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

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

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>

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

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

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

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

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, $i = 1, \ldots, n_x$, 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 $]x, x+1]$ per scrivere la verosimiglianza delle osservazioni

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

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

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

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

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

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

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

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

Osservazione

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 interpolazione lineare

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

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

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

Osservazione

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