Journal of Multivariate Analysis 158 (2017) 103–116

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Nonparametric tests for multi-parameter M-estimators

John E. Kolassa [a,*], John Robinson [b]

[a] Department of Statistics and Biostatistics, Rutgers University, 110 Frelinghuysen Road, Piscataway, NJ 08854, USA

[b] School of Mathematics and Statistics, University of Sydney, Carslaw Building F07, Eastern Avenue, Camperdown NSW 2006, Australia

ARTICLE INFO

Article history:
Received 26 November 2016
Available online 25 April 2017

AMS subject classifications:
62G09
62G10
62G20

Keywords:
Empirical saddlepoint
Tilted bootstrap
Regression
Non-linear regression
Generalized linear models

ABSTRACT

We consider likelihood ratio like test statistics based on M-estimators for multi-parameter hypotheses for some commonly used parametric models where the assumptions on which the standard test statistics are based are not justified. The nonparametric test statistics are based on empirical exponential families and permit us to give bootstrap methods for the tests. We further consider saddlepoint approximations to the tail probabilities used in these tests. This generalizes earlier work of Robinson et al. (2003) in two ways. First, we generalize from bootstraps based on resampling vectors of both response and explanatory variables to include bootstrapping residuals for fixed explanatory variables, resulting in a surprising result for the weighted resampling. Second, we obtain a theorem for tail probabilities under weak conditions providing essential justification for the approximation to bootstrap results for both cases. We use as examples linear regression, non-linear regression and generalized linear models under models with independent and identically distributed residuals or vectors of observations, giving numerical illustrations of the results.

1. Introduction

Let $Y_1(\theta), \ldots, Y_n(\theta)$ be a sample of independent and identically distributed random vectors, with $Y_j(\theta)$ from a distribution F on the sample space $\mathcal{Y}$. Suppose that θ satisfies

$$\mathrm{E}\Big[\sum_{j=1}^{n} \psi_j\{Y_j(\theta), \theta\}\Big] = 0 \tag{1}$$

and consider test statistics based on T, the M-estimate of θ, defined by the solution of

$$\sum_{j=1}^{n} \psi_j\{Y_j(\theta), T\} = 0, \tag{2}$$

where ψ_j are assumed to be smooth functions from $\mathcal{Y} \times \mathbb{R}^p$ to $\mathbb{R}^p$. The functions ψ_j are often chosen to make an analysis more robust.

We have, in particular, two cases in mind, where, for example, in linear regression with response variables Z_j and explanatory variables X_j, $Y_j(\theta) = (Z_j, X_j^\top)^\top$ and

$$\psi_j\{Y_j(\theta), t\} = X_j(Z_j - t^\top X_j),$$

* Corresponding author.
E-mail addresses: kolassa@stat.rutgers.edu (J.E. Kolassa), john.robinson@sydney.edu.au (J. Robinson).

http://dx.doi.org/10.1016/j.jmva.2017.04.004
0047-259X/

or $Y_j(\theta) = Z_j - \theta^\top x_j$ and

$$\psi_j\{Y_j(\theta), t\} = x_j\{Y_j(\theta) + (\theta - t)^\top x_j\} = x_j(Z_j - t^\top x_j),$$

for fixed $X_j = x_j$. Note that it is $Y_j(\theta)$ that are identically distributed allowing resampling.

Let $\theta = (\theta_1^\top, \theta_2^\top)^\top$, where $\theta_1 \in \mathbb{R}^{p_1}$ and $\theta_2 \in \mathbb{R}^{p_2}$, $p_1 + p_2 = p$, and suppose we wish to test the null hypothesis

$$\mathcal{H}_0 : \theta_2 = \theta_{20}.$$

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 standard likelihood theory for estimation and inference is available. However, when the sample size is moderate to small or when the model is incorrectly specified, the p-values obtained from the asymptotic theory can be very inaccurate. [10] proposed a new likelihood like statistic based on an empirical exponentially tilted distribution considering only the case $\psi_j = \psi$. Assuming that the density of $\sum_{j=1}^n \psi\{Y_j(\theta), \theta\}$ exists, they gave a saddlepoint approximation with *relative* error of order $O(n^{-1})$. This method can only be used when F is known. Further, they considered a formal approach to empirical likelihood ratio tests using bootstrap tilting. The saddlepoint approximation to the distribution of the bootstrap statistic requires a proof of the result without the restrictive condition that a density exists. This proof, given in Section 6, is of an entirely different character from that of [10].

The two purposes of this paper are to justify the formal approach for saddlepoint approximations of [10] for empirical likelihood tests and to consider score functions ψ_j which change with each observation. We note that [4] obtained tests in the case of one-dimensional parameters for identically distributed score functions but their methods could not be extended to the case of multi-dimensional parameters. In Section 2, a test statistic related to that from exponential families is derived from the cumulant generating function of the left hand side of the estimating Eq. (2) when the distribution of $Y_j(\theta)$ is known under the null hypothesis. If the distribution is not known, a tilted empirical distribution satisfying the null hypothesis is obtained as an approximation and its cumulant generating function is used to obtain a natural test statistic. We use weighted bootstrap sampling from this tilted empirical distribution to obtain p-values for the test. The theorem of Section 3 gives a saddlepoint approximation of this bootstrap p-value and could be used instead of resampling. Bootstrap sampling requires a double optimization for each bootstrap replicate and so is extremely computationally intensive, so the saddlepoint approximation may be useful as an alternative. Note that the nonparametric approach depends only on $\psi_j\{Y_j(\theta), t\}$ for all $j \in \{1, \ldots, n\}$. These functions may have been derived from some parametric model, but this model is not used except to give these estimating functions. In Section 4 we provide applications to three special cases, linear regression, robust non linear regression and robust generalized linear models. In Section 5 we give numerical results to illustrate the accuracy of the approximations for some important cases of tests and compare the power of the tests to the power of the standard tests in two cases.

2. A nonparametric test

First consider the simpler case in which the distribution F of $Y_j(\theta)$ is known. Denote the cumulant generating function of $\sum_{j=1}^n \psi_j\{Y_j(\theta), t\}$ by

$$nK(\tau, t) = \sum_{j=1}^n K_j(\tau, t) = \sum_{j=1}^n \ln\big\{\mathrm{E}\big(\exp[\tau^\top \psi_j\{Y_j(\theta), t\}]\big)\big\}. \tag{3}$$

Let $T = (T_1^\top, T_2^\top)^\top$ be the M-estimator, the solution to

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

Consider a test statistic based on the function h defined by

$$h(t_2) = \inf_{t_1} \sup_{\tau}\{-K(\tau, t)\} = -K[\tau\{t(t_2)\}, t(t_2)], \tag{4}$$

where $t(t_2) = (t_1(t_2)^\top, t_2^\top)^\top$ for

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

Note that $h(\theta_{20}) = 0$. So a test can be based on $h(T_2)$. This is the statistic considered in [10]. In Section 3.2 of [7] it is shown that, in the case of generalized linear models with the classical score statistic when $t = t_2$, the test based on $h(t_2)$ reduces to the likelihood ratio statistic.

In practice, the distribution underlying the data sample $Y_1(\theta), \ldots, Y_n(\theta)$ is often unknown, and hence K is unknown, and a nonparametric approach is required. An empirical exponential likelihood, equivalent to a tilted bootstrap, provides empirical versions of the test of $\mathcal{H}_0 : \theta_2 = \theta_{20}$. We consider weighted empirical distributions

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

where the weights are chosen to minimize the backward Kullback–Leibler distance, $\sum_{k=1}^n w_k \ln(nw_k)$, between the empirical distribution and the weighted empirical distribution subject to

$$\sum_{k=1}^n w_k \frac{1}{n} \sum_{j=1}^n \psi_j\{Y_k(\theta), \theta\} = 0 \quad \text{and} \quad \sum_{k=1}^n w_k = 1, \tag{5}$$

where $\theta = (\theta_1^\top, \theta_{20}^\top)^\top$, as in [3]. So we find stationary values of

$$\sum_{k=1}^n w_k \ln(nw_k) - \beta^\top \sum_{k=1}^n w_k \frac{1}{n} \sum_{j=1}^n \psi_j\{Y_k(\theta), \theta\} + \gamma \left(\sum_{k=1}^n w_k - 1 \right) \tag{6}$$

with respect to w_k, β, γ and θ_1, with $\theta_2 = \theta_{20}$. Differentiating with respect to each w_k, together with the constraints (5), leads to

$$w_k = \frac{1}{n} \exp\Big[\beta^\top \frac{1}{n} \sum_{j=1}^n \psi_j\{Y_k(\theta), \theta\} - \kappa(\beta, \theta)\Big], \tag{7}$$

where

$$\kappa(\beta, \theta) = \ln\Big(\frac{1}{n} \sum_{k=1}^n \exp\Big[\beta^\top \frac{1}{n} \sum_{j=1}^n \psi_j\{Y_k(\theta), \theta\}\Big]\Big). \tag{8}$$

Then (6) reduces to

$$\sum_k w_k \ln(nw_k) = -\kappa(\beta, \theta). \tag{9}$$

So the minimum of (6) under the constraints (5), is $-\kappa[\beta\{\theta(\theta_{20})\}, \theta(\theta_{20})]$, where $\theta(\theta_{20}) = (\theta_1(\theta_{20})^\top, \theta_{20}^\top)^\top$,

$$\beta(\theta) = \arg\sup_\beta -\kappa(\beta, \theta) \tag{10}$$

and

$$\theta_1(\theta_{20}) = \arg\inf_{\theta_1} -\kappa\{\beta(\theta), \theta\}. \tag{11}$$

To obtain a test statistic for $\mathcal{H}_0 : \theta_2 = \theta_{20}$, take the empirical exponential family,

$$\hat{F}(y) = \sum_{k=1}^n w_k(\theta_{20}) \mathbf{1}[Y_k\{\theta(\theta_{20})\} \le y], \tag{12}$$

where, from (7), using (10) and (11),

$$w_k(\theta_{20}) = \exp\Big(\beta\{\theta(\theta_{20})\}^\top \frac{1}{n} \sum_{j=1}^n \psi_j[Y_k\{\theta(\theta_{20}), \theta(\theta_{20})\}] - \kappa[\beta\{\theta(\theta_{20})\}, \theta(\theta_{20})]\Big). \tag{13}$$

Note that when $\psi_j\{Y_j(\theta), \theta\}$ are identically distributed, the results in (6)–(13) take the same form as in [10].

Consider tilted bootstrap sampling by drawing Y_j^* from $\hat{F}$. For any value $t = (t_1^\top, t_2^\top)^\top$, denote the cumulant generating function of $\sum_{j=1}^n \psi_j(Y_j^*, t)$ by

$$n\hat{K}(\tau; t) = \sum_{j=1}^n \ln\Big(\sum_{k=1}^n w_k(\theta_{20}) \exp\big(\tau^\top \psi_j[Y_k\{\theta(\theta_{20})\}; t]\big)\Big). \tag{14}$$

Then, as in (4), the test statistic is based on the function $\hat{h}(.)$ defined by

$$\hat{h}(t_2) = \inf_{t_1} \sup_\tau \{-\hat{K}(\tau; t)\} = -\hat{K}[\tau\{t(t_2)\}; t(t_2)], \tag{15}$$

where $t(t_2) = (t_1(t_2)^\top, t_2^\top)^\top$,

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

Note again that $\hat{h}(\theta_{20}) = 0$. Then the test can be based on $\hat{h}(T_2)$, where $T = (T_1^\top, T_2^\top)^\top$ is the solution of (2). A bootstrap version of the test uses the distribution of $\hat{h}(T_2^*)$, for $T^* = (T_1^{*\top}, T_2^{*\top})^\top$, the solution of

$$\sum_{j=1}^{n} \psi_j(Y_j^*, t) = 0.$$

Note that $\hat{h}(T_2^*)$ needs to be calculated from (15) for each T_2^*. Now $\hat{h}(T_2^*)$ has distribution

$$\hat{F}_h(x) = \Pr\{\hat{h}(T_2^*) \le x\},$$

which can be estimated by a Monte Carlo simulation. A series of B bootstrap samples are drawn from $\hat{F}$. If T_n^{*b} denotes the M-estimator for the bth such sample, $b \in \{1, \ldots, B\}$, then the Monte Carlo approximation of the p-value of the test is

$$\frac{1 + \sum_{b=1}^{B} \mathbf{1}\{\hat{h}(T_2^{*b}) \ge \hat{h}(T_2)\}}{B+1}.$$

3. A saddlepoint approximation

We will derive a saddlepoint approximation for the distribution $h(T_2)$ which can be applied directly to give an approximation to the distribution of $\hat{h}(T_2^*)$.

Suppose that for any t, $L_{jt} = \psi_j\{Y_j(\theta), t\}$ has a derivative

$$M_{jt} = \partial\psi_j\{Y_j(\theta), t\}/\partial t,$$

with probability 1. We write

$$\bar{L}_t = n^{-1}\sum_{j=1}^{n} \psi_j\{Y_j(\theta), t\},$$

$$\bar{M}_t = n^{-1}\sum_{j=1}^{n} \partial\psi_j\{Y_j(\theta), t\}/\partial t,$$

$$\bar{Q}_t = n^{-1}\sum_{j=1}^{n} (\psi_j\{Y_j(\theta), t\} - \mathrm{E}[\psi_j\{Y_j(\theta), t\}])(\psi_j(Y_j(\theta), t) - \mathrm{E}[\psi_j\{Y_j(\theta), t\}])^\top.$$

Define

$$\hat{L}_t = \bar{M}_t^{-1}\bar{L}_t,$$

whenever $|\bar{M}_t| \ne 0$, where $|\cdot|$ denotes the determinant. Then a first step in Newton's solution to $\bar{L}_T = 0$, starting from t, is $t - \hat{L}_t$.

Consider the following assumptions for some θ a solution of (1):

(A1) $|\mathrm{E}(\bar{M}_\theta)| \ne 0$, and for some $\gamma > 0$,
$\mathrm{E}(\bar{M}_t)$ is continuous at all $t \in B_\gamma^p(\theta)$, a cube with side length 2γ of dimension p centered at θ.

(A2) The elements of $\psi_j\{Y_1(\theta), t\}$ and its first two derivatives with respect to t exist and are continuous with probability 1 and for some $\upsilon > 0$, $K(\tau, t)$ is finite for $\tau \in B_\upsilon^p(0)$ for all $t \in B_\gamma^p(\theta)$.

(A3) $0 < c < |\Sigma_\tau|^{1/2} < C$ and if $\varphi_\tau(\xi) = \mathrm{E}\{\exp(i\xi^\top U^\tau)\}$, then $|\varphi_\tau(\xi)| < 1 - \rho$, for $\rho > 0$ and for all $0 < c < |\xi| < Cn^{d/2}$, where $d = p + q$.

(A4) $\sum_{j=1}^{n} \exp(\zeta^\top V_{j\theta}) < \infty$ for $|\zeta| < c$ for some $c > 0$.

Under (A1), θ_0, the solution of (1), is the unique solution in $B_\gamma^p(\theta_0)$. Conceptually, to consider the distribution of $\hat{L}_\theta$, we need to consider the joint distribution of $\bar{L}_\theta$ and $\bar{M}_\theta$, transform to the joint distribution of $\hat{L}_\theta$ and $\bar{M}_\theta$, and then obtain the marginal distribution of $\hat{L}_\theta$. However, in some cases the joint distribution of $\bar{M}_\theta$ and $\bar{L}_\theta$ is supported on a manifold of dimension smaller than the space in which it is embedded. To this end, we construct vectors $V_{j\theta}$ such that $U_{j\theta} = (L_{j\theta}, V_{j\theta})$ are independent random vectors with positive definite covariance matrix and all elements of $\bar{M}_{j\theta}$ are smooth functions of $(L_{j\theta}, V_{j\theta})$. Let $\bar{V}_\theta = \sum_{j=1}^{n} V_{j\theta}/n$ and let the dimensions of the components of $U_{j\theta}$ be p and q, respectively. Let $Z_\theta = (\hat{L}_\theta, \bar{V}_\theta) = g(\bar{L}_\theta, \bar{V}_\theta)$. Let F_{U_j} be the distribution of $U_{j\theta}$ under F and define the tilted variable U_j^τ to have distribution function

$$F_j^\tau(\ell, \upsilon) = \int_{(\ell', \upsilon') \le (\ell, \upsilon)} e^{\tau^\top \ell' - K_j(\tau, \theta)} dF_{U_j}(\ell', \upsilon').$$

Let $\mu_\tau = \sum_{j=1}^n \mathrm{E}(U_j^\tau/n)$, where, from (1), μ_0 has 0 in the first p components, and let $\Sigma_\tau = \sum_{j=1}^n \mathrm{cov}(U_j^\tau/n)$. The smoothness condition, (A3), is used in order to apply an Edgeworth expansion for $\bar{U}_{\tau(\theta)}$ and (A4) permits us to show that $|\bar{M}_\theta|$ exceeds a finite bound with exponentially small probability. The proof of the following theorem is given in Section 6.

Theorem 1. *Under assumptions* (A1)–(A4),

$$\Pr\{2h(T_2) \geq u^2\} = \bar{Q}_{p_2}(nu^2)\{1 + O(n^{-1})\} + n^{-1}c_n u^{p_2} e^{-nu^2/2}\left\{\frac{G(u)-1}{u^2}\right\} \tag{16}$$

and

$$\Pr\{2h(T_2) \geq u^2\} = \bar{Q}_{p_2}(n\hat{u}^2)\{1 + O(n^{-1})\} \tag{17}$$

where $Q_{p_2} = 1 - \bar{Q}_{p_2}$ is the distribution function of a chi-squared variate with p_2 degrees of freedom,

$$c_n = \frac{n^{p_2/2}}{2^{p_2/2-1}\Gamma(p_2/2)}, \qquad \hat{u} = u - \ln\{G(u)/nu\},$$

and

$$G(u) = \int_{S_{p_2}} \delta(u, s)ds = 1 + u^2 k(u), \tag{18}$$

for

$$\delta(u, s) = \frac{\Gamma(p_2/2)J\{t(t_2)\}|h''(\theta_{20})|^{1/2}J_1(t_2)J_2(t_2)}{2\pi^{p_2/2}u^{p_2-1}|\Sigma_{L\tau\{t(t_2)\}}|^{1/2}|H_{11}\{t(t_2)\}|^{1/2}}, \tag{19}$$

where $\Sigma_{L\tau(t)}$ is the sub-matrix of $\Sigma_{\tau(t)}$ from the first p rows and columns, $[h''(\theta_{20})]^{-1} = [\{\mathrm{E}(\bar{M}_{\theta(\theta_{20})})\}^{-1}\mathrm{E}(\bar{Q}_{\theta(\theta_{20})})\{\mathrm{E}(\bar{M}_{\theta(\theta_{20})})\}^{-1}]_{22} = \mathrm{var}(T_2)$, (r, s) are the polar coordinates corresponding to $\{h''(\theta_{20})\}^{1/2}(t_2 - \theta_{20})$, r is the radial component and $s \in S_{p_2}$, the p_2-dimensional sphere of unit radius, $H_{11}(t) = d^2K\{\tau(t), t\}/dt_1^2$,

$$J\{t(t_2)\} = |\mathrm{E}_{\tau\{t(t_2)\}}\bar{M}_{t(t_2)}| = [\partial^2 K\{\tau(t), t\}/\partial\tau\partial t]_{t=t(t_2)},$$

$J_1(t_2) = r^{p_2-1}$ and $J_2(t_2) = ru/[h'(t_2)^\top\{h''(\theta_{20})\}^{1/2}(t_2 - \theta_{20})]$, $u = \sqrt{2h(t_2)}$, $k(u)$ is bounded and the order terms are uniform for $u < \epsilon$ for some $\epsilon > 0$.

4. Application to special cases

In this section we apply the Monte Carlo and saddlepoint approximations for the bootstrap methods of Sections 2 and 3 to the special cases of linear regression, nonlinear regression and generalized linear models. These models are considered with fixed and random explanatory variables with corresponding regression and correlation bootstrap methods.

4.1. Linear regression

Consider the linear regression model

$$Z_j = \theta^\top X_j + R_j$$

for $j \in \{1, \ldots, n\}$. Under a correlation model we assume that $Y_j = (Z_j, X_j^\top)^\top$ are independent identically distributed random variables from F. For the regression model we assume that $X_j = x_j$ are fixed and take $Y_j(\theta)$ as R_j, assumed to be independent identically distributed random variables from a distribution F.

Using a score function

$$\psi_j\{Y_j(\theta), t\} = X_j\phi(Z_j - t^\top X_j), \tag{20}$$

in the correlation case, and

$$\psi_j\{Y_j(\theta), t\} = x_j\phi\{Z_j - \theta^\top x_j + (\theta - t)^\top x_j\}, \tag{21}$$

for the regression case, where $\phi(x) = x,\ -b \leq x \leq b$, and $\phi(x)$ is constant for $x \leq -b$ and $x \geq b$. A bootstrap approach is needed for both the correlation and regression models. The correlation model takes weighted bootstrap samples from $(Z_1, X_1^\top)^\top, \ldots, (Z_n, X_n^\top)^\top$ and the regression model draws weighted bootstrap samples from residuals $Z_1 - \theta^\top x_1, \ldots, Z_n - \theta^\top x_n$.

4.1.1. Correlation model

We sample $(Z_1^*, X_1^{*\top}), \ldots, (Z_n^*, X_n^{*\top})$ from the tilted empirical joint distribution

$$\hat{F}_{\theta_{20}}(y, x) = \sum_{k=1}^{n} w_k(\theta_{20})\mathbf{1}\{(z_k, x_k^\top)^\top \le (y, x^\top)^\top\},$$

where $w_k(\theta_{20})$ is given in (13) with $\kappa(\beta, \theta)$ from (8) and $\theta(\theta_{20})$ defined in (10) and (11). Using (20), the cumulative generating function of the independent identically distributed random variables $\psi_j\{(Z_j^*, X_j^{*\top})^\top, t\} = X_j^*\phi(Z_j^* - t^\top X_j^*)$ is

$$\hat{K}(\tau, t) = \ln \sum_{k=1}^{n} w_k(\theta_{20}) e^{\tau^\top X_k \phi(Z_k - t^\top X_k)}. \tag{22}$$

Note that in this case ψ_j does not depend on j. Solve $\sum_{j=1}^{n} X_j^*\phi(Z_j^* - t^\top X_j^*) = 0$ to get the bootstrap estimate $t^* = (t_1^{*\top}, t_2^{*\top})^\top$ and obtain $\hat{h}(t_2^*)$, calculating this from (15) for each t^* and solve $\sum_{j=1}^{n} X_j\phi(Z_j - \theta^\top X_j)$ to get $t = (t_1^\top, t_2^\top)^\top$ and so the test statistic $\hat{h}(t_2)$. Then a Monte Carlo approximation to the bootstrap approximation to the p-value can be obtained as at the end of Section 2. A saddlepoint approximation to the bootstrap approximation can be obtained from (16) or (17) of the theorem taking $\hat{K}(\tau, t)$ from (22).

4.1.2. Regression model

For the regression model, we will, without loss of generality, consider the case where the first element of the p elements of x_k is 1 and $\bar{x} = (1, 0, \ldots, 0)^\top$. The first constraint of (5) is then

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

leading to

$$w_k = \frac{1}{n} e^{\beta_1 \phi(Z_k - \theta^\top x_k) - \kappa(\beta_1, \theta)},$$

from (7) and (8), where

$$\kappa(\beta_1, \theta) = \ln\left\{\frac{1}{n}\sum_{k=1}^{n} e^{\beta_1 \bar{\phi}(Z_k - \theta^\top x_k)}\right\}$$

and where, if $\beta = (\beta_1, \beta_2^\top)^\top$, the $p - 1$ vector β_2 is not estimable. Now we may take $\beta(\theta) = (\beta_1(\theta), 0^\top)^\top$ from (10) as the solution to

$$\frac{\partial \kappa(\beta_1, \theta)}{\partial \beta_1} = \sum_{k=1}^{n} \phi(Z_k - \theta^\top x_k) e^{\beta_1 \phi(Z_k - \theta^\top x_k)} = 0. \tag{23}$$

Also, $\theta_1(\theta_0)$ from (11), is the solution to $d\kappa\{\beta_1(\theta), \theta\}/d\theta_1 = 0$, so

$$\frac{d\beta_1(\theta)}{d\theta_1}\frac{1}{n}\sum_{k=1}^{n} \phi(Z_k - \theta^\top x_k) e^{\beta_1(\theta)\phi(Z_k - \theta^\top x_k)} - \beta_1(\theta)\frac{1}{n}\sum_{k=1}^{n} x_k^{(1)}\mathbf{1}(|Z_k - \theta^\top x_k| < b) e^{\beta_1(\theta)\phi(Z_k - \theta^\top x_k)} = 0,$$

where $x_k^{(1)}$ contains the first p_1 elements of x_k. The first term here is zero from (23), so $\beta_1\{\theta(\theta_0)\} = 0$ and it follows that $w_k = 1/n$. We require that $\theta(\theta_0)$ satisfies $\sum_{k=1}^{n} \phi\{Z_k - \theta(\theta_0)^\top x_k\} = 0$. This does not uniquely define $\theta(\theta_0)$, but we propose using the vector $(\theta_1(\theta_{20})^\top, \theta_{20}^\top)^\top$, where $\theta_1(\theta_{20})$ is the solution to

$$\sum_{k=1}^{n} x_k^{(1)} \phi(Z_k - \theta^\top x_k) = 0. \tag{24}$$

Bootstrap replicates $R_1^*, \ldots, R_n^*$ are obtained by sampling from the empirical distribution given by (12), which here can be written

$$\hat{F}_{\theta_{20}}(x) = \sum_{k=1}^{n} \frac{1}{n}\mathbf{1}\{Z_k - \theta(\theta_{20})^\top x_k \le x\}. \tag{25}$$

Then

$$\psi_j(R_j^*, t) = x_j \phi[R_j^* + \{\theta(\theta_{20}) - t\}^\top x_j].$$

The cumulant generating function of $\sum_{j=1}^n \psi_j(R_j^*, t)$ is

$$\hat{K}(\tau, t) = \sum_{j=1}^n \ln\left(\frac{1}{n}\sum_{k=1}^n e^{\tau^\top x_j \phi[Z_k - \theta(\theta_{20})^\top x_k + \{\theta(\theta_{20})-t\}^\top x_j]}\right)$$

and we can calculate the statistic $\hat{h}(t_2)$ and the bootstrap replicates $\hat{h}(t_2^*)$ as in (15), where $t = (t_1^\top, t_2^\top)^\top$ is the solution of $\sum_{j=1}^n \psi_j[Y_j\{\theta(\theta_{20}), t\}] = 0$ and $t^* = (t_1^{*\top}, t_2^{*\top})^\top$ is the solution of $\sum_{j=1}^n x_j \phi[R_j^* - \{\theta(\theta_{20}) - t\}^\top x_j] = 0$. Then Monte Carlo approximations to the p-value can be obtained as at the end of Section 2 or saddlepoint approximations can be obtained from (16) or (17).

4.2. Nonlinear regression

Consider the non-linear regression model

$$Z_k = g(\theta, X_k) + R_k(\theta)$$

for $k \in \{1, \ldots, n\}$, where $g(\theta, X_k)$ is a known function. We assume that either $R_k(\theta)$ are independent identically distributed random variables from a distribution F for the regression case, or that $(Z_k, X_k^\top)^\top$ are independent identically distributed random vectors. Using a score function

$$\psi_j\{Y_j(\theta), \theta\} = \frac{\partial g(\theta, X_j)}{\partial \theta}\,\phi\{R_j(\theta)\},$$

where the parameter $\theta = \theta(F)$ is the solution of

$$\mathrm{E}\Big[\sum_{j=1}^n \psi_j\{Y_j(\theta), \theta\}\Big] = 0,$$

we have the estimate T of θ, as the solution of

$$\sum_{j=1}^n \psi_i\{Y_j(\theta), t\} = \sum_{j=1}^n \frac{\partial g(t, X_j)}{\partial t}\,\phi\{R_j(t)\} = 0.$$

Now that $\psi_j\{Y_j(\theta), \theta\}$ is defined the regression and correlation cases can be obtained in exactly the same way as in Section 4.1.

4.3. Generalized linear models

Generalized linear models allow us to model the relationship between the predictors and a function of the mean of the response for continuous and discrete response variables. The response variables $Z_1, \ldots, Z_n$ are supposed to come from a distribution belonging to the exponential family, such that $\mathrm{E}(Z_j) = \mu_j$ and $\mathrm{var}(Z_j) = V(\mu_j)$ for all $j \in \{1, \ldots, n\}$ and, for all $i \in \{1, \ldots, n\}$,

$$\eta_j = g(\mu_j) = X_j^\top \theta, \tag{26}$$

where $\theta \in \mathbb{R}^p$ is the vector of parameters, $X_i \in \mathbb{R}^p$, and g is the link function.

We generalize a class of robust estimators of θ of [2] using (2) so that we can consider both fixed and random explanatory variables. Take

$$\psi_j\{Y_j(\theta), \theta\} = \phi(R_j)\frac{\partial \mu_j}{\partial \theta}\{V(\mu_j)\}^{-1/2} - a(\theta), \tag{27}$$

where

$$a(\theta) = \frac{1}{n}\sum_{j=1}^n \mathrm{E}\{\phi(R_j)\}\frac{\partial \mu_j}{\partial \theta}\{V(\mu_j)\}^{-1/2},$$

$R_j = (Z_j - \mu_j)/V^{1/2}(\mu_j)$ and ϕ is defined in Section 4.1. Details of the calculation for $a(\theta)$ are given in [2] for the Poisson and binomial models and tests for parametric cases are developed in [7]. We will only consider here the case $\phi(x) = x$, so $a(\theta) = 0$.

As in the previous cases, we can consider fixed $X_j = x_j$ and $Y_j(\theta) = R_j$ or random X_j and $Y_j(\theta) = (Z_j, X_j^\top)^\top$. Then the $Y_j(\theta)$ are considered as independent and identically distributed variables (although in the regression case they are independent with common variance but not identically distributed) and we can proceed as in the two previous special cases to get bootstrap approximations via either Monte Carlo sampling or saddlepoint approximation, by using the results of Sections 2 and 3 for the score functions given here. If the explanatory variables are constrained to a lattice, condition (A3) is violated and the discreteness may cause some inaccuracies.

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

$\hat{u}$	.2	.3	.4	.5	.6	.7
BSMC	0.5150	0.2099	0.0567	0.0110	0.0024	0.0003
SPLR	0.5065	0.2121	0.0603	0.0119	0.0013	0.0001
SPBN	0.5043	0.2103	0.0596	0.0117	0.0013	0.0001
χ^2_2	0.4493	0.1653	0.0408	0.0067	0.0007	0.0001

Table 2
Comparison of bootstrap Monte Carlo results with saddlepoint approximation for linear regression under the correlation model.

u	0.1	0.2	0.3	0.4	0.5	0.6
BSMC	0.9448	0.6937	0.3587	0.11267	0.0306	0.0050
SPLR	0.9490	0.7004	0.3643	0.1307	0.0323	0.0055
SPBN	0.9486	0.6982	0.3610	0.1284	0.0313	0.0052
χ^2_3	0.9402	0.6594	0.3080	0.0937	0.0186	0.0024

5. Numerical illustrations

We will consider these numerical examples in three subsections. First we demonstrate the accuracy of the saddlepoint approximation to the distribution of bootstrap tail probabilities. Second, we use tail probability approximations to check that the bootstrap is appropriate for obtaining p-values for tests, by demonstrating via simulation that, under the null hypothesis, the p-values obtained by the bootstrap method are uniformly distributed. Third, we demonstrate that using this statistic with the bootstrap approximation results in an increase in the power of the test for models deviating from the standard when compared with standard tests, while retaining the power of standard tests when the standard model holds. The code used is available at http://www.maths.usyd.edu.au/u/johnr/SPBS.

5.1. Saddlepoint approximation to the bootstrap

We indicated at the end of Section 2 the calculation to obtain the Monte Carlo bootstrap approximation. To obtain the saddlepoint approximations we calculate directly from Theorem 1.

We note here that the method for a Monte Carlo approximation to the integral in (18) is given in a remark in Section 2 of [6] and a further method can be based on approximate numerical integration over the sphere given by [5].

We illustrate the numerical accuracy of the saddlepoint approximation to the tail areas in three models. First, for Table 1, consider a Poisson regression model, $Y_i \sim \mathcal{P}(\mu_i)$, where for each $i \in \{1, \ldots, n\}$, $\ln(\mu_i) = x_i^\top\theta$, where $x_i = (1, x_{i2}, \ldots, x_{i5})^\top$ and $\theta = (2, 1, 0, 0, 0)^\top$, under the correlation model. The $(x_{i2}, \ldots, x_{i5})$ are generated from a uniform distribution on $(0, 1)$ for each sample and then Z_i are obtained as negative binomial variables with expectation μ_i and size parameter 5, for $i \in \{1, \ldots, 40\}$, giving data from an overdispersed distribution. We consider the estimator for the parameter θ defined by the estimating Eq. (2) using the scores (27) with $\mu_i = \exp(X_i^\top\theta)$, and $V(\mu_i) = \mu_i$ and consider the test of $\mathcal{H}_0 : (\theta_4, \theta_5) = (0, 0)$. Second, for Table 2, consider a linear regression with $x_i = (1, x_{i2}, \ldots, x_{i5})^\top$ for $i \in \{1, \ldots, 40\}$ and $\theta = (3, 2, 0, 0, 0)^\top$, with exponential errors under the correlation model, with the predictor variables again uniform and test the hypothesis $\mathcal{H}_0 : (\theta_3, \theta_4, \theta_5) = (0, 0, 0)$. Third, for Table 3, consider linear regression under a regression model for a 3×2 factorial with 5 replicates, so the explanatory variables are fixed and bootstrapping pairs would not be appropriate, with $\theta = (4, 2, 1, 1, 0, 0)^\top$ and exponential errors, where we test the null hypothesis of no interaction or $\mathcal{H}_0 : (\theta_5, \theta_6) = (0, 0)$.

In each of the tables we use 10,000 bootstrap samples to obtain Monte Carlo approximations to $\Pr\{\hat{h}(t_2^*) > u^2/2\}$ given by BSMC, the bootstrap Monte Carlo approximation, we give the first order chi-squared approximation and the Lugannani–Rice and Barndorff-Nielsen saddlepoint approximations SPLR, defined to be (16), and SPBN, defined to be (17), using the Genz approximation [5] to the integral (18). The tables demonstrate remarkable accuracy of the saddlepoint approximations compared to the first order chi-squared approximation, demonstrating that the second order correction of (16) or (17) is required to obtain appropriate accuracy in the tail. When the sample size is decreased to 20 for Tables 1 and 2 and replicates are decreased to 3 for Table 3, results still show a large improvement for the saddlepoint approximations over the simple χ^2_2 approximation, but, as might be expected for small sample sizes, relative errors are not as small, particularly for the more extreme tails.

5.2. Accuracy of the bootstrap approximation

The last subsection demonstrated that the saddlepoint approximation can be used in place of the Monte Carlo bootstrap approximation. So we can use the saddlepoint approximation to the bootstrap p-values for Monte Carlo samples drawn from a given model under the null hypothesis to check that these are uniformly distributed. First we consider plots of the ordered

Table 3
Comparison of bootstrap Monte Carlo results with saddlepoint approximation for linear regression under the regression model.

u	0.1	0.2	0.3	0.4	0.5	0.6	0.7
BSMC	0.8386	0.5216	0.2365	0.0806	0.0215	0.0031	0.0006
SPLR	0.8473	0.5200	0.2352	0.0793	0.0200	0.0038	0.0005
SPBN	0.8469	0.5194	0.2348	0.0791	0.0200	0.0038	0.0005
χ^2_2	0.8607	0.5488	0.2592	0.0907	0.0235	0.0045	0.0006

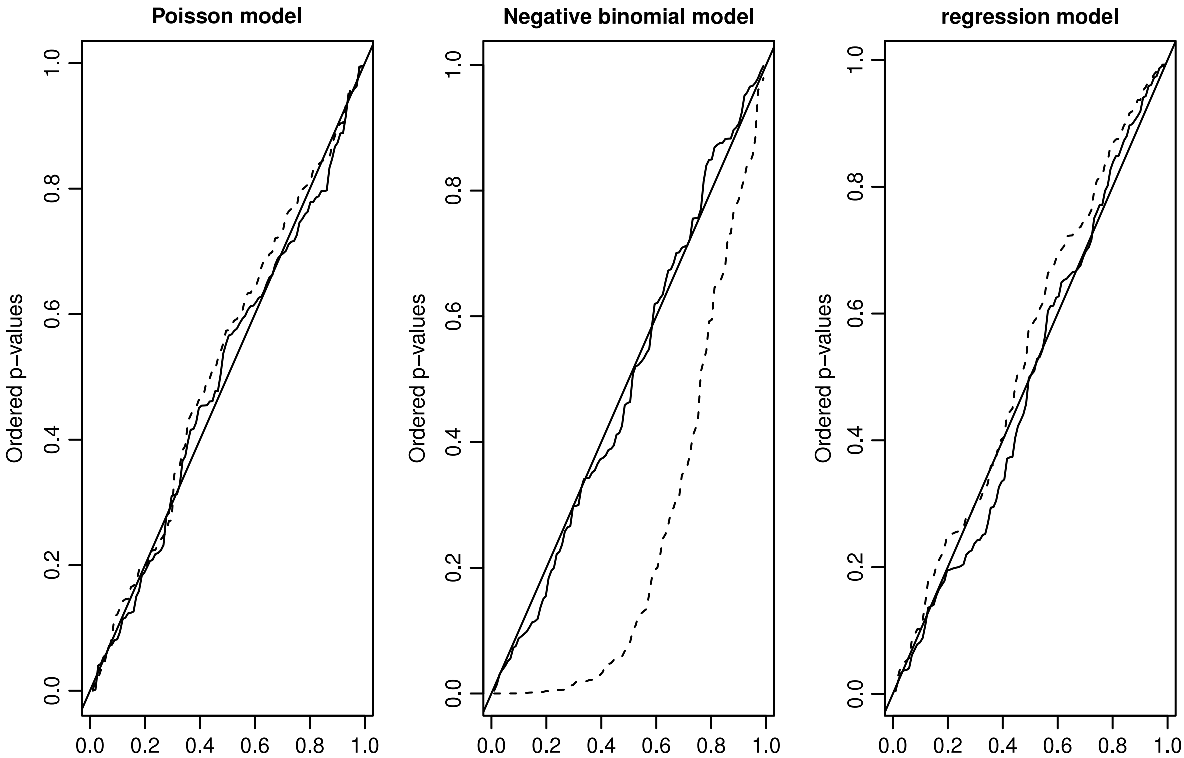

Fig. 1. The first panel plots the ordered p-values obtained from 100 Monte Carlo sets obtained under the null hypothesis $(\theta_4, \theta_5) = (0, 0)$ for test statistics under the Poisson model with mean $\exp(\theta^\top x_j), j \in \{1, \dots, 40\}$, with $\theta^\top = (2, 1, 0, 0, 0)$, from the bootstrap test (solid line) and the likelihood ratio test (dashed line). The second panel gives the same plots based on 100 simulations under the null hypothesis when the data is generated under a negative binomial model with the same mean and scale parameter 5. The third panel gives p-values for 100 Monte Carlo sets for a test of no interaction in 5 replicates of a 3×2 factorial design with exponential errors.

p-values from the nonparametric test proposed here and those for the standard test using the standard likelihood ratio test with an assumption of a Poisson model. The first panel of Fig. 1 shows that under the Poisson model both the bootstrap test and the likelihood ratio test give uniformly distributed p-values under the null hypothesis. The second panel has samples drawn from a negative binomial model with the same mean as for the first panel. Here the bootstrap p-values are uniformly distributed but the p-values from a standard analysis assuming a Poisson model are far from uniform, implying that major errors result from using this standard analysis when the errors are negative binomial. The third panel has plots of p-values from tests based on the statistic h and for the standard F statistic for 100 simulated samples for the model described in Section 5.1 to give Table 3. Both statistics have uniformly distributed p-values under the hypothesis of no interaction.

5.3. Power comparisons

First, in Table 4, we compare the power of the bootstrap test compared with the likelihood ratio test based on the assumption of a Poisson model for data generated under a Poisson model with mean $\exp(\theta^\top x_j), j \in \{1, \dots, 40\}$ and under a negative binomial model with the same mean and scale parameter 5, where $\theta^\top = (2, 1, 0, \theta_4, \theta_5)$. The upper table illustrates that there is no loss of power in using the bootstrap method compared to the likelihood ratio test when the Poisson model is true. The lower table demonstrates that, in the case where the errors are negative binomial, the likelihood ratio test gives major errors but that the bootstrap method is appropriate and retains a reduced power compared with the power under the Poisson model.

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

(θ_4, θ_5)	(0, 0)	(.1,.1)	(.2,.2)	(.3,.3)	(.4,.4)	(.5,.5)
Poisson model						
Bootstrap	0.04	0.09	0.32	0.70	0.96	0.99
Likelihood ratio	0.05	0.09	0.33	0.77	0.98	0.99
Negative binomial model						
Bootstrap	0.04	0.06	0.04	0.22	0.26	0.54
GLM power	0.44	0.50	0.67	0.87	0.85	0.95

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 × 2 with null hypothesis of no interaction.

(θ_5, θ_6)	(0, 0)	(.2,.2)	(.4,.4)	(.6,.6)	(.8,.8)	(1, 1)
Bootstrap test	0.052	0.176	0.464	0.734	0.874	0.942
F test	0.040	0.152	0.428	0.694	0.846	0.932

Finally we consider comparing the powers for the bootstrap test and the standard F test for the example of 5 replicates of a 3×2 design with exponential errors raised to power 1.5. For each test we took 500 simulated samples for the model described in Section 5.1 for Table 3 with values of (θ_5, θ_6) as shown in Table 5. The powers shown in Table 5 demonstrate a moderate increase in power for the bootstrap test. Similar simulations for the case of normal errors show no loss of power under the model for which we might expect the F test to be optimal.

6. Proof of Theorem 1

Define $\hat{L}_\theta = \bar{M}_\theta^{-1}\bar{L}_\theta$. Let $Z_\theta = (\hat{L}_\theta, \bar{V}_\theta) = g(\bar{L}_\theta, \bar{V}_\theta)$. Choose $\epsilon > 0$ such that $\epsilon < \frac{1}{4n}\sum_{j=1}^n |\mathrm{E}_0\left[\psi_j'(Y_j, \theta_0)\right]|$, choose $\gamma > 0$ and $D > 0$, and define the set E by

$$E = \left\{(Y_1, \ldots, Y_n) : |\bar{M}_\theta| > \epsilon, \max|\bar{M}_\theta'| < D, |\hat{L}_\theta| < \frac{3}{4}\gamma, \text{ for } \theta \in B_\gamma^p(\theta_0)\right\}, \tag{28}$$

where $B_\gamma^p(\theta)$ denotes a cube of dimension p, side 2γ centered at θ. Then the conditions (A1)–(A4) together with Cramér's large deviation theorem [8, Theorem 15, Chapter 3] ensure that

$$\mathrm{Pr}_0(E) > 1 - e^{-cn}$$

for some $c > 0$ depending only on ϵ, γ, and D.

Lemma 1 of [1] ensures that for $(Y_1, \ldots, Y_n) \in E$, there is a unique solution T of $\sum_{j=1}^n \psi_j(Y_j, \theta) = 0$ in $B_\gamma^p(\theta_0)$.

Since densities for $Y_1, \ldots, Y_n$ might not exist, we find the probability of the tail event $\mathcal{F} = \{T : h(T_2) \geq \lambda\}$ by partitioning the space of (T, V_θ) into small regions, approximating $\mathrm{Pr}_0(\mathcal{F})$ by summing probabilities of the appropriate small regions and by approximating this sum by an integral. To do this we need to bound the probabilities of these small regions in the space of (T, V_θ) by probabilities of regions in the space of $\bar{U}_\theta$.

We note that $E \subset \{T \in B^p_{\frac{3}{4}\gamma}(\theta_0)\}$, so

$$\mathrm{Pr}_0\{h(T_2) \geq \lambda\} = \mathrm{Pr}_0\left\{T \in B^p_{\frac{3}{4}\gamma}(\theta_0) \cap \mathcal{F}\right\} + O(e^{-cn}). \tag{29}$$

For $\mathbf{j} = \mathbb{Z}^p$, let

$$\mathcal{B}_\mathbf{j} = \{x \in R^p : (j_\ell - 1/2)\delta < x_\ell \leq (j_\ell + 1/2)\delta\} \cap \left\{B^p_{\frac{3}{4}\gamma}(\theta_0)\right\},$$

for $\mathbf{k} = \mathbb{Z}^q$ let

$$\mathcal{D}_\mathbf{k} = \{x \in R^q : (k_\ell - 1/2)\delta < x_\ell \leq (k_\ell + 1/2)\delta\} \cap \{(-c, c)^q\}$$

and let $\mathcal{E}_\mathbf{j} = \mathcal{B}_\mathbf{j} \cap \mathcal{F}$. Then, under (A4), we have

$$\mathrm{Pr}_0\{h(T_2) \geq \lambda\} = \sum_\mathbf{j}\sum_\mathbf{k} \mathrm{Pr}_0\{(T, \bar{V}_{\theta_\mathbf{j}}) \in \mathcal{E}_\mathbf{j} \times \mathcal{D}_\mathbf{k}\} + O(e^{-cn}). \tag{30}$$

For a set $\mathcal{B}$ let $\mathcal{B}_\eta = \{x + y : x \in \mathcal{B}, y^\top y < \eta^2\}$. The following lemma is a simpler version of Lemma 1 of [4].

Lemma 1. *For $\theta_{\mathbf{j}} \in \mathcal{E}_{\mathbf{j}}$ with $0 < \delta < \gamma/4$, there is a $C > 0$, depending only on D and ϵ of (28), such that for δ chosen so that $C\delta < 1/4$,*

$$\{\hat{L}_{\theta_{\mathbf{j}}} \in (\mathcal{E}_{\mathbf{j}} - \theta_{\mathbf{j}})^-\} \subset \{T \in \mathcal{E}_{\mathbf{j}}\} \subset \{\hat{L}_{\theta_{\mathbf{j}}} \in (\mathcal{E}_{\mathbf{j}} - \theta_{\mathbf{j}})^+\}, \tag{31}$$

where $(\mathcal{E}_{\mathbf{j}} - \theta_{\mathbf{j}})^- = [\{(\mathcal{E}_{\mathbf{j}} - \theta_{\mathbf{j}})^c\}_{C\delta^2}]^c$ and $(\mathcal{E}_{\mathbf{j}} - \theta_{\mathbf{j}})^+ = (\mathcal{E}_{\mathbf{j}} - \theta_{\mathbf{j}})_{C\delta^2}$.

Proof. We have, for $T \in \mathcal{E}_{\mathbf{j}}$,

$$0 = \hat{L}_T = \bar{L}_{\theta_{\mathbf{j}}} + (T - \theta_{\mathbf{j}})\bar{M}_{\mathbf{j}} + O(\delta^2),$$

so $\hat{L}_{\theta_{\mathbf{j}}} = (\theta_{\mathbf{j}} - T) + O(\delta^2)$ and we can find C such that $\hat{L}_{\theta\mathbf{j}} \in (\mathcal{E}_{\mathbf{j}} - \theta_{\mathbf{j}})^+$. Also, for any $\delta \in (0, 3\gamma/4)$, we can choose C such that

$$\sup_{\theta'_{\mathbf{j}} \in \mathcal{E}_{\mathbf{j}}} |\bar{M}_{\theta'}^{-1}\bar{M}_{\theta_{\mathbf{j}}} - I_p| < C\delta.$$

If $\hat{L}_{\theta_{\mathbf{j}}} \in (\mathcal{E}_{\mathbf{j}} - \theta_{\mathbf{j}})^-$ and δ is such that $C\delta < 1/4$, then from Lemma 1 of [1], $L_T = 0$ has a unique solution $T \in \mathcal{E}_{\mathbf{j}}$. □

For $\theta_{\mathbf{j}} \in \mathcal{E}_{\mathbf{j}}$, let $\mathcal{A}^-_{\mathbf{jk}} = (\mathcal{E}_{\mathbf{j}} - \theta_{\mathbf{j}})^- \times \mathcal{D}_{\mathbf{k}}$ and $\mathcal{A}^+_{\mathbf{jk}} = (\mathcal{E}_{\mathbf{j}} - \theta_{\mathbf{j}})^+ \times \mathcal{D}_{\mathbf{k}}$, then the lemma applied to the probability of a cube gives

$$\Pr_0\{(\hat{L}_{\theta_{\mathbf{j}}}, \bar{V}_{\theta_{\mathbf{j}}}) \in \mathcal{A}^-_{\mathbf{jk}}\} < \Pr_0\{(T, \bar{V}_{\theta_{\mathbf{j}}}) \in \mathcal{E}_{\mathbf{j}} \times \mathcal{D}_{\mathbf{k}}\} < \Pr_0\{(\hat{L}_{\theta_j}, \bar{V}_{\theta_{\mathbf{j}}}) \in \mathcal{A}^+_{\mathbf{jk}}\}.$$

Writing $u = (\ell, v)$, let

$$e_d(u, F^{\tau(\theta)}) = \frac{\exp(-nu^{*T}u^*/2)}{(2\pi/n)^{(p+q)/2}|\Sigma_{\tau(\theta)}|^{1/2}}\left\{1 + \sum_{l=1}^{d} Q_{ln}(u^*\sqrt{n})\right\}, \tag{32}$$

where $u^* = \Sigma^{-1/2}_{\tau(\theta)}(u - \mu_{\tau(\theta)})$. This is the Edgeworth approximation to the tilted probabilities, as introduced by [9]. Let $(\hat{L}_\theta, \bar{V}_\theta) = g_\theta(\bar{L}_\theta, \bar{V}_\theta)$ and let

$$P_{\mathbf{jk}} = \Pr_0\{(\hat{L}_{\theta_{\mathbf{j}}}, \bar{V}_{\theta_{\mathbf{j}}}) \in \mathcal{A}^-_{\mathbf{jk}}\} = \Pr_0\{(\bar{L}_{\theta_{\mathbf{j}}}, \bar{V}_{\theta_{\mathbf{j}}}) \in g^{-1}_{\theta_{\mathbf{j}}}(\mathcal{A}^-_{\mathbf{jk}})\}$$

and using Theorem 1 of [9],

$$\begin{aligned} P_{\mathbf{jk}} &= e^{nK\{\tau(\theta_{\mathbf{j}}),\theta_{\mathbf{j}}\}}\left[\int_{g^{-1}_{\theta_{\mathbf{j}}}(\mathcal{A}^-_{\mathbf{jk}})} e^{-n\ell^\top\tau(\theta_{\mathbf{j}})} e_d\{(\ell, v), F^{\tau(\theta_{\mathbf{j}})}\}d\ell dv + R_{\mathbf{jk}}\right] \\ &= e^{nK\{\tau(\theta_{\mathbf{j}}),\theta_{\mathbf{j}}\}}\left[\int_{\mathcal{A}^-_{\mathbf{jk}}} J_{\theta_{\mathbf{j}}}(y)e^{-n\ell(y)^\top\tau(\theta_{\mathbf{j}})} e_d\{g^{-1}_{\theta_{\mathbf{j}}}(y), F^{\tau(\theta_{\mathbf{j}})}\}dy + R_{\mathbf{jk}}\right], \end{aligned}$$

where $J_{\theta_{\mathbf{j}}}(y)$ is the Jacobian of the transformation $y = g_{\theta_{\mathbf{j}}}(\ell, v)$ and we write $g^{-1}_{\theta_{\mathbf{j}}}(y) = (\ell(y), v(y))$ and $R_{\mathbf{jk}}$ corresponding to the residual from Theorem 1 of [9]. Noting that $\ell\{(0, v)\} = 0$ and $v\{(0, v)\} = v$, that $\exp\{n\ell(y)^\top\tau(\theta_{\mathbf{j}})\} = 1 + O(n\delta)$ and that $\mathrm{vol}(\mathcal{A}^-_{\mathbf{jk}}) = \mathrm{vol}(\mathcal{E}_{\mathbf{j}} \times \mathcal{D}_{\mathbf{k}})\{1 + O(\delta)\}$, we have, taking $\delta = n^{-2}$,

$$P_{\mathbf{jk}} = e^{nK\{\tau(\theta_{\mathbf{j}}),\theta_{\mathbf{j}}\}}[J_{\theta_{\mathbf{j}}}\{(0, v_{\mathbf{k}})\}e_d\{(0, v_{\mathbf{k}}), F^{\tau(\theta_{\mathbf{j}})}\}\mathrm{vol}(\mathcal{E}_{\mathbf{j}} \times \mathcal{D}_{\mathbf{k}})\{1 + O(n^{-1})\} + R_{\mathbf{jk}}],$$

where $v_{\mathbf{k}} \in \mathcal{D}_{\mathbf{k}}$. Consider

$$\mathcal{I}_{\mathbf{jk}} = \int_{\mathcal{E}_{\mathbf{j}} \times \mathcal{D}_{\mathbf{k}}} e^{nK\{\tau(t),t\}}J_t\{(0, v)\}e_d\{(0, v), F^{\tau(t)}\}dtdv. \tag{33}$$

Note that $(\theta_{\mathbf{j}}, v_{\mathbf{k}})$ is chosen arbitrarily in $\mathcal{E}_{\mathbf{j}} \times \mathcal{D}_{\mathbf{k}}$, so by the intermediate value theorem, we can choose $(\theta_{\mathbf{j}}, v_{\mathbf{k}})$ such that

$$\mathcal{I}_{\mathbf{jk}} = e^{nK\{\tau(\theta_{\mathbf{j}}),\theta_{\mathbf{j}}\}}J_{\theta_{\mathbf{j}}}\{(0, v_{\mathbf{k}})\}e_d\{(0, v_{\mathbf{k}}), F^{\tau(\theta_{\mathbf{j}})}\}\mathrm{vol}(\mathcal{E}_{\mathbf{j}} \times \mathcal{D}_{\mathbf{k}}).$$

So

$$P_{\mathbf{jk}} = \mathcal{I}_{\mathbf{jk}}(1 + O(n^{-1})) + e^{nK\{\tau(\theta_{\mathbf{j}}),\theta_{\mathbf{j}}\}}|\Sigma_{\tau(\theta_{\mathbf{j}})}|^{-1/2}(n/2\pi)^{d/2}R_{\mathbf{jk}}. \tag{34}$$

Now from (30),

$$\Pr_0(T \in \mathcal{F}) = \mathcal{I}(1 + O(n^{-1})) + R^*,$$

where, noting that integration outside $\mathcal{F} \times (-c, c)^q$ is exponentially small,

$$\mathcal{I} = \int_{\mathcal{F}}\int_{R^q} e^{nK\{\tau(t),t\}}J_t\{(0, v)\}e_d\{(0, v), F^{\tau(t)}\}dtdv \tag{35}$$

and

$$R^* = \sum_{\mathbf{j},\mathbf{k}} e^{nK\{\tau(\theta_\mathbf{j}),\theta_\mathbf{j}\}} R_{\mathbf{jk}}. \tag{36}$$

Note that in the integral $\mathcal{I}$, integration over the region complementary to $B^p_{\frac{3}{4}\gamma}(\theta_0)$ is incorporated in the $O(n^{-1})$ term as is the $O(e^{-cn})$ term of (30).

First consider $\mathcal{I}$. The second term of the Edgeworth expansion in (32) integrates over v to 0 and the higher terms can be incorporated in the $O(n^{-1})$ term. So

$$\mathcal{I} = \int_{\mathcal{F}} \int_{R^q} \frac{e^{nK\{\tau(t),t\} - nu_2^{*T} u_2^*/2} J_t(0, v)}{(2\pi/n)^{(p+q)/2} |\Sigma_{\tau(t)}|^{1/2}} dt dv. \tag{37}$$

Now using a Laplace approximation for the integral over v, we get

$$\mathcal{I} = \int_{\{t_2 : h(t_2) \geq \lambda\}} \int_{R^{p_1}} \frac{e^{nK\{\tau(t),t\}} J(t)}{(2\pi/n)^{p/2} |\Sigma_{L\tau(t)}|^{1/2}} dt_1 dt_2 \{1 + O(1/n)\}, \tag{38}$$

where $J(t) = |\mathrm{E}_{\tau(t)} \bar{M}_t|$. The relative error is from the Laplace approximation and from replacing $J_t(0, \mathrm{E}_{\tau(t)}[V_t])$ by $J(t)$.

Next we can use a Laplace approximation to the integral with respect to t_1, to obtain

$$\mathcal{I} = \int_{\{t_2 : h(t_2) \geq \lambda\}} \frac{e^{-nh(t_2)} J\{t(t_2)\}}{(2\pi/n)^{p_2} |\Sigma_{L\tau\{t(t_2)\}}|^{1/2} |H_{11}\{t(t_2)\}|^{1/2}} dt_2 \{1 + O(n^{-1})\}, \tag{39}$$

where

$$t(t_2) = (t_1(t_2), t_2) = \arg\inf_{t_1} [-K\{\tau(t), t\}], \tag{40}$$

$$h(t_2) = -K[\tau\{t(t_2)\}, t(t_2)], \tag{41}$$

$$H(t) = \frac{d^2 K\{\tau(t), t\}}{dt^2} \tag{42}$$

and $H_{11}(t)$ is the matrix of the first p_1 rows and columns of $H(t)$.

To reduce the integral in $\mathcal{I}$ to the form in the theorem we need some preliminary results. From (40), by noting that

$$\partial K\{\tau(t), t\}/\partial\tau = 0, \tag{43}$$

we have

$$0 = \left[\frac{dK\{\tau(t), t\}}{dt_1}\right]_{t=t(t_2)} = \left[\frac{\partial K\{\tau(t), t\}}{\partial t_1}\right]_{t=t(t_2)}. \tag{44}$$

So from (41), using (44) and (43),

$$h'(t_2) = -(t_1'(t_2), I_{p_2}) \left[\frac{\partial K\{\tau(t), t\}}{\partial t}\right]_{t=t(t_2)} = -\left[\frac{\partial K\{\tau(t), t\}}{\partial t_2}\right]_{t=t(t_2)}. \tag{45}$$

So, since the unconstrained minimum of $K(\tau(t), t)$ is 0, $h'(\theta_{20}) = 0$. Further

$$h''(t_2) = -(t_1'(t_2), I_{p_2}) \left[\tau'(t) \frac{\partial^2 K\{\tau(t), t\}}{\partial\tau \partial t_2} + \frac{\partial^2 K\{\tau(t), t\}}{\partial t \partial t_2}\right]_{t=t(t_2)}. \tag{46}$$

From (43)

$$\tau'(t) = -\frac{\partial^2 K\{\tau(t), t\}}{\partial t \partial\tau} \left[\frac{\partial^2 K\{\tau(t), t\}}{\partial\tau^2}\right]^{-1}, \tag{47}$$

so, from the definition (42),

$$\begin{aligned} H(t) &= \tau'(t) \frac{\partial^2 K\{\tau(t), t\}}{\partial\tau \partial t} + \frac{\partial^2 K\{\tau(t), t\}}{\partial t^2} \\ &= \frac{\partial^2 K\{\tau(t), t\}}{\partial t^2} - \frac{\partial^2 K\{\tau(t), t\}}{\partial t \partial\tau} \left[\frac{\partial^2 K\{\tau(t), t\}}{\partial\tau^2}\right]^{-1} \frac{\partial^2 K\{\tau(t), t\}}{\partial\tau \partial t}. \end{aligned} \tag{48}$$

Differentiating (44) with respect to t_2 gives

$$(t_1'(t_2), I_{p_2}) \left[\tau'(t) \frac{\partial^2 K\{\tau(t), t\}}{\partial\tau \partial t_1} + \frac{\partial^2 K\{\tau(t), t\}}{\partial t \partial t_1}\right] = 0,$$

so from (47) and (48),

$$t_1'(t_2) = -H_{21}\{t(t_2)\}H_{11}\{t(t_2)\}^{-1}, \tag{49}$$

where the subscripts 1 and 2 on H refer to the first p_1 and last p_2 rows and columns of $H(t)$. Now, from (46), using (47)–(49), we have

$$h''(t_2) = -H_{22}(t_2) + H_{21}\{t(t_2)\}H_{11}\{t(t_2)\}^{-1}H_{12}\{t(t_2)\}. \tag{50}$$

Now note that

$$\frac{\partial K\{\tau(t), t\}}{\partial t} = \frac{1}{n}\sum_{j=1}^{n} \tau(t)^\top \mathrm{E}\Big\{\frac{\partial \psi_j(Y_j, t)}{\partial t} e^{\tau(t)^\top \psi_j(Y_j,t)}\Big\} \Big/ \mathrm{E}\Big\{e^{\tau(t)^\top \psi_j(Y_j,t)}\Big\}$$

and so, since $\tau\{\theta(\theta_{20})\} = 0$,

$$\left[\frac{\partial^2 K\{\tau(t), t\}}{\partial t^2}\right]_{t=\theta(\theta_{20})} = 0.$$

Further,

$$\begin{aligned}\left[\frac{\partial^2 K\{\tau(t), t\}}{\partial \tau \partial t}\right]_{t=\theta(\theta_{20})} &= \left(\frac{1}{n}\sum_{j=1}^{n} \mathrm{E}\Big[\frac{\partial \psi_j(Y_j, t)}{\partial t} e^{\tau(t)^\top \psi_j(Y_j,t) - K\{\tau(t),t\}}\Big]\right)_{t=\theta(\theta_{20})} \\ &= \frac{1}{n}\sum_{j=1}^{n} \mathrm{E}\left[\frac{\partial \psi_j(Y_j, t)}{\partial t}\right]_{t=\theta(\theta_{20})} = \mathrm{E}\{\bar{M}_{\theta(\theta_{20})}\}\end{aligned}$$

and

$$\left[\frac{\partial^2 K\{\tau(t), t\}}{\partial \tau^2}\right]_{t=\theta(\theta_{20})} = \mathrm{E}\{\bar{Q}_{\theta(\theta_{20})}\}.$$

So from (50), $h''(\theta_{20}) = ([\mathrm{E}\{\bar{M}_{\theta(\theta_{20})}\}]^{-1}\mathrm{E}\{\bar{Q}_{\theta(\theta_{20})}\}[\mathrm{E}\{\bar{M}_{\theta(\theta_{20})}\}]^{-1})^{-1}_{22}$.

After making the transformations $y = \{h''(\theta_{20})\}^{1/2}(t_2 - \theta_{20})$, $y \to (r, s)$ and $(r, s) \to (u, s)$, with Jacobians $|h''(\theta_{20})|^{1/2}$, $J_1(t_2) = r^{p_2-1}$ and $J_2(t_2) = ru/(h'(t_2)^\top[h''(\theta_{20})]^{1/2}(t_2 - \theta_{20}))$, respectively, the results (16) and (17) follow as in the proofs of Theorem 1 of [10] and Theorem 2 of [6], if we show that

$$R^* = \mathcal{I}O(n^{-1}), \tag{51}$$

where $\mathcal{I}$ and R^* are defined in (35) and (36), respectively.

Theorem 1 of [9] shows that

$$|R_{\mathbf{jk}}| \le R_{1\mathbf{jk}} + R_{1\mathbf{jk}},$$

where

$$R_{1\mathbf{jk}} = Cn^{(p+q)/2}\Sigma^{-1/2}_{\boldsymbol{\tau}(\theta_{\mathbf{j}})}\mathrm{vol}[\{g^{-1}_{\theta_{\mathbf{j}}}(\mathcal{A}^-_{\mathbf{jk}})\}_{2\alpha}]\Big[\eta_s(\hat{\boldsymbol{\tau}})n^{-(s-2)/2} + |\Sigma_{\boldsymbol{\tau}(\theta_{\mathbf{j}})}|^{1/2}n^{d/2}\alpha^{-d}q_{\boldsymbol{\tau}}(1/\alpha)\Big] \tag{52}$$

and

$$R_{2\mathbf{jk}} = n^{(p+q)/2}|\Sigma_{\boldsymbol{\tau}(\theta_{\mathbf{j}})}|^{-1/2}\mathrm{vol}[\{g^{-1}_{\theta_{\mathbf{j}}}(\mathcal{A}^-_{\mathbf{jk}})\}_{2\alpha} - \{g^{-1}_{\theta_{\mathbf{j}}}(\mathcal{A}^-_{\mathbf{jk}})\}_{-2\alpha}]. \tag{53}$$

Here $\eta_s\{\boldsymbol{\tau}(\theta_{\mathbf{j}})\}$ is the standardized sth moment under $F^{\boldsymbol{\tau}(\theta_{\mathbf{j}})}$ and so is bounded by (A2). The definition of $q_{\boldsymbol{\tau}}(1/\alpha)$, given in (1.24) of [9], and (A3) imply that $q_{\boldsymbol{\tau}}(n^2) = O(e^{-cn})$. In E, the determinant of $\bar{M}_{\theta_{\mathbf{j}}}$ is bounded and bounded away from 0, so for some C_1, C_2 and C_3,

$$g^{-1}_{\theta_{\mathbf{j}}}\{(\mathcal{A}^-_{\mathbf{jk}})_{C_1\alpha}\} \supset \{g^{-1}_{\theta_{\mathbf{j}}}(\mathcal{A}^-_{\mathbf{jk}})\}_{2\alpha}$$

and

$$\mathrm{vol}[\{g^{-1}_{\theta_{\mathbf{j}}}(\mathcal{A}^-_{\mathbf{jk}})\}_{2\alpha}] \le C_2\mathrm{vol}\{(\mathcal{A}^-_{\mathbf{jk}})_{C_1\alpha}\} \le C_2\mathrm{vol}[\{(\mathcal{B}_{\mathbf{j}} \cap \mathcal{F}) \times \mathcal{D}_{\mathbf{k}}\}_{C_3\alpha}],$$

noting that $\delta^2 = n^{-4}$. So

$$R_{1\mathbf{jk}} = \mathrm{vol}[\{(B_{\mathbf{j}} \cap \mathcal{F}) \times \mathcal{D}_{\mathbf{k}}\}_{C_3\alpha}]O(n^{(p+q)/2-(d+1)/2}).$$

Also $\mathrm{vol}[\{g^{-1}_{\theta_{\mathbf{j}}}(\mathcal{A}^-_{\mathbf{jk}})\}_{2\alpha} - \{g^{-1}_{\theta_{\mathbf{j}}}(\mathcal{A}^-_{\mathbf{jk}})\}_{-2\alpha}] \le 4\alpha\,\mathrm{sur}\{g^{-1}_{\theta_{\mathbf{j}}}(\mathcal{A}^-_{\mathbf{jk}})\}$, so

$$R_{2\mathbf{jk}} = \mathrm{vol}\{(\mathcal{A}^-_{\mathbf{jk}})_{C_1\alpha}\}O(1/n) = \mathrm{vol}[\{(\mathcal{B}_{\mathbf{j}} \cap \mathcal{F}) \times \mathcal{D}_{\mathbf{k}}\}_{C_3\alpha}]O(n^{(p+q)/2}\alpha/\delta).$$

So, since $\delta = n^{-2}$, if $\alpha = n^{-(d+5)/2}$ and $d = q/2 + 1$,

$$|R_{\mathbf{jk}}| = \mathrm{vol}[\{(\mathcal{B}_{\mathbf{j}} \cap \mathcal{F}) \times \mathcal{D}_{\mathbf{k}}\}_{C_3\alpha}]O(n^{p/2-1}).$$

Using the same device as in [6], let $\mathcal{G}_{\mathbf{jk}} = (\mathcal{B}_{\mathbf{j}} \cap \mathcal{F}_{C_3\alpha}) \times \mathcal{D}_{\mathbf{k}}$ and let $\mathcal{G}^*_{\mathbf{jk}}$ be the union of $\mathcal{G}_{\mathbf{jk}}$ and the $3^{p+q} - 1$ reflections across each lower dimensional face of the containing $\mathcal{B}_{\mathbf{j}} \times \mathcal{D}_{\mathbf{k}}$. Then $\{(\mathcal{B}_{\mathbf{j}} \cap \mathcal{F}) \times \mathcal{D}_{\mathbf{k}}\}_{C_3\alpha} \subset \mathcal{G}^*_{\mathbf{jk}}$ and so

$$\mathrm{vol}[\{(\mathcal{B}_{\mathbf{j}} \cap \mathcal{F}) \times \mathcal{D}_{\mathbf{k}}\}_{C_3\alpha}] \leq 3^{p+q}\mathrm{vol}(\mathcal{G}_{\mathbf{jk}}).$$

So

$$R^* = \sum_{\mathbf{jk}} e^{nK\{\boldsymbol{\tau}(\theta_{\mathbf{j}}),\theta_{\mathbf{j}}\}}\mathrm{vol}(\mathcal{G}_{\mathbf{jk}})O(n^{p/2-1}).$$

So summing over $\mathbf{k}$, noting that $\sum_{\mathbf{k}} \mathrm{vol}(\mathcal{D}_{\mathbf{k}}) = (2c)^q$, gives

$$R^* = \sum_{\mathbf{j}} r(\theta_{\mathbf{j}})\mathrm{vol}(\mathcal{B}_{\mathbf{j}} \cap \mathcal{F}_{C_3\alpha})O(n^{-1}),$$

for

$$r(\theta) = \frac{e^{nK\{\tau(\theta),\theta\}}J(\theta)}{|\Sigma_{\tau(\theta)}|^{1/2}(2\pi/n)^{p/2}},$$

since, in E, $J(\theta)$ and $\det(\Sigma_{\tau(\theta)})$ are bounded and bounded away from 0. So

$$R^* = \int_{\mathcal{F}_{C_3\alpha}} r(\theta)d\theta O(n^{-1}) = \int_{\mathcal{F}} r(\theta)d\theta O(n^{-1}) = \mathcal{I}O(n^{-1}),$$

since $\alpha = n^{-q/2-3}$. $\square$

Acknowledgments

We wish to thank Chris Field and Elvezio Ronchetti for contributions to earlier parts of this work. John Kolassa was supported in part by NSF DMS 0906569. John Robinson was supported by ARC DP0773345.

References

[1] A. Almudevar, C. Field, J. Robinson, The density of multivariate M-estimates, Ann. Statist. (ISSN: 00905364) 28 (2000) 275–297. URL http://www.jstor.org/stable/2673990.
[2] E. Cantoni, E. Ronchetti, Robust inference for generalized linear models, J. Amer. Statist. Assoc. (ISSN: 01621459) 96 (2001) 1022–1030. URL http://www.jstor.org/stable/2670248.
[3] T.J. DiCiccio, J.P. Romano, Nonparametric confidence limits by resampling methods and least favorable families, Int. Statist. Rev. (ISSN: 03067734) 58 (1990) 59–76. URL http://www.jstor.org/stable/1403474.
[4] C. Field, J. Robinson, E. Ronchetti, Saddlepoint approximations for multivariate M-estimates with applications to bootstrap accuracy, Ann. Inst. Statist. Math. 60 (2008) 205–224.
[5] A. Genz, Fully symmetric interpolatory rules for multiple integrals over hyper-spherical surfaces, J. Comput. Appl. Math. (ISSN: 0377-0427) 157 (2003) 187–195. http://dx.doi.org/10.1016/S0377-0427(03)00413-8, URL http://www.sciencedirect.com/science/article/pii/S0377042703004138.
[6] J. Kolassa, J. Robinson, Saddlepoint approximations for likelihood ratio like statistics with applications to permutation tests, Ann. Statist. (ISSN: 0090-5364) 39 (2011) 3357–3368. URL http://dx.doi.org/10.1214/11-AOS945.
[7] S.N. Lô, E. Ronchetti, Robust and accurate inference for generalized linear models, J. Multivariate Anal. 100 (2009) 2126–2136. URL http://EconPapers.repec.org/RePEc:eee:jmvana:v:100:y:2009:i:9:p:2126-2136.
[8] V. Petrov, Sums of Independent Random Variables, Springer, New York, 1975.
[9] J. Robinson, T. Hoglund, L. Holst, M.P. Quine, On approximating probabilities for small and large deviations in $\mathbb{R}^d$, Ann. Probab. (ISSN: 00911798) 8 (1990) 727–753. URL http://www.jstor.org/stable/2244314.
[10] J. Robinson, E. Ronchetti, G.A. Young, Saddlepoint approximations and tests based on multivariate M-estimates, Ann. Statist. (ISSN: 00905364) 31 (2003) 1154–1169. URL http://www.jstor.org/stable/3448455.