STIMATORE DELLA FUNZIONE DI SOPRAVVIVENZA E SUE PROPRIETÀ

Per stimare la funzione di sopravvivenza $S(x)$, $x = 0, 1, \dots, \omega$, di un modello di sopravvivenza non parametrico, si esprime la funzione di sopravvivenza come prodotto di probabilità condizionate di sopravvivenza

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

Infatti, indicata con T_0 la durata aleatoria di vita per un individuo alla nascita si ha:

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

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 $x = 0, 1, \dots, \omega - 1$

Si ottiene la seguente stima della funzione di sopravvivenza $S(x)$, $x = 0, 1, \dots, \omega$

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

Sia

$\tilde{p}_x$ lo stimatore di p_x del quale $\hat{p}_x$ è la stima, $x = 0, 1, \dots, \omega - 1$

Indichiamo con

$$\tilde{S}(x) = \prod_{j<x} \tilde{p}_j , \qquad x = 0, 1, \dots, \omega$$

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. $\tilde{p}_j$, $j = 0, 1, \dots, \omega - 1$

Siano $\boldsymbol{\mathcal{I}} = \{n'_0, n'_1, \dots, n'_{\omega-1}, \}$ le esposizioni nelle diverse classi di età.

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

Condizionatamente a $\boldsymbol{\mathcal{I}} = \{n'_0, n'_1, \dots, n'_{\omega-1}, \}$ *, i n.a.*

$\tilde{p}_x$, $x = 0, 1, \dots, \omega - 1$ siano stocasticamente indipendenti

e siano

$$E(\tilde{p}_x | \mathcal{I}) = p_x \qquad Var(\tilde{p}_x | \mathcal{I}) = \frac{p_x (1 - p_x)}{n'_x} \qquad x = 0, 1, \dots, \omega - 1$$

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

$$E\left(\tilde{S}(x) \middle| \mathcal{I}\right) = E\left(\prod_{j<x} \tilde{p}_j \middle| \mathcal{I} \right) = \prod_{j<x} p_j = S(x) \qquad x = 0, 1, \dots, \omega - 1$$

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 $\tilde{S}(x)$

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

Osservazione.

Per stimare la funzione di sopravvivenza della durata aleatoria di vita da una età minima $a > 0$

$$ {}_tP_a = P(T_a > t) \qquad t = 0, 1, \dots, \omega - a, $$

essendo $T_a = (T_0 - a) | T_0 > a$ la durata aleatoria di vita per un individuo di età a si ha:

$$ {}_tP_a = P(T_a > t) = P(T_0 - a > t | T_0 > a) = \frac{P(T_0 > a + t)}{P(T_0 > a)} = \frac{S(a + t)}{S(a)} $$

Poiché

$$ {}_tP_a = \frac{S(a + t)}{S(a + t - 1)} \cdot \frac{S(a + t - 1)}{S(a + t - 2)} \cdot \ \dots \ \cdot \frac{S(a + 1)}{S(a)} = p_{a+t-1} \cdot p_{a+t-2} \cdot \ \cdots \ \cdot p_a = \prod_{j=0}^{t-1} p_{a+j} $$

si esprime la funzione di sopravvivenza ${}_tP_a = P(T_a > t)$ come prodotto di probabilità condizionate di sopravvivenza e, analogamente a quanto visto per la funzione di sopravvivenza $S(x)$, è possibile valutare la varianza del relativo stimatore.