

# Chapter 1

## Probabilidad

### 1.1 Reglas de probabilidad

#### 1.1.1 Axiomas de Kolmogorov

Son los fundamentos matemáticos sobre los que se construye la teoría de la probabilidad Diremos que  $P(\cdot)$  es una función de probabilidad de conjuntos si se cumple lo siguiente: Sea  $A, A_1, \dots A_n$  eventos de un espacio muestral  $S$

I.  $P(A) \geq 0, \quad \forall A \in B$

II.  $P(S) = 1$

III. Si  $A_1, A_2, A_3, \dots$  es una colección de eventos mutuamente excluyentes, entonces:

$$P(A_1 \cup A_2 \cup A_3 \dots) = \sum_{i=1}^{\infty} P(A_i)$$

#### 1.1.2 Ley de la suma y Ley del producto

Para eventos que no son mutuamente excluyentes:

- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A \cap B) = P(A) * P(B|A)$

#### 1.1.3 Probabilidad Condicional

Para dos eventos cualesquiera  $A$  y  $A_1$ , con  $P(A_1) > 0$ , se tiene que la probabilidad condicional de que ocurra  $A$  dado que ocurrió  $A_1$  está dada por

$$P(A|A_1) = \frac{P(A \cap A_1)}{P(A_1)}$$

### 1.1.4 Ley de la Probabilidad Total

Sean los eventos  $A_1, A_2, \dots, A_k$  del espacio muestral  $S$ . Si los eventos forman una partición de  $S$  entonces para cualquier evento  $B$ , se tiene que

$$P(B) = \sum_{i=1}^k P(B|A_i) * P(A_i)$$

### 1.1.5 Teorema de Bayes

El Teorema de Bayes nos dice cómo actualizar nuestra creencia inicial ( $P(A)$ ) a una nueva creencia ( $P(AB)$ ) después de observar nueva evidencia ( $B$ ). La actualización depende de qué tan probable sea observar la evidencia si nuestra creencia inicial es correcta ( $P(BA)$ ) y de la probabilidad general de observar la evidencia ( $P(B)$ ).

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$

### 1.1.6 Dependencia e independencia de eventos

Decimos que dos eventos  $A$  y  $B$  son independientes si  $P(A|B) = P(A)$ , de lo contrario son dependientes.

## 1.2 Variables aleatorias y distribuciones

Algunas definiciones clave:

- **Variable Aleatoria**

Dado un experimento aleatorio con espacio muestral  $S$ , se define una variable aleatoria  $X$  como una función que asigna a cada elemento del espacio muestral  $a \in S$  un y solo un número real  $x \in \mathbb{R}$ . Estas variables aleatorias pueden ser numéricas o categóricas, por el momento solo estudiaremos las variables numéricas que pueden dividirse en discretas o continuas.

- **Función Masa de Probabilidad**

Si  $X$  es una variable aleatoria discreta que puede tomar valores en un conjunto  $S = \{x_1, x_2, \dots\}$  entonces la función Masa de probabilidad  $P_x(X) = P(X = x)$

- **Función densidad de probabilidad** A diferencia de la función masa de probabilidad para variables discretas, la FDP no da la probabilidad directamente de un valor específico, ya que para variables continuas, la probabilidad de tomar exactamente un valor es infinitesimalmente pequeña (cero). Formalmente, si  $X$  es una v.a continua, su función densidad de

probabilidad  $f_x(X)$  es una función tal que para cualquier intervalo  $[a, b]$  la probabilidad de que  $X$  tome un valor dentro del intervalo es

$$P(a \leq x \leq b) = \int_a^b f(x) dx$$

- **Función de distribución acumulada**

$$F_x(x) = P(X \leq x)$$

- **Esperanza** La esperanza matemática o valor esperado de una v.a es una medida del valor central de su distribución de probabilidad. **Para una v.a discreta:**

$$E(X) = \sum_i x_i * P(X = x_i)$$

**Para una v.a continua:**

$$E(X) = \int_{-\infty}^{\infty} x * f_x(x) dx$$

- **Varianza** Cuantifica cuánto se alejan los valores individuales de la variable de su valor esperado. **Para una v.a discreta**

$$Var(X) = \sum_i (x_i - \mu)^2 p_i$$

**Para una v.a continua**

$$E(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f_x(x) dx$$

## 1.3 Algunas distribuciones importantes

### 1.3.1 Distribución Binomial:

Es una distribución de probabilidad discreta que describe el número de éxitos en  $n$  ensayos de Bernoulli.

$$\begin{aligned} P(X = k) &= \binom{n}{k} p^k (1-p)^{n-k} & Var(X) &= np(1-p) \\ E(X) &= np \end{aligned}$$

### 1.3.2 Distribución Poisson:

La distribución de Poisson es una distribución de probabilidad discreta que modela el número de eventos que ocurren en un intervalo 1 de tiempo o espacio fijo.

$$\begin{aligned} P(X = x) &= \frac{e^{-\lambda} \lambda^x}{x!} & Var(X) &= \lambda \\ E(X) &= \lambda \end{aligned}$$

### 1.3.3 Distribución Normal:

La distribución normal es fundamental para realizar inferencias sobre poblaciones basadas en muestras.

$$\begin{aligned} f(x; \mu, \sigma) &= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} & Var(X) &= \sigma^2 \\ E(X) &= \mu \end{aligned}$$

### 1.3.4 Distribución Gamma:

La distribución gamma es una herramienta estadística versátil con aplicaciones en diversos campos debido a su capacidad para modelar variables continuas, positivas y con diferentes formas de asimetría.

$$\begin{aligned} f_x(\alpha, \beta) &= \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & Var(x) &= \alpha\beta^2 \\ \text{para } x > 0 \\ E(X) &= \alpha\beta \end{aligned}$$

## Chapter 2

# Inferencia

### 2.1 Muestreo Aleatorio y Sesgo

¿Por qué esto es importante para trabajar en IA? En inteligencia artificial y aprendizaje automático un muestreo aleatorio correcto garantiza que el modelo generalice bien. El sesgo en los datos puede llevar modelos injustos, inexactos o no robustos (como modelos que discriminan por falta de diversidad en los datos)

#### 2.1.1 Muestreo Aleatorio:

El muestreo aleatorio es una técnica para seleccionar una muestra de una población de manera que cada individuo tenga la misma probabilidad de ser seleccionado. Hay varios tipos de muestreo aleatorio entre los cuales están:

- **Muestreo aleatorio simple:** Se aplica cuando cada individuo tiene la misma probabilidad de ser elegido
- **Muestreo estratificado:** La población se divide en grupos llamados estratos y se toma una muestra aleatoria en cada grupo.
- **Muestreo sistemático:** Se elige un punto de inicio aleatorio y luego se seleccionan elementos cada cierto número fijo.

#### 2.1.2 Sesgo (Bias):

Es un tipo de error sistemático en los datos o métodos de análisis y produce resultados que no representan correctamente a la población. Hay varios tipos de sesgo entre los cuales están:

- **Sesgo de selección:** Ocurre cuando la muestra no es representativa de la población.
- **Sesgo de medición:** Errores sistemáticos en cómo se recopilan los datos.

- **Sesgo de confirmación:** Cuando la información recopilada es interpretada de manera que favorece

### 2.1.3 Estimadores puntuales

Un estimador puntual es una función de una muestra aleatoria que proporciona un único valor numérico como aproximación de un parámetro poblacional desconocido. En el contexto de la inteligencia artificial, los estimadores son fundamentales para la inferencia estadística, la construcción de modelos predictivos y la validación de su rendimiento sobre conjuntos de datos finitos. Los estimadores puntuales pueden calcularse de dos formas:

- **Método de máxima verosimilitud:** Formalmente si  $X_1, \dots, X_n \sim f(x; \theta)$  entonces un estimador puntual para  $\hat{\theta}$  es el valor que maximiza

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta)$$

o su logaritmo

$$\log(L(\theta)) = \sum_{i=1}^n \log f(x_i; \theta)$$

- **Método de los momentos:** Sea  $X_1, \dots, X_n$  una muestra aleatoria de una población con función de probabilidad  $f(x, \theta)$  con  $\theta$  un vector de parámetros  $\theta = (\theta_1, \dots, \theta_k)$  los estimadores de momentos se hallan al igualar los momentos muestrales, con los respectivos momentos poblacionales y luego resolver  $k$  ecuaciones con  $k$  incógnitas:

$$\begin{array}{ll} m_1 = \frac{1}{n} \sum_{i=1}^n x_i & E(X) \\ m_2 = \frac{1}{n} \sum_{i=1}^n x_i^2 & E(X^2) \\ \dots & \dots \\ m_k = \frac{1}{n} \sum_{i=1}^n x_i^k & E(X^k) \end{array}$$