Probabilità e Test Diagnostici

Riccardo Ievoli

Dipartimento di Scienze dell'Ambiente e della Prevenzione

CdL in Biotecnologie

Per il presente materiale si ringrazia la collega Prof.ssa L. Grisotto

Probabilità

Definizioni di probabilità

Definizione classica

Se un esperimento può dar luogo a n esiti che si escludono a vicenda e che sono ugualmente possibili, e se m di questi esiti hanno la caratteristica A, la probabilità di A è data dal rapporto $\frac{m}{n}$: $P(A) = \frac{m}{n}$

Esempio. Probabilità di estrarre una figura da un mazzo di 40 carte:12/40=3/10

Definizione frequentista

Se si ripete un eseprimento un gran numero di volte n e se un certo evanto con caratteristica A si verifica m volte; la frquenza relativa $\frac{m}{n}$ sarà approssimativamente uguale all probabilita' di A: $P(A) = \frac{m}{n}$

Esempio. Probabilità che una moneta bilanciata dia testa: 1/2

Definizioni di probabilità

Definizione soggettiva:

La probabilità dell'evento *A* misura il grado di fiducia che un individuo ripone nel verificarsi di determinati eventi, in base alle proprie conoscenze

Esempio. Probabilità che a Pasquetta piova = 0,5

Esempi

Mazzo di carte

P(pescare un asso) = 4/40 = 0.1 -> 10%

Lancio di un dado

P(ottenere 6) = 1/6 = 0.16 -> 16%

Esame di statistica

P(passare) = 14/16 = 0.92 -> 92%

Eventi complementari

Due eventi A ed $\bar{A}$ si dicono *complementari* quando il verificarsi dell'uno esclude il verificarsi del secondo ma uno dei due si verificherà di sicuro

Ad esempio nel lancio di un dado:

Evento A, esce il 6 -> evento $\bar{A}$, non esce il 6

$$P(\bar{A}) = 1 - P(A)$$

Esempio del lancio di un dato

$$P(A) = \frac{1}{6} \longrightarrow P(\bar{A}) = 1 - \frac{1}{6} = \frac{5}{6}$$

Probabilità dell’unione di due eventi

Se A e B sono eventi *incompatibili* (non possono verificarsi contemporaneamente):

$$P(A\ o\ B) = \mathrm{P}(A) + P(B)$$

Se A e B non sono incompatibili:

$$P(A\ o\ B) = \mathrm{P}(A) + P(B) - \mathrm{P}(\mathrm{A\ e\ B})$$

Probabilità dell'unione di due eventi

Esempio eventi incompatibili

Nel lancio del dado:

Evento A = esce il 2 $\quad P(A) = \frac{1}{6}$

Evento B = esce il 3 o il 4 $\quad P(B) = \frac{2}{6}$

$$P(A \text{ o } B) = \mathrm{P}(A) + P(B) = \frac{1}{6} + \frac{2}{6} = \frac{3}{6} = \frac{1}{2}$$

Probabilità dell'unione di due eventi

Esempio eventi non **incompatibili**

	Polvere	Altri allergeni
Pomodoro	2	3
Altri allergeni	15	80

Qual è la probabilità di essere allergico al Pomodoro o alla polvere?

$$P(A\ o\ B) = \mathrm{P}(A) + P(B) - \mathrm{P(A\ e\ B)}$$

P(pomodoro) + P(polvere) - P(pomodoro e polvere) =

5/100 + 17/100 - 2/100 = 20/100

P(essere allergico al pomodoro o alla polvere)=20/100

Probabilità dell'intersezione di due eventi

Se A e B sono *indipendenti* (il realizzarsi dell'uno non influenza la probabilità che si realizzi l'altro):

$$P(A \; e \; B) = P(A) \times P(B)$$

Se A e B non sono indipendenti:

$$P(A \; e \; B) = P(A) \times P(B|A) = P(B) \times P(A|B)$$

dove $P(B|A)$ è la *probabilità condizionata* di B dato A e

$P(A|B)$ è la probabilità condizionata di A dato B

Probabilità dell'intersezione di due eventi

Esempio

Supponiamo di estrarre due carte da gioco da un mazzo di carte francesi, avendo cura di reinserire nel mazzo la prima carta dopo l'estrazione; consideriamo i due seguenti eventi:

A: esce una carta di cuori

B: esce una figura

Calcoliamo la probabilità composta dei due eventi

$$P(A\ e\ B) = P(A) \times P(B) = \frac{13}{52} \times \frac{12}{52} = \frac{1}{4} \times \frac{3}{13} = \frac{3}{52}$$

Probabilità dell'intersezione di due eventi

Esempio

	Pressione alta	Pressione bassa	Totale
Beve vino rosso	7	3	10
Non beve vino rosso	25	65	90
Totale	32	68	100

P(pressione alta e bevitore) = ?

$$P(A\ e\ B) = P(A) \times P(B|A) = P(B) \times P(A|B)$$

P(bevitore) x P(pressione alta|bevitore)

(10/100) x (7/10) = 7/100

Teorema di Bayes

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

Dove $P(A)$ è la probabilità a priori di A

$P(A|B)$ è la probabilità a posteriori di A

$\frac{P(B|A)}{P(B)}$ è la verosimiglianza di B dato A

$P(B) = P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A})$ per cui il teorema di Bayes può essere scritto come:

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A})}$$

Test diagnostici

Test diagnostici

Per fare una diagnosi, il medico o il professionista sanitario sottopone i pazienti a test diagnostici.

La maggior parte dei test clinici (segni, sintomi, test di laboratorio, radiologia, ecc.) è «IMPRECISA» (soggetta a incertezza)

Per interpretarli è necessario conoscere la VALIDITÀ, vale a dire la capacità di classificare gli individui tenendo conto di quale sia la verità.

Test Diagnostici

- Nel campo delle scienze sanitarie le leggi e i concetti di probabilità vengono ampiamente usati per la valutazione dei test diagnostici.

- In ambito sanitario vengono comunemente utilizzati test diagnostici per determinare l’eventuale presenza di una malattia in un soggetto.

A cosa siamo interessati?

- Valutazione del rendimento di un test diagnostico

- L’uso di un test diagnostico per valutare la probabilità di malattia

A cosa siamo interessati?

- Valutazione del rendimento di un test diagnostico

TEST DIAGNOSTICI

Per valutare i test diagnostici dobbiamo quindi classificare gli individui rispetto alla presenza o meno della malattia e al risultato positivo o negativo ad un test diagnostico.

Possiamo sintetizzare i nostri dati in nella seguente tabella:

	Malattia		
Risultato del test	**Presente (M)**	**Assente (S)**	**Totale**
Positivo (+)	a	b	a+b
Negativo (-)	c	d	c+d
Totale	a+c	b+d	n

TEST DIAGNOSTICI

Per valutare i test diagnostici dobbiamo quindi classificare gli individui rispetto alla presenza o meno della malattia e al risultato positivo o negativo ad un test diagnostico.

Possiamo sintetizzare i nostri dati in nella seguente tabella:

	Malattia		
Risultato del test	**Presente (M)**	**Assente (S)**	**Totale**
Positivo (+)	a	b	a+b
Negativo (-)	c	d	c+d
Totale	a+c	b+d	n

Come misurare l'efficacia di un test?

Esempio

		Malati	Sani	Totale
Test 1	Positivo	40	50	90
	Negativo	10	900	910
Test 2	Positivo	0	0	0
	Negativo	50	950	1000
Test 3	Positivo	20	0	20
	Negativo	30	950	980
Totale		50	950	1000

- Calcoliamo la proporzione di diagnosi corrette?

Esempio

		Malati	Sani	Totale
Test 1	Positivo	40	50	90
	Negativo	10	900	910
Test 2	Positivo	0	0	0
	Negativo	50	950	1000
Test 3	Positivo	20	0	20
	Negativo	30	950	980
Totale		50	950	1000

- Test 1: 94%
- Test 2: 95%
- Test 3: 97%

Allora.. come misuro l'efficacia?

- Non c'è un indice solo che in modo semplice metta nella condizione di valutare l'efficacia del test!
- Perché ci sono due fattori da misurare:
  - La capacità del test di individuare correttamente la presenza della malattia quando il paziente è malato;
  - La capacità del test di segnalarne l'assenza quando il paziente è realmente sano!

Valutazione del test diagnostico

Le **probabilità condizionate** sono particolarmente utili quando si tratta di descrivere il rendimento di un test diagnostico.

Sensibilità e Specificità

	Malattia		
Risultato del test	**Presente (M)**	**Assente (S)**	**Totale**
Positivo (+)	a	b	a+b
Negativo (-)	c	d	c+d
Totale	a+c	b+d	n

Sensibilità**:** probabilità che il test risulti positivo dato che il soggetto è malato (M)

$$P(+|M) = \frac{a}{a+c}$$

Sensibilità e Specificità

Risultato del test	**Malattia** Presente (M)	Assente (S)	**Totale**
Positivo (+)	a	b	a+b
Negativo (-)	c	d	c+d
Totale	a+c	b+d	n

Specificità**:** probabilità che il test risulti negativo dato che il soggetto è sano (S)

$$P(-|S) = \frac{d}{b+d}$$

Sensibilità e Specificità

Risultato del test	**Malattia** Presente (M)	Assente (S)	Totale
Positivo (+)	a	b	a+b
Negativo (-)	c	d	c+d
Totale	a+c	b+d	n

Probabilità di falsi positivi**:** il test risulta positivo su un soggetto sano

P(falso positivo) = P(+|S) = 1-P(-|S) = 1-specificità

$$P(+|S) = \frac{b}{b+d}$$

Sensibilità e Specificità

Risultato del test	**Malattia** **Presente (M)**	 **Assente (S)**	 **Totale**
Positivo (+)	a	b	a+b
Negativo (-)	c	d	c+d
Totale	a+c	b+d	n

***Probabilità di falsi negativi*:** il test è negativo su un soggetto ammalato

P(falso negativo) = P(-|M) = 1-P(+|M) = 1-sensibilità

$$P(-|M) = \frac{c}{a+c}$$

Esempio

		Malati	Sani	Totale
Test 1	Positivo	40	50	90
	Negativo	10	900	910
Test 2	Positivo	0	0	0
	Negativo	50	950	1000
Test 3	Positivo	20	0	20
	Negativo	30	950	980
Totale		50	950	1000

- Test 1:
  - $P(+|M)=80\%$
  - $P(-|S)=95\%$
- Test 2:
  - $P(+|M)=0\%$
  - $P(-|S)=100\%$
- Test 3:
  - $P(+|M)=40\%$
  - $P(-|S)=100\%$

- Quando il test ha un’eccellente SENSIBILITA’, un risultato negativo del test permette di escludere la diagnosi di malattia, perché la probabilità di un risultato falsamente negativo è bassa!
- Il risultato positivo di un test molto SPECIFICO permette di confermare la diagnosi, perché la probabilità di un risultato falsamente positivo è bassa!

La scelta del test migliore non è sempre facile!!!

- Dobbiamo tener conto di molteplici fattori…
- Per farlo a volte è necessario dare un giudizio sull'importanza relativa di sensibilità se specificità nei vari casi.
- A volte bisogna valutare costi e benefici: Anche se un test ha performance diagnostiche superiori, potrebbe non essere preferito se il costo o il tempo richiesto per la somministrazione è troppo elevato rispetto ai benefici clinici marginali.

Se a decidere è la statistica…

- **Indice di Youden (J)**

$$J = sens + spec - 1$$

- L'indice varia da -1 a 1, dove:
  - **1** indica un test perfetto, con una sensibilità e specificità pari a 1.
  - **0** indica che il test non fornisce alcuna informazione (equivalente a una scelta casuale).
  - Un valore negativo indica che il test è peggiore della scelta casuale.

Esempio

		Malati	Sani	Totale
Test 1	Positivo	40	50	90
	Negativo	10	900	910
Test 2	Positivo	0	0	0
	Negativo	50	950	1000
Test 3	Positivo	20	0	20
	Negativo	30	950	980
Totale		50	950	1000

- Test 1:
  - P(+|M)=80%
  - P(-|S)=95%

  J=0.8+0.95-1=0.75
- Test 2:
  - P(+|M)=0%
  - P(-|S)=100%

  J=0+1-1=0
- Test 3:
  - P(+|M)=40%
  - P(-|S)=100%

  J=0.40+1-1=0.40

Sensibilità e Specificità
Esempio

	Alzheimer		
Risultato del test	**Presente (M)**	**Assente (S)**	**Totale**
Positivo (+)	436	5	441
Negativo (-)	14	495	509
Totale	450	500	950

Sensibilità = P(+|M) = 436/(436+14) = 0,97

Specificità = P(-|S) = 495/(495+5) = 0,99

Falsi positivi = 5

P(falso positivo) =1 - specificita'= 1- 0,99 = 0,01

Falsi negativi = 14

P(falso negativo) = 1 - sensibilita' = 1- 0,97 = 0,03

J=0.97+0.99-1=0.96

Sensibilità e Specificità

La ***Sensibilità*** e la ***Specificità*** misurano la validita' di un test diagnostico

I valori di specificità e sensibilità possono essere ottenuti attraverso esperimenti ad hoc.

A cosa siamo interessati?

- L’uso di un test diagnostico per valutare la probabilità di malattia

- Il paziente è risultato positivo al test,

  Qual è la probabilità che sia malato?

Aggiornamento della probabilità

Ok, ma se risulto positivo al test, sono malato?

cosa pensiamo prima ⟶ probabilità pre-test

↓

\+ informazione del test

↓

cosa pensiamo dopo ⟶ probabilità post-test

probabilità post-test = probabilità pre-test & informazione del test

⇩

TEOREMA DI BAYES!

Teorema di Bayes

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

Dove $P(A)$ è la probabilità a priori di A

$P(A|B)$ è la probabilità a posteriori di A

$\frac{P(B|A)}{P(B)}$ è la verosimiglianza di B dato A

$P(B) = P(B|A) \times P(A) + P(B|\bar{A}) \times P(\bar{A})$ per cui il teorema di Bayes può essere scritto come:

$$P(A|B) = \frac{P(A) \times P(B|A)}{P(B|A) \times P(A) + P(B|\bar{A}) \times P(\overline{A})}$$

Aggiornamento della probabilità

Ok, ma se risulto positivo al test, sono malato? $P(M|+)$?

cosa pensiamo prima $\longrightarrow$ probabilità pre-test = $P(M)$

+ informazione del test = $\frac{P(+|M)}{P(+)}$

cosa pensiamo dopo $\longrightarrow$ probabilità post-test = $P(M|+)$

probabilità post-test = probabilità pre-test & informazione del test

$$P(M|+) = P(M) \times \frac{P(+|M)}{P(+)}$$

Valore predittivo di un test diagnostico

Valore predittivo positivo (VPP)

$$P(M|+) = P(M) \times \frac{P(+|M)}{P(+)} = P(M) \times \frac{P(+|M)}{P(+|M) \times P(M) + P(+|S) \times P(S)}$$

Valore predittivo negativo (VPN)

$$P(S|-) = P(S) \times \frac{P(-|S)}{P(-)} = P(S) \times \frac{P(-|S)}{P(-|S) \times P(S) + P(-|M) \times P(M)}$$

P(M) = prevalenza della condizione o malattia

P(S) = 1 - P(M)

Valore predittivo di un test diagnostico
Esempio

	Alzheimer		
Risultato del test	**Presente (M)**	**Assente (S)**	**Totale**
Positivo (+)	436	5	441
Negativo (-)	14	495	509
Totale	450	500	950

Sensibilità = P(+|M) = 436/(436+14) = 0,97

Specificità = P(-|S) = 495/(495+5) = 0,99

P(M) = 450/950 = 0,47

P(+) = 441/950 = 0,46

Valore predittivo positivo $P(M|+) = P(M) \times \frac{P(+|M)}{P(+)} = 0{,}47 \times \frac{0{,}97}{0{,}46} = 0{,}99$

$P(M|+) = \frac{a}{a+b} = \frac{436}{441} = 0{,}99$

Valore predittivo di un test diagnostico
Esempio

	Alzheimer		
Risultato del test	**Presente (M)**	**Assente (S)**	**Totale**
Positivo (+)	436	5	441
Negativo (-)	14	495	509
Totale	450	500	950

Sensibilità = P(+|M) = 436/(436+14) = 0,97

Specificità = P(-|S) = 495/(495+5) = 0,99

P(S) = 500/950 = 0,53

P(-) = 509/950 = 0,54

Valore predittivo negativo $P(S|-) = P(S) \times \frac{P(-|S)}{P(-)} = 0{,}53 \times \frac{0{,}99}{0{,}54} = 0{,}97$

$P(S|-) = \frac{d}{c+d} = \frac{495}{509} = 0{,}97$

Valore predittivo di un test diagnostico

La prevalenza di Alzheimer nella popolazione di interesse (eta' maggiore di 65 anni) è del 4,4%

Un test diagnostico ha una sensibilità del 97% e una specificità dell'99%

Qual è la probabilità che un soggetto sia malato se è risultato positivo al test?

Dati:

P(M) = 0,044; P(S) = 0,956; Sensibilita: P(+|M) = 0,97;

Specificita' P(-|S) = 0,99; P(+|S) = 0,01

$$P(M|+) = P(M) \times \frac{P(+|M)}{P(+|M) \times P(M) + P(+|S) \times P(S)}$$

$$P(M|+) = 0{,}044 \times \frac{0{,}97}{0{,}97 \times 0{,}044 + 0{,}01 \times 0{,}956} = 0{,}82$$

Valore predittivo di un test diagnostico

La prevalenza di Alzheimer nella popolazione di interesse (eta’ maggiore di 65 anni) è del 4,4%

Un test diagnostico ha una sensibilità del 97% e una specificità dell’99%

Qual è la probabilità che un soggetto **non** sia malato se è risultato negativo al test?

Dati:

P(M) = 0,044; P(S) = 0,956; Sensibilita: P(+|M) = 0,97; P(-|M) = 0,03

Specificita’ P(-|S) = 0,99;

$$P(S|-) = P(S) \times \frac{P(-|S)}{P(-|S) \times P(S) + P(-|M) \times P(M)}$$

$$P(S|-) = 0{,}956 \times \frac{0{,}99}{0{,}99 \times 0{,}956 + 0{,}03 \times 0{,}044} = 0{,}99$$

La decisione diagnostica

La decisione diagnostica (VPP, VPN) non dipende solo dall'efficacia del test (sensibilità e specificità), ma anche dalla probabilità a priori.

Maggiore è la probabilità a priori, maggiore è il valore predittivo del test (a parità di sensibilità e specificità).