Generando normal de lo anormal:
Un ensayo del Teorema del Límite Central[⋆,⋆⋆]

Olivares Espinosa Cesar[a] (Estudiante MCIC)

ARTICLE INFO

Keywords:
Teorema del Límite Central
Ley de los Números Grandes
Demostraciones
Python 3

ABSTRACT

El Teorema del Límite Central es una de las aplicaciones de la probabilidad más importantes ya que nos permite utilizar técnicas de distribuciones normales en datos derivados de distribuciones no necesariamente normales. Este ensayo tiene como finalidad comprender el teorema, revisando la teoría probabilística necesaria para su correcta compresión y demostración, así como una seríe de simulaciones realizadas en Python 3 para visualizar el desempeño del teorema con distintas distribuciones.

1. Introducción

En el mundo de los modelos probablisticos se dan los casos cuando no es posible calcular una probabilidad o una expectación de forma directa se pueden utilizar métodos como los siguientes:

- Simulaciones utilizando el método de Monte Carlo.
- Métodos con limites de desigualdades.
- Aproximaciones con teoremas de límites.

Aquí es donde entra el denominado Teorema del Límite Central (TLC) el cual es un método probabilístico muy importante, ya que nos permite obtener, bajo ciertas condiciones, una distribución normal de datos, derivados de una variable aleatoria que puede o no tener un comportamiento normal. El escrito se divide en distintas secciones:

- Una discusión de la teoría sobre la cual se basa el TLC comenzando por las desigualades de Markov y Chebyshev y su relación con el comportamiento de la media y varianza de una distribución, siguiendo con la Ley de los Números Grandes, la convergencia intuitiva en la que caen los modelos probabilisticos y como se relaciona esta con el TLC, la Ley Débil de los Números Grandes donde se suponen las condiciones para conseguir una distribución normal para finalmente derivarse en el TLC.
- Una explicación más profunda del Teorema del Limite Central, el por que es importante, como se deriva y sustena en la teoría de la sección prevía, así como las ventajas que brinda y el por que es uno de los modelos de la teória de la probabilidad más importante.
- Una serie de demostraciones utilizando el lenguaje de programación de alto nivel Python, para visualizar el desempeño de este teorema con distintas distribuciones, no necesariamente normales, la influencia de sus parámetros y su *tuning* para conseguir la distribución normal.

Para una mejor compresión del teorema es recomendable tener conocimientos previos acerca de la Distrución Normal y Muestreo de una Distribución Estadística, los cuales quedan fuera del alcance de este escrito.

[⋆] Ensayo para la materia de Probabilidad, Procesos Aleatorios e Inferencia.
[⋆⋆] Centro de Investigación en Computación.
ces.oe97@gmail.com (O.E. Cesar)
ORCID(s):

2. Fundamentos Teóricos

2.1. Desigualdad de Markov

Consideremos una variable aleatoria X que solo toma valores no negativos ($X \leq 0$), una variable fija a y una segunda variable aleatoria Y_a definida por:

$$Y_a = \begin{cases} 0 & X < a \\ a & X \geq a \end{cases}$$

Se puede ver que de esto se deriva,

$$Y_a \geq X$$
$$\mathbf{E}[Y_a] \leq \mathbf{E}[X]$$

La masa de probabilidad de la Función de Densidad de Probabilidad (PDF) de la variable aleatoria X en el intervalo [0,a] es asignada a 0 y el resto es asignado al valor a, esto ocasiona que la masa de probabilidad este orientada hacia la izquierda, el valor esperado solo puede descender, por lo tanto:

$$E[X] \geq E[Y_a] = aP(Y_a = a) = aP(X \geq a)$$

De otra forma,

$$E[Y_a] = aP(Y_a = a) = aP(X \geq a)$$

De la cual se obtiene,

$$aP(X \geq a) \leq E[X]$$

La desigualdad de Markov nos asegura que si una variable aleatoria no negativa tiene una media de **bajo valor**, entonces la probabilidad de que se tome un valor alto también sera baja.

2.2. Desigualdad de Chebyshev

Consideremos una variable aleatoria con un valor de media μ y varianza σ^2 de la forma $(X - \mu)^2$ y aplicando la desigualdad de Markov con un valor $a = c^2$.Obtenemos,

$$P((X - \mu)^2 \geq c^2) \leq \frac{E[(X - \mu)^2}{c^2} = \frac{\sigma^2}{c^2}$$

Notemos que,

$$(X - \mu)^2 \geq c^2 = |X - \mu| \geq c$$

Por lo que podemos completar la desigualdad como,

$$P(|X - \mu| \geq c) = P((X - \mu)^2 \geq c^2) \leq \frac{\sigma^2}{c^2}$$

Introducimos la función g para generar las derivaciones de la desigualdad de Markov,

$$g(x) = \begin{cases} 0 & |x - \mu| < c, \\ c^2 & |x - \mu| \geq c \end{cases}$$

Notamos que $(x - \mu)^2 \geq g(x)$ para toda,

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f_x(x) dx \geq \int_{-\infty}^{\infty} g(x) f_x(x) dx = c^2 P(|X - \mu| \geq c)$$

Reescribiendo $c = k\sigma$, donde k es positiva,

$$P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

Por lo tanto la probabilidad de que la variable aleatoria X tome un valor a su media es a lo mas $\frac{1}{k^2}$ donde k es el número de desviaciones estándar.

Finalizando, la desigualdad de Chebyshev afirma que si una variable aleatoria tiene una **varianza baja**, entonces la probabilidad de que tome un valor alejado de su media también es baja.

2.3. Ley de los Números Grandes

Imaginemos el escenario donde lanzamos una moneda, los primeros tiros puede que salga una cantidad mayor de águila o sol, pero si realizamos muchos tiros, cerca de la mitad de los tiros corresponderán a cada uno de los posibles resultados, esta es una forma intuitiva de comprender la probabilidad, la Ley de los Números Grandes explica como un modelo probabilístico llega a la misma propiedad, y el que tan rápido lo hace lo explica el Teorema del Límite Central, el cual se verá más adelante.

2.4. Ley Débil de los Números Grandes

Supongamos una variable aleatoria X, para la cual su media $E[X] = \mu$ es desconocida y $X_1, X_2, ..., X_n$ son n valores independientes medidos de X, son independientes idénticamente distribuidos(i.i.d), con la misma Función de Densidad de Probabilidad (PDF) de X.
La media muestral de la secuencia se utiliza para estimar E[X]:

$$M_n = \frac{X_1 + ... + X_n}{n} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

$$E[M_n] = \frac{E[X_1] + ... + E[X_n]}{n} = \frac{n\mu}{n} = \mu$$

Usando la independencia,

$$var(M_n) = \frac{var(X_1 + ... + X_n)}{n^2} = \frac{var(X_1) + ... + var(X_n)}{n^2}$$
$$= \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

Usando la desigualdad de Chebyshev

$$P(|M_n - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}, \quad \forall\epsilon > 0$$

Para toda $\epsilon > 0$ el término derecho de la desigualdad tiende a cero conforme n crece. Sean $X_1, X_2, ...$ variables aleatorias i.i.d. con media μ. Para toda $\epsilon > 0$, tenemos:

$$P(|M_n - \mu| \geq \epsilon) = P(|\frac{X_1 + ... + X_n}{n} - \mu| \geq \epsilon) \to 0, n \to \infty$$

La ley débil de los números grandes nos afirma que para una n lo suficientemente grande, la masa de distribución de M_n se concentra cerca de la media μ.Si consideramos un intervalo positivo $[\mu - \epsilon, \mu + \epsilon]$ alrededor de μ, entonces hay una alta posibilidad de que M_n caiga dentro del intervalo, mientras $n \to \infty$ la probabilidad converge a 1, si ϵ es muy pequeña se debe esperar más (una n más grande), para asegurar que M_n tendrá una alta probabilidad de caer en el intervalo, por lo tanto, asegura que la media muestral de una gran número de variables aleatorias i.i.d es muy cercana a la media verdadera (con un alta probabilidad).

3. El Teorema del Límite Central

De acuerdo a la Ley Débil de los Números Grandes, la varianza tiende a cero, la varianza de la suma incrementa al infinito y la distribución de S_n no converge a nada importante.

$$S_n = X_1 + ... + X_n = n(M_n)$$

Una vista intermedia se obtiene con la desviación $S_n - n\mu$ de S_n de su media $n\mu$ y escalandola por un factor proporcional a $\frac{1}{\sqrt{n}}$ y mantiene la varianza a un nivel constante. El teorema del límite central afirma que la distribución de la variable aleatoria escalada se acerca a la distribución normal.
Sea $X_1, X_2, ...$ una secuencia de variables aleatorias i.i.d. con media μ y varianza σ^2. Se define,

$$Z(n) = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{X_1 + ... + X_n - n\mu}{\sigma\sqrt{n}}$$

Se obtiene,

$$E[Z_n] = \frac{E[X_1 + ... + X_n] - n\mu}{\sigma\sqrt{n}} = 0$$

y,

$$var(Z_n) = \frac{var(X_1 + ... + X_n}{\sigma^2 n} = \frac{var(X_1) + ... + var(X_n)}{\sigma^2 n} = \frac{n\sigma^2}{\sigma^2 n} = 1$$

El teorema del límite central es muy general, aparte de la independencia y la suposición de que la media y varianza son finitos, no tiene requerimientos para la distribución de X_i, esta puede ser discreta, continua o mixta.
Nos indica que la sumatoría de un gran número de variables aleatorias i.i.d. es aproximadamente normal, por lo que se aplica en situaciones donde un efecto aleatorio de un gran número de pequeños, pero independientes factores aleatorias, como lo es el ruido en los circuitos eléctricos.
Igualmente elimina la necesidad de modelos probabilísticos detallados y manipulaciones tediosas de las PMF y PDF, si no que permite el cálculo de ciertas probabilidades simplemente refiriendo a la tabla de la CDF normal, solo son requeridas las medias y varianzas.

4. Demostraciones del Teorema del Límite Central

Para esto se utiliza un entorno de Python 3, en este caso se programo en una Notebook de Jupyter, el código esta disponible en: `https://github.com/colivaresе/The-Central-Limit-Theorem.git`

Importamos las librerías necesarias

```
import numpy as np
import seaborn as sns
import statistics as stat
import matplotlib.pyplot as plt
```

Comenzamos haciendo la prueba con datos derivados de una distribución normal y graficamos los valores

```
norm_dist_data = np.random.normal(size=10000)
sns.displot(norm_dist_data,kde = True)
```

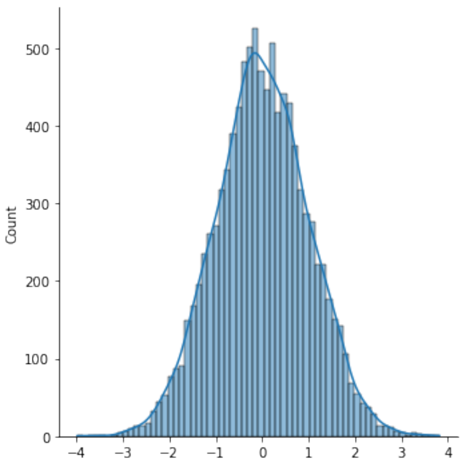


Figure 1: Distribución normal aleatoria

Ahora se crea una función para tomar muestras aleatorias de los datos, donde podremos indicar el tamaño de la mustra y el número de valores a tomar,

```
def sample_mean(dataset, sample_size, n_samples):
sample_means = []
for i in range(n_samples):
    sample = np.random.choice(dataset,
                              size=sample_size,
                              replace=False)
    sample_mean = stat.mean(sample)
    sample_means.append(sample_mean)
return sample_means
```

Comenzemos tomando 10 muestras con 10 valores cada una,

```
norm_plot_10 = sns.displot(sample_mean(norm_dist_data,10,10),kde=True)
```

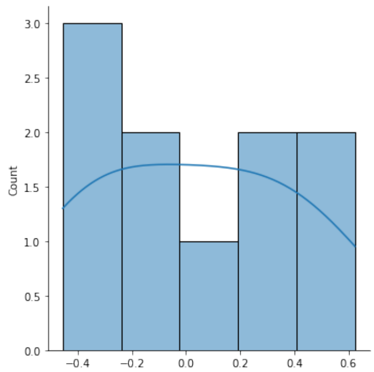


Figure 2: Distribución normal aleatoria

La distribución de la muestra no tiene aún un comportamiento normal, mantengamos 10 muestras, pero con 1000 valores cada una.

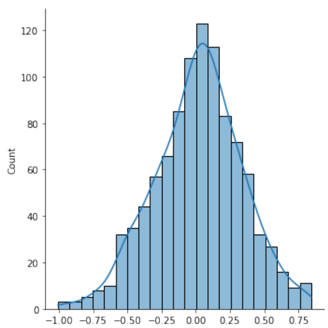


Figure 3: Distribución normal aleatoria

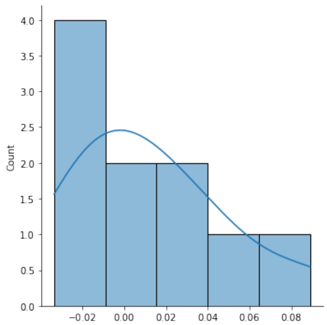
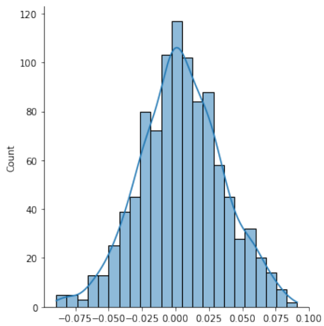


Figure 4: Distribución normal aleatoria

Se aprecia que hay una apariencia de distribución normal aumentando el número de valores por muestra, cambiemos los valores tomando 1000 muestras, con 10 valores cada una para ver el comportamiento.

La distribución no se aprecia y tiene sentido, empezara a tener una distribución normal mientras el número de muestras n aumente, como se vio en el ejercicio anterior, ahora aumentemos ambos valores a 1000. La distribución es normal,



pero recordemos que se deriva de datos con una distribución normal, probemos ahora con una que no cumple con esta propiedad.
Generemos datos con una distribución no normal.

```
from scipy.stats import skewnorm
sk_dist_data = skewnorm.rvs(12,size=10000)
sns.displot(sk_dist_data,kde=True)
```

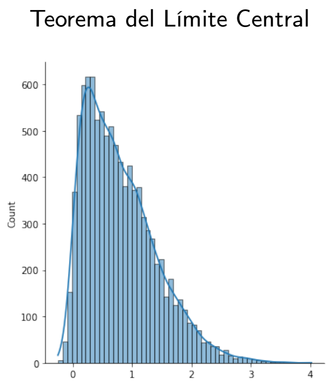


Figure 5: Distribución normal aleatoria

Repitamos el procedimiento tomando 10 muestras con 1000 valores, 1000 muestras con 1000 valores y ambos con 1000.

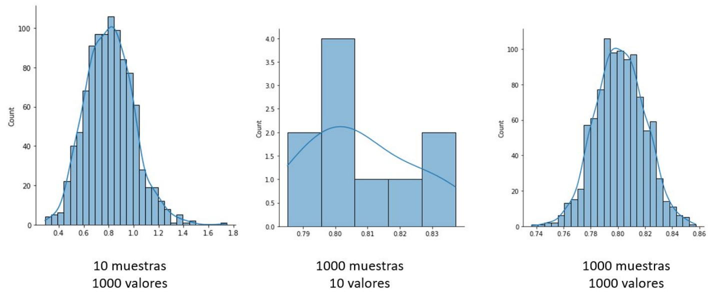


Figure 6: Distribución normal aleatoria

La distribución se aprecia como normal a pesar de proceder de datos claramente no normales, siendo esta la principal aportación del teorema.
Pero esto no podría funcionar para todo, recordemos que se asume que la varianza es finita y utilizar una distribución que no cumple con esta característica no derivaría en una distribución normal, probemos con una Distribución de Pareto con un valor de 1000 muestras y 1000 valores.

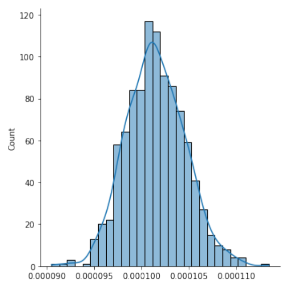


Figure 7: Distribución de Pareto

5. Conclusiones

El Teorema del Límite Central sirve como uno de los resultados más importantes de la teoría de la probabilidad, permitiéndonos asegurar que de la suma de un gran número de variables aleatorias independientes es posible obtener una distribución aproximadamente normal y utilizar técnicas de análisis probabilístico para distribuciones normales sin comprometer los resultados.
Puede ser utilizado para cualquier distribución de la cual se pueda calcular la media y la varianza, si no se pueden obtener el resultado de distribución no tendrá el mejor desempeño como ocurre en la Distribución de Pareto y aunque la teoría nos muestra en que caso es y no es útil es necesario hace runa experimentación comom la realizada pra poder visualizar y comprender de mejor forma el teorema, se dice de una regla que para que para poder obtener un valor bueno de distribución normal el valor de n debe ser al menos de 30, y esto es una generalización muy conservadora, la verdad es que dependiendo del problema y de los datos se tendrá que encontrar un valor óptimo para poder utilizar técnicas de distribución normal y facilitar el ejercicio.

6. Referencias

Bertsekas, D. P., Tsitsiklis, J. N. (2008). Introduction to probability: Dimitri P. Bertsekas and John N. Tsitsikis. Belmont, Massachussets: Athena Scientific.