



INFOMAAT
CENTRO RIPETIZIONI
SINCE 2011

UNIVERSITÀ · STATISTICA E PROBABILITÀ

Probabilità condizionata e teorema di Bayes

Esercitazione svolta — 9 problemi di probabilità applicata, con alberi di calcolo, tabelle di contingenza e soluzioni passo passo.

9 problemi · alberi di Bayes · soluzioni complete

Nove problemi, dal calcolo diretto all'inversione di Bayes

Ogni problema riporta la consegna riformulata in italiano e una soluzione completa. Nei problemi di Bayes la struttura del calcolo è resa esplicita con un albero delle probabilità.

1	Fattori di rischio cardiovascolare	CONTINGENZA
2	Polmonite o embolia in pronto soccorso	BAYES
3	Quale équipe ha fatto la diagnosi?	BAYES
4	Parità e aumento di peso in gravidanza	CONTINGENZA
5	Da cosa dipende il vomito?	BAYES
6	Quanto è affidabile il test?	BAYES
7	Valore predittivo di un test negativo	BAYES
8	Quale veleno ha ucciso la cavia?	BAYES
9	Sensibilità, specificità e valori predittivi	DA COMPLETARE

ESERCIZIO 1 ★☆☆

Fattori di rischio cardiovascolare

In uno studio su 10 000 persone che hanno avuto un infarto del miocardio si analizzano due fattori di rischio: il **fumo** e l'**ipertensione**. Nel campione risultano 5 000 fumatori e 2 000 ipertesi; fra i 2 000 ipertesi, 1 500 sono anche fumatori.

Calcola: **a)** la probabilità che una persona sia fumatrice; **b)** la probabilità che sia fumatrice sapendo che è ipertesa; **c)** la probabilità che presenti entrambi i fattori di rischio.

SOLUZIONE

I dati si organizzano in una tabella a doppia entrata. Conoscendo i totali e l'incrocio fra fumo e ipertensione, ricaviamo tutte le caselle per differenza.

	Fuma (F)	Non fuma (F ^c)	Totale
Iperteso (I)	1 500	500	2 000
Non iperteso (I ^c)	3 500	4 500	8 000
Totale	5 000	5 000	10 000

Tabella di contingenza: i dati noti (1 500, 2 000, 5 000, 10 000) determinano tutte le altre caselle.

a) La probabilità di essere fumatori è la frequenza relativa dei fumatori sul totale:

$$P(F) = \frac{5000}{10000} = \frac{1}{2} = 0.5$$

b) Ci limitiamo ai soli ipertesi (2 000) e contiamo quanti fra loro fumano (1 500):

$$P(F | I) = \frac{P(F \cap I)}{P(I)} = \frac{1500/10000}{2000/10000} = \frac{1500}{2000} = 0.75$$

c) La probabilità di avere entrambi i fattori è la casella d'incrocio sul totale:

$$P(F \cap I) = \frac{1500}{10000} = 0.15$$

RISULTATO

a) P(F)	0.5
b) P(F I)	0.75
c) P(F ∩ I)	0.15

Polmonite o embolia in pronto soccorso

In un gruppo di 152 pazienti giunti al pronto soccorso con dolore toracico, 120 ricevono diagnosi di **polmonite (N)** e 32 di **embolia polmonare (E)** — le due diagnosi coprono tutti i 152 pazienti. La **dispnea (D)** è presente in 92 pazienti con polmonite e in 20 con embolia.

Calcola: **a)** $P(E)$; **b)** $P(E|D)$; **c)** $P(\text{polmonite senza dispnea})$; **d)** la probabilità che due pazienti abbiano entrambi la polmonite; **e)** che, presi due pazienti, uno abbia la polmonite e l'altro l'embolia; **f)** che due pazienti presentino entrambi polmonite senza dispnea.

SOLUZIONE

	Dispnea (D)	No dispnea (D^c)	Totale
Polmonite (N)	92	28	120
Embolia (E)	20	12	32
Totale	112	40	152

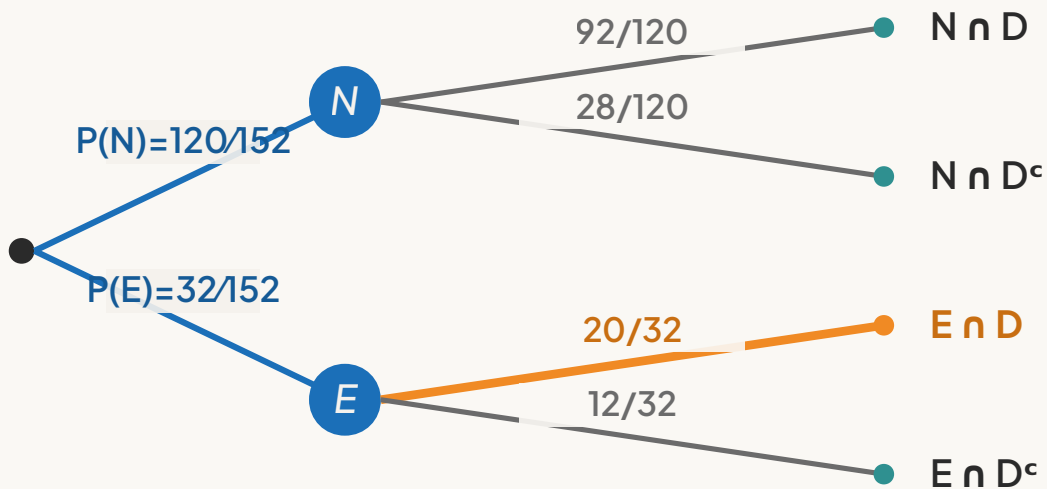
Tabella di contingenza dei 152 pazienti.

a) Embolia su totale:

$$P(E) = \frac{32}{152} \approx 0.21$$

b) Vogliamo invertire il condizionamento: data la dispnea, quanto vale la probabilità di embolia. È un problema di Bayes; l'albero rende esplicite le ramificazioni.

Albero diagnosi → dispnea (punto b)



Per Bayes (oppure leggendo direttamente la colonna D della tabella: 20 dei 112 dispnoici hanno embolia):

$$P(E | D) = \frac{P(D | E)P(E)}{P(D | N)P(N) + P(D | E)P(E)} = \frac{20}{112} \approx 0.179$$

c) Polmonite senza dispnea è la casella $N \cap D^c$:

$$P(N \cap D^c) = \frac{28}{152} \approx 0.184$$

Per i punti d-f si considerano due pazienti estratti in modo *indipendente* (popolazione ampia), usando le probabilità marginali.

d) Entrambi con polmonite:

$$\left(\frac{120}{152} \right)^2 \approx 0.623$$

e) Uno con polmonite e uno con embolia: i due ordini possibili raddoppiano il prodotto.

$$2 \cdot \frac{120}{152} \cdot \frac{32}{152} \approx 0.332$$

f) Entrambi con polmonite senza dispnea:

$$\left(\frac{28}{152} \right)^2 \approx 0.034$$

RISULTATO

a) $P(E)$	0.21
b) $P(E D)$	0.179
c) $P(N \cap D^c)$	0.184
d)	0.623
e)	0.332
f)	0.034

ESERCIZIO 3 ★★☆☆

Quale équipe ha fatto la diagnosi?

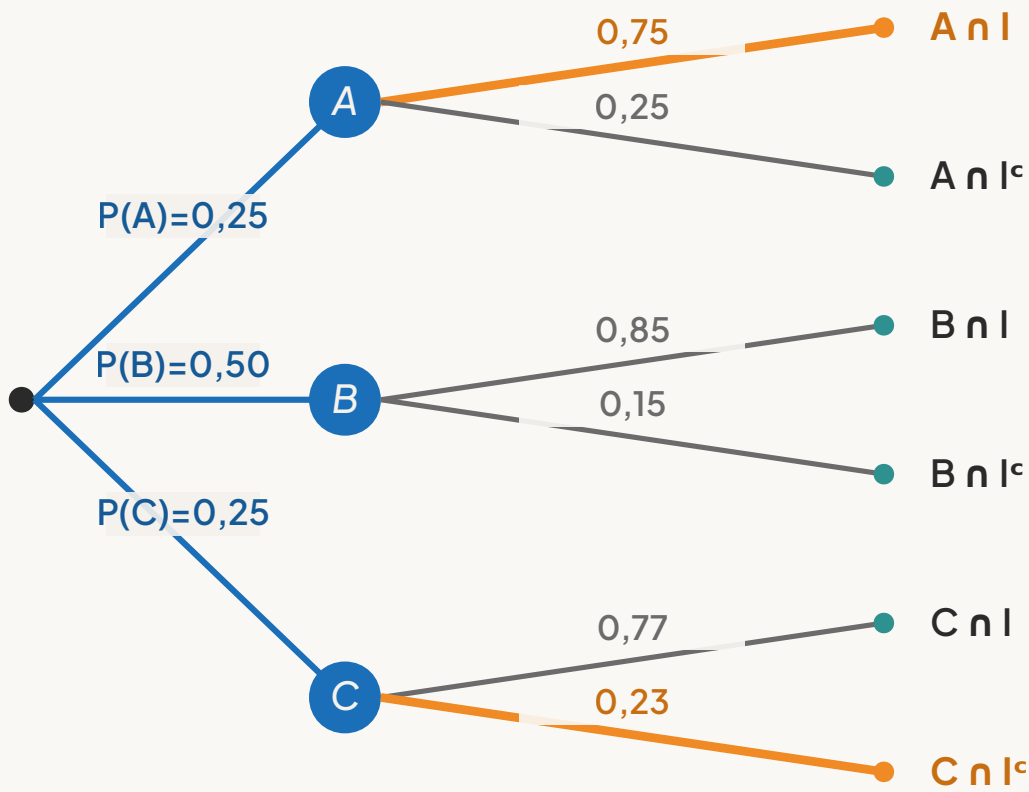
In un poliambulatorio operano tre équipes di medicina di base: **A**, **B** e **C**. L'équipe A segue il 25% degli assistiti, la B il 50% e la C il restante 25%. La quota di pazienti diagnosticati già alla **prima visita (I)** è, rispettivamente, 75%, 85% e 77%.

Calcola: **a)** la probabilità che un paziente diagnosticato alla prima visita sia seguito dall'équipe A; **b)** la probabilità che un paziente NON diagnosticato alla prima visita sia seguito dall'équipe C.

SOLUZIONE

Conosciamo le probabilità a priori delle tre équipes e le condizionate dell'esito: è la struttura tipica del teorema di Bayes. L'albero mostra i sei percorsi; in arancione i rami che servono per le due domande.

Équipe → esito prima visita



a) Per Bayes, la probabilità a posteriori che il paziente venga dall'équipe A:

$$\begin{aligned}
 P(A | I) &= \frac{P(I | A)P(A)}{P(I | A)P(A) + P(I | B)P(B) + P(I | C)P(C)} \\
 &= \frac{0,75 \cdot 0,25}{0,75 \cdot 0,25 + 0,85 \cdot 0,50 + 0,77 \cdot 0,25} = \frac{0,1875}{0,805} \approx 0,233
 \end{aligned}$$

b) Lavoriamo sull'evento complementare I^c (non diagnosticato), con le probabilità di non-diagnosi 0,25 / 0,15 / 0,23:

$$P(C | I^c) = \frac{0,23 \cdot 0,25}{0,25 \cdot 0,25 + 0,15 \cdot 0,50 + 0,23 \cdot 0,25} = \frac{0,0575}{0,195} \approx 0,295$$

RISULTATO

a) $P(A | I) \approx 0,233$

b) $P(C | I^c) \approx 0,295$

ESERCIZIO 4 ★☆☆

Parità e aumento di peso in gravidanza

Su un campione di 350 gestanti, 130 sono **primipare (P)** — quindi 220 sono **multipare (M)**.
Fra le primipare, 70 hanno superato l'aumento di peso raccomandato (**sovrappeso, S**); fra le multipare, 90 hanno superato lo stesso limite.

Calcola: **a)** $P(\text{primipara e sovrappeso})$; **b)** $P(\text{multipara e sovrappeso})$; **c)** $P(\text{multipara} \mid \text{ha superato il peso})$; **d)** $P(\text{primipara} \mid \text{non ha superato il peso})$.

SOLUZIONE

	Sovrappeso (S)	No sovrappeso (S^c)	Totale
Primipara (P)	70	60	130
Multipara (M)	90	130	220
Totale	160	190	350

Tabella di contingenza delle 350 gestanti.

a) Primipara e sovrappeso (casella incrocio sul totale):

$$P(P \cap S) = \frac{70}{350} = 0.2$$

b) Multipara e sovrappeso:

$$P(M \cap S) = \frac{90}{350} \approx 0.257$$

c) Ci limitiamo a chi è in sovrappeso (160) e contiamo le multipare (90):

$$P(M \mid S) = \frac{90}{160} \approx 0.563$$

d) Ci limitiamo a chi NON è in sovrappeso (190) e contiamo le primipare (60):

$$P(P \mid S^c) = \frac{60}{190} \approx 0.316$$

RISULTATO

a) $P(P \cap S)$	0.2
b) $P(M \cap S)$	≈ 0.257
c) $P(M \mid S)$	≈ 0.563
d) $P(P \mid S^c)$	≈ 0.316

ESERCIZIO 5 ★★☆☆

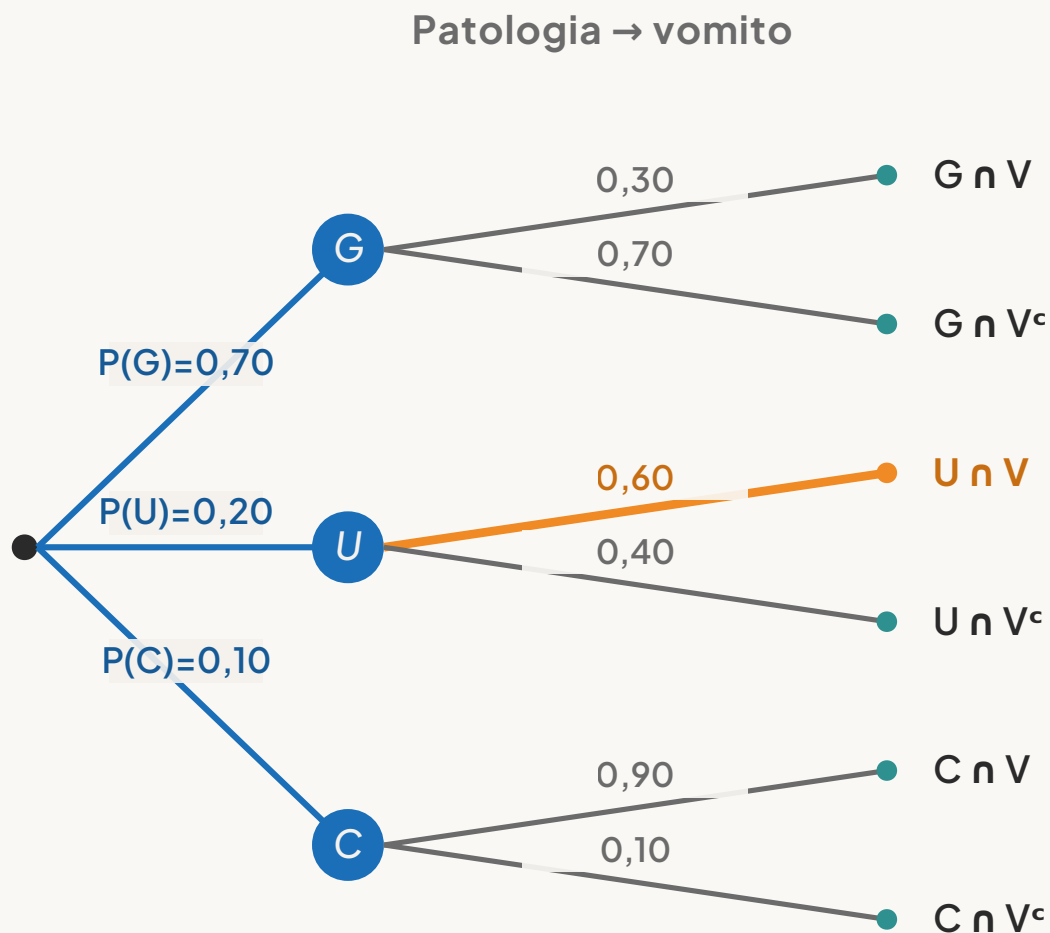
Da cosa dipende il vomito?

In un reparto di gastroenterologia il 70% dei ricoverati ha una **gastrite (G)**, il 20% un'**ulcera (U)** e il 10% un **tumore dello stomaco (C)**. Il **vomito (V)** è presente nel 30% delle gastriti, nel 60% delle ulcere e nel 90% dei tumori.

Un paziente arriva con vomito. Calcola: **a)** la probabilità che il vomito sia dovuto a un'ulcera;
b) la probabilità che sia dovuto a una gastrite.

SOLUZIONE

Tre cause possibili con probabilità a priori note e la presenza di vomito come effetto: di nuovo Bayes. Calcoliamo prima la probabilità totale di vomito (somma dei rami che portano a V).



$$P(V) = 0.30 \cdot 0.70 + 0.60 \cdot 0.20 + 0.90 \cdot 0.10 = 0.21 + 0.12 + 0.09 = 0.42$$

a) Probabilità che il vomito sia dovuto all'ulcera:

$$P(U | V) = \frac{0.60 \cdot 0.20}{0.42} = \frac{0.12}{0.42} \approx 0.286$$

b) Probabilità che sia dovuto alla gastrite:

$$P(G | V) = \frac{0.30 \cdot 0.70}{0.42} = \frac{0.21}{0.42} = 0.5$$

RISULTATO

a) $P(U | V) \approx 0.286$

b) $P(G | V) = 0.5$

ESERCIZIO 6 ★★☆☆

Quanto è affidabile il test?

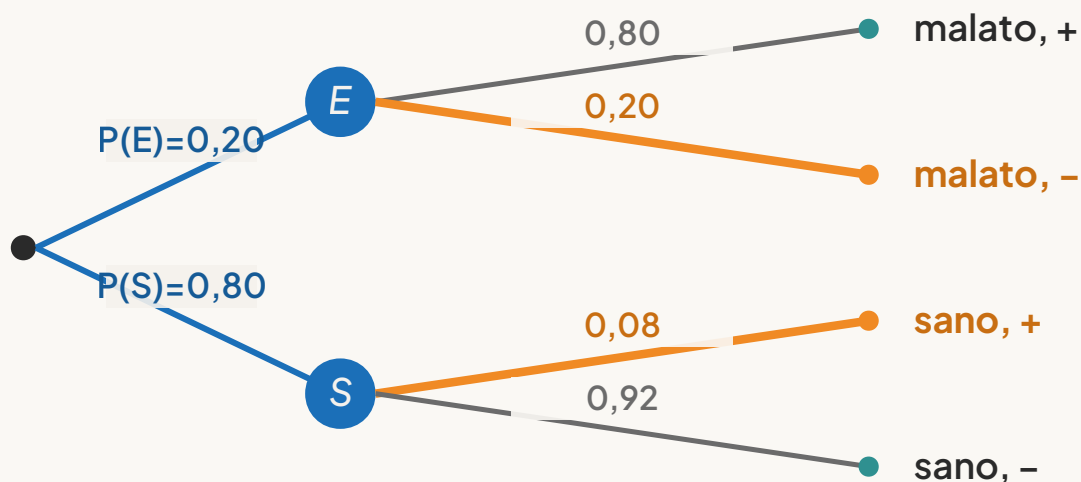
Un test diagnostico ha **sensibilità** dell'80% e **specificità** del 92%; la **prevalenza** della malattia è del 20%.

Calcola: **a)** la probabilità che una persona risultata **positiva** non sia in realtà malata; **b)** la probabilità che una persona risultata **negativa** sia comunque malata.

SOLUZIONE

Sensibilità = $P(+|malato) = 0,80$; specificità = $P(-|sano) = 0,92$, da cui $P(+|sano) = 0,08$ e $P(-|malato) = 0,20$. L'albero parte dallo stato reale (malato E / sano S) e si dirama nell'esito del test.

Stato reale → esito del test



a) Si chiede $P(\text{sano} \mid \text{positivo})$, cioè il complemento del valore predittivo positivo. La probabilità totale di positivo è:

$$P(+) = 0.80 \cdot 0.20 + 0.08 \cdot 0.80 = 0.16 + 0.064 = 0.224$$
$$P(S \mid +) = \frac{P(+ \mid S)P(S)}{P(+)} = \frac{0.08 \cdot 0.80}{0.224} = \frac{0.064}{0.224} \approx 0.286$$

b) Si chiede $P(\text{malato} \mid \text{negativo})$, complemento del valore predittivo negativo. La probabilità totale di negativo è $1 - 0.224 = 0.776$:

$$P(E \mid -) = \frac{P(- \mid E)P(E)}{P(-)} = \frac{0.20 \cdot 0.20}{0.776} = \frac{0.04}{0.776} \approx 0.052$$

RISULTATO

a) $P(\text{sano} \mid +) \approx 0.286$

b) $P(\text{malato} \mid -) \approx 0.052$

ESERCIZIO 7 ★★☆☆

Valore predittivo di un test negativo

Un test per un tumore ha un tasso di **falsi positivi** pari a 0,05 e di **falsi negativi** pari a 0,10.

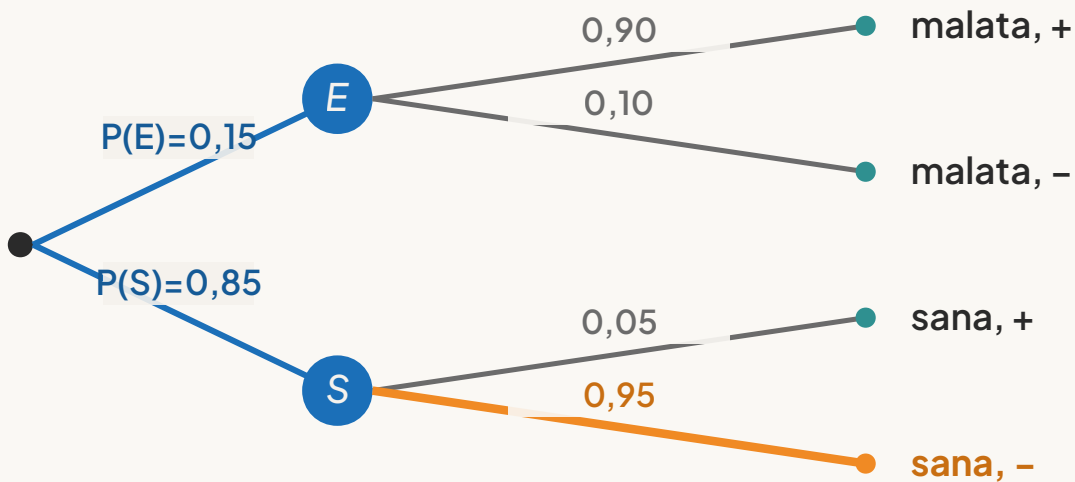
Una paziente con probabilità **pre-test** di malattia pari a 0,15 ottiene un risultato **negativo**.

Calcola la probabilità che non sia malata.

SOLUZIONE

Dai tassi di errore: falso positivo $P(+ \mid \text{sana}) = 0,05 \rightarrow$ specificità $P(- \mid \text{sana}) = 0,95$; falso negativo $P(- \mid \text{malata}) = 0,10 \rightarrow$ sensibilità $P(+ \mid \text{malata}) = 0,90$. La probabilità pre-test è la prevalenza individuale: $P(E) = 0,15$.

Stato reale → esito del test



Si chiede $P(\text{sana} \mid \text{negativo})$, cioè il valore predittivo negativo. Probabilità totale di esito negativo:

$$P(-) = 0,95 \cdot 0,85 + 0,10 \cdot 0,15 = 0,8075 + 0,015 = 0,8225$$

$$P(S \mid -) = \frac{P(- \mid S)P(S)}{P(-)} = \frac{0,95 \cdot 0,85}{0,8225} = \frac{0,8075}{0,8225} \approx 0,982$$

RISULTATO

$$P(\text{sana} \mid -) \approx 0,982$$

ESERCIZIO 8 ★★☆☆

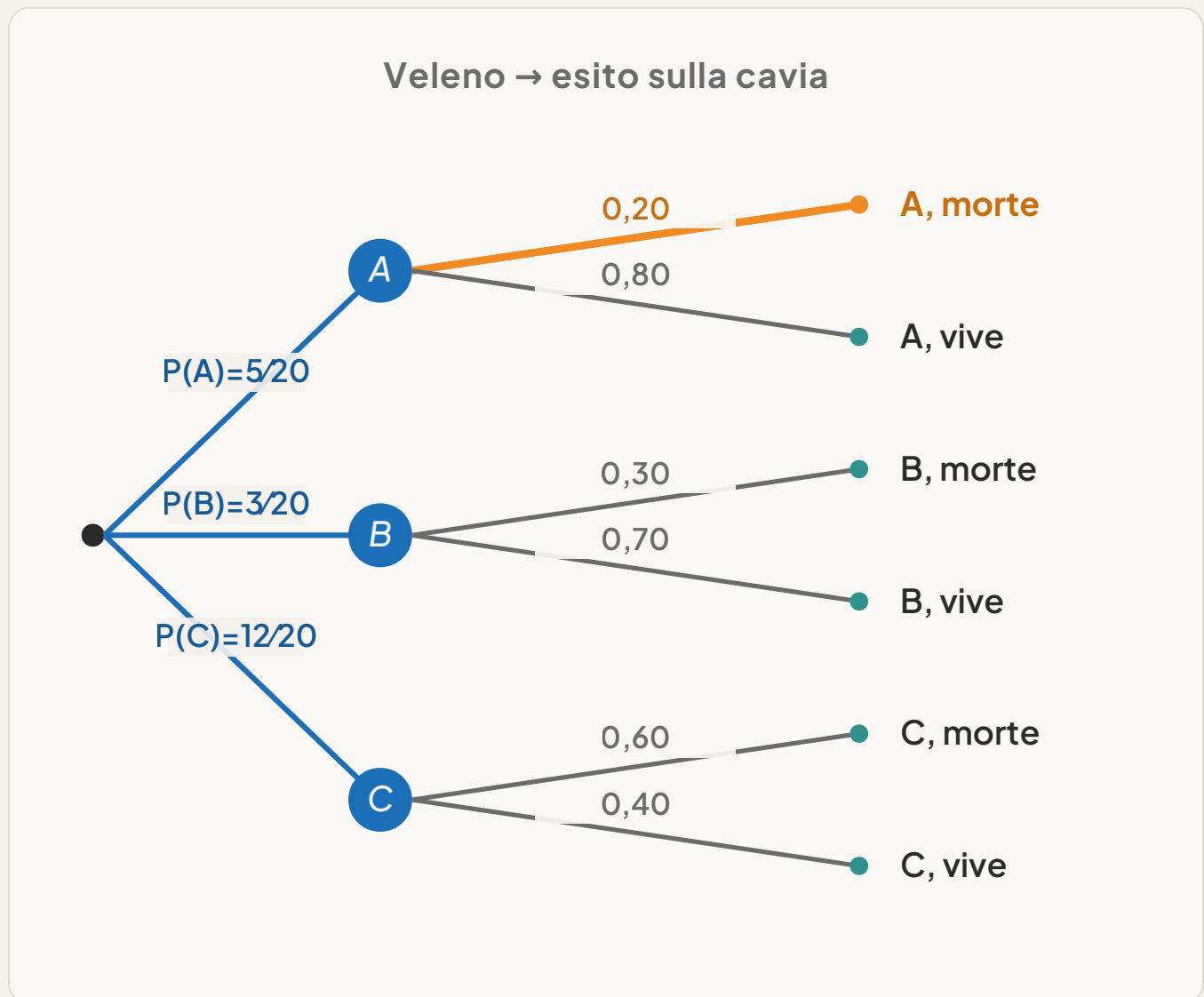
Quale veleno ha ucciso la cavia?

In un laboratorio ci sono 20 provette con tre veleni diversi: 5 contengono il veleno **A**, 3 il **B** e le restanti 12 il **C**. Iniettati a una cavia, i veleni sono letali nel 20% dei casi (A), nel 30% (B) e nel 60% (C). Le provette vengono mescolate per errore: si inietta un veleno a caso a una cavia, che **muore**.

Calcola la probabilità che il veleno iniettato fosse l'A.

SOLUZIONE

Le probabilità a priori sono le frequenze delle provette: $P(A)=5/20=0,25$; $P(B)=3/20=0,15$; $P(C)=12/20=0,60$. La morte è l'effetto osservato: applichiamo Bayes.



Probabilità totale di morte (somma dei tre rami letali):

$$P(M) = 0.25 \cdot 0.20 + 0.15 \cdot 0.30 + 0.60 \cdot 0.60 = 0.05 + 0.045 + 0.36 = 0.455$$

$$P(A | M) = \frac{P(M | A)P(A)}{P(M)} = \frac{0.25 \cdot 0.20}{0.455} = \frac{0.05}{0.455} \approx 0.110$$

RISULTATO

$$P(A | M) \approx 0.110$$

ESERCIZIO 9 ★★★

Sensibilità, specificità e valori predittivi

Un test per il diabete viene applicato a un gruppo di 10 000 persone, usando come soglia una glicemia a digiuno di 130 mg/dl. La prevalenza del diabete è dell'11%.

Calcola sensibilità (S), specificità (E), valore predittivo positivo (VPP) e valore predittivo negativo (VPN).

SOLUZIONE

! Dati insufficienti nella traccia. Sensibilità e specificità sono proprietà intrinseche del test e si ricavano solo dalla tabella 2x2 dei risultati (veri/falsi positivi e negativi), che la traccia originale non riporta. Con prevalenza e numerosità si fissano i totali di colonna, ma non le singole caselle: il problema, così com'è, non è risolvibile numericamente.

	Malato (E)	Sano (E ^c)	Totale
Test +	VP	FP	—
Test -	FN	VN	—
Totale	1 100	8 900	10 000

Struttura della tabella 2x2. Con $N = 10\,000$ e prevalenza 11% i totali di colonna sono 1 100 malati e 8 900 sani, ma le quattro caselle (VP, FP, FN, VN) NON sono fornite dalla traccia.

Le formule da applicare una volta nota la tabella sono:

$$S = \frac{VP}{VP + FN}, \quad E = \frac{VN}{VN + FP}$$

$$VPP = \frac{VP}{VP + FP}, \quad VPN = \frac{VN}{VN + FN}$$

Mandami la tabella 2x2 dell'esercizio originale (o i valori di VP, FP, FN, VN) e completo il calcolo con i numeri.