Matemática para Ciencia de Datos

Variables Aleatorias

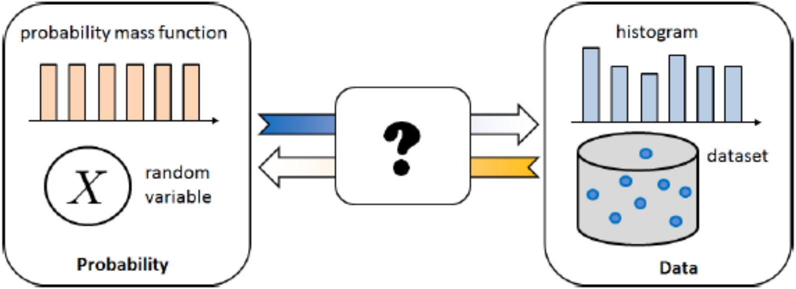
Dr. Daniel Alexis Gutierrez Pachas (dgutierrezp@ucsp.edu.pe)

26 de noviembre de 2023

Departamento de Ciencia de la Computación, Universidad Católica San Pablo, Arequipa, Perú.

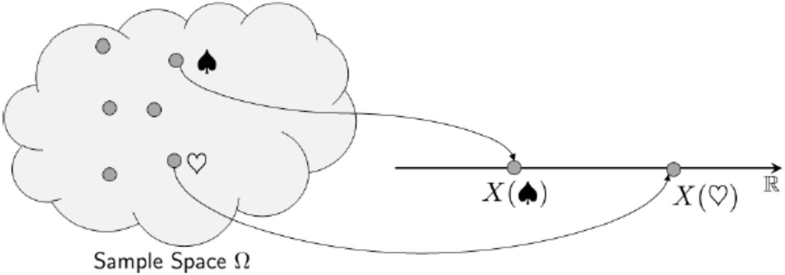
El panorama de la probabilidad y los datos

A menudo vemos la probabilidad y el análisis de datos como dos diferentes entidades. Sin embargo, la probabilidad y el análisis de datos son inseparables. El objetivo del curso es vincular estas dos áreas.



Definición

La variable aleatoria X es una función. La entrada a la función es un resultado del espacio muestral Ω, mientras que la salida es un número en la recta real.



En general, una variable aleatoria X puede tomar un número finito o infinito (numerable) de valores.

Ejemplo 1

Supongamos una moneda se lanza dos veces. Definamos X a la variable aleatoria que cuenta el número de caras.

Ejemplo 2

Entre los 5 estudiantes en un curso con coeficiente de rendimiento (CR) mayor a 8.5. Dos estudiantes serán sorteados para recibir una beca. Sean los CR de estos estudiantes dados por: 8.8; 9.2; 8,9; 9.5; 9.0. Sea X la variables aleatoria que defina la media de los estudiantes sorteados.

(a) Liste los posibles valores de X.

(b) Liste el evento $\{X \geq 9\}$.

Ejemplo 3

Un hombre posee cuatro llaves en su bolsillo. Como la habitación está oscura, el no consigue ver cual es la llave correcta y experimenta con cada una de ellas hasta encontrar la correcta. Defina la variable aleatoria X que represente el número de llaves experimentadas hasta conseguir abrir la puerta (inclusive la correcta).

Distribución de probabilidad (Probability mass function - PMF)

PMF de una variable aleatoria X es una función que especifíque la probabilidad de obtener un número $X(\xi) = x$. Denotamos PMF como:

$$f(x) = P(X = x).$$

El conjunto de todos los posibles estados de X es denotado por $X(\Omega)$. Una PMF debe satisfacer

$$\sum_{x \in X(\Omega)} f(x) = 1.$$

La función de distribución acumulada es:

$$F(x) = P(X \leq x).$$

Ejemplo 4

Supongamos que lanzamos un dado y definimos la variable aleatoria X como el numero de intentos hasta obtener 6. Determinar $f(x)$ y $F(x)$.

Ejemplo 5

Una urna contiene cinco bolas numeradas 1, 2, 2, 8 y 8. Si una persona selecciona dos bolas al azar y definimos la variable aleatoria X como la suma obtenida por las dos bolas. Determinar $f(x)$ y $F(x)$.

Valor esperado, Varianza y desviación estándar

Calculamos el valor esperado de X como sigue:

$$\mathbb{E}(X) = \sum x \cdot f(x).$$

Luego la varianza, se obtiene como:

$$\text{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

Una formula alternativa para calcular $\text{Var}(X)$ es

$$\text{Var}(X) = \mathbb{E}\left(X^2\right) - [\mathbb{E}(X)]^2.$$

Luego, la desviación estándar es

$$sd(X) = \sqrt{\text{Var}(X)}.$$

Ejemplo 6

Sea una variable aleatoria X tal que

x	1	2	3	4	5
$P(X = x)$	p^2	p^2	p	p	p^2

Calcular $E(X)$ y $Var(X)$.

Generalización de una variable aleatoria

Supongamos que $Y = g(X)$, siendo que g es una función. Entonces,

$$E(g(X)) = \sum g(x) \cdot f(x),$$

en el cual f es la función de probabilidad de la variable aleatoria X.

Si $Y = aX + b$, cuando a y b son constantes reales. Entonces,

$$E(Y) = aE(X) + b,$$

y

$$\mathrm{Var}(Y) = a^2 \mathrm{Var}(X).$$

Ejemplo 7

En una tienda, la cantidad de productos vendidos en un día por uno de los funcionarios es una variable aleatoria X con la siguiente distribución de probabilidad

x	0	1	2	3	4	5	6
$P(X = x)$	0.1	0.4	0.2	0.1	0.1	0.05	0.05

Cada vendedor recibe comisiones de venta, distribuidas de la siguiente forma: si él vende hasta dos productos en un día, él gana una comisión de S/.10 por producto vendido. A partir de la tercera venta la comisión es de S/. 50. Determinar

(a) El número medio de productos vendidos.

(b) La comisión media.

Ejemplo 8

Un supermercado hace la siguiente promoción. Cuando el cliente pasar a pagar lanza un dado y en caso de salir 6 tiene un descuento del 30 % sobre el total de la cuenta. Si sale 5 el descuento es del 20 %. Si sale 4 el descuento es del 10 % y se ocurre 1,2, o 3 es descuento es del 5 %. Determinar el descuento medio concedido.

Distribución de Bernoulli

Un ensayo de Bernoulli es un experimento aleatorio en el que sólo se pueden obtener dos resultados (éxito y fracaso). Estos ensayos están modelados por una variable aleatoria que puede tomar sólo dos valores, 0 y 1. El 1 representa el éxito y 0 el fracaso. La probabilidad de éxito es p mientras que para el fracaso es $q = 1 - p$. Además,

$$\mathbb{E}(X) = p \quad \text{y} \quad \mathrm{Var}(X) = pq.$$

Distribución Binomial

La distribución binomial describe el número de aciertos en una serie de n experimentos independientes de Bernoulli. La variable aleatoria X asume valores en el conjunto $\{0, 1, \ldots, n\}$. Entonces

$$P(X = k) = \mathrm{Comb}(n, k) p^k q^{n-k}, \ \ \forall\, k = 0, 1, \ldots, n.$$

Además

$$\mathbb{E}(X) = np \quad \text{y} \quad \mathrm{Var}(X) = npq.$$

Distribución Geométrica

La distribución geométrica describe el número de intentos necesarios hasta conseguir el primer acierto. La variable aleatoria asume valores en el conjunto $\{1, 2, \ldots\}$. Entonces,

$$P(X = k) = q^{k-1}p, \;\; \forall\, k = 1, 2, \ldots.$$

Además

$$\mathbb{E}(X) = \frac{1}{p} \quad \text{y} \quad \mathrm{Var}(X) = \frac{1-p}{p^2}.$$

Ejemplo 9

El 20 % de los trabajadores de una empresa irá a la huelga. Se seleccionan 5 trabajadores de dicha empresa. Supongamos que X es la variable aleatoria que define el Número de asistentes a la huelga entre los 5 seleccionados. Determinar

(a) La probabilidad de que al menos tres vayan a la huelga.

(b) La probabilidad de que todos vayan a la huelga.

(c) La probabilidad de que no vaya ninguno.

Ejemplo 10

Andrés y Pedro se plantean el juego de lanzar un dado. Se considera que el jugador gana cuando el resultado del dado es cuatro o seis, y recibe diez soles. En otro caso, no recibe nada. Cada apuesta (un lanzamiento) es de cinco soles.

(a) Si Andrés juega en cinco ocasiones, ¿Cual es la probabilidad de que acierte a lo sumo una vez? ¿Cual es el número medio de aciertos en esas cinco ocasiones?

(b) Si Pedro juega tantas veces como sea necesario hasta conseguir acertar una vez. ¿Cual es la probabilidad de que tenga que jugar al menos tres veces?. ¿Cual es el número medio de veces que tiene que jugar para conseguir acertar?

(c) ¿Cual será el beneficio medio de cada jugador?