1. Computing p-values

Model $f(y;\psi,\lambda)$

Goal: p-value for testing ψ

Solution: correction to likelihood ratio statistic via maximum likelihood or score type statistic

$$
\begin{aligned}
p-\text{value} &\doteq \Phi(r)+\phi(r)\left(\frac{1}{r}-\frac{1}{q}\right)\\
&\doteq \Phi(r+\frac{1}{r}\log\frac{q}{r})\\
r &= \pm\surd[2\{\ell(\hat{\psi},\hat{\lambda})-\ell(\psi,\hat{\lambda}_\psi)\}]
\end{aligned}
$$

Example: $N(\theta, \theta^2)$

Value for θ ($n = 5$)

	0.6	0.8	1.6	1.8	2.0
Exact	0.9976	0.8863	0.0794	0.0416	0.0228
r^*, q (or u)	0.9976	0.8855	0.0787	0.0412	0.0225
r^*, Sk	0.9975	0.8837	0.0768	0.0400	0.0217
$\Phi(r)$	0.9960	0.8465	0.0513	0.0249	0.0128

Example: a $(2, 1)$ exponential family

$$f(y; \theta) = \exp\{-\frac{e^{-\theta}}{\theta} y_1 - \theta y_2\};\ \psi = 1,\ n = 1$$

	-3.0	-1.96	-0.67	0.67	1.96	3.0
Exact	0.0013	0.0250	0.251	0.749	0.9750	0.9987
r^*, q	0.0012	0.0221	0.227	0.740	0.9759	0.9987
r^*, u	0.0012	0.0229	0.236	0.742	0.9759	0.9987

Example: Sample of size 9 from $N(\mu, \sigma^2)$;
$\psi = \mu + \sigma^2/2$, plot p-value vs ψ:

$$q = \frac{|\ell_{;V}(\hat{\theta}) - \ell_{;V}(\hat{\theta}_\psi) \quad \ell_{\lambda;V}(\hat{\theta}_\psi)|}{|\ell_{\theta;V}(\hat{\theta}\,)|} \quad \frac{|j_{\theta\theta}(\hat{\theta}\,)|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}}$$

$$\begin{aligned} \ell_{;V}(\theta) &= \left\{\frac{\partial \ell(\theta; y)}{\partial \underline{v}_1}, \ldots, \frac{\partial \ell(\theta; y)}{\partial \underline{v}_k}\right\} \\ &= \sum_{i=1}^{n} \ell_{;y_i}(\theta; y^0) V_i \end{aligned}$$

$$\begin{aligned} V &= -\left(\frac{\partial z}{\partial y}\right)^{-1} \left(\frac{\partial z}{\partial \theta}\right)\Bigg|_{y^0, \hat{\theta}^0} \\ z &= z(y; \theta) = \{z_1(y_1, \theta), \ldots, z_n(y_n, \theta)\} \end{aligned}$$

Example: a $(2,1)$ exponential family

$$f(y;\theta) = \exp\{-\tfrac{e^{-\theta}}{\theta}y_1 - \theta y_2\}$$
$$\ell(\theta) = -\ \frac{e^{-\theta}}{\theta}y_1 - \theta y_2$$

1. pivotal: $z_1 = \dfrac{e^{-\theta}}{\theta}y_1,\ z_2 = \theta(y_2 - 1)$

$$\begin{aligned}\left(\frac{\partial z}{\partial y}\right)^{-1} &= \begin{pmatrix} e^{-\theta}/\theta & 0 \\ 0 & \theta \end{pmatrix}^{-1} \\ \frac{\partial z}{\partial \theta} &= \left\{-\frac{e^{-\theta}}{\theta}(1+\frac{1}{\theta})y_1, \quad y_2 - 1\right\}^T\end{aligned}$$

2.

$$\begin{aligned}V &= -\left(\frac{\partial z}{\partial y}\right)^{-1}\left(\frac{\partial z}{\partial \theta}\right)\Bigg|_{y^0,\hat{\theta}^0} \\ &= \left\{(1+\frac{1}{\hat{\theta}})y_1, -\frac{1}{\hat{\theta}}(y_2 - 1)\right\}\end{aligned}$$

3.

$$\ell(\theta) = -\frac{e^{-\theta}}{\theta}y_1 - \theta y_2$$

$$\begin{aligned}
\ell_{;V} &= -\frac{e^{-\theta}}{\theta}(1+\frac{1}{\widehat{\theta}})y_1 + \frac{\theta}{\widehat{\theta}}(y_2 - 1) \\
\ell_{\theta;V} &= \\
\ell_{\theta\theta} &= \\
\ell_\theta(\widehat{\theta}) &= 0 \quad \Rightarrow y_2 = \frac{e^{-\widehat{\theta}}}{\widehat{\theta}}(1+\frac{1}{\widehat{\theta}})y_1 = y_2 \\
q &= \left[\left\{\frac{e^{-\theta}/\theta}{e^{-\widehat{\theta}}/\widehat{\theta}} - \frac{\theta}{\widehat{\theta}}\right\}y_2 + \frac{\theta}{\widehat{\theta}}\right]\left(y_2 + \frac{2y_2}{\widehat{\theta}} - \frac{1}{\widehat{\theta}}\right)^{-1} \\
& \qquad \left\{\frac{y_2(2/\widehat{\theta}+2+\widehat{\theta})}{1+\widehat{\theta}}\right\}^{1/2} \\
r &= \pm\sqrt{2}\{\ell(\widehat{\theta}) - \ell(\theta)\}
\end{aligned}$$

Note: an explicit form for an approximate ancillary a is given in Barndorff-Nielsen and Wood, 1998, from which u can be calculated

Some special cases

Model	q (from either approach)
one parameter exponential	$(\hat{\theta} - \theta) j(\hat{\theta})^{1/2}$
one parameter location	$\ell'(\theta)\{j(\hat{\theta})\}^{-1/2}$
canonical exponential	$(\hat{\psi} - \psi)\{j_p(\hat{\psi})\}^{1/2} \frac{\lvert j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})\rvert^{1/2}}{\lvert j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)\rvert^{1/2}}$
location regression	$\ell_p'(\psi)\{j_p(\hat{\psi})\}^{-1/2} \frac{\lvert j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)\rvert^{1/2}}{\lvert j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})\rvert^{1/2}}$
marginal posterior	$-\ell_p'(\psi)\{j_p(\hat{\psi})\}^{-1/2} \frac{\lvert j_{\lambda\lambda}(\psi,\hat{\lambda}_\psi)\rvert^{1/2}}{\lvert j_{\lambda\lambda}(\hat{\psi},\hat{\lambda})\rvert^{1/2}} \frac{\pi(\hat{\psi},\hat{\lambda})}{\pi(\psi,\hat{\lambda}_\psi)}$

Brief review of available applications

1. Illustrative examples

Highly simplified models, simulations, comparison to first order approximations and to exact results. A relatively large number of these types of examples, especially in one parameter problems. E.g. Cauchy model, sample size 1, true location parameter 0:

exact	0.25	0.06	0.03	0.016	0.0106
$\Phi(r^*)$	0.23	0.06	0.03	0.014	0.0094
$\Phi(r)$	0.12	0.005	0.001	0.0003	0.0001
score					0.4800
Wald					10^{-196}

See, e.g., list in Reid CJS 1996

2. Case studies

Moderately realistic examples, comparison of third order approximations with usual normal approximations, typically using data sets from the literature; sometimes includes simulations.

Fraser, Wong and Wu (1999, JASA):

$$\underline{y} = \underline{x}(\beta) + \sigma \underline{e}$$

with $x(\beta)$ known, possibly nonlinear, and distribution of e either normal or t. Four simulation studies, four case studies.

Ruggero Bellio (1999, thesis):

$$y_{ij} = \mu(x_i, \beta) + e_{ij}$$

$$e_{ij} \sim N(0, g^2(x_i, \beta, \rho)$$

Four case studies, compare first and third order approximations, comparison to bootstrap, use of r^* for model checking.

Also Skovgaard's (1996, 1999) version applied in inverse regression problems and in mixed random effects models (main nonlinear regression, normal errors).

Butler and co-authors, esp. Huzurbazar

Construct an exponential family for which the null hypothesis corresponds to the test statistic of interest, use exponential family versions (saddlepoint approximations)

3. Nearly automatic software

Maple programs used for examples in Fraser, Reid and Wu (1999, Biometrika); user provides model function, pivotal statistic; software computes sample space derivatives. Computation of profile log-likelihood and its curvature needs a fair bit of interaction from the user.

Alessandra Brazzale has built S-PLUS libraries for:

1. logistic regression and loglinear models (builds on `glm` with `family=binomial/Poisson`
2. marginal inference for linear non-normal regression
3. conditional inference for nonlinear regression model of Bellio

Available through `Statlib`, requires from user only the model function and the parameter of interest. Two main difficulties are profile likelihood and patching the middle (but

More general models

- Butler (1998ff) Saddlepoint approximations for first passage time distributions (not as closely tied to likelihood theory). Also Huzurbazar has applied similar ideas for complex survival data models.

- Kolassa and Tanner (1999, Bcs) combines exponential family version (conditional inference) with MCMC

- Empirical likelihood is Bartlett-correctable, but the small sample behaviour is not as good as for regular likelihood (Davison, Corcoran and Spady). Improved versions suggested.

Outstanding problems

Robustness

-what happens to the theory if the model is incorrect (even a little)

-does comparison of first order and higher order approximations provide a check on the adequacy of the model

-can higher order asymptotics be used for improved diagnostics

Semi- and non-parametric models

-comparison to bootstrap; theoretical and examples

-higher order results for semi-parametric models (Mykland, 1999, Annals)

Comparison to MCMC etc

-Relatively complex model, relatively large amount of computation, does MCMC need a 'reality check', does asymptotics have anything to offer?

"Real" applications

-e.g.(1) Brain activations during complex tasks (Nature Neuroscience, April) 12 men, 12 women; comparison of maximal activation areas re t-tests

-e.g.(2) Prospective study of type II diabetes (JAMA, May) 12,000 study participants, 1400 cases of diabetes; stratified by gender and race the small group had 161 cases in 976 participants

Some theoretical problems

higher order asymptotics for dependent data

higher order asymptotics for heavy tails

large (even infinite) numbers of nuisance parameters (Barndorff-Nielsen, 1998)

connection to large deviations (B-N/Wood, Jensen 1995)

more complex random effects/ mixed models

model choice

Selected references

Skovgaard, 1990 Annals, 1996 Bernoulli, 2000 SJS

Barndorff-Nielsen and Cox 1994

Barndorff-Nielsen and Wood, 1998 Bernoulli

Fraser and Reid, 1998 Biometrika, 2000 JSPI, `www.utstat.utoronto.ca/reid`

Reid, 1996, CJS

Acknowledgements and thanks

Martha Cerilla, Victor Perez-Abreu, and the local organizing committee

Don Fraser

Ib Skovgaard, Ruggero Bellio, Alessandre Brazzale

Andrey Feuerverger, David Andrews, Augustine Wong, Radford Neal