

Sul rischio relativo

Alessandra Nardi

alenardi@mat.uniroma2.it

5 giugno 2017

Il concetto di rischio è strettamente legato al tempo
Per darne una definizione formale iniziamo col chiarire cosa intendiamo per tempo di sopravvivenza specificando

- un'origine in cui inizia l'osservazione del paziente
- l'evento di interesse
- una unità di misura

Definiamo poi come

T : tempo di sopravvivenza

l'intervallo tra l'origine e il momento aleatorio in cui si verifica l'evento

Ne deriva che T è una variabile aleatoria a valori positivi.

T è caratterizzata da alcune funzioni del tempo
La densità di probabilità

$$f(t) = \lim_{\Delta \rightarrow 0^+} \frac{Pr \{t \leq T < t + \Delta\}}{\Delta}$$

da cui

$$Pr \{t \leq T < t + \Delta\} \simeq f(t)dt$$

La funzione di rischio

$$\lambda(t) = \lim_{\Delta \rightarrow 0} \frac{Pr \{t \leq T < t + \Delta | T \geq t\}}{\Delta}$$

$$Pr \{t \leq T < t + \Delta | T \geq t\} \simeq \lambda(t)dt$$

La funzione di sopravvivenza

$$S(t) = Pr \{T \geq t\}$$

da cui

$$\lambda(t)dt = \frac{f(t)dt}{S(t)}$$

o in altri termini

$$Pr \{t \leq T < t + \Delta | T \geq t\} = \frac{Pr \{t \leq T < t + \Delta\}}{Pr \{T \geq t\}}$$

Questo implica che probabilità e rischio sono simili solo se $S(t) \simeq 1$

Apriamo una piccola parentesi....cos'è una densità di probabilità ?
Se i valori osservabili come risultato del nostro esperimento sono nel discreto (pochi) potremo assegnare ad ognuno una probabilità di essere osservato, facendo attenzione a garantire che la loro somma sia 1.

Se il risultato del nostro esperimento o della nostra osservazione è una variabile continua
non è più possibile assegnare una probabilità ad ogni singolo valore osservabile (sono davvero troppi!)

Passiamo a considerare invece che singoli valori degli intervalli (per quanto piccoli ognuno di essi conterrà sempre un numero infinito di punti) e definiamo

$$P(x \leq X < x + dx) = \int_x^{x+dx} f(t)dt \approx f(x)dx$$

La funzione $f(x)$ prende il nome di densità e distribuisce la massa di probabilità unitaria sull'insieme dei valori osservabili.

La probabilità di osservare valori compresi in un dato intervallo è l'area corrispondente sotto questa curva

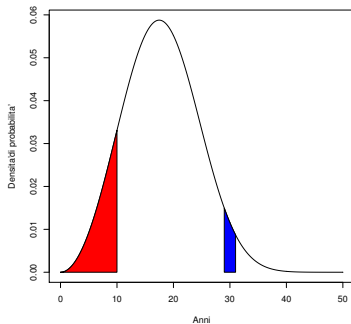


Figura: Densità di probabilità

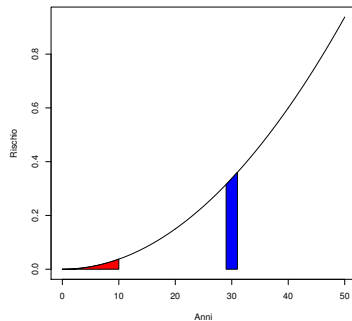


Figura: Funzione di rischio

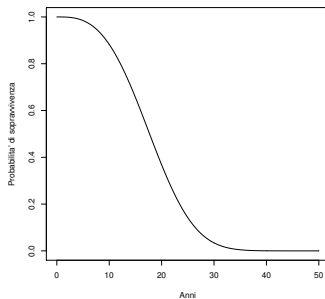


Figura: Funzione di sopravvivenza

Compreso che densità e rischio sono funzioni del tempo diventa chiaro che quando ignoriamo t stiamo implicitamente assumendo che il rischio sia costante almeno nell'intervallo di tempo che stiamo considerando.

Questa assunzione, di fatto molto restrittiva e spesso irrealistica se l'intervallo non è breve, è di fatto equivalente a quella di ipotizzare per T un modello esponenziale caratterizzato proprio da

$$\lambda(t) = c$$

dove c è una costante positiva. Vediamo alcuni esempi di distribuzioni esponenziali

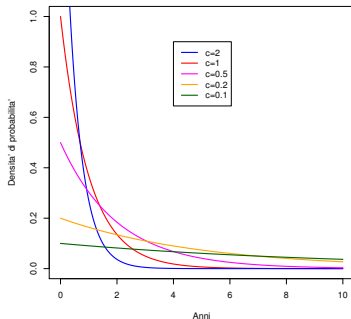


Figura: Modello esponenziale:
densità di probabilità

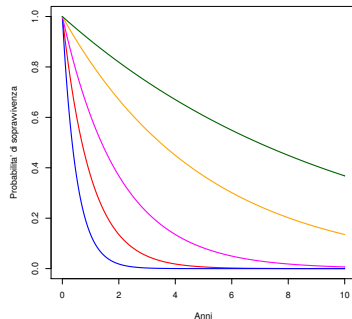


Figura: Modello esponenziale:
funzioni di sopravvivenza

La stima della densità come quella della funzione di rischio sono molto complesse e spesso associate ad una bassa precisione. Ne segue che si preferisce in genere stimare la funzione di sopravvivenza, probabilità valutate su intervalli di tempo oppure misure relative, prima fra tutte il rischio relativo. Proviamo a discuterne sulla base di uno studio prospettico randomizzato.

"The Physicians' Health Study" è uno studio clinico randomizzato condotto allo scopo di valutare il possibile effetto di riduzione della mortalità cardiovascolare legato ad un uso regolare e continuato di aspirina

Ciascun medico che partecipò allo studio prese a giorni alterni una pasticca di aspirina o un semplice placebo senza essere a conoscenza di quale sostanza stesse realmente assumendo

Riportiamo nella seguente tabella i risultati relativi ad un rapporto preliminare (N.Engl.J.Med., 1988)

	Attacco cardiaco fatale	Attacco cardiaco non fatale	Nessun attacco cardiaco	Totale
Placebo	18	171	10845	11034
Aspirina	5	99	10933	11037
Totale	23	270	21778	22071

Nel nostro studio siamo particolarmente interessati a leggere le differenze tra il gruppo di medici che hanno assunto aspirina ed il gruppo di controllo a cui è stato somministrato un semplice placebo. Calcoliamo allora le frequenze relative separatamente per i due gruppi, cioè le distribuzioni di frequenze relative dell'esito, condizionatamente al tipo di trattamento. I totali di riferimento sono quelli marginali corrispondenti alla numerosità totale del gruppo dei controlli e dei "trattati".

	Attacco cardiaco fatale	Attacco cardiaco non fatale	Nessun attacco cardiaco	Totale
Placebo	18	171	10845	11034
Aspirina	5	99	10933	11037
Totale	23	270	21778	22071

	Attacco cardiaco fatale	Attacco cardiaco non fatale	Nessun attacco cardiaco	Totale
Placebo	0.0016	0.0155	0.9829	1
Aspirina	0.0005	0.0090	0.9905	1
Totale	0.0010	0.0122	0.9868	1

Proviamo a rispondere ad alcuni quesiti

- Cosa stimano le frequenze relative che abbiamo calcolato?
- Esiste una qualche differenza tra i due trattamenti? Come la misuriamo?
- Cosa siamo realmente in grado di stimare sulla base dei dati osservati e cosa no?

Focalizziamo la nostra attenzione sull'evento cardiaco senza distinguere se fatale o meno .

	Attacco cardiaco	Nessun attacco cardiaco	Totale
Placebo	189	10845	11034
Aspirina	104	10993	11037
Totale	293	21778	22071

Per valutare l'entità dell'effetto dell'aspirina come trattamento di prevenzione indichiamo con $\pi_{AC|A}$ e $\pi_{AC|P}$ le probabilità di avere un attacco cardiaco fatale se sottoposti a terapia preventiva rispettivamente a base di aspirina e beta-carotene. Queste probabilità si riferiscono all'intera popolazione (o se volete ad un paziente estratto casualmente) e all'arco temporale dello studio, 8 anni dall'inizio del trattamento.

Le stime che otteniamo sono rispettivamente $104/11037 = 0.0094$ e $189/11037 = 0.0171$.

Come confrontiamo le due probabilità stimate?
Tre sono le misure possibili

- La differenza

$$\pi_{AC|P} - \pi_{AC|A} \quad 0.0171 - 0.0094 = 0.0077$$

- Il rapporto

$$\frac{\pi_{AC|P}}{\pi_{AC|A}} \quad \frac{0.0171}{0.0094} = 1.82$$

- L'odds ratio

$$\frac{\pi_{AC|P}/(1-\pi_{AC|P})}{\pi_{AC|A}/(1-\pi_{AC|A})} \quad \frac{0.0171/0.9829}{0.0094/0.9906} = 1.83$$

Il rapporto $\frac{\pi_{AC|P}}{\pi_{AC|A}}$ può essere considerato un rischio relativo?
Ricordiamo che

$$\lambda(t)dt = \frac{f(t)dt}{S(t)}$$

Ne deriva che probabilità e rischio sono simili se l'evento è raro:
 $S(t) \simeq 1$ e quindi $\lambda(t)dt \simeq f(t)dt$,

Attenzione tuttavia al fatto che le nostre probabilità fanno riferimento ad un intervallo di 8 anni mentre il rischio è riferito ad un intervallo infinitesimo.

Quindi il rapporto $\frac{\pi_{AC|P}}{\pi_{AC|A}}$ può essere assimilato ad un rischio relativo solo sotto l'ulteriore assunzione di un rischio costante nel tempo,
 $\frac{\lambda_{AC|P} \times 8 \text{anni}}{\lambda_{AC|A} \times 8 \text{anni}}$

Torniamo sul significato di odds per osservare che l'odds di un attacco cardiaco assumendo aspirina è

$$\frac{\pi_{AC|A}}{(1-\pi_{AC|A})} = \frac{104/11037}{10933/11037} = \frac{104}{10933}$$

Mentre la probabilità viene stimata dal rapporto tra chi ha avuto un attacco cardiaco e il totale dei pazienti, la stima dell'odds è basata sul rapporto tra chi ha avuto e chi non ha avuto un attacco cardiaco. Si tratta di un linguaggio comune tra ... scommettitori

Se l'evento è raro, $1 - \pi_{AC|A} \simeq 1$, e il rischio costante, odds ratio e rischio relativo tendono a coincidere.

L'odds ratio diventa essenziale se invece di uno studio prospettico randomizzato analizziamo dati che provengono da uno studio caso-controllo

Doll e Hill nel 1952 dimostrarono per la prima volta una relazione significativa tra fumo e cancro polmonare. I dati si riferiscono ad uno studio retrospettivo caso-controllo condotto in Inghilterra

Numero medio giornaliero di sigarette	Cancro polmonare	Controlli	Totale
Nessuna	7	61	68
< 5	55	129	184
5 – 14	489	570	1059
15 – 24	475	431	906
25 – 49	293	154	447
≥ 50	38	12	50
Totale	1357	1357	2714

Lo studio caso-controllo parte dal reclutamento di un campione di casi, nel nostro caso di pazienti affetti da cancro polmonare. A questo campione viene affiancato un campione di individui *sani* e su entrambi vengono raccolte retrospettivamente le informazioni relative ai fattori rischio in esame, nel nostro caso il fumo. Notate come la logica causa-effetto sia invertita, adesso partiamo conoscendo l'evento (effetto) ed è la causa che lo ha generato ad essere aleatoria. Come interpretare i risultati?

Si tratta di un disegno sperimentale altamente informativo per la disponibilità in tempi brevi di un elevato numero di casi e la possibilità di indagare simultaneamente l'effetto di diversi fattori di rischio.

Ha tuttavia due punti deboli: ci consente di stimare solo misure relative ed è estremamente suscettibile a possibili distorsioni nel reclutamento (selection bias).

Vediamo alcuni punti importanti su cui riflettere

Circa la scelta dei casi:

- dal momento che cerchiamo una relazione causa-effetto è preferibile individuare patologie omogenee per eziologia;
- casi prevalenti o incidenti? La prima soluzione potrebbe portare a selezionare i pazienti sopravvissuti da una serie più ampia di casi incidenti con il rischio di confondere fattori che influenzano lo sviluppo e la prognosi della patologia. L'antigene HLA A2 sembrava inizialmente legato allo sviluppo della leucemia acuta mentre si è poi rilevato un fattore prognostico per la sopravvivenza del paziente (Rogentine et al.1972,1973);
- nei casi prevalenti cause ed affetti della patologia possono sovrapporsi. Il paziente potrebbe cambiare stile di vita a seguito della diagnosi. Necessità di una attenta ricostruzione retrospettiva non sempre facile.

Circa scelta dei controlli

- su base ospedaliera oppure selezioniamo un campione casuale dalla popolazione?
- la scelta dei controlli è strettamente legata alle modalità con cui abbiamo scelto i casi. L'obiettivo primario è garantire la comparabilità dei gruppi che richiede ad esempio omogeneità nei test diagnostici;
- attenzione al rischio di selezionare i casi (o i controlli) in base a caratteristiche legate ai fattori di rischio in esame (maggiore attenzione diagnostica in presenza di sintomi non esclusivamente legati alla patologia in esame);
- bisognerebbe evitare che i controlli presentino patologie la cui eziologia è simile alla malattia in studio o che interessano lo stesso organo (cancro polmonare, broncopatie croniche e fumo);
- l'idea che casi e controlli debbano essere il più possibile simili (a parte la patologia oggetto di studio) che appartiene agli studi prospettici randomizzati, è completamente errata e inapplicabile in uno studio caso-controllo (*overmatching*).

Torniamo allo studio di Doll e Hill

Numero medio giornaliero di sigarette	Cancro polmonare	Controlli	Totale
< 5	62	190	252
≥ 5	1295	1167	2462
<i>Totale</i>	1357	1357	2714

Cosa siamo in grado di stimare sulla base dei dati osservati?

I rapporti $1357/2714 = 0.5$, $252/2714 = 0.093$, $62/252 = 0.25$
 $1295/2462 = 0.53$ hanno un valore inferenziale?

Il campione osservato non è stato estratto casualmente dalla popolazione e non consente la stima delle caratteristiche della popolazione stessa.

Siamo in grado di stimare solo quantità relative

Obiettivo dello studio è stimare il rapporto $\frac{\pi_{CP|F}}{\pi_{CP|NF}}$ ma in questo disegno sperimentale l'elemento aleatorio è l'esposizione o meno al fumo in pazienti di cui è nota la presenza o meno della malattia. In altri termini potremmo stimare il rapporto $\frac{\pi_{F|CP}}{\pi_{F|NCP}}$ diverso dal precedente.

Le proprietà dell'odds ratio ci vengono in aiuto

$$\frac{\frac{\pi_{F|CP}}{(1-\pi_{F|CP})}}{\frac{\pi_{F|NCP}}{(1-\pi_{F|NCP})}} = \frac{\frac{1295/1357}{62/1357}}{\frac{1167/1357}{190/1357}} = \frac{1295 \times 190}{62 \times 1167} = \frac{1295/2462}{1167/2462} = \frac{\frac{\pi_{CP|F}}{(1-\pi_{CP|F})}}{\frac{\pi_{CP|NF}}{(1-\pi_{CP|NF})}}$$

L'elegante simmetria dell'odds ratio implica che numericamente le due quantità coincidono pur mantenendo un significato diverso.

Questo implica che pur stimando $\frac{\frac{\pi_{F|CP}}{(1-\pi_{F|CP})}}{\frac{\pi_{F|NCP}}{(1-\pi_{F|NCP})}}$ noi otteniamo

indirettamente una stima della quantità di interesse $\frac{\frac{\pi_{CP|F}}{(1-\pi_{CP|F})}}{\frac{\pi_{CP|NF}}{(1-\pi_{CP|NF})}}$ pari

a $\frac{246050}{72354} = 3.40$

Quando l'odds ratio può essere considerato una approssimazione del rischio relativo?

Se $\pi_{CP|F} \simeq 1$ e $\pi_{CP|NF} \simeq 1$ cioè se l'evento è raro

$$\frac{\frac{\pi_{CP|F}}{(1-\pi_{CP|F})}}{\frac{\pi_{CP|NF}}{(1-\pi_{CP|NF})}} \simeq \frac{\pi_{CP|F}}{\pi_{CP|NF}}$$

Se inoltre il rischio di sviluppare un tumore può essere considerato costante

$$\frac{\pi_{CP|F}}{\pi_{CP|NF}} \simeq \frac{\lambda_{CP|F} \times \Delta t}{\lambda_{CP|NF} \times \Delta t}$$

Realistica?