
TP 2.2 GENERADORES DE NÚMEROS PSEUDOALEATORIOS DE DISTINTAS DISTRIBUCIONES DE PROBABILIDAD

Ornella Colazo
Universidad Tecnológica Nacional
Ingeniería en Sistemas
Legajo 47864
ornecolazo@gmail.com

Diego Navarro
Universidad Tecnológica Nacional
Ingeniería en Sistemas
Legajo 48029
navarrodiego201513@gmail.com

Matias Petrich
Universidad Tecnológica Nacional
Ingeniería en Sistemas
Legajo 46852
matias.petrich@gmail.com

Ramiro Cordoba
Universidad Tecnológica Nacional
Ingeniería en Sistemas
Legajo 46824
ramirocordobautn@gmail.com

Matias Ferullo
Universidad Tecnológica Nacional
Ingeniería en Sistemas
Legajo 48039
matias.ferullo1@gmail.com

25 de Abril, 2023

ABSTRACT

Para llevarse a cabo un estudio de simulación es imprescindible contar con variables aleatorias. Estas variables, depende el contexto, necesitan una determinada distribución de probabilidad. En este informe detallamos como, a través del método del rechazo y de la transformada inversa, podemos obtener variables aleatorias distribuidas de distintas formas más allá de la Uniforme.

1 Introducción.

La simulación está ampliamente reconocida como una técnica efectiva para construir pronósticos, evaluar riesgos, animar e ilustrar la evolución de un sistema en muchas áreas. Cuando existe incertidumbre en el comportamiento de algunos de los componentes del modelo de simulación, estos componentes aleatorios deben modelarse utilizando distribuciones de probabilidad y/o procesos estocásticos que son generados durante la simulación.

Un proceso estocástico es una secuencia de variables aleatorias $(X(t), t \in T)$, y que dependen del parámetro t , que toma valores en el conjunto índice T o de instantes de tiempo. Puesto que la palabra estocástico es un sinónimo de aleatoriedad, entonces un proceso estocástico es un sistema que nos permite darle seguimiento a un fenómeno aleatorio a través del tiempo. Cada valor obtenido mediante la variable aleatoria definida, nos dará información de lo que sucede con el fenómeno aleatorio conforme transcurre el tiempo. A cada valor posible se le llama un estado y a los distintos cambios de un estado a otro transición. Se denomina espacio de estados del proceso, S , al conjunto de todos los posibles valores de las variables aleatorias que componen el proceso.

Una variable aleatoria es una función que asigna un número real a cada uno de los posibles resultados de un experimento o característica de interés susceptibles de ser observados o medidos en una población objetivo.

La generación de valores de las variables aleatorias, tienen una naturaleza enteramente numérica y deben configurarse mediante la contribución de números aleatorios.

Al considerar procesos estocásticos, estos pueden ser de variables discretas (numerables) o continuas (no numerable), de tipo aleatorio. Definimos una función $F_x(x)$ llamada función de distribución acumulada de x .

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t) dt \quad (1)$$

En este caso, la integral representa la probabilidad de que la variable aleatoria X tome un valor menor o igual que x . $F_X(x)$ se define en el intervalo $0 \leq F_X(x) \leq 1$ y $f_X(t)$ representa el valor de la función de densidad de probabilidad de la variable X cuando $X = t$.

Si la variable aleatoria es discreta, entonces x tendrá valores específicos y $F_X(x)$ será una función escalonada. Si $F_X(x)$ es continua en el dominio de x , entonces esta función se podrá diferenciar, para lo cual se define $f_X(x) = dF_X(x)/dx$.

Los valores de tales variables uniformemente distribuidas en el intervalo $(0, 1)$ juegan un papel muy importante en la generación de los valores de variables aleatorias, obtenidas a partir de otros tipos de distribución de probabilidad.

Como se ha estudiado en informes anteriores, descubrimos que la aleatoriedad de los números no puede conseguirse tan fácilmente. Es por esto que pudimos determinar que nuestro generador de números pseudoaleatorios de Python, cumple con las condiciones requeridas para ser usado en estudios de Simulación. En este análisis debemos generar valores de variables aleatorias, que como hemos explicado, son necesarios para los procesos estocásticos. Si usáramos los valores que determina nuestro generador elegido, estaríamos empleando valores uniformes continuos entre 0 y 1. Sin embargo, para poder asociar los valores de las variables aleatorias a los procesos estocásticos, estos pueden estar determinados por cualquier distribución de probabilidad. La duda que se plantea es: ¿cómo hacemos para generar valores de distintas distribuciones?. Nuestro trabajo para este informe es, a través del Método de la Inversa y el Método del Rechazo, generar valores de distintas distribuciones.

2 Desarrollo.

2.1 Métodos

2.1.1 Método de la Transformada Inversa

El método de la Transformada Inversa, es el método más utilizado en la obtención de variables aleatorias para experimentos de simulación. Para aplicar éste método suponga que queremos generar el valor de una variable aleatoria discreta X con función de masa de probabilidad:

$$P(X = x_j) = p_j, j = 0, 1, \dots, \sum_j p_j = 1 \quad (2)$$

Para esto, generamos un número aleatorio R ; es decir, R está distribuido uniformemente en $(0, 1)$

$$X = \begin{cases} x_1 & \text{si } R < p_0 \\ x_2 & \text{si } p_0 \leq R < p_0 + p_1 \\ \vdots & \\ x_j & \text{si } \sum_{i=1}^{j-1} p_i \leq R < \sum_{i=1}^j p_i \\ \vdots & \end{cases} \quad (3)$$

Como $P(a \leq U < b) = b - a$ para $0 < a < b < 1$, tenemos que

$$P(X = x_j) = P\left(\sum_{i=1}^{j-1} p_i \leq R < \sum_{i=1}^j p_i\right) \quad (4)$$

y entonces X tiene la distribución deseada.

Algoritmo. Método de la transformada inversa para variables aleatorias discretas.

1. Generar un número aleatorio R .
2. Si $R < p_0$ hacer $X=x_0$ y terminar.

3. Si $R < p_0 + p_1$ hacer $X = x_1$ y terminar.
4. Si $R < \sum_{i=1}^j p_i$ hacer $X = x_j$ y terminar

Si los $x_i, i \geq 0$, están ordenados de modo que $x_0 < x_1 < x_2 < \dots$ y si F denota la función de distribución de X , entonces $F_X(x_k) = \sum_{i=0}^k p_i$:

X será igual a x_j , si $F_X(x_{j-1}) \leq R < F_X(x_j)$.

En otras palabras, después de generar un número aleatorio R determinamos el valor de X hallando el intervalo $(F_X(x_{j-1}), F_X(x_j))$ en el que está R .

Ahora consideremos si la variable que se desea generar es una variable aleatoria continua con función de distribución F .

Sea R una variable aleatoria uniforme en $(0, 1)$. Para cualquier función de distribución continua F , invertible, la variable aleatoria X definida como

$$X = F_X^{-1}(r) \quad (5)$$

donde F_X^{-1} es la transformación inversa.

para generar una variable aleatoria X a partir de la función de distribución continua F , generamos un número aleatorio R y hacemos entonces la transformada inversa.

2.1.2 Método del Rechazo

Otro procedimiento para generar valores de variables aleatorias de distribuciones de probabilidad no uniformes, es el método de rechazo. Este método consiste primeramente en generar un valor de la variable aleatoria y en seguida probar que dicho valor simulado proviene de la distribución de probabilidad que se está analizando. Para comprender la lógica de este método, suponga que $f(x)$ es una distribución de probabilidad acotada y con rango finito, es decir, $a \leq x \leq b$. De acuerdo a esta función de probabilidad, la aplicación del método de rechazo implica el desarrollo del siguiente algoritmo:

Algoritmo.

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

$$cf(x) \leq 1, a \leq x \leq b \quad (6)$$

2. Generar dos números uniformes R_1 y R_2 .
3. Determinar el valor de la variable aleatoria x de acuerdo a la siguiente relación lineal de R_1 : $x = a + (b - a) R_1$
4. Evaluar la función de probabilidad en $x = a + (b - a) R_1$.
5. Determinar si la siguiente desigualdad se cumple:

$$R_2 \leq cf(x = a + (b - a) R_1) \quad (7)$$

Se utiliza $x = a + (b - a) R_1$ si la respuesta es afirmativa como un valor simulado de la variable aleatoria. De lo contrario, es necesario regresar al paso 1 tantas veces como sea necesario

La teoría sobre la que se apoya este método se basa en el hecho de que la probabilidad de que $R_2 \leq cf(x)$ es exactamente $cf(x)$. Por consiguiente, si un número es escogido al azar de acuerdo a $x = a + (b - a) R_1$ y rechazado si $R_2 > cf(x)$, entonces la distribución de probabilidad de las x aceptadas será exactamente $f(x)$. Todas las x aceptadas estarán distribuidas uniformemente entre a y b . Autores explican que este método podría ser ineficiente para distribuciones de probabilidad con modas muy grandes.

2.2 Distribuciones de Probabilidad

2.2.1 Distribución Uniforme

La distribución uniforme es una distribución de probabilidad continua en la que la probabilidad de obtener un valor dentro de un intervalo es proporcional a la longitud de dicho intervalo. El intervalo de valores se define mediante dos parámetros: a y b , que representan los límites inferior y superior del intervalo, respectivamente. La función de densidad

de probabilidad de la distribución uniforme es constante dentro del intervalo y cero fuera de él. Los valores de los parámetros a y b determinan la altura de la función de densidad en el intervalo y su área total es igual a 1.

Función de probabilidad y gráficas: La función de densidad de probabilidad (PDF) de la distribución uniforme se define como:

$$f(x) = \begin{cases} \frac{1}{b-a} & ; a \leq x \leq b \\ 0 & ; \text{fuera del intervalo } (a,b) \end{cases} \quad (8)$$

donde a y b son los límites inferior y superior del intervalo. La función de probabilidad acumulada (CDF) se puede obtener integrando la PDF desde el límite inferior hasta un valor x :

$$F(x) = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a} \quad ; 0 \leq F(x) \leq 1 \quad (9)$$

La gráfica de la PDF es una línea horizontal en el intervalo $[a, b]$ y cero fuera de él. La gráfica de la CDF es una función lineal creciente en el intervalo $[a, b]$, que comienza en cero en a y termina en uno en b .

El valor esperado y la variancia de una variable aleatoria uniformemente distribuída están dados por las siguientes expresiones:

$$EX = \int_a^b \frac{1}{b-a} x dx = \frac{b+a}{2} \quad (10)$$

$$VX = \int_a^b \frac{(x-EX)^2}{b-a} dx = \frac{(b-a)^2}{12} \quad (11)$$

Método Transformada Inversa La función inversa de la distribución uniforme se puede obtener resolviendo la ecuación $F(x) = r$ para x , donde r es un número aleatorio uniforme generado en el intervalo $[0, 1]$. La solución de esta ecuación es:

$$x = a + (b-a)r \quad 0 \leq r \leq 1 \quad (12)$$

donde a y b son los límites inferior y superior del intervalo, respectivamente.

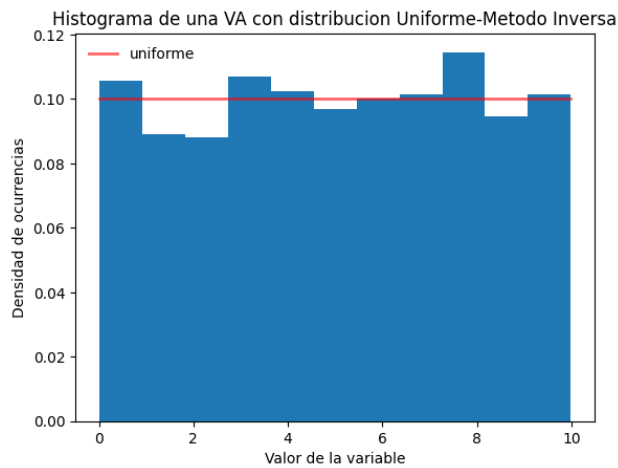


Figure 1: Histograma Distribución Uniforme - Inversa

```
1 def UniformeMI(numeros, a, b) -> list:
2     num = []
3     for r in numeros:
4         x = a+(b-a)*r
5         num.append(x)
6     return num
```

Listing 1: Código en Python del Método de la Inversa para la Distribución Uniforme

Método del rechazo El método del rechazo es un método general para generar números aleatorios de cualquier distribución de probabilidad a partir de una distribución uniforme. Para generar un número aleatorio de una distribución de probabilidad $f(x)$, se propone una función de densidad de probabilidad $g(x)$ que sea fácil de generar aleatoriamente y que encierre completamente la PDF de la distribución de probabilidad $f(x)$. Luego, se generan dos números aleatorios: uno de la distribución uniforme y otro de la distribución $g(x)$. Si el número generado de la distribución uniforme es menor o igual a la razón $f(x)/(Mg(x))$, donde M es una constante que satisface $Mg(x) \geq f(x)$ para todo x , entonces se acepta el número generado de la distribución $g(x)$ como una muestra de la distribución de probabilidad $f(x)$. Si no, se rechaza y se empieza de nuevo.

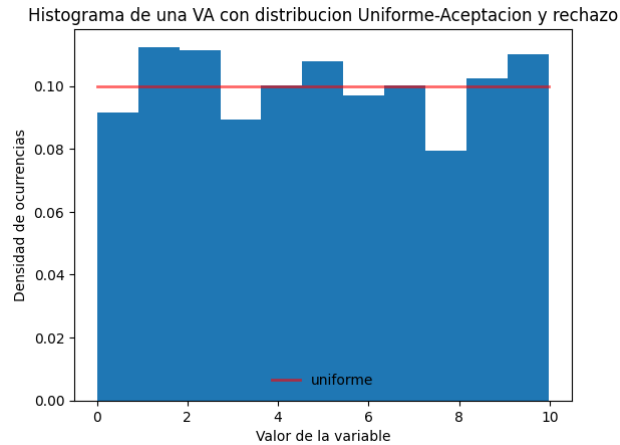


Figure 2: Histograma Distribución Uniforme - Rechazo

```

1 def UniformeR(numeros, a, b) -> list:
2     num = []
3     M = 1/(b-a)
4     for U in numeros:
5         while True:
6             V = np.random.uniform(0, 1)
7             T = a+((b-a)*V)
8             if (M*U <= 1/(b-a)):
9                 num.append(T)
10                break
11    return num
    
```

Listing 2: Código en Python del Método del Rechazo para la Distribución Uniforme

2.2.2 Distribución Exponencial

Transformada Inversa, Algoritmo del Método del Rechazo, Gráfica y Test La distribución exponencial es una distribución de probabilidad continua que se utiliza comúnmente en estadística para modelar eventos que ocurren aleatoriamente en el tiempo. Se utiliza ampliamente en áreas como la teoría de colas, el análisis de supervivencia y la teoría de la probabilidad.

Parámetros: La distribución exponencial tiene un parámetro λ , que es la tasa de eventos. La tasa de eventos representa la frecuencia promedio de ocurrencia de un evento en un período de tiempo dado. También se puede pensar en ella como la tasa de decaimiento de la probabilidad de que ocurra un evento.

Función de probabilidad: La función de densidad de probabilidad (PDF) de la distribución exponencial se define como:

$$f(x) = \lambda e^{-\lambda x} \quad ; \lambda > 0 \text{ y } x \geq 0 \quad (13)$$

La función de distribución acumulada (CDF) de la distribución exponencial se define como:

$$F(x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x} \quad (14)$$

El valor esperado y la variancia de una variable aleatoria exponencialmente distribuída están dados por las siguientes expresiones:

$$EX = \int_0^{\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda} \quad (15)$$

$$VX = \int_0^{\infty} (x - \frac{1}{\lambda})^2 \lambda e^{-\lambda x} dx = \frac{1}{\lambda^2} = (EX)^2 \quad (16)$$

Método Transformada Inversa La función de distribución acumulada inversa se calcula resolviendo la ecuación $F(x) = u$ para x , donde u es un valor aleatorio uniforme entre 0 y 1. Despejando x , se obtiene la siguiente ecuación:

$$x = -\frac{1}{\lambda} \log(r) = -EX \log(r) \quad \text{con } r = e^{-\lambda x} \quad (17)$$

Gráficos: La gráfica de la función de densidad de probabilidad de la distribución exponencial tiene forma de una curva que comienza en el eje x en 0, aumenta rápidamente al principio y luego disminuye a medida que se acerca al eje x positivo. La forma de la curva depende del valor del parámetro λ .

La gráfica de la función de distribución acumulada de la distribución exponencial es una curva creciente que comienza en 0 y se aproxima a 1 a medida que x se acerca a infinito.

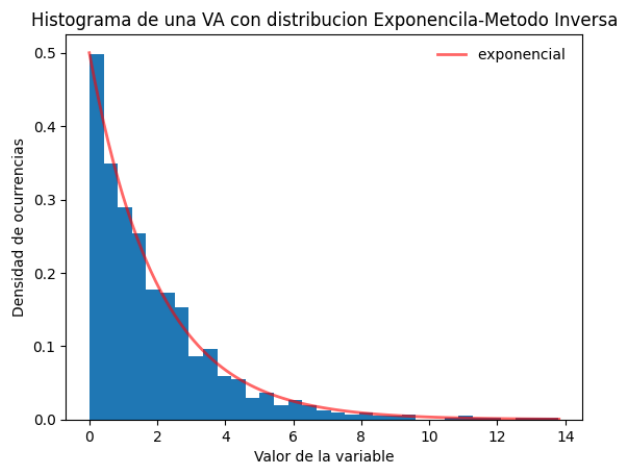


Figure 3: Histograma Distribución Exponencial - Metodo Inversa

Método Rechazo El método de rechazo se utiliza para generar valores aleatorios de la distribución exponencial. El método utiliza una distribución de probabilidad de prueba (en este caso, la distribución uniforme) para generar valores aleatorios que se utilizan para generar valores de la distribución exponencial. Para generar valores aleatorios de la distribución exponencial utilizando el método de rechazo, se sigue este proceso:

1. Se define una función que calcula la función de densidad de probabilidad de la distribución exponencial.
2. Se define una constante M que es un factor de escala utilizado para asegurarse de que la función de densidad de probabilidad de la distribución de prueba sea siempre mayor que la función de densidad de probabilidad de la distribución exponencial en el rango de interés.
3. Se generan valores aleatorios de la distribución de prueba y se calcula el valor correspondiente de la distribución exponencial. Si el valor de la distribución exponencial es menor que el valor de la distribución de prueba multiplicado por la constante M , se acepta el valor generado.

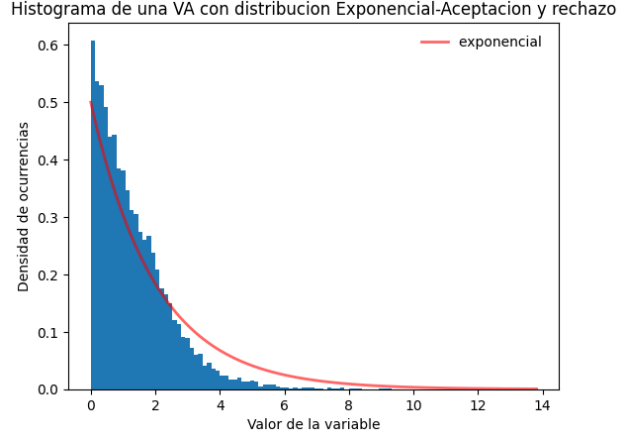


Figure 4: Histograma Distribución Exponencial - Rechazo

2.2.3 Distribución Normal

La distribución normal basa su utilidad en el teorema del limite central. Este teorema postula que, la distribución de probabilidad de la suma de N valores de variables aleatoria x_i independientes pero idénticamente distribuidos, con medias respectivas μ y variancias σ_i^2 se aproxima asintóticamente a una distribución normal, a medida que N se hace muy grande, y que dicha distribución tiene como media y variancia respectivamente, a:

$$\mu = \sum_{i=1}^N \mu_i \quad (18)$$

$$\sigma^2 = \sum_{i=1}^N \sigma_i^2 \quad (19)$$

Decimos que una variable aleatoria \mathbf{X} tiene una distribución normal de parámetros μ y σ , donde $\sigma > 0$ y notamos que $X \sim N(\mu, \sigma)$ cuando su función de densidad de probabilidad es:

$$f_X(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (20)$$

- μ : es la media.
- σ : es la desviación típica.
- σ^2 : es la varianza.

Si los parámetros de la distribución normal tienen los valores $\mu_x = 0$ y $\sigma_x = 1$, la función de distribución recibirá el nombre de distribución normal estándar, con función de densidad:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}, -\infty < z < \infty \quad (21)$$

Cualquier distribución normal se puede convertir a la estándar, mediante la siguiente sustitución:

$$z = \frac{x - \mu_x}{\sigma_x} \quad (22)$$

A fin de simular una distribución normal con media μ_x y variancia σ_x^2 dadas, se debe proponer la siguiente interpretación matemática del teorema del limite central. Si r_1, r_2, \dots, r_N representan variables aleatorias independientes, cada una de las cuales posee la misma distribución de probabilidad caracterizada por $E(r_i) = \theta$ y $Var(r_i) = \sigma^2$, entonces:

$$\lim_{N \rightarrow \infty} P[a < \frac{\sum_{i=1}^N r_i - N\theta}{\sqrt{N}\sigma} < b] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}z^2} dz \quad (23)$$

Donde:

$$E\left(\sum_{i=1}^N r_i\right) = N\theta \quad \text{Var}\left(\sum_{i=1}^N r_i\right) = N\sigma^2 \quad z = \frac{\sum_{i=1}^N r_i - N\theta}{\sigma\sqrt{N}}$$

El procedimiento para simular valores normales requiere el uso de la suma de K valores de variable aleatoria distribuidas uniformemente; esto es, la suma de r_1, r_2, \dots, r_k , con cada r_i definida en el intervalo $0 < r_i < 1$. Aplicando la convención notacional de la forma matemática del teorema del límite central, así como nuestros conocimientos previos de la distribución uniforme, encontramos que: $\theta = \frac{1}{2}$, $\sigma = \frac{1}{\sqrt{12}}$

$$z = \frac{\sum_{i=1}^K r_i - K/2}{\sqrt{K/12}} \quad (24)$$

Si ahora igualamos las ecuaciones (6) y (8), obtenemos la ecuación que nos proporciona una formula simple para generar valores de variable aleatoria normalmente distribuidos, cuya media sea igual a μ_x y variancia σ_x^2 . Para generar un solo valor de x bastara con sumar K números aleatorios definidos en el intervalo de 0 a 1. Realizando las sustituciones de los valores correspondientes, se obtiene un determinado valor particular de x. Dicho procedimiento puede repetirse tantas veces como valores de variable aleatoria normalmente distribuidos se requieran.

$$x = \sigma_x \left(\frac{12}{K}\right)^{\frac{1}{2}} \left(\sum_{i=1}^K r_i - \frac{K}{2}\right) + \mu_x \quad (25)$$

Determinar el valor de K: El valor de K que debe aplicarse a la formula usualmente queda determinado al establecer las condiciones de balance entre la eficiencia del cómputo y la precisión. Pero al tener en cuenta al teorema del limite central, es deseable que K corresponda a un número muy grande.

En la práctica de simulación se recomienda una $K = 10$, como el menor valor deseable. Sin embargo, con $K = 12$ se logra una cierta ventaja computacional, ya que en la ecuación (9) se puede evitar una multiplicación constante.

Transformada Inversa: La transformada inversa de una distribución normal es un método utilizado para generar valores aleatorios de una distribución normal estándar ($\mu = 0$ y $\sigma^2 = 1$) a partir de números aleatorios generados a partir de una distribución uniforme en el intervalo $[0,1]$.

El método se basa en la función de distribución acumulativa inversa (también conocida como la función cuantil) de la distribución normal estándar. Esta función toma un valor de probabilidad acumulada como entrada (es decir, un número entre 0 y 1), y devuelve el valor correspondiente en la distribución normal estándar que tiene esa probabilidad acumulada.

Para generar valores aleatorios de una distribución normal estándar usando el método de la transformada inversa, se sigue los siguientes pasos:

1. Genera un número aleatorio U a partir de una distribución uniforme en el intervalo $[0,1]$.
2. Calcula el valor correspondiente de la función de distribución acumulativa inversa de la distribución normal estándar utilizando U. Este valor será un número aleatorio generado a partir de una distribución normal estándar.
3. Repite los pasos 1 y 2 para generar tantos números aleatorios de una distribución normal estándar como sea necesario.

El método de la transformada inversa es un método eficaz para generar valores aleatorios de una distribución normal estándar, ya que es fácil de implementar y no requiere una gran cantidad de cálculos complejos.

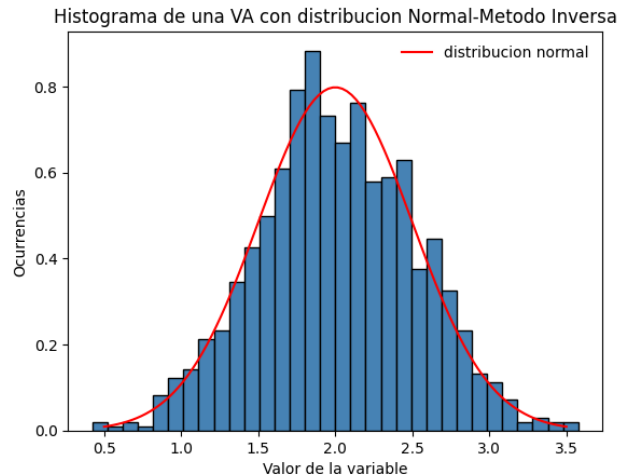


Figure 5: Histograma Distribución Normal - Inversa

```
1 def NormalInversa(numeros, mu, sigma) -> list:
2     return [mu + sqrt(2) * sigma * erfinv(2 * r - 1) for r in numeros]
```

Listing 3: Código en Python del Metodo de la inversa para la Distribución Normal

Algoritmo del Metodo del Rechazo: El método del rechazo es un método para generar valores aleatorios de una distribución normal (con μ y σ^2) a partir de valores aleatorios generados a partir de una distribución uniforme en un intervalo determinado.

El método se basa en la propiedad de que cualquier función de densidad de probabilidad puede ser acotada superiormente por una función de densidad de probabilidad más simple. En el caso de la distribución normal, una función de densidad de probabilidad que se puede utilizar para acotarla es la función de densidad de probabilidad de una distribución uniforme.

Para generar valores aleatorios de una distribución normal utilizando el método del rechazo, se sigue los siguientes pasos:

1. Selecciona una función de densidad de probabilidad simple que pueda acotar la distribución normal en el intervalo de interés. En este caso, la función de densidad de probabilidad de una distribución uniforme es adecuada para este propósito.
2. Generar un valor aleatorio X a partir de la distribución uniforme en el intervalo de interés.
3. Generar un valor aleatorio Y a partir de una distribución uniforme en el intervalo $[0, f(X)]$, donde $f(X)$ es la función de densidad de probabilidad de la distribución normal en el valor X .
4. Si Y es menor o igual que la función de densidad de probabilidad de la distribución normal en el valor X , entonces el valor generado de X es aceptado. De lo contrario, se vuelve al paso 2 y se genera otro valor de X .

Este proceso se repite hasta que se han generado suficientes valores aleatorios de la distribución normal.

El método del rechazo es una técnica útil para generar valores aleatorios de una distribución normal, especialmente cuando se trata de distribuciones no estándar o de alta dimensionalidad.

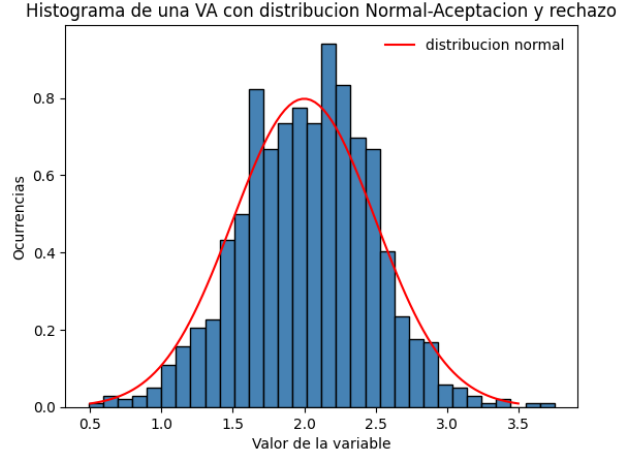


Figure 6: Histograma Distribución Normal - Rechazo

```

1 def NormalR(numeros, mu, des) -> list:
2     a, b = mu - (10 * des), mu + (10 * des)
3     M = 1 / (des * sqrt(2 * pi))
4     num = []
5     for U in numeros:
6         while True:
7             V = np.random.uniform(0, 1)
8             T = a + ((b - a) * V)
9             if M * U <= exp(-((T - mu) ** 2) / (2 * des ** 2)):
10                 num.append(T)
11                 break
12     return num
    
```

Listing 4: Código en Python del Metodo de la inversa para la Distribución Normal

2.2.4 Distribución Binomial

Las variables aleatorias definidas por el numero de eventos exitosos en una sucesión de n ensayos independientes de Bernoulli, para los cuales la probabilidad de éxito es p en cada ensayo, siguen una distribución binomial. Este modelo estocástico también se puede aplicar al proceso de muestreo aleatorio con reemplazo, cuando los elementos muestreados tienen solo dos tipos de atributos. Esto quiere decir, si ocurre o no un determinado suceso. El diseño de una muestra aleatoria de n elementos es análoga a n ensayos de Bernoulli, en los que x es un valor binomial que esta denotando al numero de elementos de una muestra de tamaño n con atributos idénticos. Es esta la analogía que sitúa la distribución binomial como uno de los modelos mas importantes en las áreas del muestreo estadístico y del control de calidad.

La distribución binomial proporciona la probabilidad de que un evento o acontecimiento tenga lugar x veces en un conjunto de n ensayos, donde la probabilidad para la distribución binomial se puede expresar de la siguiente manera:

$$f(x) = \binom{n}{x} p^x q^{n-x} \quad (26)$$

Donde x se toma como un entero definido en el intervalo finito de $0, 1, 2, \dots, n$, y al que se le asocia el valor $q = (1 - p)$. El valor esperado y la variancia de la variable binomial X son:

$$EX = np \quad VX = npq$$

La segunda expresión implica que la variancia de las variables binomiales siempre tiene un valor menor al de la media. Aun más, nótese como se define en la ecuación (26) la distribución de $(n - x)$ con un valor esperado correspondiente de np .

Cuando se conoce la media y la variancia, resulta inmediata la determinación de p y de n , las cuales pueden calcularse como sigue:

$$p = \frac{(EX - VX)}{EX} \quad (27)$$

$$n = \frac{EX^a}{(EX - VX)} \quad (28)$$

La distribución normal proporciona, cuando n es muy grande, una buena aproximación para la distribución binomial. Puesto que con la distribución normal resulta posible manipular valores negativos, a fin de hacer un buen uso de tal distribución la probabilidad de registrar observaciones negativas deberá ser despreciablemente pequeña.

En la practica esto significa que el valor esperado deberá ser por lo menos tres veces mayor que la desviación estándar, o sea:

$$np \geq 3(npq)^{\frac{1}{2}} \Rightarrow n \geq \frac{9q}{p} \quad (29)$$

Los valores de variable aleatoria con distribución binomial se pueden generar de muy diversos modos, aunque uno de los métodos mas simples, que en el caso de que el valor de n sea moderado resulta uno de los métodos mas eficientes, es el basado en la reproducción de ensayos de Benoulli, siguiendo el método de rechazos. Este método empieza con valores conocidos de p y n y consiste en generar n números aleatorios después de fijar x_0 igual a cero. Para cada numero aleatorio $r_i (1 \leq i \leq n)$ se efectúa una prueba y la variable x_i se incrementa de acuerdo con el siguiente criterio:

$$\begin{cases} x_i = x_{i-1} + 1 & , si \quad r_i \leq p \\ x_i = x_{i-1} & , si \quad r_i > p \end{cases} \quad (30)$$

después de haberse generado n números aleatorios, el valor de x_n será igual al valor de la variable aleatoria con distribución binomial X . Este procedimiento se puede repetir tantas veces como valores binomiales se requieran.

Algoritmo del Metodo del Rechazo: El método de rechazo es una técnica que se utiliza para simular valores aleatorios de una distribución de probabilidad utilizando una distribución más simple. En el caso de la distribución binomial, se puede utilizar una distribución uniforme como distribución más simple para simular valores de la distribución binomial. La idea básica es que la distribución uniforme se elige para simular valores aleatorios que se encuentran en un rango limitado (en este caso, el rango es de 0 a 1). Luego, se encuentra una constante de proporcionalidad c , que se utiliza para ajustar la distribución uniforme para que cubra la distribución binomial en ese rango.

Una vez que se ha encontrado c , se generan dos números aleatorios, u y v , de la distribución uniforme. Si v es menor o igual a c veces la función de densidad de la distribución binomial para el valor correspondiente de u , entonces se acepta el valor de x como una realización de la distribución binomial. Si no, se rechaza y se comienza de nuevo.

Este proceso se repite hasta que se haya generado el número deseado de valores aleatorios.

Es importante tener en cuenta que la eficacia de este método depende de la elección de la constante c , que debe ser elegida de manera cuidadosa para minimizar el número de valores rechazados y maximizar la eficiencia de la simulación.

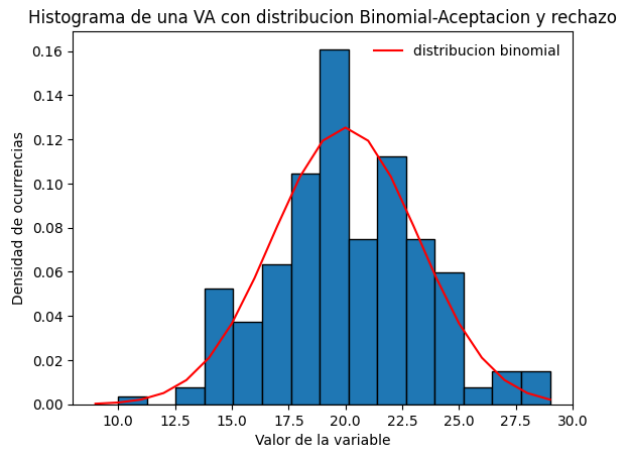


Figure 7: Histograma Distribución Binomial - Rechazo

Test

2.2.5 Distribución Pascal

La distribución de Pascal (también conocida como distribución binomial negativa) es una distribución de probabilidad discreta que describe el número de ensayos independientes de Bernoulli necesarios para obtener r éxitos, donde se

detiene el proceso de ensayo cuando se alcanza el r -ésimo éxito. Esta distribución es muy útil para modelar situaciones en las que se desea conocer el número de pruebas necesarias para alcanzar una meta específica.

Cuando los procesos de ensayos de Bernoulli se repiten hasta lograr que ocurran k éxitos ($k > 1$), la variable aleatoria que caracteriza al número de fallas tendrá una distribución de Pascal. Por consiguiente, los valores de variables aleatorias con distribución de Pascal coinciden esencialmente con la suma de k valores de variable aleatoria con distribución geométrica, en este caso, k es un número entero y la distribución recibe el nombre de distribución de Pascal. En consecuencia, la distribución geométrica constituye un caso particular de la distribución de Pascal, especificada para $k = 1$.

La función de distribución de probabilidad para una distribución de Pascal está dada por:

$$F(k) = \binom{k+x-1}{x} p^k q^x \quad x = 0, 1, 2, \dots \quad (31)$$

donde k es el número total de éxitos en una sucesión de $k+x$ ensayos, con x el número de fallas que ocurren antes de obtener k éxitos. El valor esperado y la variancia de X se representa con:

$$EX = \frac{kq}{p} \quad (32)$$

$$VX = \frac{kq}{p^2} \quad (33)$$

Tanto la distribución geométrica como la de Pascal se caracterizan por una sobredispersión, esto es, $VX > EX$.

Para una media y una variancia dadas, se pueden determinar los parámetros p y k de la siguiente manera:

$$p = \frac{EX}{VX} \quad (34)$$

$$k = \frac{EX^2}{VX - EX} \quad (35)$$

Sin embargo, puede suceder que el proceso de simulación se complique considerablemente cuando resulte que en la ecuación de k el valor que se obtenga al efectuar el cómputo de k no sea un entero.

Cuando k es un entero, los valores de la variable aleatoria con distribución de Pascal se pueden generar con sólo considerar la suma de k valores con distribución geométrica. En consecuencia

$$x = \frac{\sum_{i=1}^k \log(r_i)}{\log(q)} = \frac{\log(\prod_{i=1}^k r_i)}{\log(q)} \quad (36)$$

viene a ser un valor de variable aleatoria con distribución de Pascal, una vez que su magnitud se redondea con respecto al menor entero más próximo al valor calculado.

Algoritmo del Metodo del Rechazo: El método del rechazo para la generación de números pseudoaleatorios con distribución binomial negativa utilizando la distribución uniforme como distribución auxiliar implica la generación de números aleatorios uniformes y su transformación. Luego, se compara el valor propuesto con un valor generado aleatoriamente a partir de una distribución uniforme y, si es aceptado, se utiliza como un número aleatorio con distribución binomial negativa. Si es rechazado, se generan nuevos números aleatorios uniformes y se repite el proceso hasta que se acepte un valor propuesto.

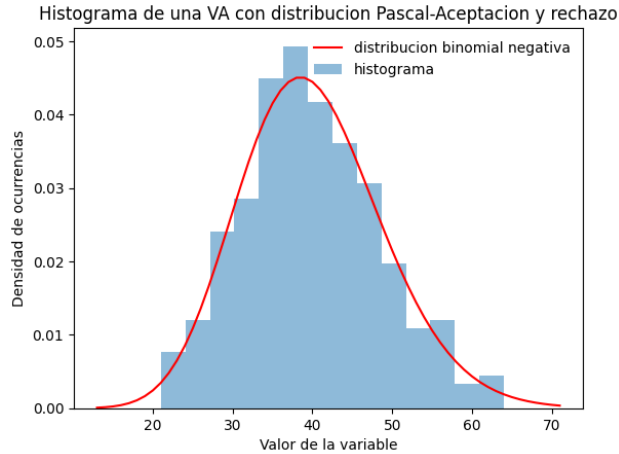


Figure 8: Histograma Distribución Pascal - Rechazo

2.2.6 Distribución Gamma

La distribución de probabilidad gamma es una distribución de probabilidad continua que se utiliza para modelar variables aleatorias positivas y asimétricas. Esta distribución es ampliamente utilizada en estadística y ciencias de la ingeniería para modelar tiempos de espera, tiempos de falla, tamaños de partículas, entre otras aplicaciones.

La distribución de probabilidad gamma tiene dos parámetros: el parámetro de forma (α) y el parámetro de escala (β). La función de densidad de probabilidad (PDF) de la distribución gamma se define como:

$$P(x) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x} \quad (37)$$

El parámetro α determina la forma de la distribución gamma y el parámetro β determina la escala de la distribución. A medida que α aumenta, la distribución gamma se vuelve más simétrica y se desplaza hacia la derecha. A medida que β aumenta, la distribución gamma se estrecha y se concentra alrededor de su valor esperado.

La distribución es limitada en su dominio, aceptando solo valores positivos de x .

Algoritmo del Método del Rechazo y Gráfica Se creó una simulación en python de esta distribución por método de rechazo. Para generar una muestra que conforme con cierto grado de aceptación la distribución gamma se usó la función de densidad de probabilidad gamma como límite superior, y se probaron números aleatorios a partir de una distribución exponencial hasta encontrar un número aceptado por el algoritmo, esto luego es repetido una cantidad de veces decidida previo a iniciar la simulación.

Esta muestra es después graficada en un histograma y comparada con la función de densidad que la limita.

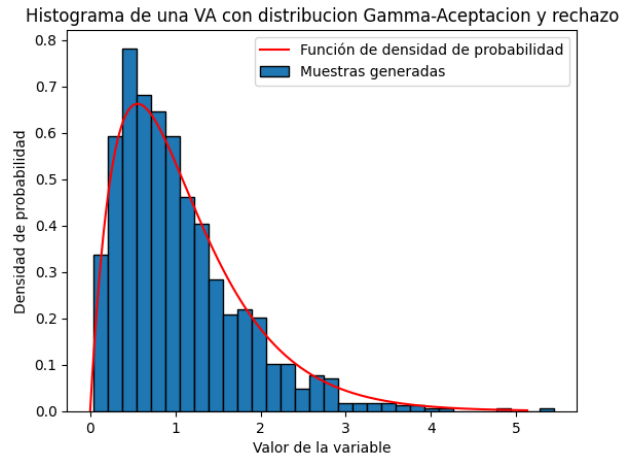


Figure 9: Histograma Distribución Gamma - Rechazo

2.2.7 Distribución Empírica Discreta

La Distribución Empírica Discreta es una función finita basada en observaciones empíricas, donde se asigna a cada valor una probabilidad, tal que la suma de todas las probabilidades de como resultado 1.

La distribución empírica discreta es una herramienta útil para describir la distribución de datos discretos y puede utilizarse para realizar inferencias estadísticas y tomar decisiones basadas en datos empíricos. Sin embargo, hay que tener en cuenta que esta distribución se basa en observaciones empíricas y puede ser sensible al tamaño y la calidad de la muestra utilizada.

Algoritmo del Método del Rechazo, Gráfica y Test Para la simulación codificada en Python del método para esta distribución, se toman n posibles sucesos y sus probabilidades, donde la suma de estas da como resultado 1 o 100%, y se programó el método de rechazo en base a una distribución uniforme y usando de límite superior la gráfica formada por las probabilidades de cada suceso.

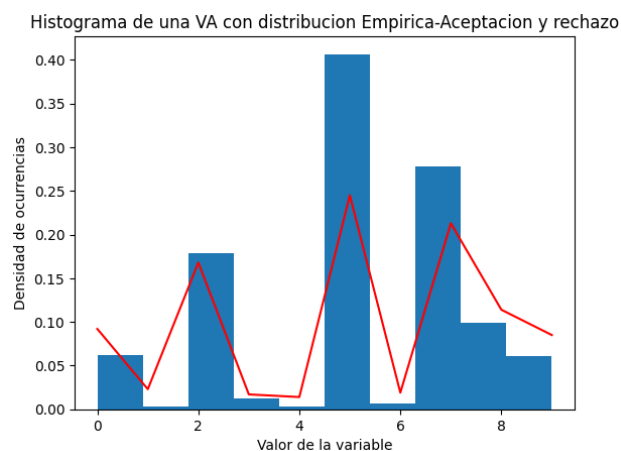


Figure 10: Histograma Distribución Empírica Discreta - Rechazo

2.2.8 Distribución Poisson

La distribución de Poisson es una distribución de probabilidad discreta, o sea, se toma un número de valor finito, esta distribución se aplica a las ocurrencias de algún evento durante un periodo determinado. Es decir, es una distribución de probabilidad discreta en la que solo es necesario conocer los eventos y cuál es su frecuencia media de ocurrencia para poder conocer la probabilidad de que ocurran.

La fórmula para calcular las probabilidades que provienen de un proceso de Poisson es:

$$P(X) = \frac{\lambda^x e^{-\lambda}}{x!} \quad (38)$$

Los parámetros de la función son:

- $P(X)$ es la probabilidad de X aciertos
- λ es el número esperado de aciertos basado en datos históricos
- e es el logaritmo natural aproximadamente igual a 2,718
- x es el número de aciertos por unidad, normalmente por unidad de tiempo.

Dependiendo la bibliografía, el numero esperado de asiertos basados en datos historicos tambien puede indicarse con la letra griega μ .

La distribución de Poisson es una distribución binomial que está limitada al solo depender de un parámetro, y este es el número esperado de eventos que ocurrirán en un intervalo fijado, o sea, la frecuencia de los eventos.

Si en una función binomial sus parámetros tienden a infinito y a cero, la distribución límite obtenida es la de Poisson.

Una propiedad importante de la distribución de Poisson es que, la suma de “n” variables de Poisson independientes tendrán como resultado también una variable de Poisson, siendo su parámetro la suma del valor de los parámetros originales.

Para utilizar la distribución de Poisson, deben cumplirse ciertos supuestos. Estos son: la probabilidad de un éxito, λ , no cambia dentro del intervalo, no puede haber éxitos simultáneos dentro del intervalo y, por último, que la probabilidad de un éxito entre intervalos es independiente, el mismo supuesto de la distribución binomial.

Algoritmo del Método del Rechazo, Gráfica y Test Para poder realizar la simulación en Python se utilizó la fórmula de la distribución de Poisson mencionada anteriormente, y luego se programó el método del rechazo. Donde, en el caso de la distribución de Poisson se generó a partir de una distribución uniforme. El objetivo era generar números aleatorios que siguieran una distribución de Poisson con parámetro λ , utilizando una distribución uniforme como generador de números aleatorios. En este caso, la función de densidad de Poisson se utilizó como límite superior, y se generaron números aleatorios uniformes hasta que se encontró uno que estuviera por debajo de la función de densidad de Poisson. Este proceso se repitió hasta que se generó la cantidad deseada de números aleatorios de Poisson. Luego, se graficó la distribución de Poisson generada junto con la distribución de Poisson teórica para comparar los resultados.

Los valores utilizados para las variables λ y X fueron extraídos de ejemplos encontrados en internet.

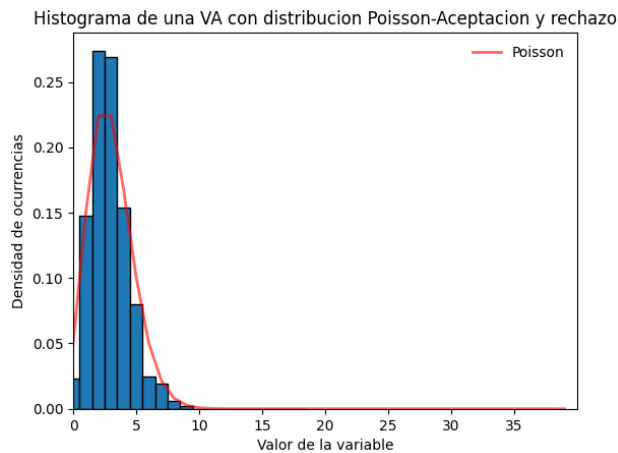


Figure 11: Histograma Distribución Poisson - Rechazo

2.2.9 Distribución Hipergeométrica

La distribución hipergeométrica es una distribución discreta que modela el número de eventos en una muestra de tamaño fijo cuando se conoce el número total de elementos en la población de la cual proviene la muestra. Cada elemento

de la muestra puede tener dos resultados posibles, que se refieren a si el elemento pertenece o no a la categoría A en la población original de la que se tomó la muestra. Las muestras no tienen reemplazo, por lo que cada elemento de la muestra es diferente. Cuando se elige un elemento de la población, no se puede volver a elegir. Por lo tanto, la probabilidad de que un elemento sea seleccionado aumenta con cada ensayo, presuponiendo que aún no haya sido seleccionado.

La distribución hipergeométrica puede ser utilizada para muestras obtenidas de poblaciones relativamente pequeñas, como ya ha sido mencionado antes, sin reemplazo.

La formula de la distribucion hipergeometrica es la siguiente

$$P(X = x) = \frac{\binom{N}{x} \binom{M-N}{n-x}}{\binom{M}{n}} \quad (39)$$

Los parametros de la formula son:

- $P(X = x)$ es la probabilidad de obtener x elementos del evento en la muestra de tamaño n sin reemplazo.
- M es el tamaño de la población
- N es el conteo de elementos del evento en la población
- n es el tamaño de la muestra.

Para dar una explicacion un poco mas detallada, a continuacion se describe cada bloque que presenta la formula

- $\binom{N}{x}$ es el número de maneras diferentes en que se pueden seleccionar x elementos con un evento en la población de tamaño N .
- $\binom{M-N}{n-x}$ es el número de maneras diferentes en que se pueden seleccionar $n - x$ elementos sin evento en la población restante de tamaño $M - N$.
- $\binom{M}{n}$ es el número total de muestras posibles de tamaño n que se pueden seleccionar sin reemplazo de la población de tamaño M .

Algoritmo del Método del Rechazo y Gráfica En caso de la simulacion realizada en Python, se utilizo esa misma formula con los mismos parametros, aunque dependiendo la bibliografia, los parametros pueden tener otros nombres. Cabe destacar que la fórmula es la misma, independientemente de los nombres que se le den a los parámetros. Lo que importa es entender qué representa cada uno de ellos en el contexto del problema que se está modelando.

Para evaluar generacion de numeros pseudoaleatorios con la distribucion hipergeometrica se utilizo, en este caso, el método del rechazo.

Para calcular el metodo del rechazo en Python se calculo la densidad máxima de la distribución hipergeométrica y así poder establecer un criterio en el método del rechazo. En la simulacion realizada se generaron pares de números aleatorios U y V , siendo estos distribuidos uniformemente en $[0, 1]$ Luego, se calcula una variable aleatoria X mediante la distribución uniforme discreta, ya que la utilizamos como distribucion de prueba. Si X cumple con con la condicion, se acepta como una realización de distribución hipergeométrica que se quiere generar. Si no se cumple la condición, se rechaza X y se repite el proceso para obtener un nuevo valor de X .

Los valores utilizados para las variables M , N y n fueron extraidos de ejemplos encontrados en internet.

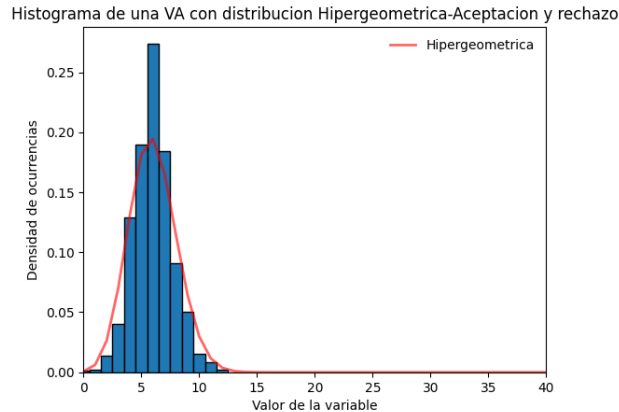


Figure 12: Histograma Distribución Hipergeométrica - Rechazo

3 Conclusión

3.1 Análisis de las gráficas generadas.

Al utilizar los métodos de la Transformada Inversa y el de Rechazo-Aceptación, los números aleatorios generados con distribución uniforme, se convierten en números aleatorios con distintas distribuciones. Esto se puede observar en todas las gráficas de las variables aleatorias generadas, ya que en los distintos métodos los histogramas se asejeman, aproximadamente, a la gráfica de la distribución densidad. Es decir que cuando necesitemos de distribuciones de probabilidad, que no sea la uniforme, podemos hacer uso de cualquiera de los métodos, según corresponda, para obtener nuestra distribución deseada.

4 Bibliografía

<https://support.minitab.com/es-mx/minitab/21/help-and-how-to/probability-distributions-random-data-and-resampling-analyses/supporting-topics/distributions/hypergeometric-distribution/>
<https://www.sdelosol.com/blog/tendencias/distribucion-de-poisson/> <https://openstax.org/books/introducción-estadística-empresarial/pages/4-4-distribucion-de-poisson> <https://bookdown.org/content/944ffa0f-050e-47cb-afaa-4dff15a9ed00/ProEstoc.html> Olivares, J. (2007). Generación de valores de las variables aleatorias.