Bioestadística: Cálculo de Probabilidades

M. González

Departamento de Matemáticas. Universidad de Extremadura

1. Considera una familia con una madre, un padre, y dos hijos. Sea $A_1 = \{$ la madre tiene la enfermedad E $\}$, $A_2 = \{$ el padre tiene la enfermedad E $\}$, $A_3 = \{$ el primer hijo tiene la enfermedad E $\}$, $A_4 = \{$ el segundo hijo tiene la enfermedad E $\}$, $B = \{$ al menos un hijo tiene la enfermedad E $\}$, $C = \{$ al menos uno de los padres tiene la enfermedad E $\}$, $D = \{$ al menos un miembro de la familia tiene la enfermedad E $\}$.

(a) ¿Qué significa $A_1 \cup A_2$? ¿Qué significa $A_1 \cap A_2$?

(b) ¿Son A_3 y A_4 incompatibles?

(c) ¿Qué significa $A_3 \cup B$? ¿Qué significa $A_3 \cap B$?

(d) Expresa C en términos de A_1, A_2, A_3, A_4.

(e) Expresa D en términos de B y C.

(f) ¿Qué significa A_1^c? ¿Qué significa A_2^c?

(g) Expresa C^c en términos de A_1, A_2, A_3, A_4.

(h) Expresa D^c en términos de B y C.

1. Considera una familia con una madre, un padre, y dos hijos. Sea $A_1 = \{$ la madre tiene la enfermedad E $\}$, $A_2 = \{$ el padre tiene la enfermedad E $\}$, $A_3 = \{$ el primer hijo tiene la enfermedad E $\}$, $A_4 = \{$ el segundo hijo tiene la enfermedad E $\}$, $B = \{$ al menos un hijo tiene la enfermedad E $\}$, $C = \{$ al menos uno de los padres tiene la enfermedad E $\}$, $D = \{$ al menos un miembro de la familia tiene la enfermedad E $\}$.

(i) Supongamos que la probabilidad de que cada hijo tenga la enfermedad E es 0.2, mientras que en un 10% de las familias ambos hijos tienen la enfermedad E. ¿Cuál es la probabilidad de que en una familia al menos un hijo tenga la enfermedad E?

Solución

$P(A_3) = 0.2$, $P(A_4) = 0.2$, $P(A_3 \cap A_4) = 0.1$

$$P(A_3 \cup A_4) = P(A_3) + P(A_4) - P(A_3 \cap A_4) = 0.2+0.2-0.1=0.3$$

1. Considera una familia con una madre, un padre, y dos hijos. Sea $A_1 = \{$ la madre tiene la enfermedad E $\}$, $A_2 = \{$ el padre tiene la enfermedad E $\}$, $A_3 = \{$ el primer hijo tiene la enfermedad E $\}$, $A_4 = \{$ el segundo hijo tiene la enfermedad E $\}$, $B = \{$ al menos un hijo tiene la enfermedad E $\}$, $C = \{$ al menos uno de los padres tiene la enfermedad E $\}$, $D = \{$ al menos un miembro de la familia tiene la enfermedad E $\}$.

(j) Supongamos que en un 10% de las familias la madre tiene la enfermedad E; en un 10% de las familias el padre tiene la enfermedad E; y en un 2% de las familias ambos, el padre y la madre, tienen la enfermedad E. ¿Son independientes los sucesos A_1 y A_2?

Solución

$P(A_1) = 0.1$, $P(A_2) = 0.1$, $P(A_1 \cap A_2) = 0.02$

$$P(A_1 \cap A_2) \neq P(A_1)P(A_2) \qquad \text{DEPENDIENTES}$$

1. Considera una familia con una madre, un padre, y dos hijos. Sea $A_1 = \{$ la madre tiene la enfermedad E $\}$, $A_2 = \{$ el padre tiene la enfermedad E $\}$, $A_3 = \{$ el primer hijo tiene la enfermedad E $\}$, $A_4 = \{$ el segundo hijo tiene la enfermedad E $\}$, $B = \{$ al menos un hijo tiene la enfermedad E $\}$, $C = \{$ al menos uno de los padres tiene la enfermedad E $\}$, $D = \{$ al menos un miembro de la familia tiene la enfermedad E $\}$.

(k) ¿Cuál es la probabilidad de que el padre tenga la enfermedad E si la madre la padece?

Solución

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_1)} = \frac{0.02}{0.1} = 0.2$$

$P(A_1) = 0.1$, $P(A_2) = 0.1$, $P(A_1 \cap A_2) = 0.02$

1. Considera una familia con una madre, un padre, y dos hijos. Sea $A_1 = \{$ la madre tiene la enfermedad E $\}$, $A_2 = \{$ el padre tiene la enfermedad E $\}$.

(l) ¿Cuál es la probabilidad de que el padre tenga la enfermedad E si la madre no la tiene?

Solución

$$\begin{aligned} P(A_2|A_1^c) &= \frac{P(A_1^c \cap A_2)}{P(A_1^c)} = \frac{P(A_2 \setminus (A_1 \cap A_2))}{1 - P(A_1)} \\ &= \frac{P(A_2) - P(A_1 \cap A_2)}{1 - P(A_1)} = \frac{0.1 - 0.02}{1 - 0.1} = 0.089 \end{aligned}$$

$A \subseteq B$, $P(B \setminus A) = P(B) - P(A)$

$P(A_1) = 0.1$, $P(A_2) = 0.1$, $P(A_1 \cap A_2) = 0.02$

$$A_1, A_2, \ldots, A_{n-2}, A_{n-1}, A_n$$

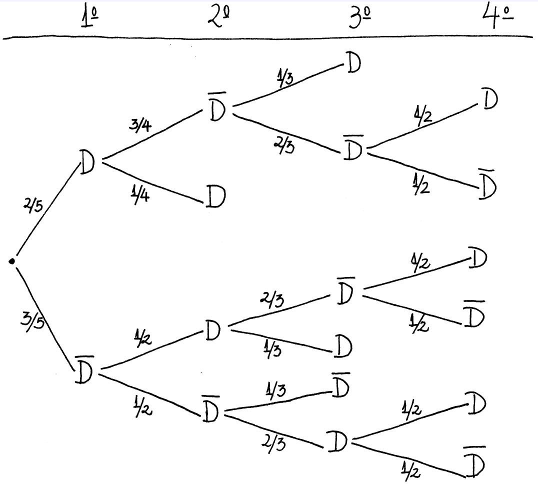
$$P(A_n \cap A_{n-1} \cap \cdots \cap A_2 \cap A_1) = P(A_n | A_{n-1} \cap \cdots \cap A_1) P(A_{n-1} | A_{n-2} \cap \cdots \cap A_1) \cdots P(A_2 | A_1) P(A_1)$$

14. Una caja contiene 5 tubos de ensayo, de los cuales dos son defectuosos. Se prueban los tubos unos tras otros hasta que se descubren los dos defectuosos. ¿Cuál es la probabilidad de que se suspenda el proceso:

(a) en la segunda prueba?

(b) en la tercera prueba?

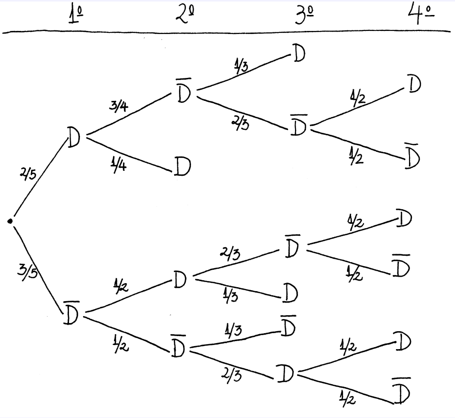
14.



$$\Omega = \{DD, D\bar{D}D, D\bar{D}\bar{D}D, D\bar{D}\bar{D}\bar{D}, \bar{D}D\bar{D}D, \bar{D}D\bar{D}\bar{D}, \bar{D}DD, \bar{D}\bar{D}\bar{D}, \bar{D}\bar{D}DD, \bar{D}\bar{D}D\bar{D}\}$$

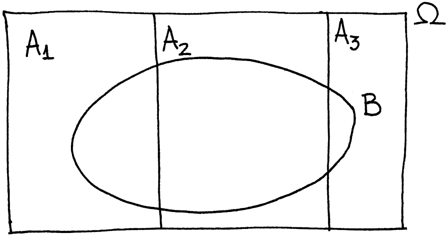
Cálculo de Probabilidades: Teorema de la Multiplicación

14.



$$
\begin{aligned}
P(DD) &= P(D_1 \cap D_2) = P(D_2|D_1)P(D_1) = \frac{1}{4}\frac{2}{5} \\
P(D\bar{D}D) &= P(D_1 \cap \bar{D}_2 \cap D_3) = P(D_3|\bar{D}_2 \cap D_1)P(\bar{D}_2|D_1)P(D_1) = \frac{1}{3}\frac{3}{4}\frac{2}{5} \\
&\ldots
\end{aligned}
$$

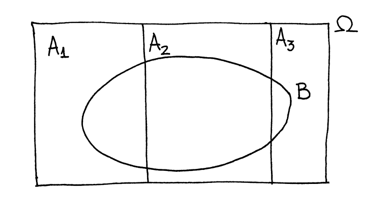
Cálculo de Probabilidades: Teorema de la Probabilidad Total



$$P(B) = P(B|A_1)\cdot P(A_1) + P(B|A_2)\cdot P(A_2) + P(B|A_3)P(A_3)$$

15. Un paciente con un conjunto de síntomas puede tener cualquiera de las tres enfermedades A_1, A_2 ó A_3 con probabilidades 0.5, 0.4 y 0.1, respectivamente. Para precisar el diagnóstico se somete al paciente a un análisis de sangre que da positivo (y notaremos por B a este hecho) en las personas que padecen A_1, A_2 ó A_3 con probabilidades 0.3, 0.98 y 0.2, respectivamente.

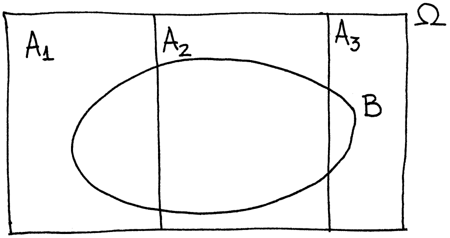
(a) ¿En qué porcentaje de la población de pacientes con tales síntomas el análisis da positivo?



$P(A_1) = 0.5$, $P(A_2) = 0.4$, $P(A_3) = 0.1$

$P(B|A_1) = 0.3$, $P(B|A_2) = 0.98$, $P(B|A_3) = 0.2$

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \\ &= 0.3 * 0.5 + 0.98 * 0.4 + 0.2 * 0.1 = 0.562 \end{aligned}$$

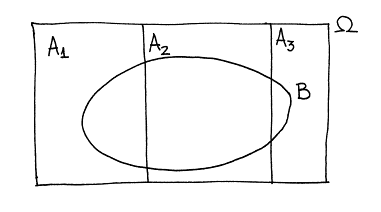


$$P(A_i|B) = \frac{P(B|A_i)\cdot P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)\cdot P(A_2) + P(B|A_3)\cdot P(A_3)}$$

$i = 1, 2, 3$

15.

(b) Si a una persona con los síntomas se le realiza el análisis y da positivo, ¿cuál es la enfermedad más probable?

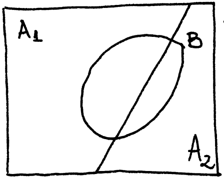


$P(A_1) = 0.5$, $P(A_2) = 0.4$, $P(A_3) = 0.1$

$P(B|A_1) = 0.3$, $P(B|A_2) = 0.98$, $P(B|A_3) = 0.2$

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \\ &= \frac{0.3 * 0.5}{0.562} = 0.267 \\ P(A_2|B) &= 0.698 \\ P(A_3|B) &= 0.036 \end{aligned}$$

Cálculo de Probabilidades: Teorema de Bayes



$\Omega \quad P(A_1) \quad P(A_2) \quad P(B|A_1) \quad P(B|A_2)$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)}$$

$i = 1, 2$

TEST DE DIAGNÓSTICO

E (estar enfermo) $\quad S = E^c$ (estar sano)

D_+ (diagnóstico enfermo) $\quad D_- = D_+^c$ (diagnóstico sano)

$P(E)$ — PREVALENCIA

SENSIBILIDAD: $P(D_+|E)$ $\quad$ FALSO-NEGATIVO: $P(D_-|E)$

ESPECIFICIDAD: $P(D_-|S)$ $\quad$ FALSO-POSITIVO: $P(D_+|S)$

- VALOR PREDICTIVO POSITIVO: $P(E|D_+)$
- VALOR PREDICTIVO NEGATIVO: $P(S|D_-)$

Cálculo de Probabilidades: Teorema de Bayes

TEST DE DIAGNÓSTICO

E (estar enfermo) $S = E^c$ (estar sano)

D_+ (diagnóstico enfermo) $D_- = D_+^c$ (diagnóstico sano)

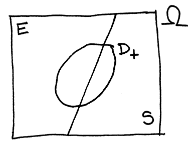
$P(E)$ — PREVALENCIA

SENSIBILIDAD: $P(D_+|E)$ FALSO-NEGATIVO: $P(D_-|E)$

ESPECIFICIDAD: $P(D_-|S)$ FALSO-POSITIVO: $P(D_+|S)$

- VALOR PREDICTIVO POSITIVO:

$$P(E|D_+) = \frac{P(D_+|E)P(E)}{P(D_+|E)P(E) + P(D_+|S)P(S)}$$



- VALOR PREDICTIVO NEGATIVO:

$$P(S|D_-) = \frac{P(D_-|S)P(S)}{P(D_-|E)P(E) + P(D_-|S)P(S)}$$

18. Para estudiar la eficacia de un nuevo test para el diagnóstico de un tipo particular de cáncer que lo padece el 1% de las mujeres de edad avanzada, se aplicó el mismo a un grupo amplio de mujeres con tal tipo de cáncer y a otro grupo de mujeres sanas, obteniéndose la siguiente tabla:

		RESULTADO DEL TEST		
		D_+	D_-	
PRESENCIA DEL CÁNCER	SI(E)	850	150	1000
	NO(S)	45	1455	1500

(a) Estima la sensibilidad y la especificidad de la nueva prueba diagnóstica.

$$\textit{SENSIBILIDAD} \quad P(D_+|E) \simeq \frac{850}{1000} = 0.85$$

$$\textit{FALSO} - \textit{NEGATIVO} \quad P(D_-|E) \simeq 0.15$$

$$\textit{ESPECIFICIDAD} \quad P(D_-|S) \simeq \frac{1455}{1500} = 0.97$$

$$\textit{FALSO} - \textit{POSITIVO} \quad P(D_+|S) \simeq 0.03$$

18.

(b) Si tomada una persona y aplicado el test, este da positivo, ¿qué probabilidad tiene de padecer la enfermedad?

$P(E) = 0.01$

$P(D_+|E) = 0.85 \quad P(D_-|E) = 0.15 \quad P(D_-|S) = 0.97 \quad P(D_+|S) = 0.03$

$$\begin{aligned} P(E|D_+) &= \frac{P(D_+|E)P(E)}{P(D_+|E)P(E) + P(D_+|S)P(S)} \\ &= \frac{0.85 * 0.01}{0.85 * 0.01 + 0.03 * 0.99} = 0.223 \end{aligned}$$

$P(S|D_+) = 0.777$

18.

(c) ¿Cuál es el valor predictivo negativo de esta prueba diagnóstica?

$P(E) = 0.01$

$P(D_+|E) = 0.85 \quad P(D_-|E) = 0.15 \quad P(D_-|S) = 0.97 \quad P(D_+|S) = 0.03$

$$\begin{aligned} P(S|D_-) &= \frac{P(D_-|S)P(S)}{P(D_-|E)P(E) + P(D_-|S)P(S)} \\ &= \frac{0.97 * 0.99}{0.15 * 0.01 + 0.97 * 0.99} = 0.998 \end{aligned}$$

$P(E|D_-) = 0.002$

INFLUENCIA DE LA PREVALENCIA, $P(E)$, EN LOS VALORES PREDICTIVOS POSITIVO, $P(E|D_+)$, Y NEGATIVO, $P(S|D_-)$.

$P(E)$	90	80	70	60	50	40	30	20	10	5	1
VPP	99.6	99.1	98.5	97.7	96.6	95	92.4	87.6	75.9	59.9	22.3
VPN	41.8	61.8	73.5	81.2	86.6	90.7	93.8	96.3	98.3	99.2	99.8