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

Mateo, Lara
Cátedra Simulación
Universidad Tecnológica Nacional
Zeballos 1341, S2000
`1laramateo@gmail.com`

Retrivi, Vittorio
Cátedra Simulación
Universidad Tecnológica Nacional
Zeballos 1341, S2000
`retrovitto@gmail.com`

10 de Junio de 2020

ABSTRACT

Se presenta un estudio sobre las diferentes distribuciones de probabilidad junto a una guía acerca de como generar números pseudoaleatorios con estas distribuciones.

***K*eywords** Simulación · Generadores pseudoaleatorios · Distribuciones de Probabilidad · Uniforme · Exponencial · Gamma · Normal · Pascal · Binomial · Hipergeométrica · Poisson · Empírica

1 Introducción

Cuando se establecen las bases racionales subyacentes al empleo de los métodos existentes para generar valores de variables estocásticas en una computadora digital, se parte de dos problemas un tanto divergentes. Estos dos problemas de tipo distinto se pueden clasificar como deterministas, es decir, no probabilísticos o bien, como estocásticos. Se ha popularizado el término Monte Carlo como sinónimo para el concepto simulación de procesos estocásticos. Sin embargo, conviene anotar que en el pasado, este término se aplicó tan sólo al emplear los métodos de simulación estocástica para la resolución de problemas estrictamente deterministas.

En un principio, los métodos de simulación estocástica fueron aplicados por los matemáticos y científicos relacionados con las áreas de la Física, para resolver ciertos problemas deterministas que se podían expresar mediante ecuaciones matemáticos de importancia reconocida, existe la posibilidad de que, una vez encontrado un proceso estocástico cuya distribución de probabilidad o cuyos parámetros satisfagan las propiedades matemáticas que se requieran, quedan resueltas las ecuaciones que caracterizan a estos problemas. Aún más, desde un punto de vista computacional pudiera resultar más eficiente construir tal tipo de procesos a la vez que generar su estadística, empleando la computadora en lugar de seguir los métodos convencionales.

2 Marco Teórico y Desarrollo

Antes de aventurarnos al mundo de la probabilidad y estadística, resulta conveniente clarificar el lenguaje específico que conlleva esta disciplina, a continuación definiremos conceptos múltiples que nos ayudaran a afianzarnos con la terminología y teoría.

2.1 Distribución 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 la probabilidad de que dicho suceso ocurra. La distribución de probabilidad está definida sobre el conjunto de todos los sucesos y cada uno de los sucesos es el rango de valores de la variable aleatoria. También puede decirse que tiene una relación estrecha con las distribuciones de frecuencia. De hecho, una distribución de probabilidades puede comprenderse como una frecuencia teórica, ya que describe cómo se espera que varíen los resultados.

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.

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.

De estas se diferencian:

1. **Variable aleatoria discreta:** Es aquella que solo toma ciertos valores (frecuentemente enteros) y que resulta principalmente del conteo realizado.
2. **Variable aleatoria continua:** Es aquella que resulta generalmente de la medición y puede tomar cualquier valor dentro de un intervalo dado

Nosotros vamos a trabajar con las siguientes distribuciones:

Distribución de Probabilidad Continua

- Distribución de Probabilidad de Uniforme
- Distribución de Probabilidad Exponencial
- Distribución de Probabilidad Gamma
- Distribución de Probabilidad Normal

Distribución de Probabilidad Discreta

- Distribución de Probabilidad de Pascal
- Distribución de Probabilidad Binomial
- Distribución de Probabilidad Hiper-geométrica
- Distribución de Probabilidad de Poisson
- Distribución de Probabilidad Empírica Discreta

2.1.1 Esperanza Matemática o Media

La media es un promedio estándar que a menudo se denomina promedio.

Sea X una variable aleatoria con un número finito de resultados finitos $x_1, x_2, ..., x_k$ que ocurren con probabilidades $p_1, p_2, ..., p_k$, respectivamente. La expectativa de X se define como:

$$\mathrm{E}(x) = \mu = \sum_{i=1}^{k} x_i\, p_i = x_1p_1 + x_2p_2 + \cdots + x_kp_k. \tag{1}$$

Dado que la suma de todas las probabilidades p_i es 1 ($p_1 + p_2 + ... + p_k = 1$), el valor esperado es el promedio ponderado de x_i valores , siendo los valores p_i los pesos.

Si todos los resultados x_i son equiprobables (es decir, $p_1 = p_2 = ... = p_k$, entonces el promedio ponderado se convierte en el promedio simple. Si los resultados x_i no son equiparables, entonces el promedio simple debe ser reemplazado por el promedio ponderado, que tiene en cuenta el hecho de que algunos resultados son más probables que otros.

2.1.2 Varianza

En teoría de la probabilidad y estadística, la varianza es la expectativa de la desviación al cuadrado de una variable aleatoria de su media. Informalmente, mide hasta qué punto un conjunto de números (aleatorios) se extienden desde su valor promedio.

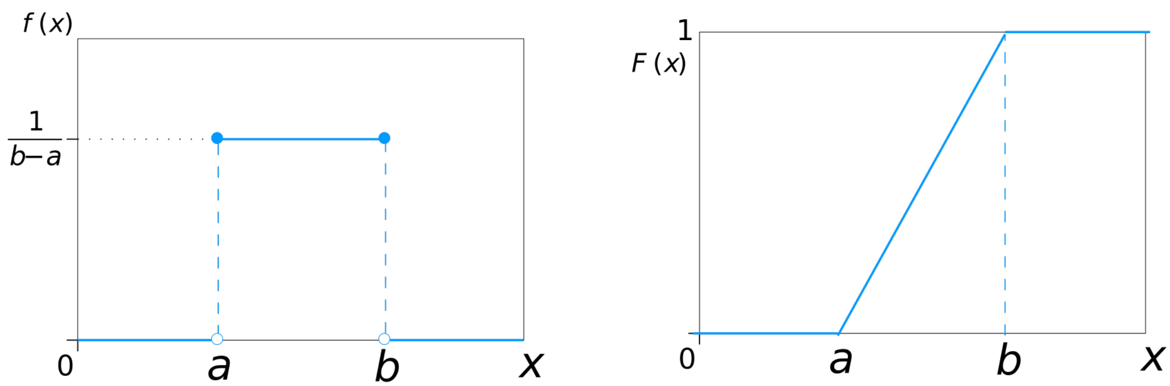
La variación tiene un papel central en la estadística, donde algunas ideas que la usan incluyen estadística descriptiva, inferencia estadística, prueba de hipótesis, bondad de ajuste y muestreo de Monte Carlo. La varianza es una herramienta importante en las ciencias, donde el análisis estadístico de datos es común.

La varianza de una variable aleatoria X es el valor esperado de la desviación al cuadrado de la media de X,$\mu = E(X)$:

$$\mathrm{V}(X) = \mathrm{E}\left[(X-\mu)^2\right]. \tag{2}$$

2.2 Distribución de Probabilidad Uniforme

En teoría de probabilidad y estadística, la distribución uniforme continua es una familia de distribuciones de probabilidad para variables aleatorias continuas, tales que para cada miembro de la familia, todos los intervalos de igual longitud en la distribución en su rango son igualmente probables.



(a) Gráfica de función de densidad de probabilidad

(b) Gráfica de función de distribución acumulada

Figure 1: Ejemplo de distribución Uniforme

2.2.1 Parámetros de entrada

El dominio está definido por dos parámetros, a y b, que son sus valores mínimo y máximo.
La distribución es comúnmente escrita de forma abreviada como $U(a, b)$.

2.2.2 Función de probabilidad

Matemáticamente, la función de densidad de probabilidad de la distribución uniforme continua se define:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{para } a \leq x \leq b, \\ 0 & \text{para } x < a \text{ o } x > b, \end{cases} \tag{3}$$

2.2.3 Esperanza Matemática y Varianza

$$E(x) = \mu = \int_a^b \frac{1}{b-a} x dx = \frac{b+a}{2} \tag{4}$$

$$V(x) = \sigma^2 = \int_a^b \frac{(x-\mu)^2}{b-a} dx = \frac{(b-a)^2}{12} \tag{5}$$

2.2.4 Función de distribución acumulada

La función de la distribución acumulada $F(x)$, para una variable aleatoria X uniformemente distribuida, se puede representar por:

$$F(x) = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a} \quad 0 \leq F(x) \leq 1 \tag{6}$$

También podemos calcularla utilizando los valores de la Esperanza Matemática y Varianza.

$$F(x) = \begin{cases} 0 & \text{para } x - \mu < -\sigma\sqrt{3} \\ \frac{1}{2}\left(\frac{x-\mu}{\sigma\sqrt{3}} + 1\right) & \text{para } -\sigma\sqrt{3} \leq x - \mu < \sigma\sqrt{3} \\ 1 & \text{para } x - \mu \geq \sigma\sqrt{3} \end{cases} \tag{7}$$

2.2.5 Generación de números aleatorios: Transformación inversa

Desarrollo para obtenerlos

Partiendo de la función de la distribución acumulada $F(x)$:

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

Llegando finalmente a

$$\boldsymbol{x = r(b-a) + a \quad 0 \leq r \leq 1} \tag{8}$$

2.2.6 Implementación en código

```
def randomUniform(a, b):
    return rand() * (b - a) + a
```

2.2.7 Test 1: Comparación de Gráficas

Comparamos la gráfica que resultó de los números generados por nuestro código con los generados por una librería de Python.

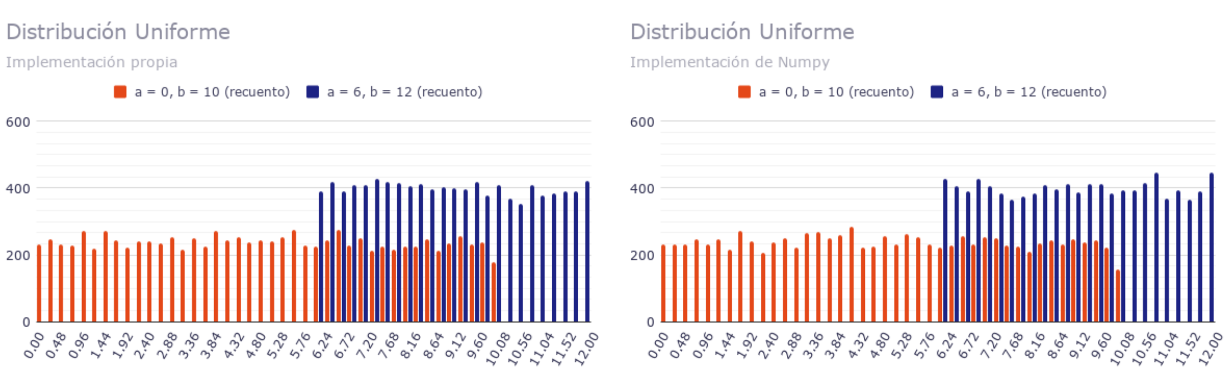


Figure 2: Comparación de la distribución Uniforme

2.2.8 Test 2: Comparación de los parámetros estadísticos

Generamos una serie de 10000 números pseudoaleatorios con la función en Python expuesta anteriormente y utilizando los parámetros $a = 6$, $b = 12$.

Luego:

$$\mu = \frac{b+a}{2} = \frac{12+6}{2} = 9$$

$$\sigma^2 = \frac{(b+a)^2}{12} = \frac{(12-6)^2}{12} = 3$$

Tras colocar la serie obtenida en una planilla de cálculo, calculamos los mismos parámetros, y los comparamos

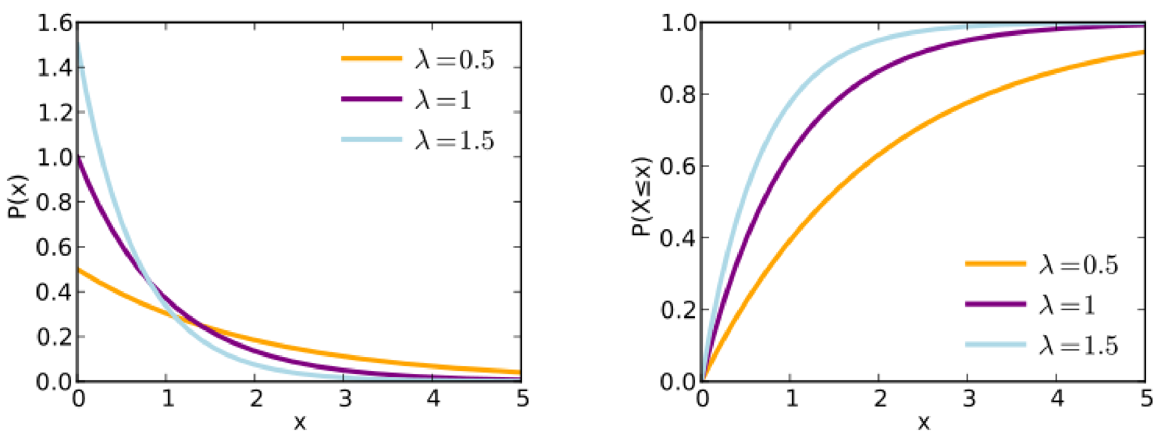
- $\mu = 8.96978392$
- $\sigma^2 = 3.003162722$

En consecuencia, aprueba el test.

2.3 Distribución de Probabilidad Exponencial

La distribución exponencial es la distribución de probabilidad del tiempo entre eventos en un proceso de punto de Poisson, es decir, un proceso en el que los eventos ocurren de forma continua e independiente a una tasa promedio constante. Es un caso particular de la distribución gamma. Es el análogo continuo de la distribución geométrica, y tiene la propiedad clave de no tener memoria. Además de ser utilizado para el análisis de procesos puntuales de Poisson, la distribución se encuentra presente en otros contextos.

La distribución exponencial no es lo mismo que la familia exponencial de distribuciones. Esta última es una gran clase de distribuciones de probabilidad en la cual se incluye la distribución exponencial como uno de sus miembros, pero además en ella se encuentra también a la distribución normal, la distribución binomial, la distribución gamma, poisson, entre otras.



(a) Gráfica de función de densidad de probabilidad

(b) Gráfica de función de distribución acumulada

Figure 3: Ejemplo de distribución Exponencial

2.3.1 Parámetro de entrada

La distribución es escrita comunmente y de forma abreviada como $\varepsilon(\lambda)$ o $Exp(\lambda)$

El único parámetro de entrada de esta distribución de probabilidad es $\lambda > 0$.

2.3.2 Función de probabilidad

Matemáticamente, la función de densidad de probabilidad de la distribución de Probabilidad Exponencial se define:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0 & x < 0. \end{cases} \tag{9}$$

2.3.3 Esperanza Matemática y Varianza

$$E(x) = \mu = \int_0^\infty x\lambda e^{-\lambda x}dx = \frac{1}{\lambda} \tag{10}$$

$$V(x) = \sigma^2 = \int_0^\infty (x - \frac{1}{\lambda})^2 \lambda e^{-\lambda x}dx = \frac{1}{\lambda^2} \tag{11}$$

notamos que

$$V(x) = \sigma^2 = \mu^2 \tag{12}$$

2.3.4 Función de distribución acumulada

La función de distribución acumulativa de x está dada por:

$$F(x) = \int_0^x \lambda e^{-\lambda t}dt = 1 - e^{\lambda x} \tag{13}$$

2.3.5 Generación de números aleatorios: Transformación inversa

Desarrollo para obtenerlos

Debido a la simetría que existe entre la distribución uniforme sigue que la intercambiabilidad de $F(x)$ y $1 - F(x)$. Por lo tanto,

$$r = e^{-\lambda x} \tag{14}$$

y consecuentemente

$$\mathbf{x = -(\frac{1}{\lambda}) \log r = -\mu \log r \quad 0 \leq r \leq 1} \tag{15}$$

2.3.6 Implementación en código

```
def randomExponential(lamb):
    return -(1 / lamb) * (math.log(rand()))
```

2.3.7 Test 1: Comparación de Gráficas

Comparamos la gráfica que resultó de los números generados por nuestro código con los generados por una librería de Python.

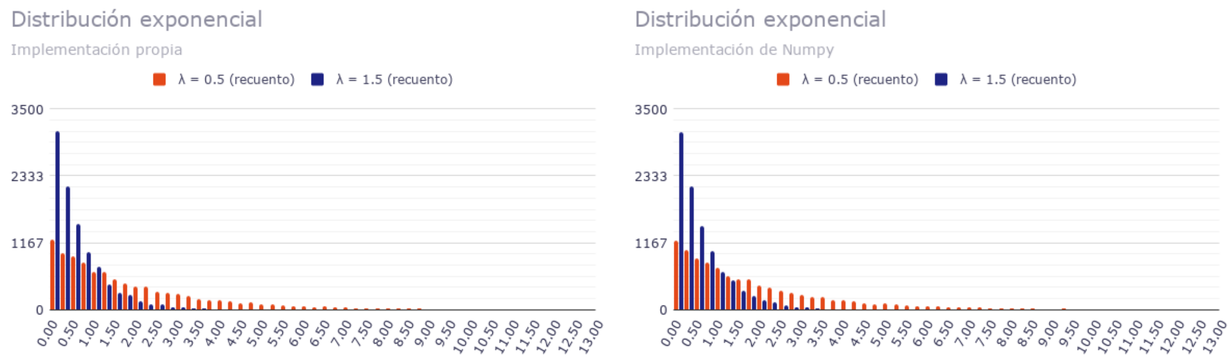


Figure 4: Comparación de la distribución Exponencial

2.3.8 Test 2: Comparación de los parámetros estadísticos

Generamos una serie de 10000 números pseudoaleatorios con la función en Python expuesta anteriormente y utilizando el parámetro $\lambda = 1.5$.

Luego:

$$\mu = \frac{1}{\lambda} = \frac{1}{1.5} = \frac{2}{3} = 0.\overline{6}$$

$$\sigma^2 = \mu^2 = 0.\overline{4}$$

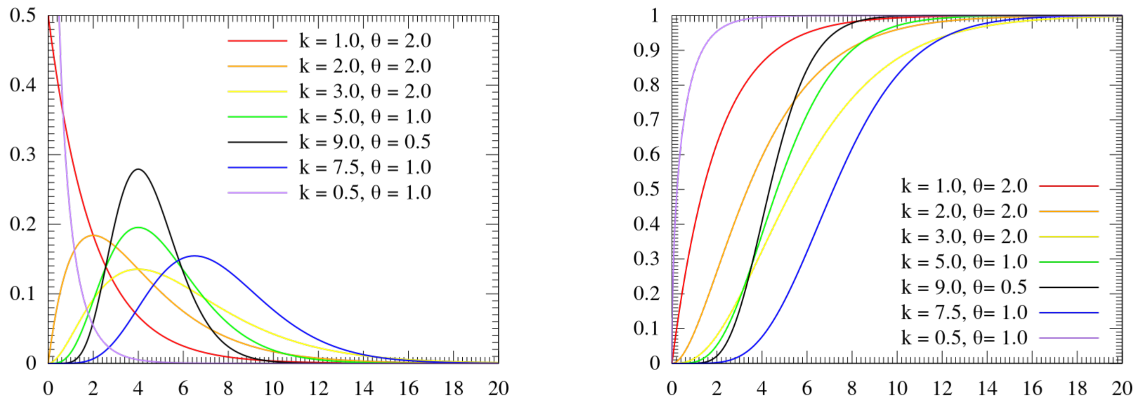
Tras colocar la serie obtenida en una planilla de cálculo, calculamos los mismos parámetros, y los comparamos

- $\mu = 0.6688043137$
- $\sigma^2 = 0.4421217418$

En consecuencia, aprueba el test.

2.4 Distribución de Probabilidad Gamma

La distribución gamma es una familia de dos parámetros de distribuciones de probabilidad continua.



(a) Gráfica de función de densidad de probabilidad

(b) Gráfica de función de distribución acumulada

Figure 5: Ejemplo de distribución Gamma

2.4.1 Parámetros de entrada

La distribución gamma se puede parametrizar en términos de un parámetro de forma k y un parámetro de escala inversa $\alpha = 1/\theta$ denominado parámetro de velocidad . Una variable aleatoria X que está distribuida en gamma con forma k y velocidad α se denota: $X \sim \Gamma(k, \alpha) \equiv \mathrm{Gamma}(k, \alpha)$

2.4.2 Función de probabilidad

La distribución Gamma está descrita por la siguiente función de densidad:

$$f(x; k, \alpha) = \frac{\alpha^k x^{k-1} e^{-\alpha x}}{\Gamma(k)} \quad \text{para} x > 0 \quad k, \alpha > 0, \tag{16}$$

Donde $\Gamma(\alpha)$ es la función gamma. Para todos los enteros positivos, $\Gamma(\alpha) = (\alpha - 1)!$.

2.4.3 Esperanza Matemática y Varianza

$$E(x) = \frac{k}{\alpha} \tag{17}$$

$$V(x) = \frac{k}{\alpha^2} \tag{18}$$

Notemos que, si $k = 1$, entonces la distribución gamma es idéntica a la distribución exponencial; mientras que si k es un entero positivo, entonces la misma coincidirá con la distribución Erlang. Además, a medida que k se incrementa, la distribución gamma tiende de forma asintótica a la distribución normal.

2.4.4 Función de distribución acumulada

La función de distribución acumulada para dicha distribución se corresponde con la función gamma regularizada:

$$F(x; \alpha, \beta) = \int_0^x f(u; \alpha, \beta)\, du = \frac{\gamma(\alpha, \beta x)}{\Gamma(\alpha)}, \tag{19}$$

2.4.5 Generación de números aleatorios

Para generar valores de variables aleatorias con distribución gamma y con un valore esperado y variancia dados, podemos emplear las siguientes fórmulas a fin de determinar los parámetros de la función de densidad de probabilidad gamma:

$$\alpha = \frac{E(X)}{V(X)}$$

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

Dado a que la función de distribución acumulada para la función gamma no se encuentra expresada de una forma explícita, no podemos hacer uso de la función inversa para la generación de números aleatorios. Debemos entonces, considerar un método alternativo para la generación de los mismos. Para ello, es posible hacer uso del mismo proceso aleatorio sobre el cual se basa la distribución de Erlang.

El método consiste en tomar la suma de k valores de variables aleatorias con distribución exponencial $x_1, x_2, ..., x_k$ cuyo valor esperado es el mismo e igual a $1/\alpha$. En consecuencia, el valor de la variable aleatoria (Erlang) x se puede expresar como:

$$\boldsymbol{x = \sum_{i=1}^{k} x_i = -\frac{1}{\alpha}\sum_{i=1}^{k} \log r_i} \tag{20}$$

Notese que en esta implementación, solo se consideran los valores en los cuales k es un entero positivo.

2.4.6 Implementación en código

```
def randomGamma(alpha, k):
    tr = 1.0
    for _ in range(0, k):
        tr *= rand()
    return -(1 / alpha) * math.log(tr)
```

2.4.7 Test 1: Comparación de Gráficas

Comparamos la gráfica que resultó de los números generados por nuestro código con los generados por una librería de Python.

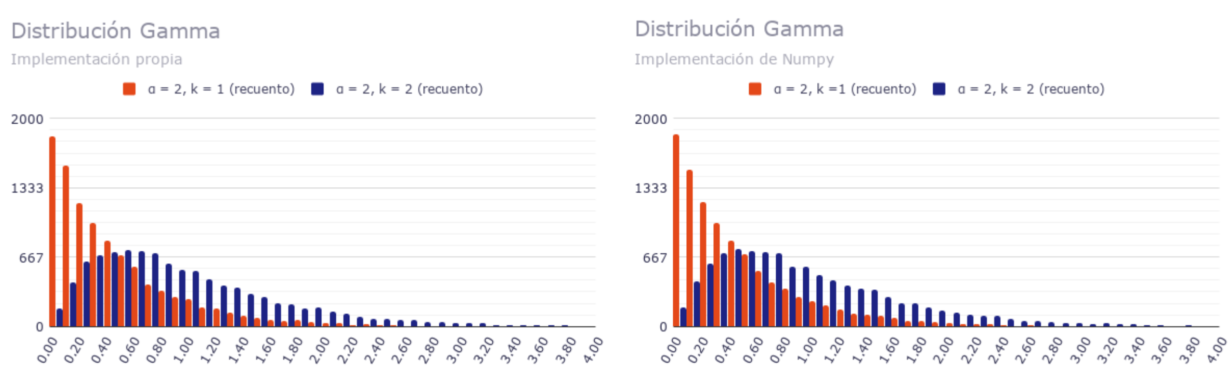


Figure 6: Comparación de la distribución Gamma

2.4.8 Aplicaciones en diferentes disciplinas

La distribución gamma se ha utilizado para modelar el tamaño de las reclamaciones de seguro y las precipitaciones, lo que significa que las reclamaciones de seguro agregadas y la cantidad de lluvia acumulada en un depósito se modelan mediante un proceso gamma, al igual que la distribución exponencial genera un proceso de Poisson.

La distribución gamma también se usa para modelar errores en modelos de regresión de Poisson de niveles múltiples, porque la combinación de la distribución de Poisson y una distribución gamma es una distribución binomial negativa.

En la comunicación inalámbrica, la distribución gamma se usa para modelar el desvanecimiento de múltiples rutas de la potencia de la señal.

En oncología, la distribución por edad de la incidencia de cáncer a menudo sigue la distribución gamma, mientras que los parámetros de forma y escala predicen, respectivamente, el número de eventos de impulso y el intervalo de tiempo entre ellos.

En neurociencia, la distribución gamma se usa a menudo para describir la distribución de los intervalos entre espigas.

En la expresión de genes bacterianos, el número de copias de una proteína expresada constitutivamente a menudo sigue la distribución gamma, donde el parámetro de escala y forma son, respectivamente, el número medio de explosiones por ciclo celular y el número medio de moléculas de proteína producidas por un solo ARNm durante Es toda la vida.

En genómica, la distribución gamma se aplicó en el paso de llamada máxima (es decir, en reconocimiento de señal) en el análisis de datos ChIP-chip y ChIP-seq.

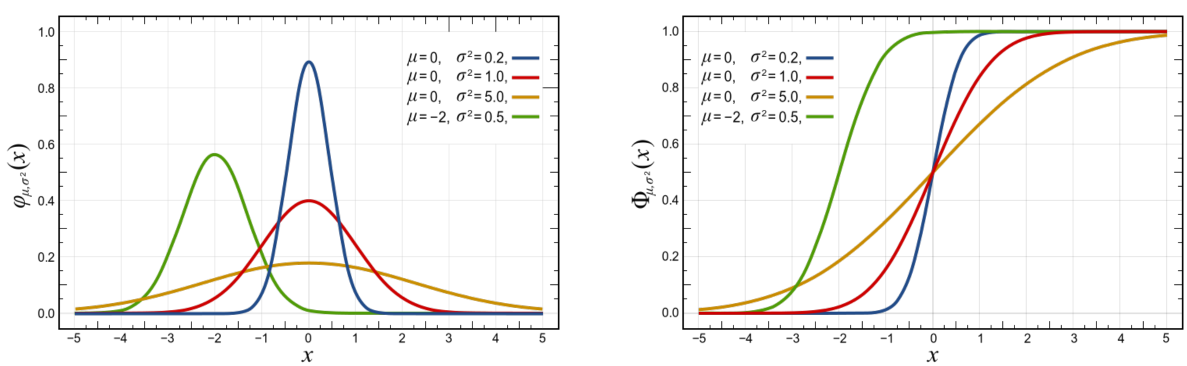
La distribución gamma se usa ampliamente como un conjugado previo en las estadísticas bayesianas. Es el conjugado anterior para la precisión (es decir, inversa de la varianza) de una distribución normal. También es el conjugado anterior para la distribución exponencial.

2.5 Distribución de Probabilidad Normal

Se llama distribución normal, distribución de Gauss, distribución gaussiana o distribución de Laplace-Gauss, a una de las distribuciones de probabilidad de variable continua que con más frecuencia aparece en estadística y en la teoría de probabilidades.

La gráfica de su función de densidad tiene una forma acampanada y es simétrica respecto de un determinado parámetro estadístico. Esta curva se conoce como campana de Gauss y es el gráfico de una función gaussiana.

La importancia de esta distribución radica en que permite modelar numerosos fenómenos naturales, sociales y psicológicos. Mientras que los mecanismos que subyacen a gran parte de este tipo de fenómenos son desconocidos, por la enorme cantidad de variables incontrolables que en ellos intervienen, el uso del modelo normal puede justificarse asumiendo que cada observación se obtiene como la suma de unas pocas causas independientes.



(a) Gráfica de función de densidad de probabilidad

(b) Gráfica de función de distribución acumulada

Figure 7: Ejemplo de distribución Normal

2.5.1 Parámetros de entrada

El parámetro μ es la media o expectativa de la distribución (y también su mediana y modo); y σ es su desviación estándar. La varianza de la distribución es σ^2.
Se dice que una variable aleatoria con una distribución gaussiana se distribuye normalmente y se denomina desviación normal.
La distribución es a menudo escrita en forma abreviada como $N(\mu, \sigma^2)$ o $\mathcal{N}(\mu, \sigma^2)$.

2.5.2 Función de probabilidad

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{21}$$

2.5.3 Esperanza Matemática y Varianza

$$E(x) = \mu \tag{22}$$

$$V(x) = \sigma^2 \tag{23}$$

2.5.4 Función de distribución acumulada

$$\Phi_{\mu,\sigma^2}(x) = \int_{-\infty}^{x} \varphi_{\mu,\sigma^2}(u)\,du = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(u-\mu)^2}{2\sigma^2}}\,du, \quad x \in \mathbb{R}. \tag{24}$$

donde:

- μ es la media (también puede ser la mediana, la moda o el valor esperado, según aplique)
- σ es la desviación típica [estándar es un anglicismo]
- σ^2 la varianza
- φ representa la función de densidad de probabilidad

2.5.5 Distribución Normal Estándar

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

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \quad -\infty < z < \infty \tag{25}$$

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

$$z = \frac{x-\mu}{\sigma} \tag{26}$$

2.5.6 Generación de números aleatorios: TLC

Puesto que no es posible expresar la distribución acumulada de la distribución normal en forma explícita, entonces no es posible utilizar para la generación de números al azar, el método de la transformada inversa.

En lugar de este método, se puede hacer uso del teorema del límite central, el cual establece que la suma de n variables aleatorias independientes se aproxima a una distribución normal a medida que n se aproxima al infinito.

Desarrollo para obtenerlos

Si $r_1, r_2, ..., r_i$ representan variables aleatorias independientes, cada una de las cuales posee la misma distribución de probabilidad caracterizada por por $E(r_i) = \theta$ y $V(r_i) = \sigma^2$, entonces

$$\lim_{N\to\infty} P\left[a < \frac{\sum_{i=1}^{N} r_i - N\theta}{\sqrt{N}\sigma} < b\right] = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{1}{2}z^2} dz \tag{27}$$

donde:

$$E(\sum_{i=1}^{N} r_i) = N\theta, \tag{28}$$

$$V(\sum_{i=1}^{N} r_i) = N\sigma^2, \tag{29}$$

$$z = \frac{\sum_{i=1}^{N} r_i - N\theta}{\sigma\sqrt{N}} \tag{30}$$

El procedimiento para simular valores normales utilizando computadoras requiere el uso de la suma de K valores de variable aleatoria distribuidos uniformemente; esto es, la suma de $r_1, r_2, ..., r_k$, con cada r_i definida en el intervalo $0 < r_i < 1$.

Aplicando la convención notacional de la forma matemática del TLC (Teorema del Límite Central), así como nuestros conocimientos previos de la distribución uniforme, encontramos que:

$$\theta = \frac{a+b}{2} = \frac{0+1}{2} = \frac{1}{2}, \tag{31}$$

$$\sigma = \frac{b-a}{\sqrt{12}} = \frac{1}{\sqrt{12}}, \tag{32}$$

$$z = \frac{\sum_{i=i}^{K} r_{i-\frac{K}{2}}}{\sqrt{\frac{K}{12}}} \tag{33}$$

Por definición, z es un valor de variable aleatoria con distribución normal estándar que se puede escribir en la forma sugerida por la ecuación (26), donde x es un valor de variable aleatoria distribuido en forma normal que se va a simular, con media μ y varianza σ^2.

Igualando las ecuaciones (33) y (26) obtendremos:

$$\frac{x-\mu}{\sigma} = \frac{\sum_{i=1}^{K} r_u - \frac{K}{2}}{\sqrt{\frac{K}{2}}}, \tag{34}$$

y resolviendo para x, se tiene que:

$$\boldsymbol{x = \sigma(\frac{12}{K})^{\frac{1}{2}}(\sum_{i=1}^{K} r_i - \frac{K}{2}) + \mu} \tag{35}$$

2.5.7 Implementación en código

```
def randomNormal(mu, sigma, k):
    sum = 0
    for _ in range(0, k):
        sum += rand()
    return sigma * math.sqrt((12 / k)) * (sum - (k / 2)) + mu
```

2.5.8 Test 1: Comparación de Gráficas

Comparamos la gráfica que resultó de los números generados por nuestro código con los generados por una librería de Python.

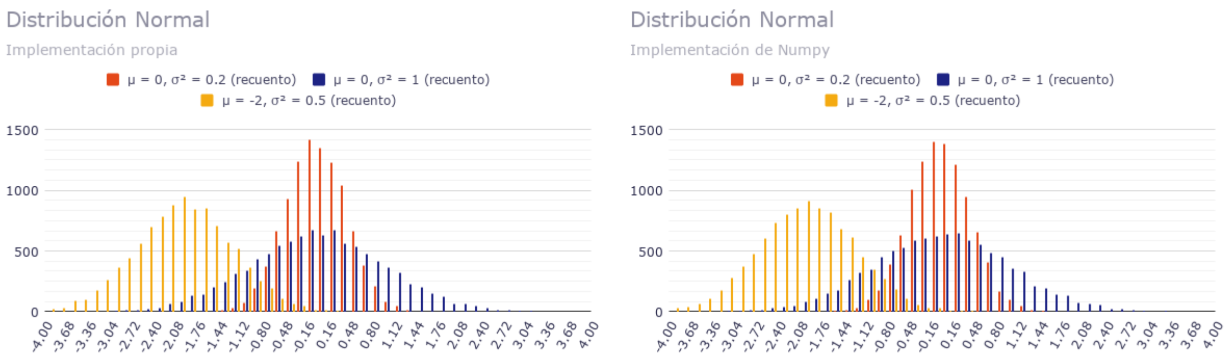


Figure 8: Comparación de la distribución Normal

2.5.9 Test 2: Comparación de los parámetros estadísticos

Generamos una serie de 10000 números pseudoaleatorios con la función en Python expuesta anteriormente y utilizando los parámetros $\mu = -2$, $\sigma^2 = 0.5$. Luego de colocar la serie obtenida en una planilla de cálculo, calculamos los mismos parámetros, y los comparamos

- $\mu = -1.992248532$
- $\sigma^2 = 0.5024611969$

En consecuencia, aprueba el test.

2.5.10 Aplicaciones en diferentes disciplinas

Algunos ejemplos de variables asociadas a fenómenos naturales que siguen el modelo de la normal son:

- Caracteres morfológicos de individuos como la estatura
- Caracteres fisiológicos como el efecto de un fármaco
- caracteres sociológicos como el consumo de cierto producto por un mismo grupo de individuos
- Caracteres psicológicos como el coficiente intelectual
- Nivel de ruido en telecomunicaciones
- Errores cometidos al medir ciertas magnitudes

La distribución normal también aparece en muchas áreas de la propia estadística. Por ejemplo, la distribución muestral de las medias muestrales es aproximadamente normal, cuando la distribución de la población de la cual se extrae la muestra no es normal. Además, la distribución normal maximiza la entropía entre todas las distribuciones con media y varianza conocidas, lo cual la convierte en la elección natural de la distribución subyacente a una lista de datos resumidos en términos de media muestral y varianza. La distribución normal es la más extendida en estadística y muchos tests estadísticos están basados en una "normalidad" más o menos justificada de la variable aleatoria bajo estudio.

En probabilidad, la distribución normal aparece como el límite de varias distribuciones de probabilidad continuas y discretas.

2.6 Distribución de Probabilidad Pascal

La distribución binomial negativa es una distribución de probabilidad discreta que modela el número de fallas en una secuencia de ensayos de Bernoulli independientes e idénticamente distribuidos antes de que ocurra un número específico (no aleatorio) de éxitos (denotado r). Por ejemplo, podemos definir lanzar un 6 en un dado como un éxito, y lanzar cualquier otro número como un fracaso, y preguntar cuántas tiradas fallidas ocurrirán antes de que veamos el tercer éxito (r = 3). En tal caso, la distribución de probabilidad del número de no-6 que aparecen será una distribución binomial negativa.

La distribución de Pascal (después de Blaise Pascal) y la distribución de Polya (para George Pólya) son casos especiales de la distribución binomial negativa. Una convención entre ingenieros, climatólogos y otros es usar "binomio negativo" o "Pascal" para el caso de un parámetro de tiempo de parada r con valor entero r, y usar "Polya" para el caso con valor real.

2.6.1 Parámetros de entrada

- $r > 0$: número de fallas hasta que se detiene el experimento
- $p \epsilon [0, 1]$:probabilidad de éxito en cada experimento

2.6.2 Función de probabilidad

Su función de probabilidad es

$$f(k; r, p) \equiv \Pr(X = k) = \binom{k + r - 1}{k} p^r (1 - p)^k \tag{36}$$

donde k es el número de fracasos

2.6.3 Esperanza Matemática y Varianza

$$E(x) = \mu = \frac{pr}{1 - p} \tag{37}$$

$$V(x) = \sigma^2 = \frac{pr}{(1 - p)^2} \tag{38}$$

2.6.4 Función de distribución acumulada

La función de distribución acumulativa se puede expresar en términos de la función beta incompleta regularizada:

$$F(k; r, p) \equiv \Pr(X \leq k) = 1 - I_p(k + 1, r) = I_{1-p}(r, k + 1) \tag{39}$$

También se puede expresar en términos de la función de distribución acumulativa de la distribución binomial.

$$F(k; r, p) = F_{binomial}(k; n = k + r, p) \tag{40}$$

2.6.5 Generación de números aleatorios: Implementación en código

```python
def randomPascal(k, q):
    multiplier = 1
    for _ in range(0, k):
        multiplier *= rand()
    x = math.log(multiplier) / math.log(q)
    return math.floor(x)
```

2.6.6 Test 1: Comparación de Gráficas

Comparamos la gráfica que resultó de los números generados por nuestro código con los generados por una librería de Python.

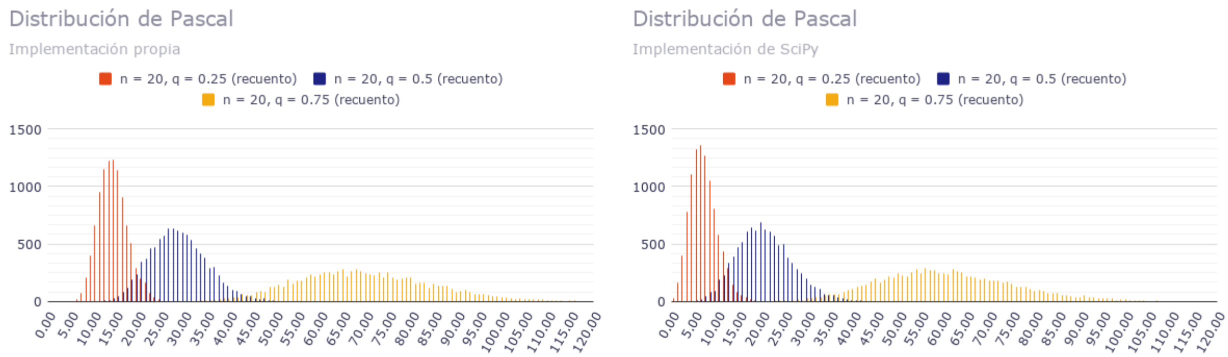


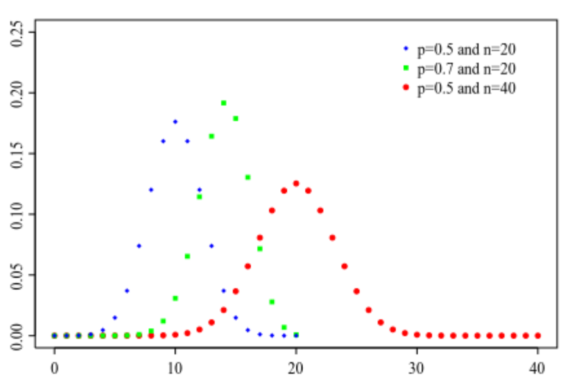
Figure 9: Comparación de la distribución de Pascal

Nótese que, si bien la forma de de las gráficas de la implementación propia en comparación con las generadas mediante SciPy coinciden, estas y por alguna extraña razón, se encuentran con un corrimiento horizontal hacia la derecha en aproximadamente 6 unidades.

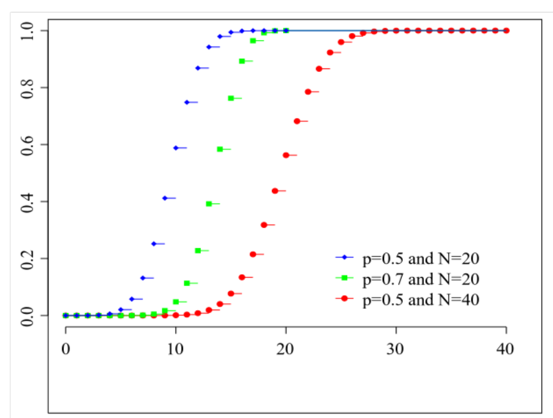
2.7 Distribución de Probabilidad Binomial

En estadística, la distribución binomial es una distribución de probabilidad discreta 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 el anterior experimento se repite n veces, de forma independiente, y se trata de calcular la probabilidad de un determinado número de éxitos. Para $n = 1$, la binomial se convierte, de hecho, en una distribución de Bernoulli.



(a) Gráfica de función de densidad de probabilidad



(b) Gráfica de función de distribución acumulada

Figure 10: Ejemplo de distribución Binomial

2.7.1 Parámetros de entrada

Una variable aleatoria que sigue una distribución binomial recibe los parámetros n y p, se escribe de la siguiente forma: $B(n, p)$, donde

- n: número de ensayos.
- p: probabilidad de éxito.

2.7.2 Función de probabilidad

Su función de probabilidad está dada por:

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad 0 \leq p \leq 1 \tag{41}$$

donde $x = \{0, 1, 2, \ldots, n\}$,

y siendo $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ las combinaciones de n sobre x (n elementos tomados de x en x).

2.7.3 Esperanza Matemática y Varianza

$$E(x) = np \tag{42}$$

$$V(x) = np(1-p) \tag{43}$$

2.7.4 Función de distribución acumulada

La función de distribución acumulada puede ser expresada como:

$$F(k; n, p) = \Pr(X \leq k) = \sum_{i=0}^{\lfloor k \rfloor} \binom{n}{i} p^i (1-p)^{n-i}, \tag{44}$$

donde k es el "piso" debajo de k, es decir, el mayor entero menor o igual a k.

2.7.5 Generación de números aleatorios

La presente distribución no posee función inversa de su correspondiente función de distribución acumulada. Sin embargo, existen muchas otras maneras de generar valores aleatorios que sigan dicha distribución.

Uno de ellos, se basa en la producción de ensayos de Bernoulli, siguiendo el método de rechazos.

Se fija $x_0 = 0$, y conociendo los valores de p y de n, se generan n números aleatorios y para cada uno de ellos se efectúa una prueba. La variable x_i se incrementa de acuerdo con el siguiente criterio:

$$x_i = x_{i-1} + 1 \quad si \; r_i \leq p$$

$$x_i = x_{i-1} \quad si \; r_i > p$$

Finalmente, luego de haberse generado n números aleatorios, el valor de x_n será igual al valor de la variable aleatoria con distribución binomial x.

2.7.6 Implementación en código

```python
def randomBinomial(n, p):
    x = 0
    for _ in range(0, n):
        if (rand() - p) <= 0:
            x += 1
    return x
```

2.7.7 Test 1: Comparación de Gráficas

Comparamos la gráfica que resultó de los números generados por nuestro código con los generados por una librería de Python.

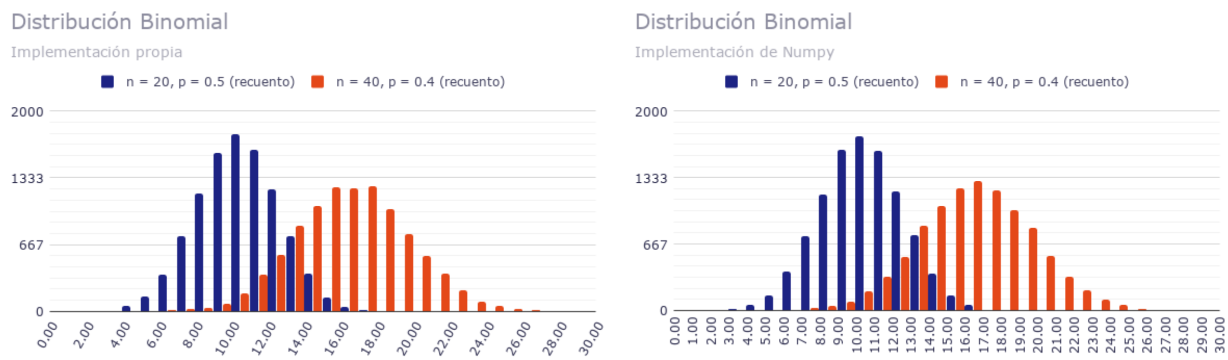


Figure 11: Comparación de la distribución Binomial

2.7.8 Test 2: Comparación de los parámetros estadísticos

Generamos una serie de 10000 números pseudoaleatorios con la función en Python expuesta anteriormente y utilizando los parámetros $n = 40$, $p = 0.4$.

Luego:

$$\mu = np = 40 * 0.4 = 16$$

$$\sigma^2 = np(1 - p) = 40 * 0.4(1 - 0.4) = 9.6$$

Tras colocar la serie obtenida en una planilla de cálculo, calculamos los mismos parámetros, y los comparamos

- $\mu = 16.0257$
- $\sigma^2 = 9.63443951$

En consecuencia, aprueba el test.

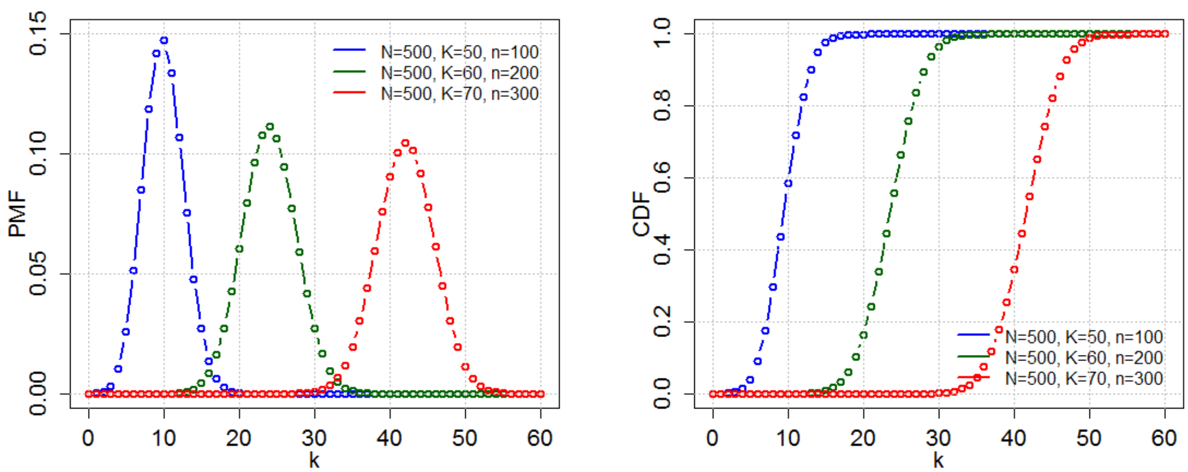
2.8 Distribución de Probabilidad Hipergeométrica

La distribución hipergeométrica es una distribución de probabilidad discreta que describe la probabilidad de k éxitos (sorteos aleatorios para los cuales el objeto dibujado tiene una característica específica) en sorteos n, sin reemplazo, de una población finita de tamaño N que contiene exactamente K objetos con esa característica, en donde cada sorteo es un éxito o un fracaso.

Por el contrario, la distribución binomial describe la probabilidad de éxito de k en sorteos de n con reemplazo.

En resumen, las siguientes condiciones caracterizan la distribución hipergeométrica:

- El resultado de cada sorteo (los elementos de la población que se está muestreando) se puede clasificar en una de dos categorías mutuamente excluyentes (por ejemplo, Aprobar / Reprobar o Empleado / Desempleado).
- La probabilidad de éxito cambia en cada sorteo, ya que cada sorteo disminuye la población (muestreo sin reemplazo de una población finita).



(a) Gráfica de función de densidad de probabilidad (b) Gráfica de función de distribución acumulada

Figure 12: Ejemplo de distribución hipergeométrica

2.8.1 Parámetros de entrada

- N es el tamaño de la población, $N\epsilon\,\{0, 1, 2, ...\}$
- K es el número de estados de éxito en la población,
  $K\epsilon\,\{0, 1, 2, ..., N\}$
- n es el número de sorteos (es decir, cantidad extraída en cada prueba),
  $n\epsilon\,\{0, 1, 2, ..., N\}$

2.8.2 Función de probabilidad

Una variable aleatoria X sigue la distribución hipergeométrica si su función de probabilidad viene dada por

$$p_X(k) = \Pr(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \tag{45}$$

donde

- k es el número de éxitos observados,
- $\binom{a}{b}$ es un coeficiente binomial.

2.8.3 Esperanza Matemática y Varianza

Sea $p = \frac{K}{N}$ la proporción de éxitos en el lote,

$$E(x) = n\frac{K}{N} = \frac{n}{p} \tag{46}$$

$$V(x) = n\frac{K}{N}\frac{(N-K)}{N}\frac{N-n}{N-1} = np(1-p)\frac{N-n}{N-1} \tag{47}$$

2.8.4 Generación de números aleatorios: Implementación en código

```
def randomHypergeometric(N, n, p):
    x = 0.0
    for _ in range(1, n):
        r = rand()
        if (r - p) <= 0:
            s = 1
            x += 1
        else:
            s = 0
        p = (N * p - s) / (N - 1)
        N -= 1
    return x
```

2.8.5 Test 1: Comparación de Gráficas

Comparamos la gráfica que resultó de los números generados por nuestro código con los generados por una librería de Python.

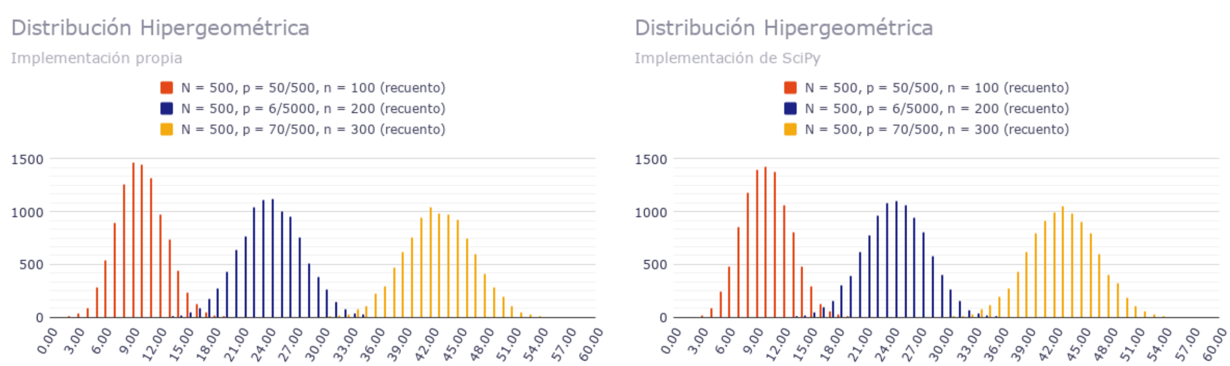
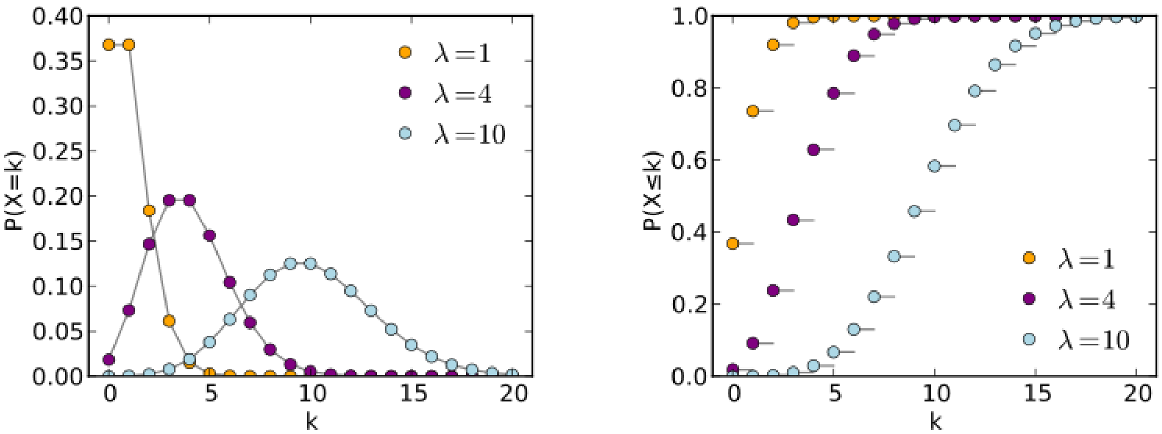


Figure 13: Comparación de la distribución Hipergeométrica

2.9 Distribución de Probabilidad Poisson

La distribución de Poisson es una distribución de probabilidad discreta que expresa, a partir de una frecuencia de ocurrencia media, la probabilidad de que ocurra un determinado número de eventos durante cierto período de tiempo. Concretamente, se especializa en la probabilidad de ocurrencia de sucesos con probabilidades muy pequeñas, o sucesos «raros».



(a) Gráfica de función de densidad de probabilidad

(b) Gráfica de función de distribución acumulada

Figure 14: Ejemplo de distribución de poisson

2.9.1 Parámetro de entrada

El único parámetro de entrada de esta distribución de probabilidad es $\lambda > 0$. Representa el número de veces que se espera que ocurra el fenómeno durante un intervalo dado. La distribución es comúnmente escrita de forma abreviada como $\mathrm{Pois}(\lambda)$.

2.9.2 Función de probabilidad

La probabilidad de x ocurrencias está dada por:

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, 2, ... \tag{48}$$

2.9.3 Esperanza Matemática y Varianza

La distribución posee la particularidad de que:

$$E(x) = V(x) = \lambda \tag{49}$$

2.9.4 Función de distribución acumulada

$$\frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor!} \text{para } k \geq 0 \tag{50}$$

donde $\Gamma(x, y)$ es la función gamma incompleta

2.9.5 Generación de números aleatorios

Dado a que no existe la inversa de la función de distribución acumulada para dicha distribución, no es posible realizar este método para la generación de números aleatorios. Sin embargo, podemos hacer uso de la relación que existe entre las distribuciones exponenciales y de Poisson.

Un método para la generación de valores de variables aleatorias con distribución Poisson es la de considerar la generación de intervalos $t_1, t_2, t_3, ...$, distribuidos en forma exponencial con un valor esperado igual a 1. Una vez generados estos intervalos aleatorios, se acumulan hasta que exceda al valor de λ.

A modo de ejemplo, se detalla la implementación en lenguaje Python.

2.9.6 Implementación en código

```
def randomPoisson(lamb):
    x = 0
    multiplier = 1.0
    b = math.exp(-lamb)
    while True:
        multiplier *= rand()
        if multiplier > b:
            x += 1
        else:
            break
    return x
```

2.9.7 Test 1: Comparación de Gráficas

Comparamos la gráfica que resultó de los números generados por nuestro código con los generados por una librería de Python.

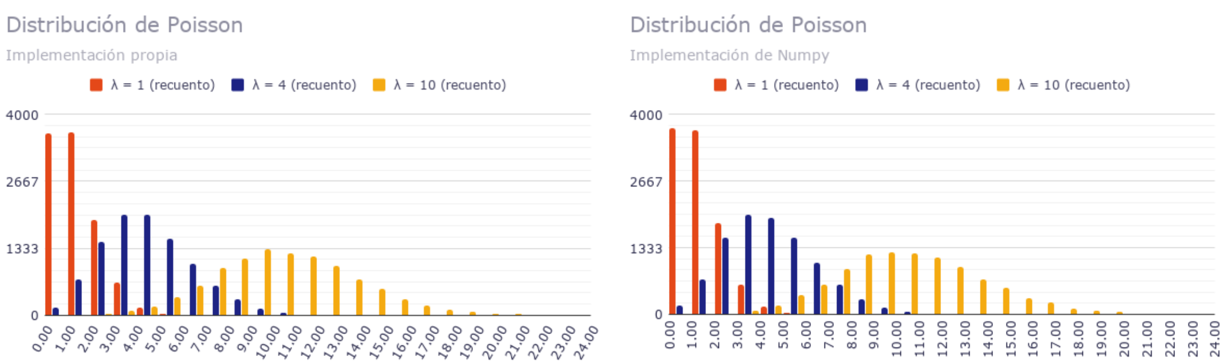


Figure 15: Comparación de la distribución de Poisson

2.9.8 Test 2: Comparación de los parámetros estadísticos

Generamos una serie de 10000 números pseudoaleatorios con la función en Python expuesta anteriormente y utilizando el parámetro $\lambda = 4$.

Luego:

$$\mu = \sigma^2 = \lambda = 4$$

Tras colocar la serie obtenida en una planilla de cálculo, calculamos los mismos parámetros, y los comparamos

- $\mu = 3.91479216$
- $\sigma^2 = 4.0028$

En consecuencia, aprueba el test.

2.10 Distribución de Probabilidad Empírica Discreta

Una función de distribución empírica es la función de distribución asociada con la medida empírica de una muestra. Esta función de distribución acumulativa es una función de paso que salta $\frac{1}{n}$ en cada uno de los n puntos de datos. Su valor en cualquier valor especificado de la variable medida es la fracción de observaciones de la variable medida que son menores o iguales al valor especificado.

La función de distribución empírica es una estimación de la función de distribución acumulativa que generó los puntos en la muestra. Converge con la probabilidad 1 a esa distribución subyacente, según el teorema de Glivenko-Cantelli. Existen varios resultados para cuantificar la tasa de convergencia de la función de distribución empírica con la función de distribución acumulativa subyacente.

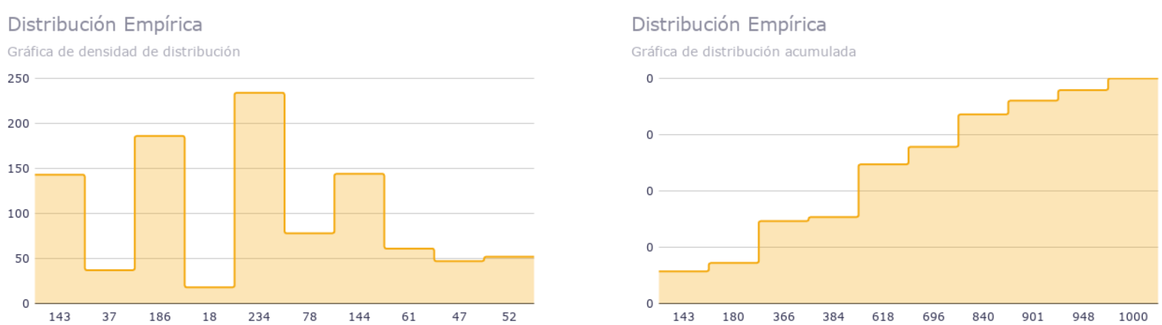


Figure 16: Ejemplo de distribución Empírica Discreta

2.10.1 Esperanza Matemática y Varianza

La media de la distribución empírica es un estimador imparcial de la media de la distribución de la población.

$$E(X) = \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right) \tag{51}$$

$$\begin{aligned} \mathrm{V}(X) &= \mathrm{E}\left[(X - \mathrm{E}[X])^2\right] \\ &= \mathrm{E}\left[(X - \bar{x})^2\right] \\ &= \frac{1}{n}\left(\sum_{i=1}^{n} (x_i - \bar{x})^2\right) \end{aligned} \tag{52}$$

2.10.2 Función de distribución acumulada

Supongamos que $(X_1, ..., X_n)$ sean variables aleatorias reales independientes, distribuidas idénticamente, con la función de distribución acumulativa común $F(t)$. Entonces la función de distribución empírica se define como

$$\widehat{F}_n(t) = \frac{\text{número de elementos en la muestra} \leq t}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{X_i \leq t} \tag{53}$$

2.10.3 Generación de números aleatorios: Implementación en código

```python
def randomEmpirical():
    p = [0.143, 0.037, 0.186, 0.018, 0.234, 0.078, 0.144, 0.061, 0.047, 0.052]
    r = rand()
    acum = 0
    x = 1
    for i in p:
        acum += i
        if (r < acum):
            break
        else:
            x += 1
    return x
```

3 Conclusiones

En este trabajo, se realizó una introducción hacia la generación de variables pseudoaleatorias que sigan una distribución de probabilidad conocida. Como podemos ver, haciendo referencia al código en Python expuesto, la implementación de estos algoritmos se realizan de una forma simple y logrando con ello resultados muy cercanos a los objetivos estudiados.

El uso de estos conocimientos podrá ser de gran utilidad a la hora de simular una amplia gama de sucesos del mundo real y de naturalezas diferentes.

References

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

[2] Thomas Naylor. *Técnicas de Simulación en Computadoras*.

[3] Raúl Coss Bu. *Simulación - Un enfoque práctico*. Limusa.

[4] Inverse transform. `https://en.wikipedia.org/wiki/Inverse_transform_sampling`.

[5] Variance. `https://en.wikipedia.org/wiki/Variance`.

[6] Probability distribution. `https://en.wikipedia.org/wiki/Probability_distribution`.

[7] Uniform distribution. `https://en.wikipedia.org/wiki/Uniform_distribution_(continuous)`.

[8] Distribución uniforme. `https://es.wikipedia.org/wiki/Distribución_uniforme_continua`.

[9] Exponential distribution. `https://en.wikipedia.org/wiki/Exponential_distribution`.

[10] Distribución exponencial. `https://es.wikipedia.org/wiki/Distribución_exponencial`.

[11] Gamma distribution. `https://en.wikipedia.org/wiki/Gamma_distribution#Properties`.

[12] Distribución normal. `https://es.wikipedia.org/wiki/Distribución_normal`.

[13] Negative binomial distribution. `https://en.wikipedia.org/wiki/Negative_binomial_distribution`.

[14] Binomial distribution. `https://en.wikipedia.org/wiki/Binomial_distribution`.

[15] Distribución binomial. `https://es.wikipedia.org/wiki/Distribución_binomial`.

[16] Hypergeometric distribution. `https://en.wikipedia.org/wiki/Hypergeometric_distribution`.

[17] Poisson distribution. `https://en.wikipedia.org/wiki/Poisson_distribution`.

[18] Distribución de poisson. `https://es.wikipedia.org/wiki/Distribución_de_Poisson`.

[19] Empirical distribution. `https://en.wikipedia.org/wiki/Empirical_distribution_function`.