Probabilidad condicional y conjunta y teorema de Bayes

Fabián Abarca Calderón

IE0405 – Modelos Probabilísticos de Señales y Sistemas

2

Tema I

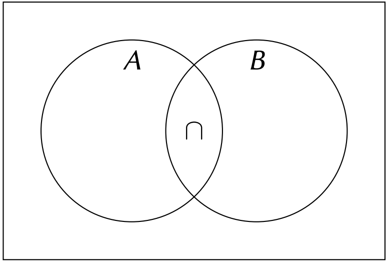
UNIVERSIDAD DE COSTA RICA

EIE Escuela de Ingeniería Eléctrica

Una gran cantidad de “experimentos” y cálculos de probabilidad cotidianos pueden ser descritos por medio de las herramientas de la probabilidad condicional y conjunta, el teorema de Bayes y los eventos independientes, que representan un conocimiento básico y útil de la probabilidad.

Introducción

Algunos eventos no son mutuamente excluyentes debido a elementos comunes en el espacio de probabilidad. Estos elementos corresponden a la ocurrencia *simultánea* o *conjunta* o *coincidente* de los eventos no excluyentes. Para dos eventos A y B, los elementos comunes forman el evento $A \cap B$.



Antecedente de la probabilidad conjunta y condicional II

- La probabilidad **conjunta** describe la probabilidad $P(A \cup B)$ de que ocurra A **o** B.
- La probabilidad **condicional** describe la probabilidad $P(A \cap B)$ de que ocurra A **y** B, y describe cómo cambia la probabilidad de un evento particular dada la ocurrencia de otro.

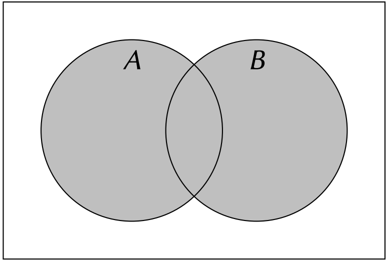
Cuando los eventos sí son mutuamente excluyentes, estos resultados ya se conocen:

- $P(A \cup B) = P(A) + P(B)$
- $P(A \cap B) = P(\varnothing) = 0$

Probabilidad conjunta

La probabilidad $P(A \cup B)$ se llama la *probabilidad conjunta para dos eventos* A y B que se intersecan en el espacio de muestras, y es equivalente a

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)} \tag{1}$$

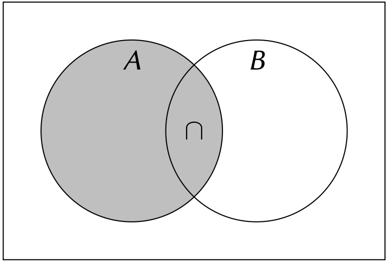


Probabilidad condicional

Dado algún evento A con una probabilidad $0 < P(A) \leq 1$, se define la probabilidad condicional de un evento B, dado A, por

$$\boxed{P(B \mid A) = \frac{P(A \cap B)}{P(A)}} \tag{2}$$

Esto implica que la probabilidad de un evento B se redefine con la previa ocurrencia de A.



La expresión (2) se puede reescribir como $P(A \cap B) = P(A)P(B \mid A)$ y puede leerse intuitivamente como "la probabilidad de que sucedan A **y** B es igual a la probabilidad de que suceda A por la probabilidad de que suceda B dado A".

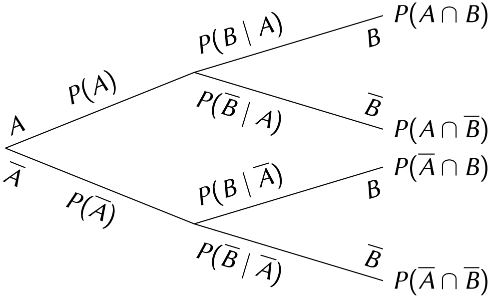


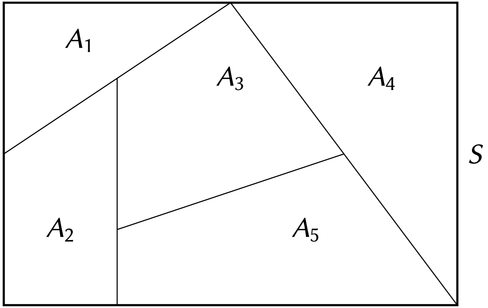
Figura: Es posible construir un árbol binario de decisiones de dos niveles para ver las combinaciones posibles y su relación numérica.

La probabilidad total

El conjunto de N eventos $\{A_n\}$, $n = 1, 2, \ldots, N$, se llama una **partición** si cumple que

Son exhaustivos $\displaystyle\bigcup_{n=1}^{N} A_n = S$

Son mutuamente excluyentes $A_m \cap A_n = \varnothing$ para todo $m \neq n$



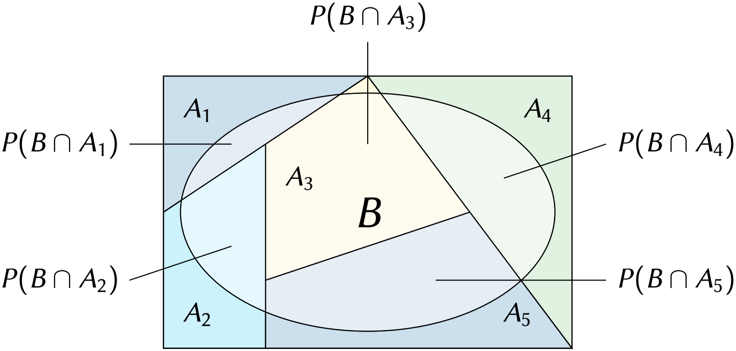
La probabilidad $P(B)$ de cualquier evento B definido sobre un espacio S puede expresarse en términos de probabilidades condicionales.
Si $\{A_n\}$ es una partición, y además

$$\begin{aligned}
B &= B \cap S && \text{(identidad de conjuntos)} \\
&= B \cap \left(\bigcup_{n=1}^{N} A_n \right) && \text{(definición de partición)} \\
&= \bigcup_{n=1}^{N} (B \cap A_n) && \text{(ley de distributividad)}
\end{aligned} \tag{3}$$

Como los eventos $\{B \cap A_n\}$ son mutuamente excluyentes, aplica el tercer axioma de la probabilidad, y utilizando (3) entonces

$$\begin{aligned} P(B) &= P(B \cap S) \\ &= P\left[\bigcup_{n=1}^{N}(B \cap A_n)\right] \\ &= \sum_{n=1}^{N} P(B \cap A_n) \\ \Aboxed{P(B) &= \sum_{n=1}^{N} P(B \mid A_n)P(A_n)} \end{aligned} \tag{4}$$

que se conoce como la *probabilidad total del evento B.*



$$\boxed{P(B) = \sum_{n=1}^{5} P(B \cap A_n) = \sum_{n=1}^{5} P(B \mid A_n) P(A_n)}$$

¿Cuál es la probabilidad de sacar una bolincha (canica) roja de una bolsa, si hay tres bolsas con la siguiente distribución?

- Bolsa 1: 20 rojas, 20 azules, 10 blancas
- Bolsa 2: 10 rojas, 25 azules, 15 blancas
- Bolsa 3: 30 rojas, 15 azules, 5 blancas

Si $R = \{\text{sacar una bolincha roja}\}$ y $B_i = \{\text{escoger la } i\text{-ésima bolsa}\}$, por la probabilidad total se tiene que

$$\begin{aligned} P(R) &= \sum_{i=1}^{3} P(R \mid B_i)P(B_i) \\ &= \frac{1}{3}(0{,}40) + \frac{1}{3}(0{,}20) + \frac{1}{3}(0{,}60) = 0{,}40 \end{aligned}$$

Teorema de Bayes

Como uno de los resultados más útiles de la teoría de probabilidad, el teorema de Bayes permite *actualizar* el conocimiento o *recalcular* la probabilidad de un evento de interés cuando encontramos nueva evidencia de su ocurrencia.

> *La prueba médica (la evidencia) dio negativa. ¿Cuál es la probabilidad de que realmente no tenga la enfermedad?*

Sean A y B eventos en los que se asume inicialmente una dependencia de B a la ocurrencia de A tal que, si $P(A) \neq 0$

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \tag{5}$$

También existe una relación inversa en que, si $P(B) \neq 0$, se cumple que

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

Del álegebra de conjuntos se sabe que $P(B \cap A) = P(A \cap B)$, y es posible igualar (5) y (6), resultando:

Regla de la probabilidad condicional inversa

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} \tag{7}$$

Una ecuación equivalente se obtiene de una sustitución de $P(B)$ en términos de una **probabilidad total**:

$$\boxed{P(A_n \mid B) = \frac{P(B \mid A_n)P(A_n)}{P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_N)P(A_N)}} \tag{8}$$

donde $\{A_n\}$ es una partición universal para $n = 1, 2, \ldots, N$.

> Una intuición importante de esta ecuación es que la ocurrencia de B puede deberse a múltiples factores (en este caso, cualquiera entre $A_1, \ldots, A_N$), pero estamos interesados en la *relación* con **uno** de ellos en particular, A_n.

Nota: esta *relación* puede, o no, ser de **causalidad**.

Teorema de Bayes con probabilidad total II

$$\boxed{P(A_n \mid B) = \frac{P(B \mid A_n)P(A_n)}{P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_N)P(A_N)}}$$

$P(A_n)$ son *probabilidades a priori*, dado que se conocen para cada evento A_n antes de la ejecución del experimento.

$P(B \mid A_n)$ son *probabilidades condicionales directas* o *probabilidades de transición* en teoría de telecomunicaciones. Típicamente son conocidas antes de ejecutar el experimento.

$P(A_n \mid B)$ son *probabilidades a posteriori* o *probabilidades condicionales inversas*, dado que se investigan después de la ejecución del experimento, cuando se obtiene un evento B.

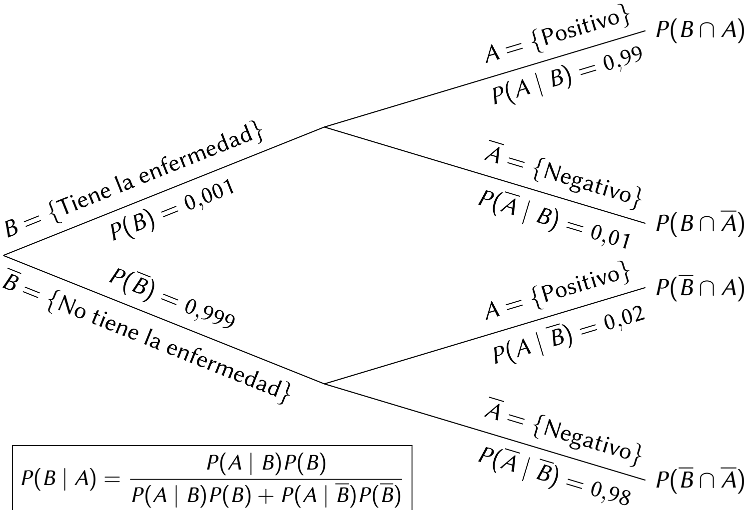
Ejemplo de incidencia de una enfermedad poco común I

Se ha desarrollado un examen de diagnóstico para una enfermedad extraña que afecta solo a 1 de cada 1000 adultos. En análisis estadísticos médicos de clasificación binaria (sí o no) se define:

Sensitividad Un resultado positivo implica que el individuo efectivamente tiene la enfermedad en el 99 % de los casos (también llamada *probabilidad de detección*). Un resultado contrario es un **falso negativo**.

Especificidad Un individuo sin la enfermedad dará resultado negativo 98 % de las veces (también llamada *tasa negativa verdadera*). Un resultado contrario es un **falso positivo**.

¿Cuál es la probabilidad de que el paciente sí tenga la enfermedad si el diagnóstico es positivo?



$$\boxed{P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B})}}$$

Ejemplo de incidencia de una enfermedad poco común III

$$
\begin{aligned}
P(B \mid A) &= \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid \overline{B})P(\overline{B})} \\
&= \frac{(0{,}99)(0{,}001)}{(0{,}99)(0{,}001) + (0{,}02)(0{,}999)} \\
&= 0{,}0472 \approx 5\,\%
\end{aligned}
$$

La probabilidad de que el paciente sí tenga la enfermedad es del 5 %…a pesar de que el resultado del examen fue positivo.

¿Por qué es tan baja esta probabilidad, con una *sensitividad* de la prueba del 99 %?

Ejemplo de incidencia de una enfermedad poco común IV

Cuadro: Otros valores de sensitividad y especificidad para algunos exámenes de enfermedades comunes

Enfermedad	Sensitividad	Especificidad
Cáncer de próstata	85 %	30 %
Cáncer de mama	75 %	92 %
Cáncer de colon	86 %	93 %
COVID-19 BioMedomics	89 %	91 %

Ejemplo del apagón en el sistema eléctrico I

Un barrio de Heredia experimenta un apagón. Una ingeniera de operación y mantenimiento de ESPH está cerca de ahí e inmediatamente sospecha de cuatro orígenes de la falla: (A_1) en la línea de transmisión Colima - Heredia, (A_2) en el transformador de la subestación de Heredia, (A_3) en la línea de distribución hacia San Pablo o en (A_4) el transformador del poste.

> Sabe la ingeniera que ante una falla en el transformador de subestación siempre habrá una desconexión permanente. ¿Cuál es la probabilidad de que la causa de la desconexión permanente haya sido una falla en (A_2) el trafo de subestación?

Nota: Las protecciones del sistema eléctrico ejecutan dos operaciones ante una falla: o (B) una desconexión permanente o ($\overline{B}$) un "recierre" luego de un tiempo prudencial si la falla ha desaparecido.

Los datos conocidos para el último mes son los siguientes:

	Falla	Casos	Recierre	Desconexión
A_1	Línea de transmisión	3	2	1
A_2	Trafo de subestación	3	0	3
A_3	Línea de distribución	16	9	7
A_4	Trafo de poste	8	0	8

Para encontrar la probabilidad de cada falla, analizamos su frecuencia relativa con los datos provistos (se descartan aquí otros tipos de fallas). Por tanto

- $P(A_1) = 3/30 = 0{,}10$
- $P(A_2) = 3/30 = 0{,}10$
- $P(A_3) = 16/30 = 0{,}5333$
- $P(A_4) = 8/30 = 0{,}2666$

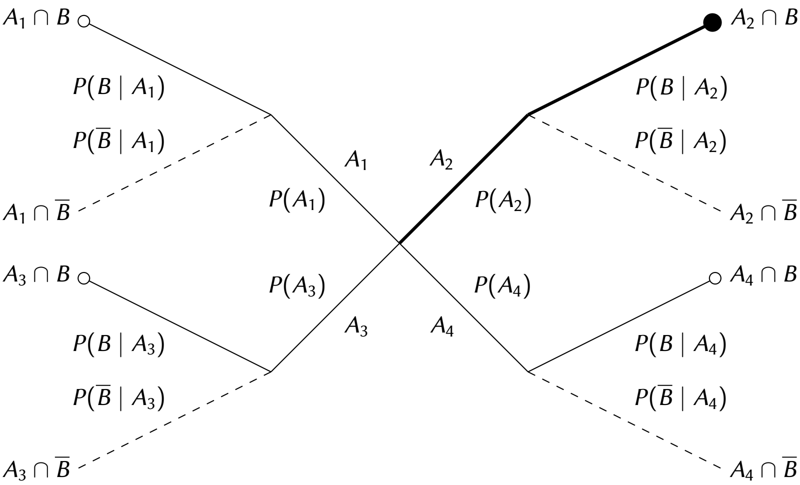
Sean $B = \{\text{desconexión permanente}\}$ y $\overline{B} = \{\text{recierre}\}$. La probabilidad que buscamos es una proporción entre el evento de interés y todas las posibilidades juntas:

$$\begin{array}{c}\text{Probabilidad de que} \\ A_2 \text{ sea la causa de } B\end{array} = \frac{\begin{array}{c}\text{Probabilidad de} \\ \text{que sucedan } A_2 \textbf{ y } B\end{array}}{\begin{array}{c}\text{Suma de las probabilidades de} \\ \text{todas las combinaciones } A_n \textbf{ y } B\end{array}}$$

y que puede expresarse como

$$P(A_2 \mid B) = \boxed{\frac{P(A_2 \cap B)}{P\left(\bigcup_{n=1}^{4} A_n \cap B\right)}} = \frac{P(A_2)P(B \mid A_2)}{P(B)} \tag{9}$$

que es otra forma de escribir la regla de Bayes. La gráfica a continuación muestra los “caminos” posibles y el de interés está resaltado.



$A_1 \cap B$
$P(B \mid A_1)$
$P(\overline{B} \mid A_1)$
$A_1 \cap \overline{B}$
$A_3 \cap B$
$P(B \mid A_3)$
$P(\overline{B} \mid A_3)$
$A_3 \cap \overline{B}$
A_1
$P(A_1)$
A_2
$P(A_2)$
$P(A_3)$
A_3
$P(A_4)$
A_4
$A_2 \cap B$
$P(B \mid A_2)$
$P(\overline{B} \mid A_2)$
$A_2 \cap \overline{B}$
$A_4 \cap B$
$P(B \mid A_4)$
$P(\overline{B} \mid A_4)$
$A_4 \cap \overline{B}$

Con $P(B \mid A_1) = 1/3 = 0{,}3333$, $P(B \mid A_2) = 1$, $P(B \mid A_3) = 7/16 = 0{,}4375$, $P(B \mid A_4) = 1$,

$$\begin{aligned}
P(A_2 \mid B) &= \frac{P(A_2)P(B \mid A_2)}{P(A_1)P(B \mid A_1) + P(A_2)P(B \mid A_2) + P(A_3)P(B \mid A_3) + P(A_4)P(B \mid A_4)} \\
&= \frac{(0{,}1)(1)}{(0{,}1)(0{,}3333) + (0{,}1)(1) + (0{,}5333)(0{,}4375) + (0{,}2666)(1)} \\
&= 0,1579
\end{aligned}$$

que representa una probabilidad quizá más baja de lo esperado. A pesar de que una falla en el transformador de subestación **siempre** provoca una desconexión permanente (y el consiguiente apagón), no son fallas comunes, y por eso su probabilidad sigue siendo baja. En este problema las probabilidades más alta son las de fallas de líneas de distribución $P(B \mid A_3) = 0{,}3684$ (expuestas a ramas, choques, etc.) y los trafos de poste $P(B \mid A_4) = 0{,}4210$, que son menos casos pero siempre implican desconexión.