TEOREMA DEL LÍMITE CENTRAL

Nexus-Probability

CURSO 2 (PROBABILIDAD II)

PARTE 5 / LECCIÓN 2

Este teorema es de uma importancia para quienes estudiamos probabilidad, pues tiene una amplia gama de aplicaciones para simplificar el cálculo de ciertas probabilidades y aproximar algunas distribuciones.

Definición 1 (Teorema central del límite) *Sea $X_1, X_2, \ldots$ una sucesión infinita de variables aleatorias independientes e idénticamente distribuidas, con media μ y varianza finita σ^2. Entonces la función de distribución de la variable aleatoria*

$$Z_n = \frac{(X_1 + \cdots + X_n) - n\mu}{\sqrt{n\sigma^2}}$$

tiende a la función de distribución normal estándar cuando n tiende a infinito.

Definición 2 (Teorema de Moivre-Laplace) *Suponga que se tiene una sucesión infinita de ensayos independientes de un experimento aleatorio. Sea A un evento de este experimento aleatorio con probabilidad de ocurrencia $p > 0$. Sean n_A el número de ocurrencias del evento de interés en los primeros n ensayos del experimento. Entonces para cualesquiera números reales $a < b$,*

$$lim_{n\to\infty} P(a < \frac{n_A/n - p}{\sqrt{p(1-p)/n}} < b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Ejercicios

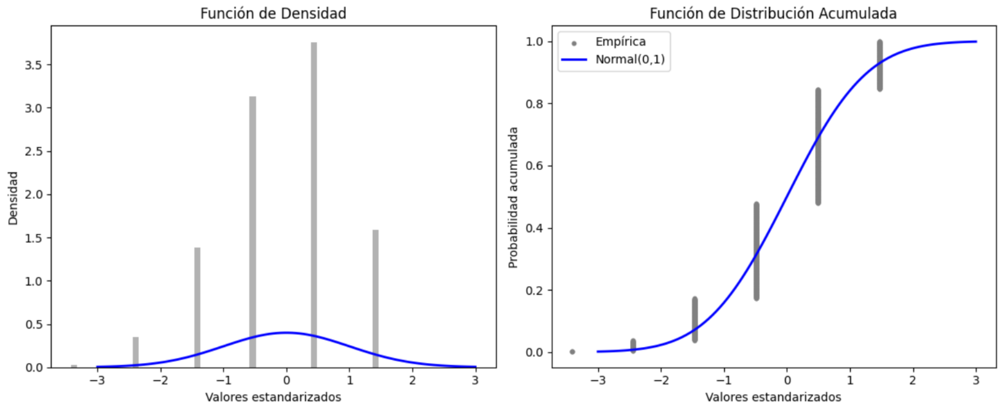
Los siguientes ejercicios propuestos tendrán solución en **Python**, por lo que te invitamos a ejecutar el código en tu computadora.

Ejercicio 1 *Usando phyton comprabamos mediante simulación el teorema central del límite en el caso de variables aleatorias $Ber(p)$ con $p = 0.7$. Como en el enunciado del teorema, el parámetro n corresponde al número de sumando $X_1 + \cdots + X_n$. El parámetro $k = 1000$ se usa para generar k valores al azar de la variable aleatoria suma $X_1 + \cdots + X_n$ y así aproximar su función de distribución.*

Solución.

```
import numpy as np
import matplotlib.pyplot as plt
import scipy.stats as stats

# N mero de simulaciones y de sumandos por simulaci n
k = 1000  # N mero de valores simulados
n = 5  # N mero de sumandos

# Par metro de la distribuci n Bernoulli
p = 0.7

# Generaci n de datos
s = np.array([np.sum(np.random.binomial(1, p, n)) for _ in range(k)])

# C lculo de la media, varianza y estandarizaci n
media = n * p
varianza = n * p * (1 - p)
s_estandarizado = (s - media) / np.sqrt(varianza)

# Graficaci n de la funci n de densidad
fig, ax = plt.subplots(1, 2, figsize=(12, 5))

# Histograma de la muestra y comparaci n con la normal est ndar
x = np.linspace(-3, 3, 100)
ax[0].hist(s_estandarizado, bins=50, density=True, alpha=0.6, color='gray')
ax[0].plot(x, stats.norm.pdf(x, 0, 1), 'b-', lw=2)
ax[0].set_title("Funci n de Densidad")
ax[0].set_xlabel("Valores estandarizados")
ax[0].set_ylabel("Densidad")

# Funci n de distribuci n emp rica vs normal est ndar
ecdf = np.sort(s_estandarizado)
y_ecdf = np.arange(1, k + 1) / k
```

```
ax[1].scatter(ecdf, y_ecdf, c='gray', s=10, label="Emp rica")
ax[1].plot(x, stats.norm.cdf(x, 0, 1), 'b-', lw=2, label="Normal(0,1)")
ax[1].set_title("Funci n de Distribuci n Acumulada")
ax[1].set_xlabel("Valores estandarizados")
ax[1].set_ylabel("Probabilidad acumulada")
ax[1].legend()

plt.tight_layout()
plt.show()
```



Ejercicio 2 *Se lanza un dado repetidas veces y sean $X_1, X_2, \ldots$ los resultados de estos lanzamientos. Es razonable suponer que estas variables aleatorias son independientes y con idéntica distribución uniforme en el conjunto $\{1, 2, 3, 4, 5, 6\}$. En particular, la esperanza es $\mu = 3.5$ y la varianza es $\sigma^2 = 2.916$. Por la ley de los grandes números, sabemos que el promedio parcial*

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

se aproxima a la media 3.5 conforme n crece.
¿Cuántas veces debe lanzarse el dado de tal forma que $\bar{X}$ se encuentre entre 3 y 4 con una probabilidad de 0.99?

Solución.

Se busca el valor de n tal que

$$P(3 \leq \bar{X} \leq 4) = 0.99.$$

Restando en cada lado de las desigualdades la media μ y dividiendo entre $\sqrt{\sigma^2/n}$, la igualdad anterior es equivalente a la ecuación

$$P\left(\frac{3-3.5}{\sqrt{\sigma^2/n}} \leq \frac{\bar{X}-3.5}{\sqrt{\sigma^2/n}} \leq \frac{4-3.5}{\sqrt{\sigma^2/n}}\right) = 0.99.$$

Por el teorema central del límite, la probabilidad indicada es aproximadamente igual a

$$P\left(\frac{-0.5}{\sqrt{\sigma^2/n}} \leq Z \leq \frac{0.5}{\sqrt{\sigma^2/n}}\right),$$

en donde Z es una variable aleatoria con distribución normal estándar. Es decir, tenemos ahora la ecuación de aproximación

$$\Phi\left(\frac{0.5}{\sqrt{\sigma^2/n}}\right) - \Phi\left(\frac{-0.5}{\sqrt{\sigma^2/n}}\right) = 0.99.$$

```python
    import scipy.stats as stats
import numpy as np

# Datos del problema
mu = 3.5
sigma2 = 2.916  # Varianza
sigma = np.sqrt(sigma2)  # Desviaci n est ndar
probabilidad = 0.99  # Probabilidad objetivo

# Valores de la desigualdad
limite_inferior = 3
limite_superior = 4

# Transformaci n a la normal est ndar
z_inferior = -0.5  # (3 - 3.5)
z_superior = 0.5   # (4 - 3.5)

# CDF de la normal est ndar
def encontrar_n():
    for n in range(1, 10000):  # Probamos valores de n
        desviacion_muestral = np.sqrt(sigma2 / n)
        phi_superior = stats.norm.cdf(z_superior / desviacion_muestral)
        phi_inferior = stats.norm.cdf(z_inferior / desviacion_muestral)
        if phi_superior - phi_inferior >= probabilidad:
            return n  # Devolvemos el primer n que satisface la condici n

# Encontramos el valor de n
n_requerido = encontrar_n()
```

```
print(f"El n mero m nimo de lanzamientos requerido es: {
    n_requerido}")
```

```
El número mínimo de lanzamientos requerido es: 78
```

Ejercicio 3 *Se desea diseñar un estacionamiento de coches para un conjunto de 200 departamentos que se encuentran en construcci ´on. Suponga que para cada departamento, el número de automóviles será de 0, 1 o 2, con probabilidades 0.1, 0.6 y 0.3, respectivamente. Se desea que con una certeza del 95 % haya espacio disponible para todos los coches cuando los departamentos se vendan. ¿Cuántos espacios de estacionamiento deben construirse?*

Solución.

Sean $X_1, \dots, X_{200}$ las variables aleatorias que denotan el número de automóviles que poseen los futuros dueños de los departamentos. Podemos suponer que estas variables aleatorias discretas son independientes unas de otras y todas ellas tienen la misma distribuci ´on de probabilidad:

$$P(X = 0) = 0.1, P(X = 1) = 0.6, P(X = 2) = 0.3$$

De esta forma la v.a suma $X_1, \dots, X_{200}$ denota el total de autos que habrá en el complejo de departamentos. Se desconoce la distribución de esta variable aleatoria, sin embargo se desea encontrar el valor de n tal que

$$P(X_1 + \cdots + X_{200} \leq n) = 0.95$$

. Haremos uso del teorema central del lı́mite para resolver este problema, y para ello se necesita calcular la esperanza y varianza de X. Puede comprobarse que $E(X) = 1.2$ y $Var(X) = 0.36$, cantidades que denotaremos por μ y σ^2 respectivamente. La ecuación planteada es entonces

$$P(X_1 + \cdots + X_{200} \leq n) = 0.95$$

, donde la incógnita es el valor de n. Restando en ambos lados de la igualsas 200μ y dividiendo entre $\sqrt{200\sigma^2}$, tenemos que

$$P(\frac{X_1 + \cdots + X_{200} - 200\mu}{\sqrt{200\sigma^2}} \leq \frac{n - 200\mu}{\sqrt{200\sigma^2}} = 0.95)$$

Por el teorema central del límite, la probabilidad indicada es aproximada a $\Phi((n - 200\mu)/\sqrt{200\sigma^2})$. Así tenemos,

$$\Phi(\frac{n - 200\mu}{\sqrt{200\sigma^2}}) = 0.95$$

Siguiendo con las cuentas y usando la tabla de distribución normal obtenemos que $n = 253.99$. Es decir, el tama~no del estacionamiento debe ser de aproximadamente 254 lugares.