Tarea1_D_SSF

Luis Gerardo Hernández Román

Agosto 2024

1 Distribuciones de probabilidad

Las distribuciones de probabilidad son funciones matemáticas que describen la probabilidad de que ocurra cada posible valor de una variable aleatoria. Existen dos tipos principales de distribuciones de probabilidad: discretas y continuas.

1.1 Distribuciones de Probabilidad Discretas

Una distribución de probabilidad discreta se aplica a variables aleatorias que pueden tomar un número finito o contable de valores. Las probabilidades asociadas a cada valor suman 1.

- Distribución de Bernoulli: Modelo de un experimento con dos posibles resultados: éxito (con probabilidad p) y fracaso (con probabilidad $1 - p$).

  Ejemplo: Lanzamiento de una moneda que da cara o cruz.

- Distribución Binomial: Describe el número de éxitos en n ensayos independientes, cada uno con probabilidad de éxito p.

  Ejemplo: Número de caras en 10 lanzamientos de una moneda justa.

- Distribución de Poisson: Modela el número de eventos que ocurren en un intervalo fijo de tiempo o espacio cuando estos eventos ocurren con una tasa media constante y de forma independiente.

  Ejemplo: Número de llamadas telefónicas que recibe una central en una hora.

- Distribución Geométrica: Describe el número de ensayos necesarios para obtener el primer éxito en una serie de ensayos independientes, cada uno con probabilidad p.

  Ejemplo: Número de lanzamientos de una moneda hasta obtener la primera cara.

- Distribución Binomial Negativa: Extiende la distribución geométrica para modelar el número de ensayos necesarios para obtener un número fijo de éxitos.

  Ejemplo: Número de lanzamientos necesarios para obtener 3 caras.

- Distribución Categórica: Modela un experimento con más de dos resultados posibles, cada uno con su propia probabilidad.

  Ejemplo: Resultado de tirar un dado, donde cada cara tiene una probabilidad de 1/6.

1.2 Distribuciones de Probabilidad Continuas

Una distribución de probabilidad continua se aplica a variables aleatorias que pueden tomar cualquier valor en un intervalo continuo. Las probabilidades se expresan como áreas bajo una curva.

- Distribución Normal: También conocida como distribución gaussiana, es la distribución más común en la estadística. Tiene una forma de campana y está definida por su media μ y desviación estándar σ.

  Ejemplo: Altura de personas adultas en una población.

- Distribución Uniforme Continua: Describe una variable aleatoria continua que tiene la misma probabilidad de tomar cualquier valor dentro de un intervalo.

  Ejemplo: Tiempo de espera de un autobús que llega en cualquier momento entre 0 y 10 minutos.

- Distribución Exponencial: Modela el tiempo entre eventos en un proceso de Poisson, donde los eventos ocurren de forma continua e independiente a una tasa constante.

  Ejemplo: Tiempo hasta el próximo terremoto en una región.

- Distribución Gamma: Extiende la distribución exponencial para modelar el tiempo hasta que ocurren k eventos en un proceso de Poisson.

  Ejemplo: Tiempo hasta la tercera llamada telefónica en una central.

- Distribución Beta: Se utiliza para modelar la distribución de probabilidades de una proporción o porcentaje.

  Ejemplo: Probabilidad de éxito en un experimento de proporción desconocida.

2 Medidas de tendencia central

La media y la desviación estándar son dos medidas fundamentales en estadística para describir la distribución de un conjunto de datos. A partir de la desviación estándar se puede obtener la varianza

2.1 Media Aritmética

La media aritmética (o simplemente ”media”) es el promedio de un conjunto de valores. Por lo general en distribuciones gaussianas o normales la media aritmética esta muy cerca de la mediana o moda. Se calcula sumando todos los valores y dividiendo la suma por el número de valores, de manera matemática podemos escribir;

Sea $x_1, x_2, \ldots, x_n$ un conjunto de n valores, la media aritmética μ (o $\bar{x}$) se calcula como:

$$\mu = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

2.2 Desviación Estándar y Varianza

La desviación estándar mide la dispersión de los datos con respecto a la media. Es la raíz cuadrada de la varianza, que es la media de las diferencias al cuadrado entre cada valor y la media.

Primero, se calcula la varianza σ^2 como la media de las diferencias al cuadrado entre cada valor y la media:

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

La desviación estándar σ es la raíz cuadrada de la varianza:

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

3 Problema

Tenemos 10000 datos aleatorios entre 1.65 y 1.75. Realizamos un histograma con python para observar que tipo de distribución siguen los datos.

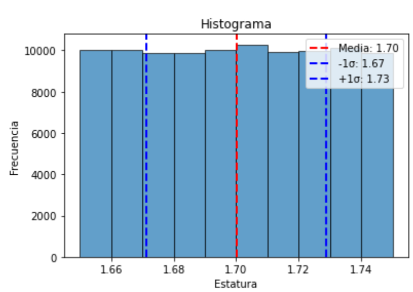

Figure 1: Histograma de los datos aleatorios proporcionados por el profesor [1]

En la Figure 1 se observa una distribución uniforme de los datos, se pudo obtener la media y la desviación estándar de los datos, pero, ¿esto es importante en este caso?. La respuesta es no, Como los datos en este caso son de igual probable para cualquiera de los datos de la distribución, no tiene sentido obtenerlos.

Para poder obtener una distribución normal con datos aleatorios podemos usar el código Tarea1_D2.py del link [1]. En este código se obtienen 10000 datos aleatorios, que se ajustan en un rango de 1.60 a 1.80. La media es de 1.70 y se estable un desviación estándar de 0.05. El resultado se muestran en la Figura 2, donde observamos la distribución normal de los datos en un rango de 1.5 a 1.9 aproximadamente.

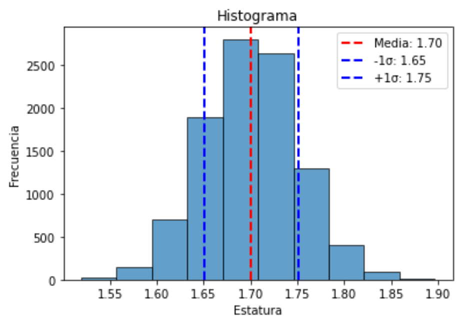

Figure 2: Histograma de los datos aleatorios obtenidos con el codigo de python[1]

La mayor parte de los datos se encuentran entre $\mu-\sigma$ y $\mu+\sigma$ lo que representa una buena señal de que los datos siguen la distribucion normal.

References

[1] GitHub Disponible en: `https://github.com/Luis1g/SSF-Oto24LuisGerardoHernandezRoman/blob/main/altura5.dat`