STIMA DI MODELLI DI SOPRAVVIVENZA NON PARAMETRICI

- Rilevazione della mortalità in ambito attuariale
- Esposizione attuariale e frequenze di decesso
- Riferimento: anno di vita, anno di polizza e anno di calendario
- Frequenze di decesso per tavole selezionate
- Stima di modelli di sopravvivenza non parametrici – uscite soltanto per morte
- Modello di sopravvivenza con più cause di eliminazione
- Stima di modelli di sopravvivenza non parametrici – uscite per morte e altra causa
- Stimatore della funzione di sopravvivenza e sue proprietà
- Confronto della tavola di sopravvivenza stimata con una tavola standard
- Perequazione mediante tavole standard
- Perequazione con leggi di sopravvivenza
- Altre formule di perequazione utilizzate in ambito attuariale
- Stima dei parametri di una formula di perequazione
- Perequazione mediante modelli lineari generalizzati

RILEVAZIONE DELLA MORTALITÀ IN AMBITO ATTUARIALE

Rilevazioni trasversali (cross-sectional studies)

Si individua il <u>gruppo di studio</u>, cioè un gruppo di individui per i quali interessa studiare la sopravvivenza (tipicamente gli assicurati di una compagnia di assicurazione, gli iscritti ad un fondo pensione, …).

Si fissa un <u>periodo di osservazione</u> durante il quale viene osservato il gruppo di studio; di solito si osserva la collettività per 3-5 anni.

All'inizio dell'osservazione ci saranno individui già presenti, ai quali se ne aggiungeranno altri durante il periodo di osservazione;

alcuni individui possono uscire per causa diversa dal decesso durante l'osservazione, per esempio perché è scaduto il contratto di assicurazione, oppure perché è stata riscattata la polizza;

ci saranno individui ancora in vita al termine dell'osservazione.

Si può considerare come <u>istante iniziale</u> l'età minima di ingresso in assicurazione, oppure un'età minima a partire dalla quale si dispone di osservazioni.

Generalmente si ha a che fare con dati **incompleti**:

se non è osservato l'istante iniziale, l'osservazione è detta troncata a sinistra

se non è osservato il decesso, l'osservazione è detta censurata a destra

Obiettivo: stimare q_x o m_x per $x = a, a+1, \ldots, \omega$ essendo a l'età minima.

Osservazione: Nei modelli di sopravvivenza non parametrici la stima del modello avviene separatamente per ciascuna classe di età

Supponiamo di disporre di dati individuali esatti, cioè per ogni individuo osservato i sono noti:

- data di nascita
- data di ingresso in osservazione
- data di uscita dall'osservazione
- causa di uscita, che può essere: fine osservazione (survival)
decesso (death)
altra causa (withdrawal)

Per ogni individuo osservato i si determina il **vettore delle età**:

$$(y_i, z_i, \theta_i, \phi_i,)$$

essendo

y_i l'età esatta (anno intero + frazione d'anno) di ingresso in osservazione

z_i l'età esatta che l'individuo i avrà alla data in cui terminerà la sua osservazione (può essere la data di fine rilevazione della collettività, oppure la data di scadenza della polizza); è detta **età di uscita pianificata**

θ_i l'età esatta di uscita per morte ($\theta_i = 0$ se l'individuo i non è uscito per morte)

ϕ_i l'età esatta di uscita per altra causa ($\phi_i = 0$ se l'individuo i non è uscito per altra causa)

$(y_i, z_i]$ è detto **intervallo di osservazione pianificata** per l'individuo i

Per ogni individuo osservato si determinano le classi di età $]x, x+1]$ per le quali l'individuo ha contribuito all'osservazione

Con riferimento alla classe di età $]x, x+1]$ ed all'individuo i, caratterizzato dal vettore delle età $(y_i, z_i, \theta_i, \phi_i,)$

l'individuo i non contribuisce alla osservazione per la classe di età $]x, x+1]$ se:

- $y_i \geq x+1$
- $z_i \leq x$
- $0 < \theta_i \leq x$ oppure $0 < \phi_i \leq x$

Se l'individuo i contribuisce alla osservazione per la classe di età $]x, x+1]$

tale osservazione, relativa alla classe di $]x, x+1]$, è riassunta da un vettore detto vettore delle durate

Per ogni classe di età $]x, x+1]$ e per ogni individuo i che contribuisce alla osservazione per tale classe di età si determina il **vettore delle durate**

$$(r_i, s_i, t_i, k_i,)$$

essendo

$x + r_i$ l'età esatta di ingresso in osservazione nella classe di età $]x, x+1]$ con $0 \le r_i < 1$ e

$$r_i = \begin{cases} 0 & se\ y_i \le x \\ y_i - x & se\ x < y_i < x+1 \end{cases}$$

$x + s_i$ l'età esatta pianificata di uscita dalla osservazione per la classe di età $]x, x+1]$ con $0 < s_i \le 1$ e

$$s_i = \begin{cases} 1 & se\ z_i \ge x+1 \\ z_i - x & se\ x < z_i < x+1 \end{cases}$$

$x + t_i$ l'età esatta di uscita per morte se $\theta_i = x + t_i$, altrimenti $t_i = 0$

$x + k_i$ l'età esatta di uscita per altra causa se $\phi_i = x + k_i$, altrimenti $k_i = 0$

ESPOSIZIONE ATTUARIALE E FREQUENZE DI DECESSO

Per stimare q_x ovvero m_x, per $x = a,\, a+1, \ldots, \omega$,sono state introdotte, in ambito attuariale, le stime ottenute rapportando il numero di decessi osservati ad una qualche misura di esposizione.

Con riferimento alla classe di età $]x, x+1]$ e con riferimento agli individui i che contribuiscono alla osservazione per tale classe di età si definiscono

$S = \{i \mid \text{l'individuo } i \text{ è in vita all'età } x + s_i\}$ *survival*

$D = \{i \mid \text{l'individuo } i \text{ esce per morte all'età } x + t_i\}$ *death*

$W = \{i \mid \text{l'individuo } i \text{ esce per altra causa all'età } x + k_i\}$ *withdrawal*

Sia

$\theta_x = \# D$ il numero di individui che decedono nella classe di età $]x, x+1]$

Stime per $q_x = \dfrac{d_x}{l_x}$ per $x = a, a+1, \dots, \omega$

Def. **Frequenza (grezza) di decesso**

$$\overset{o}{q}_x = \frac{\theta_x}{E_x}$$

essendo E_x l'**esposizione attuariale o numero iniziale di esposti al rischio**

$$E_x = \sum_{i \in S \cup D \cup W} (1 - r_i) - \sum_{i \in S} (1 - s_i) - \sum_{i \in W} (1 - k_i)$$

Stime per $m_x = \dfrac{d_x}{L_x}$ per $x = a, a+1, \dots, \omega$

Def. **Frequenza (grezza) centrale di decesso**

$$\overset{o}{m}_x = \frac{\theta_x}{E_x^C}$$

essendo E_x^C il **numero centrale di esposti al rischio**

$$E_x^C = \sum_{i \in S \cup D \cup W} (1 - r_i) - \sum_{i \in S} (1 - s_i) - \sum_{i \in W} (1 - k_i) - \sum_{i \in D} (1 - t_i)$$

Osservazione

Ricordando la relazione $L_x = l_x - d_x(1-\bar{t}_x)$ si nota che si ha

$$E_x^C = E_x - \sum_{i \in D}(1-t_i)$$

Giustificazione di Cantelli

Sia D_x il n.a. dei decessi nella classe di età $]x, x+1]$; si stima q_x con il metodo dei momenti, ponendo

$$E(D_x) = \theta_x$$

Poiché

$$E(D_x) = \sum_{i \in S \cup D \cup W} {}_{1-r_i}q_{x+r_i} - \sum_{i \in S} {}_{1-s_i}q_{x+s_i} - \sum_{i \in W} {}_{1-k_i}q_{x+k_i}$$

nell'ipotesi di interpolazione iperbolica si ha ${}_{1-r}q_{x+r} = (1-r)\,q_x$, quindi

$$E(D_x) = q_x\left[\sum_{i \in S \cup D \cup W}(1-r_i) - \sum_{i \in S}(1-s_i) - \sum_{i \in W}(1-k_i)\right] = q_x\, E_x$$

Da cui si ottiene la stima con il metodo dei momenti: $\overset{o}{q}_x = \dfrac{\theta_x}{E_x}$

Osservazioni

- Nella valutazione si ipotizza uguale mortalità per i soggetti che rimangono nella collettività, per i nuovi ingressi e per coloro che escono per altra causa

- La giustificazione di Cantelli è errata in quanto le valutazioni relative ad uno stesso individuo sono fatte in stati di informazione diversi

- θ_x è il numero di decessi osservati con età esatta nella classe di età $]x, x+1]$.

  Poiché si contano i decessi nell'anno che inizia con l'età esatta x e termina con l'età esatta $x+1$ si dice che si prende come **riferimento l'anno di vita** compreso tra due compleanni.

  Per coerenza, anche nella valutazione dell'esposizione si prende come riferimento l'anno di vita. Infatti, disponendo di dati individuali esatti si determina per ciascun individuo la sua esposizione nell'anno di vita, infatti si ha

$$E_x = \sum_{i \in S} (s_i - r_i) + \sum_{i \in W} (k_i - r_i) + \sum_{i \in D} (1 - r_i)$$

Nota: per gli individui che decedono il contributo all'esposizione è $\sum_{i \in D} (1 - r_i)$

RIFERIMENTO: ANNO DI VITA, ANNO DI POLIZZA E ANNO DI CALENDARIO

Talvolta non è disponibile l'informazione sulla data di nascita dell'individuo osservato e non si è allora in grado di calcolare l'età esatta in cui si verifica un determinato evento (per esempio la morte o l'uscita per altra causa).

Le valutazioni non possono allora essere fatte prendendo come riferimento <u>l'anno di vita</u>.

Ciò tipicamente avviene quando alla stipulazione della polizza si attribuisce all'assicurato una età arrotondata (intera); in tal caso si dirà che si prende come riferimento <u>l'anno di polizza</u>.

Un'altra eventualità si ha nel caso in cui agli individui sia attribuita un'età arrotondata ad una certa data (per esempio all'1/1), come avviene spesso per i fondi pensione; in tal caso si dirà che si prende come riferimento <u>l'anno di calendario</u>.

<u>Definizione</u>

Diremo che un individuo ha **età arrotondata** (intera) x ad una certa data (per es. un anniversario di polizza oppure all'1 gennaio) se in quella data ha età esatta nell'intervallo

$$\left[x - \frac{1}{2}, \, x + \frac{1}{2} \right[$$

Conteggio dei decessi prendendo come riferimento l'anno di polizza

Alla stipulazione della polizza si attribuisce all'assicurato l'età arrotondata (intera).

Si definisce **data di nascita di valutazione**, la data in cui giorno e mese coincidono con il giorno ed il mese di stipulazione della polizza, l'anno è dato da

anno di stipulazione della polizza - età arrotondata alla stipulazione della polizza

Utilizzando tale data di nascita di valutazione al posto della data di nascita e conoscendo

la data di ingresso in osservazione,

la data di uscita dall'osservazione e la causa di uscita

si è in grado di determinare il vettore delle età $(y_i, z_i, \theta_i, \phi_i,)$

Con riferimento alla classe di età $]x, x+1]$ si è in grado di determinare il vettore delle durate $(r_i, s_i, t_i, k_i,)$

essendo

r_i la durata di tempo dall'anniversario di polizza precedente all'ingresso in osservazione nella classe di età $]x, x+1]$ con $0 \le r_i < 1$ e

$$r_i = \begin{cases} 0 & se\ y_i \le x \\ y_i - x & se\ x < y_i < x+1 \end{cases}$$

s_i la durata di tempo dall'anniversario di polizza precedente all'uscita dalla osservazione per la classe di età $]x, x+1]$ con $0 < s_i \le 1$ e

$$s_i = \begin{cases} 1 & se\ z_i \ge x+1 \\ z_i - x & se\ x < z_i < x+1 \end{cases}$$

Poiché, nei casi di uscita per morte e, rispettivamente, per altra causa

$x+t_i$ non è più l'età esatta di uscita per morte, ma t_i è la durata di tempo dall'anniversario di polizza precedente all'uscita per morte

$x+k_i$ non è più l'età esatta di uscita per altra causa, ma k_i è la durata di tempo dall'anniversario di polizza precedente all'uscita per altra causa

si dice che si prende come **riferimento l'anno di polizza** in quanto si considerano gli eventi che avvengono nell'anno di polizza.

Si ha allora che

θ_x è il numero di decessi osservati per gli assicurati nell'anno di polizza $]x, x+1]$

Anche nella valutazione dell'esposizione si prende come riferimento l'anno di polizza, conteggiando le esposizioni tra due anniversari di polizza.

$$E_x = \sum_{i \in S}(s_i - r_i) + \sum_{i \in W}(k_i - r_i) + \sum_{i \in D}(1 - r_i)$$

Le stime

$$\overset{o}{q}_x = \frac{\theta_x}{E_x} \qquad \overset{o}{m}_x = \frac{\theta_x}{E_x^C} \qquad \text{per } x = a, a+1, \dots, \omega$$

forniscono delle stime per, rispettivamente,

$$q_{x+f} \qquad m_{x+f}$$

Essendo

$$-\frac{1}{2} \le f < \frac{1}{2}$$

Nel caso di distribuzione uniforme delle stipulazioni delle polizze rispetto agli anni di vita, ovvero di distribuzione uniforme dei compleanni rispetto all'anno di polizza, si può assumere

$$f = 0$$

e quindi $\overset{o}{q}_x$ e $\overset{o}{m}_x$ forniscono delle stime per, rispettivamente, q_x e m_x

Conteggio dei decessi prendendo come riferimento l’anno di calendario

Nelle assicurazioni collettive o nei fondi pensione, c’è una data prefissata, per esempio l’1/1, chiamata **data di valutazione del fondo** ed agli assicurati viene attribuita l’età arrotondata (intera) in tale data.

Si dice allora che si prende come **riferimento l’anno di calendario** dall’1/1 al 31/12.

È come se tutte le polizze fossero stipulate nella stessa data ed a tutti gli assicurati si attribuisse l’età arrotondata in tale data.

Analogamente a quanto visto nel caso in cui si prenda come riferimento l’anno di polizza, si definisce per ogni individuo osservato, la data di nascita di valutazione, il vettore delle età $(y_i, z_i, \theta_i, \phi_i,)$ e, con riferimento alla classe di età $]x, x+1]$ il vettore delle durate $(r_i, s_i, t_i, k_i,)$

Poiché generalmente l’intervallo di osservazione inizia all’1/1 di un certo anno e termina al 31/12 di qualche anno dopo, per ogni individuo osservato si ha $r_i = 0$ e $s_i = 1$.

Poiché, nei casi di uscita per morte e, rispettivamente, per altra causa

$x+t_i$ non è più l’età esatta di uscita per morte, ma t_i è la durata di tempo dalla data di valutazione del fondo precedente all’uscita per morte

$x+k_i$ non è più l’età esatta di uscita per altra causa, ma k_i è la durata di tempo dalla data di valutazione del fondo precedente all’uscita per altra causa

si dice che si prende come **riferimento l’anno di calendario** in quanto si considerano gli eventi che avvengono nell’anno di calendario.

Si ha allora che

θ_x è il numero di decessi osservati per gli assicurati nell’anno di calendario $]x, x+1]$

Anche nella valutazione dell’esposizione si prende come riferimento l’anno di calendario, conteggiando le esposizioni tra due date di valutazione del fondo.

Le stime

$$\overset{o}{q}_x = \frac{\theta_x}{E_x} \qquad \overset{o}{m}_x = \frac{\theta_x}{E_x^C} \qquad \text{per } x = a, a+1, \dots, \omega$$

forniscono delle stime per, rispettivamente,

$$q_{x+f} \qquad m_{x+f}$$

Essendo

$$-\frac{1}{2} \le f < \frac{1}{2}$$

Nel caso di distribuzione uniforme dei compleanni nell'anno di calendario, si può assumere

$$f = 0$$

e quindi $\overset{o}{q}_x$ e $\overset{o}{m}_x$ forniscono delle stime per, rispettivamente, q_x e m_x

Talvolta si attribuisce come età arrotondata intera ad una data di valutazione del fondo, l'età raggiunta all'ultimo compleanno.

Definizione

Diremo che un individuo ha **età troncata** (intera) x all'1/1 di un certo anno se in quella data ha età esatta nell'intervallo $[x, x+1[$

θ_x è il numero di decessi osservati con età troncata x all'1/1 precedente al decesso

Le stime $\overset{o}{q}_x = \dfrac{\theta_x}{E_x}$ $\overset{o}{m}_x = \dfrac{\theta_x}{E_x^C}$ per $x = a, a+1, \ldots, \omega$

forniscono delle stime per, rispettivamente, q_{x+f} m_{x+f}

essendo

$$0 \leq f < 1$$

Nel caso di distribuzione uniforme dei compleanni nell'anno di calendario, si può assumere

$$f = \frac{1}{2}$$

e quindi $\overset{o}{q}_x$ e $\overset{o}{m}_x$ forniscono delle stime per, rispettivamente, $q_{x+1/2}$ e $m_{x+1/2}$

FREQUENZE DI DECESSO PER TAVOLE SELEZIONATE

Un modello di sopravvivenza selezionato è definito mediante una famiglia di funzioni di sopravvivenza

$$S(t;x) \qquad t \geq 0 \qquad x = a, a+1, \ldots$$

dove

x è l'età (intera) di ingresso in assicurazione

t è l'antidurata dell'assicurazione

Una **tavola di mortalità selezionata** è definita da un insieme di sequenze del tipo

$$l_{[a]} \quad l_{[a]+1} \quad l_{[a]+2} \quad \ldots$$

$$l_{[a+1]} \quad l_{[a+1]+1} \quad l_{[a+1]+2} \quad \ldots$$

$$\ldots$$

$$l_{[x]} \quad l_{[x]+1} \quad l_{[x]+2} \quad \ldots$$

$$\ldots$$

dove

$$l_{[x]+t} = l_{[x]} \cdot S(t;x)$$

$$x = a, a+1, \ldots \qquad t = 0, 1, \ldots$$

Per stimare un modello di sopravvivenza selezionato si determinano le frequenze di decesso per le diverse età di ingresso in assicurazione e antidurate.

Poiché l'età (intera) di ingresso in assicurazione è l'età arrotondata all'emissione della polizza e l'antidurata è il numero di anni in cui l'individuo è presente in assicurazione, si prende come <u>riferimento l'anno di polizza</u>.

Con riferimento all'età arrotondata x di ingresso in assicurazione ed all'intervallo di antidurate $]t, t+1]$,il vettore delle durate $(r_i, s_i, t_i, k_i,)$ dell'individuo i che contribuisce alla osservazione per l'intervallo di antidurate $]t, t+1]$ è così definito

r_i con $0 \leq r_i < 1$ tale che $t + r_i$ è l'antidurata esatta all'ingresso in osservazione nell'intervallo di antidurate $]t, t+1]$

s_i con $0 < s_i \leq 1$ tale che $t + s_i$ è l'antidurata esatta di uscita dall'osservazione dell'intervallo di antidurate $]t, t+1]$

t_i con $0 < t_i \leq 1$ tale che $t + t_i$ è l'antidurata esatta di uscita per morte

k_i con $0 < k_i \leq 1$ tale che $t + k_i$ è l'antidurata esatta di uscita per altra causa

Si definisce

$\theta_{[x]+t}$ il numero di decessi osservati per gli assicurati entrati in assicurazione all'età arrotondata x e con antidurata esatta in $]t, t+1]$

$E_{[x]+t}$ il numero iniziale di esposti al rischio nell'anno di polizza $]t, t+1]$, per gli assicurati entrati in assicurazione all'età arrotondata x

$$E_{[x]+t} = \sum_{i \in S \cup D \cup W} (1 - r_i) - \sum_{i \in S} (1 - s_i) - \sum_{i \in W} (1 - k_i)$$

Si definisce la **frequenza di decesso**

$$\overset{o}{q}_{[x]+t} = \frac{\theta_{[x]+t}}{E_{[x]+t}}$$

che fornisce una stima di $q_{[x+f]+t}$ con $-\frac{1}{2} \leq f < \frac{1}{2}$

Nel caso di distribuzione uniforme dei compleanni nell'anno di polizza si può assumere $f = 0$ e quindi $\overset{o}{q}_{[x]+t}$ fornisce una stima di $q_{[x]+t}$

Indicato con

$E^{C}_{[x]+t}$ il numero centrale di esposti al rischio nell'anno di polizza $]t, t+1]$, per gli assicurati entrati in assicurazione all'età arrotondata x

$$E^{C}_{[x]+t} = \sum_{i \in S \cup D \cup W} (1 - r_i) - \sum_{i \in S} (1 - s_i) - \sum_{i \in W} (1 - k_i) - \sum_{i \in D} (1 - t_i)$$

si definisce la **frequenza centrale di decesso**

$$\overset{o}{m}_{[x]+t} = \frac{\theta_{[x]+t}}{E^{C}_{[x]+t}}$$

che fornisce una stima di $m_{[x+f]+t}$

Poiché l’effetto della selezione si esaurisce entro un certo numero t' di anni

$$q_{[x]} < q_{[x-1]+1} < q_{[x-2]+2} < \ldots < q_{[x-t']+t'} = q_{[x-t'-1]+t'+1} = q_{[x-t'-2]+t'+2} = \ldots = q_{[a]+x-a}$$

si definiscono le tavole selezionate ridotte

$$\begin{array}{cccccc}
l_{[a]} & l_{[a]+1} & l_{[a]+2} & \cdots & l_{[a]+t'-1} & l_{(a+t')} \\
l_{[a+1]} & l_{[a+1]+1} & l_{[a+1]+2} & \cdots & l_{[a+1]+t'-1} & l_{(a+t'+1)} \\
\vdots & & & & & \vdots \\
l_{[x]} & l_{[x]+1} & l_{[x]+2} & \cdots & l_{[x]+t'-1} & l_{(x+t')} \\
\vdots & & & & & \vdots
\end{array}$$

dove $l_{[a]}$ è la radice della tavola e $l_{[x]}$ è tale che

$$l_{[x]} \cdot S(t';x) = l_{(x+t')} \qquad x = a+1, a+2, \ldots$$

Indichiamo con

$$q_{(x)} = q_{[x-t']+t'} = q_{[x-t'-1]+t'+1} = \ldots = q_{[a]+x-a}$$ con $x = a+t$ e $t = t', t'+1, \ldots$

Per stimare

$$q_{(x)} = q_{[x-t']+t'} = q_{[x-t'-1]+t'+1} = \dots = q_{[a]+x-a} \qquad \text{con } x = a+t \quad \text{e} \quad t = t', t'+1, \dots$$

si continua a prendere come riferimento l'anno di polizza e si considerano le frequenze di decesso

$$\overset{o}{q}_{(x)} = \frac{\theta_{(x)}}{E_{(x)}}$$

essendo

$$\theta_{(x)} = \theta_{[x-t']+t'} + \theta_{[x-t'-1]+t'+1} + \dots + \theta_{[a]+x-a}$$

$$E_{(x)} = E_{[x-t']+t'} + E_{[x-t'-1]+t'+1} + \dots + E_{[a]+x-a}$$

oppure le frequenze centrali di decesso

$$\overset{o}{m}_{(x)} = \frac{\theta_{(x)}}{E^C_{(x)}}$$

essendo

$$E^C_{(x)} = E^C_{[x-t']+t'} + E^C_{[x-t'-1]+t'+1} + \dots + E^C_{[a]+x-a}$$

STIMA DI MODELLI DI SOPRAVVIVENZA NON PARAMETRICI
USCITE SOLTANTO PER MORTE

Con riferimento alla classe di età $]x, x+1]$ supponiamo di avere osservato n_x individui e di disporre di dati individuali esatti riassunti, per ogni individuo i che contribuisce alla osservazione per tale classe di età, dal vettore delle durate

$$(r_i, s_i, t_i) \qquad\qquad i = 1, \dots, n_x$$

essendo

$x + r_i$ l'età di ingresso in osservazione nella classe di età $]x, x+1]$ con $0 \le r_i < 1$

$x + s_i$ l'età di uscita pianificata dalla osservazione per la classe di età $]x, x+1]$ con $0 < s_i \le 1$

$x + t_i$ l'età di uscita per morte se $\theta_i = x + t_i$, altrimenti $t_i = 0$

<u>Nota</u>: se il riferimento è l'anno di vita, $x + r_i$, $x + s_i$ e $x + t_i$ sono età esatte, se il riferimento è l'anno di polizza (o l'anno di calendario), x è l'età arrotondata all'anniversario di polizza ed r_i, s_i e t_i sono durate riferire all'anno di polizza (o all'anno di calendario).

Stima con il metodo dei momenti

Si definiscono i n.a.

$$D_i = \begin{cases} 1 & \text{se l'individuo } i \text{ decede nella classe di età } x \\ 0 & \text{altrimenti} \end{cases} \qquad i = 1, \ldots, n_x$$

Si definisce il n.a. D_x dei decessi nella classe di età $]x, x+1]$

$$D_x = \sum_{i=1}^{n_x} D_i$$

Nell'ipotesi che il modello di descrizione della sopravvivenza sia lo stesso per ogni i si ha

$$E(D_x) = \sum_{i=1}^{n_x} E(D_i) = \sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i}$$

Sia d_x il numero dei decessi osservati nella classe di età $]x, x+1]$;
si può scrivere quindi l'equazione dei momenti

$$E(D_x) = d_x \qquad \Leftrightarrow \qquad \sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i} = d_x$$

A) Nell'ipotesi $r_i = 0$ e $s_i = 1$ per ogni i si ha

$$\sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i} = d_x \quad \Leftrightarrow \quad q_x \, n_x = d_x$$

$$\Rightarrow \quad \hat{q}_x = \frac{d_x}{n_x}$$

La stima ottenuta coincide con la stima di massima verosimiglianza di q_x in ipotesi di distribuzione Binomiale(n_x, q_x) per il n.a. D_x

B) Nell'ipotesi di <u>interpolazione lineare</u> con $r_i = 0$ per ogni i si ha

$$\sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i} = d_x \quad \Leftrightarrow \quad \sum_{i=1}^{n_x} {}_{s_i} q_x = d_x \quad \Leftrightarrow \quad \sum_{i=1}^{n_x} s_i \, q_x = d_x$$

$$\Rightarrow \quad \hat{q}_x = \frac{d_x}{\sum_{i=1}^{n_x} s_i}$$

essendo in tale ipotesi ${}_s q_x = s \, q_x$

C) Nell'ipotesi di <u>interpolazione esponenziale</u>

$$ {}_{s-r}q_{x+r} = 1-(1-q_x)^{s-r} $$

che richiede di risolvere l'equazione dei momenti numericamente

D) Nell'ipotesi di <u>interpolazione iperbolica</u> con $s_i = 1$ per ogni i si ha

$$ \sum_{i=1}^{n_x} {}_{s_i-r_i}q_{x+r_i} = d_x \quad \Leftrightarrow \quad \sum_{i=1}^{n_x} {}_{1-r_i}q_{x+r_i} = d_x \quad \Leftrightarrow \quad \sum_{i=1}^{n_x} (1-r_i)\, q_x = d_x $$

$$ \Rightarrow \quad \hat{q}_x = \frac{d_x}{\sum_{i=1}^{n_x}(1-r_i)} $$

essendo in tale ipotesi ${}_{1-r}q_{x+r} = (1-r)\, q_x$

Per risolvere l'equazione dei momenti in forma chiusa

$$ \sum_{i=1}^{n_x} {}_{s_i-r_i}q_{x+r_i} = d_x $$

si formula, in generale, la seguente ipotesi: ${}_{s-r}q_{x+r} = (s-r)\, q_x$

Si ha allora

$$\sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i} = d_x \quad \Leftrightarrow \qquad \sum_{i=1}^{n_x} (s_i - r_i)\, q_x = d_x$$

$$\Rightarrow \quad \hat{q}_x = \frac{d_x}{\sum_{i=1}^{n_x} (s_i - r_i)}$$

dove $\sum_{i=1}^{n_x} (s_i - r_i)$ è detta **esposizione totale pianificata** nella classe di età $]x, x+1]$

Sia

$$\tilde{q}_x = \frac{D_x}{\sum_{i=1}^{n_x} (s_i - r_i)}$$

lo stimatore del quale la stima $\hat{q}_x$ è il valore osservato

Sotto l'ipotesi

$${}_{s-r} q_{x+r} = (s - r)\, q_x$$

$\tilde{q}_x$ è non distorto, infatti $E(\tilde{q}_x) = q_x$

Inoltre, in ipotesi di indipendenza stocastica dei n.a. D_i, $i = 1, \dots, n_x$, si ha

$$Var(\tilde{q}_x) = \frac{\sum_{i=1}^{n_x} {}_{s_i-r_i}q_{x+r_i}\left(1 - {}_{s_i-r_i}q_{x+r_i}\right)}{\left[\sum_{i=1}^{n_x} (s_i - r_i)\right]^2} = \frac{\sum_{i=1}^{n_x} (s_i - r_i) q_x \left(1 - (s_i - r_i) q_x\right)}{\left[\sum_{i=1}^{n_x} (s_i - r_i)\right]^2} = \frac{q_x \sum_{i=1}^{n_x} (s_i - r_i) - q_x^2 \sum_{i=1}^{n_x} (s_i - r_i)^2}{\left[\sum_{i=1}^{n_x} (s_i - r_i)\right]^2}$$

Sostituendo al posto di q_x il valore stimato $\hat{q}_x$ si ottiene una stima di $Var(\tilde{q}_x)$.

Se si considera invece l'ipotesi cosiddetta "binomiale", cioè

$$E(\tilde{q}_x) = q_x \qquad\qquad Var(\tilde{q}_x) = \frac{q_x(1 - q_x)}{\sum_{i=1}^{n_x} (s_i - r_i)}$$

si ottiene la seguente stima della $Var(\tilde{q}_x)$

$$V\hat{a}r(\tilde{q}_x) = \frac{\hat{q}_x(1 - \hat{q}_x)}{\sum_{i=1}^{n_x} (s_i - r_i)}$$

Stima con il metodo della massima verosimiglianza

Con riferimento alla classe di età $\left]x, x+1\right]$ per scrivere la verosimiglianza delle osservazioni

$$\left(r_i, s_i, t_i\right) \qquad\qquad i = 1, \ldots, n_x$$

definiamo, per ogni $i = 1, 2, \ldots, n_x$, i n.a.

$T^{(i)}$ durata aleatoria di vita dell'individuo i nell'intervallo di età $\left]x, x+1\right]$

Nota: $T^{(i)}$ ha determinazioni $\left]r_i, s_i\right]$

Indicato con $T^{(i)}_{x+r_i}$ la durata aleatoria di vita dell'i-esimo individuo in vita all'età $x + r_i$ si ha

$$T^{(i)} = \min\left(T^{(i)}_{x+r_i} + r_i, s_i\right)$$

Nell'ipotesi che per ogni individuo i la durata aleatoria di vita sia descritta dallo stesso modello di sopravvivenza, dotato di funzione di densità, si ha

$$P(T^{(i)} \le t) = \begin{cases} 0 & t \le r_i \\ F_{x+r_i}(t - r_i) & r_i < t < s_i \\ 1 & t \ge s_i \end{cases}$$

ed è inoltre $P(T^{(i)} = s_i) = 1 - F_{x+r_i}(s_i - r_i) = 1 - {}_{s_i - r_i}q_{x+r_i} = {}_{s_i - r_i}p_{x+r_i}$

Quindi se l'individuo i decede con età esatta $x + t_i$ la verosimiglianza di tale osservazione è

$$f_{x+r_i}(t_i - r_i) = {}_{t_i - r_i}p_{x+r_i}\ \mu(x + t_i)$$

Se invece l'individuo i raggiunge in vita l'età di uscita pianificata $x + s_i$, la verosimiglianza di tale osservazione è

$${}_{s_i - r_i}p_{x+r_i}$$

Si definiscono

$S=\{i \mid \text{l'individuo } i \text{ è in vita all'età } x+s_i\}$ *survival*

$D=\{i \mid \text{l'individuo } i \text{ esce per morte all'età } x+t_i\}$ *death*

In ipotesi di indipendenza stocastica dei n.a. $T^{(i)}$ la verosimiglianza delle osservazioni è

$$L=\prod_{i\in S} {}_{s_i-r_i}p_{x+r_i}\cdot\prod_{i\in D} f_{x+r_i}(t_i-r_i)=\prod_{i\in S} {}_{s_i-r_i}p_{x+r_i}\cdot\prod_{i\in D} {}_{t_i-r_i}p_{x+r_i}\;\mu(x+t_i)$$

A) Nell'ipotesi di <u>interpolazione esponenziale</u> $\mu(x+t)=\mu_x$, $0<t\le 1$

$$ {}_{s_i-r_i}p_{x+r_i}=e^{-\mu_x(s_i-r_i)}$$

la verosimiglianza è

$$L=\prod_{i\in S} e^{-\mu_x(s_i-r_i)}\cdot\prod_{i\in D}\left(e^{-\mu_x(t_i-r_i)}\;\mu_x\right)=\exp\left[-\mu_x\left(\sum_{i\in S}(s_i-r_i)+\sum_{i\in D}(t_i-r_i)\right)\right]\cdot(\mu_x)^{d_x}$$

dove $d_x=\#D$

La log-verosimiglianza è allora

$$l = \log L = \left[-\mu_x \left(\sum_{i \in S} (s_i - r_i) + \sum_{i \in D} (t_i - r_i) \right) \right] + d_x \log(\mu_x)$$

Risolvendo l'equazione di verosimiglianza si trova

$$\hat{\mu}_x = \frac{d_x}{\sum_{i \in S} (s_i - r_i) + \sum_{i \in D} (t_i - r_i)}$$

dove $\sum_{i \in S} (s_i - r_i) + \sum_{i \in D} (t_i - r_i)$ è detta **esposizione esatta**

Se poniamo

$$E_x^C = \sum_{i \in S} (s_i - r_i) + \sum_{i \in D} (t_i - r_i)$$

si ha

$$L = e^{-\mu_x E_x^C} \cdot (\mu_x)^{d_x}$$

<u>Osservazione</u>

Sia D_x il n.a. dei decessi nella classe di età $]x, x+1]$, in ipotesi di distribuzione di Poisson di parametro $\mu_x E_x^C$ si ha

$$P(D_x = d_x) = \frac{\left(\mu_x E_x^C\right)^{d_x}}{d_x!} e^{-\mu_x E_x^C} = \frac{\left(E_x^C\right)^{d_x}}{d_x!} e^{-\mu_x E_x^C} \mu_x^{\ d_x} \propto L$$

Pertanto ai fini della stima di massima verosimiglianza dell’intensità istantanea di mortalità sono equivalenti le ipotesi esponenziale e di distribuzione di Poisson per il n.a. dei decessi D_x con $E(D_x) = \mu_x E_x^C$

B) Nell’ipotesi di <u>interpolazione lineare</u>

$$_{s_i - r_i} p_{x+r_i} = \frac{1 - s_i\ q_x}{1 - r_i\ q_x} \qquad \mu(x + t_i) = \frac{q_x}{1 - t_i\ q_x}$$

indicato con $d_x = \# D$, la verosimiglianza è

$$L = \prod_{i \in S} \frac{1 - s_i\ q_x}{1 - r_i\ q_x} \cdot \prod_{i \in D} \left(\frac{1 - t_i\ q_x}{1 - r_i\ q_x} \cdot \frac{q_x}{1 - t_i\ q_x} \right) = \prod_{i \in S \cup D} (1 - r_i\ q_x)^{-1} \prod_{i \in S} (1 - s_i\ q_x) \quad q_x^{\ d_x}$$

Dalla log-verosimiglianza

$$l = \log L = -\sum_{i \in S \cup D} \log(1 - r_i\, q_x) + \sum_{i \in S} \log(1 - s_i\, q_x) + d_x \log(q_x)$$

si ottiene l'equazione di log-verosimiglianza che può essere risolta per via numerica

$$\sum_{i \in S \cup D} \frac{r_i}{1 - r_i\, q_x} - \sum_{i \in S} \frac{s_i}{1 - s_i\, q_x} + \frac{d_x}{q_x} = 0$$

<u>Osservazione</u>

Soluzioni in forma chiusa possono essere ottenute per dati particolari (p. es. se $r_i = 0$ e $s_i = 1$ per ogni i) e nel caso di particolari dati raggruppati (per es. se nel caso $s_i = 1$ per ogni i, si considera una comune età media di ingresso r con $0 < r < 1$ per ogni i, oppure nel caso $r_i = 0$ per ogni i, se si considera una comune età media di uscita pianificata s con $0 < s < 1$ per ogni i)

Vedi London, cap. 7

MODELLO DI SOPRAVVIVENZA CON PIÙ CAUSE DI ELIMINAZIONE

Sia una collettività di individui soggetta a m cause di eliminazione $\alpha_1, \dots, \alpha_m$

Si definiscono i n.a.

T_x durata di permanenza nella collettività per un individuo presente nella collettività all'età x

C con determinazioni $1, \dots, m$, tale che $(C = j) \Leftrightarrow$ "l'individuo esce per la causa α_j"

Si introduce un modello probabilistico per la coppia di n.a. (T_x, C)

Sia

${}_t q_x^{(\alpha_j)} = P(T_x \le t, C = j)$ probabilità che l'individuo presente nella collettività all'età x esca dalla collettività per la causa α_j entro l'età $x+t$

${}_t q_x^{(\tau)} = P(T_x \le t)$ probabilità che l'individuo presente nella collettività all'età x esca dalla collettività entro l'età $x+t$ per una qualsiasi causa

${}_t p_x^{(\tau)} = 1 - {}_t q_x^{(\tau)}$ probabilità che l'individuo presente nella collettività all'età x sia presente all'età $x+t$

Si ha

$$ {}_t q_x^{(\tau)} = P(T_x \le t) = \sum_{j=1}^{m} P(T_x \le t, C = j) = \sum_{j=1}^{m} {}_t q_x^{(\alpha_j)} $$

Si definisce **intensità di eliminazione per una qualunque causa**

$$ \mu^{(\tau)}(x+t) = \lim_{\Delta t \to 0} \frac{P(T_x \le t + \Delta t | T_x > t)}{\Delta t} $$

Si ha

$$ {}_t p_x^{(\tau)} = \exp\left(-\int_0^t \mu^{(\tau)}(x+u) du \right) $$

Si definisce **intensità di eliminazione per la causa** α_j

$$ a\mu^{(\alpha_j)}(x+t) = \lim_{\Delta t \to 0} \frac{P(T_x \le t + \Delta t, C = j | T_x > t)}{\Delta t} $$

Si ha

$$ \mu^{(\tau)}(x+t) = \sum_{j=1}^{m} a\mu^{(\alpha_j)}(x+t) $$

Poiché

$$a\mu^{(\alpha_j)}(x+t)=\lim_{\Delta t\to 0}\frac{P(T_x\le t+\Delta t, C=j \mid T_x>t)}{\Delta t}=\lim_{\Delta t\to 0}\frac{{}_{t+\Delta t}q_x^{(\alpha_j)}-{}_t q_x^{(\alpha_j)}}{\Delta t}\frac{1}{{}_t p_x^{(\tau)}}$$

se esiste finito

$$\lim_{\Delta t\to 0}\frac{{}_{t+\Delta t}q_x^{(\alpha_j)}-{}_t q_x^{(\alpha_j)}}{\Delta t}$$

e poniamo

$$f_{T,C}(t,j)=\lim_{\Delta t\to 0}\frac{{}_{t+\Delta t}q_x^{(\alpha_j)}-{}_t q_x^{(\alpha_j)}}{\Delta t}$$

la distribuzione congiunta della coppia di n.a. (T_x, C), si ha

$$a\mu^{(\alpha_j)}(x+t)=\frac{f_{T,C}(t,j)}{{}_t p_x^{(\tau)}} \qquad f_{T,C}(t,j)={}_t p_x^{(\tau)}\cdot a\mu^{(\alpha_j)}(x+t)$$

e quindi per la probabilità di eliminazione per la causa α_j

$${}_t q_x^{(\alpha_j)}=P(T_x\le t, C=j)=\int_0^t f_{T,C}(u,j)du=\int_0^t {}_u p_x^{(\tau)}\cdot a\mu^{(\alpha_j)}(x+u)du \quad j=1,\ldots,m$$

Osservazione

Disponendo di osservazioni sulla coppia di n.a. (T_x, C) si possono stimare le intensità di eliminazione per le varie cause $a\mu^{(\alpha_j)}(t),\ t \geq 0,\ j = 1,\dots,m$, e quindi la distribuzione congiunta $f_{T,C}(t, j)\ t \geq 0,\ j = 1,\dots,m$

Se si è invece interessati a stimare un modello di sopravvivenza in relazione alla mortalità nella collettività osservata, tenendo conto che sono presenti anche altre cause di uscita, occorre introdurre ulteriori ipotesi.

Siano i n.a.

$T_x^{(j)}$ durata di permanenza nella collettività per un individuo di età x fino al verificarsi dell'uscita per la causa α_j

Si ha

$$T_x = \min\left(T_x^{(1)}, \dots, T_x^{(m)}\right) \qquad \text{e} \qquad (C = j) \Leftrightarrow \left(T_x = T_x^{(j)}\right)$$

In relazione alle distribuzioni marginali dei n.a. $T_x^{(j)},\ j=1,\dots,m$ si definisce

Def.: **intensità marginale di eliminazione per la causa** α_j

$$\mu^{(\alpha_j)}(x+t)=\lim_{\Delta t\to 0}\frac{P\left(T_x^{(j)}\le t+\Delta t \,\middle|\, T_x^{(j)}>t\right)}{\Delta t}$$

Def.: **probabilità assoluta di sopravvivenza** e, rispettivamente, **di eliminazione**

$${}_t p_x'^{(\alpha_j)}=\exp\left(-\int_0^t \mu^{(\alpha_j)}(x+u)du\right) \qquad {}_t q_x'^{(\alpha_j)}=1-{}_t p_x'^{(\alpha_j)}$$

<u>Osservazione</u>

Sono dette <u>probabilità relative di eliminazione</u> per le diverse cause

$${}_t q_x^{(\alpha_j)}=P(T_x\le t, C=j)=1-{}_t p_x^{(\alpha_j)}=1-\exp\left(-\int_0^t a\mu^{(\alpha_j)}(x+u)du\right) \qquad j=1,\dots,m$$

In generale si ha

$$\mu^{(\alpha_j)}(t) \neq a\mu^{(\alpha_j)}(t) \qquad j = 1, \ldots, m \qquad t \geq 0$$

Nell'<u>ipotesi</u>

$$\mu^{(\alpha_j)}(t) = a\mu^{(\alpha_j)}(t) \qquad j = 1, \ldots, m \qquad t \geq 0$$

sussiste la seguente **relazione di Karup**

$$_t p_x^{(\tau)} = \prod_{j=1}^{m} {}_t p_x'^{(\alpha_j)}$$

e si ha allora

$$f_{T,C}(t, j) = {}_t p_x^{(\tau)} \cdot a\mu^{(\alpha_j)}(x+t) = \left(\prod_{j=1}^{m} {}_t p_x'^{(\alpha_j)} \right) \mu^{(\alpha_j)}(x+t)$$

Si dimostra che se i n.a. $T_x^{(1)}, \ldots, T_x^{(m)}$ sono stocasticamente indipendenti allora

$$\mu_x^{(\alpha_j)}(t) = a\mu_x^{(\alpha_j)}(t) \qquad t \geq 0$$

Nella realtà i n.a. $T_x^{(1)}, \ldots, T_x^{(m)}$ presentano delle relazioni di dipendenza.

Si può allora porre il problema se a partire dai dati osservati, che consentono di stimare la distribuzione congiunta di (T_x, C), si è in grado di stimare pure la distribuzione congiunta di $\left(T_x^{(1)}, \ldots, T_x^{(m)}\right)$

Senza introdurre ipotesi aggiuntive sui legami di dipendenza tra tali n.a., la risposta è negativa.

Sussiste infatti il problema della non identificabilità:

> Esistono diverse distribuzioni congiunte di $\left(T_x^{(1)}, \ldots, T_x^{(m)}\right)$ che danno luogo alla stessa distribuzione congiunta della coppia di n.a. (T_x, C)

STIMA DI MODELLI DI SOPRAVVIVENZA NON PARAMETRICI
USCITE PER MORTE E PER ALTRA CAUSA

Con riferimento alla classe di età $]x, x+1]$ supponiamo di avere osservato n_x individui e di disporre di dati individuali esatti riassunti, per ogni individuo i che contribuisce alla osservazione per tale classe di età, dal vettore delle durate

$$(r_i, s_i, t_i, k_i,) \qquad\qquad i = 1, \dots, n_x$$

essendo

$x + r_i$ l'età di ingresso in osservazione nella classe di età $]x, x+1]$ con $0 \le r_i < 1$

$x + s_i$ l'età di uscita pianificata dalla osservazione per la classe di età $]x, x+1]$ con $0 < s_i \le 1$

$x + t_i$ l'età di uscita per morte se $\theta_i = x + t_i$, altrimenti $t_i = 0$

$x + k_i$ l'età esatta di uscita per altra causa se $\phi_i = x + k_i$, altrimenti $k_i = 0$

Nota: il riferimento può essere sia l'anno di vita, sia l'anno di polizza (o l'anno di calendario).

Modello di sopravvivenza a due cause di eliminazione: morte ed altra causa

Sia una collettività di individui soggetta a due cause di eliminazione:

d morte

w altra causa

Sia

T_x durata di permanenza nella collettività per un individuo presente nella collettività all'età x

$$T_x = \min\left(T_x^{(d)}, T_x^{(w)}\right)$$

essendo

$T_x^{(d)}$ durata di permanenza nella collettività finché non si ha l'uscita per morte, per un individuo presente nella collettività all'età x

$T_x^{(w)}$ durata di permanenza nella collettività finché non si ha l'uscita per altra causa, per un individuo presente nella collettività all'età x

<u>Ipotesi</u>: **uscite non informative**

$$\mu^{(d)}(t) = a\mu^{(d)}(t) \qquad t \geq 0$$

$$\mu^{(w)}(t) = a\mu^{(w)}(t) \qquad t \geq 0$$

essendo

$$\mu^{(d)}(x+t) = \lim_{\Delta t \to 0} \frac{P\left(T_x^{(d)} \leq t + \Delta t \middle| T_x^{(d)} > t\right)}{\Delta t}$$

$$\mu^{(w)}(x+t) = \lim_{\Delta t \to 0} \frac{P\left(T_x^{(w)} \leq t + \Delta t \middle| T_x^{(w)} > t\right)}{\Delta t}$$

le <u>intensità marginali di eliminazione</u> e

$$a\mu^{(d)}(x+t) = \lim_{\Delta t \to 0} \frac{P\left(T_x \leq t + \Delta t, C = 1 \middle| T_x > t\right)}{\Delta t}$$

$$a\mu^{(w)}(x+t) = \lim_{\Delta t \to 0} \frac{P\left(T_x \leq t + \Delta t, C = 2 \middle| T_x > t\right)}{\Delta t}$$

le <u>intensità di uscita per le varie cause</u>, dove $C = 1$ denota l’evento “uscita per morte” e $C = 2$ denota l’evento uscita per altra causa

In ipotesi di uscite non informative sussiste inoltre la **relazione di Karup**

$$ {}_t p_x^{(\tau)} = {}_t p_x'^{(d)} \; {}_t p_x'^{(w)} $$

essendo

$$ {}_t p_x'^{(d)} = \exp\left(-\int_0^t \mu^{(d)}(x+u)\,du \right) \qquad {}_t p_x'^{(w)} = \exp\left(-\int_0^t \mu^{(w)}(x+u)\,du \right) $$

${}_t p_x^{(\tau)} = 1 - {}_t q_x^{(\tau)}$ probabilità che l'individuo presente nella collettività all'età x sia presente all'età $x+t$

${}_t q_x^{(\tau)} = P(T_x \le t)$ probabilità che l'individuo presente nella collettività all'età x esca dalla collettività entro l'età $x+t$ per una qualsiasi causa

Si ottengono allora le seguenti espressioni per le probabilità di uscita per morte e, rispettivamente, per altra causa

$$ {}_{s-r}q_{x+r}^{(d)} = P\left(T_{x+r} \leq s-r, C=1\right) = \int_r^s f_{T,C}(u-r,1)du = \int_r^s {}_{u-r}p_{x+r}^{(\tau)} \cdot a\mu^{(d)}(x+u)du $$

$$ = \int_r^s {}_{u-r}p_{x+r}^{\prime(d)} \cdot {}_{u-r}p_{x+r}^{\prime(w)} \cdot \mu^{(d)}(x+u)du $$

$$ {}_{s-r}q_{x+r}^{(w)} = P\left(T_{x+r} \leq s-r, C=2\right) = \int_r^s f_{T,C}(u-r,2)du = \int_r^s {}_{u-r}p_{x+r}^{(\tau)} \cdot a\mu^{(w)}(x+u)du $$

$$ = \int_r^s {}_{u-r}p_{x+r}^{\prime(d)} \cdot {}_{u-r}p_{x+r}^{\prime(w)} \cdot \mu^{(w)}(x+u)du $$

Ed è inoltre

$$ {}_{s-r}q_{x+r}^{(\tau)} = {}_{s-r}q_{x+r}^{(d)} + {}_{s-r}q_{x+r}^{(w)} $$

Stima con il metodo dei momenti

n_x individui osservati in relazione alla classe di età $]x, x+1]$

$$(r_i, s_i, t_i, k_i,) \qquad i = 1, \ldots, n_x$$

Si definiscono i n.a.

D_x n.a. dei decessi nella classe di età $]x, x+1]$

W_x n.a. delle uscite per altra causa nella classe di età $]x, x+1]$

Siano

d_x il numero dei decessi osservati nella classe di età $]x, x+1]$;

w_x il numero di uscite per altra causa osservate nella classe di età $]x, x+1]$;

Si ha

$$E(D_x) = \sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i}^{(d)} \qquad E(W_x) = \sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i}^{(w)}$$

dove ${}_{s_i - r_i} q_{x+r_i}^{(d)}$ e ${}_{s_i - r_i} q_{x+r_i}^{(w)}$ sono le probabilità di uscita per morte e per altra causa

Le equazioni dei momenti sono allora

$$\begin{cases} E(D_x) = d_x \\ E(W_x) = w_x \end{cases} \quad \Leftrightarrow \quad \begin{cases} \sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i}^{(d)} = d_x \\ \sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i}^{(w)} = w_x \end{cases}$$

Nelle ipotesi

$$ {}_{s-r} q_{x+r}^{(d)} = (s-r)\, q_x^{(d)} \qquad {}_{s-r} q_{x+r}^{(w)} = (s-r)\, q_x^{(w)}$$

si ottengono le seguenti stime:

$$\hat{q}_x^{(d)} = \frac{d_x}{\sum_{i=1}^{n_x} (s_i - r_i)} \qquad \hat{q}_x^{(w)} = \frac{w_x}{\sum_{i=1}^{n_x} (s_i - r_i)}$$

delle probabilità di uscita, rispettivamente, per morte e per altra causa

Se però l’obiettivo è stimare una tavola di mortalità, tenendo conto che sulla collettività agiscono due cause di uscita, si devono stimare le probabilità assolute

$$ {}_t p_x'^{(d)} = \exp\left(-\int_0^t \mu^{(d)}(x+u)\,du \right) \qquad {}_t p_x'^{(w)} = \exp\left(-\int_0^t \mu^{(w)}(x+u)\,du \right)$$

Nell'ipotesi di uscite non informative le equazioni dei momenti sono

$$\begin{cases} E(D_x) = d_x \\ E(W_x) = w_x \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i}^{(d)} = d_x \\ \sum_{i=1}^{n_x} {}_{s_i - r_i} q_{x+r_i}^{(w)} = w_x \end{cases} \Leftrightarrow \begin{cases} \sum_{i=1}^{n_x} \int_{r_i}^{s_i} {}_{u-r_i} p_{x+r_i}'^{(d)} \cdot {}_{u-r_i} p_{x+r_i}'^{(w)} \cdot \mu^{(d)}(x+u) du = d_x \\ \sum_{i=1}^{n_x} \int_{r_i}^{s_i} {}_{u-r_i} p_{x+r_i}'^{(d)} \cdot {}_{u-r_i} p_{x+r_i}'^{(w)} \cdot \mu^{(w)}(x+u) du = w_x \end{cases}$$

A) Nell'ipotesi di <u>distribuzione uniforme</u> per le probabilità assolute di uscita per morte e, rispettivamente, per altra causa si ha

$${}_{u-r_i} p_{x+r_i}'^{(d)} = \frac{1 - u\, q_x'^{(d)}}{1 - r_i\, q_x'^{(d)}} \qquad \mu^{(d)}(x+u) = \frac{q_x'^{(d)}}{1 - u\, q_x'^{(d)}}$$

$${}_{u-r_i} p_{x+r_i}'^{(w)} = \frac{1 - u\, q_x'^{(w)}}{1 - r_i\, q_x'^{(w)}} \qquad \mu^{(w)}(x+u) = \frac{q_x'^{(w)}}{1 - u\, q_x'^{(w)}}$$

Le equazioni dei momenti diventano

$$\begin{cases} \displaystyle\sum_{i=1}^{n_x} \frac{q_x'^{(d)}\left[(s_i - r_i) - \frac{\left(s_i^2 - r_i^2\right)}{2} q_x'^{(w)}\right]}{\left(1 - r_i\, q_x'^{(d)}\right)\left(1 - r_i\, q_x'^{(w)}\right)} = d_x \\ \displaystyle\sum_{i=1}^{n_x} \frac{q_x'^{(w)}\left[(s_i - r_i) - \frac{\left(s_i^2 - r_i^2\right)}{2} q_x'^{(d)}\right]}{\left(1 - r_i\, q_x'^{(d)}\right)\left(1 - r_i\, q_x'^{(w)}\right)} = w_x \end{cases}$$

Il sistema può essere risolto per via numerica ottenendo le stime

$$\hat{q}_x'^{(d)} \qquad \hat{q}_x'^{(w)}$$

Nel caso particolare $r_i = 0$ e $s_i = 1$ per ogni i , le equazioni dei momenti diventano

$$\begin{cases} n_x \, q_x'^{(d)} \left[1 - \frac{1}{2} q_x'^{(w)} \right] = d_x \\ n_x \, q_x'^{(w)} \left[1 - \frac{1}{2} q_x'^{(d)} \right] = w_x \end{cases}$$

Il sistema può essere risolto in forma chiusa ottenendo le stime

$$\hat{q}_x'^{(d)} = \frac{b - \sqrt{b^2 - 2 n_x d_x}}{n_x} \qquad \text{con} \qquad b = n_x + \frac{d_x}{2} - \frac{w_x}{2}$$

$$\hat{q}_x'^{(w)} = \frac{b - \sqrt{b^2 - 2 n_x w_x}}{n_x} \qquad \text{con} \qquad b = n_x - \frac{d_x}{2} + \frac{w_x}{2}$$

B) Nell'ipotesi di <u>intensità di uscita per morte e per altra causa costanti</u> si ha

$$\mu^{(d)}(x+u)=\mu_x^{(d)} \qquad {}_{u-r_i}p'^{(d)}_{x+r_i} = \exp\left[-\mu_x^{(d)}(u-r_i)\right]$$

$$\mu^{(w)}(x+u)=\mu_x^{(w)} \qquad {}_{u-r_i}p'^{(w)}_{x+r_i} = \exp\left[-\mu_x^{(w)}(u-r_i)\right]$$

Le equazioni dei momenti diventano

$$\begin{cases} \dfrac{\mu_x^{(d)}}{\mu_x^{(d)}+\mu_x^{(w)}} \displaystyle\sum_{i=1}^{n_x} \left(1-\exp\left[-(s_i-r_i)\left(\mu_x^{(d)}+\mu_x^{(w)}\right)\right]\right)=d_x \\ \dfrac{\mu_x^{(w)}}{\mu_x^{(d)}+\mu_x^{(w)}} \displaystyle\sum_{i=1}^{n_x} \left(1-\exp\left[-(s_i-r_i)\left(\mu_x^{(d)}+\mu_x^{(w)}\right)\right]\right)=w_x \end{cases}$$

Il sistema può essere risolto per via numerica ottenendo le stime

$$\hat{\mu}_x^{(d)} \qquad \hat{\mu}_x^{(w)}$$

Nel caso particolare $r_i = 0$ e $s_i = 1$ per ogni i , le equazioni dei momenti diventano

$$\begin{cases} \dfrac{\mu_x^{(d)}}{\mu_x^{(d)} + \mu_x^{(w)}} \left(1 - \exp\left[-\left(\mu_x^{(d)} + \mu_x^{(w)}\right)\right]\right) n_x = d_x \\ \dfrac{\mu_x^{(w)}}{\mu_x^{(d)} + \mu_x^{(w)}} \left(1 - \exp\left[-\left(\mu_x^{(d)} + \mu_x^{(w)}\right)\right]\right) n_x = w_x \end{cases}$$

Il sistema può essere risolto in forma chiusa ottenendo le stime

$$\hat{\mu}_x^{(d)} = -\log\left(\frac{n_x - d_x - w_x}{n_x}\right)^{\frac{d_x}{d_x + w_x}} \qquad \hat{\mu}_x^{(w)} = -\log\left(\frac{n_x - d_x - w_x}{n_x}\right)^{\frac{w_x}{d_x + w_x}}$$

Da queste si ottengono le stime

$$\hat{q}_x'^{(d)} = 1 - \hat{p}_x'^{(d)} = 1 - \left(\frac{n_x - d_x - w_x}{n_x}\right)^{\frac{d_x}{d_x + w_x}} \qquad \hat{q}_x'^{(w)} = 1 - \hat{p}_x'^{(w)} = 1 - \left(\frac{n_x - d_x - w_x}{n_x}\right)^{\frac{w_x}{d_x + w_x}}$$

Stima con il metodo della massima verosimiglianza

Con riferimento alla classe di età $]x, x+1]$ per scrivere la verosimiglianza delle osservazioni

$$(r_i, s_i, t_i, k_i,) \qquad\qquad i = 1, \dots, n_x$$

definiamo, per ogni $i = 1, 2, \dots, n_x$, i n.a.

$T^{(i)}$ durata aleatoria di permanenza dell'individuo i nella collettività tra le età $]x, x+1]$

<u>Nota</u>: $T^{(i)}$ ha determinazioni $]r_i, s_i]$

Indicato con $T^{(i)}_{x+r_i}$ la durata di permanenza nella collettività per l'i-esimo individuo presente all'età $x + r_i$ si ha

$$T^{(i)} = \min\left(T^{(i)}_{x+r_i} + r_i, s_i\right)$$

Se l'individuo i è presente nella collettività all'età di uscita pianificata:

$$T^{(i)} = s_i \qquad P(T^{(i)} = s_i) = {}_{s_i - r_i}p^{(\tau)}_{x+r_i} = {}_{s_i - r_i}p'^{(d)}_{x+r_i} \cdot {}_{s_i - r_i}p'^{(w)}_{x+r_i}$$

Se l'individuo i esce per morte all'età esatta $x + t_i$:

$$T^{(i)} = t_i \qquad f_{T,C}(t_i - r_i, 1) = {}_{t_i - r_i}p^{(\tau)}_{x+r_i} \cdot \mu^{(d)}(x + t_i) = {}_{t_i - r_i}p'^{(d)}_{x+r_i} \cdot {}_{t_i - r_i}p'^{(w)}_{x+r_i} \cdot \mu^{(d)}(x + t_i)$$

Se l'individuo i esce per altra causa all'età esatta $x + k_i$:

$$T^{(i)} = k_i \qquad f_{T,C}(k_i - r_i, 2) = {}_{k_i - r_i}p^{(\tau)}_{x+r_i} \cdot \mu^{(w)}(x + k_i) = {}_{k_i - r_i}p'^{(d)}_{x+r_i} \cdot {}_{k_i - r_i}p'^{(w)}_{x+r_i} \cdot \mu^{(w)}(x + k_i)$$

Si definiscono

$S = \{i \mid \text{l'individuo } i \text{ è in vita all'età } x + s_i\}$ *survival*

$D = \{i \mid \text{l'individuo } i \text{ esce per morte all'età } x + t_i\}$ *death*

$W = \{i \mid \text{l'individuo } i \text{ esce per altra causa all'età } x + k_i\}$ *withdrawal*

In ipotesi di indipendenza stocastica dei n.a. $T^{(i)}$ la verosimiglianza delle osservazioni è

$$L = \prod_{i \in S} P(T^{(i)} = s_i) \cdot \prod_{i \in D} f_{T,C}(t_i - r_i, 1) \cdot \prod_{i \in W} f_{T,C}(k_i - r_i, 2)$$

$$= \prod_{i \in S} {}_{s_i - r_i} p^{(\tau)}_{x+r_i} \cdot \prod_{i \in D} {}_{t_i - r_i} p^{(\tau)}_{x+r_i} \cdot \mu^{(d)}(x + t_i) \cdot \prod_{i \in W} {}_{k_i - r_i} p^{(\tau)}_{x+r_i} \cdot \mu^{(w)}(x + k_i)$$

$$= \prod_{i \in S} {}_{s_i - r_i} p'^{(d)}_{x+r_i} \cdot {}_{s_i - r_i} p'^{(w)}_{x+r_i} \cdot$$

$$\cdot \prod_{i \in D} {}_{t_i - r_i} p'^{(d)}_{x+r_i} \cdot {}_{t_i - r_i} p'^{(w)}_{x+r_i} \cdot \mu^{(d)}(x + t_i) \cdot$$

$$\cdot \prod_{i \in W} {}_{k_i - r_i} p'^{(d)}_{x+r_i} \cdot {}_{k_i - r_i} p'^{(w)}_{x+r_i} \cdot \mu^{(w)}(x + k_i)$$

$$= \prod_{i \in S} {}_{s_i - r_i} p'^{(d)}_{x+r_i} \cdot \prod_{i \in D} {}_{t_i - r_i} p'^{(d)}_{x+r_i} \cdot \mu^{(d)}(x + t_i) \cdot \prod_{i \in W} {}_{k_i - r_i} p'^{(d)}_{x+r_i} \cdot$$

$$\cdot \prod_{i \in S} {}_{s_i - r_i} p'^{(w)}_{x+r_i} \cdot \prod_{i \in D} {}_{t_i - r_i} p'^{(w)}_{x+r_i} \cdot \prod_{i \in W} {}_{k_i - r_i} p'^{(w)}_{x+r_i} \cdot \mu^{(w)}(x + k_i)$$

Quindi

$$L = L^{(d)} \cdot L^{(w)}$$

con

$$L^{(d)} = \prod_{i \in S} {}_{s_i - r_i} p'^{(d)}_{x+r_i} \cdot \prod_{i \in D} {}_{t_i - r_i} p'^{(d)}_{x+r_i} \cdot \mu^{(d)}(x + t_i) \cdot \prod_{i \in W} {}_{k_i - r_i} p'^{(d)}_{x+r_i}$$

e

$$L^{(w)} = \prod_{i \in S} {}_{s_i - r_i} p'^{(w)}_{x+r_i} \cdot \prod_{i \in D} {}_{t_i - r_i} p'^{(w)}_{x+r_i} \cdot \prod_{i \in W} {}_{k_i - r_i} p'^{(w)}_{x+r_i} \cdot \mu^{(w)}(x + k_i)$$

Quindi per stimare le probabilità assolute $p'^{(d)}_x$ si porrà

$$\max L^{(d)}$$

mentre per stimare le probabilità assolute $p'^{(w)}_x$ si porrà

$$\max L^{(w)}$$

Stima di massima verosimiglianza delle probabilità assolute $p_x'^{(d)}$

$$L^{(d)} = \prod_{i \in S} {}_{s_i - r_i} p'^{(d)}_{x+r_i} \cdot \prod_{i \in D} {}_{t_i - r_i} p'^{(d)}_{x+r_i} \cdot \mu^{(d)}(x + t_i) \cdot \prod_{i \in W} {}_{k_i - r_i} p'^{(d)}_{x+r_i}$$

Si nota che le informazioni sulle uscite per altra causa sono trattate come le informazioni sulla sopravvivenza all'età di uscita pianificata

A) Nell'ipotesi di <u>interpolazione esponenziale</u> $\mu^{(d)}(x+t) = \mu_x^{(d)}$, $0 < t \le 1$

$${}_{s_i - r_i} p'^{(d)}_{x+r_i} = e^{-\mu_x^{(d)}(s_i - r_i)}$$

la verosimiglianza è

$$L^{(d)} = \prod_{i \in S} e^{-\mu_x^{(d)}(s_i - r_i)} \cdot \prod_{i \in D} \left(e^{-\mu_x^{(d)}(t_i - r_i)} \mu_x^{(d)} \right) \cdot \prod_{i \in W} e^{-\mu_x^{(d)}(k_i - r_i)}$$

$$= \exp\left[-\mu_x^{(d)} \left(\sum_{i \in S} (s_i - r_i) + \sum_{i \in D} (t_i - r_i) + \sum_{i \in W} (k_i - r_i) \right) \right] \cdot \left(\mu_x^{(d)} \right)^{d_x}$$

dove $d_x = \# D$

La log-verosimiglianza è allora

$$\log L^{(d)} = \left[-\mu_x^{(d)} \left(\sum_{i \in S} (s_i - r_i) + \sum_{i \in D} (t_i - r_i) + \sum_{i \in W} (k_i - r_i) \right) \right] + d_x \log\left(\mu_x^{(d)}\right)$$

Risolvendo l'equazione di verosimiglianza si trova

$$\hat{\mu}_x^{(d)} = \frac{d_x}{\sum_{i \in S} (s_i - r_i) + \sum_{i \in D} (t_i - r_i) + \sum_{i \in W} (k_i - r_i)}$$

dove $\sum_{i \in S} (s_i - r_i) + \sum_{i \in D} (t_i - r_i) + \sum_{i \in W} (k_i - r_i)$ è detta **esposizione totale esatta**

<u>Osservazione</u>: l'esposizione totale esatta coincide con il numero centrale degli esposti al rischio secondo l'impostazione attuariale

$$E_x^C = \sum_{i \in S \cup D \cup W} (1 - r_i) - \sum_{i \in S} (1 - s_i) - \sum_{i \in W} (1 - k_i) - \sum_{i \in D} (1 - t_i) = \sum_{i \in S} (s_i - r_i) + \sum_{i \in D} (t_i - r_i) + \sum_{i \in W} (k_i - r_i)$$

Quindi la frequenza (grezza) centrale di decesso $\overset{o}{m}_x = \frac{\theta_x}{E_x^C}$

coincide con la stima di massima verosimiglianza $\hat{\mu}_x^{(d)}$

<u>Osservazione</u>

Sia D_x il n.a. dei decessi nella classe di età $]x, x+1]$, in ipotesi di distribuzione di Poisson di parametro $\mu_x^{(d)} E_x^C$ si ha

$$P(D_x = d_x) = \frac{\left(\mu_x^{(d)} E_x^C\right)^{d_x}}{d_x!} e^{-\mu_x^{(d)} E_x^C} = \frac{\left(E_x^C\right)^{d_x}}{d_x!} e^{-\mu_x^{(d)} E_x^C} \left(\mu_x^{(d)}\right)^{d_x} \propto L^{(d)}$$

Pertanto ai fini della stima di massima verosimiglianza dell’intensità istantanea di mortalità sono equivalenti le ipotesi esponenziale e di distribuzione di Poisson per il n.a. dei decessi D_x con $E(D_x) = \mu_x E_x^C$

B) Nell'ipotesi di <u>interpolazione lineare</u>

$$ {}_{s_i - r_i} p'^{(d)}_{x+r_i} = \frac{1 - s_i \; q'^{(d)}_x}{1 - r_i \; q'^{(d)}_x} \qquad \mu^{(d)}(x + t_i) = \frac{q'^{(d)}_x}{1 - t_i \; q'^{(d)}_x} $$

indicato con $d_x = \# D$, la verosimiglianza è

$$ L^{(d)} = \prod_{i \in S} \frac{1 - s_i \; q'^{(d)}_x}{1 - r_i \; q'^{(d)}_x} \cdot \prod_{i \in D} \left(\frac{1 - t_i \; q'^{(d)}_x}{1 - r_i \; q'^{(d)}_x} \cdot \frac{q'^{(d)}_x}{1 - t_i \; q'^{(d)}_x} \right) \cdot \prod_{i \in W} \frac{1 - k_i \; q'^{(d)}_x}{1 - r_i \; q'^{(d)}_x} $$

$$ = \prod_{i \in S \cup D \cup W} \left(1 - r_i \; q'^{(d)}_x\right)^{-1} \prod_{i \in S} \left(1 - s_i \; q'^{(d)}_x\right) \prod_{i \in W} \left(1 - k_i \; q'^{(d)}_x\right) \quad q_x^{\,d_x} $$

Dalla log-verosimiglianza

$$ \log L^{(d)} = - \sum_{i \in S \cup D \cup W} \log\left(1 - r_i \; q'^{(d)}_x\right) + \sum_{i \in S} \log\left(1 - s_i \; q'^{(d)}_x\right) + \sum_{i \in W} \log\left(1 - k_i \; q'^{(d)}_x\right) + d_x \log\left(q'^{(d)}_x\right) $$

si ottiene l'equazione di log-verosimiglianza che può essere risolta per via numerica

$$ \sum_{i \in S \cup D \cup W} \frac{r_i}{1 - r_i \; q'^{(d)}_x} - \sum_{i \in S} \frac{s_i}{1 - s_i \; q'^{(d)}_x} - \sum_{i \in W} \frac{k_i}{1 - k_i \; q'^{(d)}_x} + \frac{d_x}{q'^{(d)}_x} = 0 $$

Stime di massima verosimiglianza per dati raggruppati

Con riferimento alla classe di età $]x, x+1]$ nel caso di dati raggruppati, i dati sono:

n_x numero di individui osservati

d_x numero di decessi osservati

w_x numero di individui usciti per altra causa

Con riferimento all'i-esimo, $i = 1, \ldots, n_x$, si definisce il seguente n.a.

$$C^{(i)} = \begin{cases} 1 & \text{se l'individuo } i \text{ esce per morte} \\ 2 & \text{se l'individuo } i \text{ esce per altra causa} \\ 0 & \text{altrimenti} \end{cases} \qquad \begin{array}{l} P(C^{(i)} = 1) = q_x^{(d)} \\ P(C^{(i)} = 2) = q_x^{(w)} \\ P(C^{(i)} = 0) = 1 - q_x^{(d)} - q_x^{(w)} \end{array}$$

In ipotesi di indipendenza stocastica dei n.a. $C^{(i)}$ la verosimiglianza delle osservazioni è

$$L = \prod_{i \in D} P(C^{(i)} = 1) \cdot \prod_{i \in W} P(C^{(i)} = 2) \cdot \prod_{i \in S} P(C^{(i)} = 0) = \left(q_x^{(d)}\right)^{d_x} \left(q_x^{(w)}\right)^{w_x} \left(1 - q_x^{(d)} - q_x^{(w)}\right)^{n_x - d_x - w_x}$$

Dalla log-verosimiglianza

$$\log L = d_x \log\left(q_x^{(d)}\right) + w_x \log\left(q_x^{(w)}\right) + \left(n_x - d_x - w_x\right)\log\left(1 - q_x^{(d)} - q_x^{(w)}\right)$$

si ottiene il sistema di equazioni di verosimiglianza che fornisce le stime delle probabilità di eliminazione per causa di morte e, rispettivamente, per altra causa.

$$\begin{cases} \dfrac{d_x}{q_x^{(d)}} - \dfrac{n_x - d_x - w_x}{1 - q_x^{(d)} - q_x^{(w)}} = 0 \\ \dfrac{w_x}{q_x^{(w)}} - \dfrac{n_x - d_x - w_x}{1 - q_x^{(d)} - q_x^{(w)}} = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} \hat{q}_x^{(d)} = \dfrac{d_x}{n_x} \\ \hat{q}_x^{(w)} = \dfrac{w_x}{n_x} \end{cases}$$

Le stime coincidono con quelle ottenute con il metodo dei momenti nel caso particolare $r_i = 0$ e $s_i = 1$ per ogni $i = 1, 2, \dots, n_x$

Se si vogliono stimare le probabilità assolute, si devono formulare le opportune ipotesi.

In ipotesi di uscite non informative si ha

$$q_x^{(d)} = \int_0^1 {}_u p_x'^{(d)} \cdot {}_u p_x'^{(w)} \cdot \mu^{(d)}(x+u)du$$

$$q_x^{(w)} = \int_0^1 {}_u p_x'^{(d)} \cdot {}_u p_x'^{(w)} \cdot \mu^{(w)}(x+u)du$$

A) Nell'ipotesi di <u>distribuzione uniforme</u> per le probabilità assolute di uscita per morte e, rispettivamente, per altra causa si ha

$${}_u p_x'^{(d)} = 1 - u\, q_x'^{(d)} \qquad \mu^{(d)}(x+u) = \frac{q_x'^{(d)}}{1 - u\, q_x'^{(d)}}$$

$${}_u p_x'^{(w)} = 1 - u\, q_x'^{(w)} \qquad \mu^{(w)}(x+u) = \frac{q_x'^{(w)}}{1 - u\, q_x'^{(w)}}$$

e quindi

$$q_x^{(d)} = q_x'^{(d)}\left(1 - \frac{1}{2} q_x'^{(w)}\right) \qquad q_x^{(w)} = q_x'^{(w)}\left(1 - \frac{1}{2} q_x'^{(d)}\right)$$

La log-verosimiglianza diventa allora

$$\log L = d_x \log\left(q_x^{(d)}\right) + w_x \log\left(q_x^{(w)}\right) + (n_x - d_x - w_x)\log\left(1 - q_x^{(d)} - q_x^{(w)}\right)$$

$$= d_x \log\left(q_x'^{(d)}\right) + d_x \log\left(1 - \frac{1}{2} q_x'^{(w)}\right) + w_x \log\left(q_x'^{(w)}\right) + w_x \log\left(1 - \frac{1}{2} q_x'^{(d)}\right) +$$

$$+ (n_x - d_x - w_x)\log\left(1 - q_x'^{(d)}\left(1 - \frac{1}{2} q_x'^{(w)}\right) - q_x'^{(w)}\left(1 - \frac{1}{2} q_x'^{(d)}\right)\right)$$

Risolvendo il sistema di equazioni di verosimiglianza si ottengono le stesse stime ottenute con il metodo dei momenti nel caso particolare $r_i = 0$ e $s_i = 1$ per ogni $i = 1, 2, \ldots, n_x$.

$$\hat{q}_x'^{(d)} = \frac{b - \sqrt{b^2 - 2n_x d_x}}{n_x} \qquad \text{con} \qquad b = n_x + \frac{d_x}{2} - \frac{w_x}{2}$$

$$\hat{q}_x'^{(w)} = \frac{b - \sqrt{b^2 - 2n_x w_x}}{n_x} \qquad \text{con} \qquad b = n_x - \frac{d_x}{2} + \frac{w_x}{2}$$

A) Nell'ipotesi di <u>intensità di uscita per morte e per altra causa costanti</u> si ha

$$\mu^{(d)}(x+u)=\mu_x^{(d)} \qquad {}_u p_x'^{(d)} = \exp\left[-\mu_x^{(d)}u\right]$$

$$\mu^{(w)}(x+u)=\mu_x^{(w)} \qquad {}_u p_x'^{(w)} = \exp\left[-\mu_x^{(w)}u\right]$$

e quindi

$$q_x^{(d)} = \frac{\mu_x^{(d)}}{\mu_x^{(d)}+\mu_x^{(w)}}\left(1-\exp\left[-\left(\mu_x^{(d)}+\mu_x^{(w)}\right)\right]\right) \qquad q_x^{(w)} = \frac{\mu_x^{(w)}}{\mu_x^{(d)}+\mu_x^{(w)}}\left(1-\exp\left[-\left(\mu_x^{(d)}+\mu_x^{(w)}\right)\right]\right)$$

Sostituendo nella log-verosimiglianza

$$\log L = d_x \log\left(q_x^{(d)}\right)+ w_x \log\left(q_x^{(w)}\right)+(n_x-d_x-w_x)\log\left(1-q_x^{(d)}-q_x^{(w)}\right)$$

e derivando, si ottiene il sistema di equazioni di verosimiglianza le cui soluzioni coincidono con le stime ottenute con il metodo dei momenti nel caso particolare $r_i=0$ e $s_i=1$ per ogni $i=1, 2, \dots, n_x$.

$$\hat{\mu}_x^{(d)} = -\log\left(\frac{n_x-d_x-w_x}{n_x}\right)^{\frac{d_x}{d_x+w_x}} \qquad \hat{\mu}_x^{(w)} = -\log\left(\frac{n_x-d_x-w_x}{n_x}\right)^{\frac{w_x}{d_x+w_x}}$$

STIMATORE DELLA FUNZIONE DI SOPRAVVIVENZA E SUE PROPRIETÀ

Per stimare la funzione di sopravvivenza $S(x)$, $x = a, a+1, \dots, \omega$, di un modello di sopravvivenza non parametrico, si esprime la funzione di sopravvivenza come prodotto di probabilità condizionate di sopravvivenza

$$S(x) = \frac{S(x)}{S(x-1)} \frac{S(x-1)}{S(x-2)} \cdots \frac{S(1)}{S(0)} = p_{x-1} p_{x-2} \cdots p_0 = \prod_{j<x} p_j$$

infatti

$$\frac{S(j)}{S(j-1)} = \frac{P(T_0 > j)}{P(T_0 > j-1)} = \frac{P(T_0 > j, T_0 > j-1)}{P(T_0 > j-1)} = P(T_0 > j \mid T_0 > j-1) = p_{j-1}$$

Siano

$\hat{p}_x = 1 - \hat{q}_x$ la stima di p_x

n'_x l'esposizione nella classe di età $]x, x+1]$

con $x = a, a+1, \dots, \omega - 1$

Si ottiene la seguente stima della funzione di sopravvivenza $S(x)$, $x = a, a+1, \dots, \omega$

$$\hat{S}(x) = \hat{p}_{x-1} \hat{p}_{x-2} \cdots \hat{p}_0$$

Sia

$\tilde{p}_x$ lo stimatore di p_x del quale $\hat{p}_x$ è la stima, $x = a, a+1, \dots, \omega - 1$

Indichiamo con

$$\tilde{S}(x) = \prod_{j<x} \tilde{p}_j \,, \qquad x = a, a+1, \dots, \omega$$

lo stimatore del quale $\hat{S}(x)$ è la stima.

Per valutare speranza matematica e varianza dello stimatore $\tilde{S}(x)$ occorre formulare delle ipotesi sui n.a. $\tilde{p}_j$, $j = a, a+1, \dots, \omega - 1$

Siano $\mathcal{I} = \{n'_a, n'_{a+1}, , \dots, n'_{\omega-1}\}$ le esposizioni nelle diverse classi di età.

Si formulano le seguenti <u>ipotesi</u> sui n.a. $\tilde{p}_x$

Condizionatamente a $\mathcal{I} = \{n'_a, n'_{a+1}, , \dots, n'_{\omega-1}\}$ *, i n.a.*

$\tilde{p}_x$, $x = a, a+1, \dots, \omega - 1$ siano stocasticamente indipendenti

e siano

$$E(\tilde{p}_x | \mathcal{I}) = p_x \qquad Var(\tilde{p}_x | \mathcal{I}) = \frac{p_x\,(1 - p_x)}{n'_x} \qquad x = a, a+1, \dots, \omega - 1$$

Risulta allora che $\tilde{S}(x) = \prod_{j<x} \tilde{p}_j$ è uno stimatore non distorto, infatti

$$E\left(\tilde{S}(x)\middle|\mathcal{I}\right) = E\left(\prod_{j<x} \tilde{p}_j \middle| \mathcal{I}\right) = \prod_{j<x} p_j = S(x) \qquad\qquad x = a, a+1, \ldots, \omega$$

La varianza dello stimatore $\tilde{S}(x)$ è

$$Var\left(\tilde{S}(x)\middle|\mathcal{I}\right) = [S(x)]^2 \left[\prod_{j<x}\left(1 + \frac{q_j}{p_j n'_j}\right) - 1\right]$$

e può essere approssimata da

$$Var\left(\tilde{S}(x)\middle|\mathcal{I}\right) \cong [S(x)]^2 \sum_{j<x} \frac{q_j}{p_j n'_j}$$

dalla quale si ottiene la **formula di Greenwood**, che fornisce una stima della varianza dello stimatore $\tilde{S}(x)$

$$V\hat{a}r\left(\tilde{S}(x)\middle|\mathcal{I}\right) = \left[\hat{S}(x)\right]^2 \sum_{j<x} \frac{\hat{q}_j}{\hat{p}_j n'_j}$$

CONFRONTO DELLA TAVOLA DI SOPRAVVIVENZA STIMATA CON UNA TAVOLA STANDARD

Siano

$$\hat{q}_x, \qquad x = a, a+1, \dots, \omega - 1$$

le stime delle probabilità di morte q_x, $x = a, a+1, \dots, \omega - 1$, di un modello di sopravvivenza non parametrico.

Ci si pone il problema se il fenomeno della mortalità osservata nella collettività possa essere descritto da una tavola di mortalità proveniente da altre esperienze statistiche. Tale tavola viene allora detta **tavola "standard"** e la indichiamo con

$$q'_x, \qquad x = a, a+1, \dots, \omega - 1$$

Per verificare se la tavola standard accosta bene le osservazioni si sottopone a verifica d'ipotesi la seguente ipotesi nulla:

H_0 : $\quad q_x = q'_x \qquad x = a, a+1, \dots, \omega - 1$

Con riferimento alla classe di età $]x, x+1]$ siano

D_x	n.a. di decessi
d_x	numero di decessi osservati
E_x	il numero di esposti al rischio

Per costruire la funzione test formuliamo le seguenti ipotesi

$$E(D_x) = E_x\, q_x \qquad\qquad Var(D_x) = E_x\, q_x\,(1-q_x)$$

Sotto l'ipotesi nulla i n.a.

$$Z_x = \frac{D_x - E_x\, q'_x}{\sqrt{E_x\, q'_x(1-q'_x)}} \qquad\qquad x = a,\, a+1, \dots,\, \omega-1$$

hanno distribuzione approssimata $N(0,1)$. In ipotesi di indipendenza stocastica dei n.a. Z_x , $x = a,\, a+1, \dots,\, \omega-1$, sia ha che il n.a.

$$\sum_{x=a}^{\omega-1} (Z_x)^2$$

ha distribuzione approssimata chi-quadrato con n gradi di libertà, essendo n il numero di classi di età.

Fissato un livello di significatività α si determina $\chi^2_{n,1-\alpha}$ tale che $P\left(\chi^2_n > \chi^2_{n,1-\alpha}\right) = 1-\alpha$ si rifiuta l'ipotesi nulla se

$$\sum_{x=a}^{\omega-1} (z_x)^2 > \chi^2_{n,1-\alpha}$$

essendo $\sum_{x=a}^{\omega-1} (z_x)^2$ la determinazione osservata del n.a. $\sum_{x=a}^{\omega-1} (Z_x)^2$

Tale test potrebbe non rilevare un buon accostamento della tavola standard a i dati osservati, e quindi non fare rifiutare l'ipotesi nulla, nelle seguenti situazioni:

– esistenza di scostamenti eccessivamente elevati per alcune età, controbilanciati da scostamenti molto ridotti per altre età;

– numero eccessivo di scostamenti tutti dello stesso segno (conseguenza di una mortalità rilevata "uniformemente" maggiore o minore di quella attesa in base alla tavola standard);

– gruppi eccessivamente numerosi di età consecutive con scostamenti tutti dello stesso segno.

Per evidenziare tali problematiche si utilizzano altri test, quale per esempio il test delle deviazioni cumulate.

Sotto l'ipotesi nulla

H_0 : $\quad q_x = q'_x \qquad x = a, a+1, \dots, \omega - 1$

si ha

$$E\left(D_x - E_x\ q'_x\right) = 0 \qquad \text{per ogni } x = a, a+1, \dots, \omega - 1$$

Consideriamo le deviazioni cumulate nell'intervallo di età da x_1 a x_2

$$\sum_{x=x_1}^{x_2}\left(D_x - E_x\ q'_x\right)$$

In ipotesi di indipendenza stocastica dei n.a. D_x si ha

$$E\left[\sum_{x=x_1}^{x_2}\left(D_x - E_x\ q'_x\right)\right] = 0 \qquad Var\left[\sum_{x=x_1}^{x_2}\left(D_x - E_x\ q'_x\right)\right] = \sum_{x=x_1}^{x_2} Var\left(D_x\right) = \sum_{x=x_1}^{x_2} E_x\ q'_x (1 - q'_x)$$

La distribuzione della deviazione cumulata standardizzata

$$\frac{\sum_{x=x_1}^{x_2}\left(D_x - E_x\ q'_x\right)}{\sqrt{\sum_{x=x_1}^{x_2} E_x\ q'_x(1-q'_x)}}$$

può essere approssimata mediante una $N(0,1)$ e quindi, fissato un livello di significatività α si determina il quantile $z_{1-\alpha/2}$ della distribuzione normale standard e si rifiuta l'ipotesi nulla se

$$\left|\frac{\sum_{x=x_1}^{x_2}\left(d_x - E_x\ q'_x\right)}{\sqrt{\sum_{x=x_1}^{x_2} E_x\ q'_x(1-q'_x)}}\right| > z_{1-\alpha/2}$$

Tale analisi va ripetuta su diversi intervalli di età che evidenziano criticità nell'accostamento della tavola standard ai dati osservati.

PEREQUAZIONE MEDIANTE TAVOLE STANDARD

Se è stata rifiutata l'ipotesi nulla sulla bontà di accostamento della tavola standard ai dati, si può decidere di "adattare" la tavola standard ai dati osservati.

Si ipotizza quindi un legame funzionale tra le probabilità di morte della collettività in esame

$$q_x, \; x = a, a+1, \ldots, \omega - 1,$$

e le probabilità di morte riportate nella tavola standard

$$q'_x, \; x = a, a+1, \ldots, \omega - 1$$

Si assume che tale funzione

$$q_x = f(x, q'_x)$$

dipenda da alcuni parametri che devono essere stimati, per esempio con il metodo dei minimi quadrati pesati.

Siano

$$\hat{q}_x, \qquad x = a, a+1, \dots, \omega-1$$

le stime delle probabilità di morte q_x, $x = a, a+1, \dots, \omega-1$, di un modello di sopravvivenza non parametrico.

Dall'analisi grafica dei rapporti $\dfrac{\hat{q}_x}{q'_x}, \qquad x = a, a+1, \dots, \omega-1$

si individua un possibile legame funzionale $q_x = f\left(x, q'_x\right)$.

Per esempio, nel caso di andamento approssimativamente lineare si può ipotizzare

$$q_x = q'_x (a + b\,x)$$

oppure

$$q_x = a\,q'_x + b$$

Più in generale si può ipotizzare anche un legame con due tavole standard $\{q'_x\}$ e $\{q''_x\}$

$$q_x = a'\,q'_x + a''\,q''_x$$

In alternativa, Lidstone ha proposto di considerare per l'analisi grafica la seguente trasformazione

$$\log\left(\frac{p'_x}{\hat{p}_x}\right), \qquad\qquad x = a,\, a+1, \ldots,\, \omega - 1$$

con

$$p'_x = 1 - q'_x \qquad\qquad \hat{p}_x = 1 - \hat{q}_x$$

in quanto i $\dfrac{p'_x}{\hat{p}_x}$ presentano un andamento più regolare rispetto ai $\dfrac{\hat{q}_x}{q'_x}$

Se evidenziano un andamento approssimativamente costante si può ipotizzare

$$\log\left(\frac{p'_x}{p_x}\right) = c \quad \Rightarrow \quad q_x = 1 - \frac{p'_x}{e^c}$$

Dopo avere individuato la funzione $f(x;a,b,\ldots)$ che esprime il legame tra q_x e q'_x, per la stima dei parametri della funzione si può utilizzare, per esempio, il metodo dei minimi quadrati.

$$\min_{a,b,\ldots} F(a,b,\ldots) \qquad \text{con } F(a,b,\ldots)=\sum_{x=a}^{\omega-1} w_x [\hat{q}_x - f(x;a,b,\ldots)]^2$$

essendo

$w_x = 1$ nel caso di minimi quadrati non pesati

$w_x = \dfrac{n'_x}{\hat{q}_x}$ nel caso di minimi quadrati pesati

Si noti che nell'ipotesi $Var(D_x)=n'_x\, q_x\,(1-q_x)$ si ha

$$Var\left(\frac{D_x}{n'_x}\right)=\frac{q_x\,(1-q_x)}{n'_x}\cong\frac{q_x}{n'_x}$$

quindi il peso $w_x = \dfrac{n'_x}{\hat{q}_x}$ è approssimativamente pari al reciproco della varianza dello stimatore di q_x

Se la funzione $f(x;\, a, b, \ldots)$ è lineare, le stime dei minimi quadrati dei parametri si ottengono agevolmente risolvendo un sistema lineare.

Sia u_x tale che il legame funzionale tra q_x e q'_x sia espresso mediante la funzione lineare

$$f(x;\, a, b) = a + b\,x$$

Per esempio $u_x = \dfrac{q_x}{q'_x}$ nel caso in cui sia $q_x = q'_x (a + b\,x)$

Si ha allora

$$\min_{a,b} F(a,b) \qquad \text{con } F(a,b) = \sum_{x=a}^{\omega-1} w_x [\hat{u}_x - (a + b\,x)]^2$$

essendo $\hat{u}_x = \dfrac{\hat{q}_x}{q'_x}$

Si ottiene:

$$\begin{cases} \dfrac{\partial F}{\partial a} = 0 \\ \dfrac{\partial F}{\partial b} = 0 \end{cases} \Leftrightarrow \begin{cases} \sum\limits_{x=a}^{\omega-1} w_x [\hat{u}_x - (a + b\,x)] = 0 \\ \sum\limits_{x=a}^{\omega-1} w_x [\hat{u}_x - (a + b\,x)]\, x = 0 \end{cases}$$

PEREQUAZIONE CON LEGGI DI SOPRAVVIVENZA

Siano

$$\hat{q}_x, \qquad x = a, a+1, \dots, \omega - 1$$

le stime delle probabilità di morte q_x, $x = a, a+1, \dots, \omega - 1$, di un modello di sopravvivenza non parametrico ottenute secondo un approccio di stima di tipo non parametrico.

Tali stime presentano usualmente delle irregolarità spesso imputabili alla limitata numerosità della popolazione, in particolare in alcune classi di età.

Tali irregolarità possono essere rimosse mediante opportune procedure di perequazione.

Due obiettivi sono alla base della scelta di una procedura di perequazione:

- la <u>regolarità</u> (o smoothness) delle stime perequate al variare dell'età;
- l'<u>accostamento</u> (o goodness of fit) delle stime perequate alle stime originali.

La perequazione con leggi di sopravvivenza o perequazione analitica consiste nel sostituire alle stime iniziali le stime ottenute mediante un modello analitico di mortalità (per es. il modello di Gompertz).

Il procedimento di perequazione analitica si articola in due fasi:

1. verifica (mediante analisi grafica) della possibilità di accostamento fornita dalla legge di sopravvivenza considerata
2. stima dei parametri della legge di sopravvivenza scelta

Analisi grafica di modelli di sopravvivenza

Si devono individuare dei legami di tipo lineare, per esplorare mediante grafici le possibilità di accostamento del modello ai dati.

Modello di Gompertz

$$\mu(x) = \beta\, e^{\alpha x} \qquad \alpha > 0 \qquad \beta > 0 \qquad x > 0$$

Si ha

$$\log \mu(x) = \log \beta + \alpha\, x$$

Si considera allora il grafico dei punti

$$(x, \log \hat{m}_x) \qquad x = a,\, a+1, \ldots, \omega - 1$$

essendo $\hat{m}_x$ le stime delle intensità istantanee di mortalità ottenute in un approccio non parametrico;

se il grafico dei punti presenta un andamento approssimativamente lineare, il modello di Gompertz si presta a descrivere la mortalità nella collettività in esame.

Il coefficiente angolare e l'intercetta della retta interpolante forniscono delle stima preliminari per i parametri $\alpha > 0$ e $\beta > 0$

Un altro legame lineare può essere ottenuto considerando le probabilità di sopravvivenza p_x

Dalla

$$S(x) = \exp\left(\frac{\beta}{\alpha}\left(1 - e^{\alpha x}\right)\right) \qquad x \geq 0, \qquad \text{si ha} \qquad p_x = \frac{S(x+1)}{S(x)} = \exp\left(\frac{\beta}{\alpha}\left(1 - e^{\alpha}\right)e^{\alpha x}\right)$$

e quindi

$$\log(-\log p_x) = \log\left(\frac{\beta}{\alpha}\left(e^{\alpha} - 1\right)\right) + \alpha x$$

Si considera allora il grafico dei punti

$$\left(x, \log(-\log \hat{p}_x)\right) \qquad x = a, a+1, \ldots, \omega - 1$$

Il coefficiente angolare e l'intercetta della retta interpolante forniscono delle stima preliminari per i parametri $\alpha > 0$ e $\beta > 0$

Un altro grafico che può indicare se il modello di Gompertz si presta a descrivere la mortalità nella collettività in esame è il seguente

$$\left(x, \frac{\log \hat{p}_{x+1}}{\log \hat{p}_x} \right) \qquad x = a, a+1, \dots, \omega - 2$$

infatti

$$\frac{\log p_{x+1}}{\log p_x} = e^{\alpha}$$

quindi se i punti del grafico hanno un andamento approssimativamente costante, il modello di Gompertz potrebbe essere adatto.

Modello di Makeham

$$\mu(x) = \delta + \beta\, e^{\alpha x} \qquad \alpha > 0 \quad \beta > 0 \quad \delta > 0 \qquad x > 0$$

Si ha

$$\log(\mu(x+1) - \mu(x)) = \log\left(\beta\left(e^{\alpha} - 1\right)\right) + \alpha\, x$$

Se il grafico dei punti

$$\left(x, \log(\hat{m}_{x+1} - \hat{m}_x)\right) \qquad x = a, a+1, \dots, \omega - 1$$

presenta un andamento approssimativamente lineare, il modello di Makeham si presta a descrivere la mortalità nella collettività in esame.

Il coefficiente angolare e l'intercetta della retta interpolante forniscono delle stima preliminari $\hat{\alpha}$ e $\hat{\beta}$ per i parametri α e β,rispettivamente.

Per una stima preliminare di δ si può considerare una media delle quantità

$$\hat{m}_x - \hat{\beta}\, e^{\hat{\alpha} x} \qquad x = a, a+1, \dots, \omega - 1$$

Un altro legame lineare può essere ottenuto considerando le probabilità di sopravvivenza p_x

Dalla

$$S(x) = \exp\left(\frac{\beta}{\alpha}\left(1-e^{\alpha x}\right) - \delta x\right) \qquad x \geq 0,$$

si ha

$$p_x = \frac{S(x+1)}{S(x)} = \exp\left(\frac{\beta}{\alpha}\left(1-e^{\alpha}\right)e^{\alpha x} - \delta\right)$$

Indicato con $\Delta \log p_x = \log p_{x+1} - \log p_x$ si ha

$$\frac{\Delta \log p_{x+1}}{\Delta \log p_x} = e^{\alpha}$$

Se il grafico dei punti $\left(x, \frac{\Delta \log \hat{p}_{x+1}}{\Delta \log \hat{p}_x}\right)$ $x = a, a+1, \dots, \omega - 2$

presenta un andamento approssimativamente costante, il modello di Makehanm si presta a descrivere la mortalità nella collettività in esame.

Come stima preliminare di α si può considerare il logaritmo della media di valori

$$\frac{\Delta \log \hat{p}_{x+1}}{\Delta \log \hat{p}_x} \qquad x = a, a+1, \dots, \omega - 2$$

Dalla $$\Delta \log p_x = \log p_{x+1} - \log p_x = -\frac{\beta}{\alpha}\left(1 - e^{\alpha}\right)^2 e^{\alpha x}$$

si individua come stima preliminare di β la media dei seguenti valori

$$-\frac{\Delta \log \hat{p}_x \cdot \hat{\alpha}}{\left(1 - e^{\hat{\alpha}}\right)^2 e^{\hat{\alpha} x}} \qquad x = a, a+1, \dots, \omega - 1$$

Infine, dalla

$$\log p_x = \frac{\beta}{\alpha}\left(1 - e^{\alpha}\right) e^{\alpha x} - \delta$$

Si ottiene come stima preliminare di δ la media dei seguenti valori

$$\frac{\hat{\beta}}{\hat{\alpha}}\left(1 - e^{\hat{\alpha}}\right) e^{\hat{\alpha} x} - \log \hat{p}_x \qquad x = a, a+1, \dots, \omega - 1$$

ALTRE FORMULE DI PEREQUAZIONE UTILIZZATE IN AMBITO ATTUARIALE

Formula di Barnett

$$\frac{q_x}{1-q_x} = A + H\,x + B\,c^x \qquad\qquad A, H, B, c > 0$$

Le quantità

$$\frac{q_x}{1-q_x}$$

sono dette odds.

Formula di Wilkie

$$\frac{q_x}{1-q_x} = \exp\bigl(pol(x)\bigr)$$

dove $pol(x)$ è un polinomio in x, spesso lineare o di grado 2

Tali espressioni, che esprimono legami funzionali tra gli odds e le età, possono essere viste come formule perequative che costituiscono casi particolari della seguente espressione più generale:

<u>Formula Gompertz-Makeham di tipo</u> (r, s)

$$GM_{\boldsymbol{\alpha}}^{r,s}(x) = \sum_{i=1}^{r} \alpha_i \, x^{i-1} + \exp\left(\sum_{i=r+1}^{r+s} \alpha_i \, x^{i-r-1} \right)$$

dove r e s sono interi positivi

$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_r, \alpha_{r+1}, \ldots, \alpha_{r+s})$ è un vettore di coefficienti

Se $r = 0$ si ha solamente il termine esponenziale

$$GM_{\boldsymbol{\alpha}}^{0,s}(x) = \exp\left(\sum_{i=1}^{s} \alpha_i \, x^{i-1} \right)$$

Se $s = 0$ si ha solamente il termine polinomiale

$$GM_{\boldsymbol{\alpha}}^{r,0}(x) = \sum_{i=1}^{r} \alpha_i \, x^{i-1}$$

Se $(r,s)=(0,2)$ si ha $GM_{\boldsymbol{\alpha}}^{0,2}(x)=\exp(\alpha_1+\alpha_2\,x)$

e si trova quindi una formula di tipo Gompertz

$$GM_{\boldsymbol{\alpha}}^{0,2}(x)=\exp(\alpha_1+\alpha_2\,x)=e^{\alpha_1}e^{\alpha_2 x}=\beta\,e^{\alpha\,x}$$

Se $(r,s)=(1,2)$ si ha $GM_{\boldsymbol{\alpha}}^{1,2}(x)=\alpha_1+\exp(\alpha_2+\alpha_3\,x)$

e si trova quindi una formula di tipo Makeham

$$GM_{\boldsymbol{\alpha}}^{1,2}(x)=\alpha_1+\exp(\alpha_2+\alpha_3\,x)=\alpha_1+e^{\alpha_2}e^{\alpha_3 x}=\delta+\beta\,e^{\alpha\,x}$$

Se $(r,s)=(2,2)$ si ha $GM_{\boldsymbol{\alpha}}^{2,2}(x)=\alpha_1+\alpha_2\,x+\exp(\alpha_3+\alpha_4\,x)$

e si trova quindi una formula di tipo Barnett

$$GM_{\boldsymbol{\alpha}}^{2,2}(x)=\alpha_1+\alpha_2\,x+\exp(\alpha_3+\alpha_4\,x)=\alpha_1+\alpha_2\,x+e^{\alpha_3}e^{\alpha_4 x}=A+H\,x+BC^{\,x}$$

Se $(r,s)=(0,n)$ si ha $GM_{\boldsymbol{\alpha}}^{0,n}(x)=\exp\left(\sum_{i=1}^{n}\alpha_i\,x^{i-1}\right)$

e si trova quindi la formula di Wilkie.

STIMA DEI PARAMETRI DI UNA FORMULA DI PEREQUAZIONE

Dopo avere individuato una legge di sopravvivenza adatta a descrivere la mortalità nella collettività, oppure una formula adatta per perequare le stime iniziali

$\hat{q}_x$ oppure $\hat{m}_x$ $\qquad x = a, a+1, \dots, \omega-1$

si devono stimare i parametri.

<u>Metodo dei minimi quadrati</u>

$$\min_{\alpha,\beta,\dots} F(\alpha,\beta,\dots) \qquad \text{con } F(\alpha,\beta,\dots) = \sum_{x=a}^{\omega-1} w_x [\hat{u}_x - f(x;\alpha,\beta,\dots)]^2$$

essendo

$w_x = \dfrac{n'_x}{\hat{q}_x}$ nel caso di minimi quadrati pesati, con n'_x esposizione nella classe di età x

$\hat{u}_x$ una opportuna trasformazione dei $\hat{q}_x$ oppure degli $\hat{m}_x$ tale che la funzione f sia lineare nei parametri del modello; infatti se f è lineare, le stime dei minimi quadrati dei parametri si ottengono agevolmente risolvendo un sistema lineare.

Per esempio, nel caso del modello di Gompertz si ha

$$\log \mu(x) = \log \beta + \alpha \; x$$

Quindi si può considerare il seguente problema

$$\min_{\alpha, \beta} \sum_{x=a}^{\omega-1} w_x \left[\log \hat{m}_x - \log \beta - \alpha \left(x + \frac{1}{2} \right) \right]^2$$

Si noti che, poiché $\hat{m}_x$ ha il significato di stima dell'intensità istantanea di mortalità costante nella classe di età $]x, x+1]$, dovendo "attribuirla" ad una precisa età nella classe $]x, x+1]$ si considera l'età $x + \frac{1}{2}$

<u>Metodo della massima verosimiglianza</u>

Tratteremo la stima dei parametri mediante il metodo della massima verosimiglianza nell'ambito dei GLM.

PEREQUAZIONE MEDIANTE MODELLI LINEARI GENERALIZZATI

Siano

$\hat{q}_x$ oppure $\hat{m}_x$ $\quad x = a, a+1, \ldots, \omega - 1$

le stime iniziali di una tavola di sopravvivenza ottenute in un approccio di tipo non parametrico

n'_x l'esposizione (es. il numero iniziale di esposti al rischio) nella classe di età x

Definiamo dei GLM per perequare le stime iniziali.

Un GLM è definito dalle seguenti ipotesi:

- **ipotesi probabilistiche**: distribuzioni delle variabili risposta appartenenti alla famiglia esponenziale lineare
- **ipotesi strutturali**: struttura di regressione e funzione di collegamento

Illustriamo alcuni modelli probabilistici e le conseguenti ipotesi strutturali adatte per la perequazione delle stime iniziali.

Modelli con distribuzione binomiale scalata

La distribuzione Binomiale scalata è una distribuzione della famiglia esponenziale lineare. Infatti, se

$$X \approx B(n,\ p) \qquad \Rightarrow \qquad P(X = x) = \binom{n}{x} p^{x} (1 - p)^{n-x} \qquad \text{con } x = 0, 1, \ldots, n$$

si ha che il n.a. $Y = \dfrac{X}{n}$ ha distribuzione Binomiale scalata: $Y \approx B(n,\ p)/n$

$$P(Y = y) = \binom{n}{n\,y} p^{n\,y} (1 - p)^{n-ny} \qquad \text{con } y = 0, \frac{1}{n}, \ldots, 1$$

Poiché

$$P(Y = y) = \binom{n}{n\,y} \left(\frac{p}{1-p}\right)^{n\,y} (1 - p)^{n} = \binom{n}{n\,y} \exp\left\{ n \left[y \log\left(\frac{p}{1-p}\right) + \log(1 - p) \right] \right\}$$

è una distribuzione della famiglia esponenziale lineare con

parametro canonico $\theta = \log\left(\dfrac{p}{1-p}\right)$ funzione comulante $b(\theta) = \log\left(1 + e^{\vartheta}\right)$

peso $\omega = n$ parametro di dispersione $\phi = 1$

Consideriamo le osservazioni

$$y_x = \hat{q}_x$$

ed i pesi

$\omega_x = \lfloor n'_x \rfloor$ dati dalle esposizioni troncate

con $x = a, a+1, \dots, \omega - 1$.

Siano

Y_x i n.a. variabili risposta, $x = a, a+1, \dots, \omega - 1$

- **ipotesi probabilistiche**: Y_x stoc. indip. con distribuzione Binomiale scalata con

  pesi ω_x parametro di dispersione $\phi = 1$

  parametro canonico $\theta = \log\left(\dfrac{q_x}{1-q_x}\right)$ funzione comulante $b(\theta) = \log\left(1 + e^{\vartheta}\right)$

  Si ha allora:

$$E(Y_x) = b'(\theta) = \frac{e^{\vartheta}}{1+e^{\vartheta}} = q_x \qquad Var(Y_x) = \frac{1}{\omega_x} b''(\theta) = \frac{1}{\omega_x} q_x (1 - q_x)$$

- **ipotesi strutturali**

Funzione di collegamento $g(q_x) = \eta_x$ con g funzione monotona, derivabile e η_x previsore lineare

Funzione di collegamento canonica o *logit* o *log-odds*

$$g(q_x) = \log\left(\frac{q_x}{1-q_x}\right)$$

Funzione log-log complementare

$$g(q_x) = \log(-\log(1-q_x))$$

Funzione Probit

$g(q_x) = \Phi^{-1}(q_x)$ essendo Φ la funzione di ripartizione della distribuzione normale standard

Previsore lineare $\eta_x = \boldsymbol{z}'_x \boldsymbol{\beta}$ con $\boldsymbol{z}_x$ vettore delle determinazioni delle variabili esplicative relative alla classe di età x

Se si tiene conto soltanto dell'età, si ha usualmente: $\eta_x = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_m x^m$

Esempio: il modello di Gompertz

Abbiamo visto che per il modello di Gompertz si ha

$$\log(-\log p_x) = \log\left(\frac{\beta}{\alpha}\left(e^{\alpha}-1\right)\right) + \alpha\, x$$

Si può stimare tale modello con un GLM per le osservazioni $y_x = \hat{q}_x\,,\ x = a, a+1, \ldots, \omega-1$

Variabili risposta: Y_x con distribuzione Binomiale scalata con $E(Y_x) = q_x$ e pesi $\omega_x = \lfloor n'_x \rfloor$ dati dalle esposizioni troncate

Funzione di collegamento: log-log complementare $g(q_x) = \log(-\log(1-q_x))$

Previsore lineare: $\eta_x = \beta_0 + \beta_1\, x$ essendo

$$\begin{cases} \beta_0 = \log\left(\dfrac{\beta}{\alpha}\left(e^{\alpha}-1\right)\right) \\ \beta_1 = \alpha \end{cases}$$

Il modello può essere esteso considerando $\eta_x = \beta_0 + \beta_1\, x + \beta_2\, x^2 + \ldots + \beta_m\, x^m$

cioè una formula di perequazione del tipo: $GM_{\boldsymbol{\alpha}}^{r,0}(x) = \sum_{i=1}^{r} \alpha_i\, x^{i-1}$

Esempio: <u>il modello di Wilkie</u>

In tale modello si ipotizza $$\frac{q_x}{1-q_x} = \exp\left(pol(x)\right)$$

dove $pol(x)$ è un polinomio in x, spesso lineare o di grado 2

Si può stimare tale modello con un GLM per le osservazioni $y_x = \hat{q}_x$, $x = a, a+1, \ldots, \omega - 1$

<u>Variabili risposta</u>: Y_x con distribuzione Binomiale scalata con $E(Y_x) = q_x$ e pesi $\omega_x = \lfloor n'_x \rfloor$ dati dalle esposizioni troncate

<u>Funzione di collegamento</u>: logit $g(q_x) = \log\left(\frac{q_x}{1-q_x}\right)$

<u>Previsore lineare</u>: $\eta_x = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_m x^m$

Poiché

$$g(q_x) = \eta_x \Leftrightarrow \log\left(\frac{q_x}{1-q_x}\right) = \beta_0 + \beta_1 x + \ldots + \beta_m x^m \Leftrightarrow \frac{q_x}{1-q_x} = \exp\left(\beta_0 + \beta_1 x + \ldots + \beta_m x^m\right)$$

Si ha una formula di perequazione del tipo: $$GM_{\boldsymbol{\alpha}}^{0,s}(x) = \exp\left(\sum_{i=1}^{s} \alpha_i x^{i-1}\right)$$

Modelli con distribuzione di Poisson

Sia

$$Y \approx Poi(\mu) \qquad \Rightarrow \qquad P(Y=y)=\frac{\mu^y}{y!}e^{-\mu} \qquad \text{con } y=0,1,\dots$$

È una distribuzione della famiglia esponenziale lineare, infatti

$$P(Y=y)=\frac{\mu^y}{y!}e^{-\mu}=\frac{1}{y!}\exp\{y\log(\mu)-\mu\}$$

parametro canonico $\theta=\log(\mu)$ funzione comulante $b(\theta)=e^{\vartheta}$

peso $\omega=1$ parametro di dispersione $\phi=1$

Abbiamo visto che se D_x è n.a. dei decessi nella classe di età $]x,x+1]$, in ipotesi di distribuzione di Poisson di parametro $\mu_x^{(d)}E_x^C$

$$P(D_x=d_x)=\frac{\left(\mu_x^{(d)}E_x^C\right)^{d_x}}{d_x!}e^{-\mu_x^{(d)}E_x^C}=\frac{\left(E_x^C\right)^{d_x}}{d_x!}e^{-\mu_x^{(d)}E_x^C}\left(\mu_x^{(d)}\right)^{d_x}\propto L^{(d)}$$

essendo $L^{(d)}$ la funzione di verosimiglianza, con parametro l’intensità istantanea di mortalità e E_x^C il numero centrale di esposti al rischio.

Con riferimento alla classe di età $]x, x+1]$ siano

$$E_x^C = \sum_{i \in S}(s_i - r_i) + \sum_{i \in D}(t_i - r_i) + \sum_{i \in W}(k_i - r_i)$$

il numero centrale di esposti al rischio

D_x il n.a. dei decessi con distribuzione di Poisson di parametro $\mu_x E_x^C$

Si ha

$$P(D_x = d_x) = \frac{\left(\mu_x E_x^C\right)^{d_x}}{d_x!} e^{-\mu_x E_x^C} = \frac{1}{d_x!} \exp\left\{d_x \log\left(\mu_x E_x^C\right) - \mu_x E_x^C\right\}$$

$$= \frac{\left(E_x^C\right)^{d_x}}{d_x!} \exp\left\{d_x \log(\mu_x) - \mu_x E_x^C\right\} = \frac{\left(E_x^C\right)^{d_x}}{d_x!} \exp\left\{E_x^C\left[\frac{d_x}{E_x^C}\log(\mu_x) - \mu_x\right]\right\}$$

cioè una distribuzione della famiglia esponenziale lineare con

parametro canonico $\theta_x = \log(\mu_x)$ funzione comulante $b(\theta) = e^{\vartheta}$

peso $\omega = E_x^C$ parametro di dispersione $\phi = 1$

e con variabili risposta $\dfrac{d_x}{E_x^C}$

Consideriamo le osservazioni

$$y_x = \hat{m}_x = \frac{d_x}{E_x^C}$$

ed i pesi

$\omega_x = E_x^C$ numeri centrali di esposti al rischio

con $x = a, a+1, \dots, \omega - 1$.

Siano

Y_x i n.a. variabili risposta, $x = a, a+1, \dots, \omega - 1$

- **ipotesi probabilistiche**: Y_x stoc. indip. con distribuzione di Poisson con

  pesi $\omega_x = E_x^C$ parametro di dispersione $\phi = 1$

  parametro canonico $\theta_x = \log(\mu_x)$ funzione comulante $b(\theta) = e^{\vartheta}$

  Si ha allora:

$$E(Y_x) = b'(\theta_x) = e^{\vartheta_x} = \mu_x \qquad Var(Y_x) = \frac{1}{E_x^C} b''(\theta_x) = \frac{1}{E_x^C} e^{\vartheta_x} = \frac{\mu_x}{E_x^C}$$

- **ipotesi strutturali**

<u>Funzione di collegamento</u> $g(\mu_x) = \eta_x$ con g funzione monotona, derivabile e η_x previsore lineare

Funzione di collegamento canonica logaritmo

$$g(\mu_x) = \log(\mu_x)$$

<u>Previsore lineare</u> $\eta_x = \mathbf{z}'_x \boldsymbol{\beta}$ con $\mathbf{z}_x$ vettore delle determinazioni delle variabili esplicative relative alla classe di età x

Se si tiene conto soltanto dell'età, si ha usualmente: $\eta_x = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_m x^m$

Esempio: il modello di Gompertz

Abbiamo visto che per il modello di Gompertz si ha $\log \mu(x) = \log \beta + \alpha\, x$

Si può stimare tale modello con un GLM per le osservazioni

$$y_x = \hat{m}_x = \frac{d_x}{E_x^C}, \qquad x = a, a+1, \dots, \omega - 1$$

Variabili risposta: Y_x con distribuzione di Poisson con

$E(Y_x) = \mu_x$ e pesi $\omega_x = E_x^C$ i numeri centrali di esposti al rischio

Funzione di collegamento: logaritmo $g(\mu_x) = \log(\mu_x)$

Previsore lineare: $\eta_x = \beta_0 + \beta_1\, x$ essendo

$$\begin{cases} \beta_0 = \log(\beta\ ) \\ \beta_1 = \alpha \end{cases}$$

Il modello può essere esteso considerando $\eta_x = \beta_0 + \beta_1\, x + \beta_2\, x^2 + \dots + \beta_m\, x^m$

cioè una formula di perequazione del tipo: $GM_{\alpha}^{r,0}(x) = \sum_{i=1}^{r} \alpha_i\, x^{i-1}$

RIFERIMENTI BIBLIOGRAFICI

D. London, Survival models and their estimation, Actex publications, 1997 (Cap. 5, 6, 7, 9)

E. Pitacco, Matematica e tecnica attuariale delle assicurazioni sulla durata di vita, Lint, 2002 (App. A)

Forfar, D.O. et al. (1988), On graduation by mathematical formula, JIA, 115, 41-97

Renshaw, A.E. (1991), Actuarial graduation practice and generalized linear and non-linear models, JIA, 118, 295-312