

Modelli analitici descrittivi

Lezione 4: calcolo delle probabilità

1 Esercizio 4-1:

Durante l'ultima sessione di appelli, il 20% degli studenti non ha passato l'esame di algebra, il 30% quello di statistica e il 10% non ha superato né l'uno né l'altro. Estratto uno studente a caso:

- che probabilità c'è che sia stato bocciato in statistica sapendo che lo è stato in algebra?
- che probabilità c'è che sia stato bocciato in algebra sapendo che lo è stato in statistica?
- che probabilità c'è che sia stato bocciato in almeno uno dei due esami?

Soluzione Consideriamo gli eventi A = “bocciatura in algebra” e S = “bocciatura in statistica”: sappiamo che $P[A] = 0.2$, $P[S] = 0.3$ e $P[A \cap S] = 0.1$.

- La probabilità che uno studente sia stato bocciato in statistica sapendo che lo è stato in algebra è

$$P[S|A] = \frac{P[S \cap A]}{P[A]} = \frac{0.1}{0.2} = 50\%$$

- La probabilità che uno studente sia stato bocciato in algebra sapendo che lo è stato in statistica è

$$P[A|S] = \frac{P[A \cap S]}{P[S]} = \frac{0.1}{0.3} = 33.\bar{3}\%$$

- La probabilità che uno studente sia stato bocciato in almeno uno dei due esami è

$$P[A \cup S] = P[A] + P[S] - P[A \cap S] = 0.2 + 0.3 - 0.1 = 40\%$$

2 Esercizio 4-2:

Un turista vuol sapere se farà tempo bello (evento A) o brutto (evento \bar{A}) per decidere come vestirsi durante una gita. Consulta il barometro, che può dare i seguenti responsi:

- sereno (evento S)
- variabile (evento V)
- pioggia (evento P)

i quali coprono tutti i casi possibili e sono tra loro incompatibili. Dall'esperienza storica, si sa che le probabilità assolute dei due tipi di tempo in questa stagione sono $p(A) = 0.40$ e $p(\bar{A}) = 0.60$, mentre le probabilità condizionate dei tre responsi rispetto al tempo sono:

E	$p(E A)$	E	$p(E \bar{A})$
S	0.60	S	0.20
V	0.25	V	0.30
P	0.15	P	0.50

Che cosa conviene fare al turista?

Soluzione Per la soluzione algebrica, vedi i lucidi della lezione 4. Per la soluzione con Excel, vedi il file `Es4-2.xls`.

3 Esercizio 4-3:

Si vuole costruire un filtro anti-spam. È noto da indagini precedenti che il 30% dei messaggi di posta elettronica sono di spam. Fra i messaggi di spam, un'alta percentuale (l'80%) contiene parole appartenenti a una data lista, mentre solo il 7% dei messaggi normali contiene tali parole. Si calcoli la probabilità che un messaggio appena ricevuto e contenente tali parole sia un messaggio di spam.

Soluzione Il punto fondamentale è impostare bene il problema. Può essere di aiuto a questo scopo distinguere fra eventi “causa” (C_i) ed eventi “effetto” (E_j). In realtà la distinzione è spesso arbitraria, e comunque il teorema di Bayes non ne ha affatto bisogno, ma l'utilità sta nel riordinare le idee. In questo caso, ipotizziamo che lo spam sia la causa e la presenza di parole chiave l'effetto, e che siano legate in modo statisticamente significativo, ma non deterministico. Infatti, ci sono messaggi di spam senza parole chiave (cioè si presenta la causa senza l'effetto) e messaggi non di spam con le parole chiave (cioè si presenta l'effetto senza la causa).

Il file `Es4-3.xls` riporta il procedimento di soluzione del problema con Excel. Ovviamente, i calcoli si potrebbero fare anche a mano, specialmente per un problema così piccolo. Il file, come tutti i successivi, adotta le seguenti convenzioni (assolutamente arbitrarie):

1. le celle che contengono i dati sono evidenziate in giallo, mentre le altre celle contengono formule, dunque risultati intermedi o finali;
2. le celle che contengono i risultati finali sono evidenziate in azzurro;
3. le cause sono associate alle righe, mentre gli effetti alle colonne.

Sia le cause sia gli effetti devono formare una partizione dello spazio campionario, cioè classificare tutti i casi possibili senza sovrapposizioni. Nel caso in questione, abbiamo solo due possibili cause, complementari:

- il messaggio è di spam (C),
- il messaggio non è di spam (c),

e due possibili effetti, complementari:

- il messaggio contiene parole chiave (E),
- il messaggio non contiene parole chiave (\bar{E}).

I dati sono

- la probabilità assoluta della prima causa, $P[C]$;
- le probabilità condizionate del primo effetto rispetto alle due cause, $P[E|C]$ e $P[E|\bar{C}]$;

Il risultato richiesto è la probabilità condizionata della prima causa rispetto al primo effetto, $P[C|E]$.

Bisogna quindi passare dalla tabella delle probabilità condizionate degli effetti rispetto alle cause a quella delle probabilità condizionate delle cause rispetto agli effetti, passando per quella delle probabilità congiunte. Si veda il file `Ex4-3.xls` per i vari passaggi: si tratta sempre di sottrazioni dall'unità (per calcolare probabilità di eventi complementari), prodotti (per le probabilità congiunte), somme (per le probabilità totali di eventi incompatibili) e rapporti (per le probabilità condizionate).

4 Esercizio 4-4:

Un'azienda produce semiconduttori con un procedimento che è esposto a molte possibili cause di errore. Effettua quindi un test su ogni pezzo prodotto. Il 20% dei pezzi ha un difetto. Se il pezzo è integro, supera il test con probabilità 90%; se è difettoso, fallisce il test con probabilità 70%. Sapendo che un semiconduttore ha superato il test, qual è la probabilità che sia integro?

Soluzione Per l'impostazione e le convenzioni adottate nel file `Es4-4.xls`, vedi la soluzione dell'esercizio precedente. Come prima, ci sono due cause: pezzo difettoso (C) e pezzo integro (\bar{C}), e due effetti, test fallito (E) e test superato (\bar{E}). I dati sono la probabilità assoluta della prima causa e le probabilità condizionate del primo effetto rispetto alla prima causa e del secondo effetto rispetto alla seconda causa. I dati sono quindi leggermente diversi rispetto all'esercizio precedente. Il risultato richiesto è la probabilità condizionata della seconda causa rispetto al secondo effetto.

5 Esercizio 4-5:

Un gruppo di 72 consumatori abituali di cocaina viene diviso in tre sottogruppi uguali da 24 pazienti, che vengono sottoposti a diversi trattamenti per la disintossicazione: desipramina (un antidepressivo, di cui si vuole valutare l'efficacia come farmaco per la disintossicazione), litio (che è il trattamento standard per la dipendenza da cocaina) e un placebo. I risultati dello studio, misurati dal numero di pazienti ricaduti nella dipendenza, sono riportati nella seguente tabella.

	Ricaduta	
	sì	no
desipramina	10	14
litio	18	6
placebo	20	4

Si valuti l'efficacia del nuovo trattamento.

Soluzione Il file `Es4-5.xls` riporta la risoluzione secondo lo schema e le convenzioni descritti negli esercizi precedenti. Questa volta ci sono tre cause, che corrispondono ai tre trattamenti, e due effetti: ricaduta o no. I dati sono le frequenze assolute di tutti e sei gli eventi intersezione. Si possono trasformare in probabilità congiunte dividendole per 72, cioè per il numero totale di casi. I risultati utili per

rispondere alla domanda sono le probabilità condizionate delle cause rispetto agli effetti (in particolare, la probabilità di aver subito il nuovo trattamento per i pazienti ricaduti) e degli effetti rispetto alle cause (in particolare, la probabilità di ricaduta per ciascun trattamento).

6 Esercizio 4-6:

Dei componenti metallici possono essere prodotti assoggettandoli a una temperatura semplice o a due temperature consecutive. Nel secondo caso, il componente risulta più resistente allo sforzo. Si eseguono su un campione di 200 componenti, alcuni temprati una volta e altri due, delle prove sotto sforzo, contando i componenti che escono danneggiati dalle prove e quelli che rimangono intatti. La tabella seguente riporta i risultati delle prove.

Temperature	Componenti	
	rotti	intatti
una	40	45
due	15	100

Si valutino le probabilità assolute, congiunte e condizionate, per dedurne la probabilità che un componente danneggiato sia stato temprato una volta sola o due.

Soluzione Il file `Es4-6.xls` riporta la risoluzione secondo lo schema e le convenzioni descritti negli esercizi precedenti. La struttura di questo caso è molto simile a quella dell'esercizio precedente, salvo che le cause sono due anziché tre.

Nota Conoscendo i costi delle due fasi di temperatura e il costo della sostituzione di un componente danneggiato, se ne potrebbe dedurre se sia preferibile temprare i componenti una volta e sostituirli più spesso o due volte e sostituirli più raramente.