Derivación del Teorema de Bayes

El Teorema de Bayes se utiliza para revisar probabilidades previamente calculadas cuando se posee nueva información. Desarrollado por Thomas Bayes en el siglo XVII, el teorema de Bayes es una extensión de lo que se ha aprendido hasta ahora acerca de la probabilidad condicional.

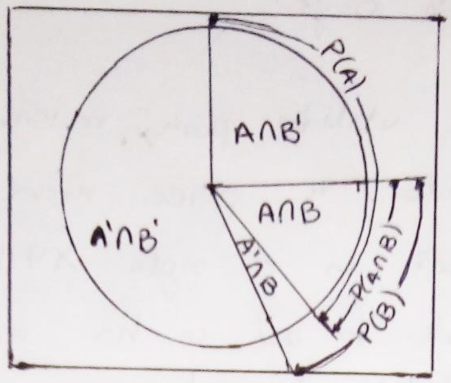
Probabilidad Condicional

En la teoría de la probabilidad, la probabilidad condicional es una medida de la probabilidad de un evento (se produce en alguna situación particular) dado que se ha producido otro evento. Si el evento de interés es A y se sabe o se supone que ocurrió el evento B, "la probabilidad condicional" de A dado B, o "la probabilidad de A en la condición B" generalmente se escribe como $P(A|B)$.

El concepto de probabilidad condicional es uno de los conceptos más fundamentales y uno de los más importantes en la teoría de la probabilidad. Pero las probabilidades condicionales pueden ser bastante resbaladizas y requieren una interpretación cuidadosa. Por ejemplo, no es necesario que exista una relación causal entre A y B y no tienen que ocurrir simultaneamente.

Derivacion formal

Formalmente, $P(A|B)$ se define como la probabilidad de A deacuerdo con una nueva función de probabilidad en el espacio muestral de modo que los resultados que no están en B tienen una probabilidad de 0 y son consistentes con todas las medidas de probabilidad originales.



Gráfica circular de Venn describiendo probabilidades condicionales.

Sea Ω un espacio muestral con eventos elementales w. Supongamos que nos informan que el evento $B \subseteq \Omega$ ha ocurrido. Se debe asignar una nueva distribución de probabilidad en w para reflejar esto. Para los eventos en B, es razonable suponer que las nuevas magnitudes relativas de las probabilidades se conservarán. Para un factor de escala constante α, la nueva distribución por lo tanto satisfará:

1) $w \in B: P(w|B) = \alpha P(w)$

2) $w \notin B: P(w|B) = 0$

3) $\sum_{w \in \Omega} P(w|B) = 1$

Sustituyendo 1 y 2 en 3 para seleccionar α.

$$\begin{aligned}
1 &= \sum_{w \in \Omega} P(w|B) \\
&= \sum_{w \in B} P(w|B) + \underbrace{\cancel{\sum_{w \notin B} P(w|B)}}_{0} \\
&= \alpha \sum_{w \in B} P(w) \\
&= \alpha \cdot P(B)
\end{aligned}$$

$$\alpha = \frac{1}{P(B)}$$

Así que la nueva distribución de probabilidad es

1) $\omega \in B: P(\omega|B) = \frac{P(\omega)}{P(B)}$

2) $\omega \notin B: P(\omega|B) = 0$

Ahora para un evento general A,

$$P(A|B) = \sum_{\omega \in A \cap B} P(\omega|B) + \underbrace{\cancel{\sum_{\omega \in A \cap B^c} P(\omega|B)}}_{0}$$

$$P(A|B) = \sum_{\omega \in A \cap B} \frac{P(\omega)}{P(B)}$$

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

Derivación del Teorema de Bayes

Este teorema se puede derivar usando la probabilidad condicional de $P(A|B)$ que se puede representar como:

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

multiplicando por $P(B)$ obtenemos:

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

También podemos realizar esta misma manipulación para $P(B|A)$

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

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

Observe cómo ahora tenemos dos representaciones alternativas de la probabilidad conjunta $P(A,B)$ que podemos igualar entre sí:

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

y ahora dividimos por $P(B)$

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}, \quad \text{si } P(B) \neq 0$$

Esto se define como el teorema de Bayes. Escencialmente, lo que hemos hecho es relacionar dos ecuaciones de probabilidad condicionales diferentes pero relacionados entre sí.

Digamos que tenemos una hipótesis 'θ' y algunos datos 'y' para apoyar u oponerse a esta hipótesis. Si hacemos que nuestra X (desde arriba) sea igual a θ entonces nuestra regla de Bayes se convierte en:

$$P(\theta|y) = \frac{P(y|\theta)\,P(\theta)}{P(y)}$$

¿Qué significa esto?

$P(\theta)$ representa nuestra creencia "previa" de cual es la hipótesis antes de que veamos datos. Este previo es informado por nuestra opinión de "experto" sobre la hipótesis y por lo tanto es subjetivo. Pero la belleza de esta regla, es que nos permite relacionar este anterior, $P(\theta)$, con una "creencia actualizada" de la hipótesis $P(\theta|y)$ una vez que tengamos datos a considerar. En términos bayesianos, esta creencia actualizada se llama "posterior".