Métricas para distribuciones de Probabilidad

Rodrigo Gerardo Trejo Arriaga

30 de marzo de 2024

Distribución Normal

La distribución normal se caracteriza por dos parámetros: la media μ y la desviación estándar σ. Estos se calculan como sigue:

- **Media** (μ): La media de una población se calcula como:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i$$

  donde N es el número total de valores en el conjunto de datos y x_i es cada valor individual.

- **Desviación estándar** (σ): La desviación estándar se calcula usando la fórmula:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2}$$

  donde μ es la media del conjunto de datos, N es el número total de valores, y x_i es cada valor individual.

Tendencia Central

Para una distribución normal con media μ y desviación estándar σ, las métricas son:

- **Media**: μ
- **Mediana**: μ
- **Moda**: μ

Dispersión

Para una distribución normal con media μ y desviación estándar σ:

- **Rango**: No definido específicamente ya que teóricamente es $(-\infty, \infty)$. En el caso discreto, diferencia entre el valor máximo y mínimo observado.
- **Varianza** (σ^2): σ^2
- **Desviación estándar** (σ): σ
- **Coeficiente de variación** (CV): $\frac{\sigma}{\mu}$ (para $\mu \neq 0$)

Posición Relativa

La posición relativa dentro de una distribución normal puede determinarse mediante la función de distribución acumulativa (CDF). Los percentiles y cuartiles se calculan como sigue:

- **Percentil** (P): Un valor x que cumple con la condición de que P

  $$x = \mu + Z(P) \cdot \sigma$$

  donde $Z(P)$ es el valor Z para el percentil P de la distribución normal estándar.
- **Cuartil**: Los cuartiles son casos especiales de percentiles; por ejemplo, el primer cuartil (25 %) es $Q1 = \mu + Z(0{,}25) \cdot \sigma$, el segundo cuartil (mediana, 50 %) es $Q2 = \mu$, y el tercer cuartil (75 %) es $Q3 = \mu + Z(0{,}75) \cdot \sigma$.

Para calcular cualquier percentil P en una distribución normal, se utiliza la función de distribución acumulativa (CDF) inversa ajustada por la media μ y la desviación estándar σ:

- **Percentil** P:

  $$x_P = \mu + Z(P) \cdot \sigma$$

  Donde x_P es el valor correspondiente al percentil P y $Z(P)$ es el valor de la puntuación Z para el percentil P en la distribución normal estándar.

Los cuartiles son casos específicos de percentiles (25 %, 50 %, 75 %, 100 %), y se calculan con la misma fórmula ajustando P para 0.25, 0.5, 0.75 y 1.

Forma

La distribución normal es simétrica, por lo que sus métricas de forma son:

- **Asimetría (Skewness)**: 0, indicando una distribución perfectamente simétrica.

Ver la sección de notas.

Curtosis y Entropía

Para una distribución normal con media μ y desviación estándar σ:

- **Curtosis**: 3 (La curtosis se mide a menudo relativa a una distribución normal, que tiene una curtosis de 3. Por lo tanto, el exceso de curtosis sería 0).
- **Entropía**: $\frac{1}{2}\ln(2\pi e\sigma^2)$, donde e es la base del logaritmo natural.

Distribución Gamma

La distribución gamma se caracteriza por los parámetros de forma k y escala θ. La estimación de estos parámetros no se realiza mediante una fórmula directa, sino a través de métodos estadísticos avanzados como la estimación de máxima verosimilitud o el ajuste de momentos.

- **Nota**: Los parámetros k y θ suelen estimarse a partir de los datos utilizando técnicas estadísticas complejas que van más allá de cálculos directos.

Tendencia Central

Para una distribución gamma con parámetros de forma k y escala θ:

- **Media**: $k\theta$
- **Mediana**: La mediana para la distribución gamma no tiene una expresión simple cerrada.
- **Moda**: Para $k > 1$, Moda $= (k-1)\theta$. Para $k \leq 1$, la moda no está bien definida.

Dispersión

Para una distribución gamma con parámetros de forma k y escala θ:

- **Rango**: $(0, \infty)$ para $k, \theta > 0$. En el caso discreto, diferencia entre el valor máximo y mínimo observado.
- **Varianza**: $k\theta^2$
- **Desviación estándar**: $\sqrt{k}\theta$
- **Coeficiente de variación**: $\frac{\sqrt{k}\theta}{k\theta} = \frac{1}{\sqrt{k}}$

Posición Relativa

Los percentiles y cuartiles de la distribución gamma también se basan en su CDF, pero no hay fórmulas simples para su cálculo directo y generalmente se requieren métodos numéricos:

- **Percentil y Cuartil**: Se determinan a través de la inversa de la CDF de la distribución gamma, que generalmente se calcula mediante software estadístico debido a la complejidad matemática de la función.

```python
import scipy.stats as stats

# Parametros de la distribucion Gamma: forma (k) y escala (theta)
k = 2.0  # Parametro de forma
theta = 3.0  # Parametro de escala

# Calcular cuartiles
cuartil_1 = stats.gamma.ppf(0.25, k, scale=theta)
cuartil_2 = stats.gamma.ppf(0.50, k, scale=theta)  # Mediana
cuartil_3 = stats.gamma.ppf(0.75, k, scale=theta)
cuartil_4 = stats.gamma.ppf(1.00, k, scale=theta)  # Maximo

print("Cuartil 1 (Q1):", cuartil_1)
print("Cuartil 2 (Mediana, Q2):", cuartil_2)
print("Cuartil 3 (Q3):", cuartil_3)
print("Cuartil 4 (Maximo):", cuartil_4)

# Calcular un percentil especifico, por ejemplo, el percentil 90
percentil_90 = stats.gamma.ppf(0.90, k, scale=theta)
print("Percentil 90:", percentil_90)

# Para calcular cualquier otro percentil, cambia el 0.90 del ejemplo anterior
# por el valor correspondiente.
```

Este código utiliza la función `ppf` (Percent Point Function, la inversa de la CDF) de la distribución gamma en SciPy para calcular los cuartiles y percentiles.

Forma

La asimetría de la distribución gamma depende de su parámetro de forma k:

- **Asimetría (Skewness)**: $\frac{2}{\sqrt{k}}$, indicando que la distribución puede ser asimétrica dependiendo del valor de k. Para valores grandes de k, la distribución se aproxima a una forma simétrica.

Curtosis y Entropía

Para una distribución gamma con parámetros de forma k y escala θ:

- **Curtosis**: $\frac{6}{k}$
- **Entropía**: $k + \ln(\theta) + \ln(\Gamma(k)) + (1-k)\psi(k)$, donde $\Gamma(k)$ es la función gamma y $\psi(k)$ es la función digamma.

Distribución Uniforme

La distribución uniforme está definida por dos parámetros: a y b, que son los límites inferior y superior de la distribución, respectivamente.

- **Límite inferior** (a): El valor mínimo en el rango de valores posibles.
- **Límite superior** (b): El valor máximo en el rango de valores posibles.

Tendencia Central

Para una distribución uniforme definida entre a y b:

- **Media**: $\frac{a+b}{2}$
- **Mediana**: $\frac{a+b}{2}$
- **Moda**: No está bien definida ya que todos los valores dentro del intervalo $[a, b]$ son igualmente probables.

Dispersión

Para una distribución uniforme definida entre a y b:

- **Rango**: $b - a$
- **Varianza**: $\frac{(b-a)^2}{12}$
- **Desviación estándar**: $\sqrt{\frac{(b-a)^2}{12}}$
- **Coeficiente de variación**: $\frac{\sqrt{\frac{(b-a)^2}{12}}}{\frac{a+b}{2}}$ (para $a, b > 0$)

Posición Relativa

Para la distribución uniforme, los percentiles y cuartiles se pueden calcular de forma más directa debido a la naturaleza uniforme de la distribución:

- **Percentil** (P): Dado un percentil P (como fracción, por ejemplo, 0.25 para el 25 %), el valor correspondiente x es:
$$x = a + P \cdot (b - a)$$
- **Cuartil**: De manera similar, cada cuartil se calcula ajustando P para 0.25, 0.5, y 0.75 respectivamente.

Los percentiles y cuartiles se calculan directamente de la forma:

- **Percentil** P:
$$x_P = a + P \cdot (b - a)$$
Donde x_P es el valor correspondiente al percentil P (como fracción, por ejemplo, 0.25 para el 25 %).

Para los cuartiles, P se ajusta para 0.25, 0.5, 0.75 y 1, respectivamente.

Forma

La distribución uniforme es simétrica entre a y b, por lo que sus métricas de forma son:

- **Asimetría (Skewness)**: 0, reflejando su simetría perfecta.

Ver la sección de notas.

Curtosis y Entropía

Para una distribución uniforme definida entre a y b:

- **Curtosis**: $-\frac{6}{5}$ (o el exceso de curtosis es $-\frac{6}{5}$, ya que la curtosis de una distribución uniforme perfecta es menor en comparación con una normal).
- **Entropía**: $\ln(b - a)$

Notas

La asimetría (skewness) de un conjunto de datos se puede calcular utilizando la siguiente fórmula:

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_i - \bar{x}}{s} \right)^3$$

donde n es el número de observaciones, x_i son los valores individuales, $\bar{x}$ es la media del conjunto de datos, y s es la desviación estándar.

Este cálculo puede realizarse fácilmente en Python para conjuntos de datos tanto normalmente como uniformemente distribuidos, utilizando bibliotecas como SciPy o Pandas. A continuación se muestra un ejemplo de código en Python para calcular la asimetría de un conjunto de datos:

```
import numpy as np
from scipy.stats import skew

# Ejemplo de conjunto de datos
data = np.array([1, 2, 3, 4, 5, 6, 7, 8, 9, 10])

# Cálculo de la asimetría
data_skewness = skew(data)

print("Asimetría del conjunto de datos:", data_skewness)
```