Accurate directional inference for vector parameters

Nancy Reid

June 23, 2016

with Don Fraser, Nicola Sartori, Anthony Davison
## Parametric models and likelihood

- **model** $f(y; \theta)$,
- **data** $y = (y_1, \ldots, y_n)$
- **log-likelihood function**
- **parameter of interest** $\theta \in \mathbb{R}^p$

### independent observations

$p < n$

- **log-likelihood function**
  
  \[ \ell(\theta; y) = \log f(y; \theta) + a(y) \]

- **parameter of interest**
  \[ \theta = (\psi, \lambda), \quad \psi \in \mathbb{R}^d \]

### likelihood inference

\[ w(\psi) = 2\{\ell(\hat{\psi}, \hat{\lambda}) - \ell(\psi, \hat{\lambda}_\psi)\} \sim \chi^2_d \]

### Likelihood Contours

- **Likelihood Contours**

### Variance Reduction

\[ w_B(\psi) = \frac{w(\psi)}{1 + B(\psi)} \sim \chi^2_d \quad O_p(n^{-2}) \]

\[ B(\psi) = \mathbb{E}\{w(\psi)\}/d \]

\[ w^*(\psi) = w(\psi) \left\{ 1 - \frac{\log \gamma(\psi)}{w(\psi)} \right\} \]

Skovgaard, 2001
Example: $2 \times 3$ contingency table

- activity amongst psychiatric patients

<table>
<thead>
<tr>
<th></th>
<th>Affective disorders</th>
<th>Schizophrenics</th>
<th>Neurotics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retarded</td>
<td>12</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>Not retarded</td>
<td>18</td>
<td>17</td>
<td>25</td>
</tr>
</tbody>
</table>

- model: log-linear $y \sim \text{Poisson}$, $\log\{E(y)\} = X\theta$, $\theta \in \mathbb{R}^6$

- log-likelihood $\ell(\theta; y) = \theta^T X^T y - 1^T e^{X\theta} = \theta^T u - c(\theta)$

- $\theta = (\psi, \lambda)$, $\psi \in \mathbb{R}^2$, $\lambda \in \mathbb{R}^4$

- $(\psi_1, \psi_2)$ interaction parameters

- $H_0 : \psi = \psi_0 = (0, 0)$ independence

- log-likelihood $\ell(\psi, \lambda; y) = \psi^T u_1 + \lambda^T u_2 - c(\psi, \lambda)$
Testing $\psi = 0$

$$w(\psi_0) \sim \chi^2_2$$

$p$-value 0.047

$w^*(\psi_0)$

0.048

directional 0.050

exact conditional 0.051
Linear exponential families

- model \( f(y; \theta) = \exp\{\varphi(\theta)^T u(y) - c\{\varphi(\theta)\} - d(y)\} \quad y_1, \ldots, y_n \)
i.i.d.

- sufficient statistic \( f(u; \theta) = \int_{y: u(y) = u} f(y; \theta) dy = \exp\{\varphi(\theta)^T u - nc\{\varphi(\theta)\} - \tilde{d}(u)\} \)

- reduce dimension from \( n \) to \( p \) by marginalization

- linear parameter of interest \( \varphi(\theta) = \theta = (\psi, \lambda) \)

- model \( f(u_1, u_2; \psi, \lambda) = \exp\{\psi^T u_1 + \lambda^T u_2 - nc(\psi, \lambda) - \tilde{d}(u)\} \)

- conditional density
  \( f(u_1 \mid u_2; \psi) = \exp\{\psi^T u_1 - n\tilde{c}_2(\psi) - \tilde{d}_2(u_1)\} \)

- reduce dimension from \( n \) to \( d \) by conditioning
... conditional density

\[ f(u_1 \mid u_2; \psi) = \exp\{\psi^T u_1 - n\tilde{c}_2(\psi) - \tilde{d}_2(u_1)\} \]

\[ f(u_1 \mid u_2; \psi) = \frac{f(u_1, u_2; \psi, \lambda)}{f(u_2; \psi, \lambda)} \propto f(u; \psi, \lambda) \quad \text{with } u_2 \text{ held fixed} \]

\[ u_2 \text{ held fixed } \iff \hat{\lambda}_\psi \text{ held fixed} \quad \hat{\theta}_\psi = (\psi, \hat{\lambda}_\psi) \]

centering: \[ s = u - u^0 \quad s^0 = s(y^0) = 0 \]

saddlepoint approximation:

\[ f(s_1 \mid s_2; \psi) \doteq c \exp[\ell(\hat{\theta}^0_\psi; s) - \ell(\hat{\theta}(s); s)]|j\{\hat{\theta}(s); s\}|^{-1/2}, \quad s \in L^0_\psi \]

plane \[ L^0_\psi = \{s \in \mathbb{R}^p : s_2 = 0\} = \{s \in \mathbb{R}^p : \hat{\lambda}_\psi = \hat{\lambda}^0_\psi\} \]

tilted log-likelihood: \[ \ell(\theta; s) = \psi^T s_1 + \lambda^T s_2 + \ell(\theta; y^0) \]

\[ \hat{\theta}(s) : \partial \ell(\theta; s)/\partial \theta = 0 \quad \text{m.l.e. as a function of the variable} \]
... conditional density

\[ f(s_1 \mid s_2; \psi) \doteq c \exp\left[ \ell(\hat{\theta}_0; s) - \ell(\hat{\theta}(s); s) \right] |j\{\hat{\theta}(s); s}\}|^{-1/2}, \quad s \in L_\psi^0 \]

<table>
<thead>
<tr>
<th>Affective disorders</th>
<th>Schizophrenics</th>
<th>Neurotics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Retarded</td>
<td>12</td>
<td>13</td>
</tr>
<tr>
<td>Not retarded</td>
<td>18</td>
<td>17</td>
</tr>
</tbody>
</table>

\( s_1 \in \mathbb{R}^2, \ s_2 \in \mathbb{R}^4 \)  

\( L_\psi^0: \) all 2 × 3 tables with the same row and column totals  

full model Poisson
Directional tests

- measure the **directed departure** from $H_0$ in $L^0_{\psi}$

$s_2 = 0$ on $L^0_{\psi}$

- $s_\psi$: expected value of $s_1$, under $H_0$

- $s^0$: observed value of $s_1$ = 0 from centering

- $L^*_\psi$: line through these two points

$L^*_\psi = ts^0 + (1 - t)(s_\psi - s^0), \quad t \in \mathbb{R}$

Relative log likelihood

\[ \psi_1 \]

\[ \psi_2 \]

-3 -2 -1 0 1 2 3

-20 2 4 6

S1

S2

S(t)
... directional tests \[ L^*_{\psi} = ts^0 + (1 - t)s_{\psi} \]

null hypothesis of independence \( t = 0 \)
observed value of \( s \) \( t = 1 \)

\( p \)-value compares probability from \( \times \) to \( \infty \) to that from \( \bigcirc \) to \( \infty \)
along the line in the sample space

like a 2-sided \( p \)-value \( \Pr(\text{response} > \text{observed} | \text{response} > 0) \)
... directional $p$-value

- **$p$-value**
  
  $$ p-value = \frac{\int_{\infty}^{\infty} t^{d-1} f\{s(t); \psi\} dt}{\int_{0}^{\infty} t^{d-1} f\{s(t); \psi\} dt} $$

  $s(t)$ along the line $L^*_\psi$

- $t^{d-1}$ from change to polar coordinates $||s||$, conditional on $s/||s||$

- **Simplifications:**

  1: $L^*_\psi \subset L^0_\psi \subset \mathbb{R}^p$

  2: ratio of two integrals – drop any terms that don’t depend on $t$

  3: saddlepoint approximation

  $$ f(s; \psi) \doteq c \exp[\ell(\hat{\theta}^0_\psi; s) - \ell\{\hat{\theta}(s); s\}]|j\{\hat{\theta}(s); s\}|^{-1/2}, \quad s \in L^0_\psi $$
Introduction
Linear exponential families
Directional testing
More general models
Conclusion

$2 \times 3$ table

\begin{align*}
\begin{array}{ccc}
t = 0 & \quad t = 0.5 & \quad t = 1 \\
10 & 10 & 10 \\
20 & 20 & 20 \\
11.0 & 11.5 & 7.5 \\
19.0 & 18.5 & 22.5 \\
12 & 13 & 5 \\
18 & 17 & 25 \\
14 & 16 & 0 \\
16 & 14 & 30
\end{array}
\end{align*}
... $2 \times 3$ table
### 2 × 3 table simulations

<table>
<thead>
<tr>
<th></th>
<th>Nominal</th>
<th>0.010</th>
<th>0.025</th>
<th>0.050</th>
<th>0.100</th>
<th>0.250</th>
<th>0.500</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRT</td>
<td>0.011</td>
<td>0.028</td>
<td>0.055</td>
<td>0.107</td>
<td>0.260</td>
<td>0.510</td>
<td></td>
</tr>
<tr>
<td>Skovgaard, 2001</td>
<td>0.010</td>
<td>0.024</td>
<td>0.050</td>
<td>0.100</td>
<td>0.250</td>
<td>0.501</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Nominal</th>
<th>0.750</th>
<th>0.900</th>
<th>0.950</th>
<th>0.975</th>
<th>0.990</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRT</td>
<td>0.757</td>
<td>0.905</td>
<td>0.952</td>
<td>0.974</td>
<td>0.992</td>
<td></td>
</tr>
<tr>
<td>Skovgaard, 2001</td>
<td>0.752</td>
<td>0.902</td>
<td>0.950</td>
<td>0.973</td>
<td>0.992</td>
<td></td>
</tr>
</tbody>
</table>
Example: comparison of normal variances

\begin{figure}
\centering
\includegraphics[width=\textwidth]{example.png}
\end{figure}

Likelihood ratio statistic \quad 0.0042

Directional \quad 0.0389

Skovgaard, 2001 \quad 0.0622

Bartlett’s test \quad 0.0136

$F$-test
Simulations

3 groups, 10 observations per group
Simulations

3 groups, 5 observations per group
Simulations

1000 groups, 5 observations per group

dimension $\psi$ 999; dimension nuisance 1001
Example: covariance selection

- model \( y_i \sim N_q(\mu, \Lambda^{-1}) \)
  - inverse covariance matrix

- linear exponential family
  \[
  \ell(\theta; y) = \frac{n}{2} \log |\Lambda| - \frac{1}{2} \text{tr}(\Lambda y^T y) + 1^T y \xi - \frac{n}{2} \xi^T \Lambda \xi
  \]
  \( \theta = (\xi, \Lambda) = (\Lambda \mu, \Lambda) \)

- \( H_0: \) some off-diagonal elements of \( \Lambda \) are 0
  \( \psi_1 = \cdots = \psi_d = 0 \)
  - conditional independence

- need constrained m.l.e: use fitConGraph in ggm

- \( \hat{\Lambda}^{-1}(t) = t\hat{\Lambda}^{-1} + (1 - t)\hat{\Lambda}_0^{-1} \)
  - m.l.e. along the line

- \( f\{s(t); \psi\} \propto |t\hat{\Lambda}^{-1} + (1 - t)\hat{\Lambda}_0^{-1}|(n-q-2)/2 \)
... covariance selection

<table>
<thead>
<tr>
<th></th>
<th>0.010</th>
<th>0.025</th>
<th>0.050</th>
<th>0.100</th>
<th>0.250</th>
<th>0.500</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>0.011</td>
<td>0.026</td>
<td>0.050</td>
<td>0.101</td>
<td>0.248</td>
<td>0.498</td>
</tr>
<tr>
<td>LRT</td>
<td>0.055</td>
<td>0.105</td>
<td>0.170</td>
<td>0.270</td>
<td>0.487</td>
<td>0.730</td>
</tr>
<tr>
<td>Directional</td>
<td>0.007</td>
<td>0.018</td>
<td>0.036</td>
<td>0.074</td>
<td>0.196</td>
<td>0.422</td>
</tr>
<tr>
<td>Skovgaard, 2001</td>
<td>0.678</td>
<td>0.852</td>
<td>0.919</td>
<td>0.955</td>
<td>0.980</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>0.750</th>
<th>0.900</th>
<th>0.950</th>
<th>0.975</th>
<th>0.990</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>0.749</td>
<td>0.899</td>
<td>0.949</td>
<td>0.974</td>
<td>0.990</td>
</tr>
<tr>
<td>LRT</td>
<td>0.895</td>
<td>0.967</td>
<td>0.985</td>
<td>0.994</td>
<td>0.998</td>
</tr>
<tr>
<td>Directional</td>
<td>0.678</td>
<td>0.852</td>
<td>0.919</td>
<td>0.955</td>
<td>0.980</td>
</tr>
<tr>
<td>Skovgaard, 2001</td>
<td>0.750</td>
<td>0.900</td>
<td>0.950</td>
<td>0.975</td>
<td>0.990</td>
</tr>
</tbody>
</table>

Covariance matrix $11 \times 11$; dimension of $\psi = 45$

first order Markov dependence
Nonlinear exponential families

- \( f(y; \theta) = \exp\{\varphi(\theta)^T u(y) - c(\varphi) - d(y)\} \)

- \( f(u; \theta) = \exp\{\varphi(\theta)^T u - nc(\varphi) - \tilde{d}(u)\} \quad n \downarrow p \)

- centering: \( s = u - u^0 \),
  tilted log-likelihood \( \ell(\theta; s) = \varphi(\theta)^T s + \ell(\theta; y^0) \)

- parameter of interest: \( \psi = \psi(\theta) = \psi(\theta(\varphi)) \)

- no reduction in dimension by conditioning

- can eliminate nuisance parameter by Laplace integration
Testing a value for $\psi$

- $\theta = (\psi, \lambda); \varphi = \varphi(\theta)$ is the canonical parameter of the TEM

- with $\psi$ fixed by the hypothesis, we can integrate out the nuisance parameter by Laplace approximation

- define the same plane in the score space $L_0^\psi$, where $\hat{\varphi}_\psi$ is fixed at its observed value

$$f(s; \psi) = c \exp\{\ell(\hat{\varphi}_\psi^0; s) - \ell(\hat{\varphi}(s); s)\}|j_{\varphi\varphi}(\hat{\varphi}(s); s)|^{-1/2}|j_{\lambda\lambda}(\hat{\varphi}_\psi^0; s)|^{1/2},$$

$s \in L_0^\psi$

$$p\text{-value} = \frac{\int_1^\infty t^{d-1}f\{s(t); \psi\}dt}{\int_0^\infty t^{d-1}f\{s(t); \psi\}dt}$$

$s(t)$ along the line $L_\psi^*$
Example: marginal independence

- \( y_i \sim N_q(\mu, \Sigma), \quad H_0 : \) some entries of \( \Sigma \) are 0

- estimate \( \Sigma \) under \( H_0 \) using fitCovGraph

- \( \ell(\Sigma) = \frac{n-1}{2} \left[ \log |\varphi(\Sigma)| - \frac{n-1}{2} \text{tr} \{ \varphi(\Sigma)S\} \right], \quad \varphi(\Sigma) = \Sigma^{-1} \)

- \( S = \) sample covariance

- \( f\{s(t); \psi\} \propto |t\hat{\Sigma} + (1 - t)\hat{\Sigma}_0|^{(n-q-2)/2} |j_{(\lambda\lambda)}\{\hat{\Sigma}_0; s(t)\}|^{1/2} \)

- \( j_{(\lambda\lambda)} \) needs \( \partial \ell(\Sigma) / \partial (\sigma_{jk}) \) for the non-zero elements

- \( j_{\lambda j\lambda k}\{\hat{\Sigma}_0; s(t)\} = \)

  \[ \frac{n-1}{2} \left( \text{tr}(A_{kj} + t \left[ \text{tr}\{(A_{kj} + A_{jk})(\hat{\Sigma}_0^{-1}\hat{\Sigma} - I_q)\} \right] \right) \]

- \( A_{kj} = \Sigma_0^{-1}(\partial \Sigma / \partial \lambda_k)\Sigma_0^{-1}(\partial \Sigma / \partial \lambda_j) \)
<table>
<thead>
<tr>
<th></th>
<th>Nominal</th>
<th>0.010</th>
<th>0.025</th>
<th>0.050</th>
<th>0.100</th>
<th>0.250</th>
<th>0.500</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRT</td>
<td>1.00</td>
<td></td>
<td></td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Directional</td>
<td>0.007</td>
<td>0.024</td>
<td>0.049</td>
<td>0.097</td>
<td>0.250</td>
<td>0.497</td>
<td></td>
</tr>
<tr>
<td>Skovgaard, 2001</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.006</td>
<td>0.026</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Nominal</th>
<th>0.750</th>
<th>0.900</th>
<th>0.950</th>
<th>0.975</th>
<th>0.990</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRT</td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td></td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>Directional</td>
<td>0.749</td>
<td>0.904</td>
<td>0.951</td>
<td>0.975</td>
<td>0.990</td>
<td></td>
</tr>
<tr>
<td>Skovgaard, 2001</td>
<td>0.099</td>
<td>0.225</td>
<td>0.333</td>
<td>0.440</td>
<td>0.570</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Nominal</th>
<th>0.010</th>
<th>0.025</th>
<th>0.050</th>
<th>0.100</th>
<th>0.250</th>
<th>0.500</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRT</td>
<td>0.065</td>
<td>0.122</td>
<td>0.199</td>
<td>0.311</td>
<td>0.538</td>
<td>0.786</td>
<td></td>
</tr>
<tr>
<td>Directional</td>
<td>0.010</td>
<td>0.024</td>
<td>0.049</td>
<td>0.098</td>
<td>0.251</td>
<td>0.496</td>
<td></td>
</tr>
<tr>
<td>Skovgaard, 2001</td>
<td>0.003</td>
<td>0.010</td>
<td>0.022</td>
<td>0.050</td>
<td>0.150</td>
<td>0.352</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Nominal</th>
<th>0.750</th>
<th>0.900</th>
<th>0.950</th>
<th>0.975</th>
<th>0.990</th>
</tr>
</thead>
<tbody>
<tr>
<td>LRT</td>
<td>0.927</td>
<td>0.979</td>
<td>0.991</td>
<td>0.997</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>Directional</td>
<td>0.751</td>
<td>0.898</td>
<td>0.950</td>
<td>0.975</td>
<td>0.989</td>
<td></td>
</tr>
<tr>
<td>Skovgaard, 2001</td>
<td>0.622</td>
<td>0.819</td>
<td>0.894</td>
<td>0.943</td>
<td>0.974</td>
<td></td>
</tr>
</tbody>
</table>
Tangent exponential model

- Every model \( f(y; \theta) \) on \( \mathbb{R}^n \) can be approximated by an exponential family model:

\[
f_{TEM}(s; \theta) = \exp\{\varphi(\theta)^T s + \ell^0(\theta)\} h(s)
\]

- \( s \) is a score variable on \( \mathbb{R}^p \)
- \( \ell^0(\theta) = \ell(\theta; y^0) \) is the observed log-likelihood function
- \( \varphi(\theta) = \varphi(\theta; y^0) \) is the canonical parameter \( \in \mathbb{R}^p \) to be described
- matches log-likelihood function at \( y^0 \), and its first derivative on the sample space, at \( y^0 \)
- implements conditioning on an approximate ancillary statistic by construction contrast with exp fam
Aside: canonical parameter $\varphi(\theta)$

- if $f(y; \theta)$ is an exponential family, $\varphi$ is sitting in the model

- if not find a pivotal quantity $z_i = z_i(y_i; \theta)$ with a fixed distribution

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

$$\varphi(\theta) = \varphi(\theta; y^0) = \sum_{i=1}^{n} \frac{\partial \ell(\theta; y^0)}{\partial y_i} V_i = \ell; V(\theta; y^0)$$

example: $(y_i - \mu)/\sigma$

a vector of length $p$
Example: Box-Cox model for regression

- \( y_i(\gamma) = x_i^T \beta + \sigma z_i, \quad i = 1, \ldots, n \)

- \( y_i(\gamma) = \begin{cases} \frac{(y_i^\gamma - 1)}{\gamma}, & \gamma \neq 0 \\ \log y_i, & \gamma = 0 \end{cases} \)

- \( \varphi(\theta) = \varphi(\theta; y^0) = \sum_{i=1}^{n} \frac{\partial \ell(\theta; y^0)}{\partial y_i} V_i \)

- \[
\frac{\partial \ell_i(\theta)}{\partial y_i} = -\frac{\{y_i(\gamma) - x_i^T \beta\}}{\sigma^2} \frac{\partial y_i(\gamma)}{\partial y_i} + \frac{\gamma - 1}{y_i}, \\
V_i = y_i^{1-\hat{\gamma}} \left[ x_i^T, \frac{y_i(\hat{\gamma}) - x_i^T \hat{\beta}}{\hat{\sigma}}, \frac{y_i^{\hat{\gamma}} - \hat{\gamma} y_i^{\hat{\gamma}} \log y_i - 1}{\hat{\gamma}^2} \right], \]

row vector of length \( p \)
Simulations

3 × 4 factorial with 4 replications

\[ \text{dim}(\theta) = 14; \quad \text{dim}(\psi) = 6; \quad H_0: \text{no interaction} \]
Conclusion

- different way to assess vector parameters

- incorporates information in the direction of departure

- easy to compute: two model fits, plus 1-d numerical integration

- accurate conditionally, by construction, and unconditionally

- can be used in models of practical interest

- exponential family model not necessary – easily generalized using approximate exponential family model
... conclusion
... conclusion
... conclusion
References


Davison, A.C., Fraser, D.A.S., Reid, N., Sartori, N. (2014). Accurate directional inference for vector parameters in linear exponential families. *JASA*