Higher order asymptotics in practice

Nancy Reid, University of Toronto

www.utstat.utoronto.ca/reid/research
Likelihood inference in parametric models

Example: \( f(y; \theta) = \exp\{- (y - \theta) - e^{-(y - \theta)}\} \)

\[
\ell(\theta) = \log f(y; \theta)
\]

\( y = y^0 = 21.5 \)
Likelihood inference in parametric models

Example: \( f(y; \theta) = \exp\{-(y - \theta) - e^{-y}\} \)

\[ p\text{-value function } p(\theta) = F(y^0; \theta) \]
Likelihood inference in parametric models

Example: \( f(y; \theta) = \exp\{- (y - \theta) - e^{-(y-\theta)}\} \)

- **log-likelihood function** \( \ell(\theta) = \log f(y; \theta) \)
- **score function** \( \ell'(\theta) \)
- **maximum likelihood estimate** \( \hat{\theta} \)
- **observed information** \( j(\hat{\theta}) = -\ell''(\hat{\theta}) \)
Likelihood inference in parametric models

Example: \( f(y; \theta) = \exp\{- (y - \theta) - e^{-(y-\theta)}\} \)

log-likelihood function \( \ell(\theta) = \log f(y; \theta) \)

score function \( \ell'(\theta) \)

maximum likelihood estimate \( \hat{\theta} \)

observed information \( j(\hat{\theta}) = -\ell''(\hat{\theta}) \)

score \( s(\theta) = \ell'(\theta)\{j(\hat{\theta})\}^{-1/2} \)

max. lik. \( q(\theta) = (\hat{\theta} - \theta)\{j(\hat{\theta})\}^{1/2} \)

likelihood root \( r(\theta) = \sqrt{2\{\ell(\hat{\theta})-\ell(\theta)\}} \)

\( \xrightarrow{d} N(0, 1) \)
Third order approximation

\[ r^* = r + \frac{1}{r} \log \frac{s}{r} \]

\[ r^* \overset{d}{\rightarrow} N(0, 1) \]
Likelihood statistics as pivots

score statistic \( s(\theta) = \ell'(\theta)\{j(\widehat{\theta})\}^{-1/2} \)

standardized m.l.e. \( q(\theta) = (\widehat{\theta} - \theta)\{j(\widehat{\theta})\}^{1/2} \)

likelihood root \( r(\theta) = \sqrt{2\{\ell(\widehat{\theta}) - \ell(\theta)\}} \)

First order \( p \)-values

\[
p(\theta) = \Phi\{r(\theta)\} = \Phi\{q(\theta)\} = \Phi\{s(\theta)\}
\]

Third order \( p \)-values

\[
p(\theta) = \Phi\{r^*(\theta)\} = \Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{Q}\right)
\]

\[
r^*(\theta) = r + \frac{1}{r} \log \frac{Q}{r}
\]

\[
Q = \{\ell,\widehat{\theta}(\theta) - \ell,\theta(\theta)\} j(\widehat{\theta})^{-1/2}
\]
Example: $Y \sim Po(\theta), \theta > b$

$$f(y; \theta) = \frac{1}{y!} \theta^y e^{-\theta} = \frac{1}{y!}(\mu + b)^y e^{-(\mu + b)}$$

$b$ is the background, $\mu > 0$ corresponds to a new signal

Fraser, Reid, Wong 2004

$p$-values:

- upper 0.0005993
- lower 0.0002170
- mid 0.0004081
- $r^*$ 0.0003779
- $r$ 0.0004416
- $\tilde{\theta}$ 0.0062427
Example: $2 \times 2$ table


<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Stayed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>1</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>Women</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>20</td>
<td>26</td>
</tr>
</tbody>
</table>
Example: 2 × 2 table


<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Stayed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>1</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>Women</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>20</td>
<td>26</td>
</tr>
</tbody>
</table>

Model: $Y_1 \sim Bin(19, p_1), Y_2 \sim Bin(7, p_2)$

$\theta = (\psi, \lambda)$

$\psi = \log\left\{\frac{p_1(1 - p_2)}{p_2(1 - p_1)}\right\}$

$H_0 : p_1 = p_2, \text{ or } \psi = 0$
Example: $2 \times 2$ table


<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Stayed</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Men</td>
<td>1</td>
<td>18</td>
<td>19</td>
</tr>
<tr>
<td>Women</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
<td>20</td>
<td>26</td>
</tr>
</tbody>
</table>

Model: $Y_1 \sim Bin(19, p_1), Y_2 \sim Bin(7, p_2)$

$\theta = (\psi, \lambda)$

$\psi = \log\{p_1(1 - p_2)/p_2(1 - p_1)\}$

$H_0: p_1 = p_2$, or $\psi = 0$

$p$-value: $\Phi(q) = 0.002$

$\Phi(r) = 0.0003$

$\Phi(r^*) = 0.0005$
```r
> library("cond")
> astro.glm <- glm(astro~gender,family=binomial)
> astro.cond <- cond.glm(astro.glm,offset=gender)
> summary(astro.cond)

FORMULA: astro ~ gender
FAMILY : binomial
OFFSET : gender

COEFFICIENTS

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>uncond.</td>
<td>-3.80666</td>
<td>1.32449</td>
</tr>
<tr>
<td>cond.</td>
<td>-3.55074</td>
<td>1.23712</td>
</tr>
</tbody>
</table>

CONFIDENCE INTERVALS
level = 95 %

<table>
<thead>
<tr>
<th>Approximation</th>
<th>lower</th>
<th>two-sided</th>
<th>upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE normal approximation</td>
<td>-6.40261</td>
<td>-1.210710</td>
<td></td>
</tr>
<tr>
<td>CMLE normal approximation</td>
<td>-5.97544</td>
<td>-1.126030</td>
<td></td>
</tr>
<tr>
<td>Directed deviance</td>
<td>-7.06294</td>
<td>-1.525520</td>
<td></td>
</tr>
<tr>
<td>Modified directed deviance</td>
<td>-6.37130</td>
<td>-1.315430</td>
<td></td>
</tr>
<tr>
<td>Modified directed deviance (cont. corr.)</td>
<td>-7.71524</td>
<td>-0.881396</td>
<td></td>
</tr>
</tbody>
</table>

DIAGNOSTICS:

| INF NP | 0.126 0.234 |

Approximation based on 20 points

> summary(astro.cond,test=0,digits=4)

FORMULA: astro ~ gender
FAMILY : binomial
OFFSET : gender

COEFFICIENTS

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
```

10
uncond. -3.807 1.324
cond. -3.551 1.237

HYPOTHESIS TESTING
hypothesis : coef( gender ) = 0

<table>
<thead>
<tr>
<th>statistic</th>
<th>tail prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLE normal approximation</td>
<td>-2.874 0.0020260</td>
</tr>
<tr>
<td>CMLE normal approximation</td>
<td>-2.870 0.0020510</td>
</tr>
<tr>
<td>Directed deviance</td>
<td>-3.447 0.0002838</td>
</tr>
<tr>
<td>Modified directed deviance</td>
<td>-3.298 0.0004872</td>
</tr>
<tr>
<td>Modified directed deviance (cont. corr.)</td>
<td>-2.903 0.0018450</td>
</tr>
</tbody>
</table>

DIAGNOSTICS:
INF NP
0.126 0.234

Approximation based on 20 points

> plot(astro.cond)
Nuisance parameters: \( \theta = (\psi, \lambda) \)

score statistic \( s(\psi) = \ell_p'(\psi)\{j_p(\tilde{\psi})\}^{-1/2} \)

standardized m.l.e. \( q(\psi) = (\tilde{\psi} - \psi)\{j_p(\tilde{\psi})\}^{1/2} \)

likelihood root \( r(\psi) = \sqrt{2\{\ell_p(\tilde{\psi}) - \ell_p(\psi)\}} \)

First order \( p \)-values

\[ p(\psi) = \Phi\{r(\psi)\} \]
\[ = \Phi\{q(\psi)\} \]
\[ = \Phi\{s(\psi)\} \]

Third order \( p \)-values

\[ p(\psi) = \Phi\{r^*(\psi)\} \]
\[ = \Phi(r) + \phi(r) \left( \frac{1}{r} - \frac{1}{Q} \right) \]

\[ r^*(\psi) = r + \frac{1}{r} \log \frac{Q}{r} \]
\[ Q = \]
Steps

– Reduce from dimension $n$ to $p$ by conditioning, as in location family model

– Reduce from dimension $p$ to 1 by marginalizing, as in exponential family model

– Replace exact ancillary by ancillary directions $V$

– Use saddlepoint approximation at the marginalization step
The algorithm

**Have:** – data $y \in R^n$, parameter $\theta \in R^p$, parameter of interest $\psi \in R$

– likelihood function at the observation $\ell(\theta; y^0)$

**Goal:** – scalar function of $y$ that measures $\psi$

– e.g. a quantity $r(y, \psi)$ with a fixed distribution

– the distribution $\Pr\{r(Y, \psi) \geq r(y^0, \psi)\}$ to be known exactly or to a high order of approximation

**Add:** $- z_i = z_i(y_i, \theta)$, a pivotal statistic

Example: $y_i = x'_i\beta + \sigma e_i$, $z_i = e_i$

Example: $F(y_i; \theta) \sim U(0, 1)$
Compute: – the likelihood function and its derivative in the sample space

− $\ell(\theta; y^0)$

− $\varphi(\theta) = \ell; V(\theta; y^0) = \sum \ell; y_i(\theta; y^0)V_i$

$$V_i = -\left(\frac{\partial z_i}{\partial y_i}\right)^{-1}\left(\frac{\partial z_i}{\partial \theta}\right)|_{\hat{\theta}^0}$$

− $\varphi(\theta)$ is the canonical parameter for an approximating exponential family model

− this exponential family model has $p$-dimensional sample space

Compute: – the constrained m.l.e. $\hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi)$

− the likelihood root $r = \pm \sqrt{2[\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)]}$

− an improved version $r^* = r + \frac{1}{r} \log \frac{Q}{r}$

$r^*$ is the scalar variable we seek, $r^* \sim N(0, 1)$
exponential family
\[ f(y_1, \ldots, y_n; \theta) = \exp\{\theta'y_+ - nc(\theta) - \sum d(y_i)\} \]

\[
Q = \frac{\nu(\hat{\theta}) - \nu(\hat{\theta}_\psi)}{\sigma_\nu} = \frac{|\phi(\hat{\theta}) - \phi(\hat{\theta}_\psi)| |\lambda(\hat{\theta})|}{|\phi'(\hat{\theta})| |\lambda'(\hat{\theta})|^{1/2}} \frac{|j_{\theta\theta}(\hat{\theta})|^{1/2}}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|^{1/2}}
\]

\[ \phi(\theta) = \ell;V(\theta; y^0) = \sum \ell; y_i(\theta; y^0)V_i \]

\[ j_{\theta\theta}(\hat{\theta}) \text{ from fitting full model} \]

\[ j_{\lambda\lambda}(\hat{\theta}_\psi) \text{ from fitting reduced model (}\psi\text{ fixed)} \]

derivatives can be computed numerically
Example: Logistic regression (Brazzale, 2000)

- data from Davison and Hinkley (1997, Ch.7)

- binary response (presence of calcium oxalate crystals in urine)

- 77 observations, 6 covariates

- although there are 7 observations per parameter, higher order approximations make a difference

First order $p$-value : 0.01648
Third order $p$-value: 0.02664
Model

\[ \text{logit}(p_i) = \beta_0 + \beta_1 \text{gravity} + \beta_2 \text{ph} + \beta_3 \text{osmo} + \beta_4 \text{cond} + \beta_5 \text{calc} + \beta_6 \text{urea} \]

\[ y_i \sim \text{Bin}(1, p_i) \]

\( \beta_0, \ldots, \beta_5 \) can be exactly eliminated by conditioning on \( \sum y_i, \sum y_i \text{gravity}_i \), etc.

method outlined above is equivalent to 3rd order
A Bayesian version

\[ \Pr_m(\Psi \geq \psi \mid y) \doteq \Phi(r^*) \]

\[ r^* = r + \frac{1}{r} \log \frac{q_B}{r} \]

\[ r = \text{sign}(\hat{\psi} - \psi)[2\{\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\}]^{1/2} \]

\[ q_B = \ell_\theta(\hat{\theta}_\psi) \left\{ \frac{|j_{\lambda\lambda}(\hat{\theta}_\psi)|}{|j_{\theta\theta}(\hat{\theta})|} \right\}^{1/2} \frac{\pi(\hat{\theta})}{\pi(\hat{\theta}_\psi)}. \]

**Example:** simple logistic regression

\[ \text{logit}(p_i) = \alpha + \beta x_i \]

Comparison of 95% confidence intervals for \( \beta \) (Hosmer & Lemeshow, Ex. 1.1; \( y=\text{chd}, x=\text{age}, n=100 \))

<table>
<thead>
<tr>
<th></th>
<th>Lower endpoint</th>
<th>Upper endpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>First order</td>
<td>0.06377</td>
<td>0.1581</td>
</tr>
<tr>
<td>Bayesian ( r^* )</td>
<td>0.06680</td>
<td>0.1598</td>
</tr>
<tr>
<td>Frequentist ( r^* )</td>
<td>0.06524</td>
<td>0.1596</td>
</tr>
</tbody>
</table>
Choice of prior?

- noninformative prior $\pi(\theta) \propto \frac{1}{2} \beta \beta(\theta) g(\eta)$
- $g$ arbitrary
- $\eta$ orthogonal to $\beta$ with respect to expected Fisher information
- $\eta = \eta(\alpha, \beta) = \sum n_i p_i(\alpha, \beta)$
- invariant to choice of $g(\cdot)$ in this setting
**Example:** linear regression, non-normal error

\[ y_i = x_i'\beta + \sigma e_i, \ e_i \sim f(\cdot) \]

Data: House prices (Sen & Srivastava)
\( y = \) selling price, 4 covariates

95% confidence interval for coefficient of frontage:

<table>
<thead>
<tr>
<th></th>
<th>First order</th>
<th></th>
<th>Third order</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Student (3)</td>
<td>-0.07</td>
<td>0.65</td>
<td>-0.16</td>
<td>0.69</td>
</tr>
<tr>
<td>Student (5)</td>
<td>-0.09</td>
<td>0.65</td>
<td>-0.15</td>
<td>0.70</td>
</tr>
<tr>
<td>Student (7)</td>
<td>-0.08</td>
<td>0.66</td>
<td>-0.14</td>
<td>0.70</td>
</tr>
<tr>
<td>Normal</td>
<td>-0.08</td>
<td>0.66</td>
<td>-0.13</td>
<td>0.71</td>
</tr>
</tbody>
</table>
Profile and modified profile log-LIKs

Log likelihood

Profile log-likelihood

modified profile log-likelihood

Coefficient of front

-0.20 -0.05 0.10 0.25 0.40 0.55 0.70

-3.0 -2.8 -2.6 -2.4 -2.2 -2.0 -1.8 -1.6 -1.4 -1.2 -1.0 -0.8 -0.6 -0.4 -0.2 0.0

Profile log-likelihood

modified profile log-likelihood
**Example:** Type-2 censored data (Wong & Wu, 2000, TCS)

Weibull model: \( F(t; \lambda, \beta) = 1 - \exp\left\{ (-t/\lambda)^\beta \right\} \)

data \((y_{(1)} \leq \ldots \leq y_{(r)})\), \(n - r\) units still on test at end of experiment

\[ \text{(12)} \]

---

### Table 1. 90% Confidence Interval for Mean, Standard Deviation, and 0.1 Quantile

<table>
<thead>
<tr>
<th>Method</th>
<th>(\psi(\theta) = \mu)</th>
<th>(\psi(\theta) = \sigma)</th>
<th>(\psi(\theta) = y_{(1)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>AN</td>
<td>((-0.13, 0.44))</td>
<td>((0.66, 1.16))</td>
<td>((-2.70, -1.49))</td>
</tr>
<tr>
<td>LR</td>
<td>((-0.12, 0.48))</td>
<td>((0.70, 1.22))</td>
<td>((-2.62, -1.37))</td>
</tr>
<tr>
<td>Third-order</td>
<td>((-0.11, 0.51))</td>
<td>((0.72, 1.28))</td>
<td>((-2.71, -1.41))</td>
</tr>
<tr>
<td>Exact</td>
<td>((-0.11, 0.51))</td>
<td>((0.72, 1.28))</td>
<td>((-2.71, -1.41))</td>
</tr>
</tbody>
</table>
Other examples

– transformed regression $y^\lambda = x'\beta + \sigma e$

– generalized linear models: Poisson, gamma, binomial

– mean of a log-normal distribution $\psi = \mu + \sigma^2/2$

– nonlinear regression with normal error, non-constant variance (Brazzale)

– linear mixed models (Brazzale & Guolo)
Extensions not (yet) available

– complex models: hierarchical models, time dependencies, spatial dependencies

– **robustness** or model dependence

– large datasets, especially with $p >> n$
Some references


www.stat.unipd.it/~LIKASY


www.utstat.utoronto.ca/reid

www.utstat.utoronto.ca/dfraser
As an example consider sampling from the bivariate normal distribution with unknown parameter \((\theta)\) (Cox and Hinkley, 1974, Ex. 2.30; Barndorff-Nielsen, 1978). The log-likelihood function is

\[
\ell(\theta) = -\frac{n}{2} \log(1 - \theta^2) - \frac{1}{2(1 - \theta^2)} \sum (y_{1i}^2 + y_{2i}^2) + \frac{\theta}{1 - \theta^2} \sum
\]
<table>
<thead>
<tr>
<th>(\theta)</th>
<th>Exact</th>
<th>left tail</th>
<th>right tail</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.001</td>
<td>0.005</td>
</tr>
<tr>
<td>-0.9</td>
<td>(r)</td>
<td>0.0022</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(r^*)</td>
<td>0.0010</td>
<td>0.005</td>
</tr>
<tr>
<td>-0.7</td>
<td>(r)</td>
<td>0.0019</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(r^*)</td>
<td>0.0009</td>
<td>0.005</td>
</tr>
<tr>
<td>-0.5</td>
<td>(r)</td>
<td>0.0022</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(r^*)</td>
<td>0.0010</td>
<td>0.005</td>
</tr>
<tr>
<td>-0.3</td>
<td>(r)</td>
<td>0.0022</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>(r^*)</td>
<td>0.0011</td>
<td>0.005</td>
</tr>
<tr>
<td>0</td>
<td>(r)</td>
<td>0.0022</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>(r^*)</td>
<td>0.0011</td>
<td>0.005</td>
</tr>
</tbody>
</table>