Kanonisk link og sufficient stikprøvefunktion

Når $\eta_i = \theta_i$, dvs. når $g(\mu_i) = \theta_i$, kaldes $g(\cdot)$ for det kanoniske link (θ_i er den kanoniske parameter i den eksponentielle familie.)

Loglikelihoodfunktionen bliver

$$l(\beta) = \sum_i \left(\frac{w_i}{\varphi}(y_i\theta_i - b(\theta_i)) + c_i(y_i, \varphi) \right)$$

$$= \sum_i \frac{w_i}{\varphi}(y_i x_i^T\beta - b(x_i^T\beta)) + \sum_i c_i(y_i, \varphi)$$

$$= \frac{1}{\varphi}\left((\sum_i w_i y_i x_i)^T \beta - \sum_i w_i b(x_i^T\beta) \right) + \sum_i c_i(y_i, \varphi)$$

dvs. $\sum_i w_i y_i x_i$ er minimal sufficient for β, når φ er kendt. Når φ er ukendt, er $\sum_i w_i y_i x_i$ en komposant af den minimalsufficiente stikprøvefkt.

eks. $Y_i \sim \Gamma(\omega, \frac{\omega}{\mu_i})$, formparameteren ω konstant for alle obs.

$$g(\mu_i) = \eta_i = x_i^T\beta$$

tæthedsfkt. (indeks i udeladt):

$$f(y) = \frac{(\frac{\omega}{\mu})^\omega}{\Gamma(\omega)} y^{\omega-1} e^{-\frac{\omega}{\mu}y}, \quad y > 0$$

$$= \exp\left(\omega(-\frac{y}{\mu} - \ln\mu) + (\omega-1)\ln y - \ln\Gamma(\omega) + \omega\ln\omega\right)$$

$$= \exp\left(\omega\left(\theta y + \ln(-\theta)\right) + c(y,\omega)\right),$$

hvor $\mu = -\frac{1}{\theta}$.

Det kanoniske link er

$$-\frac{1}{\mu_i} = \theta_i = x_i^T\beta$$

Heraf

$$\ell(\beta) = \left(\omega\left(\left(\sum_i y_i x_i\right)^T\beta + \sum_i\left(-x_i^T\beta\right)\right) + \sum_i c(y_i,\omega)\right)$$

dvs. med ω kendt er $\sum_i y_i x_i$ min. suff. for β. Når ω er ukendt, er

$\left(\sum_i y_i x_i, \sum_i \ln y_i\right)$ min. suff. for (β,ω). □

Andre fordele ved at benytte kanonisk link

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial \mu_i}{\partial \theta_i} = \frac{\partial b'(\theta_i)}{\partial \theta_i} = b''(\theta_i)$$

$\Downarrow$

$$\frac{\partial \ell_i}{\partial \beta_j} = \frac{(y_i-\mu_i)x_{ij}}{Var[Y_i]}\, b''(\theta_i) = \frac{w_i\,(y_i-\mu_i)x_{ij}}{\varphi}$$

For $w_i = 1$ og med $\hat{\mu}_i = g^{-1}(x_i^T\hat{\beta})$ får vi

$$\sum_i \frac{\partial \ell_i}{\partial \beta_j} = 0 \iff \sum_i \frac{(y_i-\hat{\mu}_i)x_{ij}}{\varphi} = 0$$

$$\Rightarrow \sum_i y_i x_{ij} = \sum_i \hat{\mu}_i x_{ij}$$

$$\iff X^T y = X^T\hat{\mu}$$

Når (fx) 1. søjle i X er 1_n, bliver $\sum y_i = \sum \hat{\mu}_i$, dvs. summen af obs. er lig summen af de estimerede værdier.

Øvrige søjler i X svarer til, at visse linearkombinationer af hhv. y_i'erne og $\hat{\mu}_i$'erne får samme værdi.

Vedr. Fisherinformation:

Ved diff. af $\frac{\partial l_i}{\partial \beta_j}$ mht. β_k får vi

$$\frac{\partial^2 l_i}{\partial \beta_j \partial \beta_k} = - \frac{w_i x_{ij}}{\phi} \frac{\partial \mu_i}{\partial \beta_k}$$, som er uafh. af y

$$\Rightarrow -\frac{\partial^2 l}{\partial \beta_j \partial \beta_k} = -E\left[\frac{\partial^2 l}{\partial \beta_j \partial \beta_k}\right]$$, dvs.

den observerede information og den forventede information er sammenfaldende.

IRWLS-algoritme

Løsning af likelihoodligningerne:

Lad $u^{(k)}$ være scoringsfunktionen udregnet svarende til $\beta^{(k)}$ (k'te iteration)

I Newton-Raphson's iterationsligning erstattes den observerede information med Fisherinformationen, hvorved Fishers scoringsalgoritme fremkommer.
(jf. AA s. 63)

$$\beta^{(k+1)} = \beta^{(k)} + \left(I(\beta^{(k)})\right)^{-1} u^{(k)}$$

$$\Rightarrow I(\beta^{(k)})\beta^{(k+1)} = I(\beta^{(k)})\beta^{(k)} + u^{(k)} \qquad (*)$$

Højresidens k'te element, jf. note 13 s. 7 og 8

$$\sum_j \left(\sum_i \frac{x_{ih} x_{ij}}{Var[Y_i]} \left(\frac{\partial \mu_i^{(k)}}{\partial \eta_i^{(k)}} \right)^2 \right) \beta_j^{(k)} + \sum_i \frac{(y_i - \mu_i^{(k)}) x_{ih}}{Var[Y_i]} \frac{\partial \mu_i^{(k)}}{\partial \eta_i^{(k)}}$$

Med

$$\tilde{w}_i^{(k)} = \frac{1}{Var[Y_i]} \left(\frac{\partial \mu_i^{(k)}}{\partial \eta_i^{(k)}} \right)^2$$

og

$$z_i^{(k)} = \sum_j x_{ij} \beta_j^{(k)} + (y_i - \mu_i^{(k)}) \frac{\partial \eta_i^{(k)}}{\partial \mu_i^{(k)}}$$

$$= \eta_i^{(k)} + (y_i - \mu_i^{(k)}) \left(\frac{\partial \eta_i^{(k)}}{\partial \mu_i^{(k)}} \right)$$

bliver højresiden til $X^T \tilde{W}^{(k)} z^{(k)}$.

Idet $I(\beta^{(k)}) = X^T \tilde{W}^{(k)} X$, kan (*) nu skrives

$$X^T \tilde{W}^{(k)} X \beta^{(k+1)} = X^T \tilde{W}^{(k)} z^{(k)}$$

Dermed har vi

$$\beta^{(k+1)} = (X^T \tilde{W}^{(k)} X)^{-1} X^T \tilde{W}^{(k)} z^{(k)}$$

(Udtrykket er analogt til estimatet for β i en lineær normal model med vægtfaktorer, jf. opg. 5.18)

Der itereres indtil $\beta^{(k)}$ konvergerer.

(IRWLS = iterated reweighted least squares)

Bemærk, at

$$g(y_i) \approx g(\mu_i) + g'(\mu_i)(y_i - \mu_i)$$

$$= \eta_i + (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i}$$

$$= z_i$$

dvs z_i er lokalt den lineære approksimation af $g(y_i)$.

Start på IRWLS : $z_i^{(0)} = g(y_i)$
$\tilde{W}^{(0)} = I_n$

eks. IRWLS på lin. normal model : $z^{(0)} = y$

dvs. $$\beta^{(1)} = (X^T I_n X)^{-1} X^T I_n y$$
$$= (X^T X)^{-1} X^T y = \hat{\beta} \quad (\text{stabil?})$$
$$\tilde{w}_i^{(1)} = \frac{1}{\sigma^2} 1^2 \Rightarrow \tilde{W}^{(1)} = \frac{1}{\sigma^2} I_n$$
$$z_i^{(1)} = x_i^T \hat{\beta} + (y_i - x_i^T \hat{\beta})\, 1 = y_i \quad (\beta^{(1)} = \hat{\beta})$$
$$\beta^{(2)} = (X^T \tfrac{1}{\sigma^2} I_n X)^{-1} X^T \tfrac{1}{\sigma^2} I_n y$$
$$= (X^T X)^{-1} X^T y = \hat{\beta} = \beta^{(1)}$$

dvs. konvergens efter én iteration! $\square$

eks. IRWLS på logistisk regression

Fra tidl.:

$$\mathrm{Var}[Y_i] = \frac{1}{m_i} V(\mu_i) = \frac{1}{m_i} \mu_i (1-\mu_i)$$
$$g(\mu_i) = \ln \frac{\mu_i}{1-\mu_i} = \eta_i, \quad \mu_i = \frac{e^{\eta_i}}{1+e^{\eta_i}}$$
$$\eta_i = x_i^T \beta = \theta_i, \quad \text{dvs. kanonisk link } \theta_i = \mathrm{logit}\, \mu_i$$

Vidr. iterationer:

$$\tilde{w}_i^{(k)} = \frac{m_i}{\mu_i^{(k)}(1-\mu_i^{(k)})} \left(\mu_i^{(k)}(1-\mu_i^{(k)})\right)^2$$
$$= m_i \mu_i^{(k)} (1-\mu_i^{(k)})$$
$$z_i^{(k)} = \eta_i^{(k)} + (y_i - \mu_i^{(k)}) \frac{1}{\mu_i^{(k)}(1-\mu_i^{(k)})}$$
$$\left(\frac{d\eta}{d\mu} = \frac{d}{d\mu} \ln \frac{\mu}{1-\mu} = \frac{1-\mu}{\mu} \frac{(1-\mu) 1 - \mu(-1)}{(1-\mu)^2} = \frac{1}{\mu(1-\mu)} \right)$$

Start på IRWLS:

$$z_i^{(0)} = g(y_i) = \ln \frac{y_i}{1-y_i} = \ln \frac{\tilde{y}_i}{m_i - \tilde{y}_i} ,$$

men $\tilde{y}_i = 0$ eller $\tilde{y}_i = m_i$ kan forekomme, vælg i stedet

$$z_i^{(0)} = \ln \frac{\tilde{y}_i + \frac{1}{2}}{m_i - \tilde{y}_i + \frac{1}{2}} ,$$

som kaldes empirisk logit.

$\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)$

Den estimerede kurve:

$$\hat{\pi}(x) = \frac{\exp(\hat{\beta}_1 + \hat{\beta}_2 x)}{1 + \exp(\hat{\beta}_1 + \hat{\beta}_2 x)}$$

Taleksempel: Se AA s. 241 $\square$

Estimation af φ (dispersionsparameteren)

$$\tilde{\varphi} = \frac{1}{n-p} \sum_i \frac{w_i (y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} ,$$

hvor $\hat{\mu}_i$ estimeres vha. IRWLS

$\tilde{\varphi}$ er simplere at beregne end maksimum likelihood estimatet, og har større numerisk stabilitet (u/ bevis)

Udregnes $\tilde{\varphi}$ svarende til lineær normal model, får vi

$$\tilde{\phi} = \frac{1}{n-r} \sum_i \frac{1(y_i - \hat{\mu}_i)^2}{1} = \frac{1}{n-r} \sum_i (y_i - \hat{\mu}_i)^2$$

$$= s^2 ,$$

det sædvanlige variansestimat

Devians

Residualkvadratsummen i en lineær normal model

$$\sum_i (y_i - \hat{\mu}_i)^2 = \| y - \hat{\mu} \|^2$$

kaldes også for deviansen og betegnes derfor ofte D, dvs.

$$\ell(\hat{\beta}) = c - \frac{n}{2} \ln \sigma^2 - \frac{D}{2\sigma^2}$$

En række forskellige modeller fremkommer ved at ændre på den lineære prædiktor. Den mættede model svarer til estimatet $\tilde{\mu}_i = y_i$, der har devians 0. Den benyttes kun som reference i forbindelse med andre modeller. Bemærk, at

$$-2\left(l(\hat{\beta}) - l(\tilde{\beta})\right) = \frac{D-0}{\sigma^2} = \frac{D}{\sigma^2} \qquad (*)$$

($l(\cdot)$ som betegnelse dækker over forskellige loglikelihoodfunktioner.)

Model M_2 siges at være indlejret i M_1, $M_2 \subset M_1$, når M_2 er fremkommet ved tilføjelse af yderligere bånd på parametrene, $g_i(\beta) = 0$, $i = 1, 2, \ldots, p_1 - p_2$

$$W = -2\left(l(\hat{\beta}_2) - l(\hat{\beta}_1)\right) = \frac{D_2 - D_1}{\sigma^2} \sim \chi^2(p_1 - p_2)$$

under H_0: M_2 er sand

<u>For GLM</u> opstilles analogt til (*)

$$W(y) = -2\left(l(\hat{\beta}) - l(\tilde{\beta})\right)$$

$$= -2\sum_i \left(\frac{w_i}{\varphi}\left(y_i\hat{\theta}_i - b(\hat{\theta}_i) - (y_i\tilde{\theta}_i - b(\tilde{\theta}_i))\right)\right)$$

$$= \frac{D(y;\hat{\mu})}{\varphi},$$

som kaldes skaleret devians (tælleren alene kaldes deviansen).

$$D(y,\hat{\mu}) = \sum_i 2 w_i \left(y_i(\tilde{\theta}_i - \hat{\theta}_i) - (b(\tilde{\theta}_i) - b(\hat{\theta}_i)) \right)$$

$$= \sum_i d_i \text{ , hvor}$$

d_i er bidraget til deviansen fra den i'te observation

For indlejret model $M_2 \subset M_1$ gælder

$$\frac{D(y;\hat{\mu}_2) - D(y;\hat{\mu}_1)}{\varphi} \xrightarrow{d} \chi^2(p_1 - p_2) \text{ for } n \to \infty$$

ifølge den approksimative teori i AA kap. 4

(gælder dog ikke i almindelighed, når M_1 er den mættede model)

eks. $Y_i \sim N(\mu, \sigma^2)$

$$D(y;\hat{\mu}) = \sum_i 2 \cdot 1 \cdot \left(y_i(y_i - \hat{\mu}_i) - \tfrac{1}{2}(y_i^2 - \hat{\mu}_i^2) \right)$$

$$= \sum_i (y_i^2 - 2 y_i \hat{\mu}_i + \hat{\mu}_i^2)$$

$$= \sum_i (y_i - \hat{\mu}_i)^2$$

$$= D$$

$\square$

eks. $Y_i \sim P(\mu_i)$, $\theta_i = \ln \mu_i$, $b(\theta_i) = \mu_i$

$$D = D(y; \hat{\mu}) \qquad (\varphi = 1, \ w_i = 1)$$

$$= 2 \sum_i \left(y_i(\ln y_i - \ln \hat{\mu}_i) - (y_i - \hat{\mu}_i) \right)$$

$$= 2 \sum_i \left(y_i \ln \frac{y_i}{\hat{\mu}_i} - y_i + \hat{\mu}_i \right) \qquad (0 \ln 0 := 0)$$

Når $\underbrace{\sum_i y_i = \sum_i \hat{\mu}_i}$ bliver $D = 2 \sum_i y_i \ln \frac{y_i}{\hat{\mu}_i}$,

opfyldt, når det kanoniske link benyttes, og når $1_n \in C(X)$

Stat. 5.5.2010

Residualer

Pearson residual

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\frac{1}{w_i} V(\hat{\mu}_i)}}$$

Devians residual

$$r_i^D = \operatorname{sgn}(y_i - \hat{\mu}_i)\sqrt{d_i}$$

Anscombe residual

$$r_i^A = \frac{A(y_i) - A(\hat{\mu}_i)}{A'(\hat{\mu}_i)\sqrt{\frac{1}{w_i} V(\hat{\mu}_i)}} \quad , \quad A(x) = \int \frac{1}{(V(x))^{\frac{1}{3}}} dx$$

eks. $\quad Y_i \sim N(\mu, \sigma^2)$

$$r_i^P = r_i^D = r_i^A = y_i - \hat{\mu}_i$$

$\square$

eks. $\quad Y_i \sim p(\mu)$

$$r_i^P = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}} \quad *$$

$$r_i^D = \operatorname{sgn}(y_i - \hat{\mu}_i)\sqrt{2\left(y_i \ln \frac{y_i}{\hat{\mu}_i} - y_i + \hat{\mu}_i\right)}$$

$$r_i^A = \frac{\frac{3}{2}\left(y_i^{\frac{2}{3}} - \hat{\mu}_i^{\frac{2}{3}}\right)}{\hat{\mu}_i^{\frac{1}{6}}}$$

* Bemærk $\displaystyle\sum_i (r_i^P)^2 = \sum_i \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$

Pearsons χ^2