11 Oct 2001

Back to confidence and testing

Likelihood ratio test and extension of anova

Prob function $f(y|\theta)$ $\theta \in \Theta$

Data $y_1, \ldots, y_n$

Null hypothesis $H_0: \theta \in \Theta_0 \subseteq \Theta$ nested

Likelihood ratio statistic

$$\lambda_n = \frac{\max_{\theta \in \Theta_0} \prod_1^n f(y_i|\theta)}{\max_{\theta \in \Theta} \prod_1^n f(y_i|\theta)}$$

$$-2 \log \lambda_n = 2[\ell(\hat{\theta}) - \ell(\hat{\underline{\theta}}_0)]$$ difference of deviances

as $n \to \infty$

$$\sim \chi^2_{k-s} \text{ or } \chi^2_r,$$ r: number of restrictions, s: number of free ~~restrictions~~

if $\Theta_0 = \{(\theta_1, \ldots, \theta_r, \theta_{r+1}, \ldots, \theta_k); \theta_j = \theta_{0j}, j = 1, \ldots, r\}$

and model in family, i.e. model correct

There may be a sequence of nested hypotheses

$$\Theta_1 \subset \Theta_2 \subset \Theta$$

$$\ell(\theta) = \sum_i \log f(y_i|\theta)$$

11 Oct. 2001

Proof of χ^2,

a) full mle, k unknowns

Set $\underset{\sim}{D}^{k\times 1} = \sqrt{n}(\hat{\theta} - \theta_0) \sim N_k(\underset{\sim}{0}, \underset{\sim}{I}_0^{-1})$

$$\underset{\sim}{V}^{k\times 1} = \frac{1}{\sqrt{n}} \sum_i \left. \frac{\partial \log f(y_i|\theta)}{\partial \theta} \right|_{\theta_0} \quad (*)$$

Then $\underset{\sim}{V} \sim N(\underset{\sim}{0}, \underset{\sim}{I}_0)$

where $\underset{\sim}{I}_0 \sim \mathrm{var}\left\{ \left. \frac{\partial \log f(Y|\theta)}{\partial \theta} \right|_{\theta_0} \right\}$

$$= -E\left\{ \frac{\partial^2}{\partial\theta\partial\theta^\tau} \right\}$$

By Taylor expansion

$$\underset{\sim}{0}^{k\times 1} = \sum_i \left. \frac{\partial \log f(y_i|\theta)}{\partial \theta} \right|_{\hat{\theta}} \sim \sum_i \left. \frac{\partial \log f(y_i|\theta)}{\partial \theta} \right|_{\theta_0}^{k\times 1} + \sum_i \frac{\partial^2 \log f(y_i|\theta)}{\partial\theta\partial\theta^\tau}^{k\times k} (\hat{\theta} - \theta_0)^{k\times 1}$$

So $\underset{\sim}{V} \sim \underset{\sim}{I}_0 \underset{\sim}{D}$

$\underset{\sim}{D} \sim \underset{\sim}{I}_0^{-1} \underset{\sim}{V}$

11 Oct 2001

$$\ell(\theta) = \sum_i \log f(y_i|\theta)$$

$$\ell(\theta_0) \simeq \sum_i \log f(y_i|\hat{\theta}) + \left(\sum_i \frac{\partial \log f(y_i|\theta)}{\partial \theta}\Big|_{\hat{\theta}}\right)^T (\theta_0 - \hat{\theta})$$

$$+ \frac{1}{2}(\theta_0 - \hat{\theta})^T \left(\sum_i \frac{\partial^2 \log f(y_i|\theta)}{\partial \theta\, \partial \theta^T}\right)_0 (\theta_0 - \hat{\theta})$$

So $\ell(\theta_0) \simeq \ell(\hat{\theta}) - \frac{1}{2}(\theta_0 - \hat{\theta})^T n \underset{\sim}{I}_0 (\theta_0 - \hat{\theta})$

or $2[\ell(\hat{\theta}) - \ell(\theta_0)] \sim \underset{\sim}{D}^T \underset{\sim}{I}_0 \underset{\sim}{D}$

$$\sim \chi^2_k$$

$$\sim \underset{\sim}{V}^T \underset{\sim}{I}_0^{-1} \underset{\sim}{V} \quad \text{also}$$

This result can be used to:

i) ~~examine the simple hypothesis $\theta = \theta_0$~~

ii) construct a confidence region for θ_0 (using $\hat{\underset{\sim}{I}}_0$)

iii) corresponds to $r = k$ restrictions

11 Oct. 2001

Now lets suppose that there are $k-r$ free parameters only.

Have restrictions such that the original $\theta_1, \ldots, \theta_k$ are functions of $s = k-r$ new parameters

e.g. $\theta_j = g_j(\beta_1, \ldots, \beta_s) \qquad j = 1, \ldots, k$

$r = k - s$ restrictions

Suppose, for convenience

$$\begin{bmatrix} \beta_1 \\ \vdots \\ \beta_s \end{bmatrix} = \begin{bmatrix} \theta_{r+1} \\ \vdots \\ \theta_k \end{bmatrix}$$

We will need

$$\underset{\sim}{\overset{k \times s}{M}} = \left[\frac{\partial g_i}{\partial \beta_j} \right] = \begin{bmatrix} \underset{\sim}{0} \\ \underset{\sim}{I}_s \end{bmatrix}$$

Write underscore when referring to last s coordinates of θ, eg $\underline{\theta}$, $\underline{\theta}_0$

Will repeat previous development only using last s coordinates.

11 Oct. 2001

Let $\underset{s\times 1}{\underset{\sim}{U}} = \sum_i \frac{1}{\sqrt{n}} \frac{\partial \log f(y_i|\theta)}{\partial \underline{\theta}} \Big|_{\underline{\theta}_0}$ (**)

Let $\underset{\sim}{F} = \sqrt{n}(\hat{\underline{\theta}} - \underline{\theta}_0)$

$\underset{\sim}{\mathcal{I}}_0$: information matrix for $\underline{\theta}$

As before

$$\underset{\sim}{U} \sim \underset{\sim}{\mathcal{I}}_0 \underset{\sim}{F}, \quad \underset{\sim}{F} \sim \underset{\sim}{\mathcal{I}}_0^{-1} \underset{\sim}{U}$$

$$2[\ell(\hat{\underline{\theta}}_0) - \ell(\underline{\theta}_0)] \sim \underset{\sim}{F}' \underset{\sim}{\mathcal{I}}_0 \underset{\sim}{F}$$

Lets connect $\underset{\sim}{U}$ and $\underset{\sim}{V}$

From (*) and (**) $\quad \underset{\sim}{U} = \underset{\sim}{M}' \underset{\sim}{V}$ and so

$$\underset{\sim}{\mathcal{I}}_0 = \underset{\sim}{M}' \underset{\sim}{I}_0 \underset{\sim}{M} \quad (\text{****})$$

Continuing

$$2[\ell(\hat{\underline{\theta}}_0) - \ell(\underline{\theta}_0)] \sim \underset{\sim}{F}' \underset{\sim}{\mathcal{I}}_0 \underset{\sim}{F}$$

$$\sim \underset{\sim}{U}' \underset{\sim}{\mathcal{I}}_0^{-1} \underset{\sim}{U}$$

$$\sim \underset{\sim}{V}' \underset{\sim}{M} \underset{\sim}{\mathcal{I}}_0^{-1} \underset{\sim}{M}' \underset{\sim}{V}$$

11 Oct. 2001

Subtracting

$$2\left[\ell(\hat{\theta}) - \ell(\hat{\underset{\sim}{\theta}})\right] \sim \underset{\sim}{V}'\left(\underset{\sim}{I}_0^{-1} - \underset{\sim}{M}\,\underset{\sim}{D}_0^{-1}\,\underset{\sim}{M}'\right)\underset{\sim}{V} \qquad \text{(***)}$$

But $\underset{\sim}{V} \sim N_k(\underset{\sim}{0}, \underset{\sim}{I}_0)$

write $\underset{\sim}{V} = \underset{\sim}{I}_0^{1/2}\,\underset{\sim}{Z}$ with $\underset{\sim}{Z} \sim N(\underset{\sim}{0}, \underset{\sim}{I})$

So

$$\text{(***)} = \underset{\sim}{Z}'\left(\underset{\sim}{I} - \underset{\sim}{I}_0^{1/2}\,\underset{\sim}{M}\,\underset{\sim}{D}_0^{-1}\,\underset{\sim}{M}'\,\underset{\sim}{I}_0^{1/2}\right)\underset{\sim}{Z}$$

Note that $\underset{\sim}{D}_0$ from (**)

$$\underset{\sim}{I}_0^{1/2}\,\underset{\sim}{M}\,\underset{\sim}{D}_0^{-1}\,\underbrace{\underset{\sim}{M}'\,\underset{\sim}{I}_0^{1/2}\,\underset{\sim}{I}_0^{1/2}\,\underset{\sim}{M}}\,\underset{\sim}{D}_0^{-1}\,\underset{\sim}{M}'\,\underset{\sim}{I}_0^{1/2}$$

$$= \underset{\sim}{I}_0^{1/2}\,\underset{\sim}{M}\,\underset{\sim}{D}_0^{-1}\,\underset{\sim}{M}'\,\underset{\sim}{I}_0^{1/2} \qquad \text{i.e. idempotent}$$

So (***) is χ^2_ν

$$\nu = k - \text{r}\left(\underset{\sim}{I}_0^{1/2}\,\underset{\sim}{M}\,\underset{\sim}{D}_0^{-1}\,\underset{\sim}{M}'\,\underset{\sim}{I}_0^{1/2}\right)$$

$$= k - \text{tr}\left(\underset{\sim}{D}_0^{-1}\,\underset{\sim}{M}'\,\underset{\sim}{I}_0\,\underset{\sim}{M}\right)$$

$$= k - \text{tr}\left(\underset{\sim}{D}_0^{-1}\,\underset{\sim}{D}_0\right) \qquad \text{from (**)}$$

$$= k - s = r$$

The nested sequence result follows similarly.

11 Oct 2001

The <u>deviance</u> of the fitted model $f(y|\hat{\theta})$ is

$$2[\ell(\tilde{\theta}) - \ell(\hat{\theta})] \geq 0$$

where $\tilde{\theta}$ is the mle for the saturated model, viz. parameter for each observation

By a big stretch

$$2[\ell(\tilde{\theta}) - \ell(\hat{\theta})] \sim \chi^2_{n-k}$$

People often compare final deviance to

$E\chi^2_{n-k} = n-k$, but $\exists$ problems with approx.

We have seen that

$$2[\ell(\hat{\theta}) - \ell(\hat{\underline{\theta}})] \sim \chi^2_{k-s} = \chi^2_p \text{ under null}$$

i.e. the difference in deviances between two nested models has a chi-squared distribution under the null hypothesis.

$$2[\ell(\tilde{\theta}) - \ell(\hat{\underline{\theta}})] - 2[\ell(\tilde{\theta}) - \ell(\hat{\theta})]$$

H_0 $\qquad$ H

11 Oct 2001

Might have $\Theta_1 \subset \Theta_0 \subset \Theta$

free params	s_1 <	s_0 <	k
constraints/restrictions	r_1	r_0	0

$$2[\ell(\tilde{\theta}) - \ell(\hat{\underline{\theta}})] - 2[\ell(\tilde{\theta}) - \ell(\hat{\theta}_-)] = D_{\Theta_1} - D_{\Theta_0}$$

$$\sim \chi^2_{s_0 - s_1}$$

Can set up an ANODEV table

Hypothesis	Deviance	df	Deviance difference	Δ df
Θ_1	D_{Θ_1}	$n - s_1$	$D_{\Theta_1} - D_{\Theta_0}$	$s_0 - s_1$
Θ_0	D_{Θ_0}	$n - s_0$	$D_{\Theta_0} - D_{\Theta}$	$k - s_0$
Θ	D_{Θ}	$n - k$		

Deviance gets smaller as bring in more parameters

11 Oct. 2001

Advantages of deviance.

1. Invariant under 1-1 parametrizations of the model

2. Unifies a variety of model selection problems: linear model (anova), glm, graphical models ...

3. Has interpretation in terms of K-L divergence

4. Lot known about large sample distribution

11 Oct. 2001

Distribution of the deviance / likelihood ratio statistic when the model is not true

We saw that

mle $\hat{\theta} \xrightarrow{p} \theta_*$ where θ_* maximizes

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

$$\frac{1}{n} l(\theta) = \frac{1}{n} \sum \log f(Y_i|\theta)$$

$$\xrightarrow{p} E_p\{\log f(Y|\theta)\}$$

so can anticipate

$$\frac{1}{n} l(\hat{\theta}) \xrightarrow{p} E_p\{\log f(Y|\theta_*)\}$$

Likewise

$$\frac{1}{n} l(\hat{\underline{\theta}}) \longrightarrow E_p\{\log f(Y|\underline{\theta}_*)\}$$

so $-\frac{2}{n}\log \lambda_n \longrightarrow 2E_p\{\log f(Y|\theta_*)\} - 2E_p\{\log f(Y|\underline{\theta}_*)\}$

Going to ∞ when $\theta_* \notin \Theta_0$

$\underline{\theta}_*$ gives $\max_{\theta \in \Theta_0} E_p\{\log f(Y|\theta)\}$

11 Oct. 2001

When $\theta_* \in \Theta_0$

$$-2 \log \lambda_n \sim c_1 z_1^2 + \dots + c_r z_r^2$$ $\Delta = k-r$

[n] Foutz & Srivastava (1978) Canadian J. Stat 6, 273-9.

14 Oct. 2001

Example. Gauss-Markov + normal model

$$Y_i = \alpha + (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})' \underset{\sim}{\beta} + \epsilon_i \qquad \epsilon_i : IN(0, \sigma^2)$$

$$H_0 : \quad \underset{\sim}{\beta} = \underset{\sim}{\beta}_0$$

$$\theta = \begin{bmatrix} \alpha \\ \underset{\sim}{\beta} \\ \sigma \end{bmatrix}$$

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2} \| y_i - \alpha - (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})' \underset{\sim}{\beta} \|^2 \right\}$$

$$\ell(\theta) = -\frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} \| y_i - \alpha - (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})' \underset{\sim}{\beta} \|^2 + C$$

"Full" mle: $\hat{\alpha} = \bar{y}$

$$\tilde{\underset{\sim}{X}}' \tilde{\underset{\sim}{X}} \hat{\underset{\sim}{\beta}} = \tilde{\underset{\sim}{X}}' \underset{\sim}{Y} \qquad \tilde{\underset{\sim}{X}} = \underset{\sim}{X} - \bar{\underset{\sim}{X}}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \| y_i - \bar{y} - (\underset{\sim}{x}_i - \bar{\underset{\sim}{x}})' \hat{\underset{\sim}{\beta}} \|^2$$

$$\ell(\hat{\theta}) = -\frac{n}{2} \log \hat{\sigma}^2 + C'$$

14 Oct. 01

Mle under H_0: $\quad \hat{\alpha} = \bar{y}$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} \|y_i - \bar{y}\|^2$$

$$n\, l(\hat{\underline{\theta}}_0) = -\frac{n}{2}\log \hat{\sigma}^2$$

$$2\left[l(\hat{\underline{\theta}}) - l(\hat{\underline{\theta}}_0)\right] = n \log \frac{\sum_i \|y_i - \bar{y}\|^2}{\sum_i \|y_i - \bar{y} - (\underset{\sim}{x}_i - \underset{\sim}{\bar{x}})'\hat{\underset{\sim}{\beta}}\|^2}$$

Now

$$\sum_{i=1}^{n} \|y_i - \bar{y}\|^2 = \sum_{i=1}^{n} \|(\underset{\sim}{x}_i - \underset{\sim}{\bar{x}})'\hat{\underset{\sim}{\beta}}\|^2 + \sum_{i=1}^{n}\|y_i - \bar{y} - (\underset{\sim}{x}_i - \underset{\sim}{\bar{x}})'\hat{\underset{\sim}{\beta}}\|^2$$

$$SST \quad = \quad SSR \quad + \quad SSE$$

Test statistic is based on SSR/SSE

We saw its exact distribution under H_0

$$\frac{SSR/p}{SSE/(n-p-1)} \sim F_{p,\, n-p-1}$$

14 Oct. 01

Relation to previous approximation:

It was based on $n \to \infty$

$$\hat{\sigma}^2 \xrightarrow{\text{Prob}} \sigma^2$$

$$n \log \frac{SSR + SSE}{SSE} = n \log\left(1 + \frac{SSR}{SSE}\right)$$

$$\sim \frac{SSR}{SSE/n}$$

$$\sim \frac{SSR}{\sigma^2}$$

$$\sim \chi^2_p$$