General Estimation and Testing Theory (1)

D.R. Brillinger notes

26 Sept 01

Nonlinear Estimation.

Stochastic model with parameter θ

Data $\{y_i\}$

I. $$\min_{\theta} \sum_{i=1}^{n} \rho_i(y_i|\theta)$$

II. $$\sum_i \underset{\sim}{\psi}_i(y_i|\hat{\theta}) = \underset{\sim}{0}$$

where $$\underset{\sim}{\psi}_i(y_i|\theta), \text{ eg. } = \frac{\partial}{\partial\theta}\rho_i(y_i|\theta)$$

$$\gamma_i(\theta) = E\,\rho_i(Y_i|\theta), \quad \frac{1}{n}\sum_i \gamma_i(\theta) \to \gamma(\theta)$$

$$\gamma(\theta) > \gamma(\theta_0) \qquad \forall\, \theta \neq \theta_0$$

$$\underset{\sim}{\lambda}_i(\theta) = E\,\underset{\sim}{\psi}_i(Y_i|\theta), \quad \frac{1}{n}\sum_i \lambda_i(\theta) \to \lambda(\theta)$$

$$\underset{\sim}{\lambda}(\theta_0) = \underset{\sim}{0}, \quad |\underset{\sim}{\lambda}(\theta)| > 0 \text{ if } \theta \neq \theta_0$$

<u>Theorem</u>. Under regularity conditions 26 Sept 01

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\underset{\sim}{0}, \underset{\sim}{\Lambda}^{-1} \underset{\sim}{C} \underset{\sim}{\Lambda}^{-1})$$

$$\underset{\sim}{C}_i = \operatorname{Var}\{\underset{\sim}{\psi}_i(Y_i|\theta_0)\}$$

$$\underset{\sim}{C} = \lim_{n\to\infty} \frac{1}{n} \sum_i \underset{\sim}{C}_i$$

$$\underset{\sim}{\Lambda}_i = \left.\frac{\partial \underset{\sim}{\lambda}_i(\theta)}{\partial \theta}\right|_{\theta_0}$$

$$\underset{\sim}{\Lambda} = \lim_{n\to\infty} \frac{1}{n} \sum_i \underset{\sim}{\Lambda}_i$$

$$\hat{\underset{\sim}{C}} = \frac{1}{n} \sum_{i=1}^{n} \underset{\sim}{\psi}_i(y_i|\hat{\theta})\, \underset{\sim}{\psi}_i(y_i|\hat{\theta})^T$$

$$\hat{\underset{\sim}{\Lambda}} = \frac{1}{n} \sum_{i=1}^{n} \left.\frac{\partial \underset{\sim}{\psi}_i(y_i|\theta)}{\partial \theta}\right|_{\hat{\theta}}$$

Particular Cases (of theorem) 28 Sept 01

1. Nonlinear regression - correct model

Whence $\mu_i()$?

$$Y_i = \mu_i(\theta) + \epsilon_i, \quad \text{true } \theta_0$$

$$\rho_i(y_i|\theta) = [y_i - \mu_i(\theta)]^2$$

$$\underset{\sim}{\psi}_i(y_i|\theta) = -2[y_i - \mu_i(\theta)]\frac{\partial\mu_i(\theta)}{\partial\theta}$$

(Drop "−2")

$$\begin{aligned}\gamma_i(\theta) &= E[Y_i - \mu_i(\theta)]^2 \\ &= E[Y_i - \mu_i(\theta_0) + \mu_i(\theta_0) - \mu(\theta)]^2 \\ &= \sigma^2 + [\mu_i(\theta_0) - \mu(\theta)]^2\end{aligned}$$

if $E\epsilon_i = 0$, $\text{Var}\ \epsilon_i = \sigma^2$

Minimized at $\theta = \theta_0$

(Unique for identifiability)

28 Sept 01

$$\lambda_i(\theta) = E\,[Y_i - \mu_i(\theta)]\,\frac{\partial \mu_i(\theta)}{\partial \theta}$$

$$= [\mu_i(\theta_0) - \mu_i(\theta)]\,\frac{\partial \mu_i(\theta)}{\partial \theta} \quad \text{neglecting}^2$$

Value 0 at $\theta = \theta_0$.

(Identifiability)

$$\underset{\sim}{C}_i = \mathrm{Var}\,[Y_i - \mu_i(\theta)]\,\frac{\partial \mu_i(\theta)}{\partial \theta} \qquad \text{at } \theta = \theta_0$$

$$= \sigma^2\,\frac{\partial \mu_i(\theta_0)}{\partial \theta_0}\,\frac{\partial \mu_i(\theta_0)}{\partial \theta_0}^T$$

$$\underset{\sim}{\Lambda}_i = -\frac{\partial \mu_i(\theta_0)}{\partial \theta_0}\,\frac{\partial \mu_i(\theta_0)}{\partial \theta_0}^T + [\mu_i(\theta_0) - \mu_i(\theta)]\,\frac{\partial^2 \mu_i(\theta_0)}{\partial \theta\,\partial \theta^T}$$

$$= -\frac{\partial \mu_i(\theta_0)}{\partial \theta_0}\,\frac{\partial \mu_i(\theta_0)}{\partial \theta_0}^T$$

26 Sept 01

$$\underset{\sim}{C} = \sigma^2 \lim_{n\to\infty} \frac{1}{n} \sum \frac{\partial \mu_i(\theta_0)}{\partial \theta_0} \frac{\partial \mu_i(\theta_0)}{\partial \theta_0}^T$$

$$\underset{\sim}{\Lambda} = - \lim_{n\to\infty} \frac{1}{n} \sum \frac{\partial \mu_i(\theta_0)}{\partial \theta_0} \frac{\partial \mu_i(\theta_0)}{\partial \theta_0}^T$$

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N(\underset{\sim}{0}, \sigma^2 \underset{\sim}{\Lambda}^{-1})$$

$$\hat{\underset{\sim}{\Lambda}} = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \mu_i(\hat{\theta})}{\partial \hat{\theta}} \frac{\partial \mu_i(\hat{\theta})}{\partial \hat{\theta}}^T$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} [y_i - \mu_i(\hat{\theta})]^2$$

There is an S function nls()
Program more efficient if recognize linear parameters

28 Sept 01

Some details.

Compute via OLS

$y_i = \mu_i(\theta) + \epsilon_i$

True θ_0

Suppose starting value $\hat{\theta}_-$ near θ_0

$$y_i \simeq \mu_i(\theta) \simeq \mu_i(\hat{\theta}_-) + (\theta - \hat{\theta}_-) \left.\frac{\partial \mu_i}{\partial \theta}\right|_{\hat{\theta}_-}$$

linear model

Regress, $\hat{\beta} \doteq (\theta - \hat{\theta}_-)$

Next starting value $\hat{\theta}_- + \hat{\beta}$ (a correction)

Iterate til: $|\hat{\theta} - \hat{\theta}_-| \leq \varepsilon$ $\hat{\theta}_-$ updating

or

$|SS(\hat{\theta}) - SS(\hat{\theta}_-)| \leq \eta$

or both

Splus nls( )

26 Sept 01

Assess via residuals $y_i - \hat{y}_i = y_i - \mu_i(\hat{\theta})$

1. Normal prob plot
2. Index plot
3. Vs explanatories and possible explanatories
4. . . .

Difficulties

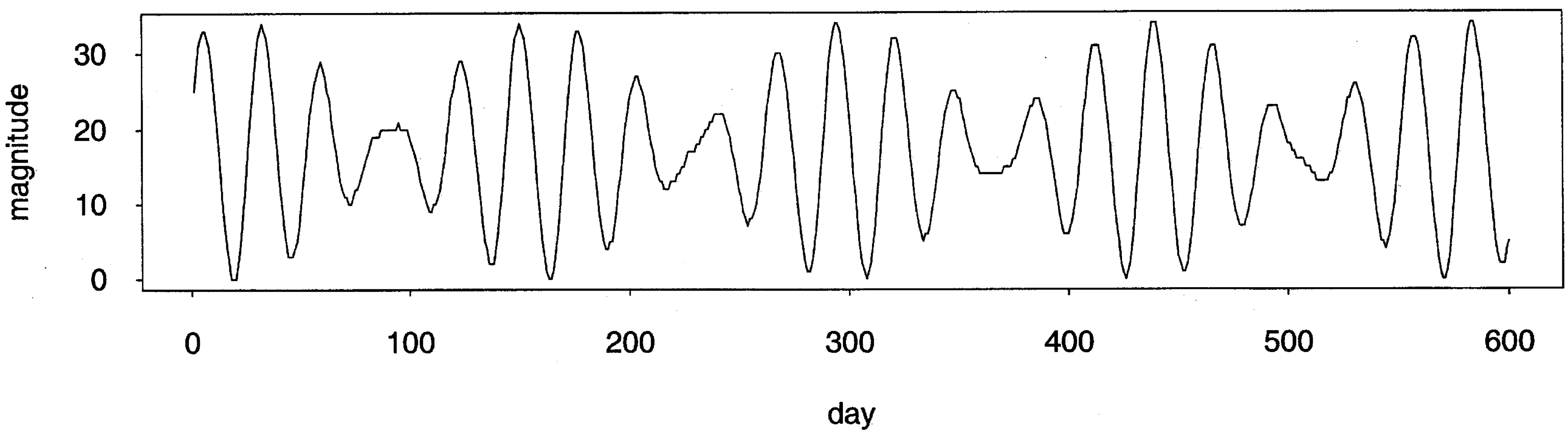
1. $\left.\dfrac{\partial \mu_i(\theta)}{\partial \theta}\right|_{\theta_0} = 0$ ridge regression

2. initial value

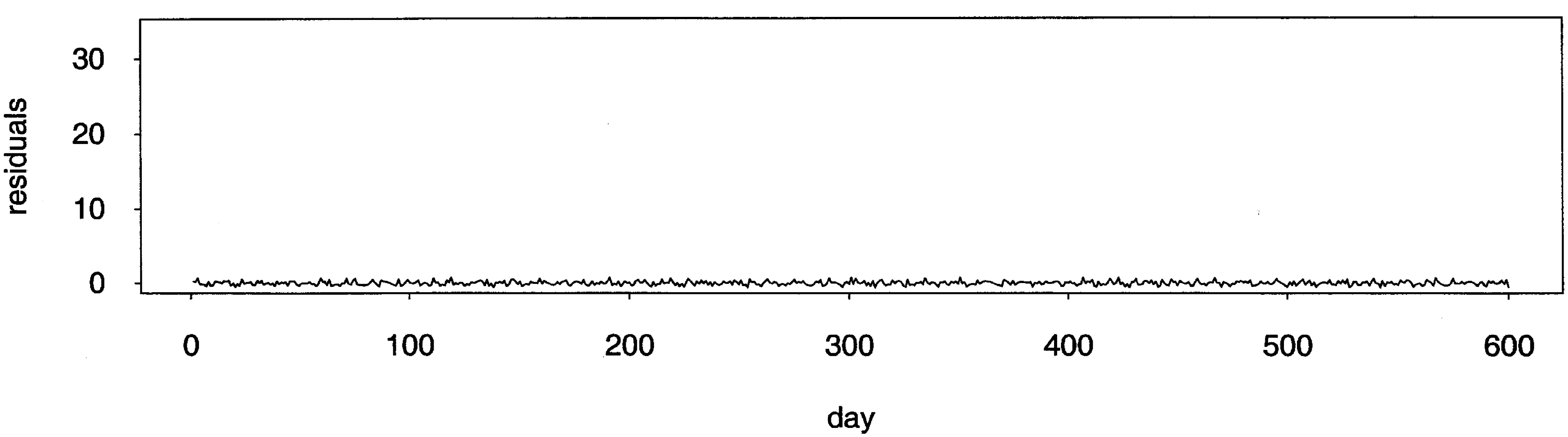
Will see later what happens when model incorrect, i.e. EY_i not of form $\mu_i(\theta)$

Constructing the regression function - NMR

Magnitude of variable star at midnight



Residuals



```
postscript(file="Po")
par(mfrow=c(2,1))
library(mass)
mag<-scan("../Data/star")
day<-1:600
plot(day,mag,type="l",main="Magnitude of variable star at midnight",xlab="day",ylab="ma
gnitude",las=1)
junk2<-data.frame(mag=mag,day=day)
junk3<-c(th1=2*pi/24,th2=2*pi/29)
junk4<-nls(mag~cbind(1,cos(th1*day),sin(th1*day),cos(th2*day),sin(th2*day)),algorithm="
plinear",data=junk2,start=junk3,trace=T)
plot(junk4$res,type="l",ylim=range(mag),main="Residuals",xlab="day",ylab="residuals",la
s=1)
graphics.off()
q()
Residual sum of squares : 54.67214
parameters:
       th1       th2
 0.2617983 0.2166595 17.08578 -1.832544 6.843451 6.073713 7.982854
formula: mag ~ cbind(1, cos(th1 * day), sin(th1 * day), cos(th2 * day), sin(th2 * day))

600 observations
```

$$y_t \sim \alpha + \beta_1 \cos \theta_1 t + \beta_2 \sin \theta_1 t + \beta_3 \cos \theta_2 t + \beta_4 \sin \theta_2 t + \epsilon_t$$

```
> stem(junk4$res)

N = 600   Median = -0.01585285
Quartiles = -0.2314647, 0.2055498

Decimal point is 1 place to the left of the colon

  -6 : 100
  -5 : 998888877666555
  -5 : 44321
  -4 : 977775555
  -4 : 444333333222221111111000
  -3 : 999999999888887766655555
  -3 : 4444433322211111111000
  -2 : 999999998888777777776666555555555555
  -2 : 444444444333333332221111000
  -1 : 999998888888777777666666666655555555
  -1 : 44444444443333222222211111100000000
  -0 : 99999999999988888888887777777666666555555
  -0 : 44433333333333222222222211110000000
   0 : 00000111111112222223333344444
   0 : 5555556666666666777777777778888888888889999999
   1 : 0000000111111122222333344444444
   1 : 555555566667777777777788889
   2 : 00000111111222222222333333444444444
   2 : 555555666667777777888899999
   3 : 0000012223
   3 : 555555666677777888999999
   4 : 00112233344
   4 : 566677778
   5 : 123333444
   5 : 66667899
   6 : 023
   6 : 6677889
   7 : 0133
   7 : 567899
   8 : 23
```

26 Sept. 01

More to nonlinear regression computations.

Can use OLS

$y_i = \mu_i(\theta) + \epsilon_i$ true θ_0

eg $\mu_i(\theta) = \alpha e^{-\beta x}$ $\theta = (\alpha, \beta)$

$\hat{\theta}$ starting value, near θ_0

$y_i \approx \mu_i(\hat{\theta}) + (\theta - \hat{\theta}) \left.\frac{\partial \mu_i}{\partial \theta}\right|_{\hat{\theta}}$

by OLS $\hat{\beta} = \hat{\theta} - \hat{\theta}$ Gauss-Newton

next starting $\theta + \hat{\beta}$ i.e. correction

Continue until convergence

eg $|\hat{\theta} - \hat{\theta}_-| \le \varepsilon$

or $|SS(\hat{\theta}) - SS(\hat{\theta}_-)| \le \eta$

Uncertainties. cp. $y_i = \alpha + \beta x_i + \epsilon_i$ $x_i = \frac{\partial f_i}{\partial \theta}$

$\hat{\beta} = \sum y_i (x_i - \bar{x}) / \sum (x_i - \bar{x})^2$

$\text{var}\,\hat{\beta} \sim \sigma^2 \left[\sum \left(\frac{\partial \mu_i}{\partial \theta}\right)\left(\frac{\partial \mu_i}{\partial \theta}\right)' \right]^{-1}$ comes from lm( )

$\hat{\sigma}^2 = \sum_i [y_i - \mu_i(\hat{\theta})]^2 / (n-p)$

27 Sept 01

Nonlinear regression - incorrect model H. White (8 JASA

Model mean is $h(\underset{\sim}{x}_i \mid \underset{\sim}{\beta})$, but

$$E\{Y_i \mid \underset{\sim}{x}_i\} = g(\underset{\sim}{x}_i)$$

$$\rho_i(y_i \mid \theta, \underset{\sim}{x}_i) = |y_i - h(\underset{\sim}{x}_i \mid \beta)|^2$$

$$E\,\rho_i(y_i \mid \theta, \underset{\sim}{x}_i) = E\{|Y_i - g(\underset{\sim}{x}_i) + g(\underset{\sim}{x}_i) - h(\underset{\sim}{x}_i \mid \beta)|^2$$

$$= \sigma^2 + |g(\underset{\sim}{x}_i) - h(\underset{\sim}{x}_i \mid \beta)|^2$$

Looking for

$$\min_\beta \frac{1}{n}\sum_{i=1}^{n} |g(\underset{\sim}{x}_i) - h(\underset{\sim}{x}_i \mid \beta)|^2$$

$$\sim \min_\beta \int |g(\underset{\sim}{x}) - h(\underset{\sim}{x} \mid \beta)|^2 \, dF(\underset{\sim}{x})$$

Call this β_0. Then

$$\hat{\beta} \xrightarrow{p} \beta_0$$

27 Sept 01

Looking for $h(\underset{\sim}{x} \mid \beta)$ closest to $g(\underset{\sim}{x})$

in norm

$$\int |g(\underset{\sim}{x}) - h(\underset{\sim}{x} \mid \beta)|^2 \, dF(\underset{\sim}{x})$$

Later we will be using K-L distance in model selection.

Heuristics of proof: 28 Sept 01

Asymptotic normality.

By Taylor series expansion.

$$\underset{\sim}{\lambda}_i(\theta) = E_{Y_i}\{\psi_i(Y_i|\theta)\}$$

$$\underset{\sim}{\Lambda}_i(\theta) = \frac{\partial \underset{\sim}{\lambda}_i(\theta)}{\partial \theta}$$

$$\frac{1}{n}\sum_{i=1}^{n} \underset{\sim}{\Lambda}_i(\theta) \rightarrow \underset{\sim}{\Lambda}(\theta), \quad \underset{\sim}{\Lambda} = \underset{\sim}{\Lambda}(\theta_0)$$

$$\underset{\sim}{C}_i = \mathrm{var}\{\underset{\sim}{\psi}_i(\underset{\sim}{Y}_i|\theta_0)\}$$

$$= E\, \underset{\sim}{\psi}_i \underset{\sim}{\psi}_i^T \quad \text{as } E\underset{\sim}{\psi}_i(Y_i|\theta_0) = \underset{\sim}{0}$$

$$\frac{1}{n}\sum \underset{\sim}{C}_i \rightarrow \underset{\sim}{C}$$

$$\underset{r\times 1}{\underset{\sim}{0}} = \sum_i \underset{r\times 1}{\underset{\sim}{\psi}_i}(y_i|\hat{\theta})$$

$$\approx \sum_i \underset{\sim}{\psi}_i^{r\times 1}(y_i|\theta_0) + \sum_i \frac{\partial \underset{\sim}{\psi}_i}{\partial \hat{\theta}}\underset{r\times 1}{(\hat{\theta}-\theta_0)}$$

28 Sept 01

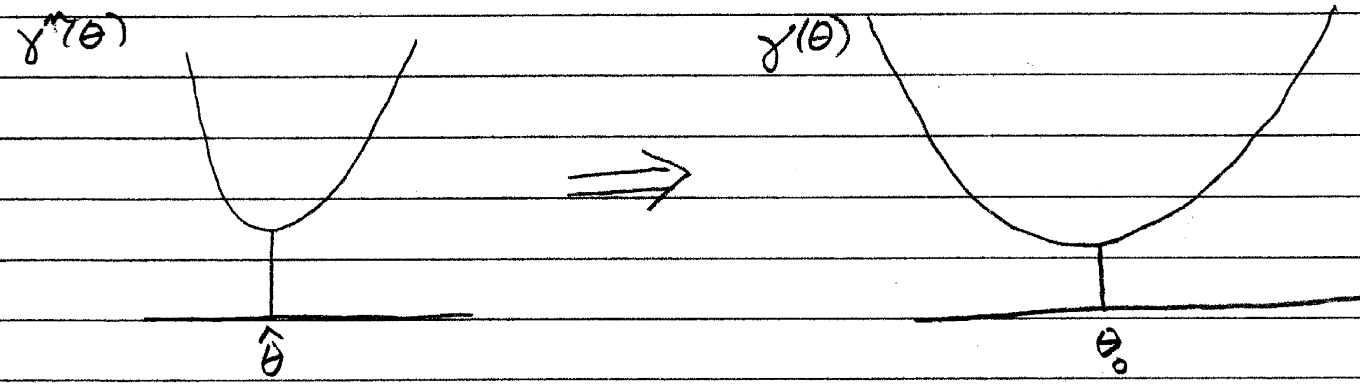
$$\underset{r\times 1}{\sqrt{n}\left(\hat{\theta}-\theta_0\right)} \sim \left[\frac{1}{n}\sum \frac{\partial \underset{\sim}{\psi}_i}{\partial \theta'}(y_i \mid \theta_0)\right]^{-1} \left[\frac{1}{\sqrt{n}} \sum \underset{\sim}{\psi}_i (y_i \mid \theta_0)\right]$$

$$\sim \underset{\sim}{\Lambda}^{-1} N_r(\underset{\sim}{0}, \underset{\sim}{C})$$

28 Sept 01

Consistency

$$\gamma^n(\theta) = \frac{1}{n}\sum_{i=1}^{n} \rho_i(y_i|\theta)$$



so anticipate that, under conditions, $\hat{\theta} \Rightarrow \theta_0$

Eg. $\theta \in \Theta$ compact, continuity

Via Law of Large Numbers in a function space

$$\sup_{\theta \in \Theta} |\gamma^n(\theta) - \gamma(\theta)| \xrightarrow{p} 0$$

ie $\varepsilon' > 0$

$$\text{Prob}\{\sup_{\theta \in \Theta} |\gamma^n(\theta) - \gamma(\theta)| > \varepsilon'\} \to 0 \quad \text{as } n \to \infty$$

28 Sept 01

Need to show $\forall \ \varepsilon > 0$

$$\text{Prob}\{ |\hat{\theta} - \theta_0| > \varepsilon \} \longrightarrow 0$$

Set $\Omega = \{ \theta \mid |\theta - \theta_0| > \varepsilon \}$

$\text{Prob}\{ |\hat{\theta} - \theta_0| > \varepsilon \}$

$$\leq \text{Prob}\left\{ \inf_{\theta \in \Omega} \frac{1}{n} \sum_i \rho(Y_i | \theta) < \frac{1}{n} \sum_i \rho(Y_i | \theta_0) \right\} \quad (*)$$

$[B$ implies $A \Rightarrow P(A) \leq P(B)]$

$$(*) = \text{Prob}\left\{ \inf_{\theta \in \Omega} \left\{ \frac{1}{n} \sum_i \rho(Y_i | \theta) - \gamma(\theta) \right\} < \left\{ \frac{1}{n} \sum_i \rho(Y_i | \theta_0) - \gamma(\theta_0) \right\} + \{ \gamma(\theta_0) - \gamma(\theta) \} \right\}$$

The last term here is negative and the first two can be made arbitrarily close to 0 by picking n large, so the probability can be made arbitrarily small.

1 Oct. 01

MLE (model correct)

$$Y_i \; : \; f_i(y|\theta) \qquad \text{true } \theta_0$$

$$\rho_i(y_i|\theta) = -\log f_i(y_i|\theta)$$

$$\underset{\sim}{\psi}_i(y_i|\theta) = -\frac{\partial}{\partial\theta}\log f_i(y_i|\theta)$$

$$= -\frac{1}{f_i(y_i|\theta)}\frac{\partial f_i(y_i|\theta)}{\partial\theta}$$

$$\lambda_i(\theta) = -\int \frac{1}{f_i(y|\theta)}\frac{\partial f_i(y|\theta)}{\partial\theta} f_i(y|\theta_0)\,dy \qquad (*)$$

Now $\displaystyle\int f_i(y|\theta)\,dy = 1$, so

$$\int \frac{\partial f_i(y|\theta)}{\partial\theta}\,dy = \underset{\sim}{0}$$

so

$$\underset{\sim}{\lambda}_i(\theta_0) = \underset{\sim}{0}$$

$$\underset{\sim}{C}_i = \mathrm{var}\{\underset{\sim}{\psi}_i(Y_i|\theta_0)\}$$

$$= E_0\left\{ \frac{\partial \log f_i(Y_i|\theta_0)}{\partial \theta} \frac{\partial \log f_i(Y_i|\theta_0)}{\partial \theta'} \right\}$$

~~differentiate~~ $$\int \frac{\partial \log f_i(y|\theta)}{\partial \theta} f_i(y|\theta)\,dy = \underset{\sim}{0}$$

$$\int \frac{\partial^2 \log f_i(y|\theta)}{\partial \theta\,\partial \theta'} f_i(y|\theta)\,dy$$

$$+ \int \frac{\partial \log f_i(y|\theta)}{\partial \theta} \frac{\partial \log f_i(y|\theta)}{\partial \theta'} f_i(y|\theta)\,dy = \underset{\sim}{0}$$

So $$\underset{\sim}{C}_i = -E\left\{ \left. \frac{\partial^2 \log f_i(y|\theta)}{\partial \theta\,\partial \theta'} \right|_0 \right\}$$

$$= \underset{\sim}{I}_i(\theta_0)$$ Fisher information

1 Oct 01

$$\underset{\sim}{\Lambda}_i = \frac{\partial \lambda_i(\theta)}{\partial \theta}\Big|_0$$

$$= -\int \frac{\partial^2 \log f_i(y|\theta)}{\partial\theta\,\partial\theta'}\Big|_0 f_i(y|\theta_0)\,dy$$

from *

$$= \underset{\sim}{I}_i(\theta_0)$$

$$\underset{\sim}{I}(\theta_0) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \underset{\sim}{I}_i(\theta_0)$$

So

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\underset{\sim}{0}, \underset{\sim}{I}(\theta_0)^{-1})$$

Can estimate via

$$\hat{\underset{\sim}{C}} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f_i(y_i|\hat{\theta})}{\partial\theta} \frac{\partial \log f_i(y_i|\hat{\theta})}{\partial\theta}^T$$

$$\geq \underset{\sim}{0}$$

Splus function nlmd)

1 Oct 01

MLE: model incorrect (Cox). Suppose true distribution is P. Given $f_i(Y|\theta)$: model

Suppose θ_* maximizes

$$E_p\{\log f_i(Y|\theta)\}$$

or

$$E_p\left\{\frac{\partial \log f_i(Y|\theta)}{\partial \theta}\right\}\Bigg|_{\theta_*} = \underset{\sim}{0}$$

Theorem.

$$\sqrt{n}(\hat{\theta}-\theta_*) \to N(\underset{\sim}{0}, \underset{\sim}{\Lambda}^{-1}\underset{\sim}{C}\underset{\sim}{\Lambda}^{-1})$$

$$\underset{\sim}{C} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E\left\{\frac{\partial \log f_i(Y|\theta)}{\partial \theta}\frac{\partial \log f_i(Y|\theta)}{\partial \theta}'\right\}\Bigg|_*$$

$$\underset{\sim}{\Lambda} = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} E\left\{-\frac{\partial^2 \log f_i(Y|\theta)}{\partial \theta\,\partial \theta'}\right\}\Bigg|_*$$

1 Oct 01

What is this $f_i(y|\theta_*)$?

Suppose given densities p and q

$$\mathbb{K}(p;q) = \int p(y) \log \frac{p(y)}{q(y)} dy = E_p \log \frac{p(Y)}{q(Y)}$$

A measure of the goodness of fit of q to p.

Kullback-Leibler distance (risk, information, divergence)
The discrimination information or negative entropy of p wrt q.

Properties.

$$\mathbb{K}(p;p) = 0$$

$$\mathbb{K}(p;q) > 0 \quad q \neq p$$
$$= 0 = \text{iff } p = q \text{ a.e.}$$

Proof.

Not symmetric
Does not satisfy triangle inequality