GENERALIZED LINEAR MODELS

- Estendono i mod di regressione ordinari contemplando risposte non normali e modellizzando funzioni del valore atteso
- Sono caratterizzati da 3 componenti

1. Componente Casuale (Random Component)
   Consiste di una v.a. Y, di cui si osservano N realizzazioni $\{y_1 \dots y_N\}$, la cui distribuz rientra nella fam esponenziale

$$f_Y(y_i ; \theta_i) = a(\theta_i)\, b(y_i) \exp\{ y_i\, Q(\theta_i) \}$$

parametro della distrib
Può variare da $y_1 \dots y_N$
$\Longrightarrow$ non identicam distrib

PARAMETRO NATURALE

2. Componente Sistematica (Systematic Component)
   Mette in relazione un vettore $(\eta_1 \dots \eta_N)$ con le variabili esplicative del modello. Detta x_{ij} il valore del j-esimo predittore dell'unità i-esima, allora

$$\eta_{ij} = \underbrace{\sum_j \beta_j x_{ij}}_{\text{PREDITTORE LINEARE}} \qquad i = 1 \dots N$$

3. LINK FUNCTION
   Connette componente sistematica e componente casuale. Detta $\mu_i = E[Y_i]$, $i = 1 \dots N$, il modello prevede che

$$\eta_i = g(\mu_i) \quad \Longrightarrow \quad g(\mu_i) = \sum_j \beta_j x_{ij}$$

monotona e differenziabile

classico lm

Oss: $g(\mu) = \mu$ IDENTITY LINK. Modello direttamente per la media

(Oss:) La link function che trasforma la media nel param naturale è detta LINK FUNCTION CANONICA

$$g(\mu_i) = Q(\theta_i) = \sum_j \beta_j x_{ij}$$

$\Longrightarrow$ Un GLM è un modello lineare per una trasformaz della media della var risp la cui distribuz ricade nella fam esponenziale

Es 1: Binomial Logit Model for BINARY DATA

Binary responses of the form $y = \begin{cases} 0 & (1-\pi) & \mathbb{P}(Y=0) \\ 1 & \pi & \mathbb{P}(Y=1) \end{cases}$

$\Rightarrow \mathbb{E}[y] = \mathbb{P}(Y=1) = \pi$ $\qquad$ successo (esempi)

Caso particolare di $Y \sim Bi(m, \pi)$ con $m = 1 \Rightarrow Y \sim Be(\pi)$

~ Componente casuale

$$f_Y(y;\pi) = \pi^y (1-\pi)^{1-y} = (1-\pi)\left[\frac{\pi}{1-\pi}\right]^y \mathbb{1}_{\{0,1\}}(y)$$

$$= (1-\pi)\exp\left\{ y \log\left(\frac{\pi}{1-\pi}\right)\right\} \mathbb{1}_{\{0,1\}}(y)$$

$\Rightarrow Y$ è fam esponenziale in cui

$\theta = \pi$

$a(\theta) = 1-\pi$

$b(y) = 1$

$Q(\theta) = \log\left(\frac{\pi}{1-\pi}\right) = \text{logit}(\pi)$ $\quad$ LINK CANONICO

Es 2: Poisson Loglinear Model for COUNT DATA

La distribuzione più comunemente usata per modellizzare i conteggi è la Poisson

$$Y \sim P(\mu) \implies \mathbb{E}[Y] = \mu$$

~ Componente Casuale

$$f_Y(y;\mu) = \frac{e^{-\mu}\mu^y}{y!} = \exp\{-\mu\} \cdot \frac{1}{y!} \exp\{y \log(\mu)\} \, \mathbb{1}_{\mathbb{N} \cup \{0\}}(y)$$

$\Rightarrow Y$ è fam esponenziale in cui

$\theta = \mu$

$a(\theta) = e^{-\mu}$

$b(y) = 1/y!$

$Q(\theta) = \log(\mu)$ $\quad$ LINK CANONICO

Es 3: GLM for CONTINUOUS RESPONSES

Chiaramente i mod generalizzati includono quelli tradiz come caso particolare

$$f_Y = \mathcal{N} \qquad Q(\mathbb{E}[Y]) = \mathbb{E}[Y] \quad (\text{link identità})$$

Ma poi ci sono tutte le altre distribuz continue $(\mathcal{G}, \mathcal{E}, \ldots)$

(Oss:) Di solito i modelli lineari trasformano Y in modo che sia distribuito $\mathcal{N}$ e con var costante, quindi ottengono le stime via LS. Nei GLM la scelta di random component e link function sono separate. Non è necessario che il link stabilizzi la varianza o produca normalità.

DEVIANZA

Per un GLM e un insieme di osservaz $\{y_1 \cdots y_N\}$ sia

$\ell(\underline{\mu}, \underline{y})$ = log verosimiglianza espressa in termini di media

$\Rightarrow \ell(\hat{\underline{\mu}}, \underline{y})$ è la log verosim massimizzata

$\ell(\underline{y}, \underline{y})$ è la log verosim massima ottenibile considerando tutti i possibili modelli (è quella del mod + generale che ha un param per ogni oss e produce il fit perfetto (interpolato) per cui $\hat{\mu} = y$)
Inutile in quanto non produce alcuna sintesi dei dati xò serve come baseline comparison

MODELLO SATURATO

$$\text{DEVIANZA} = -2\left[\ell(\hat{\underline{\mu}}, \underline{y}) - \ell(\underline{y}, \underline{y})\right] \underset{asynt}{\sim} \chi^2_{(N-p)}$$

← statistica derivante dal rapporto di verosimiglianze

p = # of model param

$\hookrightarrow \lambda = -2 \log\left(\frac{Lik\,1}{Lik\,2}\right) = -2(\log L_1 - \log L_2)$

Es: $Y_i \sim Bi(m_i, \pi_i)$, $i = 1 \dots N$
Immaginiamo di voler testare che $\pi_i = \alpha \; \forall i$, avendo un mod con un solo parametro anziché N — testare la bontà di

$\Rightarrow$ DEVIANZA $\sim \chi^2_{(N-1)}$

L'anal delle devianza è una generalizzaz dell'ANOVA per variab risposta non Normali

GLM per DATI BINARI

Sia $Y \sim Be(\pi(\underline{x})) = \begin{cases} 0 & \mathbb{P}(Y=0) = (1-\pi(\underline{x})) \\ 1 & \mathbb{P}(Y=1) = \pi(\underline{x}) \end{cases}$

per sottolineare dipendenza con

Si ha pertanto che $E[Y] = \pi(\underline{x})$ $\quad V[Y] = \pi(\underline{x})(1-\pi(\underline{x}))$

Una prima idea per modellizzare Y attraverso $\pi(\underline{x})$ è l'utilizzo di un mod lineare della forma

$\pi(\underline{x}) = \alpha + \beta x \longrightarrow$ GLM con compon casuale Binomiale e link identità

PB 1: così facendo nulla garantisce che $\pi \in [0,1]$ come dovrebbe
PB 2: se volessimo stime LS dovremmo partire da v.a. $\mathcal{N}$ (qui $Y \in \{0,1\}$) e con varianza costante (qui la varianza è tanto minore quanto più $x \to 0,1$)

$\Longrightarrow$ una relazione non lineare risponde meglio alle esigenze. In particolare, qualcosa che abbia minore impatto se $\pi \to 0,1$ che se $\pi \sim 0,5$. Storicamente la forma più usata è la logistica

$$\pi(\underline{x}) = \frac{\exp\{\alpha+\beta x\}}{1+\exp\{\alpha+\beta x\}}$$

se $\beta > 0 \Longrightarrow \pi(x) \to 1$ se $x \to \infty$
se $\beta < 0 \Longrightarrow \pi(x) \to 0$ se $x \to \infty$

$1-\pi = 1 - \frac{\exp}{1+\exp} = \frac{1+\exp-\exp}{1+\exp}$ $\quad \Rightarrow \frac{\pi}{1-\pi} = \frac{\exp}{1+\exp} \Big/ \frac{1}{1+\exp} = \exp$

ODD
rapporto tra prob successo e insuccesso

$$\frac{\pi(\underline{x})}{1-\pi(\underline{x})} = \exp\{\alpha+\beta x\} \Longrightarrow \log\left(\frac{\pi(\underline{x})}{1-\pi(\underline{x})}\right) = \alpha+\beta x$$

PARAM NATURALE

$$\Longrightarrow \text{logit}(\pi(\underline{x})) = \alpha + \beta x \qquad g(\mu_i) = \eta_i = \sum_j \beta_j x_{ij}$$

Interpretazione dei parametri $\underline{\beta}$

Il segno di β determina se $\pi(\underline{x})$ aumenta o decresce al crescere di x, analogam al coeff angolare della retta di regressione

Dalle formule sopra abbiamo capito che in corrisp di un increm unitario di x, l'odd cresce di un fattore e^{β}

In altre parole

ODDS RATIO

$$\left.\begin{aligned} \frac{\pi(X=x+1)}{1-\pi(X=x+1)} &= e^{\alpha}e^{\beta(x+1)} \\ \frac{\pi(X=x)}{1-\pi(X=x)} &= e^{\alpha}e^{\beta x} \end{aligned}\right\} \Longrightarrow \log\left(\frac{\frac{\pi(x+1)}{1-\pi(x+1)}}{\frac{\pi(x)}{1-\pi(x)}}\right) = \log\left(\frac{\cancel{e^{\alpha}}e^{\beta(x+1)}}{\cancel{e^{\alpha}}e^{\beta x}}\right) = \beta$$

ovvero e^{β} è un ODDS RATIO, e β misura l'incremento di rischio relativo (~~passaggio~~ rapporto tra prob di successo e insuccesso) in corrispondenza di un incremento unitario del regressore

⇔ facile nel caso di regressore binario (tabella di contingenza)

Probit and Inverse CDF Link Functions
Analoghe alla logistica sono altre funzioni link spesso usate per connettere $\pi(\underline{x})$ al predittore lineare

$$\pi(\underline{x}) = F(\underline{x}) \Rightarrow F^{-1}(\pi(\underline{x})) = \sum_j \beta_j x_{ij}$$

$$\pi(\underline{x}) = \Phi(\underline{x}) \Rightarrow \Phi^{-1}(\pi(\underline{x})) = \sum_j \beta_j x_{ij} \quad \text{PROBIT}$$

(GLM per DATI di tipo CONTEGGIO)
La distribuz di Poisson ha media $\mu > 0$. Abbiamo visto in preced che il log della media è il param naturale in un GLM con risp Poisson e che il log è il suo link canonico. Pertanto

$$Y \sim P(\mu) \qquad \log(\underset{E[Y]}{\mu}) = \alpha + \beta x$$

Consideraz analoghe al caso logistico valgono per l'interpretaz dei β ecc ecc.

- Spesso capita però che i dati manifestino una variabilità maggiore di quella ammessa da un mod di Poisson: OVERDISPERSION
Una delle principali cause è l'eterogeneità dei soggetti.
Di solito qto non è un pb nei soliti mod di regr perchè c'è un param apposta x la varianza. Nel caso di Binomiale e Poisson invece la varianza è una funz della media

⇒ Estensione: Negative Binomial

$$Y \sim f_Y(y; k, \mu) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{k}{\mu+k}\right)^k \left(1 - \frac{k}{\mu+k}\right)^y$$

per k fissato è un GLM

$$E[Y] = \mu \qquad V[Y] = \mu + \frac{\mu^2}{k}$$

se $\frac{1}{k} \to 0 \Rightarrow V[Y] \to \mu$ e $Y \to P$

$1/k$ DISPERSION PARAM

MOMENTS & LIKELIHOOD for GLMs

Per affrontare l'argomento della stima e dell'inferenza ci serve generalizzare la notazione fin qui introdotto per comprendere fam di variabili con più parametri

Sia quindi $\{y_1 \dots y_N\}$ un campione da Y distribuite come

NATURAL PARAMETER — DISPERSION PARAM

$$f_Y(y_i; \theta_i, \phi) = \exp\{[y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi)\} \quad (\square)$$

EXPONENTIAL DISPERSION FAMILY

Ci si riconduce alla forma precedente, per ϕ fissato, immaginando che

$$\varphi(\theta) = \theta/a(\phi)$$
$$a(\theta) = \exp\{-b(\theta)/a(\phi)\}$$
$$b(y) = \exp\{c(y, \phi)\}$$

solitamente è della forma

$$a(\phi) = \phi/\omega_i$$

pesi noti

Es: $Y_i \sim Bi(\pi, 1)$ $i = 1 \dots m \Rightarrow \omega_i = m_i$

Media e Varianza della componente casuale

Si possono ricavare espressioni generali per media e varianza a partire da $(\square)$. Sia $\ell_i = \log f_Y(y_i; \theta_i, \phi)$ e $\ell = \sum_i \ell_i$

$$\overset{(\square)}{\Longrightarrow} \ell_i = [y_i\theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi)$$

$$\Longrightarrow \frac{\partial \ell_i}{\partial \theta_i} = [y_i - b'(\theta_i)]/a(\phi) \qquad \frac{\partial^2 \ell}{\partial \theta_i^2} = -b''(\theta_i)/a(\phi)$$

Ora, sotto opportune hp di regolarità soddisfatte dalla fam exp, valgono i seguenti risultati generali relativi alle verosim

↳ Cox and Hinkley 1974

$$E\left[\frac{\partial L}{\partial \theta}\right] = 0 \qquad -E\left[\frac{\partial^2 L}{\partial \theta^2}\right] = E\left[\frac{\partial L}{\partial \theta}\right]^2$$

da cui si ricava che

i) $$E\left[\frac{y_i - b'(\theta_i)}{a(\phi)}\right] = \frac{1}{a(\phi)}\left(E[y_i] - E[b'(\theta)]\right) = 0 \Rightarrow E[Y_i] = b'(\theta_i)$$

ii) $$\frac{b''(\theta_i)}{a(\phi)} = \left(E\left[\frac{Y_i - b'(\theta_i)}{a(\phi)}\right]\right)^2 = \frac{1}{a(\phi)^2}\left(E[Y - E[Y]]\right)^2 = \frac{Var[Y_i]}{a(\phi)^2}$$

$$\Longrightarrow Var[Y_i] = b''(\theta_i)\, a(\phi)$$

$b(\cdot)$ è la funzione che determina i momenti di Y.

Esempio POISSON

$$Y \sim P(\mu) \implies f_{Y_i}(y_i;\mu) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!} = \exp\{y_i \log\mu_i - \mu_i - \log y_i!\} \overset{\theta=\log(\mu_i)}{=} \exp\{y_i\theta_i - \exp(\theta_i) - \log y_i!\}$$

$\implies \in$ exp disp family with $b(\theta_i) = \exp(\theta_i)$

$a(\phi) = 1$

$c(y_i,\phi) = -\log y_i!$

$$\implies E[Y_i] = b'(\theta_i) = \exp(\theta_i) = \mu_i$$

$$V[Y_i] = b''(\theta_i)a(\phi) = \mu_i$$

Esempio BINOMIALE

$m_i Y_i \sim Bi(m_i, \pi_i)$ $\implies f_{Y_i}(y_i; m_i, \pi_i) = \binom{m_i}{m_i y_i}\pi_i^{m_i y_i}(1-\pi_i)^{m_i - m_i y_i}$

proporz di successi anziché #

$$\overset{\theta_i=\text{logit}(\pi_i)}{=} \exp\left\{\frac{y_i\theta_i - \log(1+\exp(\theta_i))}{1/m_i} + \log\binom{m_i}{m_i y_i}\right\}$$

$\log(1-\pi_i) = -\log(1+\exp(\theta_i))$

$\implies \in$ exp disp family with $b(\theta_i) = \log(1+\exp(\theta_i))$

$a(\phi) = 1/m_i$

$c(y_i,\phi) = \log\binom{m_i}{m_i y_i}$

$$\implies E[Y_i] = b'(\theta_i) = \frac{\exp(\theta_i)}{1+\exp(\theta_i)} = \pi_i \qquad \frac{\pi/1-\pi}{1+\pi/1-\pi} = \frac{1-\pi}{1-\pi}\cdot\frac{\pi}{1-\pi+\pi} = \pi$$

$$V[Y_i] = b''(\theta_i)a(\phi) = \frac{\exp(\theta_i)}{[1+\exp(\theta_i)]^2}\cdot\frac{1}{m_i} = \frac{\pi_i(1-\pi_i)}{m_i}$$

Nota: $\eta_i = g(\mu_i) = \sum_j \beta_j x_{ij}$. Inoltre abbiamo detto che la link function tale per cui

$$\theta_i = \sum_j \beta_j x_{ij} \longleftarrow \quad g(\mu_i) = \theta_i$$

è quella canonica

Dal momento che $\mu_i = b'(\theta_i) \implies \theta_i = (b')^{-1}(\mu_i)$

$\implies$ il link canonico è l'inversa di b

Likelihood Equations for GLMs

Per N osservazioni indipendenti,

$$\ell(\underline{\beta}) = \sum_i \ell_i = \sum_i \log f_Y(y_i;\theta_i,\phi) = \sum_i \frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + \sum_i c(y_i,\phi)$$

La dipendenza da $\underline{\beta}$ e' mascherata nel fatto che $\theta = \theta(\underline{\beta})$

Le score equations sono quindi
$$\frac{\partial \ell(\underline{\beta})}{\partial \beta_j} = \sum_i \frac{\partial \ell_i}{\partial \beta_j} = 0 \quad \forall j = 1 \dots p$$

Applicando la chain rule

$$\frac{\partial \ell_i}{\partial \beta_j} = \underbrace{\frac{\partial \ell_i}{\partial \theta_i}}_{A} \cdot \underbrace{\frac{\partial \theta_i}{\partial \mu_i}}_{B} \cdot \underbrace{\frac{\partial \mu_i}{\partial \eta_i}}_{C} \cdot \underbrace{\frac{\partial \eta_i}{\partial \beta_j}}_{D}$$

A. $\dfrac{\partial \ell_i}{\partial \theta_i} = \dfrac{[y_i - b'(\theta_i)]}{a(\phi)} \underset{\mu_i = b'(\theta_i)}{=} \dfrac{y_i - \mu_i}{a(\phi)}$

D. $\dfrac{\partial \eta_i}{\partial \beta_j} = x_{ij}$ in quanto $\eta_i = \sum_j \beta_j x_{ij}$

B. $\dfrac{\partial \mu_i}{\partial \theta_i} = \dfrac{Var[Y_i]}{a(\phi)}$ in quanto $\begin{cases} \mu_i = b'(\theta_i) \Rightarrow \frac{\partial \mu}{\partial \theta} = b''(\theta) \\ Var[Y_i] = b''(\theta_i)\, a(\phi) \Rightarrow b''(\theta) = \frac{Var}{a} \end{cases}$

C. dipende dalla link function del modello

$$\Longrightarrow \frac{\partial \ell_i}{\partial \beta_j} = \frac{y_i - \mu_i}{a(\phi)} \cdot \frac{a(\phi)}{V[Y_i]} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot x_{ij} = \frac{(y_i - \mu_i)x_{ij}}{Var[Y_i]} \cdot \frac{\partial \mu_i}{\partial \eta_i}$$

$$\Longrightarrow \frac{\partial \ell}{\partial \beta_j} = \sum_i \frac{(y_i - \mu_i)x_{ij}}{Var[Y_i]} \cdot \frac{\partial \mu_i}{\partial \eta_i} = 0$$

NB: dipendenza da β nascosta nel fatto che $\mu_i = g^{-1}\left(\sum_j \beta_j x_{ij}\right)$

Oss 1: Le eq della verosim dipendono dalla distrib di Y_i solo attraverso μ_i e $V[Y_i]$

Oss 2: La varianza dipende dalla media $Var[Y_i] = \nu(\mu_i)$
- ⤳ Poisson: $\nu(\mu_i) = \mu_i$
- ⤳ Bernoulli: $\nu(\mu_i) = \mu_i(1-\mu_i)$
- ⤳ Normale: $\nu(\mu_i) = \sigma^2$ (i.e. constant)

Quando $Y_i \in$ fam exp, la relaz tra media e varianza caratterizza la distribuzione

FITTING GLM

Avevamo lasciato in sospeso come ottenere le stime dei parametri, dicendo che le score equations di solito nei GLM sono non lineari in $\underline{\beta}$ $\Rightarrow$ servono metodi iterativi

1. Newton-Raphson

Algoritmo iterativo per la soluz di sistemi non lineari

0 Parte da una guess iniziale sulla soluz

1 Approssima la funz da massimizzare in un intorno del pto di cui in (0) tramite un polinomio di secondo grado e ne determina il max

2 Itera fino a convergenza

Più precisamente, sia

$$u' = \left(\frac{\partial L(\underline{\beta})}{\partial \beta_1}, \frac{\partial L(\underline{\beta})}{\partial \beta_2}, \dots \right)$$

$$H = \left\{ h_{ab} : h_{ab} = \frac{\partial^2 L(\underline{\beta})}{\partial \beta_a \partial \beta_b} \right\} \quad \text{HESSIANA}$$

Siano quindi $u^{(t)}$ e $H^{(t)}$ le loro valutazioni in $\underline{\beta}^{(t)}$ ← guess solution for $\underline{\beta}$ all'iteraz t-esima

Allo step t, l'algoritmo approssima $L(\underline{\beta})$ nell'intorno di $\underline{\beta}^{(t)}$ come

$$L(\underline{\beta}) \approx L(\underline{\beta}^{(t)}) + u^{(t)'}(\underline{\beta}-\underline{\beta}^{(t)}) + \frac{1}{2}(\underline{\beta}-\underline{\beta}^{(t)})^T H^{(t)} (\underline{\beta}-\underline{\beta}^{(t)})$$

Risolvendo

$$\frac{\partial L(\underline{\beta})}{\partial \underline{\beta}} \approx u^{(t)} + H^{(t)}(\underline{\beta}-\underline{\beta}^{(t)}) = 0$$

← cerco punto staz (max locale dell'approx polin)

si ottiene la guess per il passo successivo

$$\underline{\beta}^{(t+1)} = \underline{\beta}^{(t)} - (H^{(t)})^{-1}\underline{u}^{(t)}$$

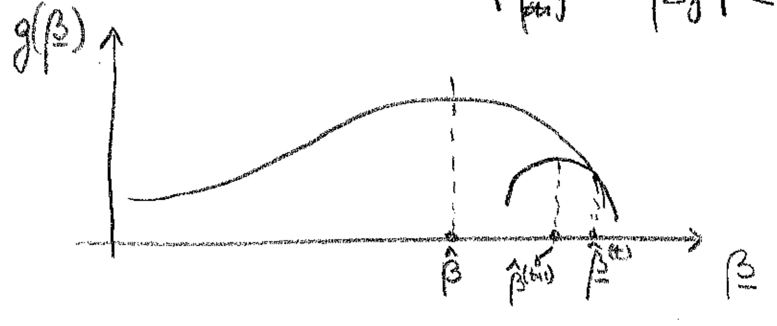
assumendo H non singolare (cmq le routine risolvono il sist lin piuttosto che calcolare l'inverso)

Si prosegue finché la differenza di $L(\underline{\beta}^{(t)})$ tra due istanti successivi è sufficientemente piccola.

Ovviamente nulla assicura che la convergenza sia al massimo globale

La convergenza avviene in genere in modo abbastanza rapido, in quanto per t grandi vale che

$$|\beta_j^{(t+1)} - \hat{\beta}_j| \leq c\, |\beta_j^{(t)} - \hat{\beta}_j|^2 \qquad c > 0$$



2. Fisher Scoring

Metodo iterativo del tutto analogo al Newton-Raphson se non nel fatto che usa il valore atteso dell'hessiana (EXPECTED INFORMATION) anziche' l'hessiana stessa (OBSERVED INFORMAT.) all'interno dell'approssimaz polinomiale al second'ordine

Inferenza per i parametri di regressione

Test di ipotesi sui singoli regressori: $H_0: \beta_j = 0$ vs $H_1: \beta_j \neq 0$

WALD TEST $\frac{(\hat{\beta}-0)^2}{se(\hat{\beta})^2} \sim \chi^2_{(1)}$

anche visto come test z asintotico $\Rightarrow \frac{\hat{\beta}-0}{se(\hat{\beta})} \sim \mathcal{N}(0,1)$

$\hookrightarrow$ GOF !

Curva ROC e tab miscl

INFERENZA per GLM

Abbiamo introdotto in precedenza il modello saturato che per ogni osservazione introduceva un diverso param. Qto modello fornisce il fit perfetto (interpolaz) ma non e' molto utile in quanto non fornisce alcuna sintesi parsimoniosa delle info del dataset, ne' si può usare x fare previsione. Tuttavia serve come termine di cfr.

Un mod saturato spiega tutta la variab tramite la componente sistematica.

Sia $\tilde{\theta}$ la stima di θ e $\tilde{\mu}_i = y_i \ \forall i$ le corrispondenti stime delle medie.
Siano inoltre $\hat{\theta}$ e $\hat{\mu}_i$ le stime ML di un generico mod insaturo.
Allora la statistica

$$G = -2 \log \frac{\text{max likelihood model}}{\text{max likelihood sat model}} = -2\left[L(\hat{\underline{\mu}}, \underline{y}) - L(\underline{y}, \underline{y})\right]$$

descrive il LOF tra i 2 modelli. Essa rappresenta il rapporto di verosim per testare la bontà del mod in esame contro generiche alternative.

Recuperando l'espressione della likelihood,

$$-2\left[L(\hat{\mu}, \underline{y}) - L(\underline{y}, \underline{y})\right] =$$

$$= 2\sum_i \left[y_i \tilde{\theta}_i - b(\tilde{\theta}_i)\right] / a(\phi) - 2\sum_i \left[y_i \hat{\theta}_i - b(\hat{\theta}_i)\right] / a(\phi)$$

Essendo $a(\phi)$ di solito pari a ϕ / ω_i,

$$= 2\sum_i \omega_i \left[y_i (\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)\right] / \phi = \underbrace{D(\underline{y}; \hat{\underline{\mu}}) / \phi}_{\text{SCALED DEVIANCE}}$$

Più grande è la scaled deviance e maggiore è il LOF che si ha considerando il modello ridotto

Per alcuni GLM la scaled deviance $\sim \chi^2$

Esempio BINOMIALE

$X_i = m_i Y_i \sim Bi(m_i, \pi_i) \implies \hat{\theta}_i = \log(\hat{\pi}_i / (1-\hat{\pi}_i))$

$$b(\hat{\theta}_i) = \log(1 + \exp\{\hat{\theta}_i\})$$
$$= -\log(1-\hat{\pi}_i)$$

Analogamente

$\tilde{\theta}_i = \log(y_i / (1-y_i))$

$b(\tilde{\theta}_i) = -\log(1-y_i)$

$a(\phi) = 1/m_i \implies \phi = 1$ e $\omega_i = m_i$

Teorema

Per testare $H_0: \theta = \theta_0$ vs $H_1: \theta \neq \theta_0$, si supponga che $X_1 \dots X_m$ iid $\sim f_X(x;\theta)$, e che $\hat{\theta}$ sia lo stimatore ML di θ e che $f(\cdot)$ soddisfi le condizioni di regolarità (..).

Allora, sotto H_0 e per $m \to \infty$

$$-2 \log \lambda(\underline{x}) \xrightarrow{\mathcal{L}} \chi^2_{(1)}$$

Dimostrazione

$$-2 \log \lambda(\underline{x}) = -2 \log \left(\frac{L(\theta_0; \underline{x})}{L(\hat{\theta}; \underline{x})} \right)$$

$$= -2 \ell(\theta_0) + 2 \ell(\hat{\theta})$$

Espandendo la log verosimiglianza in serie di Taylor attorno a $\hat{\theta}$, si ottiene

$$\ell(\theta_0) = \ell(\hat{\theta}) + \underbrace{\cancel{\ell'(\hat{\theta})}}_{0 \text{ per definiz di } \hat{\theta}}(\theta_0 - \hat{\theta}) + \ell''(\hat{\theta}) \frac{(\theta_0 - \hat{\theta})^2}{2} + \dots$$

$$= \cancel{-2\ell(\hat{\theta})} - \cancel{2}\ell''(\hat{\theta}) \frac{(\theta_0 - \hat{\theta})^2}{\cancel{2}} + \cancel{2\ell(\hat{\theta})}$$

$$\Rightarrow \quad -2 \log \lambda(\underline{x}) = -\ell''(\hat{\theta})(\theta_0 - \hat{\theta})^2$$

Ora, sotto le opportune ip di regolarità (vedi CR) si può mostrare che

$$\hat{I}_m(\hat{\theta}) = E\left[\left(\log f(\underline{x}; \hat{\theta}) \right)^2 \right] = -\ell''(\hat{\theta})$$

e che $\quad \frac{1}{m} \hat{I}_m(\hat{\theta}) \xrightarrow{P} I(\theta)$

Quindi, sotto H_0,

$$-2 \log \lambda(\underline{x}) = -\ell''(\hat{\theta})(\theta_0 - \hat{\theta})^2 = \underbrace{\underbrace{-\ell''(\hat{\theta}) \cdot \frac{1}{m}}_{\xrightarrow{P} I(\theta_0)} \frac{1}{I(\theta_0)}}_{\xrightarrow{P} 1} \cdot \underbrace{\frac{\sqrt{m^2}}{\left(\frac{1}{\sqrt{I(\theta_0)}} \right)^2} \cdot (\hat{\theta} - \theta_0)^2}_{= \left(\frac{\sqrt{m}(\hat{\theta} - \theta_0)}{1/\sqrt{I(\theta_0)}} \right)^2 \xrightarrow{\mathcal{L}} \chi^2_1}$$

asintotica efficienza MLE

Quindi per il teo di Slutzky

$$-2 \log \lambda(\underline{x}) \xrightarrow{\mathcal{L}} \chi^2_1$$

$$-2D = -2\left[\ell(\theta_0) - \ell(\hat{\theta}) \right]$$

$$\Rightarrow 2\sum_i m_i \left\{ y_i\left(\log\frac{y_i}{1-y_i} - \log\frac{\hat\pi_i}{1-\hat\pi_i}\right) + \log(1-y_i) - \log(1-\hat\pi_i)\right\}$$

moltiplico e divido per m_i

$$= 2\sum_i m_i y_i \log\frac{m_i y_i}{m_i - m_i y_i} - 2\sum_i m_i y_i \log\frac{m_i\hat\pi_i}{m_i - m_i\hat\pi_i} + 2\sum_i m_i \log\frac{1-y_i}{1-\hat\pi_i}$$

$$= 2\sum_i m_i y_i \left(\log\left(\frac{m_i y_i}{m_i - m_i y_i}\right) - \log\left(\frac{m_i\hat\pi_i}{m_i - m_i\hat\pi_i}\right)\right) + 2\sum_i \cdots$$

$$= 2\sum_i m_i y_i \left(\log\frac{m_i y_i}{m_i\hat\pi_i} + \frac{m_i - m_i\hat\pi_i}{m_i - m_i y_i}\right) + 2\sum_i m_i \log\frac{m_i - m_i y_i}{m_i - m_i\hat\pi_i}$$

$$= 2\sum_i m_i y_i \log\frac{m_i y_i}{m_i\hat\pi_i} + 2\sum_i m_i y_i \log\frac{m_i - m_i\hat\pi_i}{m_i - m_i y_i} + 2\sum_i m_i \log\frac{m_i - m_i y_i}{m_i - m_i\hat\pi_i}$$

$$= 2\sum_i \underbrace{(m_i y_i)}_{\#\ \text{successi}} \log\frac{m_i y_i}{m_i\hat\pi_i} + 2\sum_i \underbrace{(m_i - m_i y_i)}_{\#\ \text{insuccessi}} \log\frac{m_i - m_i y_i}{m_i - m_i\hat\pi_i}$$

$\Rightarrow$

LR Model Comparison using Deviance

Per i mod Poisson e Binom abbiamo visto che $\phi = 1$, pertanto

$$D(\underline{y};\hat{\underline{\mu}}) = -2\left[L(\hat{\underline{\mu}},\underline{y}) - L(\underline{y},\underline{y})\right]$$

Consideriamo ora 2 modelli M_0 con valori fittati $\hat{\underline{\mu}}_0$ e M_1 con valori fittati $\hat{\underline{\mu}}_1$, in cui M_0 sia nested in M_1. Dal momento che la verosim massimizzata su un # ristretto di valori non può assumere valori più grandi, si avrà

$$L(\hat{\underline{\mu}}_0;\underline{y}) \leq L(\hat{\underline{\mu}}_1;\underline{y}) \Rightarrow D(\underline{y};\hat{\underline{\mu}}_1) \leq D(\underline{y};\hat{\underline{\mu}}_0)$$

simpler models have larger dev.

Immaginando che valga M_1, il LR test per il cfr tra M_1 e M_0 userà la stat test

$$-2\left[L(\hat{\underline{\mu}}_0;\underline{y}) - L(\hat{\underline{\mu}}_1;\underline{y})\right] = -2\left[L(\hat{\underline{\mu}}_0;\underline{y}) - L(\underline{y};\underline{y})\right] - \left\{-2\left[L(\hat{\underline{\mu}}_1;\underline{y}) - L(\underline{y};\underline{y})\right]\right\}$$

$$= D(\underline{y};\hat{\underline{\mu}}_0) - D(\underline{y};\hat{\underline{\mu}}_1)$$

$\Rightarrow$ la LR statistics che confronta i due modelli non è altro che la diff tra le devianze.

La statistica test assume valori grandi quando M_0 produce un fit molto più insoddisfacente (significativamente più povero) di M_1

$$= 2\sum_i \omega_i\left[y_i(\hat{\theta}_{1i} - \hat{\theta}_{0i}) - b(\hat{\theta}_{1i}) + b(\hat{\theta}_{0i})\right]$$

opportune hp di regolarità generalm soddisfatte da fam exp

$$\approx \chi^2_{(p_{M_1} - p_{M_0})}$$

param M_1 — # param M_0

Residui

Come sempre se un mod produce un fitting insoddisfacente, l'analisi dei residui permette di capire dove in particolare il modello è povero.

Nel caso dei GLM, si possono considerare vari tipi di residui

1. Deviance Residuals: $\sqrt{d_i} \times \text{sign}(y_i - \hat{\mu}_i)$ $\quad d_i = 2\omega_i[y_i(\tilde{\theta}_i - \hat{\theta}_i) - b(\tilde{\theta}_i) + b(\hat{\theta}_i)]$
2. Pearson Residuals: $e_i = \dfrac{y_i - \hat{\mu}_i}{Var[Y_i]^{1/2}}$

C'è anche un analogo degli studentized residuals e dei leverages