

Probabilidad

💡 ¿Qué es la probabilidad?

La **probabilidad** cuantifica la incertidumbre de que ocurra un evento. Se define entre 0 y 1: 0 significa imposible, 1 significa seguro. Nos permite **modelar aleatoriedad**, hacer inferencias y tomar decisiones bajo incertidumbre.

🎯 Conceptos fundamentales

Espacio muestral (S)

Conjunto de todos los resultados posibles de un experimento aleatorio.

Ejemplos: Al lanzar un dado: $S = \{1,2,3,4,5,6\}$. Al lanzar una moneda: $S = \{\text{cara, sello}\}$.

Evento (A)

Subconjunto del espacio muestral que describe un resultado de interés.

Ejemplo: "Obtener un número par" en un dado: $A = \{2,4,6\}$.

⚠️ Axiomas de probabilidad

$$\text{No negatividad: } 0 \leq P(A) \leq 1 .$$

$$\text{Normalización: } P(S) = 1 .$$

Aditividad: Si A y B son mutuamente excluyentes, entonces $P(A \cup B) = P(A) + P(B)$.

Reglas útiles:

$$\text{complemento: } P(A^c) = 1 - P(A) .$$

$$\text{Regla general de la unión: } P(A \cup B) = P(A) + P(B) - P(A \cap B) .$$

Probabilidad condicional e independencia

Probabilidad condicional:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

Independencia:

A y B son independientes si $P(A \cap B) = P(A)P(B)$. Equivalente a

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

Ley de Probabilidad Total:

Si $\{B_1, B_2, \dots, B_n\}$ es una partición del espacio muestral:

$$P(A) = \sum_{i=1}^n P(A | B_i) P(B_i)$$

Útil cuando conocemos probabilidades condicionales de A en diferentes escenarios.

Teorema de Bayes:

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

Reversa la dirección de la condicional, clave para actualizar creencias con nueva evidencia. Frecuentemente se

combina con la Ley de Probabilidad Total para calcular $P(B)$.

Variables aleatorias y momentos

Variable aleatoria discreta

Toma valores contables ($0, 1, 2, \dots$). Se describe con **función de probabilidad** $p(x) = P(X = x)$.

Esperanza:

$$\mathbb{E}[X] = \sum_x x p(x)$$

Varianza:

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

Variable aleatoria continua

Toma valores en intervalos. Se describe con **densidad** $f(x)$ y $P(a \leq X \leq b) = \int_a^b f(x) dx$.

Esperanza:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Varianza:

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

Función de distribución acumulada (CDF):

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

Propiedades de esperanza y varianza

Propiedades de la Esperanza:

Linealidad: $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$

Aditividad: $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ (siempre, incluso si no son independientes)

Producto de independientes: Si X e Y son independientes: $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$

Propiedades de la Varianza:

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Si X e Y son independientes: $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$

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

Covarianza y Correlación:

Covarianza: $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$

Correlación: $\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$, con $-1 \leq \rho \leq 1$

Si X e Y son independientes, entonces $\text{Cov}(X, Y) = 0$ (el recíproco no siempre es cierto)

Distribuciones comunes

Discretas

Bernoulli(p): Un intento éxito/fallo.

$$P(X = 1) = p, \quad P(X = 0) = 1 - p,$$

$$\mathbb{E}[X] = p, \quad \text{Var}(X) = p(1 - p).$$

Binomial(n, p): Suma de n Bernoullis independientes.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k},$$

$$\mathbb{E}[X] = np, \quad \text{Var}(X) = np(1 - p).$$

Geométrica(p): Número de intentos hasta el primer éxito.

$$P(X = k) = (1 - p)^{k-1} p, \quad \mathbb{E}[X] = \frac{1}{p},$$

$$\text{Var}(X) = \frac{1-p}{p^2}.$$

Poisson(λ): Conteo de eventos raros en intervalo fijo.

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad \mathbb{E}[X] = \lambda,$$

$$\text{Var}(X) = \lambda.$$

Hipergeométrica(N, K, n): Muestreo sin reemplazo.

$$P(X = k) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

Continuas

Uniforme(a, b): Densidad constante en $[a, b]$.

$$f(x) = \frac{1}{b-a}, \quad \mathbb{E}[X] = \frac{a+b}{2},$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}.$$

Exponencial(λ): Tiempo entre eventos Poisson.

$$f(x) = \lambda e^{-\lambda x}, \quad \mathbb{E}[X] = \frac{1}{\lambda},$$

$$\text{Var}(X) = \frac{1}{\lambda^2}.$$

Normal(μ, σ^2): Distribución gaussiana (campana).

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Simétrica, suma de normales independientes es normal.

Normal Estándar: $Z \sim N(0, 1)$

Se usa tabla Z para calcular probabilidades.

Cualquier $X \sim N(\mu, \sigma^2)$ se estandariza:

$$Z = \frac{X - \mu}{\sigma}.$$

Desigualdades importantes

Desigualdad de Markov:

Para $X \geq 0$ y $a > 0$:

$$P(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Desigualdad de Chebyshev:

Para cualquier $k > 0$:

$$P(|X - \mathbb{E}[X]| \geq k) \leq \frac{\text{Var}(X)}{k^2}$$

Interpretación: La probabilidad de que X se aleje más de k unidades de su media está acotada por $\frac{\sigma^2}{k^2}$.

Por ejemplo, con $k = 2\sigma$: al menos 75% de los datos están dentro de 2 desviaciones estándar.

Teorema del Límite Central

Si X_1, X_2, \dots, X_n son variables aleatorias independientes e idénticamente distribuidas con media μ y varianza σ^2 , entonces:

$$\frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \xrightarrow{d} N(0, 1) \text{ cuando } n \rightarrow \infty$$

donde $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ es la media muestral.

Importancia: Justifica el uso de la distribución normal en estadística inferencial, sin importar la distribución original de los datos (si n es suficientemente grande, usualmente $n \geq 30$).

Ejemplo 1: Dado y eventos



Enunciado:

Se lanza un dado justo. Calcular: (a) $P(\text{número par})$, (b) $P(\text{número} > 4)$, (c) $P(\text{par y} > 4)$, (d) $P(\text{par o} > 4)$.

Espacio muestral y eventos:

S	Par (A)	Mayor a 4 (B)	$A \cap B$	$A \cup B$
{1,2,3,4,5,6}	{2,4,6}	{5,6}	{6}	{2,4,5,6}

Cálculos:

$$P(A) = \frac{3}{6} = 0.5 \quad P(B) = \frac{2}{6} = \frac{1}{3}$$

$$P(A \cap B) = \frac{1}{6}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + \frac{1}{3} - \frac{1}{6} = \frac{2}{3}$$

 **Resultado:** $P(\text{par}) = 0.5$, $P(>4) = 0.333\dots$, $P(\text{par y} > 4) = 0.166\dots$, $P(\text{par o} > 4) = 0.666\dots$

Ejemplo 2: Binomial y esperanzas

Enunciado:

Se realiza 10 lanzamientos de moneda justa ($p=0.5$). Sea X = número de caras. Calcular: (a) $P(X=7)$, (b) $\mathbb{E}[X]$ y $\text{Var}(X)$.

Fórmulas:

Binomial:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

Momentos:

$$\mathbb{E}[X] = np, \quad \text{Var}(X) = np(1-p)$$

Cálculos:

$$P(X = 7) = \binom{10}{7} (0.5)^7 (0.5)^3 = \binom{10}{7} (0.5)^{10} = 120 \times \frac{1}{1024} \approx 0.1172$$

$$\mathbb{E}[X] = 10(0.5) = 5 \quad \text{Var}(X) = 10(0.5)(0.5) = 2.5$$

Resultado: $P(X=7) \approx 0.1172$, esperanza = 5, varianza = 2.5.

Ejemplo 3: Bayes en test diagnóstico

Enunciado:

Prevalencia de una enfermedad: 2%. Test con sensibilidad 95% y especificidad 97%. Calcular $P(\text{enfermo} | \text{test positivo})$.

Datos:

$$P(E) = 0.02 \text{ (prevalencia)}$$

$$P(+|E) = 0.95 \text{ (sensibilidad)}$$

$$P(-|E^c) = 0.97 \rightarrow P(+|E^c) = 0.03 \text{ (falso positivo)}$$

Cálculos con Ley de Probabilidad Total y Bayes:

$$P(+) = P(+|E)P(E) + P(+|E^c)P(E^c) = 0.95(0.02) + 0.03(0.98) = 0.019 +$$

$$P(E|+) = \frac{P(+|E)P(E)}{P(+)} = \frac{0.019}{0.0484} \approx 0.3926$$

Interpretación: A pesar del buen test, con prevalencia baja la probabilidad posterior de estar enfermo dado un positivo es ~39.3%. ¡La prevalencia importa mucho!

Ejemplo 4: Desigualdad de Chebyshev

Enunciado:

Una variable aleatoria X tiene $E[X] = 50$ y $\text{Var}(X) = 25$. Usar Chebyshev para acotar $P(|X - 50| \geq 15)$.

Solución:

Aplicamos Chebyshev con $k = 15$:

$$P(|X - 50| \geq 15) \leq \frac{\text{Var}(X)}{k^2} = \frac{25}{15^2} = \frac{25}{225} = \frac{1}{9} \approx 0.111$$

Interpretación: Al menos 88.9% de las observaciones de X están entre 35 y 65, sin importar la distribución de X .

Combinatoria básica

Permutaciones sin repetición: Ordenar r objetos de n totales.

$$P(n, r) = \frac{n!}{(n-r)!}$$

Combinaciones: Elegir r objetos de n sin importar el orden.

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

Principio multiplicativo: Si una tarea se hace en a formas y otra en b , juntas en $a \cdot b$ formas.

Principio aditivo: Si dos tareas son mutuamente excluyentes y pueden hacerse de a y b formas respectivamente, el total de formas es $a + b$.

Errores comunes

Confundir independencia con exclusión mutua: Eventos mutuamente excluyentes no pueden ocurrir juntos ($P(A \cap B) = 0$); eventos independientes sí pueden ocurrir juntos pero no se afectan ($P(A \cap B) = P(A)P(B)$).

Olivar el complemento: A veces es más fácil calcular $P(A^c)$ y restar de 1.

Asumir equiprobabilidad sin justificación: No todos los resultados son igualmente probables en todos los experimentos.

Aplicar fórmulas fuera de contexto: Usar Binomial cuando los intentos no son independientes, o Poisson cuando λ no es constante.

Ignorar la tasa base (prevalencia): En tests diagnósticos y aplicaciones de Bayes, la prevalencia inicial cambia drásticamente las probabilidades posteriores.

Confundir $P(A|B)$ con $P(B|A)$: La dirección de la condicional importa. Bayes nos ayuda a "revertir" la condicional.

Asumir que covarianza cero implica independencia: Independencia implica covarianza cero, pero el recíproco no siempre es cierto.