PEREQUAZIONE MEDIANTE MODELLI LINEARI GENERALIZZATI

Siano

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

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

n'_x $\qquad$ 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

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

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

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

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**

<u>Funzione di collegamento</u> $g(q_x) = \eta_x$ 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

<u>Previsore lineare</u> $\eta_x = \boldsymbol{z}'_x \boldsymbol{\beta}$ 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: <u>il modello di Gompertz</u>

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

<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>: log-log complementare $g(q_x) = \log(-\log(1 - q_x))$

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

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

Esempio: il modello di Wilkie

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

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

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

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

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

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

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