TP 2.2 - GENERADORES DE NÚMEROS PSEUDOALEATORIOS DE DISTINTAS DISTRIBUCIONES DE PROBABILIDAD

Abud Santiago Elias
Legajo 47015
sabudvicco@gmail.com

Buchhamer Ariel
Legajo 46217
arielbuchhamer1@outlook.com

Castellano Marcelo
Legajo 39028
marce.geek22@gmail.com

Navarro Franco
Legajo 46387
franconavarro1889@gmail.com

20 de mayo de 2022

ABSTRACT

El siguiente documento tiene por objetivo realizar un estudio sobre distintas distribuciones de probabilidad con sus respectivas características. De cada una se desarrollarán sus conceptos, se utilizará la tecnología para mostrar su funcionamiento y finalmente testear la generación de valores de las mismas.

***K*eywords** Simulación · Números Pseudoaleatorios · Distribución · Probabilidad · Discretas · Continuas

1. Introducción

Se busca poder observar la generación de los distintos tipos de distribuciones a través de la generación de números pseudoaleatorios. Los tipos de distribuciones a desarrollar y analizar son tanto continuas como discretas, por lo que usaremos generadores específicos de cada uno. Luego de realizar las simulaciones, se procederá a aplicar el test de bondad Chi-Cuadrado a algunos de ellos para conocer su resultado.

2. Números Pseudoaleatorios

Un número pseudoaleatorio no es más que el valor de una variable aleatoria x que tiene una distribución de probabilidad uniforme definida en el intervalo (0, 1).

Los números pseudoaleatorios constituyen una parte realmente importante en la simulación de procesos estocásticos y generalmente se usan para generar el comportamiento de variables aleatorias, tanto continuas como discretas. Debido a que no es posible generar números realmente aleatorios, los consideramos como pseudoaleatorios, generados por medio de algoritmos determinísticos que requieren parámetros de arranque.

Podemos asegurar con altos niveles de confiabilidad que el conjunto de números que utilizaremos en una simulación se comportan de manera muy similar a un conjunto de números totalmente aleatorios; por ello es que se les denomina números pseudoaleatorios.

3. Test de Bondad Ji Cuadrado

El estadístico ji-cuadrado (o chi cuadrado), que tiene distribución de probabilidad del mismo nombre, sirve para someter a prueba hipótesis referidas a distribuciones de frecuencias. En términos generales, esta prueba contrasta frecuencias observadas con las frecuencias esperadas de acuerdo con la hipótesis nula. En este artículo se describe el uso del estadístico ji-cuadrado para probar la asociación entre dos variables utilizando una situación hipotética y datos simulados. Luego se describe su uso para evaluar cuán buena puede resultar una distribución teórica, cuando pretende representar la distribución real de los datos de una muestra determinada. A esto se le llama evaluar la bondad de un ajuste. Probar la bondad de un ajuste es ver en qué medida se ajustan los datos observados a una distribución teórica o esperada. Para esto, se utiliza una segunda situación hipotética y datos simulados.

En el presente documento utilizaremos estre test en algunas distribuciones de probabilidades, los cuales dividiremos en 10 intervalos, agrupando los valores y determinando las frecuencias observadas y esperadas. Como manera de verificación se realizará la sumatoria de todos los valores observados y la sumatoria de todos los valores esperados, ya que cada uno de estos valores debería de poseer un valor de 1000, debido a que se generarán 1000 números pseudoaleatorios. Para la visualizacion del valor de chi-cuadrado esperado utilizaremos la tabla con 9 grados de libertad y un 95 por ciento de confianza y la formula para obtener el valor de chi-cuadrado observado sera:

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

Donde e es el valor esperado y o es el valor observado en cada intervalo.

4. Distribuciones de Probabilidad

En teoría de la probabilidad y estadística, la distribución de probabilidad de una variable aleatoria es una función que asigna a cada suceso definido sobre la variable aleatoria, la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos el rango de valores de la variable aleatoria.

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

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

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

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

A continuación se desarrollarán distintos generadores de valores de variables aleatorias a partir de las distribuciones probabilidad, las cuales están clasificadas según el tipo de variable.

4.1. Distribuciones Continuas de Probabilidad

Una distribución continua describe las probabilidades de los posibles valores de una variable aleatoria continua. Una variable aleatoria continua es una variable aleatoria con un conjunto de valores posibles (conocido como el rango) que es infinito y no se puede contar.

4.1.1. Distribución Uniforme

Se caracteriza por ser constante, en un intervalo entre a y b sin contener al cero. Surge del estudio de las características de los errores por redondeo al registrar un conjunto de medidas sujetas a cierto nivel de precisión. Sobresale respecto a las técnicas de simulación por su simplicidad y por el hecho de que se puede emplear para simular variables aleatorias a partir de case cualquier tipo de distribución de probabilidad.

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

$$fx = \begin{cases} \frac{1}{b-a} & a < x < b \\ \\ 0, & no \quad esta \\ en & (a, b) \end{cases} \tag{2}$$

En esta ecuación X es una variable aleatoria definida en (a, b) El valor esperado y la varianza de una variable aleatoria uniformemente distribuida, se puede representar por:

$$EX = \frac{b+a}{2} \tag{3}$$

$$VX = \frac{(b-a)^2}{12} \tag{4}$$

4.1.1.1. Método de la transformación inversa

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

$$F(x) = \int_0^a \frac{1}{b-a} dt = \frac{x-a}{b-a} \quad 0 \leq F(x) \leq 1. \tag{5}$$

Lo cual podemos deducir como su función inversa a:

$$x = a + (b-a)r \quad 0 \leq r \leq 1. \tag{6}$$

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

4.1.1.2. Código en Python

```
    def uniforme(a, b, rep: int = 1):
      x = []
      for i in range(rep):
          x.append(a + (b - a) * (round(random(), 4)))
      return x
```

Se ingresa como parámetro a y b que son valores que expresan los límites de las distribuciones uniformes.

4.1.1.3. Análisis

En la gráfica se puede observar la figura formada a partir de la distribución uniforme obtenida de la generación de números pseudoaleatorios y del método de la transformación inversa. En el eje de las abscisas se encuentran los valores valores que rondan el intervalo (a,b) y en el eje de las ordenadas la frecuencia absoluta en ese intervalo.

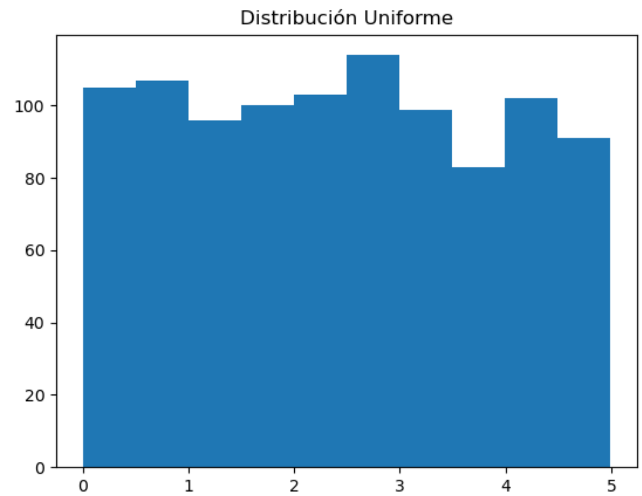

Figura 1: Frecuencias absolutas de valores generados dentro del rango ingresado.

Valor Observado	Valor Esperado	Valor Ji-Cuadrado
45	100	30.25
38	100	68.69
34	100	112.25
39	100	149.46
40	100	185.46
40	100	221.46
39	100	258.67
50	100	283.67
44	100	315.03000000000003
45	100	345.28000000000003

Cuadro 1: Valores observados,esperados y chi-cuadradado de una distribución Uniforme

4.1.1.4. Test

Valor de X2 obtenido = 345.28000000000003

Con 9 grados de libertad y con un 95 porciento de confianza se obtiene un valor de Chi cuadrado de 16.92.

Como X2 es mayor a 16.92, la distribución no pasó el test.

4.1.2. Distribución Exponencial

En caso de que la ocurrencia de un evento en un intervalo corto, y si la ocurrencia de tal evento es, estadísticamente independiente respecto a la ocurrencia de otros eventos, entonces el intervalo de tiempo entre ocurrencias de eventos de este tipo estará distribuido en forma exponencial. Se dice que una variable aleatoria X tiene una distribución exponencial, si se puede definir a su función de densidad como:

$$f(x) = ae^{-ax} \tag{7}$$

con a >0 y x >= 0. La función de distribución acumulativa de X está dada por:

$$F(x) = \int_0^a ae^{-at}dt = 1 - e^{-ax}, \tag{8}$$

y la media junto con la variancia de X se pueden expresar como

$$EX = \int_0^\infty xae^{-ax}dx = \frac{1}{a} \tag{9}$$

$$VX = \int_0^\infty (x - \frac{1}{a})^2 = \frac{1}{a^2} = (EX)^2 \quad (10)$$

4.1.2.1. Método de la transformación inversa

Existen muchas maneras para lograr la generación de valores de variables aleatorias exponenciales. Puesto que la distribución acumulativa existe explícitamente (vista en la ecuación (8)),la técnica de la transformación inversa nos permite desarrollar métodos directos para dicha generación. Debido a la simetría que existe entre la distribución uniforme sigue que la intercambiabilidad de F(x) y 1 - F(x). Por lo tanto:

$$r = e^{-ax} \quad (11)$$

y consecuentemente

$$x = -\frac{1}{a} logr = -EXlogr \quad (12)$$

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

4.1.2.2. Código en Python

```
    def exponencial(ex, rep: int = 1):
      x = []
      for i in range(rep):
          x.append(-ex * math.log(round(random(), 4)))
      return x
```

El parámetro ex que ingresa es la media de la distribución exponencial y cuyo valor es generado por el generador Mersenne twister propio de Python.

4.1.2.3. Análisis

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

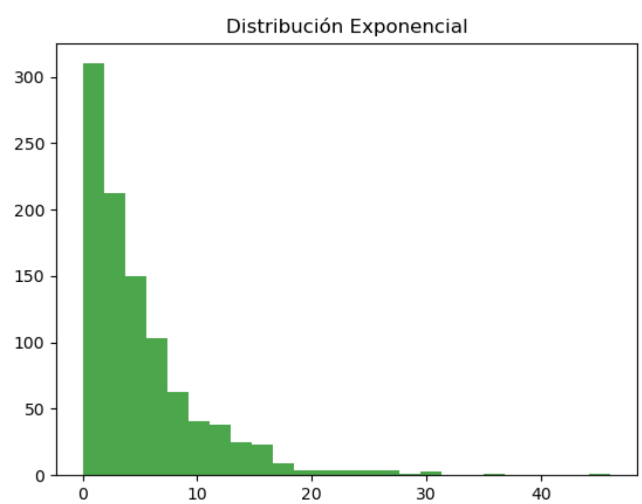

Figura 2: Frecuencias absolutas de valores generados.

Valor Observado	Valor Esperado	Valor Ji-Cuadrado
50	58.23546641575128	0.0858121416985893
60	54.84409686709124	0.5705194964456324
45	51.65022530588559	1.4267693260858056
67	48.64235034471853	8.354956358122323
48	45.80964038483559	8.459687048069659
32	43.14189461068685	11.337210546652889
32	40.62950625597428	13.170075011033358
26	38.26342800891591	17.10050347381562
50	36.035139432151084	22.512369051487404
584	582.7482523739898	22.515057814796155

Cuadro 2: Valores observados,esperados y chi-cuadradado de una distribución Exponencial

4.1.2.4. Test

Valor de X2 obtenido = 22.515057814796155

Se poseen 10 intervalos, por lo tantos tendremos 9 grados de libertad.

Con 9 grados de libertad y con un 95 porciento de confianza se obtiene un valor de Chi cuadrado de 16.92. Como X2 es mayor, no pasó el test.

4.1.3. Distribucion Gamma

Si un determinado proceso consiste de k eventos sucesivos y si el total del tiempo transcurrido para dicho proceso se puede considerar igual a la suma de k valores independientes de la variable aleatoria con distribución exponencial, cada uno de los cuales tiene un parámetro definido a la distribución de esta suma coincidirá con una distribución gamma con parámetros a y k. La función gamma está descrita mediante la siguiente función de densidad:

$$f(x) = \frac{a^k x^{(k-1)} e^{-ax}}{(k-1)!} \tag{13}$$

donde a >0, k >9 y x se considera no negativo. Respecto a la media y la variancia de esta distribución, sus corrspondientes expresiones están formuladas como sigue:

$$EX = \frac{k}{a} \tag{14}$$

$$VX = \frac{k}{a^x} \tag{15}$$

4.1.3.1. Código en Python

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

Los parámetros que se ingresan son k (cantidad de eventos sucesivos) y a (lambda).

4.1.3.2. Análisis

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

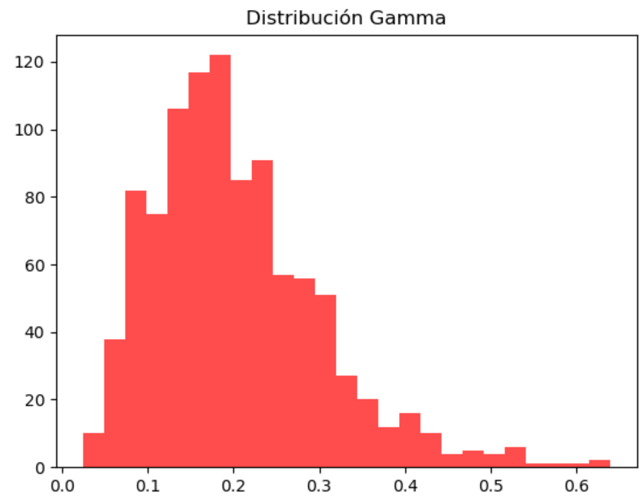

Figura 3: Frecuencias absolutas de valores generados.

4.1.4. Distribución Normal

La distribución normal basa su utilidad en el teorema del límite central. Este postula que, la distribución de probabilidad de la sima de N valores de variable aleatoria x independientes pero idénticamente distribuidos, con medias respectivas u y variancias se aproximan asintóticamente a una distribución normal, a medida que N se hace muy grande.

En consecuencia, el teorema del límite central permite el empleo de distribuciones normales para representar medidas globales operadas sobre los efectos de causas (errores) aditivas distribuidas en forma independiente sin importar la distribución de probabilidad a que obedezcan las mediciones de causas individuales.

Si la variable aleatoria X tiene una función de densidad f(x) dada como:

$$f(x) = \frac{1}{_x\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2^2}} \tag{16}$$

El valor esperado y la variancia de la distribución normal están dados por:

$$EX = \mu_x \tag{17}$$

$$EX = \sigma_x^2 \tag{18}$$

4.1.4.1. Método de la transformación inversa

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

$$\theta = \frac{a+b}{2} = \frac{0+1}{2} = \frac{1}{2}, \tag{19}$$

$$\sigma = \frac{b-a}{\sqrt{12}} = \frac{1}{\sqrt{12}}, \tag{20}$$

$$z = \frac{\sum_{i=2}^{K} r_i - K/2}{\sqrt{K/12}} \tag{21}$$

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

$$x = \sigma_x (\frac{12}{K})^{1/2} (\sum_{i=1}^{K} r_i - \frac{K}{2}) + \mu_x \tag{22}$$

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

4.1.4.2. Código en Python

```
    def normal(ex, stdx, rep: int = 1):
      x = []
      for i in range(rep):
          sm = 0
          for j in range(12):
              sm += random()
          x.append(stdx * (sm - 6) + ex)
      return x
```

Se puede observar que se ingresa como parámetro la media (ex) y la desviación estándar que es stdx.

4.1.4.3. Análisis

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

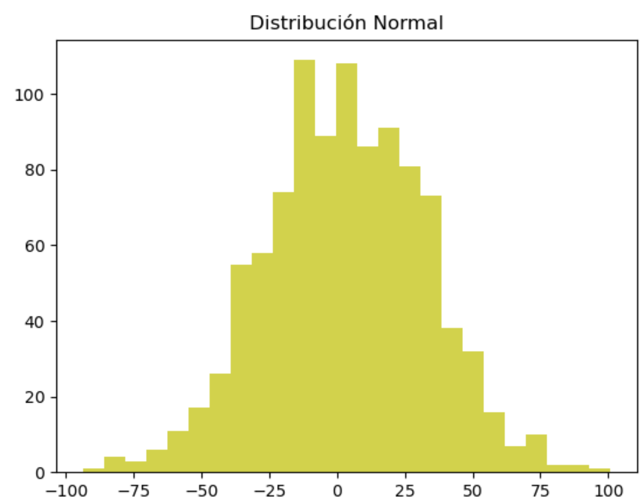

Figura 4: Frecuencias absolutas de valores generados.

4.1.4.4. Test

Valor de X2 obtenido = 9.304212174355968 Se poseen 10 intervalos, por lo tantos tendremos 9 grados de libertad. Con 9 grados de libertad y con un 95 porciento de confianza se obtiene un valor de Chi cuadrado de 16.92 El valor de chi cuadrado obtenido: 9.304212174355968 es menor que 16.92 ,por lo tanto Paso el test.

Valor Observado	Valor Esperado	Valor Ji-Cuadrado
5	2.2746694468500497	3.2652773501741432
14	14.110294410228015	3.266139476546129
47	56.34629805925615	4.816432759710728
147	144.9788279442314	4.844610229048467
244	240.5186147918511	4.895001517797601
248	257.37370714101644	5.236397639964036
198	177.6535039813194	7.566663528115143
73	79.077766308295	8.033789055589368
21	22.685503247816598	8.159019744429111
2	4.190697350629602	9.304212174355968

Cuadro 3: Valores observados,esperados y chi-cuadradado de una distribución Normal

4.2. Distribuciones Discretas de Probabilidad

Se encuentra definido un número muy significativo de distribuciones de probabilidad para variables aleatorias que solamente toman valores discretos, esto es, enteros no negativos. La distribución acumulativa de probabilidad para una variable aleatoria discreta X se define de manera muy similar a la de la ecuación

$$F(x) = P(X \leq x) = \sum_{x=0}^{x} f(x) \tag{23}$$

Donde f(x) es la frecuencia o función de probabilidad de X definida por valores enteros que:

$$F(x) = P(X = x) \tag{24}$$

Para x=0,1,2,...

Las distribuciones discretas de probabilidad son muy útiles cuando se las emplea como modelos estocásticos para ciertos procesos de conteo, ya sea sobre muestras finitas o no finitas, donde la presencia o ausencia de un atributo dicotómico está gobernada por el azar.

Las secciones siguientes contienen la descripción de técnicas para la generación de valores de variables estocásticas a partir de la mayoría de las distribuciones discretas de probabilidad más conocidas:

4.2.1. Distribución Pascal

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 binomial negativa. Por consiguiente, los valores de variables aleatorias con distribución binomial negativa 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.

La función de distribución de probabilidad para una distribución binomial negativa está dada por:

$$f(x) = (\frac{k+x-1}{x})p^k q^x \quad x = 0, 1, 2, ..., \tag{25}$$

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:

$$EX = \frac{kq}{p} \tag{26}$$

$$VX = \frac{kq}{p^2} \tag{27}$$

4.2.1.1. Código en Python

```
    def pascal(k, q, rep: int = 1):
      nx = []
      for i in range(rep):
          tr = 1
          for j in range(k):
              tr *= random()
          x = math.log(tr) // math.log(q)
          nx.append(x)
      return nx
```

Los parámetros ingresados son k que hace referencia a la cantidad de repeticiones hasta el éxito y q es la probabilidad de no tener éxito.

4.2.1.2. Análisis

En esta gráfica podemos observar la figura que se forma a partir de la distribución Pascal obtenida por la generación de números pseudoaleatorios. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones a partir de repeticiones de k éxitos con valor igual a 5 y con una probabilidad de exito de 0.6.

Además en el eje de las ordenadas se puede observar la frecuencia absoluta de los diferentes intervalos del histograma.

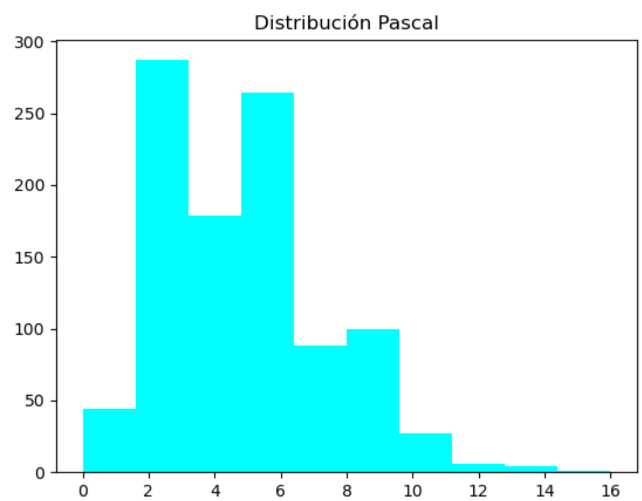

Figura 5: Frecuencias absolutas de valores generados.

4.2.2. Distribución Binomial

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 de éxito está dada por p. La función de probabilidad para la distribución binomial se puede expresar de la manera siguiente:

$$f(x) = \binom{n}{x} p^x q^{n-x} \tag{28}$$

donde x se toma como un entero definido en el intervalo finito 0,1,2,...n, y al que se le asocio el valor q = (1-p).

El valor esperado y la variancia de la variable binomial X son

$$EX = np \tag{29}$$

$$VX = npq \tag{30}$$

4.2.2.1. Código en Python

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

N representa la cantidad de ensayos independientes de Bernoulli y p representa la probabilidad de éxito.

4.2.2.2. Análisis

En esta gráfica podemos observar la figura formada a partir de la distribución Binomial obtenida por la generación de números pseudoaleatorios. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones a partir de "n.ensayos independientes de Bernoully para los cuales la probabilidad de éxito es de 0.4. En el eje de las ordenadas se encuentra la frecuencia absoluta de los diferentes intervalos del histograma.

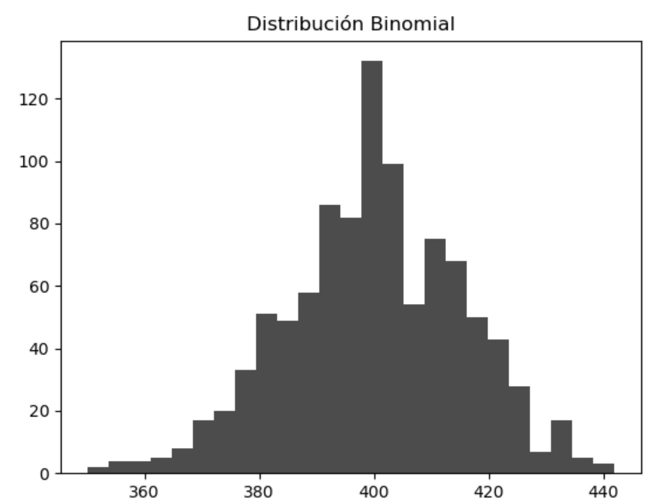

Figura 6: Frecuencias absolutas de valores generados.

4.2.2.3. Test

Valor de X2 obtenido = 5.584025744557111

Se poseen 10 intervalos, por lo tantos tendremos 9 grados de libertad.

Con 9 grados de libertad y con un 95 porciento de confianza se obtiene un valor de Chi cuadrado de 16.92

El valor de chi cuadrado obtenido: 5.584025744557111 es menor que 16.92 ,por lo tanto Paso el test.

Valor Observado	Valor Esperado	Valor Ji-Cuadrado
2	1.2183998806820666	0.5013942927964039
18	16.347797144268114	0.6683754617207881
99	98.40878879839349	0.6719272855890921
246	270.47446286358735	2.886551639102344
358	344.0787593417779	3.449797512910062
217	204.45808031184686	4.219147147926512
55	57.02444084702785	4.291017396581979
5	7.475187463117927	5.1106025919773
0	0.4601497332212823	5.570752325198582
0	0.013273419358528216	5.584025744557111

Cuadro 4: Valores observados,esperados y chi-cuadradado de una distribución Binomial

4.2.3. Distribución Hipergeométrica

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

$$f(x) = \frac{\binom{N_p}{x}\binom{N_q}{n-x}}{\binom{N}{n}} \tag{31}$$

con 0 =<x =<Np, y 0 =<n-x =<Nq. Donde x,n y N son enteros. El valor esperado y la variancia se caracterizan como sigue:

$$EX = np \tag{32}$$

$$VX = npq(\frac{N-n}{N-1}) \tag{33}$$

La generación de valores hipergeométricos involucra, substancialmente, la simulación de experimentos de muestreo sin reemplazo.

4.2.3.1. Código en Python

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

La población total es indicada con el parámetro tn, la muestra seleccionada con ns y un índice de proporción de 0.4.

4.2.3.2. Análisis

En esta gráfica podemos observar la figura formada a partir de la distribución Hipergeometrica obtenida a por la generación de números pseudoaleatorios. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones por medio de una población N (en nuestro caso N = 5000000), una muestra de dicha población de 500 y un índice de proporción de 0.4, mientras que en el eje de las ordenadas se observa la frecuencia absoluta de los diferentes intervalos del histograma.

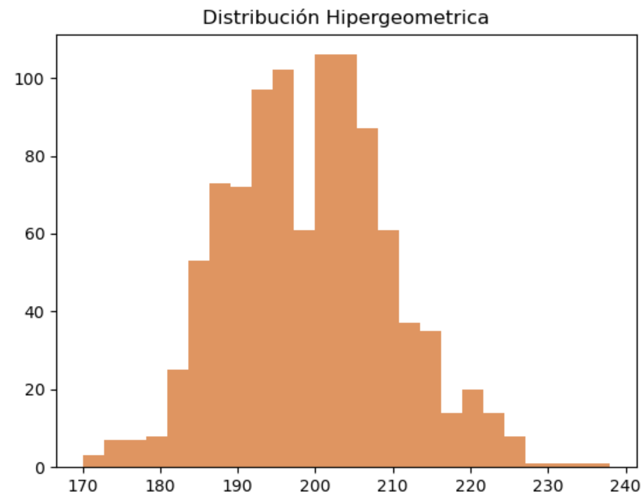

Figura 7: Frecuencias absolutas de valores generados.

4.2.4. Distribución Poisson

Si tomamos una serie de n ensayos independientes de Bernoulli, en cada uno de los cuales se tenga una probabilidad p muy pequeña relativa a la ocurrencia de un cierto evento, a medida que n tiende al infinito, la probabilidad de x ocurrencias está dada por la distribución de Poisson

$$f(x) = e^{-\delta}\frac{\delta^x}{x!} \tag{34}$$

con x = 0,1,2,... y $\delta > 0$.

Esto sucede siempre y cuando permitamos que p se aproxime a cero de manera que se satisfaga la relación $\delta = np$.

4.2.4.1. Código en Python

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

El parámetro que se ingresa a parte de las tiradas es el del valor de lambda.

4.2.4.2. Análisis

En esta gráfica podemos observar la figura que se formada a partir de la distribución Poisson obtenida a través de la generación de números pseudoaleatorios. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones por un lambda de valor 50 y en el eje de las ordenadas la frecuencia absoluta de los diferentes intervalos del histograma.

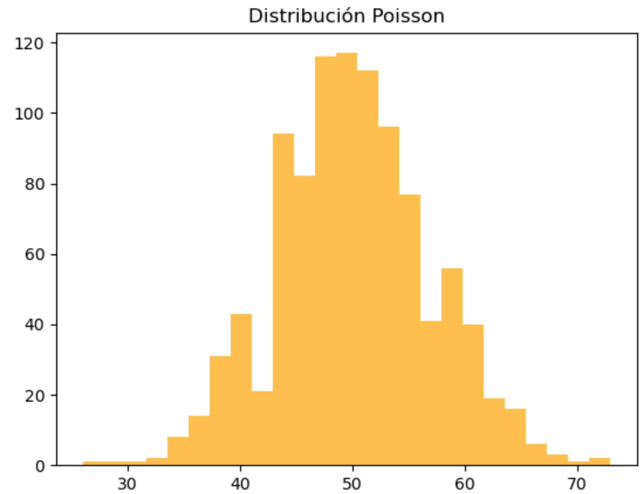

Figura 8: Frecuencias absolutas de valores generados.

4.2.4.3. Test

Valor de X2 obtenido = 4.981640986645496

Se poseen 30 intervalos, por lo tantos tendremos 29 grados de libertad.

Con 29 grados de libertad y con un 95 porciento de confianza se obtiene un valor de Chi cuadrado de 42,56

El valor de chi cuadrado obtenido: 4.981640986645496 es menor que 16.92 ,por lo tanto Paso el test.

Valor Observado	Valor Esperado	Valor Ji-Cuadrado
0	0.07112803794031959	0.07112803794031959
3	2.6146757209846765	0.1279132059085921
24	31.268611202195082	1.81755360813301
140	145.84172123781602	2.0515449977759217
305	301.3950691932913	2.0946629099433007
311	303.27871616598674	2.291241901041262
170	159.84882289088642	2.93589098157238
41	46.80178059878204	3.6551084490650516
6	7.992563228096383	4.151858753648817
0	0.8297822329966786	4.981640986645496

Cuadro 5: Valores observados,esperados y chi-cuadradado de una distribución Poisson

4.2.5. Distribución Empírica Discreta

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

$$p1 + ... + p_{i-1} < r \leq p_1 + ...p_i. \tag{35}$$

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

4.2.5.1. Código en Python

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

No se ingresara ningún parámetro, definiendo nosotros una serie de probabilidades cuya sumatoria da como resultado el valor de uno.

4.2.5.2. Análisis

En esta gráfica podemos observar la figura que se forma a partir de la distribución Empírica discreta obtenida por la generación de números pseudoaleatorios. En el eje de las abscisas se encuentran los valores obtenidos de las simulaciones desde diez probabilidades (0.273,0.037,0.195,0.009,0.124,0.058, 0.062,0.151,0.047,0.044) cuya suma da un total de 1, y en el eje de las ordenadas la frecuencia absoluta de los diferentes intervalos del histograma.

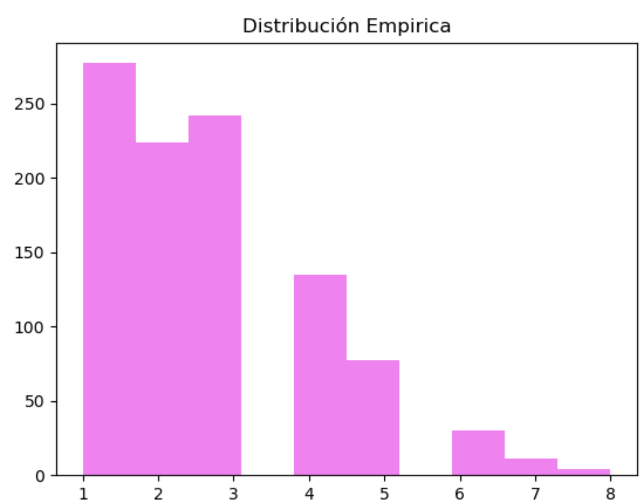

Figura 9: Frecuencias absolutas de valores generados.

4.2.5.3. Test

Valor de X2 obtenido = 3027.8842519713903

Se poseen 10 intervalos, por lo tantos tendremos 9 grados de libertad.

Con 9 grados de libertad y con un 95 porciento de confianza se obtiene un valor de Chi cuadrado de 16.92

Valor Observado	Valor Esperado	Valor Ji-Cuadrado
277	273	0.05860805860805861
224	37	945.1667161667162
242	195	956.4949212949214
135	9	2720.4949212949214
77	124	2738.3094374239536
30	58	2751.826678803264
11	62	2793.7782917064897
4	151	2936.8842519713903
0	47	2983.8842519713903
0	44	3027.8842519713903

Cuadro 6: Valores observados,esperados y chi-cuadradado de una distribución Empírica

Como X2 es mayor a 16.92, no pasó el test.

5. Conclusiones

Después de visualizar lo obtenido en las simulaciones, podemos decir que los generadores por medio del método de la transformada inversa, se distribuyen de manera pseudo aleatoria correspondiendo a cada una de las distribuciones planteadas. Del test chi cuadrado realizado a la distribución Uniforme, Exponencial, Normal, Poisson, Binomial y Empírica se puede decir que lograron resultados muy satisfactorios aunque los intervalos y las formas de calcularlos fueron muy variados. En cuanto a las distribuciones aleatorias continuas y discretas realizadas, se pueden ver los intervalos distintos de cada una y la forma gráfica que se representan, logrando mostrar las partes más importante de cada una.

6. Referencias

https://astridmll.wordpress.com/2016/09/13/numeros-pseudoaleatorios-y-sus-caracteristicas/
https://www.medwave.cl/link.cgi/Medwave/Series/MBE04/5266?ver=sindiseno

https://ansenuza.unc.edu.ar/comunidades/bitstream/handle/11086.1/1346/DistribC3