Trabajo Práctico Distribuciones de Probabilidad

Bengoechea, Guadalupe María[a], Corsetti,Ornela Milagros[a], Garcia, Narela Rosana[a], Golzman,Gabriel[a]

[a] *Universidad Tecnológica Nacional Regional Rosario*

Enunciado

Este trabajo consiste en la generación de mil valores aleatorios, mediante la aplicación del método de la transformación inversa a siete funciones distintas de distribuciones de probabilidad. Finalmente, se debe comprobar si estos valores generados conforman una distribución de probabilidad semejante a la teórica.

Palabras Clave:
distribución,probabilidad,histograma,gráfica, python,excel

1. Marco Teórico

1.1. Introducción.

Al considerar los procesos estocásticos que involucran variables continuas o discretas, pero siempre de tipo aleatorio, definimos una función $F(x)$ llamada *función de distribución acumulativa de x*, la cuál denota la probabildiad que una variable aleatoria X tome un valor menor o igual a x.

Si la variable aleatoria es discreta, entonces x tendrá valores específicos y $F(x)$ será una función escalonada.

Si $F(x)$ es continua en el dominio de x, entonces está función se podrá diferenciar, para lo cual se define $f(x) = \frac{dF(x)}{dx}$
La derivada $f(x)$ recibe el nombre de función de densidad de probabilidad.La función de distribución acumulativa se puede proponer matemáticamente como:

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

,donde $F(x)$ se define en el intervalo $0(x) \leq 1$ y f(t) representa el valor de la función de densidad de probabilidad de la variable aleatoria X cuando $X = t$

Denotaremos por r a los valores de variables aleatorias uniformes cuando $0 \leq r \leq 1$ y $F(r) = r$.
Un método básico para generar valores de variables aleatorias a partir de distribuciones de probabilidad es el *método de la transformación inversa*.

1.2. Método de la transformación inversa.

Si deseamos generar los valores de x_i de las variables aleatorias a partir de cierta estadística de población cuya función de densidad esta dada por $f(x)$, debemos en primer lugar obtener la función de distribución acumulativa $F(x)$. Puesto que $F(x)$ se define sobre el rango de 0 a 1, podemos generar números aleatorios distribuidos uniformemente y además hacer $F(x) = r$.

Entonces, x queda determinada unívocamente por $r = F(x)$.Sigue, por lo tanto, que para cualquier valor particular de r, que generemos, por ejemplo r_0, siempre es posible encontrar el valor de x; en este caso x_0, que corresponde a r_0 debido a la función inversa de F, si es conocida. Esto es,

$$x_0 = F^{-1}(r_0)$$

donde $F^{-1}(x)$ es la transformación inversa o mapeo de r sobre el intervalo unitario en el dominio de x.

Si generamos números aleatorios uniformes correspondientes a una $F(x)$ dada, podemos resumir, matemáticamente esto como:

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

entonces

$$P(X) = F(x) = P[r \leq F(x)] = P[F^{-1}(r) \leq x]$$

y consecuentemente $F^{-1}(r)$ es una variable que tiene a f(x) como función de densidad de probabilidad.

1.3. Distribuciones de probabilidad.

1.3.1. 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:

$$f_x(x) = \begin{cases} \frac{1}{b-a} & \text{si} \quad x \in [a, b] \\ 0 & \text{en otro caso} \end{cases}$$

El hecho de que una variable aleatoria tenga un comportamiento uniforme significa que intervalos de igual amplitud contenidos

---

Correos Electrnicos: guadabengoechea@gmail.com (Bengoechea, Guadalupe María), ocorsetti@frro.utn.edu.ar (Corsetti,Ornela Milagros), nargarcia@frro.utn.edu.ar (Garcia, Narela Rosana), ggolzman@frro.utn.edu.ar (Golzman,Gabriel)

en el intervalo $[a, b]$ tienen la misma probabilidad de ocurrir.

El valor esperado para la Esperanza y la Varianza en esta distribución es:

$$E(X) = \mu_x = \int_a^b \frac{x}{b-a} dx = \frac{a+b}{2}$$

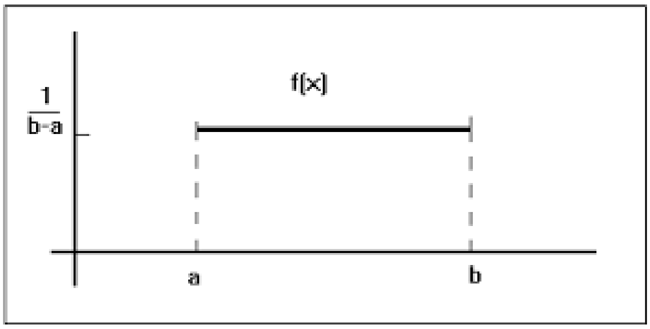
$$V(X) = \sigma_x^2 = \int_a^b \left(x - \frac{a+b}{2}\right)^2 * \frac{1}{b-a} dx = \frac{(b-a)^2}{12}$$

Aplicando el método referenciado en la sección 1.2 en $f(x)$ se obtiene que la transformación inversa de la función de densidad es:

$$x = a + (b-a)r \quad r \in [0, 1]$$

La gráfica que presenta dicha distribución es:

Figura 1: Distribución Uniforme.



1.3.2. Distribución exponencial.

Decimos que una variable aleatoria X tiene una distribución exponencial de parámetro $\alpha > 0$ y notamos $X \sim \mathrm{E}(\alpha)$ cuando su función de densidad de probabilidad es

$$f_x(x) = \begin{cases} \alpha e^{-\alpha x} & \text{si} \quad x > 0 \\ \\ 0 & \text{en otro caso} \end{cases}$$

Si $x > 0$, entonces $F_x(x) = \int_0^x \alpha e^{-\alpha t} dt = 1 - e^{-\alpha x}$.
Consecuentemente $P(X > x) = 1 - 1(-e^{-\alpha x)} = e^{-\alpha x}$

$$E(X) = \mu_x = \int_0^{+\infty} x\alpha e^{-\alpha x} dx = \frac{1}{\alpha}$$

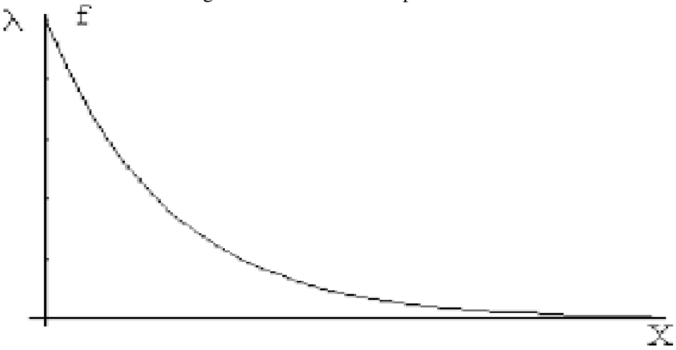
$$V(X) = \sigma_x^2 = \int_0^{+\infty} \left(x - \frac{1}{\alpha}\right)^2 \alpha e^{-\alpha x} dx = \frac{1}{\alpha^2}$$

Aplicando el método referenciado en la sección 1.2 en $f(x)$ se obtiene que la transformación inversa de la función de densidad es:

$$x = -\frac{1}{\alpha} \ln r \quad r \in [0, 1]$$

La gráfica que presenta dicha distribución es:

Figura 2: Distribución Exponencial.



1.3.3. Distribución gamma.

En estadística la distribución gamma es una distribución de probabilidad continua con dos parámetros k y λ cuya función de densidad para valores $x > 0$ es

$$f(x) = \lambda e^{\lambda x} \frac{(\lambda x)^{k-1}}{\Gamma(k)}$$

Aquí e es el número e y Γ es la función gamma. Para valores $k \in \mathbb{N}$ la función gamma es $\Gamma(k) = (k-1)!$ el factorial de $k - 1$. En este caso - por ejemplo para describir un proceso de Poisson - se llaman la distribución Erlang con un parámetro $\theta = \frac{1}{\lambda}$.

El valor esperado para la Esperanza y la Varianza en esta distribución es:

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

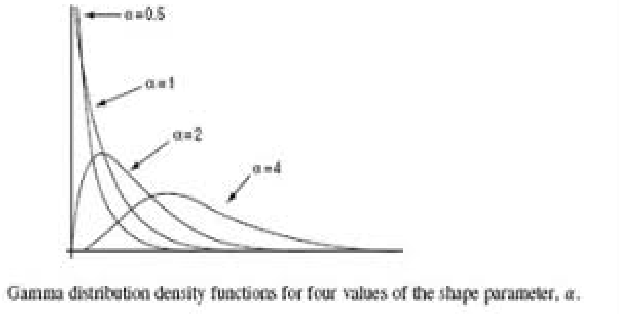
$$V(X) = \frac{k}{\lambda^2} = k\theta^2$$

Aplicando el método referenciado en la sección 1.2 en $f(x)$ se obtiene que la transformación inversa de la función de densidad es:

$$x = -\frac{1}{\alpha}(\log \prod_{i=1}^{k} r_i) \quad r \in [0, 1]$$

La gráfica que presenta dicha distribución es:

Figura 3: Distribución Gamma.



1.3.4. Distribución normal.

La distribución normal fue introducida en 1733 por Abraham De Moivre, al obtenerla como una forma límite de la distribución binomial.

Medio siglo después fue redescubierta por Laplace y Gauss para describir el comportamiento de los errores en las mediciones astronómicas, que seguían una distribución simétrica en forma de campana

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

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

La vida útil de las baterías, el diámetro interior de un anillo de pistón, los errores en instrumentos para medir longitudes, la resistencia a la ruptura de una soga, el nivel de agua en un lago y el valor de glucosa en sangre son variables aleatorias que pueden con una distribución normal

La posibilidad de poder explicar la variabilidad en situaciones tan diversas a partir de un mismo modelo, hace que la distribución normal sea una de las más importantes de la estadística.

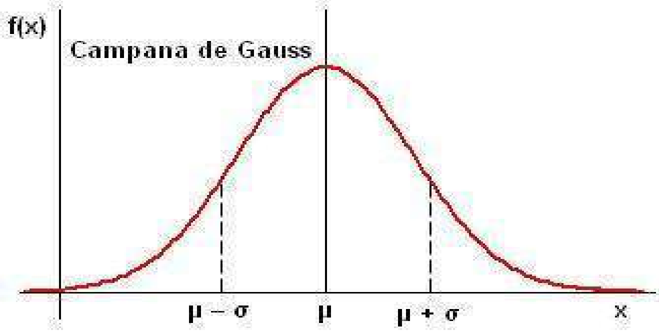
$$E(X) = \int_{-\infty}^{+\infty} x f_x(x)dx = \mu$$
$$V(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 f_x(x)dx = \sigma^2$$

Aplicando el método referenciado en la sección 1.2 en $f(x)$ se obtiene que la transformación inversa de la función de densidad es:

$$x = \sigma_x((\sum_{i=1}^{12} r_i) - 6) + \mu_x \quad r \in [0,1]$$

La gráfica que presenta dicha distribución es:

Figura 4: Distribución Normal.



1.3.5. Distribución binomial

Sea A un suceso con probabilidad p. Entonces, la variable aleatoria Y : 'cantidad de veces que se presenta el suceso A en n repeticiones independientes' tiene un comportamiento o distribución binomial con parámetros n y p. En símbolos

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

.

$$E(X) = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = ... = np$$
$$V(X) = \sum_{k=0}^{n} (k-np)^2 \binom{n}{k} p^k (1-p)^{n-k} = ... = np(1-p)$$

.

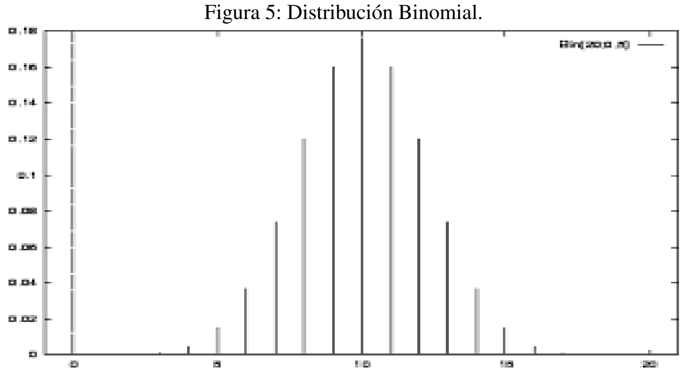
Aplicando el método referenciado en la sección 1.2 en $f(x)$ se obtiene que la transformación inversa de la función de densidad es:

$$x_i = x_{x-1} + 1 \quad si\ r_i \leq p$$
$$x_i = x_{x-1} \quad si\ r_i > p$$

Para cada número aleatorio r_i $(1 \leq i \leq n)$.

La gráfica que presenta dicha distribución es:

Figura 5: Distribución Binomial.



1.3.6. Distribución Poisson.

Decimos que una variable aleatoria X tiene una distribución de Poisson con parámetro > 0 y notamos $X \sim \mathrm{P}(\lambda)$ cuando

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

Una variable con distribución de Poisson es discreta porque su recorrido es infinito numerable (N_0).
Se puede probar que

$$\sum_{k=0}^{+\infty} \frac{\lambda^k e^{-\lambda}}{k!} = 1$$

El valor esperado para la Esperanza y la Varianza en esta distribución es:

$$E(X) = \sum_{k=0}^{+\infty} \frac{\lambda^k e^{-\lambda}}{k!}$$

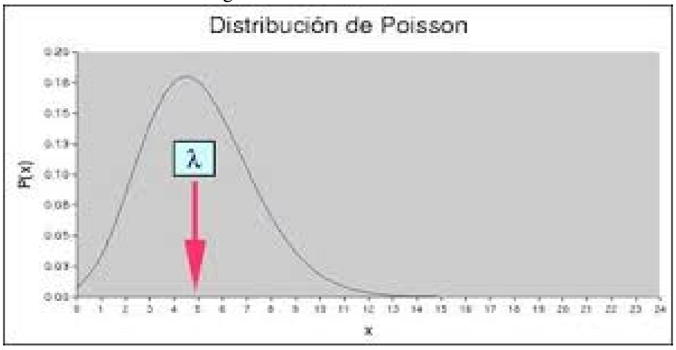
$$V(X) = \sum_{k=0}^{+\infty} (k-\lambda)^2 \frac{\lambda^k e^{-\lambda}}{k!}$$

Aplicando el método referenciado en la sección 1.2 en $f(x)$ se obtiene que la transformación inversa de la función de densidad se da por la siguiente desigualdad:

$$\prod_{i=0}^{x} r_i \geq e^{-\lambda} \geq \prod_{i=0}^{x+1} r_i$$

La gráfica que presenta dicha distribución es:

Figura 6: Distribución Poisson.



1.3.7. Distribución Empirica Discreta.

Esta distribución resulta de la aplicación de un método un tanto más general que se puede emplear para simular cualquiera de las siguientes distribuciones:

- Empirica.
- Discreta.
- Continua que puede ser aproximada mediante una distribución discreta.

Sea X una variable aleatoria discreta con $P(X = b_i) = p_i$, para lograr generar x_i se debe generar un valor de variable aleatoria r sujeto a una distribución uniforme, en el intervalo $(0, 1)$ y un conjunto de valores $x = b_i$, siempre que se satisfaga

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

El histograma de dicha distribución, varia en cuanto a las probabilidades de aparición de cada uno de sus elementos, el que tenga mas probabilidad de aparición tendrá mayor ocurrencias en el histograma. La esperanza para este caso con diez variables es:

$$E(X) = \sum_{i=1}^{10} x_i * p_i$$

La varianza es:

$$V(X) = \sum_{i=1}^{10} (x_i - \mu)^2 * p_i$$

1.4. Histograma.

Un histograma es una representación gráfica de una variable en forma de barras, donde la superficie de cada barra es proporcional a la frecuencia, ya sea absoluta o relativa, de los valores representados.

Su función es exponer gráficamente números, variables y cifras de modo que los resultados se visualicen mas clara y ordenadamente.

Normalmente, las frecuencias son representadas en el eje vertical mientras que en el horizontal se representan los valores de cada una de las variables.

2. Metodología

Para realizar los experimentos generaremos 1000 valores aleatorios en cada una de las distribuciones a analizar. Para realizar esto, codificamos los algoritmos para la generación de valores de cada una de las distribuciones.

Estos algoritmos están basados en las funciones, resultantes de aplicar el método de la sección 1.2, que se obtuvieron en el marco teórico.

Una vez obtenidos los valores necesarios, utilizamos los mismos para realizar un histograma y calculamos la Varianza y esperanza de cada una de las distribuciones, para esto utilizamos la librería *numpy* que incluye los cálculos para determinar la esperanza y la varianza

El objetivo final es hacer una comparación con los valores teóricos y gráficas proveídas por la teoría, para verificar que el total de los valores generados se corresponde con la distribución especificada.

Para la codificación de las distribuciones utilizamos los siguientes algoritmos representados con *Diagramas de Flujo*.

2.0.1. Algoritmos.

Figura 7: Algoritmo de la distribución uniforme.

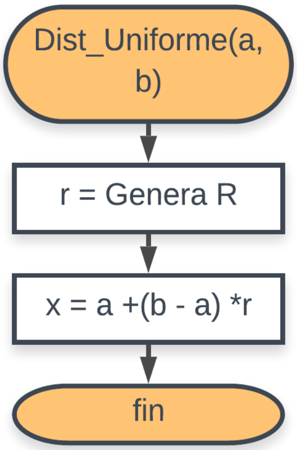


Figura 8: Algoritmo de la distribución exponencial.

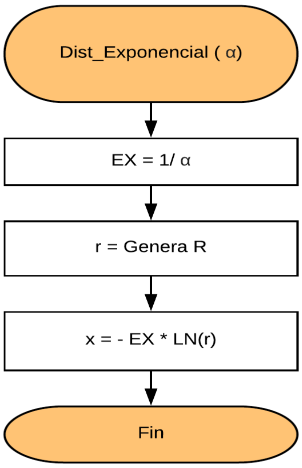


Figura 9: Algoritmo de la distribución gamma.

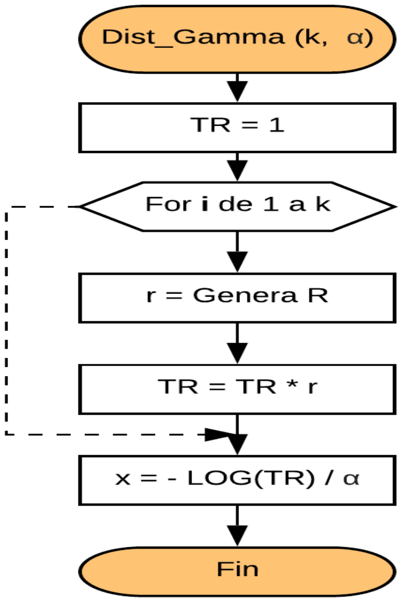


Figura 10: Algoritmo de la distribución normal.

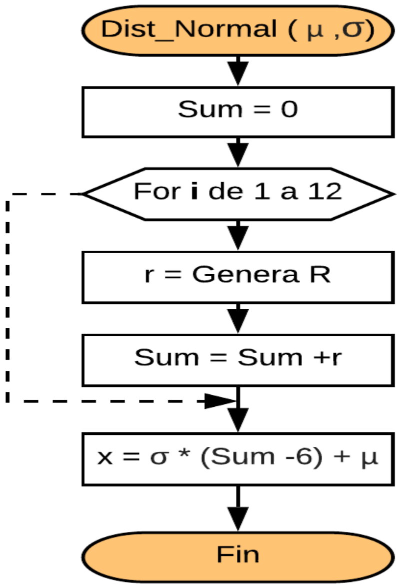


Figura 11: Algoritmo de la distribución binomial.

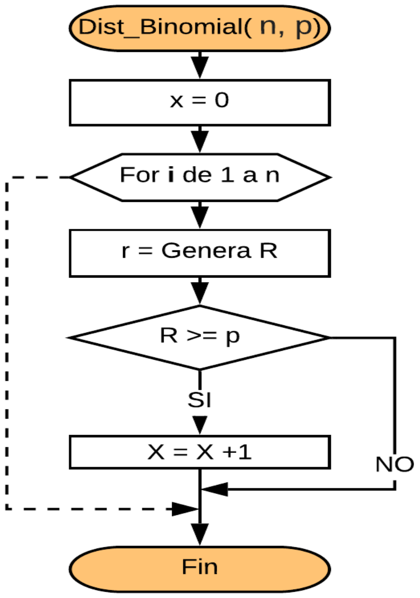


Figura 12: Algoritmo de la distribución de Poisson.

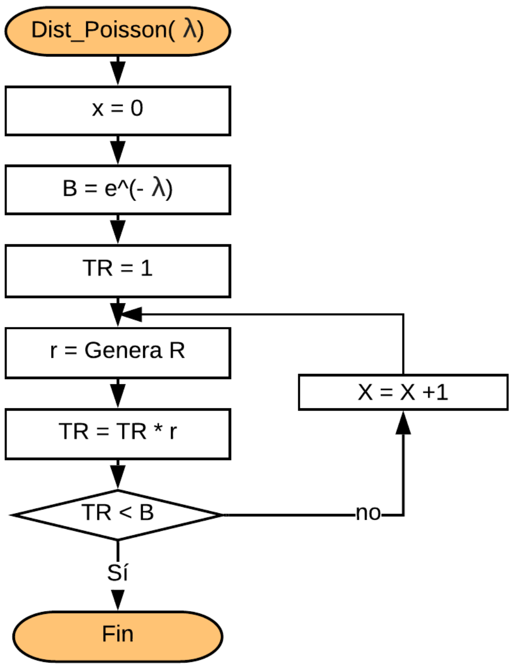
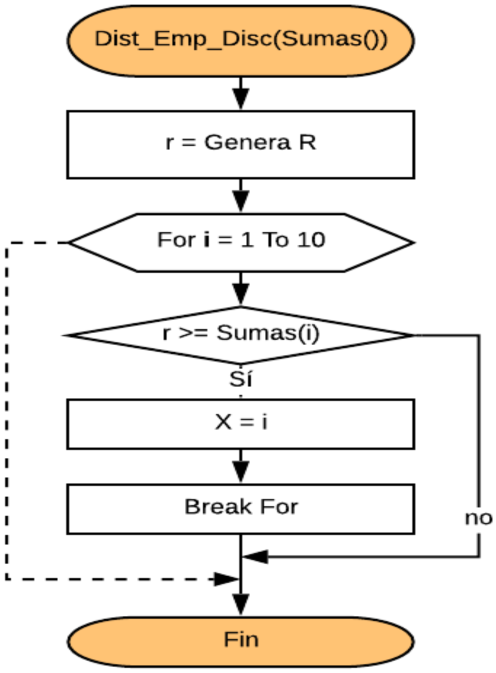


Figura 13: Algoritmo de la distribución empírica discreta.



2.1. Hipótesis

Siguiendo los algoritmos previamente definidos, una vez conseguidos y graficados los 1000 valores para cada distribución, cada histograma debe presentar la forma que le corresponda según su distribución.

- La distribución uniforme debe presentar una gráfica simil rectangular.
- La distribución exponencial presenta una gráfica donde los valores decrecen exponencialmente.
- La distribución gamma presenta una gráfica con forma de campana cuyo pico mas alto debe estar aproximadamente en el valor 1,6 (ya que tomamos $k = 5$ y $\alpha = 3$).
- La distribución normal presenta una gráfica con forma de campana cuyo pico mas alto debe estar aproximadamente en el valor 87,5 (ya que tomamos como μ dicho valor).
- La distribución binomial presenta la misma forma que el experimento anterior solo que esta se debería centrar alrededor del valor 2,5 (tomando $n = 5$ y $p = 0,5$).
- La distribución de Poisson coincide con el ítem anterior, solo que en este caso se encuentra su pico mas alto alrededor del numero 10 (valor seleccionado para λ)
- La gráfica de la distribución empírica discreta dependerá de las probabilidades que se le asigne a cada elemento.

Como tenemos la probabilidad de aparición de cada elemento podemos definir que el que se presentará con mayor ocurrencia es el elemento 1 ya que este es el que posee la mayor posibilidad de aparición, siguiéndole 3,8 y 5 respectivamente.

A su vez, las Esperanzas y Varianzas de la muestra deben aproximarse al valor teórico que le corresponda según su distribución.

3. Casos de estudio

3.1. Experimento 1-Distribución uniforme.

Elegimos dos valores de manera aleatoria para los índices a y b, en este caso el valor 5 y el valor 10 respectivamente.

Luego, aplicamos la fórmula descripta en la sección 1.3.1, la cual colocamos en el algoritmo representado por la figura 7 para obtener un valor para la variables aleatorias x_i a partir de un número r siendo este $0 \leq r \leq 1$.

El número r lo obtuvimos en *python* con la función *random.uniform* de la librería *random*.

Luego de las 1000 corridas, para elaborar la gráfica, usamos la función *hist* perteneciente a la librería *matplotlib*, junto con la descripción de la distribución de los intervalos y otros parámetros estéticos, como el color, la transparencia, etcétera.

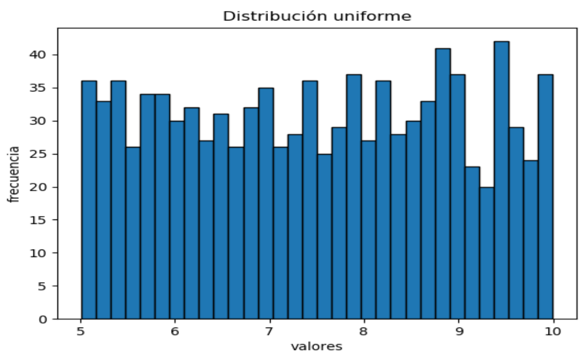


Figura 14: Histograma con distribución uniforme

3.2. Experimento 2-Distribución exponencial.

Designamos un valor α de forma azarosa, en el caso de este experimento designamos el valor tres.

Luego calculamos la Esperanza a partir de la formula 1.3.2. Usamos el procedimiento de obtención de r de la misma manera que en el experimento anterior y, luego, para obtener el valor de la variable aleatoria usamos la fórmula enunciada en la sección 1.3.2, como se puede ver en el algoritmo representado en la figura 8

Repetimos el procedimiento para las 1000 corridas y elaboramos el histograma correspondiente.

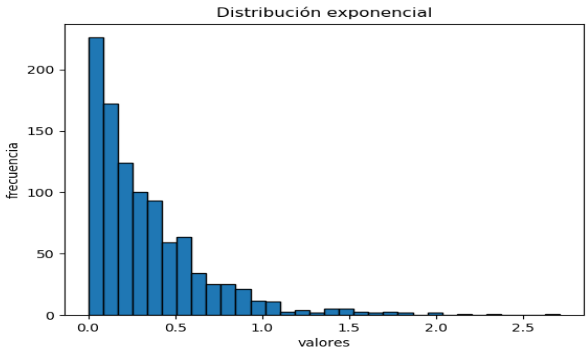


Figura 15: Histograma con distribución exponencial

3.3. Experimento 3- Distribución gamma

Establecemos primero la cantidad de veces que se va a ejecutar la productoria, en este caso, elegimos que se va a ejecutar cinco veces. Designamos con el valor tres a α e inicializamos nuestro TR en 1. Luego, multiplicamos nuestro TR inicial $TR = 1$ por el número obtenido aleatoriamente r una cantidad k de veces hasta que obtuvimos un TR para luego utilizar en la fórmula de obtención de valores de las variables aleatorias x_i de la sección 1.3.3. Este procedimiento se puede observar en la figura 9

Repetimos el procedimiento para las 1000 corridas y elaboramos el histograma correspondiente.

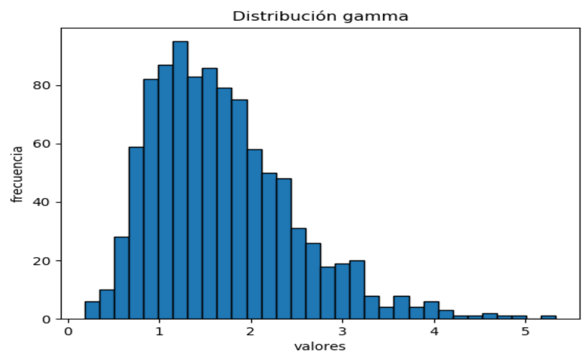


Figura 16: Histograma con distribución gamma

3.4. Experimento 4 - Distribución Normal.

Para conseguir los valores de x_i en esta distribución tenemos como parámetros los valores de μ y σ, los cuales son 87,5 y 0,5 respectivamente. Con dichos valores aplicamos el algoritmo descripto en la figura 10, el cual consiste en repetir doce veces la generación del numero aleatorio r (tomando este valores entre 0 y 1) y realizar la sumatoria de estos doce valores.

Una vez realizado esto, utilizamos esta sumatoria para incluirla en la fórmula 1.3.4.

Posteriormente, para realizar la gráfica,procedemos de la misma manera que los experimentos anteriores.

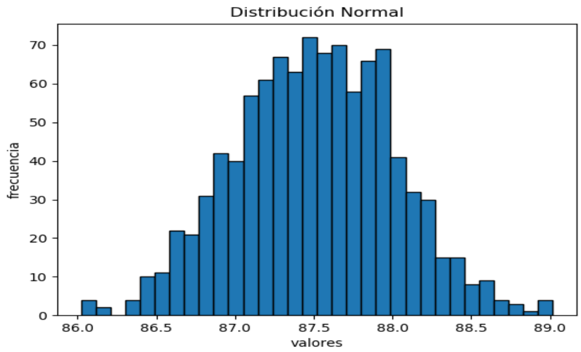


Figura 17: Histograma con distribución normal

3.5. *Experimento 5 - Distribución Binomial.*

De la misma forma que los experimentos anteriores, debemos generar los valores para x_i. Para esto, utilizamos la función 1.3.5. La cual esta representada en el algoritmo de la figura 11. Los valores de los parámetros que utilizaremos para este son $n = 5$ y $p = 0{,}5$.

El algoritmo describe una variable $x = 0$, la cual aumenta progresivamente siempre y cuando se cumpla la condición de que r (número generado aleatoriamente, siendo $r \in [0, 1]$) sea menor que la probabilidad obtenida como parámetro. Dicha comparación se repite n veces.

Luego, realizamos las acciones ya descriptas para realizar el siguiente histograma.

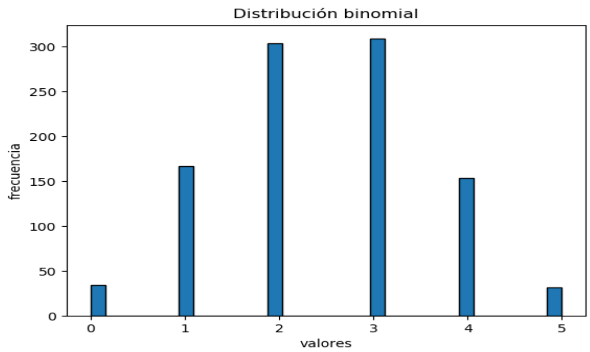


Figura 18: Histograma con distribución binomial.

3.6. *Experimento 6 - Distribución de Poisson*

Como ya describimos, usamos la función 1.3.6, descripta en el algoritmo 12, para generar los valores x_i correspondientes a esta distribución.

El algoritmo recibe como parámetro $\lambda = 10$. Como primer paso inicializa $x = 0$ y $TR = 1$ y calcula la variable B según esta descripto.

Luego se genera el valor para r, y se lo multiplica a TR, de ser este mayor o igual a B, x aumenta en 1. Esto se repite hasta que TR sea menor a B.

Luego de correr el algoritmo 1000 veces, obtenemos el siguiente histograma.

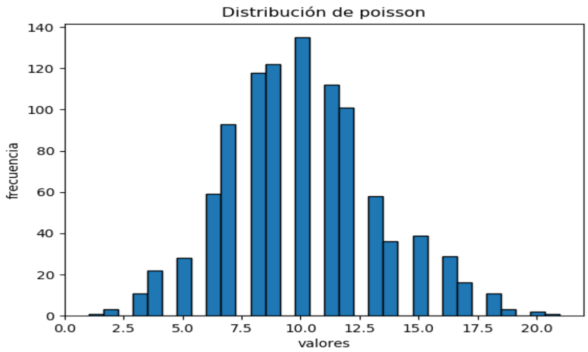


Figura 19: Histograma con distribución de poisson.

3.7. *Experimento 7-Empírica Discreta*

En este método definimos para cada una de las corridas un número aleatorio r entre 0 y 1. Cada número se corresponderá con un sector definido por las probabilidades de cada elemento.

Estos sectores se representan a través de la sumatoria de las probabilidades individuales. Por lo tanto, si un elemento se ubica en la sumatoria de i probabilidades, este corresponderá, al elemento i en cuestión. Luego de mil corridas con este algoritmo, podemos notar como los elementos con mayor probabilidad de ocurrencias se presentan mas frecuentemente aplicando el algoritmo. Por ende, el histograma refleja una mayor frecuencia para los elementos con mayor probabilidad (como, en este caso, los elementos 1,3,5 y 8) Posteriormente calculamos la esperanza y la varianza usando la librería nombrada en la sección 1.3.7.

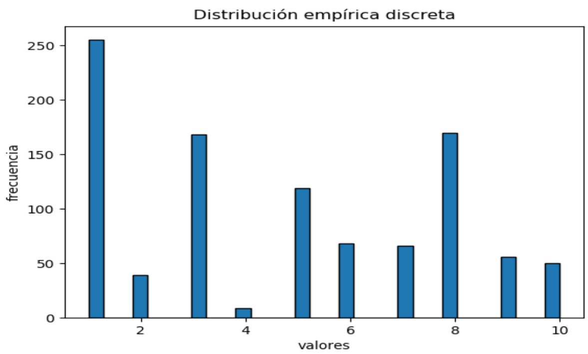


Figura 20: Histograma con distribución empírica discreta.

4. Conclusiones

Como podemos notar al hacer la comparación entre los histogramas del marco teórico y aquellos que resultaron de los experimentos enunciados en los casos de estudio, estos se asemejan entre sí.

Esto quiere decir, que con la implementación de la función, obtenida de aplicar el método de la sección 1.2, en cada una de las distribuciones obtenemos realmente valores de x_i válidos para cada una de ellas.

En cuanto a la comparación de las esperanzas podemos notar como los valores teóricos se asemejan a los valores obtenidos tras realizar las respectivas formulas de esperanzas para cada una de las distribuciones.El mismo análisis podemos realizar para las varianzas.

5. Apéndice 1

El código en Python y el excel descriptos en este proyecto pueden ser consultados en el siguiente URL.
https://github.com/Gabriel713/TP2$_{D}istribuciones$