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

A PREPRINT

Gigena Daiana
Catedra Simulacion
Legajo 39372
UTN-FRRO
daigigena3@gmail.com

Quintero Florencia
Catedra Simulacion
Legajo 44199
UTN-FRRO
florencia.quinteroo@gmail.com

Vilchez Joaquin
Catedra Simulacion
Legajo 46483
UTN-FRRO
joaquinvilchez95@gmail.com

10 de junio de 2020

ABSTRACT

El siguiente documento tiene por objetivo detallar la finalización del trabajo realizado sobre uno de tantos elementos fundamentales de las simulaciones: los generadores de números pseudoaleatorios.

1. Introducción

Una Distribución de probabilidad indica toda la gama de valores que pueden representarse como resultado de un experimento, si este se llevase a cabo. Es decir, describe la probabilidad de que un evento se realice.
Cuando abordamos el estudio de distribuciones de probabilidad, no solo lo hacemos a través de las definiciones teóricas, sino que también lo hacemos mediante algún método, formula o algoritmo que nos permita obtener valores que sigan el comportamiento de tales distribuciones de probabilidad.

2. Variables Aleatorias Discretas y Distribuciones de Probabilidad

Una variable aleatoria es discreta si su recorrido es un conjunto finito o infinito numerable (susceptible de ser contado). Las distribuciones que tienen este tipo de variable son:

- Distribución Pascal
- Distribución Binomial
- Distribución Hipergeometrica
- Distribución Poisson
- Distribución Empírica

2.1. Función de Probabilidad puntual

En cierta ocasiones resulta necesario calcular la probabilidad de que una variable aleatoria asuma un valor menor o igual a un cierto valor dado, por ello decimos: Sea X una variable aleatoria discreta con recorrido $\Re_X$ y una función de probabilidad p_X. La Función F_X tal que:

$$F_X : \mathbb{R} \to [0, 1]$$
$$x \to F_X(x)) = P(X \leq x)$$

El símbolo P(X = x) se lee 'probabilidad de que la variable aleatoria X asuma el valor x'.
Si p_X es una función de probabilidad, entonces cumple que:

1. $p_x(x) \geq 0, \; \forall \; x \; \epsilon \; \Re_x$.
2. $\sum_{x \epsilon \Re_x} p_x(x) = 1$.

2.2. Función de Probabilidad acumulada

Sea X una variable aleatoria discreta con recorrido $\Re_X$ y función de probabilidad p_X. La función F_X tal que:

$$F_X : \mathbb{R} \rightarrow [0,1]$$
$$x \rightarrow F_X(x) = P(X \leq x)$$

Es decir, $F_X(x)$ representa la probabilidad de que la variable X asuma un valor menor o igual a $x \; \epsilon \; \Re_x$ se observa que:

$$P(X \leq x) = \sum_{t \leq x \; ; \; t \; \epsilon \; \Re_x} p_x(t)$$

Cualquiera sea la variable aleatoria X, su función de probabilidad acumulada F_X se comprueba que:

1. $0 \leq F_X(x) \leq 1, \; para \; x \epsilon \mathbb{R}$
2. $F_X \; es \; no \; decreciente \; en \; \mathbb{R}$
3. $F_X \; es \; discontinua \; en \; cada \; punto \; x \; donde \; P(X = x) > 0$

2.3. Esperanza Matemática de una variable discreta

Sea X una variable aleatoria discreta con recorrido $\Re_X$ y p_X su función de probabilidad asociada. Definimos Esperanza matemática de la variable aleatoria X o media poblacional de la variable aleatoria X y la notamos E(X) o μ_X a:

- $E(X) = \mu_X = \sum_{x \epsilon \Re_x} x p_X(x)$

2.4. Varianza de una variable discreta

Sea X una variable aleatoria discreta con recorrida $\Re_X$ y p_X su función de probabilidad asociada. Se define varianza de una variable aleatoria X y la notamos V(X) o σ_X^2 a:

- $V(X) = \sigma_X^2 = \sum_{x \epsilon \Re_X} (x - \mu_X)^2 p_X(x)$

3. Variables Aleatorias Continuas y Distribuciones de Probabilidad

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. Si se definen probabilidades para variables continuas, estas son algunas de sus distribuciones:

- Distribución Uniforme
- Distribución Exponencial
- Distribución Gamma
- Distribución Normal

3.1. Función de Densidad

Sea X una variable aleatoria continua. Decimos que su función de densidad f_X de probabilidad asociada a X es cuando se cumple:

- $f_X(x) \geq 0$ para todo $x \epsilon \mathbb{R}$
- $\int_{-\infty}^{+\infty} f_X(x) dx = 1$
- $P(a \leq X \leq b) = \int_a^b f_X(x) dx$

3.2. Función de distribución acumulada

Sea X una variable aleatoria continua y f_X su densidad de probabilidad. Se denomina Función de distribución acumulada de la variable X a:

$$F_X : \mathbb{R} \to [0,1]$$
$$x \to F_X(x) = \int_{-\infty}^{x} f_x(t)dt$$

Propiedades:

1. $F_X(x) = P(X \leq x)$.
2. F_X es monótona no decreciente.
3. $F'_X(x) = f_X(x)$ en los puntos de continuidad de f_X.

3.3. Esperanza matemática de una variable continua

Sea X una variable aleatoria continua y f_X su función de densidad de probabilidad.La esperanza matemática o media poblacional de X, que se nota E(X) o μ_X es siempre que la integral sea finita:

- $E(X) = \mu_X = \int_{-\infty}^{+\infty} x f_X(x)dx$

3.4. Varianza de una variable continua

Sea X una variable aleatoria continua y fX su función de densidad de probabilidad. La varianza de X que se nota V(X) o σ_X^2 y es el numero no negativo de:

- $V(X) = \sigma_X^2 = \int_{-\infty}^{+\infty} (x - \mu_X)^2 f_X(x)dx$

4. Distribuciones de Probabilidad

4.1. Distribución Uniforme

Decimos que una variable aleatoria X se distribuye uniformemente en el intervalo [a,b] y notamos X $\sim$ U(a,b), cuando su función de densidad de probabilidad es:

$$f(x) = \begin{cases} \frac{1}{b-a} \; si \; x \, \epsilon \, |a,b| \\ \\ 0 \; en \; otro \; caso \end{cases} \tag{1}$$

El hecho de que una variable tenga este tipo de comportamiento significa que los intervalos de igual amplitud contenidos en el intervalo [a,b] tienen la misma probabilidad de ocurrir.

4.1.1. La Esperanza matemática y Varianza

La Esperanza se calcula:

$$E(X) = \mu_X = \frac{(a+b)}{2} \tag{2}$$

La Varianza se calcula:

$$V(X) = \sigma_X^2 = \frac{(b-a)^2}{12} \tag{3}$$

4.1.2. Gráfico

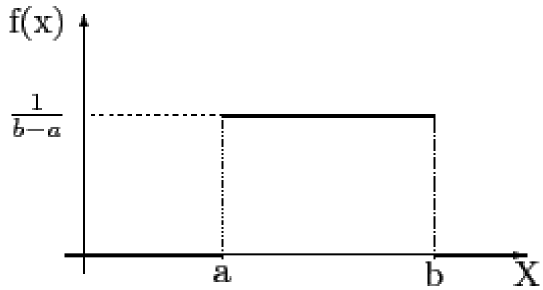


Figura 1: Función de densidad uniforme

4.1.3. Transformada Inversa

$$f(x) = \frac{1}{b-a} \qquad a \leq x \leq b \tag{4}$$

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

$$F(x) = \frac{x-a}{b-a} \tag{6}$$

$$x = F(r)^{-1} \tag{7}$$

$$r = \frac{x-a}{b-a} \tag{8}$$

$$r(b-a) = x - a \tag{9}$$

$$x = r(b-a) + a \tag{10}$$

4.1.4. Código en Python

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

4.2. Distribución Exponencial

Este tipo de distribución se utiliza mucho para describir el tiempo en eventos, más específicamente la variable aleatoria que representa al tiempo necesario para servir a la llegada.
Decimos que una variable aleatoria X tiene una distribución exponencial de parámetro λ >0 y notamos X $\sim$ E (λ) cuando su función de densidad de probabilidad es:

$$f(x) \begin{cases} \lambda\, e^{-\lambda x} \; si \; x > 0 \\ \\ 0 \; en \; otro \; caso \end{cases} \tag{11}$$

4.2.1. La Esperanza matemática y Varianza

La Esperanza se calcula:

$$E(X) = \mu_X = \frac{1}{\lambda} \tag{12}$$

La Varianza se calcula:

$$V(X) = \sigma_X^2 = \frac{1}{\lambda^2} \tag{13}$$

4.2.2. Gráfico

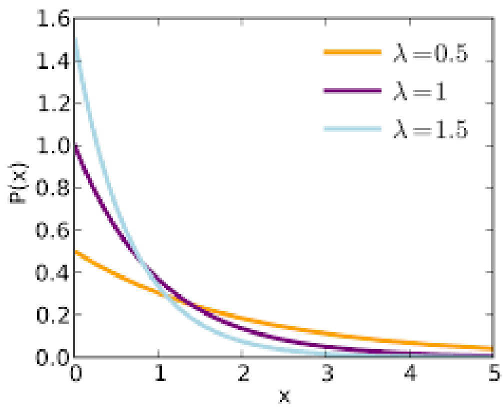


Figura 2: Función de densidad de probabilidad exponencial para tres λ diferentes

4.2.3. Transformada Inversa

$$f(x) = \alpha e^{-\alpha x} \quad \alpha > 0 \ , x \geq 0 \tag{14}$$

$$F(x) = \int_0^x \alpha e^{-\alpha x} \tag{15}$$

$$F(x) = 1 - e^{-\alpha x} \tag{16}$$

$$x = F(r)^{-1} \tag{17}$$

$$r = 1 - e^{-\alpha x} \tag{18}$$

$$1 - r = e^{-\alpha x} \tag{19}$$

$$\ln(1 - r) = \ln \ e^{-\alpha x} \tag{20}$$

$$\ln(1 - r) = -\alpha x \tag{21}$$

$$x = -\frac{\ln(1 - r)}{\alpha} \tag{22}$$

4.2.4. Código en Python

```
def EXPENT(alfa):
    ex = 1 / alfa
    x = []
    for i in range(1000):
        r = random.random()
        x += [-ex * (math.log(r))]
    return x
```

4.3. Distribución Gamma

Este tipo de probabilidad es adecuada para modelizar el comportamiento de variables aleatorias con asimetría positiva y/o los experimentos en donde está involucrado el tiempo.
Decimos que una variable aleatoria X tiene una distribución gamma, X $\sim$ τ (k,λ) con k>0 y λ >0, cuya función de densidad de probabilidad es:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} \frac{(\lambda x)^{k-1}}{\tau k} \; si \; x > 0 \\ \\ 0 \; si \; x \geq 0 \end{cases} \tag{23}$$

4.3.1. La Esperanza matemática y Varianza

Cabe aclarar, que no existe una forma explícita para describir la función acumulativa de la distribución gamma. Respectos de la media y su varianza, sus correspondientes expresiones están dadas por:
La Esperanza:

$$E(X) = \mu_X = \frac{k}{\lambda} \tag{24}$$

La Varianza:

$$V(X) = \sigma_X^2 = \frac{k}{\lambda^2} \tag{25}$$

4.3.2. Gráfico

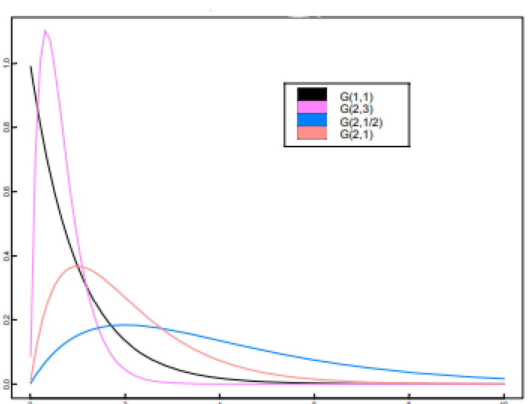


Figura 3: Generación de valores aleatorios de una Distribución Gamma

4.3.3. Código en Python

```
def GAMMA (k,a):
    x=[]
    for i in range(1, 1000):
        tr=1.0
        for j in range(1,k):
            r = random.random()
```

```
            tr=tr*r
        x.append(-(math.log10(tr))/a)
    return x
```

4.4. Distribución Normal

Esta distribución basa su utilidad en el teorema del límite central que postula que la distribución de probabilidad de la suma de N valores de variable aleatoria X independientes, pero idénticamente distribuidos.
Decimos que una variable aleatoria X tiene una distribución normal de parámetros μ σ, donde σ >0 , y notamos X $\sim$ N (μ, σ) cuando su función de densidad de probabilidad es:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{26}$$

4.4.1. La Esperanza matemática y Varianza

La Esperanza es:

$$E(X) = \mu \tag{27}$$

La Varianza es:

$$(X) = \sigma^2 \tag{28}$$

4.4.2. Gráfica

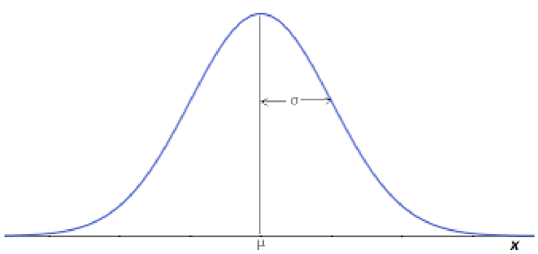


Figura 4: Función de densidad de probabilidad normal estándar

4.4.3. Código en Python

```
def NORMAL(mean, sd):
    lista_normal = []
    for j in range(1000):
        sum = 0
        for i in range(1, 12):
            r = random.random()
            sum = sum + r
        x = (sd * (sum - 6)) + mean
        lista_normal.append(x)
    return lista_normal
```

4.5. Distribución Pascal

También llamada Distribución Binomial Negativa, es una generalización de la distribución geométrica, pero en este caso la variable se define como 'cantidad de ensayos que se realizan hasta que el suceso A se presenta por r-ésima vez'. Sea X la variable aleatoria y P(A)=p. Entonces X tiene una distribución de Pascal con parámetros p y r, entonces:

$$P(X = k) = k - 1r - 1p^r(1-p)^{k-r} \; k = r, r+1, r+2, ... \tag{29}$$

4.5.1. La Esperanza matemática y Varianza

La Esperanza se calcula:

$$E(X) = \frac{r}{p} \tag{30}$$

La Varianza se calcula:

$$V(X) = \frac{r(1-p)}{p^2} \tag{31}$$

4.6. Gráfica

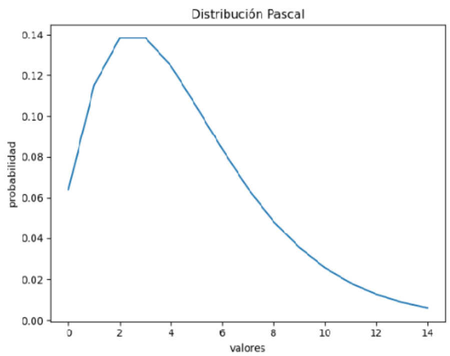


Figura 5: Gráfico Distribución Pascal con parámetros r=3, p=0.40

4.6.1. Código en Python

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

4.7. Distribución Binomial

Es una distribución que cuenta el número de éxitos en una secuencia de n ensayos de Bernoulli independientes entre sí, con una probabilidad fija p de ocurrencia del éxito entre los ensayos. Un experimento de Bernoulli se caracteriza por ser dicotómico, esto es, solo dos resultados son posibles. A uno de estos se denomina éxito y tiene una probabilidad de ocurrencia p y al otro, fracaso, con una probabilidad q = 1 - p.
En la distribución binomial, dicho experimento, se repite n veces, de forma independiente, y se trata de calcular la probabilidad de un determinado número de éxitos.
La función de densidad probabilidad para esta distribución se puede expresar de la manera siguiente:

$$p_Y(k) = P(Y = k) = nkp^k(1-p)^{n-k} \; k = 0, 1, 2, ..., n. \tag{32}$$

4.7.1. La Esperanza matemática y Varianza

La Esperanza se calcula:

$$E(X) = np \tag{33}$$

La Varianza se calcula:

$$V(X) = np(1-p) \tag{34}$$

4.8. Gráfico

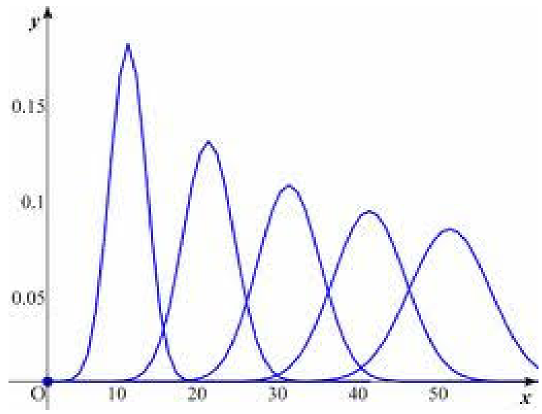


Figura 6: Gráfica de una distribución binomial(n, 0.5) para n = 20, 40, 60, 80, 100. Se puede ver la aproximación a la distribución normal.

4.8.1. Código en Python

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

4.9. Distribución Hipergeometrica

Esta distribución 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 tiene dos resultados posibles (es un evento o un no evento). 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.

$$P(X=k)=\frac{rkN-rn-k}{Nk}\ k=0,1,2,...,min(n,r) \tag{35}$$

4.9.1. La Esperanza matemática y Varianza

Sea p = r /N podemos demostrar que
La Esperanza se calcula:

$$E(X)=np \tag{36}$$

La Varianza se calcula:

$$V(X)=np(1-p)\frac{N-n}{N-1} \tag{37}$$

4.9.2. Gráfico

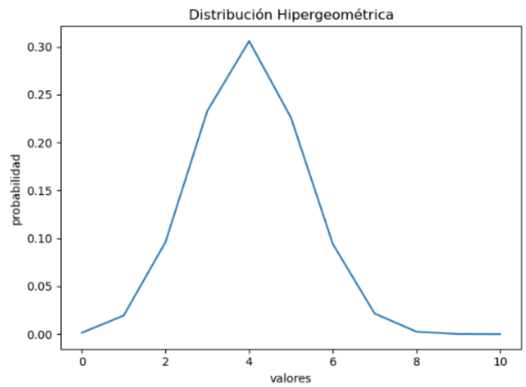


Figura 7: Gráfico de Distribución Hipergeometrica con valores N=12,r=30,n=10

4.9.3. Código en Python

```
def HIPERGEOMETRICA(tn, ns, p):
    listado_hipergeometrica = []
    for i in range(1000):
        p_variable = p
        tn_variable = tn
        x = 0
        s = 0

        for i in range(1, ns):
            r = np.random.rand()
            if (r - p_variable <= 0):
                s = 1
                x = x + 1
            else:
                s = 0
            p_variable = (tn_variable * p_variable - s) / (tn_variable - 1)
            tn_variable = tn_variable - 1

        listado_hipergeometrica.append(p_variable)

    return listado_hipergeometrica
```

4.10. Distribución Poisson

Es una distribución de probabilidad que expresa a partir de una frecuencia de ocurrencia media λ, la probabilidad que ocurra un determinado números de eventos durante un intervalo de tiempo dado o región. En general se usa para la aproximación de experimentos binomiales donde el numero de pruebas es muy alto ($n \to \infty$), pero la probabilidad de éxito es muy baja ($p \to 0$).
Decimos que una variable aleatoria X tiene una distribución de Poisson con parámetro λ >0 y notamos X $\sim$ P (λ) cuando:

$$P(X = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 0, 1, 2...n \tag{38}$$

4.10.1. La Esperanza matemática y Varianza

La Esperanza y la Varianza en este caso es:

$$E(X) = V(X) = \lambda \tag{39}$$

4.10.2. Gráfico

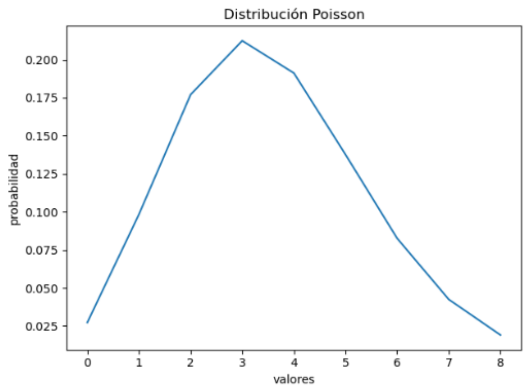


Figura 8: Gráfico Distribución de Poisson con λ=3.6

4.10.3. Código en Python

```
def POISSON(p):
    listado_poisson = []
    for i in range(1000):
        x = 0
        b = np.exp(-p)
        tr = 1
        r = np.random.rand()
        tr = tr * r

        while ((tr - b) >= 0):
            x = x + 1
            r = np.random.rand()
            tr = tr * r
        listado_poisson.append(x)

    return listado_poisson
```

4.11. Distribución Empírica Discreta

En principio, este método se centra en una tabla de frecuencias de una variable x (continua o discreta), la cual utiliza para generar números aleatorios.
La función de distribución empírica FDE se define como la distribución discreta que asigna probabilidad $\frac{1}{n}$ a cada valor x_i tal que $i = 1, 2, ..., n$:

$$FDE(x) = \frac{1}{N} \sum_{i=1}^{N} I\{x \leq x_i\} \tag{40}$$

4.11.1. Código en Python

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

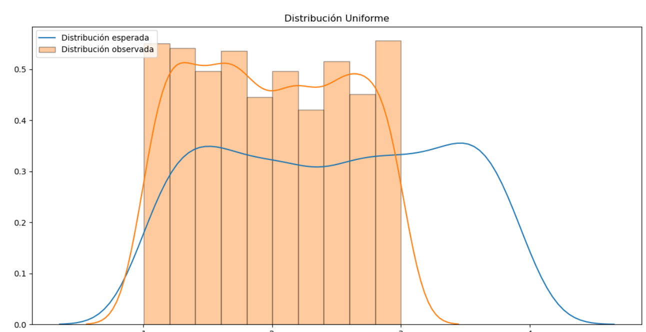
```
                break
            else:
                x = x + 1
        lista_empirica.append(x)
    return lista_empirica
```

5. Testeo

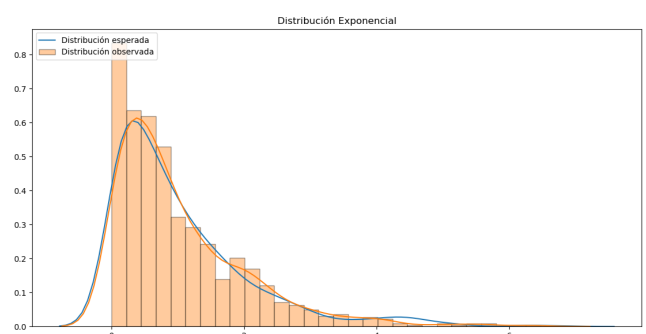
Para poder verificar que la generación de variables aleatorias se asemeja a las distribuciones nombradas anteriormente correspondientemente de cada una, los valores obtenidos se someten a distintas pruebas. Para realizar estas pruebas lo que se hizo fue graficar las distribuciones teóricas y compararlas con los datos obtenidos de cada distribución.

Como se puede observar en los siguientes gráficos, la gráfica de color azul representa la distribución teórica esperada, y la de color naranja representa un histograma de los datos obtenidos de dicha distribución asemejándose cada una a la gráfica de color azul y llegando a la conclusión de que al ser similares, los datos obtenidos respetan dicha distribución.

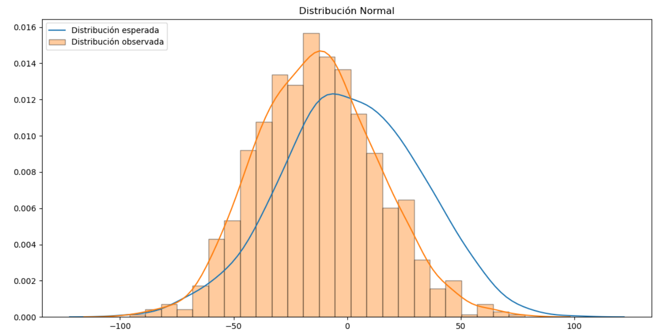
5.1. Distribución uniforme



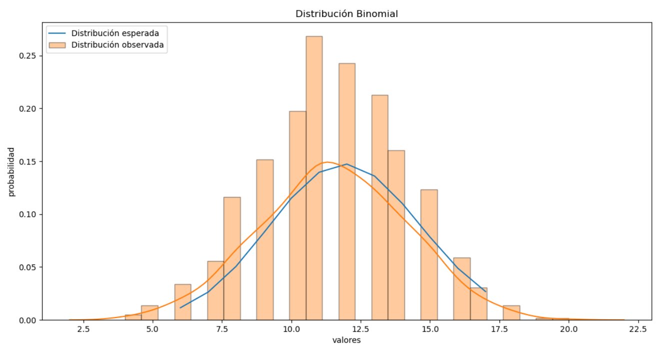
5.2. Distribución exponencial



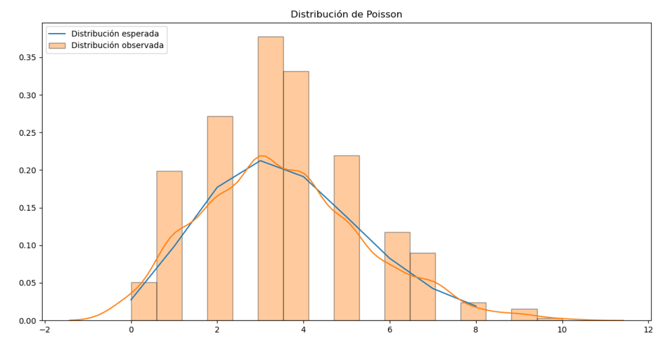
5.3. Distribución normal



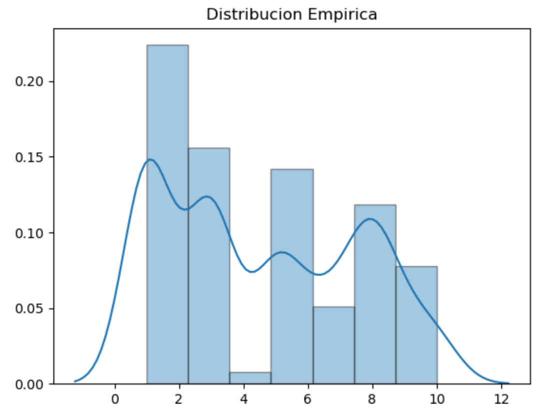
5.4. Distribución binomial



5.5. Distribución de Poisson



5.6. Distribución empírica discreta



Para el caso de la distribución empírica discreta utilizamos el Test Chi-cuadrado, (desarrollado anteriormente en el Trabajo Practico 2.1). A través de esta prueba verificamos si las variables son independientes una de la otra o no.

6. Conclusión

Podemos concluir que la generación de números pseudoaleatorios se pueden generar de distintas maneras dependiendo de la necesidad eventual que tengamos. Observamos que hay distintos algoritmos que resuelven estos problemas para simular correctamente situaciones aleatorias de la realidad, las cuales cada una se comporta de manera diferente. Al utilizar estos algoritmos en simulaciones podemos generar comportamientos de diferentes eventos de la vida real que se asemejan a las distribuciones nombradas anteriormente.

Referencias

[1] Variables Aleatorias Continuas y algunas Distribuciones de Probabilidad.Raúl D. Katz Pablo A. Sabatinelli 2013. UTN Rosario.

[2] Variables Aleatorias Discretas y algunas Distribuciones de Probabilidad. Raúl D. Katz Pablo A. Sabatinelli 2018. UTN Rosario

[3] Naylor, T.H. Técnicas de simulación en computadoras, 1982.