17.09.2015: 55-57, 14, 15 (55) (1)

$\varphi_i = a_i \varphi$ a) $\ell(\beta_0, \dots, \beta_p, \varphi) = \varphi^{-1} \sum_{i=1}^{n} \left(\frac{y_i \theta_i - a(\theta_i)}{a_i} + C(y_i, \varphi) \right)$

$\log f(y_1, \dots, y_n \mid \beta_0, \dots, \beta_p, a_1, \dots, a_n, \varphi) = [y_i \text{ are ind}] =$

$= \log \prod_{i=1}^{n} f(y_i \mid \beta_0, \dots, \beta_p, a_i, \varphi) = \sum_{i=1}^{n} \log f(y_i \mid \beta_0, \dots, \beta_p, a_i, \varphi) =$

$= \sum_{i=1}^{n} \log \left(c(y_i, \varphi a_i)\, e^{\frac{\theta_i y_i - a(\theta_i)}{a_i \varphi}} \right) = \left[\theta_i = \tilde{g}(\beta_0, \dots, \beta_p, x_i) \right] =$

$= \sum_{i=1}^{n} \left(K(y_i, \varphi a_i) + \frac{\theta_i y_i - a(\theta_i)}{a_i \varphi} \right) =$

$= \varphi^{-1} \sum_{i=1}^{n} \left(K^*(y_i, \varphi a_i) + \frac{\theta_i y_i - a(\theta_i)}{a_i} \right)$

b) φ & $\beta_0, \dots, \beta_p$ - functionally ind. Show that maximizing $\ell(\beta_0, \dots, \beta_p, \varphi)$ w.r.t $\vec{\beta}$ does not depend on φ, show the form of

$\ell_p(\varphi) = \ell(\hat{\beta}_0, \dots, \hat{\beta}_p, \varphi) = \varphi^{-1} \sum_{i=1}^{n} \left[\frac{y_i \hat{\theta}_i - a(\hat{\theta}_i)}{a_i} + c(y_i, a_i \varphi) \right]$

sep 17-09:56

$\ell(\vec{\beta}, \varphi) =: \ell(\cdot) = \varphi^{-1} \sum_{i=1}^{n} \left(\frac{y_i \theta_i - a(\theta_i)}{a_i} + c(y_i, \varphi a_i) \right)$ (2)

$\mu_i = E(y_i) = \dot{a}(\theta_i) \Rightarrow \theta_i(\vec{\beta}) \Rightarrow$

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

$\frac{\partial \ell(\cdot)}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial \ell(\cdot)}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \beta_j} \overset{(*)}{=} \left[\frac{\partial \ell(\cdot)}{\partial \theta_i} = \right.$

$= \frac{1}{\varphi} \left(\frac{y_i - \dot{a}(\theta_i)}{a_i} \right); \quad \frac{\partial \theta_i}{\partial \beta_j} = \left[\theta_i(\vec{\beta}) = \theta_i(\eta(\vec{\beta})) \right] =$

$= \frac{\partial \theta_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \theta_i}{\partial \eta_i} X_{ij}; \quad \left(f^{-1}(x) \right)' = \frac{1}{f'(x)};$

$\left(\frac{\partial \theta_i}{\partial \eta_i} \right)^{-1} = \frac{\partial \eta_i}{\partial \theta_i} = \left[\eta_i(\mu_i(\theta_i)) \right] = \frac{\partial \eta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \theta_i} =$ ($\dot{g}(\cdot)$, $\ddot{a}(\cdot)$)

$\mu_i = \dot{a}(\theta_i), \ \frac{\partial \mu_i}{\partial \theta_i} = \ddot{a}(\theta_i)$

$= \dot{g}(\mu_i) \cdot \underbrace{\ddot{a}(\theta_i)}_{V(\mu_i)} = \dot{g}(\mu_i) V(\mu_i); \quad \frac{\partial \theta_i}{\partial \eta_i} = \frac{1}{\dot{g}(\mu_i) V(\mu_i)}$

return to (*)

$\sum_{i=1}^{n} \frac{\partial \ell(\cdot)}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \beta_j} = \frac{1}{\varphi} \sum_{i=1}^{n} \frac{y_i - \dot{a}(\theta_i)}{a_i} \cdot \frac{X_{ij}}{\dot{g}(\mu_i) V(\mu_i)} = 0$

$\Leftrightarrow [\mu_i = \mu_i(\vec{\beta})] \Leftrightarrow \frac{\partial \ell(\cdot)}{\partial \beta_j} = 0$

sep 17-10:27

$$\cancel{\frac{1}{\phi}} \sum \frac{y_i - \mu_i(\vec\beta)}{a_i} \cdot \frac{1}{g'(\mu_i(\vec\beta))\, V(\mu_i(\vec\beta))}\, x_{ij} = 0 \quad (3) \quad \forall j \in \{1,\dots,p\}$$

not important!

MLE has the invariance property:

$$l(\cdot) \xrightarrow[\vec\beta]{} \max : \quad \boxed{\hat{\vec\beta}_{MLE} = \arg\max_{\vec\beta} l(\cdot)}$$

$$\vec\gamma = h(\vec\beta) \Rightarrow \hat{\vec\gamma} = h(\hat{\vec\beta}_{MLE})$$

$\hat\beta_{MLE}$ is known $\Longrightarrow$

$$\hat\theta_i = \dot a^{-1}\left(g^{-1}\left(x_i^T \hat\beta_{MLE}\right)\right)$$

in our case for example.

sep 17-10:44

c) parametrization of Gamma

$$f(y|\mu,\nu) = \frac{y^{\nu-1}}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^{\nu} \exp\left(-\frac{\nu y}{\mu}\right)$$

$$f(y|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\, y^{\alpha-1} \exp(-\beta y), \quad (4)$$

where $\alpha = \nu$, $\mu = \frac{\alpha}{\beta}$ | proceed with 1 st:

$$\log\nu - \psi(\nu) = \sum_{i=1}^{n} (z_i - \log z_i - 1)$$

$$z_i = \frac{y_i}{\hat\mu_i}; \quad \psi(\nu) = \frac{\partial \log\Gamma(\nu)}{\partial\nu}$$

di Gamma function

$$\log f(y_1,\dots,y_n|\nu,\mu_i) = \text{independence}$$

$$= \log\prod f(y_i|\nu_i,\mu_i) = \sum_{i=1}^{n} \log f(y_i|\nu_i,\mu_i) =$$

$$= \sum_{i=1}^{n}\left[(\nu-1)\log y_i - \log\Gamma(\nu) + \nu\log\nu - \nu\log\mu_i - \frac{\nu \cdot y_i}{\mu_i}\right] \Leftrightarrow$$

$$\Leftrightarrow \frac{\partial l(\cdot)}{\partial\mu_i} = -\frac{\nu}{\mu_i} + \frac{\nu y_i}{\mu_i^2} = 0 \Leftrightarrow$$

$$\frac{\cancel{\nu}}{\mu_i} = \frac{\cancel{\nu}\, y_i}{\mu_i^2} \Leftrightarrow \hat\mu_i = y_i$$

do not depend on ν,

sep 17-10:52

$\hat{\mu}_i$ do not depend on ϑ => we first find $\hat{\mu}_i = \underset{\mu_i}{\arg\max}\, \ell(\cdot)$, then plug $\hat{\mu}$ into $\ell(\cdot)$ and find stationary points w.r.t ϑ ⑤

$$\frac{\partial \ell(\cdot)}{\partial \vartheta} = \sum_{i=1}^{n} \log y_i - \Psi(\vartheta) + 1 + \log\vartheta - \log\hat{\mu}_i - \frac{y_i}{\hat{\mu}_i} = \left[\frac{y_i}{\hat{\mu}_i} = z_i\right] =$$

$$\sum_{i=1}^{n} \log z_i - \Psi(\vartheta) + 1 + \log\vartheta - z_i = 0$$

$$\Leftrightarrow \Psi(\vartheta) - \log\vartheta = \frac{\sum_{i=1}^{n} 1 + \log z_i - z_i}{n}$$

sep 17-10:59

S6 $E\left(-\frac{\partial^2(\ell)}{\partial\beta_j\,\partial\varphi}\right) = I_{\beta_j,\varphi} = 0$ ⑥

$$\frac{\partial \ell(\cdot)}{\partial \beta_j} = \frac{1}{\varphi}\sum_{i=1}^{n} \frac{y_i - \mu_i}{1} \cdot \frac{x_{ij}}{g(\mu_i)V(\mu_i)}$$

$$\frac{\partial^2 \ell(\cdot)}{\partial\beta_j\,\partial\varphi} = -\frac{1}{\varphi^2}\sum \frac{y_i-\mu_i}{1}\cdot\frac{x_{ij}}{g(\mu_i)V(\mu_i)}$$

$$E\left\{-\frac{\partial^2 \ell(\cdot)}{\partial\beta_j\,\partial\varphi}\right\} = E\left\{\frac{1}{\varphi^2}\sum_{i=1}^{n}\frac{y_i-\mu_i}{1}\cdot\frac{x_{ij}}{g(\mu_i)V(\mu_i)}\right\}$$

$$= \frac{1}{\varphi^2}\sum_{i=1}^{n}\frac{E(y_i)-\mu_i}{1}\cdot\frac{x_{ij}}{g(\mu_i)V(\mu_i)} = [E(y_i) = \mu_i] =$$

$$= 0 = I_{\beta_j,\varphi} \Rightarrow$$

β_j & φ are independent $\forall\, j \in \{0,\dots,p\}$

sep 17-11:22

S7 $I_{\beta_i \beta_j} = J_{\beta_i \beta_j} \quad \forall\, i, j \in \{0, \dots p\}$ 7

$\eta_i = \theta_i = x_i^T \beta = \dot{a}^{-1}(\mu_i)$ - for canonical link!

$\ell(\varphi, \vec{\beta}, y) = \sum_{i=1}^{n} \left[\frac{y_i \theta_i - a(\theta_i)}{\varphi} - C(y_i, \varphi) \right] =$

$= [\theta_i = x_i^T \beta] = \sum \left[\frac{y_i x_i^T \beta - a(x_i^T \beta)}{\varphi} + C(y_i, \varphi) \right] \Leftrightarrow \frac{\partial \ell(\cdot)}{\partial \beta_j} = \frac{1}{\varphi} \sum_{i=1}^{n} \left(y_i x_{ij} - \dot{a}(x_i^T \beta) x_{ij} \right)$

$y_i x_{ij}$ — does not depend on β_j

$\dot{a}(x_i^T \beta) x_{ij}$ — does not depend on y_i

$$\frac{\partial^2 \ell(\cdot)}{\partial \beta_j \partial \beta_k} = \boxed{-\frac{1}{\varphi} \sum_{i=1}^{n} \ddot{a}(x_i^T \beta) x_{ij} x_{ik}}$$

sep 17-11:27

$$E\left\{ -\frac{\partial^2 \ell(\cdot)}{\partial \beta_j \partial \beta_k} \right\} = E\left\{ \frac{1}{\varphi} \sum_{i=1} \ddot{a}(x_i^T \beta) x_{ij} x_{ik} \right\}$$

8

$$= \frac{1}{\varphi} \sum_{i=1}^{n} \ddot{a}(x_i^T \beta) x_{ij} x_{ik} =$$

$$= -\frac{\partial^2 \ell(\cdot)}{\partial \beta_j \partial \beta_k} \Leftrightarrow \underbrace{I_{\beta_j \beta_k}}_{\text{Fisher}} = \underbrace{J_{\beta_j \beta_k}}_{\text{Hessian}}$$

$\forall\, j, k \in \{0, \dots p\}$

sep 17-11:43

(14) comments — we will look at 1 dim. example

Taylor Expansion (9)

$$\ell(\beta+\delta) \approx \ell(\beta) + \dot\ell(\beta)\,\delta + \ddot\ell(\beta)\frac{\delta^2}{2} =: \ell^*(\beta+\delta)$$

$$\max_{\delta} \ell^*(\beta+\delta) \Leftrightarrow \frac{\partial \ell^*(\cdot)}{\partial \delta} = 0$$

$$\dot\ell(\beta) + \delta\,\ddot\ell(\beta) = 0 \Leftrightarrow$$

$$\delta = -\frac{\dot\ell(\beta)}{\ddot\ell(\beta)} \quad \text{- step}$$

$$\beta^{(k+1)} = \beta^{(k)} + \delta^{(k)} = \beta^{(k)} - \left(\ddot\ell(\beta^{(k)})\right)^{-1}\dot\ell(\beta^{(k)})$$

full dimensional case $\Leftrightarrow$

$$\vec\beta^{(k+1)} = \beta^{(k)} + J(\beta^{(k)})^{-1} S(\beta^{(k)})$$

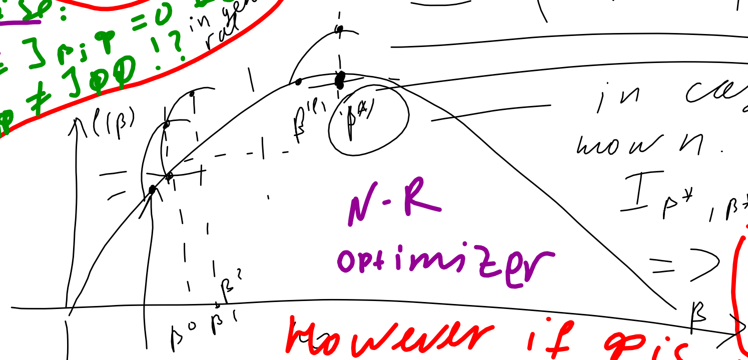
sep 17-11:46

$$= [\varphi - \text{is known}] =$$ (10)

$$= \beta^{(k)} + I(\beta^{(k)})^{-1} \cdot S(\beta^{(k)}) \quad (**)$$

in case φ is known.

Exercise: show $I_{\beta,\varphi} = J_{\beta,\varphi} = 0$!? in general & $I_{\varphi\varphi} \neq J_{\varphi\varphi}$



N-R optimizer

$$I_{\beta^*,\beta^*} = J_{\beta^*,\beta^*} \Rightarrow Var(\beta^*) = I^{-1}_{\beta^*\beta^*} \;!!!$$

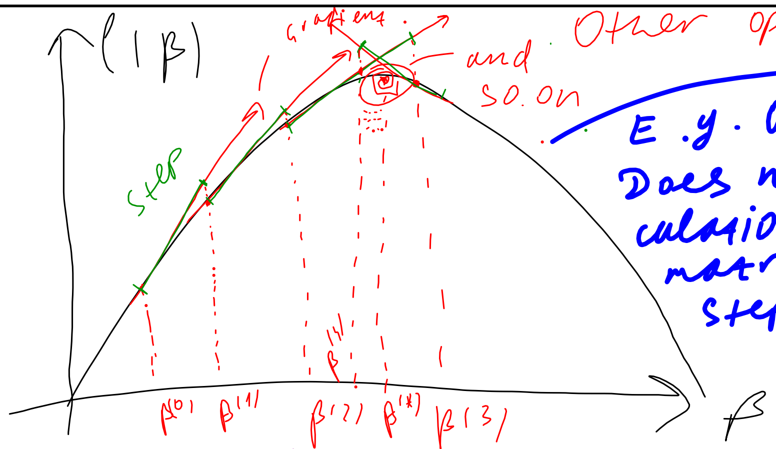
However if φ is unknown

$$J = \begin{pmatrix} J_{\beta,\beta} & 0 \\ 0 & J_{\varphi\varphi} \end{pmatrix} \neq I = \begin{pmatrix} I_{\beta,\beta} & 0 \\ 0 & I_{\varphi\varphi} \end{pmatrix}$$

generally speaking

but $J \xrightarrow{n\to\infty} I$

sep 17-11:52



Other optimizations (11)

E.g. On one hand Does not require calculations of Hessian matrix at every step, but does not automatically provide us with $Var(\beta^*)$

Grad. Optimizer

Ideas:

- replace J with $\bar{I}$ in (**)
- proceed some steps with gradient opt. then switch to Newton-Rapson
- others...

17.09.2015

sep 17-11:58