Approximating models

Nancy Reid, University of Toronto

Oxford, February 6

www.utstat.utoronto.reid/research

1. Context – Likelihood based inference

model $f(y;\theta)$, log likelihood function $\ell(\theta;y)$

$y = (y_1,\dots,y_n)$, $\ell(\theta;y) = \sum \log f(y_i;\theta)$

assume $\ell(\theta;y) = O(n)$, $\theta - \widehat{\theta} = O(n^{-1/2})$

goal to improve the approximation given by the limiting distribution

e.g. $r = \pm[2\{\ell(\widehat{\theta}) - \ell(\theta)\}]^{1/2} \xrightarrow{d} N(0,1)$

$E(r) = an^{-1/2} + O(n^{-3/2})$

$\text{var}(r) = 1 + bn^{-1} + O(n^{-3/2})$

implies

$$\frac{r - an^{-1/2}}{(1 + bn^{-1})^{1/2}} \;\dot\sim\; N(0,1)$$

is better than

$$r \;\dot\sim\; N(0,1)$$

2. Local exponential family models

$f(x;\varphi) = \exp\{\varphi x - c(\varphi) - d(x)\}$

$\ell_{;x} =$ $\qquad$ $\ell_{\varphi;x} =$

Start with arbitrary $f(y;\theta)$, $\qquad$ y, θ scalar

Expand $\ell(\theta;y)$ about $(\theta^0;y^0)$ $\qquad$ $\theta^0 =$

Represent coefficients $\ell(\theta^0;y^0)$, $\ell_{;y}(\theta^0;y^0)$, etc. by a_{ij}

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} \\ 0 & a_{11} & a_{12} & a_{13} & - \\ a_{20} & a_{21} & a_{22} & - & - \\ a_{30} & a_{31} & - & - & - \\ a_{40} & - & - & - & - \end{pmatrix}$$

1. Standardize $\theta \to (\theta - \theta^0)\hat{\jmath}^{1/2}$

$$y \to (y - y^0) a_{11} \hat{\jmath}^{-1/2}$$

$a_{20} \to -1, a_{11} \to 1, a_{ij} \to \tilde{a}_{ij}$

2. Reparametrize $\tilde{\theta} \to \tilde{\theta} + \tilde{a}_{21}\tilde{\theta}^2/2 + \tilde{a}_{31}\tilde{\theta}^3/6$
$\tilde{a}_{21} \to 0, \tilde{a}_{31} \to 0$

3. New variable $\tilde{y} \to \tilde{y} + \tilde{a}_{12}\tilde{y}^2/2 + \tilde{a}_{13}\tilde{y}^3/6$
$\tilde{a}_{12} \to 0, \tilde{a}_{13} \to 0$

4. Notation $a_{30} = -\dfrac{\alpha_3}{n^{1/2}}, \quad a_{40} = - \qquad a_{22} =$

$$\begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} & a_{04} \\ 0 & 1 & 0 & 0 & - \\ -1 & 0 & \gamma/n & - & - \\ -\alpha_3/n^{1/2} & 0 & - & - & - \\ -\alpha_4/n & - & - & - & - \end{pmatrix}$$

5. Density must integrate to 1

$\frac{3\alpha_4-5\alpha_3^2-12\gamma}{24n}$	$\frac{-\alpha_3}{2\sqrt{n}}$	$-1+\frac{-\alpha_4-2\alpha_3^2-5\gamma}{2n}$	$\frac{\alpha_3}{\sqrt{n}}$	$\frac{\alpha_4-3\alpha_3^2-}{n}$
0	1	0	0	$-$
-1	0	γ/n	$-$	$-$
$-\alpha_3/n^{1/2}$	0	$-$	$-$	$-$
$-\alpha_4/n$	$-$	$-$	$-$	$-$

$\alpha_3 =$

$\alpha_4 =$

$\gamma =$

First row is

$$-\log(2\pi) + (3\alpha_4 - 5\alpha_3^2 - 12\gamma)/24n, \quad -\alpha_3/2n^{1/2},$$

$$-1 + (\alpha_4 - 2\alpha_3^2 - 5\gamma)/2n, \quad \alpha_3/n^{1/2},$$

$$(\alpha_4 - 3\alpha_3^2 - 6\gamma)/n$$

...2 Local exponential family models

New density looks like

$$\begin{aligned} f(x,\varphi) &\doteq \phi(x-\varphi)\,\mathrm{exp}\{\ldots\} \\ &\doteq \phi(x-\varphi)\{1+\ldots\} \end{aligned}$$

with cdf

$$\begin{aligned} F(x,\varphi) &= \Phi(x-\varphi) + \\ & \phi(x-\varphi)\left[\frac{\alpha_3}{6\sqrt{n}}\{\ldots\} + \frac{\alpha_4}{24n}\{\ldots\} + \frac{\alpha_3^2}{72n}\{\ldots \right. \\ & \left. +\frac{\gamma}{4n}\left\{-2x+\varphi x^2+x^3\right\}\right] \end{aligned}$$

Free of γ at $x=0 \quad (y=y^0)$

p-value does not depend on γ

Andrews, Fraser, Wong, 2002

3. Tangent exponential model

$$p_{TEM}(x;\theta) = c|j(\hat{\varphi})|^{-1/2}\exp[\ell(\theta;y^0) - \ell(\hat{\theta}^0;y^0) + \{\varphi(\theta) - \varphi(\hat{\theta}^0)\}x]$$

,

$$\varphi = \frac{\partial \ell(\theta, y)}{\partial y}|_{y=y^0}$$

$$x = \frac{\partial \ell(\theta; y)}{\partial \theta}|_{\theta=\hat{\theta}^0}$$

$$j(\hat{\varphi}) = -\frac{\partial^2 \ell(\varphi)}{\partial \varphi^2}|_{\hat{\varphi}}$$

$\ell(\theta; y^0)$ is first column (ignoring (0,0) entry)
$\varphi(\theta)$ is second column (ignoring (0,1) entry)

These 2 columns determine the rest of the array, except the γ/n term

Easy to use p_{TEM} to get a p-value (saddlepoint type approximation)

...3 Tangent exponential model

How to get a scalar variable y? Condition on an (approximate) ancillary, so $\ell_{;y}$ is taken for fixed ancillary $a(y)$.

This can be computed by finding a vector $V = (V_1, \ldots, V_n)^T$ tangent to the ancillary at y^0:

$$\varphi(\theta) = \ell_{;V}(\theta; y)|_{y^0} = \sum \ell_{;y_i}(\theta; y_i^0)V_i$$

Example

$$y_i \sim f(y_i - \mu) \quad a_i = y_i - \widehat{\mu}, \text{say}, \; V_i = 1$$

$$\varphi(\theta) = \sum \frac{\partial \log f(y_i - \mu)}{\partial y_i}|_{y^0} = -\ell_\theta(\theta; y^0)$$

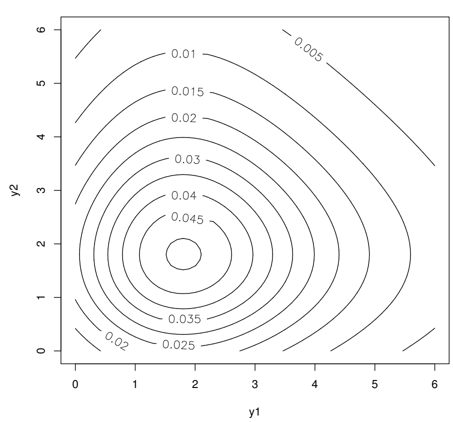
Example

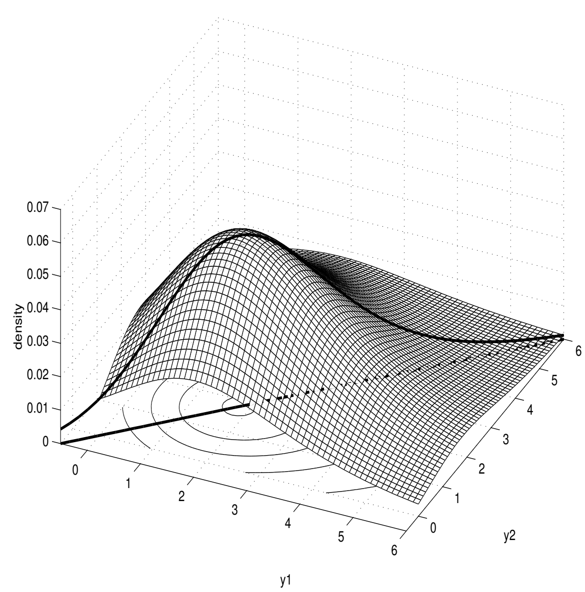
$$f(y_1, y_2; \theta) = \frac{e^{y-\theta}}{\{1 + e^{(y-\theta)}\}^2} \exp[\gamma(\theta)(y-\theta) - c\{\gamma(\theta)\}],$$

$$-1 \leq \theta \leq 1$$

$$\gamma(\theta) = 0.5 \tanh(\theta)$$

$$c(\theta) = \log\{(\pi\theta)/\sin(\pi\theta)\}$$





...3 Tangent exponential model Vector θ? Use the same approach, now

$$V = (\underline{V}_1, \underline{V}_2, \ldots, \underline{V}_n)^T$$

$\underline{V}_i$ is $1 \times d$, $\ell_{;V}(\theta)$ is also $1 \times d$

Example

$$y_i = x_i^T \beta + \sigma e_i$$

$$\underline{V}_i = (x_i^T \quad \hat{e}_i)$$

Example

$$y_i = \mu_i(\beta) + \sigma e_i$$

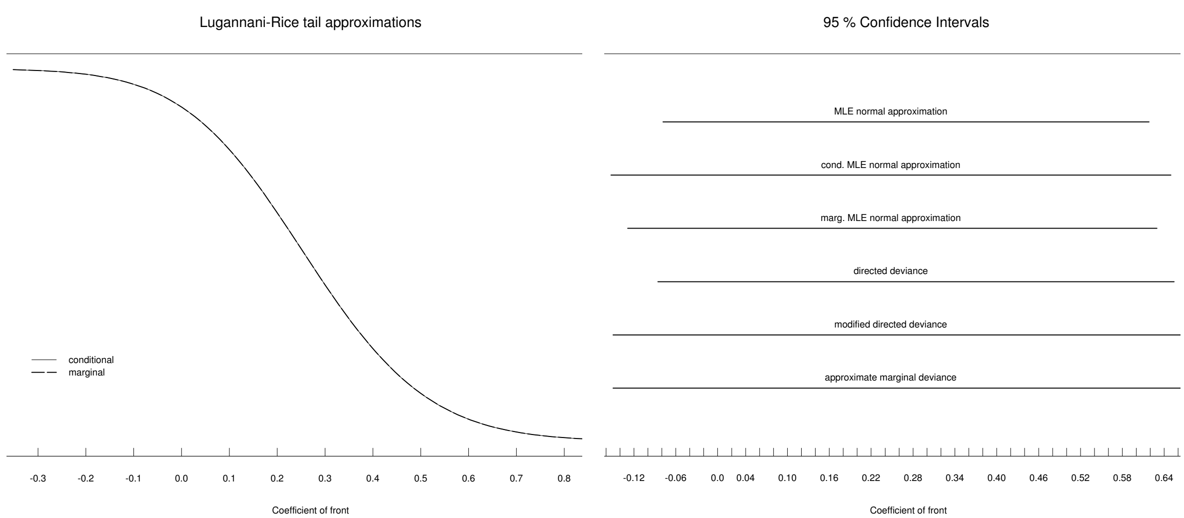
$$\underline{V}_i = \{\mu_i'(\hat{\beta}) \quad \hat{e}_i\}$$

Inference re nuisance parameters uses p_{TEM} twice to get a marginal distribution

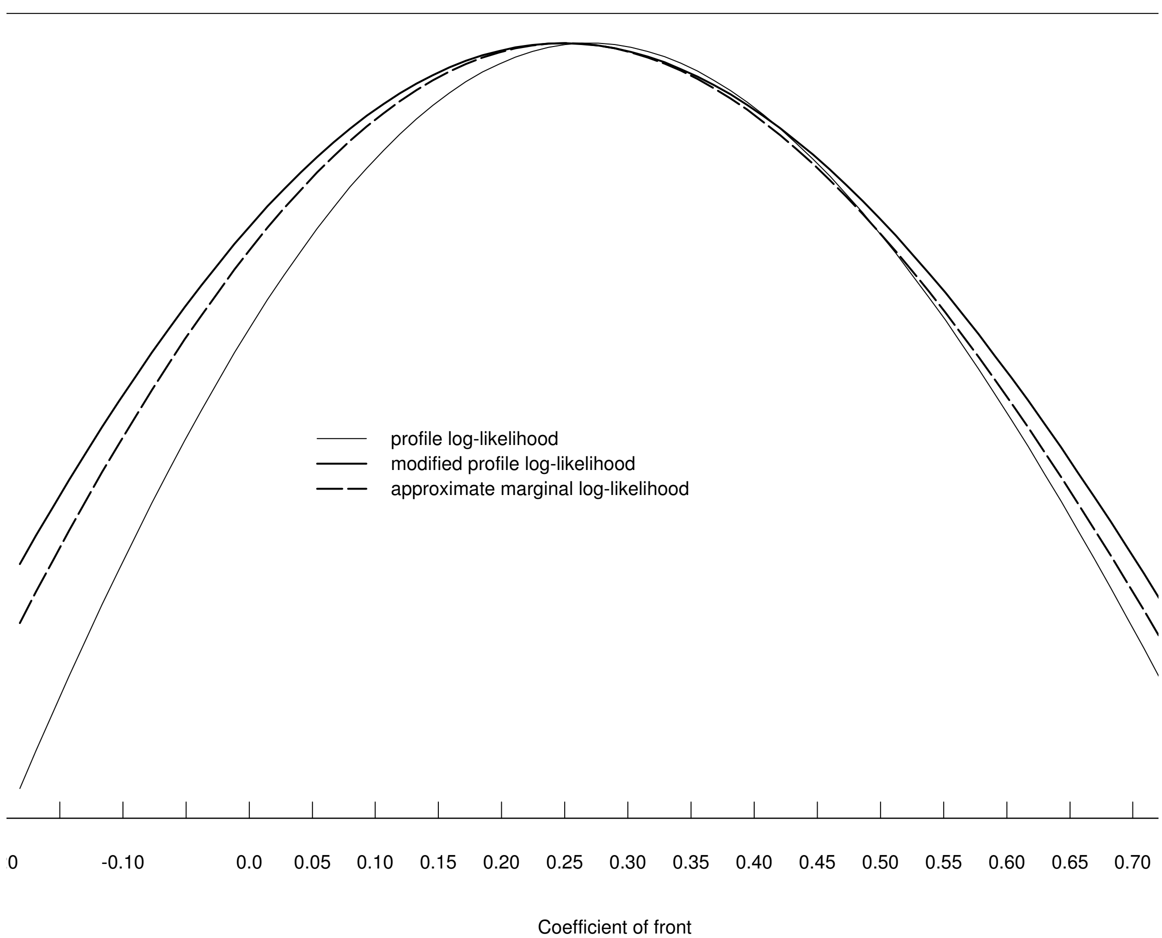
Example House price data (Srivastava and Sen); 4 covariates, 26 observations, model

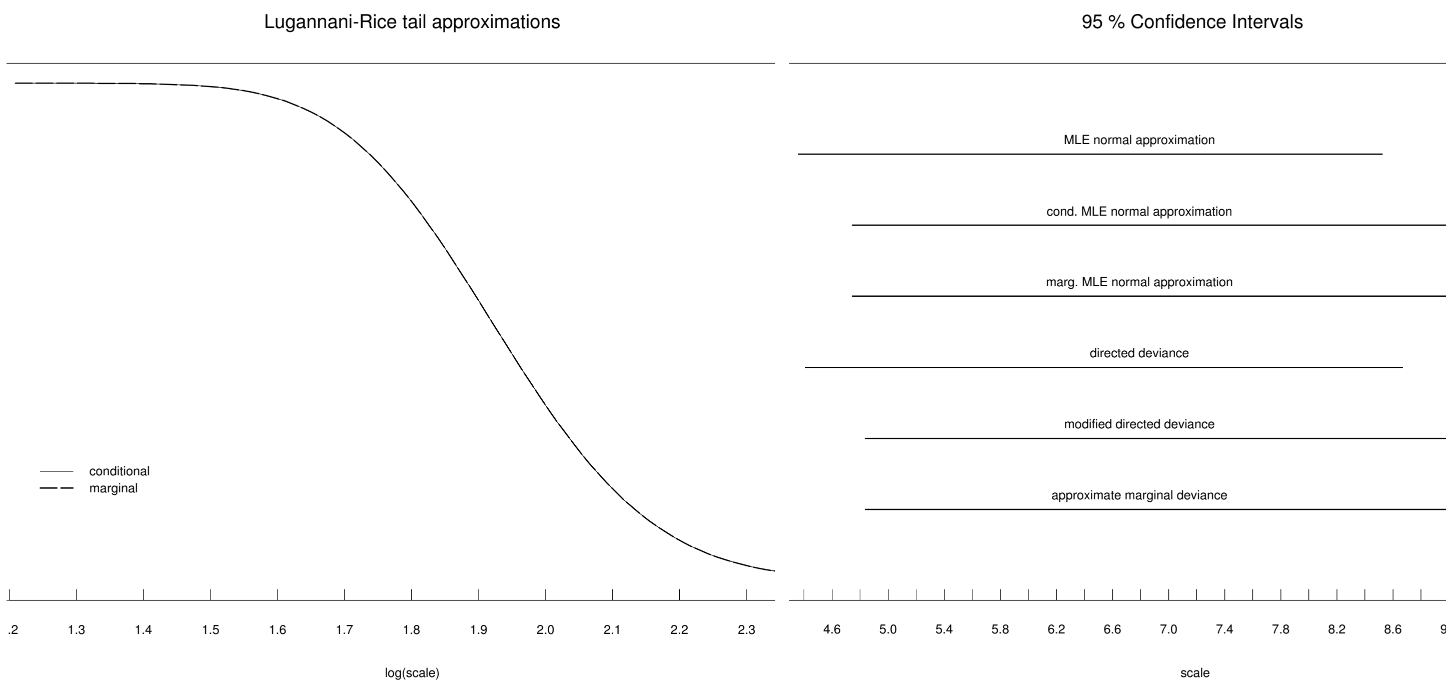
$$y_i = x_i^T \beta + \sigma e_i, \quad e_i \sim t_5$$

marginal inference for β_4 and for $\log \sigma$, (conditional on usual ancillary), uses Alessandra Brazzale's Splus library `HOA`

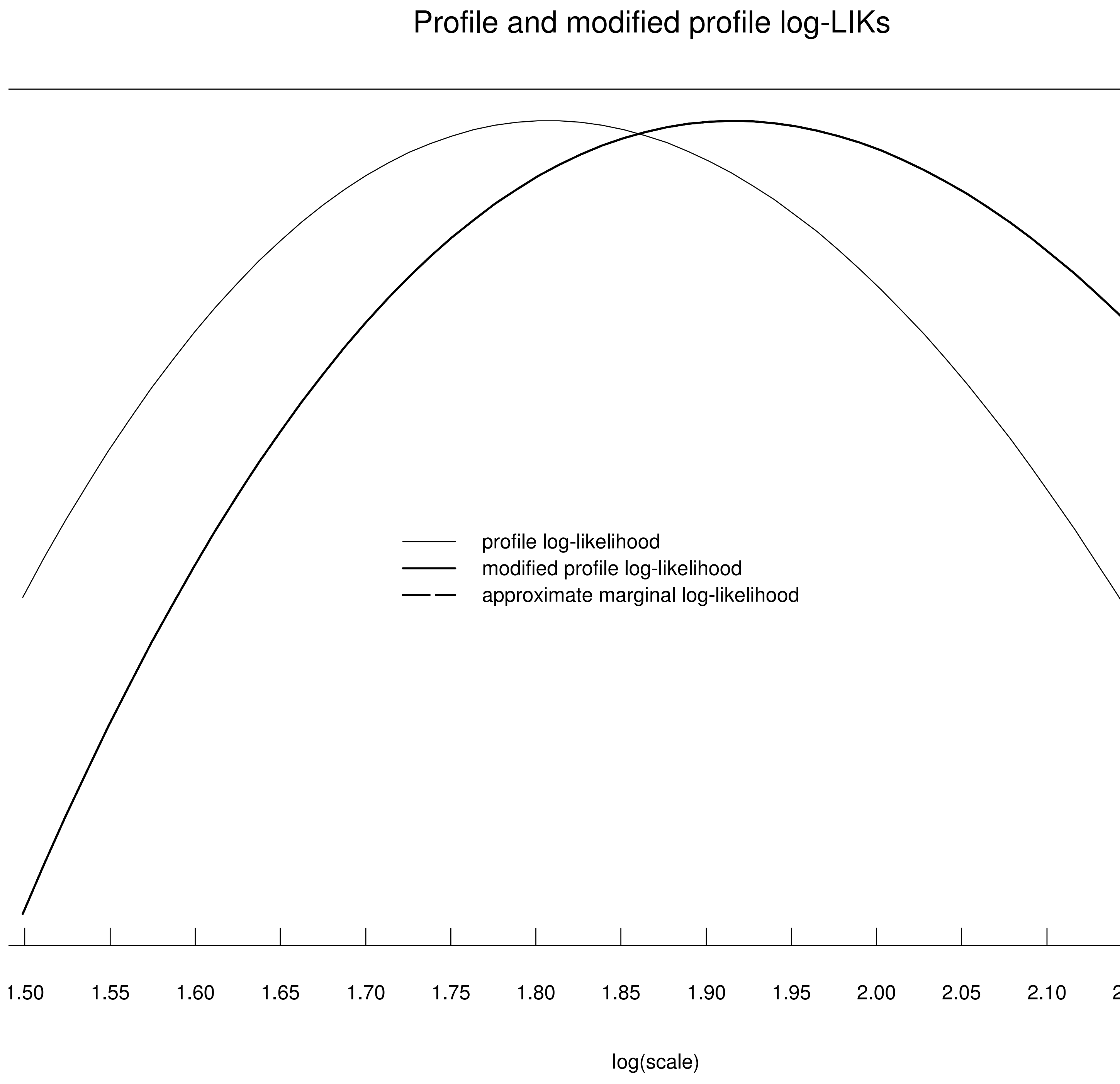


Profile and modified profile log-LIKs





Lugannani-Rice tail approximations
conditional
marginal
.2
1.3
1.4
1.5
1.6
1.7
1.8
1.9
2.0
2.1
2.2
2.3
log(scale)
95 % Confidence Intervals
MLE normal approximation
cond. MLE normal approximation
marg. MLE normal approximation
directed deviance
modified directed deviance
approximate marginal deviance
4.6
5.0
5.4
5.8
6.2
6.6
7.0
7.4
7.8
8.2
8.6
scale


Profile and modified profile log-LIKs
profile log-likelihood
modified profile log-likelihood
approximate marginal log-likelihood
1.50
1.55
1.60
1.65
1.70
1.75
1.80
1.85
1.90
1.95
2.00
2.05
2.10
log(scale)

```
> houses.marg.front <- cond.rsm(mod.obj=houses.rsm,offset=fron
> summary(houses.marg.front)

 FORMULA:  price ~ bdroom + floor + rooms + front
 FAMILY :  student
 OFFSET :  front

 COEFFICIENTS
          Value   Std. Error
uncond.  0.269959     0.177693
  cond.  0.248288     0.204779
   marg. 0.250306     0.193580

 CONFIDENCE INTERVALS
 level = 95 %
                                            lower  two-sided
    MLE normal approx.                 -0.0783128            (
    Cond. MLE normal approx.           -0.1530720            (
    Marg. MLE normal approx.           -0.1291040            (
    Directed deviance                  -0.0857505            (
    Modified directed deviance         -0.1498960            (
    Marginal directed deviance         -0.1498960            (
```

4. Local location models

$f(x;\beta) = f(x-\beta)$

$\ell_\beta = -\ell_{;x}$ $\ell_{\beta\beta} = -\ell_{\beta x} = \ell_{;xx}$

4.1 If $y \sim f(y;\theta)$ then

$$x = \int^y -\frac{F_y(y;\theta_0)}{F_\theta(y;\theta_0)} dy$$

has a density which is a location model near θ_0, $g(x-\Delta)$, say.

Satisfies $\ell_\Delta = -\ell_{;x}$, but not higher order.

This model has an exact ancillary

This ancillary can be used for the original model, for computing p-values. (This is where V above came from.)

...4. Local location models

4.2 As with exponential model we can carry this further to get an array of coefficients for the double expansion about $(y^0, \hat{\theta}^0)$ of the form:

$$\begin{pmatrix} a_{00} & 0 & -1+ & a_3/n^{1/2} & -a_4/n- \\ 0 & 1 & -a_3/n^{1/2} & a_4/n & - \\ -1 & a_3/n^{1/2} & -a_4/n+ & - & - \\ -a_3/n^{1/2} & a_4/n & - & - & - \\ -a_4/n & - & - & - & - \end{pmatrix}$$

Andrews, Fraser, Wong, 2003

A more compact notation

$$f\{x - \beta(\theta)\}, \quad \beta(\theta) = \int^{\theta} -\frac{\ell_\theta(\theta)}{\varphi(\theta)} d\theta$$

Existence (algorithm) for vector θ

Fraser, Yi, 2002

$$\begin{pmatrix} a + \frac{3\alpha_4 - 5\alpha_3^2 - 12\gamma}{24n} & 0 & -1 + \frac{5\gamma}{2n} & \frac{\alpha_3}{n^{1/2}} & \frac{-\alpha_4 - 6\gamma}{n} \\ \\ 0 & 1 & -\alpha_3/n^{1/2} & -\alpha_4/n & - \\ -1 & \alpha_3/n^{1/2} & -\frac{\alpha_4+\gamma}{n} & - & - \\ -\alpha_4/n & - & - & - & - \end{pmatrix}$$

$$x = \int^y -\frac{F_y(y;0)}{F_{;\theta}(y;0)} dy, \qquad G(x;\theta) = F\{y(x);\theta\}$$

$$\begin{aligned} G_x(x;0) &= F_y\{y(x);0\}\left\{-\frac{F_{;\theta}(y;0)}{F_y(y;0)}\right\} \\ &= -F_{;\theta}(y;0) \\ &= -G_{;\theta}(x;0) \end{aligned}$$

...4 Local location model

Bayesian analysis of location model uses flat prior for location parameter, in our case

$$\pi(\theta) \propto d\beta(\theta)$$

and this will give posterior p-values equal to those from tangent exponential model

to $O(n^{-3/2})$ if non-location term $\gamma = 0$,

to $O(n^{-1})$ if $\gamma \neq 0$

With nuisance parameters, can only obtain 'strong matching' priors for a single parameter of interest, using

$$\pi(\psi, \hat{\lambda}_\psi) \propto \left| \frac{\partial \psi}{\partial \beta} \right|^{-1}_{(\psi, \hat{\lambda}_\psi)} \times \frac{|j_{\lambda\lambda}(\hat{\theta}_\psi)|}{|\varphi_\lambda(\hat{\theta}_\psi)|}$$

Fraser & Reid, 2003

Example Location model with curved parameter of interest

$Y_1 \sim N(\theta_1, 1)$, $Y_2 \sim N(\theta_2, 1)$ independent

$\psi^2 = (R + \theta_1)^2 + \theta_2^2$; R known

$r^2 = \surd\{(R + y_1)^2 + y_2^2\}$

Bayesian posterior under usual flat prior $(\theta_1, \theta_2 | y) \sim N(y_1, y_2)$

frequentist p-value (marginal) Pr$\{r \leq r^0; \psi^0)$

Bayesian p-value Pr$\{\psi \geq \psi^0 | y)$

Will be quite different:

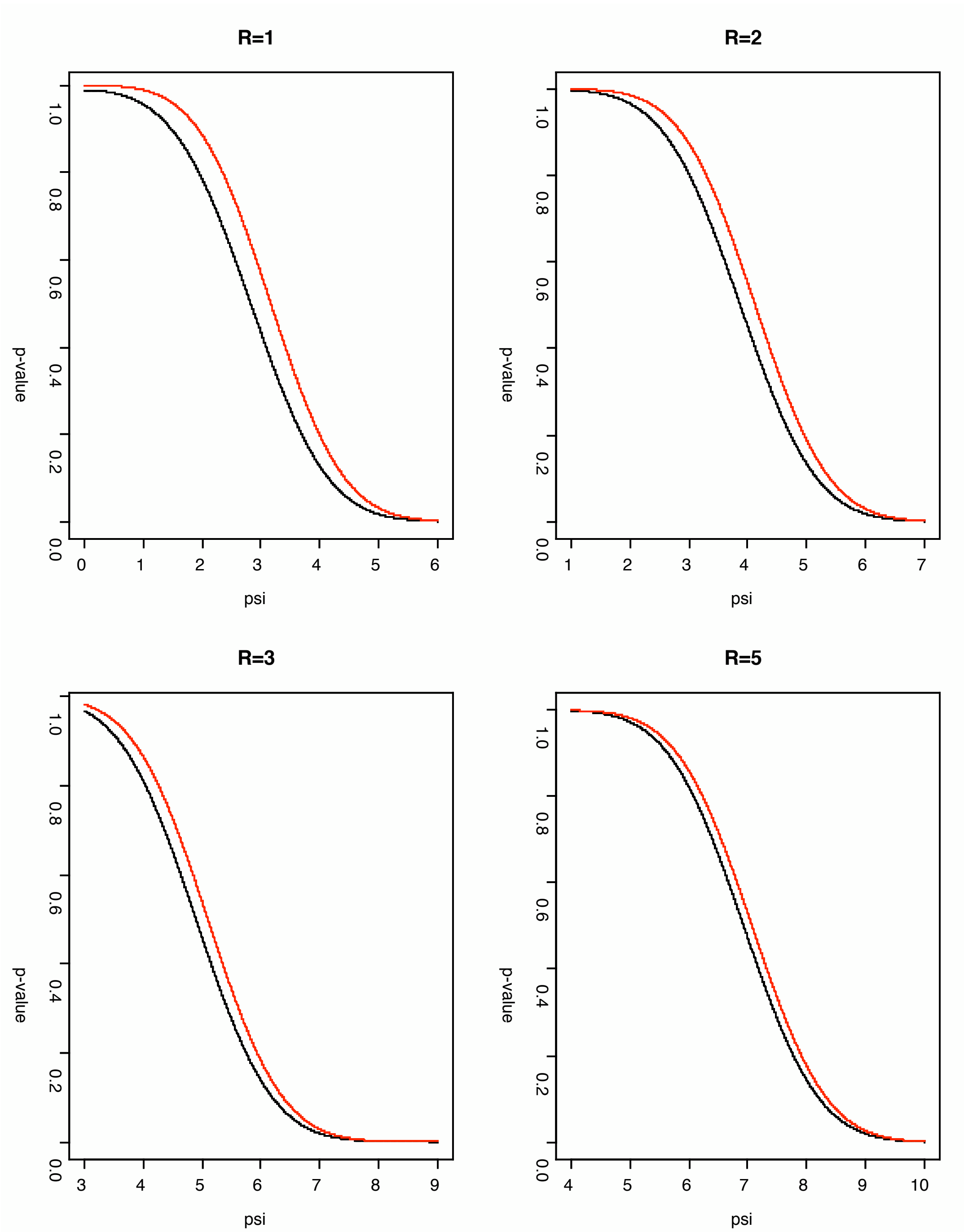
matching prior using information adjustment gives $\pi(\theta) \propto \dfrac{r}{\psi}$

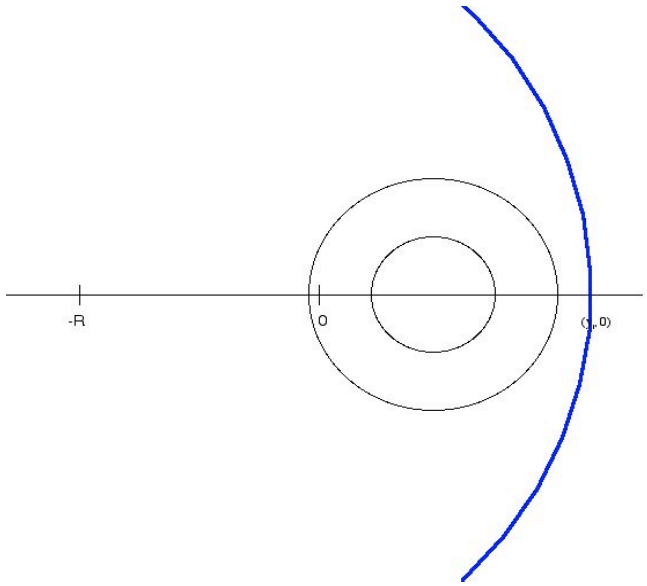
frequentist = $\Pr\{\chi_2^{2'}(\psi^0)^2 \leq (y_1 + R)^2 + y_2^2\}$

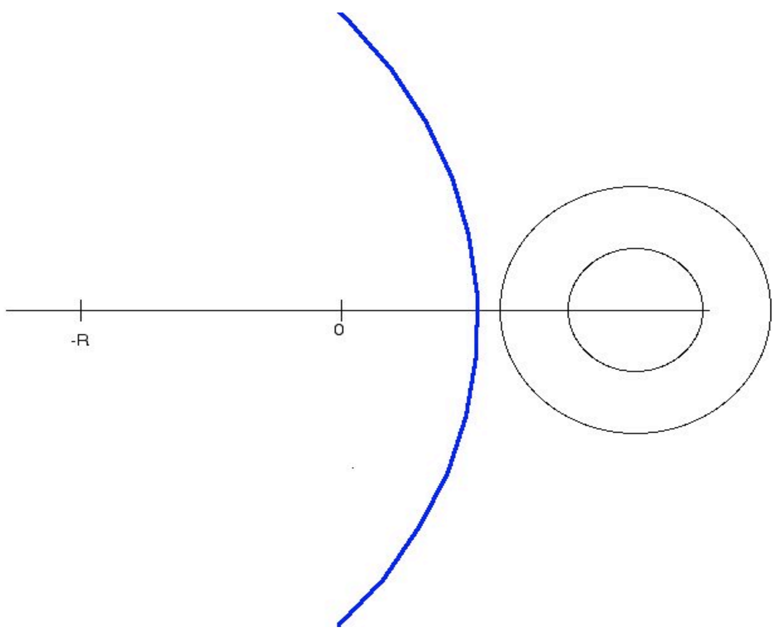
Bayesian = $\Pr\{\chi_2^{2'}((y_1 + R)^2 + y_2^2) \geq \psi^{02}\}$

Bayesian−frequentist = $\Pr\{X_1 - X_2 = 0\}$

$X_1 \sim Po((y_1 + R)^2 + y_2^2), X_2 \sim Po(\psi^{02})$


R=1
R=2
R=3
R=5
p-value
psi
0.0
0.2
0.4
0.6
0.8
1.0


-R
0


-R
0

References

Andrews, D.A., Fraser, D.A.S., Wong, A. Computation of distribution functions from likelihood information near observed data.

Brazzale, A. `http://www.isib.cnr.it/ brazzale`

Fraser, D.A.S., Reid, N. Strong matching of frequentist and Bayesian parametric inference.

Fraser, D.A.S., Yi, G. Location reparametrization and default priors for statistical analyses.

Reid, N. Asymptotics and the theory of inference.