Probabilidad y estadística para la Inteligencia Artificial: Medidas de tendencia central y dispersión

Camilo Enrique Argoty Pulido

Especialización en Inteligencia Artificial

9 de septiembre de 2025

Esperanza matemática

Esperanza matemática

Uno de los primeros conceptos desarrollados ha sido el de **esperanza matemática**.

Esperanza matemática

Uno de los primeros conceptos desarrollados ha sido el de **esperanza matemática**.

En el contexto de los juegos de azar, corresponde a la ganancia que puede esperar un jugador.

Esperanza matemática

Uno de los primeros conceptos desarrollados ha sido el de **esperanza matemática**.

En el contexto de los juegos de azar, corresponde a la ganancia que puede esperar un jugador.

Pregunta, ¿pueden ustedes calcular una ganancia esperada si juegan a la ruleta?

La esperanza de una variable aleatoria

Si X es una v.a. continua y h es una función continua de $\mathbb{R}$ en $\mathbb{R}$, la esperanza de $h(X)$ es:

La esperanza de una variable aleatoria

Si X es una v.a. continua y h es una función continua de $\mathbb{R}$ en $\mathbb{R}$, la esperanza de $h(X)$ es:

$$E[h(X)] = \int_{\mathbb{R}} h(x) f_X(x) dx$$

La esperanza de una variable aleatoria

Si X es una v.a. continua y h es una función continua de $\mathbb{R}$ en $\mathbb{R}$, la esperanza de $h(X)$ es:

$$E[h(X)] = \int_{\mathbb{R}} h(x) f_X(x) dx$$

Si X es una v.a. discreta y h igual que en el caso anterior,

La esperanza de una variable aleatoria

Si X es una v.a. continua y h es una función continua de $\mathbb{R}$ en $\mathbb{R}$, la esperanza de $h(X)$ es:

$$E[h(X)] = \int_{\mathbb{R}} h(x) f_X(x) dx$$

Si X es una v.a. discreta y h igual que en el caso anterior,

$$E[h(X)] = \sum_{i=1}^{\infty} h(x_i) p_i$$

La esperanza de una variable aleatoria

Si X es una v.a. continua y h es una función continua de $\mathbb{R}$ en $\mathbb{R}$, la esperanza de $h(X)$ es:

$$E[h(X)] = \int_{\mathbb{R}} h(x) f_X(x) dx$$

Si X es una v.a. discreta y h igual que en el caso anterior,

$$E[h(X)] = \sum_{i=1}^{\infty} h(x_i) p_i$$

Para la esperanza de X, suele usarse la letra griega μ y también se le dice la **media** de X.

La varianza de una v.a.

El problema de definir la dispersión de una v.a.

Además de entender la esperanza de una variable, se hace necesario entender qué tan dispersos están los datos alrededor de la media.

El problema de definir la dispersión de una v.a.

Además de entender la esperanza de una variable, se hace necesario entender qué tan dispersos están los datos alrededor de la media.

Para ello, sería bueno calcular la distancia media de los valores de X a μ ($E[X-\mu]$), ¿verdad?

El problema de definir la dispersión de una v.a.

Además de entender la esperanza de una variable, se hace necesario entender qué tan dispersos están los datos alrededor de la media.

Para ello, sería bueno calcular la distancia media de los valores de X a μ ($E[X-\mu]$), ¿verdad?

La verdad no, porque $(X-\mu)$ puede tomar valores positivos y negativos, los cuales se cancelan entre sí y no permiten tener una idea clara de la dispersión de los datos. ¿Qué hacer entonces?

Existen varias opciónes, por ejemplo:

Existen varias opciónes, por ejemplo:

- $E[|X - \mu|]$

Existen varias opciónes, por ejemplo:

- $E[|X - \mu|]$
- $E[(X - \mu)^2]$

Existen varias opciónes, por ejemplo:

- $E[|X-\mu|]$
- $E[(X-\mu)^2]$

¿Cuál es mejor?

Existen varias opciónes, por ejemplo:

- $E[|X - \mu|]$
- $E[(X - \mu)^2]$

¿Cuál es mejor?

La mejor es $E[(X - \mu)^2]$ por lo siguiente. Si X e Y son v.a. con medias μ y ν.

Existen varias opciónes, por ejemplo:

- $E[|X - \mu|]$
- $E[(X - \mu)^2]$

¿Cuál es mejor?

La mejor es $E[(X - \mu)^2]$ por lo siguiente. Si X e Y son v.a. con medias μ y ν.Se tiene $E[X + Y] = \mu + \nu$. Entonces

Existen varias opciónes, por ejemplo:

- $E[|X-\mu|]$
- $E[(X-\mu)^2]$

¿Cuál es mejor?

La mejor es $E[(X-\mu)^2]$ por lo siguiente. Si X e Y son v.a. con medias μ y ν.Se tiene $E[X+Y]=\mu+\nu$. Entonces

$$
\begin{aligned}
E[(X+Y-\mu-\nu)^2] &= E[\big((X-\mu)+(Y-\nu)\big)^2] = \\
&= E[(X-\mu)^2 + 2(X-\mu)(Y-\nu) + (Y-\nu)^2] = \\
&= E[(X-\mu)^2 + (Y-\nu)^2] + 2E[(X-\mu)(Y-\nu)]
\end{aligned}
$$

La dispersión de la suma de dos variables

De lo anterior se tiene que,

La dispersión de la suma de dos variables

De lo anterior se tiene que,

$$E[(X + Y - \mu - \nu)^2] = E[(X - \mu)^2] + E[(Y - \nu)^2] + 2E[(X - \mu)(Y - \nu)]$$

La dispersión de la suma de dos variables

De lo anterior se tiene que,

$$E[(X+Y-\mu-\nu)^2] = E[(X-\mu)^2] + E[(Y-\nu)^2] + 2E[(X-\mu)(Y-\nu)]$$

No parece una fórmula muy amable, pero si usamos $E[|X-\mu|]$ tampoco parece haber una fórmula mejor.

La dispersión de la suma de dos variables

De lo anterior se tiene que,

$$E[(X+Y-\mu-\nu)^2] = E[(X-\mu)^2] + E[(Y-\nu)^2] + 2E[(X-\mu)(Y-\nu)]$$

No parece una fórmula muy amable, pero si usamos $E[|X-\mu|]$ tampoco parece haber una fórmula mejor.

Lo interesante es que aparece esta cantidad:

La dispersión de la suma de dos variables

De lo anterior se tiene que,

$$E[(X+Y-\mu-\nu)^2] = E[(X-\mu)^2] + E[(Y-\nu)^2] + 2E[(X-\mu)(Y-\nu)]$$

No parece una fórmula muy amable, pero si usamos $E[|X-\mu|]$ tampoco parece haber una fórmula mejor.

Lo interesante es que aparece esta cantidad:

$$E[(X-\mu)(Y-\nu)]$$

La dispersión de la suma de dos variables

De lo anterior se tiene que,

$$E[(X+Y-\mu-\nu)^2] = E[(X-\mu)^2] + E[(Y-\nu)^2] + 2E[(X-\mu)(Y-\nu)]$$

No parece una fórmula muy amable, pero si usamos $E[|X-\mu|]$ tampoco parece haber una fórmula mejor.

Lo interesante es que aparece esta cantidad:

$$E[(X-\mu)(Y-\nu)]$$

¿Esto qué puede ser?

Comparación con el álgebra lineal

En álgebra lineal, si $\bar{x}$ e $\bar{y}$ son vectores en $\mathbb{R}^n$, se tiene que:

Comparación con el álgebra lineal

En álgebra lineal, si $\bar{x}$ e $\bar{y}$ son vectores en $\mathbb{R}^n$, se tiene que:

$$\|\bar{x}+\bar{y}\|^2 = \|\bar{x}\|^2 + \|\bar{y}\|^2 + 2\langle \bar{x}|\bar{y}\rangle,$$

Comparación con el álgebra lineal

En álgebra lineal, si $\bar{x}$ e $\bar{y}$ son vectores en $\mathbb{R}^n$, se tiene que:

$$\|\bar{x}+\bar{y}\|^2 = \|\bar{x}\|^2 + \|\bar{y}\|^2 + 2\langle \bar{x}|\bar{y}\rangle,$$

donde $\langle \bar{x}|\bar{y}\rangle$ es el **producto interno** o **producto punto** entre los vectores $\bar{x}$ e $\bar{y}$.

Comparación con el álgebra lineal

En álgebra lineal, si $\bar{x}$ e $\bar{y}$ son vectores en $\mathbb{R}^n$, se tiene que:

$$\|\bar{x}+\bar{y}\|^2 = \|\bar{x}\|^2 + \|\bar{y}\|^2 + 2\langle\bar{x}|\bar{y}\rangle,$$

donde $\langle\bar{x}|\bar{y}\rangle$ es el **producto interno** o **producto punto** entre los vectores $\bar{x}$ e $\bar{y}$.

De lo anterior, se obtiene que $E[(X-\mu)^2]$ y $E[(X-\mu)(Y-\nu)]$ tienen comportamientos parecidos a la norma al cuadrado y al producto interno del álgebra lineal.

Comparación con el álgebra lineal

En álgebra lineal, si $\bar{x}$ e $\bar{y}$ son vectores en $\mathbb{R}^n$, se tiene que:

$$\|\bar{x}+\bar{y}\|^2 = \|\bar{x}\|^2 + \|\bar{y}\|^2 + 2\langle \bar{x}|\bar{y}\rangle,$$

donde $\langle \bar{x}|\bar{y}\rangle$ es el **producto interno** o **producto punto** entre los vectores $\bar{x}$ e $\bar{y}$.

De lo anterior, se obtiene que $E[(X-\mu)^2]$ y $E[(X-\mu)(Y-\nu)]$ tienen comportamientos parecidos a la norma al cuadrado y al producto interno del álgebra lineal.

De ahí que el valor $E[(X-\mu)^2]$ sea el que aparenta ser el más apropiado para representar la dispersión de los resultados de un experimento.

Varianza y covarianza

Por lo tanto, se define la **varianza** de una v.a. X como

Varianza y covarianza

Por lo tanto, se define la **varianza** de una v.a. X como

$$Var(X) := E[(X - \mu)^2]$$

Varianza y covarianza

Por lo tanto, se define la **varianza** de una v.a. X como

$$Var(X) := E[(X - \mu)^2]$$

Ya que la varianza es el equivalente estadístico de la norma al cuadrado, la raíz cuadrada merece un nombre propio.

Varianza y covarianza

Por lo tanto, se define la **varianza** de una v.a. X como

$$Var(X) := E[(X-\mu)^2]$$

Ya que la varianza es el equivalente estadístico de la norma al cuadrado, la raíz cuadrada merece un nombre propio.

Por lo tanto, si X es una v.a. se define la **desviación (o desvío) estándar** como:

$$\sigma(X) := \sqrt{Var(X)}$$

Varianza y covarianza

Por lo tanto, se define la **varianza** de una v.a. X como

$$Var(X) := E[(X-\mu)^2]$$

Ya que la varianza es el equivalente estadístico de la norma al cuadrado, la raíz cuadrada merece un nombre propio.

Por lo tanto, si X es una v.a. se define la **desviación (o desvío) estándar** como:

$$\sigma(X) := \sqrt{Var(X)}$$

Por otro lado,Y si X e Y son v.a. se define la **covarianza** entre X e Y:

Varianza y covarianza

Por lo tanto, se define la **varianza** de una v.a. X como

$$Var(X) := E[(X-\mu)^2]$$

Ya que la varianza es el equivalente estadístico de la norma al cuadrado, la raíz cuadrada merece un nombre propio.

Por lo tanto, si X es una v.a. se define la **desviación (o desvío) estándar** como:

$$\sigma(X) := \sqrt{Var(X)}$$

Por otro lado,Y si X e Y son v.a. se define la **covarianza** entre X e Y:

$$Cov(X,Y) := E[(X-\mu)(Y-\nu)]$$

Correlación

Siguiendo con la similitud con el álgebra lineal, si $\bar{x}$ e $\bar{y}$ son vectores en $\mathbb{R}^n$, el coseno del ángulo entre $\bar{x}$ e $\bar{y}$ es el valor:

$$\rho(\bar{x}, \bar{y}) = \frac{\langle \bar{x} | \bar{y} \rangle}{(\|\bar{x}\| \|\bar{y}\|)}$$

Correlación

Siguiendo con la similitud con el álgebra lineal, si $\bar{x}$ e $\bar{y}$ son vectores en $\mathbb{R}^n$, el coseno del ángulo entre $\bar{x}$ e $\bar{y}$ es el valor:

$$\rho(\bar{x}, \bar{y}) = \frac{\langle \bar{x} | \bar{y} \rangle}{(\|\bar{x}\| \|\bar{y}\|)}$$

De aquí podemos definir una cantidad similar para v.a.

Correlación

Siguiendo con la similitud con el álgebra lineal, si $\bar{x}$ e $\bar{y}$ son vectores en $\mathbb{R}^n$, el coseno del ángulo entre $\bar{x}$ e $\bar{y}$ es el valor:

$$\rho(\bar{x}, \bar{y}) = \frac{\langle \bar{x} | \bar{y} \rangle}{(\|\bar{x}\| \|\bar{y}\|)}$$

De aquí podemos definir una cantidad similar para v.a.
Si X e Y son v.a., se define la **correlación (de Pearson)** entre X e Y como:

$$\rho(X, Y) := \frac{Cov(X, Y)}{\sqrt{Var(X) * Var(Y)}}$$

Media y varianza de distribuciones comunes

Media y varianza de variables aleatorias discretas comunes

Nombre	Parámetros	$p(x)$	Media y varianza
Uniforme	$N > 0$, entero	$\frac{1}{N}$, $x = 1, \ldots, N$	$\frac{N+1}{2}$, $\frac{N^2-1}{12}$.
Binomial	$0 < p < 1$,n entero	$\binom{n}{x} p^x (1 - p)^x$ x= 1,2,...,n	np, $np(1-p)$.
Poisson	$\lambda > 0$	$\frac{e^{-\lambda}\lambda^x}{x!}$, x entero ≥ 0	λ, λ
Geométrica	$0 < p < 1$	$(1-p)^x p$, x entero ≥ 0	$\frac{1-p}{p}$, $\frac{1-p}{p^2}$
Binomial negativa	$r > 0$, $0 < p < 1$	$\binom{r+x-1}{x} p^x (1 - p)^x$ x= 1,2,...,n	$\frac{r(1-p)}{p}$, $\frac{r(1-p)}{p^2}$

Media y varianza de variables aleatorias continuas comunes

Nombre	Parámetros	$f_X(x)$	Media y varianza
Uniforme	$a < b \in \mathbb{R}$	$\frac{1}{b-a}\mathbb{I}_{(a,b)}$	$\frac{a+b}{2}$, $\frac{(b-a)^2}{12}$
Normal	$\mu,\sigma^2 \geq 0$	$\frac{1}{\sigma\sqrt{2\pi}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$	μ, σ^2
Exponencial	$\frac{1}{\lambda} = \theta > 0$	$\lambda e^{-\lambda x}$, $x > 0$	$\frac{1}{\lambda}$, $\frac{1}{\lambda}$
Ji cuadrado	$k > 0$, grados de libertad	$\frac{(\frac{1}{2})^{\frac{k}{2}}}{\Gamma(\frac{k}{2})}x^{\frac{k}{2}-1}e^{-\frac{x}{2}}$ $x > 0$	k, $2k$
Gamma	$\alpha > 0$, $\beta > 0$	$\frac{\beta^\alpha x^{\alpha-1}e^{-\beta x}}{\Gamma(\alpha)}$, $x > 0$	$\frac{\alpha}{\beta}$, $\frac{\alpha}{\beta^2}$

Desviación estándar y Teorema de Chebychev

La raíz cuadrada de la varianza se denomina **desviación (o desvío) estándar**

Desviación estándar y Teorema de Chebychev

La raíz cuadrada de la varianza se denomina **desviación (o desvío) estándar**

El resultado que mejor ilustra y representa a la desviación estándar es el **Teorema de Chebyshev**:

Desviación estándar y Teorema de Chebychev

La raíz cuadrada de la varianza se denomina **desviación (o desvío) estándar**

El resultado que mejor ilustra y representa a la desviación estándar es el **Teorema de Chebyshev**:

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

Independencia

Independencia

En un espacio de probabilidad (Ω, p), se dice que dos eventos A y B se dicen **independientes** si:

Independencia

En un espacio de probabilidad (Ω, p), se dice que dos eventos A y B se dicen **independientes** si:

$$P(A|B) = P(A),$$

o en otras palabras:

Independencia

En un espacio de probabilidad (Ω, p), se dice que dos eventos A y B se dicen **independientes** si:

$$P(A|B) = P(A),$$

o en otras palabras:

$$P(A \cap B) = P(A)P(B)$$

Independencia

En un espacio de probabilidad (Ω, p), se dice que dos eventos A y B se dicen **independientes** si:

$$P(A|B) = P(A),$$

o en otras palabras:

$$P(A \cap B) = P(A)P(B)$$

De la misma forma, dos v.a. se dicen **independientes** si:

$$f_{Y|X}(y|x) = f_Y(y)$$

Independencia

En un espacio de probabilidad (Ω, p), se dice que dos eventos A y B se dicen **independientes** si:

$$P(A|B) = P(A),$$

o en otras palabras:

$$P(A \cap B) = P(A)P(B)$$

De la misma forma, dos v.a. se dicen **independientes** si:

$$f_{Y|X}(y|x) = f_Y(y)$$

En otras palabras,

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$$

Por otro lado, dos v.a. X e Y son **no correlacionadas** si $Cov(X, Y) = 0$.

Independencia

En un espacio de probabilidad (Ω, p), se dice que dos eventos A y B se dicen **independientes** si:

$$P(A|B) = P(A),$$

o en otras palabras:

$$P(A \cap B) = P(A)P(B)$$

De la misma forma, dos v.a. se dicen **independientes** si:

$$f_{Y|X}(y|x) = f_Y(y)$$

En otras palabras,

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$$

Por otro lado, dos v.a. X e Y son **no correlacionadas** si $Cov(X, Y) = 0$.

En álgebra lineal, que dos variables sean no correlacionadas es equivalente a que sean linealmente independientes.

Independencia

En un espacio de probabilidad (Ω, p), se dice que dos eventos A y B se dicen **independientes** si:

$$P(A|B) = P(A),$$

o en otras palabras:

$$P(A \cap B) = P(A)P(B)$$

De la misma forma, dos v.a. se dicen **independientes** si:

$$f_{Y|X}(y|x) = f_Y(y)$$

En otras palabras,

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$$

Por otro lado, dos v.a. X e Y son **no correlacionadas** si $Cov(X, Y) = 0$.

En álgebra lineal, que dos variables sean no correlacionadas es equivalente a que sean linealmente independientes.

Si X e Y son independientes, entonces son no correlacionadas pero el recíproco no necesariamente es cierto, a menos que X e Y se distribuyan normal.