# DISTRIBUCIONES DE PROBABILIDAD

Jonnathan Oswaldo Matute Curillo

Universidad Politécnica Salesiana

Investigación

Pareto - T de Student - Chi cuadrado – Beta – Gamma – Log normal

Resumen

Se presentan características de las distribuciones probabilísticos de algunos tipos de distribuciones tales como Pareto, T de Student, Chi cuadrado, Beta, Gamma y Log-normal. Para cada una de estas distribuciones proporcionan sus ecuaciones y ejemplos prácticos. Se contrastan las diferencias entre la aplicación adecuadas de cada una de estas distribuciones probabilísticas. Sabiendo que una distribución de probabilidad es aquella que permite establecer toda la gama de resultados probables de ocurrir en un experimento determinado. Es decir, describe la probabilidad de que un evento se realice en el futuro.

*Palabras clave*: Distribución Pareto, T de Student, Chi cuadrado, Beta, Gamma, Log-normal

# DISTRIBUCIONES DE PROBABILIDAD

Comenzaremos por definir lo que se entiende por variable aleatoria: una variable aleatoria X es una función que asocia un número real a cada punto del espacio muestral. Dado un experimento aleatorio cualquiera cuyos sucesos elementales posibles pueden identificarse fácilmente mediante un número real, se denomina Variable Aleatoria X, al conjunto de estos números. 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 todo el rango de valores de la variable aleatoria.

## Definición de algunos tipos de distribución

### Pareto.

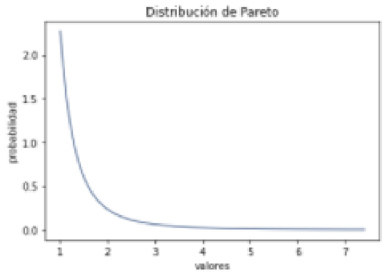
La Distribución de Pareto está dada por la función:

$$p(x; \alpha, k) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}}$$

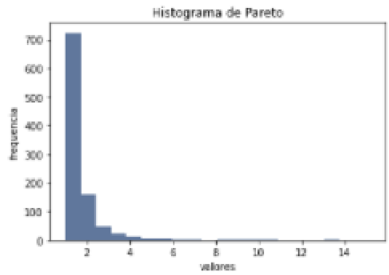
En dónde la variable $x \geqslant k$ y el parámetro $\alpha > 0$ son números reales. Esta distribución fue introducida por su inventor, Vilfredo Pareto, con el fin de explicar la distribución de los salarios en la sociedad. La Distribución de Pareto se describe a menudo como la base de la regla 80/20. Por ejemplo, el 80% de las quejas de los clientes con respecto al funcionamiento de su vehículo por lo general surgen del 20% de los componentes.

Ejemplo

```
# Graficando Pareto
k = 2.3 # parametro de forma.
pareto = stats.pareto(k)
x = np.linspace(pareto.ppf(0.01),
                pareto.ppf(0.99), 100)
fp = pareto.pdf(x) # Función de Probabilidad
plt.plot(x, fp)
plt.title('Distribución de Pareto')
plt.ylabel('probabilidad')
plt.xlabel('valores')
plt.show()
```




```
# histograma
aleatorios = pareto.rvs(1000) # genera aleatorios
cuenta, cajas, ignorar = plt.hist(aleatorios, 20)
plt.ylabel('frequencia')
plt.xlabel('valores')
plt.title('Histograma de Pareto')
plt.show()
```




***T de Student.***

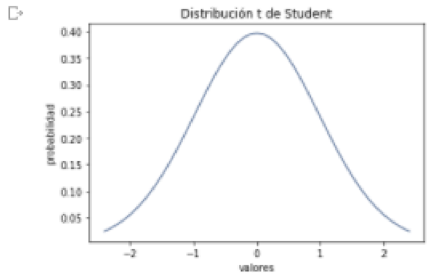
La Distribución t de Student está dada por la función:

$$p(t;n) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\left(1+\frac{t^2}{2}\right)^{-\frac{n+1}{2}}$$

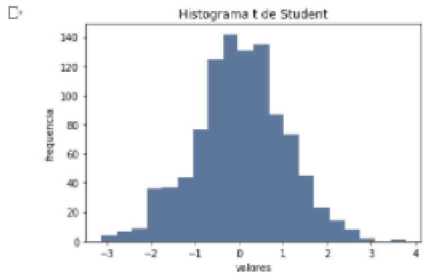
En dónde la variable t es un número real y el parámetro n es un número entero positivo. La Distribución t de Student es utilizada para probar si la diferencia entre las medias de dos muestras de observaciones es estadísticamente significativa. Por ejemplo, las alturas de una muestra aleatoria de los jugadores de baloncesto podrían compararse con las alturas de una muestra aleatoria de jugadores de fútbol; esta distribución nos podría ayudar a determinar si un grupo es significativamente más alto que el otro.

Ejemplo

```
# Graficando t de Student
df = 50 # parametro de forma.
t = stats.t(df)
x = np.linspace(t.ppf(0.01),
                t.ppf(0.99), 100)
fp = t.pdf(x) # Función de Probabilidad
plt.plot(x, fp)
plt.title('Distribución t de Student')
plt.ylabel('probabilidad')
plt.xlabel('valores')
plt.show()
```




```
# histograma
aleatorios = t.rvs(1000) # genera aleatorios
cuenta, cajas, ignorar = plt.hist(aleatorios, 20)
plt.ylabel('frequencia')
plt.xlabel('valores')
plt.title('Histograma t de Student')
plt.show()
```




***Chi cuadrado.***

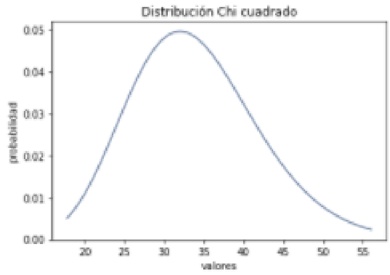
La Distribución Chi cuadrado está dada por la función:

$$p(x;n) = \frac{\left(\frac{x}{2}\right)^{\frac{n}{2}-1} e^{\frac{-x}{2}}}{2\Gamma\left(\frac{n}{2}\right)}$$

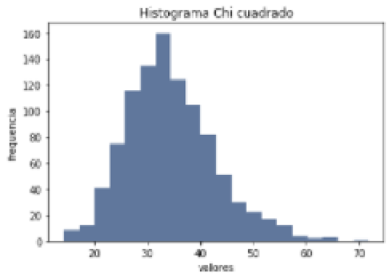
En dónde la variable x $\geqslant$ 0 y el parámetro n, el número de grados de libertad, es un número entero positivo. Una importante aplicación de la Distribución Chi cuadrado es que cuando un conjunto de datos es representado por un modelo teórico, esta distribución puede ser utilizada para controlar cuan bien se ajustan los valores predichos por el modelo, y los datos realmente observados.

Ejemplo

```
[16] # Graficando Chi cuadrado
     df = 34 # parametro de forma.
     chi2 = stats.chi2(df)
     x = np.linspace(chi2.ppf(0.01),
                     chi2.ppf(0.99), 100)
     fp = chi2.pdf(x) # Función de Probabilidad
     plt.plot(x, fp)
     plt.title('Distribución Chi cuadrado')
     plt.ylabel('probabilidad')
     plt.xlabel('valores')
     plt.show()
```




```
# histograma
aleatorios = chi2.rvs(1000) # genera aleatorios
cuenta, cajas, ignorar = plt.hist(aleatorios, 20)
plt.ylabel('frequencia')
plt.xlabel('valores')
plt.title('Histograma Chi cuadrado')
plt.show()
```




***Beta.***

La Distribución Beta está dada por la formula:

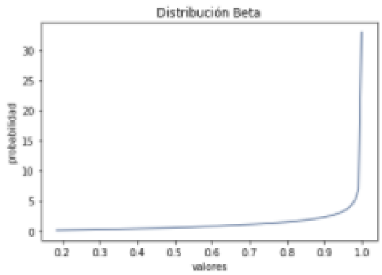
$$p(x; p, q) = \frac{1}{B(p, q)} x^{p-1} (1 - x)^{q-1}$$

En dónde los parámetros p y q son números reales positivos, la variable x satisface la condición $0 \leqslant x \leqslant 1$ y B (p, q) es la función beta. Las aplicaciones de la Distribución Beta incluyen el modelado de variables aleatorias que tienen un rango finito de a hasta b. Un ejemplo de ello es la distribución de los tiempos de actividad en las redes de proyectos. La

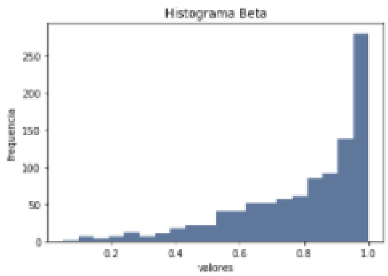
Distribución Beta se utiliza también con frecuencia como una probabilidad a priori para proporciones binomiales en el análisis bayesiano.

Ejemplo

```
# Graficando Beta
a, b = 2.3, 0.6 # parametros de forma.
beta = stats.beta(a, b)
x = np.linspace(beta.ppf(0.01),
                beta.ppf(0.99), 100)
fp = beta.pdf(x) # Función de Probabilidad
plt.plot(x, fp)
plt.title('Distribución Beta')
plt.ylabel('probabilidad')
plt.xlabel('valores')
plt.show()
```




```
# histograma
aleatorios = beta.rvs(1000) # genera aleatorios
cuenta, cajas, ignorar = plt.hist(aleatorios, 20)
plt.ylabel('frequencia')
plt.xlabel('valores')
plt.title('Histograma Beta')
plt.show()
```




***Gamma.***

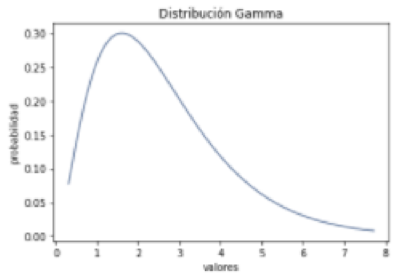
La Distribución Gamma está dada por la formula:

$$p(x; a, b) = \frac{a(ax)^{b-1} e^{-ax}}{\Gamma(b)}$$

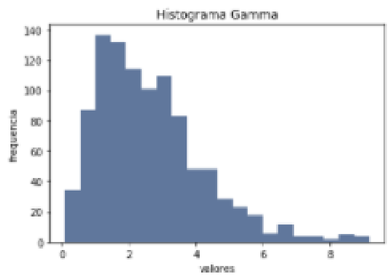
En dónde los parámetros a y b y la variable x son números reales positivos y Γ(b) es la función gamma. La Distribución Gamma comienza en el origen de coordenadas y tiene una forma bastante flexible. Otras distribuciones son casos especiales de ella.

Ejemplo

```
# Graficando Gamma
a = 2.6 # parametro de forma.
gamma = stats.gamma(a)
x = np.linspace(gamma.ppf(0.01),
                gamma.ppf(0.99), 100)
fp = gamma.pdf(x) # Función de Probabilidad
plt.plot(x, fp)
plt.title('Distribución Gamma')
plt.ylabel('probabilidad')
plt.xlabel('valores')
plt.show()
```




```
# histograma
aleatorios = gamma.rvs(1000) # genera aleatorios
cuenta, cajas, ignorar = plt.hist(aleatorios, 20)
plt.ylabel('frequencia')
plt.xlabel('valores')
plt.title('Histograma Gamma')
plt.show()
```




***Log-normal.***

La Distribución Log-normal está dada por la formula:

$$p(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-1}{2}\left(\frac{\ln x-\mu}{\sigma}\right)^2}$$

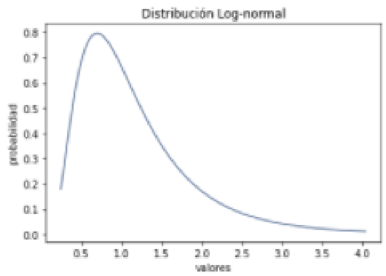
En dónde la variable x>0 y los parámetros μ y σ>0 son todos números reales. La Distribución Log-normal es aplicable a variables aleatorias que están limitadas por cero, pero tienen pocos valores grandes. Es una distribución con asimetría positiva. Algunos de los ejemplos en que la solemos encontrar son:

- El peso de los adultos.
- La concentración de los minerales en depósitos.

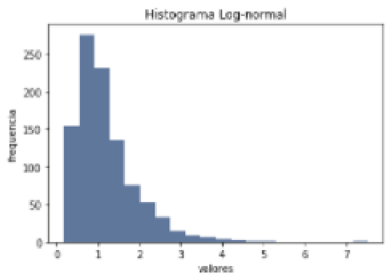
- Duración de licencia por enfermedad.
- Distribución de riqueza
- Tiempos muertos de maquinarias.

Ejemplo

```
# Graficando Log-Normal
sigma = 0.6 # parametro
lognormal = stats.lognorm(sigma)
x = np.linspace(lognormal.ppf(0.01),
                lognormal.ppf(0.99), 100)
fp = lognormal.pdf(x) # Función de Probabilidad
plt.plot(x, fp)
plt.title('Distribución Log-normal')
plt.ylabel('probabilidad')
plt.xlabel('valores')
plt.show()
```




```
# histograma
aleatorios = lognormal.rvs(1000) # genera aleatorios
cuenta, cajas, ignorar = plt.hist(aleatorios, 20)
plt.ylabel('frequencia')
plt.xlabel('valores')
plt.title('Histograma Log-normal')
plt.show()
```



Referencias

Anscombe, F.J. & W.W. Glynn, 1983. Distributions of the kurtosis statistic for normal statistics. Biometrika

Badii, M.H. & J. Castillo (eds.). 2007. Técnicas Cuantitativas en la Investigación. 348 pp. UANL, Monterrey. ISBN

Badii, M.H., J. Castillo, J. Landeros & K. Cortez. 2007a. Papel de la estadística en la investigación científica. Innovaciones de Negocios.

Badii, M.H., J. Castillo, A. Wong & J. Landeros. 2007b. Precisión de los índices estadísticos: técnicas de jacknife & bootstrap. Innovaciones de Negocios.

Badii, M.H., J. Castillos, R. Foroughbakhch & K. Cortez. 2007c. Probability and scientific research. Daena

Groeneveld, R.A. & G.Meeden, 1984. Measuring skewness and kurtosis. Statistician

Notas al pie

*Acerca de los autores

El Est. Jonnathan Matute es Estudiante de la Carrera de Ciencias de la Computación de la Universidad Politécnica Salesiana SEDE Cuenca – Ecuador, jmatutec2@est.ups.edu.ec

Bibliografía

https://towardsdatascience.com/generate-random-variable-using-inverse-transform-method-in-python8e5392f170a3

https://usmanwardag.github.io/python/astronomy/2016/07/10/inverse-transform-sampling-with-python.html

Distribuciones

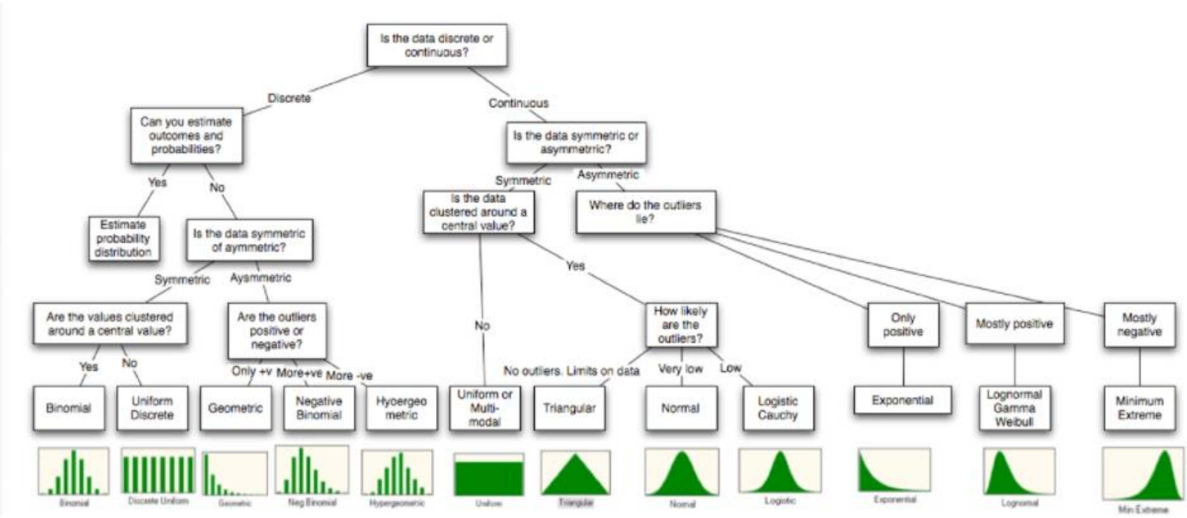



*Ilustración 1*. Un modelo que podemos seguir cuando nos encontramos con datos que necesitamos ajustar a una distribución, es comenzar con los datos sin procesar y responder a cuatro preguntas básicas acerca de los mismos, que nos pueden ayudar a caracterizarlos.