Nonparametric Tests for Multi-parameter M-estimators

John Robinson
School of Mathematics and Statistics
University of Sydney

The talk is based on joint work with John Kolassa. It follows from work over a number of years with Chris Field, Elvezio Ronchetti and Alastair Young and generalizes Robinson, Ronchetti and Young (2003) and Field, Robinson and Ronchetti (2008).
Abstract  Tests of hypotheses concerning subsets of multivariate means or coefficients in linear or generalized linear models depend on parametric assumptions which may not hold. One nonparametric approach to these problems uses the standard nonparametric bootstrap using the test statistics derived from some parametric model but basing inferences on bootstrap approximations. We derive different test statistics based on empirical exponential families and use a tilted bootstrap to give inferences. The bootstrap approximations can be accurately approximated to relative second order accuracy by a saddlepoint approximation. This generalises earlier work in two ways. First, we generalise from bootstraps based on resampling vectors of both response and explanatory variables to include bootstrapping residuals for fixed explanatory variables, and second, we obtain a theorem for tail probabilities under weak conditions justifying approximation to bootstrap results for both cases.
A likelihood based test. Let $Y_1(\theta), \cdots, Y_n(\theta)$ be iid random vectors with $Y_j(\theta)$ from a distribution $F$ on the sample space $\mathcal{Y}$. Suppose that $\theta$ satisfies

$$E\sum_{j=1}^{n} \psi_j(Y_j(\theta), \theta) = 0, \quad (1)$$

Let $\theta = (\theta_1, \theta_2)$, where $\theta_1 \in \mathbb{R}^{p_1}$ and $\theta_2 \in \mathbb{R}^{p_2}$, $p_1 + p_2 = p$, and consider a test of the null hypothesis

$$H_0 : \theta_2 = \theta_{20}.$$

If the common distribution of $Y_j(\theta)$ belongs to some parametric model, then $F$ belongs to a class of distributions such that (1) holds, with $\theta_2 = \theta_{20}$, and likelihood ratio tests may be used.
Consider test statistics based on $T = (T_1, T_2)$, the M-estimate of $\theta$, the solution of

$$\sum_{j=1}^{n} \psi_j(Y_j(\theta), T) = 0,$$

where $\psi_j$ are assumed to be smooth functions from $\mathcal{Y} \times \mathbb{R}^p$ to $\mathbb{R}^p$.

The functions $\psi_j$ are often from some parametric model, sometimes adjusted to make estimates robust. For the present we ignore robustness.

We will propose a test statistic $h(T_2)$. 
Three examples

In linear models and generalized linear models with response variables \(Z_j\) and explanatory variables \(X_j\).

Correlation model:

For iid random vectors \(Y_j(\theta) = (Z_j, X_j)\),

\[ Y_j(\theta) = (Z_j, X_j) \]

and the estimating equations are

\[ \sum_{j=1}^{n} \psi_j(Y_j(\theta), t) = \sum_{j=1}^{n} X_j(Z_j - t^T X_j) = 0. \]
Regression model:

For fixed $X_j = x_j$, as in the case of a designed experiment. Here the residuals,

$$Y_j(\theta) = Z_j - \theta^T x_j,$$

are iid random variables and the estimating equations are

$$\sum_{j=1}^{n} \psi_j(Y_j(\theta), t) = \sum_{j=1}^{n} x_j(Y_j(\theta) + (\theta - t)^T x_j)$$

$$= \sum_{j=1}^{n} x_j(Z_j - t^T x_j) = 0.$$
Generalized Linear Models: Poisson case

Response variables \( Z_j \), for \( j = 1, \ldots, n \) such that \( EZ_j = \mu_j \) and \( var(Z_j) = V(\mu_j) \) for \( j = 1, \ldots, n \) and

\[
\eta_j = \log(\mu_j) = X_j^T \theta, \quad i = 1, \ldots, n,
\]

Consider only the correlation case with \( Y_j(\theta) = (Z_j, X_j) \) iid random variables. The estimating equations are

\[
\sum_{j=1}^{n} \psi_j(Y_j(\theta), t) = \sum_{j=1}^{n} X_j(Z_j - e^{X_j^T t}) = 0.
\]
Test statistic for parametric model: $Y_j(\theta)$ has distribution $F$. Put

$$nK(\tau, t) = \sum_{j=1}^{n} K_j(\tau, t) = \sum_{j=1}^{n} \log \mathbb{E}[\exp(\tau^T \psi_j(Y_j(\theta), t))].$$

(3)

Consider a test statistic based on the function $h(.)$ defined by

$$h(t_2) = \inf_{t_1} \sup_{\tau} \{-K(\tau, t)\} = -K(\tau(t(t_2)), t(t_2)),$$

(4)

where $t(t_2) = (t_1(t_2), t_2)$ for

$$\tau(t) = \arg \sup_{\tau} [-K(\tau, t)] \text{ and } t_1(t_2) = \arg \inf_{t_1} [-K(\tau(t), t)].$$

Note $h(\theta_{20}) = 0$, $h'(\theta_{20}) = 0$ and $h''(\theta_{20}) > 0$, so $h(t_2) \geq 0.$
Tilted bootstrap: $F$ and so $K$ unknown. Consider weighted empirical distributions

$$\hat{F}(x) = \sum_{k=1}^{n} w_k I(Y_k(\theta) \leq x),$$

where $\theta = (\theta_1, \theta_2)$ and $w_k$ minimize

$$\sum_{k=1}^{n} w_k \log(nw_k),$$

subject to

$$\sum_{k=1}^{n} w_k \frac{1}{n} \sum_{j=1}^{n} \psi_j(Y_k(\theta), \theta) = 0 \text{ and } \sum_{k=1}^{n} w_k = 1,$$ (5)
So we find stationary values of

\[ \sum_{k=1}^{n} w_k \log(nw_k) - \beta^T \sum_{k=1}^{n} w_k \frac{1}{n} \sum_{j=1}^{n} \psi_j(Y_k(\theta), \theta) + \gamma\left(\sum_{k=1}^{n} w_k - 1\right) \]  

(6)

with respect to \( w_k, \beta, \gamma \) and \( \theta_1 \), with \( \theta_2 = \theta_{20} \). Equating derivatives wrt \( w_k \) to 0 gives

\[ w_k = \frac{1}{n} \exp\left(\beta^T \sum_{j=1}^{n} \psi_j(Y_k(\theta), \theta)/n - \kappa(\beta, \theta)\right), \]  

(7)

where

\[ \kappa(\beta, \theta) = \log \frac{1}{n} \sum_{k=1}^{n} \exp\left(\beta^T \sum_{j=1}^{n} \psi_j(Y_k(\theta), \theta)/n\right). \]  

(8)
Then (6) reduces to

$$\sum_k w_k \log(nw_k) = -\kappa(\beta, \theta).$$

So the minimum of (6) under the constraints is $-\kappa(\beta(\theta_0), \theta_0)$, where $\theta_0 = (\theta_1(\theta_20), \theta_20)$,

$$\beta(\theta) = \arg \sup_{\beta} -\kappa(\beta, \theta) \text{ and } \theta_1(\theta_20) = \arg \inf_{\theta_1} -\kappa(\beta(\theta), \theta)$$

and the minimizing weights are

$$\hat{w}_k = \exp \left( \beta(\theta_0)^T \sum_{j=1}^n \psi_j(Y_k(\theta_0), \theta_0)/n - \kappa(\beta(\theta_0), \theta_0) \right).$$
Test statistic for the tilted bootstrap.

To test $H_0: \theta_2 = \theta_{20}$, take the empirical exponential family,

$$\hat{F}(y) = \sum_{k=1}^{n} \hat{w}_k I\{Y_k(\theta_0) \leq y\}. \quad (9)$$

Draw $Y_1^*, \cdots, Y_n^*$ from $\hat{F}$, then $\sum_{j=1}^{n} \psi_j(Y_j^*, t)$ has cumulant generating function

$$n\hat{K}(\tau; t) = \sum_{j=1}^{n} \log \sum_{k=1}^{n} \hat{w}_k \exp \left( \tau^T \psi_j(Y_k(\theta_0); t) \right), \quad (10)$$
Then the test statistic is based on the function $\hat{h}(.)$ defined by

$$
\hat{h}(t_2) = \inf_{t_1} \sup_{\tau} \left[ -\hat{K}(\tau; t) \right] = -\hat{K}(\tau(t(t_2)); t(t_2)),
$$

(11)

where $t(t_2) = (t_1(t_2), t_2)$,

$$
\tau(t) = \arg \sup_{\tau} [-\hat{K}(\tau, t)] \text{ and } t_1(t_2) = \arg \inf_{t_1} [-\hat{K}(\tau(t); t)].
$$

Again $\hat{h}(\theta_{20}) = 0$, $\hat{h}'(\theta_{20}) = 0$ and $\hat{h}''(\theta_{20}) > 0$, so $\hat{h}(t_2) \geq 0$.

Obtain $B$ tilted bootstrap samples and for each obtain $T^* = (T^*_1, T^*_2)$ the solution of $\sum_{j=1}^{n} \psi_j(Y^*_j, t) = 0$. Calculate $\hat{h}(T^*_2)$ and the p-value for the test is the proportion of $\hat{h}(T^*_2)$ greater than $\hat{h}(t_2)$. Note that $\hat{h}(T^*_2)$ needs to be calculated from (11) for each $T^*_2$. 


A general saddlepoint approximation for multivariate M-estimates

Under mild assumptions on existence of $K(\tau, t)$ and smoothness of scores, we have two approximations, the generalised Lugannani-Rice form

$$P(2h(T_2) \geq u^2) = \bar{Q}_{p_2}(nu^2)[1 + O(n^{-1})] + n^{-1}c_n u^{p_2} e^{-nu^2/2} \left[ \frac{G(u) - 1}{u^2} \right]$$

and the generalised Barndorff-Nielsen form

$$P(2h(T_2) \geq u^2) = \bar{Q}_{p_2}(n\hat{u}^2)[1 + O(n^{-1})]$$

where $\hat{u} = u - \log G(u)/nu$. 
Here \( Q_{p_2} = 1 - \bar{Q}_{p_2} \) is the distribution function of a chi-squared variate with \( p_2 \) degrees of freedom, \( c_n \) a known constant, \( \hat{u} = u - \log G(u)/nu \), and

\[
G(u) = \int_{S_{p_2}} \delta(u, s) ds = 1 + u^2 k(u), \tag{12}
\]

for \( \delta(u, s) \) a somewhat complicated function involving derivatives of the cumulant generating function and \((r, s)\) are the polar coordinates of \( \hat{h}''(\theta_{20})^{-1/2}(t_2 - \theta_{20}) \).

These results may also be shown to hold for the bootstrap where here the Monte Carlo approximation requires heavy calculation at each step.
Application to Special Cases

For the correlation model for both linear models and generalized linear models, the values $Y_j(\theta)$ are $(Z_j, X_j)$ and the score functions $\psi_j$ do not depend on $j$.

So, replacing the general score function $\psi_j(Y_j(\theta), t)$ by $X_j(Z_j - t^T X_j)$ and $X_j(Z_j - e^{t^T X_j})$, respectively, we can proceed as in the general case.
Regression model: For the regression model we will, without loss of
genearlity, consider the case where the first element of $x_k$ is 1 and $\bar{x} = (1, 0, \cdots, 0)^T$. The first constraint of (5) is then

$$\sum_{k=1}^{n} w_k (1, 0)^T (Z_k - \theta^T x_k) = 0,$$

leading to $w_k = \frac{1}{n} e^{\beta_1 (Z_k - \theta^T x_k) - \kappa(\beta_1, \theta)}$, from (7) and (8), where

$$\kappa(\beta_1, \theta) = \log \frac{1}{n} \sum_{k=1}^{n} e^{\beta_1 (Z_k - \theta^T x_k)}.$$

If $\beta = (\beta_1, \beta_2^T)^T$, the $p - 1$ vector $\beta_2$ is not estimable.
Now we may take \( \beta(\theta) = (\beta_1(\theta), 0)^T \) as the solution to

\[
\frac{\partial \kappa(\beta_1, \theta)}{\partial \beta_1} = \sum_{k=1}^{n} (Z_k - \theta^T x_k) e^{\beta_1(Z_k - \theta^T x_k)} = 0. \tag{13}
\]

Also, \( \theta_0 \) is the solution to \( d\kappa(\beta_1(\theta), \theta)/d\theta_1 = 0 \), so

\[
\frac{d\beta_1(\theta)}{d\theta_1} \frac{1}{n} \sum_{k=1}^{n} (Z_k - \theta^T x_k) e^{\beta_1(\theta)(Z_k - \theta^T x_k)}
\]

\[
-\beta_1(\theta) \frac{1}{n} \sum_{k=1}^{n} x_k^{(1)} I(|Z_k - \theta^T x_k| < b) e^{\beta_1(\theta)(Z_k - \theta^T x_k)} = 0,
\]

where \( x_k^{(1)} \) contains the first \( p_1 \) elements of \( x_k \).
The first term here is zero from (13), so $\beta_1(\theta_0) = 0$ and it follows that $w_k = 1/n$. We require that $\theta_0$ satisfies $\sum_{k=1}^{n}(Z_k - \theta(\theta_0)^T x_k) = 0$. This does not uniquely define $\theta_0$, but we propose using the solution to

$$\sum_{k=1}^{n} x_k^{(1)}(Z_k - \theta_0^T x_k) = 0.$$  \hfill (14)

Bootstrap replicates $R_1^*, \ldots, R_n^*$ are obtained by sampling from the empirical distribution

$$\hat{F}_{\theta_0}(x) = \sum_{k=1}^{n} \frac{1}{n} I(Z_k - \theta_0^T x_k \leq x).$$ \hfill (15)
Then
\[ \psi_j(R_j^*, t) = x_j(R_j^* + (\theta_0 - t)^T x_j). \]

A Monte Carlo approximation to the tail probability of the test statistic can be obtained exactly as for the correlation case. The cumulant generating function of \( \sum_{j=1}^{n} \psi_j(R_j^*, t) \) is

\[
\hat{K}(\tau, t) = \sum_{j=1}^{n} \log \frac{1}{n} \sum_{k=1}^{n} e^{\tau^T x_j(Z_k - \theta_0^T x_k + (\theta_0 - t)x_j)}
\]

and we can calculate the statistic \( \hat{h}(t_2) \) as in (11), where \( t = (t_1^T, t_2^T)^T \) is the solution of

\[
\sum_{j=1}^{n} \psi_j(Y_j(\theta_0), t) = 0.
\]
Accuracy of saddlepoint approximation

Table 1: Comparison of bootstrap Monte Carlo results with saddlepoint approximations with an overdispersed negative binomial generated model, analysed as a Poisson model.

<table>
<thead>
<tr>
<th>( \hat{u} )</th>
<th>.2</th>
<th>.3</th>
<th>.4</th>
<th>.5</th>
<th>.6</th>
<th>.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSMC</td>
<td>0.5150</td>
<td>0.2099</td>
<td>0.0567</td>
<td>0.0110</td>
<td>0.0024</td>
<td>0.0003</td>
</tr>
<tr>
<td>SPLR</td>
<td>0.5065</td>
<td>0.2121</td>
<td>0.0603</td>
<td>0.0119</td>
<td>0.0013</td>
<td>0.0001</td>
</tr>
<tr>
<td>SPBN</td>
<td>0.5043</td>
<td>0.2103</td>
<td>0.0596</td>
<td>0.0117</td>
<td>0.0013</td>
<td>0.0001</td>
</tr>
<tr>
<td>( \chi_2^2 )</td>
<td>0.4493</td>
<td>0.1653</td>
<td>0.0408</td>
<td>0.0067</td>
<td>0.0007</td>
<td>0.0001</td>
</tr>
</tbody>
</table>
Table 2: Comparison of bootstrap Monte Carlo results with saddlepoint approximation for linear regression under the correlation model.

<table>
<thead>
<tr>
<th>u</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSMC</td>
<td>0.9448</td>
<td>0.6937</td>
<td>0.3587</td>
<td>0.11267</td>
<td>0.0306</td>
<td>0.0050</td>
</tr>
<tr>
<td>SPLR</td>
<td>0.9490</td>
<td>0.7004</td>
<td>0.3643</td>
<td>0.1307</td>
<td>0.0323</td>
<td>0.0055</td>
</tr>
<tr>
<td>SPBN</td>
<td>0.9486</td>
<td>0.6982</td>
<td>0.3610</td>
<td>0.1284</td>
<td>0.0313</td>
<td>0.0052</td>
</tr>
<tr>
<td>$\chi^2_3$</td>
<td>0.8607</td>
<td>0.5488</td>
<td>0.2592</td>
<td>0.0907</td>
<td>0.0235</td>
<td>0.0045</td>
</tr>
</tbody>
</table>
Table 3: Comparison of bootstrap Monte Carlo results with saddlepoint approximation for linear regression under the regression model.

<table>
<thead>
<tr>
<th>u</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
<th>0.6</th>
<th>0.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSMC</td>
<td>0.8386</td>
<td>0.5216</td>
<td>0.2365</td>
<td>0.0806</td>
<td>0.0215</td>
<td>0.0031</td>
<td>0.0006</td>
</tr>
<tr>
<td>SPLR</td>
<td>0.8473</td>
<td>0.5200</td>
<td>0.2352</td>
<td>0.0793</td>
<td>0.0200</td>
<td>0.0038</td>
<td>0.0005</td>
</tr>
<tr>
<td>SPBN</td>
<td>0.8469</td>
<td>0.5194</td>
<td>0.2348</td>
<td>0.0791</td>
<td>0.0200</td>
<td>0.0038</td>
<td>0.0005</td>
</tr>
<tr>
<td>$\chi^2_3$</td>
<td>0.8607</td>
<td>0.5488</td>
<td>0.2592</td>
<td>0.0907</td>
<td>0.0235</td>
<td>0.0045</td>
<td>0.0006</td>
</tr>
</tbody>
</table>
Figure 1: **Accuracy of the bootstrap approximation**
Bootstrap test (solid) and parametric test (dashed)
## Power comparisons

Table 4: Power of bootstrap and standard tests for simulations under Poisson and negative binomial models.

<table>
<thead>
<tr>
<th>$(\theta_3, \theta_4)$</th>
<th>(0,0)</th>
<th>(.1,.1)</th>
<th>(.2,.2)</th>
<th>(.3,.3)</th>
<th>(.4,.4)</th>
<th>(.5,.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Poisson Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bootstrap</td>
<td>0.04</td>
<td>0.09</td>
<td>0.32</td>
<td>0.70</td>
<td>0.96</td>
<td>0.99</td>
</tr>
<tr>
<td>Likelihood ratio</td>
<td>0.05</td>
<td>0.09</td>
<td>0.33</td>
<td>0.77</td>
<td>0.98</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Negative Binomial Model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bootstrap</td>
<td>0.04</td>
<td>0.06</td>
<td>0.04</td>
<td>0.22</td>
<td>0.26</td>
<td>0.54</td>
</tr>
<tr>
<td>GLM Power</td>
<td>0.44</td>
<td>0.50</td>
<td>0.67</td>
<td>0.87</td>
<td>0.85</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Table 5: Power of bootstrap and standard tests for simulations under regression model with errors exponential random variables to power 1.5 for 5 replicates of a $3 \times 2$ with null hypothesis of no interaction.

<table>
<thead>
<tr>
<th>$(\theta_3, \theta_4)$</th>
<th>(0,0)</th>
<th>(.2,.2)</th>
<th>(.4,.4)</th>
<th>(.6,.6)</th>
<th>(.8,.8)</th>
<th>(1,1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bootstrap test</strong></td>
<td>0.052</td>
<td>0.176</td>
<td>0.464</td>
<td>0.734</td>
<td>0.874</td>
<td>0.942</td>
</tr>
<tr>
<td><strong>F test</strong></td>
<td>0.040</td>
<td>0.152</td>
<td>0.428</td>
<td>0.694</td>
<td>0.846</td>
<td>0.932</td>
</tr>
</tbody>
</table>
A toy parametric example. Let $Y_{1j}, Y_{2j}$ be independent exponentials with means $\theta_1 + \theta_2, \theta_2$ for $j = 1, \ldots, n$ and consider a test of $H_0 : \theta_2 = \theta_{20}$. Here $F = (1 - \exp(y_1/(\theta_1 + \theta_{20}))(1 - \exp(y_2/\theta_{20})).$ The estimating equations are

$$\sum_j (Y_{1j} - t_1 - t_2) = 0 \text{ and } \sum_j^n (Y_{2j} - t_2) = 0.$$ 

A simple integration gives

$$K(\tau, t) = -\tau_1(t_1 + t_2) - \tau_2t_2 - \log(1 - \tau_1(\theta_1 + \theta_{20}))) - \log(1 - \tau_2\theta_{20}).$$

Equating the derivative of this wrt $\tau$, to 0 gives

$$\tau_1(t) = (\theta_1 + \theta_2)^{-1} - (t_1 + t_2)^{-1} \text{ and } \tau_2(t) = \theta_2^{-1} - t_2^{-1}. $$
So

\[ K(\tau(t), t) = 2 - \frac{t_1 + t_2}{\theta_1 + \theta_{20}} - \frac{t_2}{\theta_{20}} + \log \frac{t_1 + t_2}{\theta_1 + \theta_{20}} + \log \frac{t_2}{\theta_{20}}. \]

Now equating the derivative of this with respect to \( t_1 \) gives

\[ t_1(t_2) + t_2 = \theta_1 + \theta_{20} \]

and so

\[ h(t_2) = -K(\tau(t(t_2)), t(t_2)) = -1 + \frac{t_2}{\theta_{20}} - \log \frac{t_2}{\theta_{20}}. \]

Note that \( h(\theta_{20}) = 0, \ h'(\theta_{20}) = 0 \) and \( h''(\theta_{20}) = 1/\theta_{20}^2. \)