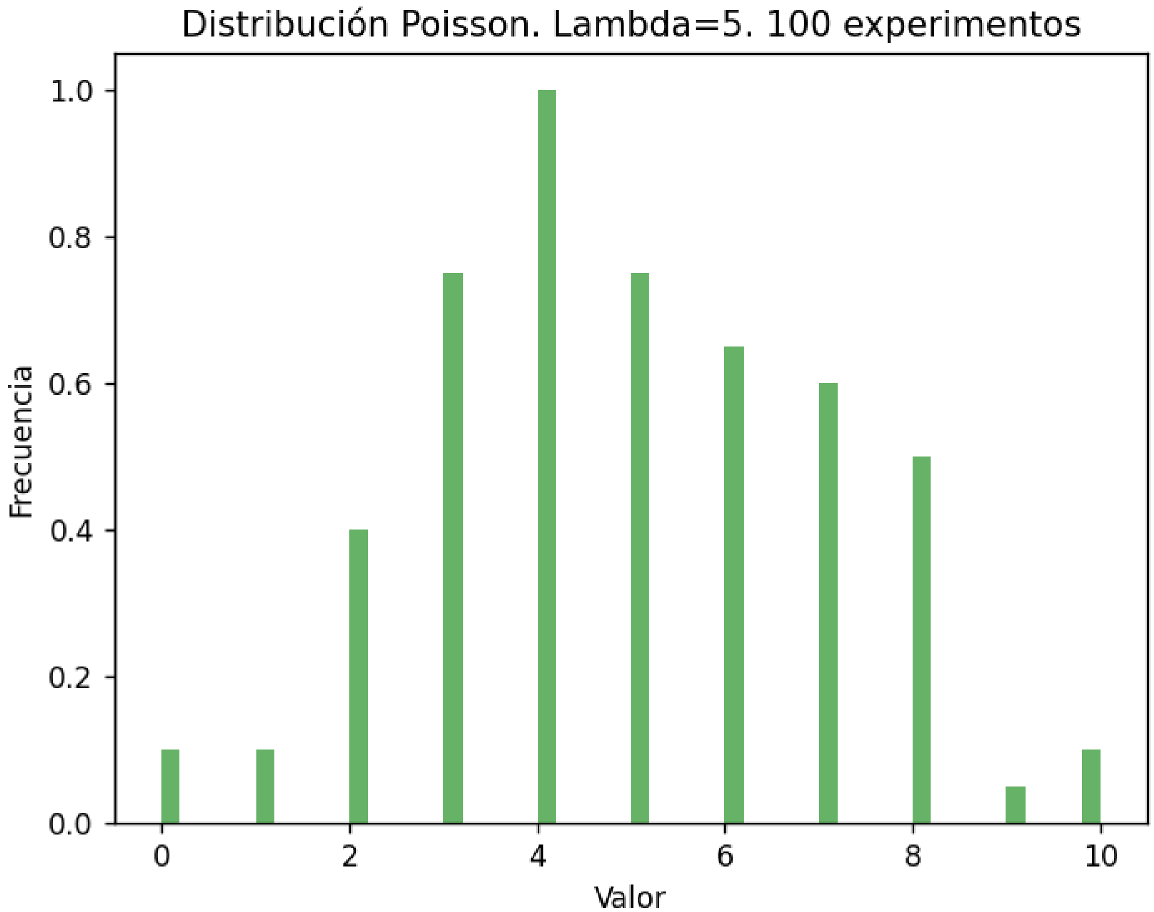
$$F(k) = \sum_{i=0}^{k} \frac{\lambda^i e^{-\lambda}}{i!}$$

4. Factorización de $e^{-\lambda}$ fuera de la suma:

$$F(k) = e^{-\lambda} \sum_{i=0}^{k} \frac{\lambda^i}{i!}$$

5. Resultado Final:

$$F(k) = e^{-\lambda} \sum_{i=0}^{k} \frac{\lambda^i}{i!}$$



2.8.4 Testeo

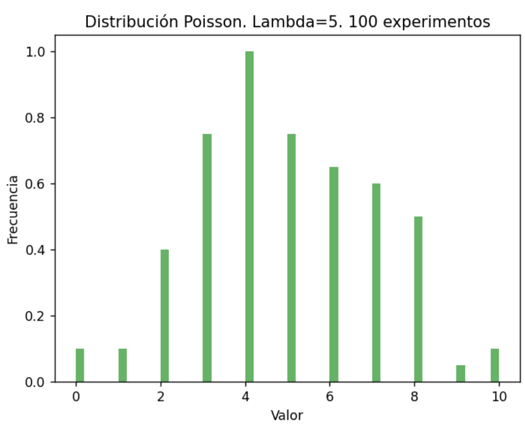


Figure 9: Gráfica de distribución Poisson utilizando GLC

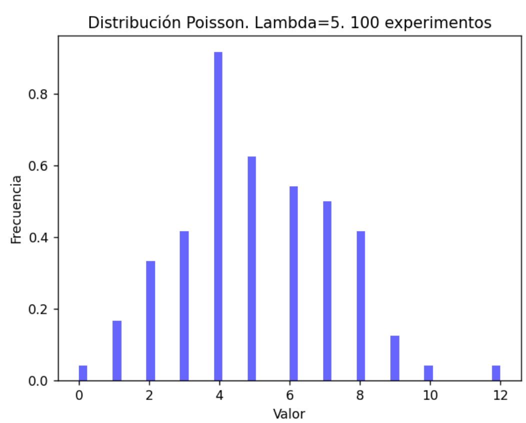


Figure 10: Gráfica de distribución Poisson utilizando el generador random de Python

2.8.5 Método de Rechazo

El método de rechazo para la distribución de Poisson se implementa usando la distribución uniforme.

1. Generar dos variables aleatorias U_1 y U_2 independientes y uniformemente distribuidas en el intervalo $[0, 1]$.

2. Calcular $Y = 0$.
3. Inicializar $k = 0$ y $p = e^{-\lambda}$.
4. Repetir:
   - Incrementar k.
   - Actualizar $Y = Y + \ln(U_1)$.
   - Si $Y > \lambda \ln(p/k)$, aceptar $k - 1$ como una muestra de la distribución de Poisson. Si no, repetir el ciclo.

2.9 Empírica Discreta

La distribución empírica discreta es una distribución de probabilidad que se utiliza para modelar datos observados directamente sin suponer una forma específica de la distribución subyacente. Esta distribución se define mediante los valores observados y sus probabilidades asociadas.

2.9.1 Parámetros de entrada

- x_i**:** Valores observados (donde $i = 1, 2, \ldots, n$).
- p_i**:** Probabilidad asociada a cada valor observado x_i tal que $\sum_{i=1}^{n} p_i = 1$.

2.9.2 Función de Masa de Probabilidad (PMF)

La función de masa de probabilidad (PMF) para la distribución empírica discreta está dada por:

$$P(X = x_i) = p_i \quad \text{para} \quad i = 1, 2, \ldots, n$$

2.9.3 Función Acumulada (CDF)

La función acumulada (CDF) para la distribución empírica discreta está dada por:

$$F(x) = \sum_{x_i \leq x} P(X = x_i)$$

Paso a paso:

1. Definición de la PMF:

$$P(X = x_i) = p_i \quad \text{para} \quad i = 1, 2, \ldots, n$$

2. Definición de la CDF:

$$F(x) = \sum_{x_i \leq x} P(X = x_i)$$

3. Sustitución de la PMF:

$$F(x) = \sum_{x_i \leq x} p_i$$

4. Ordenar los valores observados x_i de menor a mayor:

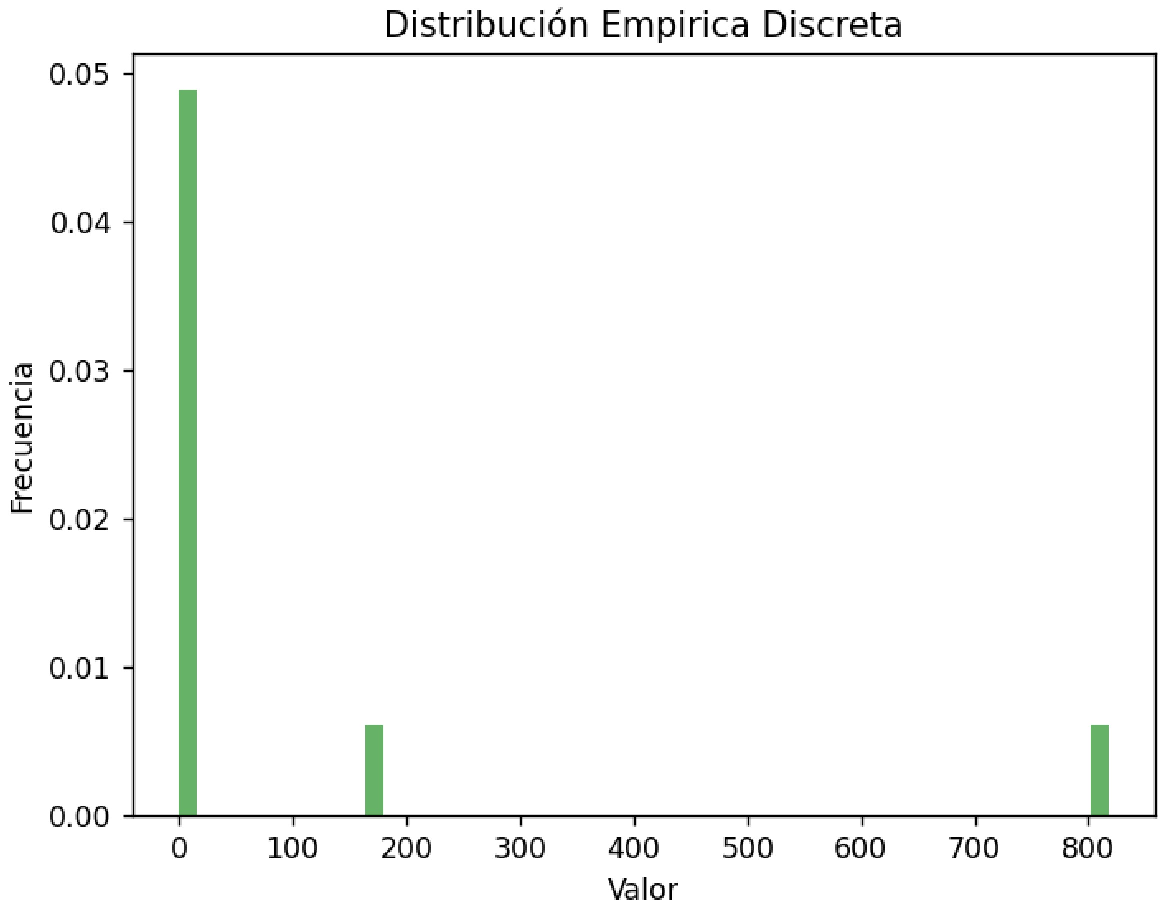
$$x_1, x_2, \ldots, x_n \quad \text{donde} \quad x_1 \leq x_2 \leq \ldots \leq x_n$$

5. Sumar las probabilidades asociadas a los valores observados menores o iguales a x:

$$F(x) = \sum_{i: x_i \leq x} p_i$$

6. Resultado Final:

$$F(x) = \sum_{i: x_i \leq x} p_i$$



2.9.4 Testeo

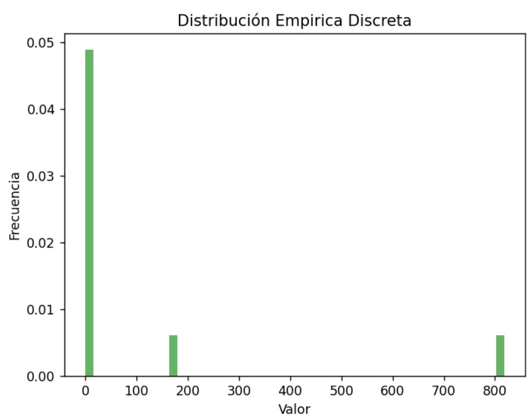


Figure 11: Gráfica de distribución empírica utilizando GLC

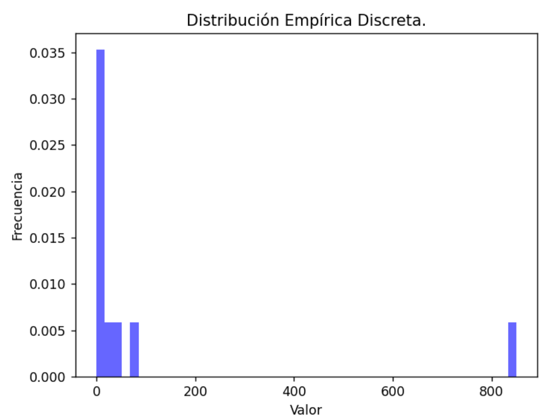


Figure 12: Gráfica de distribución empírica utilizando el generador random de Python

2.9.5 Método de Rechazo

El método de rechazo para la distribución empírica discreta es directo y utiliza la distribución uniforme.

1. Generar un número entero aleatorio i en el conjunto $\{1, 2, \ldots, k\}$ con probabilidades $p_1, p_2, \ldots, p_k$.

2. Aceptar x_i como una muestra de la distribución empírica discreta.

3 Conclusiones

Los resultados obtenidos en este estudio destacan la importancia de implementar correctamente generadores de números pseudoaleatorios para simular distribuciones de probabilidad específicas. En comparación con los métodos anteriores evaluados, se ha demostrado que la selección adecuada del generador tiene un impacto significativo en la calidad y la fiabilidad de los resultados obtenidos.

Las distribuciones de probabilidad analizadas abarcan tanto variables continuas como discretas, cada una con sus propias características y aplicaciones. Desde la distribución uniforme, utilizada como base para muchos generadores, hasta distribuciones más complejas como la normal, la hipergeométrica y la de Poisson, se ha detallado el proceso para generar números pseudoaleatorios que sigan fielmente estas distribuciones.

El Generador Lineal Congruencial (GLC) y la librería Random de Python han sido identificados como opciones robustas y confiables para la generación de números pseudoaleatorios en diversas aplicaciones. Estos métodos no solo han superado las pruebas estadísticas aplicadas, sino que también han demostrado ser adecuados para situaciones que requieren una alta calidad en la aleatoriedad, como simulaciones y aplicaciones criptográficas.

La comparación con métodos menos adecuados, como el método de los Cuadrados Medios, ha resaltado las limitaciones de estos en términos de ciclos cortos y sensibilidad a la semilla inicial. Esto subraya la importancia de seleccionar cuidadosamente el método de generación según los requisitos específicos de la aplicación.

En conclusión, este trabajo proporciona una guía práctica y teórica para la implementación efectiva de generadores de números pseudoaleatorios basados en distribuciones de probabilidad. La elección adecuada del método y su correcta implementación son fundamentales para asegurar la validez y precisión de los resultados en entornos de simulación y análisis estadístico.

References

[1] https://tereom.github.io/est-computacional-2018/numeros-pseudoaleatorios.html

[2] https://www.random.org/analysis/

[3] https://www.random.org/analysis/Analysis2005.pdf