STIMA DI MODELLI DI SOPRAVVIVENZA PARAMETRICI

USCITE PER MORTE E PER ALTRA CAUSA

Supponiamo di avere osservato n individui e di disporre di dati individuali esatti, riassunti per ogni individuo i dal vettore delle età

$$(y_i, z_i, \theta_i, \phi_i,) \qquad i = 1, \ldots, n$$

essendo

y_i l'età esatta di ingresso in osservazione

z_i l'età esatta 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)

Obiettivo: stimare la funzione di sopravvivenza di un **modello di sopravvivenza parametrico**

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

Sia

T_x durata di permanenza nella collettività per un individuo presente 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

Siano

$S^{(d)}(t) = P(T_0^{(d)} > t)$ funzione di sopravvivenza (parametrica) relativa al n.a. $T_0^{(d)}$

$S^{(w)}(t) = P(T_0^{(w)} > t)$ funzione di sopravvivenza (parametrica) relativa al n.a. $T_0^{(w)}$

Ipotesi: **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 intensità marginali di eliminazione 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 intensità di uscita per le varie cause, 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 la **relazione di Karup**

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

essendo

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

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

Stima con il metodo della massima verosimiglianza

Per scrivere la verosimiglianza delle osservazioni

$$(y_i, z_i, \theta_i, \phi_i,) \qquad\qquad i = 1, \ldots, n$$

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

$T^{(i)}$ durata aleatoria di osservazione dell'individuo i nella collettività

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

<u>Nota</u>: $T^{(i)}$ ha determinazioni $\left]0, z_i - y_i\right]$

Indicato con $T_{y_i}^{(i)}$ la durata di permanenza nella collettività per l'i-esimo individuo, si ha

$$T^{(i)} = min\left(T_{y_i}^{(i)}, z_i - y_i\right)$$

Nell'ipotesi che per ogni individuo i le durate aleatorie di permanenza nella collettività, all'età minima possibile per l'ingresso in assicurazione, siano ugualmente distribuite, si ha

$$P\left(T^{(i)} \leq t\right) = \begin{cases} 0 & t \leq 0 \\ P\left(T_{y_i}^{(i)} \leq t\right) & 0 < t < z_i - y_i \\ 1 & t \geq z_i - y_i \end{cases}$$

dove $P\left(T_{y_i}^{(i)} \leq t\right) = {}_t q_{y_i}^{(\tau)}$.

Si definiscono

$S = \{i \mid \text{l'individuo } i \text{ è presente nella collettività all'età } z_i\}$ *survival*

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

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

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

$$T^{(i)} = z_i - y_i \qquad P(T^{(i)} = z_i - y_i) = {}_{z_i - y_i}p_{y_i}^{(\tau)} = {}_{z_i - y_i}p_{y_i}^{\prime(d)} \cdot {}_{z_i - y_i}p_{y_i}^{\prime(w)} = \frac{S^{(d)}(z_i)}{S^{(d)}(y_i)} \cdot \frac{S^{(w)}(z_i)}{S^{(w)}(y_i)}$$

Se l'individuo i esce per morte all'età esatta θ_i:

$$T^{(i)} = \theta_i - y_i \qquad f_{T,C}(\theta_i - y_i, 1) = {}_{\theta_i - y_i}p_{y_i}^{(\tau)} \cdot \mu^{(d)}(\theta_i) = \frac{S^{(d)}(\theta_i)}{S^{(d)}(y_i)} \cdot \frac{S^{(w)}(\theta_i)}{S^{(w)}(y_i)} \cdot \mu^{(d)}(\theta_i)$$

Se l'individuo i esce per altra causa all'età esatta ϕ_i:

$$T^{(i)} = \phi_i - y_i \qquad f_{T,C}(\phi_i - y_i, 2) = {}_{\phi_i - y_i}p_{y_i}^{(\tau)} \cdot \mu^{(w)}(\phi_i) = \frac{S^{(d)}(\phi_i)}{S^{(d)}(y_i)} \cdot \frac{S^{(w)}(\phi_i)}{S^{(w)}(y_i)} \cdot \mu^{(w)}(\phi_i)$$

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

$$L = \prod_{i \in S} P(T^{(i)} = z_i - y_i) \cdot \prod_{i \in D} f_{T,C}\left(\theta_i - y_i, 1\right) \cdot \prod_{i \in W} f_{T,C}\left(\phi_i - y_i, 2\right)$$

$$= \prod_{i \in S} {}_{z_i - y_i} p_{y_i}^{(\tau)} \cdot \prod_{i \in D} {}_{\theta_i - y_i} p_{y_i}^{(\tau)} \cdot \mu^{(d)}\left(\theta_i\right) \cdot \prod_{i \in W} {}_{\phi_i - y_i} p_{y_i}^{(\tau)} \cdot \mu^{(w)}\left(\phi_i\right)$$

$$= \prod_{i \in S} \frac{S^{(d)}(z_i)}{S^{(d)}(y_i)} \cdot \frac{S^{(w)}(z_i)}{S^{(w)}(y_i)} \cdot$$

$$\cdot \prod_{i \in D} \frac{S^{(d)}(\theta_i)}{S^{(d)}(y_i)} \cdot \frac{S^{(w)}(\theta_i)}{S^{(w)}(y_i)} \cdot \mu^{(d)}\left(\theta_i\right) \cdot$$

$$\cdot \prod_{i \in W} \frac{S^{(d)}(\phi_i)}{S^{(d)}(y_i)} \cdot \frac{S^{(w)}(\phi_i)}{S^{(w)}(y_i)} \cdot \mu^{(w)}\left(\phi_i\right)$$

Poiché

$$S^{(d)}(\theta_i)\cdot\mu^{(d)}(\theta_i)=f^{(d)}(\theta_i) \qquad\qquad S^{(w)}(\phi_i)\cdot\mu^{(w)}(\phi_i)=f^{(w)}(\phi_i)$$

si ha

$$L=\prod_{i\in S}\frac{S^{(d)}(z_i)}{S^{(d)}(y_i)}\cdot\prod_{i\in D}\frac{S^{(d)}(\theta_i)}{S^{(d)}(y_i)}\cdot\mu^{(d)}(\theta_i)\cdot\prod_{i\in W}\frac{S^{(d)}(\phi_i)}{S^{(d)}(y_i)}\cdot$$

$$\cdot\prod_{i\in S}\frac{S^{(w)}(z_i)}{S^{(w)}(y_i)}\cdot\prod_{i\in D}\frac{S^{(w)}(\theta_i)}{S^{(w)}(y_i)}\cdot\prod_{i\in W}\frac{S^{(w)}(\phi_i)}{S^{(w)}(y_i)}\cdot\mu^{(w)}(\phi_i)$$

$$=\prod_{i\in S}\frac{S^{(d)}(z_i)}{S^{(d)}(y_i)}\cdot\prod_{i\in D}\frac{f^{(d)}(\theta_i)}{S^{(d)}(y_i)}\cdot\prod_{i\in W}\frac{S^{(d)}(\phi_i)}{S^{(d)}(y_i)}\cdot$$

$$\cdot\prod_{i\in S}\frac{S^{(w)}(z_i)}{S^{(w)}(y_i)}\cdot\prod_{i\in D}\frac{S^{(w)}(\theta_i)}{S^{(w)}(y_i)}\cdot\prod_{i\in W}\frac{f^{(w)}(\phi_i)}{S^{(w)}(y_i)}$$

Quindi

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

con

$$L^{(d)} = \prod_{i \in S} \frac{S^{(d)}(z_i)}{S^{(d)}(y_i)} \cdot \prod_{i \in D} \frac{f^{(d)}(\theta_i)}{S^{(d)}(y_i)} \cdot \prod_{i \in W} \frac{S^{(d)}(\phi_i)}{S^{(d)}(y_i)}$$

e

$$L^{(w)} = \prod_{i \in S} \frac{S^{(w)}(z_i)}{S^{(w)}(y_i)} \cdot \prod_{i \in D} \frac{S^{(w)}(\theta_i)}{S^{(w)}(y_i)} \cdot \prod_{i \in W} \frac{f^{(w)}(\phi_i)}{S^{(w)}(y_i)}$$

Quindi per stimare la funzione di sopravvivenza $S^{(d)}(t)$ si porrà

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

mentre per stimare la funzione di sopravvivenza $S^{(w)}(t)$ si porrà

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

Stima di massima verosimiglianza della funzione di sopravvivenza $S^{(d)}(t)$

$$L^{(d)} = \prod_{i \in S} \frac{S^{(d)}(z_i)}{S^{(d)}(y_i)} \cdot \prod_{i \in W} \frac{S^{(d)}(\phi_i)}{S^{(d)}(y_i)} \cdot \prod_{i \in D} \frac{f^{(d)}(\theta_i)}{S^{(d)}(y_i)}$$

dove

$$f^{(d)}(t) = -\frac{d}{d\,t} S^{(d)}(t)$$

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

Se il modello di sopravvivenza parametrico per la durata aleatoria di vita $T_0^{(d)}$ è assegnato mediante l'intensità istantanea di mortalità $\mu^{(d)}(t)$ la verosimiglianza può essere scritta nel modo seguente

$$L^{(d)} = \prod_{i \in S} \exp\left(-\int_{y_i}^{z_i} \mu^{(d)}(t) dt \right) \cdot \prod_{i \in W} \exp\left(-\int_{y_i}^{\phi_i} \mu^{(d)}(t) dt \right) \cdot \prod_{i \in D} \left[\exp\left(-\int_{y_i}^{\theta_i} \mu^{(d)}(t) dt \right) \mu^{(d)}(\theta_i) \right]$$

Con riferimento all'i-esimo individuo osservato, indichiamo con t_i la determinazione del n.a. $T^{(i)}$, si ha

$$t_i = \begin{cases} z_i - y_i & \text{se } i \in S \\ \phi_i - y_i & \text{se } i \in W \\ \theta_i - y_i & \text{se } i \in D \end{cases}$$

Consideriamo inoltre l'indicatore di evento

$$d_i = \begin{cases} 0 & \text{se } i \in S \cup W \\ 1 & \text{se } i \in D \end{cases}$$

la verosimiglianza $L^{(d)}$ relativa agli n individui osservati può allora essere scritta come segue

$$L^{(d)} = \prod_{i=1}^{n} \left[\exp\left(-\int_0^{t_i} \mu^{(d)}(y_i + s)ds \right) \left(\mu^{(d)}(y_i + t_i) \right)^{d_i} \right]$$

Indicata con

$$H_y(t) = \int_0^t \mu^{(d)}(y+s)ds$$

la funzione di rischio integrata, la verosimiglianza $L^{(d)}$ diventa

$$L^{(d)} = \prod_{i=1}^{n} \left[\exp\left(-H_{y_i}(t_i)\right) \left(\mu^{(d)}(y_i + t_i)\right)^{d_i} \right]$$

e quindi si ottiene la seguente espressione per la log-verosimiglianza

$$\log L^{(d)} = \sum_{i=1}^{n} -H_{y_i}(t_i) + \sum_{i=1}^{n} d_i \log\left(\mu^{(d)}(y_i + t_i)\right)$$

Per ottenere quindi la stima di massima verosimiglianza di un modello di sopravvivenza parametrico, disponendo di dati individuali esatti, è sufficiente specificare la struttura della intensità istantanea di mortalità $\mu^{(d)}(t)$ e della relativa funzione di rischio integrato $H_y(t)$

S.J. Richards (2012), "A handbook of parametric survival models for actuarial use", Scandinavian Actuarial Journal

V. Table 1. Some actuarial mortality laws and their corresponding integrated hazard functions.