Stima con il metodo della massima verosimiglianza

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

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

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

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

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

Nota: $T^{(i)}$ ha determinazioni $]0, s_i - r_i]$

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

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

Nell'ipotesi che per ogni individuo i la durata aleatoria di permanenza nella collettività sia descritta dallo stesso modello di sopravvivenza, si ha

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

dove $P\left(T^{(i)}_{x+r_i} \leq t\right) = {}_tq^{(\tau)}_{x+r_i}$.

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*

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

Se l'individuo i è presente nella collettività all'età di uscita pianificata (quindi $i \epsilon S$):

$$T^{(i)} = s_i - r_i \qquad P\left(T^{(i)} = s_i - r_i\right) = {}_{s_i - r_i}p_{x+r_i}^{(\tau)} = {}_{s_i - r_i}p_{x+r_i}^{\prime(d)} \cdot {}_{s_i - r_i}p_{x+r_i}^{\prime(w)}$$

Se l'individuo i esce per morte all'età esatta $x + t_i$ (quindi $i \epsilon D$):

$$T^{(i)} = t_i - r_i \qquad f_{T,C}\left(t_i - r_i, 1\right) = {}_{t_i - r_i}p_{x+r_i}^{(\tau)} \cdot \mu^{(d)}\left(x + t_i\right) = {}_{t_i - r_i}p_{x+r_i}^{\prime(d)} \cdot {}_{t_i - r_i}p_{x+r_i}^{\prime(w)} \cdot \mu^{(d)}\left(x + t_i\right)$$

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

$$T^{(i)} = k_i - r_i \qquad f_{T,C}\left(k_i - r_i, 2\right) = {}_{k_i - r_i}p_{x+r_i}^{(\tau)} \cdot \mu^{(w)}\left(x + k_i\right) = {}_{k_i - r_i}p_{x+r_i}^{\prime(d)} \cdot {}_{k_i - r_i}p_{x+r_i}^{\prime(w)} \cdot \mu^{(w)}\left(x + k_i\right)$$

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

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

Si ha

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

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

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

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

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

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

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**

Osservazione: 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 $\overset{o}{m}_x = \dfrac{\theta_x}{E_x^C}$

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

Osservazione

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

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

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

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 individuo, $i = 1, \ldots, n_x$, si definisce il seguente n.a.

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

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 $i = 1, \ldots, n_x$

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 distribuzione uniforme 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 $i = 1, \ldots, n_x$.

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

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

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

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

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 $i = 1,\dots,n_x$.

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