The Version of Record of this Manuscript has been published and is available in Communications in Statistics - Simulation and Computation (ISSN: 0361-0918) from the 21 Jun 2017 on https://www.tandfonline.com/doi/abs/10.1080/03610918.2017.1309427?journalCode=lssp20

https://doi.org/10.1080/03610918.2017.1309427

Bootstrap adjustments of signed scoring rule root statistics

V. Mameli[a] * and M. Musio[b] and L.Ventura[c]

[a]Ca' Foscari University of Venice, Department of Environmental Sciences, Informatics and Statistics, Italy

[b]University of Cagliari, Department of Mathematics and Computer Science, Italy

[c]University of Padova, Department of Statistical Sciences, Italy

March 9, 2017

Abstract

Scoring rules give rise to methods for statistical inference and are useful tools to achieve robustness or reduce computations. Scoring rule inference is generally performed through first-order approximations to the distribution of the scoring rule estimator or of the ratio-type statistic. In order to improve the accuracy of first-order methods even in simple models, we propose bootstrap adjustments of signed scoring rule root statistics for a scalar parameter of interest in presence of nuisance parameters. The method relies on the parametric bootstrap approach that avoids onerous calculations specific of analytical adjustments. Numerical examples illustrate the accuracy of the proposed method.

Keywords: Asymptotic expansions; Higher-order inference; Parametric Bootstrap; Regression models; Robustness; Tsallis scoring rule.

1 Introduction

A proper scoring rule $S(x;Q)$ provides a way of judging the quality of a quoted probability distribution Q for a random variable X in the light of its outcome x and gives a measure of how good specific probabilities are. The mathematical theory of proper scoring rules has a wide range of applications in Statistics; a review of the general theory, with applications, has been given recently in Dawid and Musio (2014).
Since every statistical decision problem induces a proper scoring rule, there is a very wide variety of these; see Dawid and Musio (2014). The most famous is the logarithmic score, which is highly connected with likelihood inference. Proper scoring rules, different from the

logarithmic score, can be used as an alternative to the full likelihood, when the interest is in increasing robustness or simplifying computations. Examples of particular interest include the *Tsallis* score that in general gives robust procedures (see Dawid et al. (2016)) and the *Hyvärinen* score that satisfies the property of homogeneity, which implies that the quoted distribution need only to be known up to the normalisation constant (see Parry et al. (2012), Ehm and Gneiting (2012)).
In a parametric setting, proper scoring rules induce natural unbiased estimating equations and they form a special case of M-estimation. We can then apply the general theory for M-estimators to proper scoring rules. Scoring rule inference is generally performed through first-order approximations to the distribution of the scoring rule estimator or of the scoring rule ratio statistic. However, there are various examples (see e.g. Dawid et al. (2016), Mameli and Ventura (2015), and references therein) which illustrate the inaccuracy of these methods, even in models with a scalar parameter, for small or moderate sample sizes.

Many attempts have been made to improve first-order approximations in the likelihood framework; see, e.g., Brazzale et al. (2007), Brazzale and Davison (2008), Young (2009), and references therein. There are two routes to obtain higher-order accuracy in parametric inference: analytic methods and parametric bootstrap.

Analytical higher-order asymptotic expansions for proper scoring rules, which encompass the classical results for likelihood quantities while allowing for the failing of the information identity, have been recently discussed in Mameli and Ventura (2015). However, the analytical adjustments of the signed scoring rule root statistic (or its profile version) require the evaluation of some quantities which are rather complex as the dimension of the parameter (or of the nuisance parameter) increases, even for simple models. To avoid onerous calculations specific of analytical procedures, we discuss bootstrap adjustments of signed scoring rule root statistics for a scalar parameter of interest in the presence of nuisance parameters. The method relies on parametric bootstrap procedures, paralleling results for likelihood statistics (see, e.g., Young (2009)). The use of parametric bootstrap has been discussed by Di Ciccio et al. (2001) for the signed roots of likelihood ratios, and has also been proposed by Aerts and Claeskens (1999) to approximate the distribution of general pseudo-log likelihood ratios. The proposed adjustment is also related to the formulation of pre-pivoting introduced by Beran (1987), Beran (1988) and refined by Lee and Young (2003).

The paper unfolds as follows. Section 2 gives a background on scoring rules, from first-order inference to higher-order analytical adjustments. The bootstrap adjustment is illustrated in Section 3. Section 4 discusses numerical examples in the context of regression models in order to illustrate the accuracy of the proposed method. Concluding remarks close the paper.

2 Background

Let X be a random variable with values in $\mathcal{X}$, and let $\mathcal{P}$ be a family of distributions over $\mathcal{X}$. Suppose a forecaster expresses the uncertainty about X by quoting a distribution Q for

it. A scoring rule $S(x;Q)$ is a special kind of loss function intended to measure the quality of the quoted distribution when an outcome x of X actually occurs. A scoring rule is said to be *proper* if, for any distribution P for X, the expected score $S(P,Q) := E_{X\sim P}S(X;Q)$ is minimised by quoting $Q = P$, and *strictly proper* if the minimum is unique. In other words, the forecaster must be honest. Every statistical decision problem induces a proper scoring rule, so there is a wide variety of these. Here, we present three cases of special interest. For further special cases see among others (Dawid (2007), Dawid and Sebastaini (1999), Gneiting and Raftery (2007)).
Let $p(\cdot)$ be the density (Radon-Nikodym derivative) $dP/d\mu$ with respect to an underlying measure μ dominating $\mathcal{P}$.
The *log-score* or *logarithmic score*, proposed by Good (1952), is given by

$$S(x;Q) = -\ln q(x), \tag{1}$$

where $q(\cdot)$ denotes the density of Q. The log-score is the only proper scoring rule that is local, i.e. it depends on the density function $q(x)$ only through its value at x (see Bernardo (1979)).
The *Tsallis* score (see Tsallis (1988), Basu et al. (1998)) is given by

$$S(x;Q) = (\gamma-1)\int d\mu(y)\cdot q(y)^{\gamma} - \gamma q(x)^{\gamma-1} \quad (\gamma > 1). \tag{2}$$

The Tsallis score gives in general robust procedures; see Basu et al. (1998) and Dawid et al. (2016).
Let $\boldsymbol{X}$ be a variable taking values in $\mathcal{X} = \mathbb{R}^k$. The *Hyvärinen score* (Hyvärinen (2005), Almeida and Gidas (1993)) is defined by

$$S(\boldsymbol{x};Q) = \Delta \ln q(\boldsymbol{x}) + \frac{1}{2}\left|\nabla \ln q(\boldsymbol{x})\right|^2 = \frac{\Delta\sqrt{q(\boldsymbol{x})}}{\sqrt{q(\boldsymbol{x})}}, \tag{3}$$

where ∇ and Δ are the gradient and the Laplacian operator on $\mathcal{X}$, respectively, and $|\boldsymbol{u}|^2 = \langle \boldsymbol{u}, \boldsymbol{u}\rangle$, where $\langle \boldsymbol{u}, \boldsymbol{v}\rangle$ denotes the inner product between vectors $\boldsymbol{u}$ and $\boldsymbol{v}$.
The Hyvärinen score depends on the predictive density only through its value, and the values of its first and second derivatives at the observation x. Furthermore, the Hyvärinen score is homogeneous, it is unchanged if $q(\cdot)$ is scaled by a positive constant. In particular, $S(x;Q)$ can be computed without knowledge of the normalising constant of the distribution Q. Parry et al. (2012) discussed a general characterization of the local scoring rules of all orders.

2.1 First-order inference

Suppose we wish to fit a parametric statistical model $F_\theta = F(x;\theta)$, indexed by a k-dimensional parameter $\theta \in \Theta \subseteq \mathbb{R}^k$ $(k \geq 1)$, based on the random sample $(x_1, \ldots, x_n)$ of

size n. Given a proper scoring rule, in the following of the paper we will implicitly identify $S(x; F_\theta)$ with $S(x; \theta)$.

For inference on θ, we might assess the goodness-of-fit by the total empirical score $S(\theta) = \sum_{i=1}^{n} S(x_i; \theta)$. The scoring rule estimator $\hat{\theta}_S$ is the value of θ which minimizes $S(\theta)$, i.e. $\hat{\theta}_S = \arg\min_\theta S(\theta)$.
Asymptotic arguments indicate that $\hat{\theta}_S \to \theta_0$ as $n \to \infty$, where θ_0 is the true parameter value. Maximum likelihood estimation, as well as composite likelihood estimation (Varin et al. (2011)), are special cases of score estimation when $S(\theta)$ is the negative log-likelihood (Dawid and Musio (2014)).

Scoring rule estimation forms a special case of M-estimation (see, e.g., Huber and Ronchetti (2009)). Let $s_\theta(x; \theta) = \partial S(x; \theta)/\partial\theta$. Under suitable regularity conditions on the scoring rule and on the statistical model, $\hat{\theta}_S$ is the solution of the scoring rule estimating equation

$$s_\theta(\theta) = \sum_{i=1}^{n} s_\theta(x_i; \theta) = 0 \ . \tag{4}$$

It can be shown that the scoring rule estimating function $s_\theta(\theta)$ is an unbiased estimating function for any proper scoring rule (see Dawid and Lauritzen (2005), Dawid and Musio (2014), and Dawid et al. (2016)). For instance, when $S(x; \theta)$ is the logarithmic score in (1), the total empirical score is $S(\theta) = -\sum_{i=1}^{n} \log q(x_i; \theta)$, the scoring rule estimating equation is the likelihood equation $s_\theta(\theta) = \sum_{i=1}^{n} (\partial \log q(x_i; \theta))/(\partial\theta)$, and the scoring rule estimator $\hat{\theta}_S$ reduces to the maximum likelihood estimator (MLE). The scoring rule estimator $\hat{\theta}_S$ is consistent and asymptotically normal with mean θ and variance

$$V(\theta) = K(\theta)^{-1} J(\theta) (K(\theta)^{-1})^\intercal \ ,$$

with $J(\theta) = E_\theta(s_\theta(\theta) s_\theta(\theta)^\intercal)$ and $K(\theta) = E_\theta(\partial s_\theta(\theta)/\partial\theta^\intercal)$. The matrix $G(\theta) = V(\theta)^{-1}$ is known as the Godambe information matrix (Godambe (1960)).

The general theory of M-estimators could be also used to derive the influence function (IF) of the estimator $\hat{\theta}_S$, which is defined as

$$IF(x; \theta, s) = K(\theta)^{-1} s_\theta(x; \theta) \ ,$$

which assess the effect of an infinitesimal contamination at x on the estimator $\hat{\theta}_S$. The estimator $\hat{\theta}_S$ will be B-robust if and only if $s_\theta(x; \theta)$ is bounded in x, for each θ. Sufficient conditions for the robustness of the Tsallis score are discussed in Dawid et al. (2016).

Asymptotic inference on the parameter vector θ, in particular concerning the construction of asymptotic hypothesis tests and confidence regions, can be found in Dawid et al. (2016). Here, we focus on the case in which θ is a parameter vector partitioned as $\theta = (\psi, \lambda)$, where ψ is a k_0-dimensional parameter of interest and λ is a $(k - k_0)$-dimensional nuisance parameter. In this case, in analogy with the full parameter case (see Dawid et al. (2016)), we can define the profile scoring rule ratio statistic for ψ as

$$W_p^S(\psi) = 2 \left(S(\hat{\theta}_S) - S(\hat{\theta}_{S\psi}) \right) \xrightarrow{L} \sum_{j=1}^{k_0} \nu_j Z_j^2, \tag{5}$$

where $\hat{\theta}_{S\psi}$ is the constrained scoring rule estimate, $\nu_1, \ldots, \nu_{k_0}$ are the eigenvalues of the matrix $(K^{\psi\psi})^{-1}G^{\psi\psi}$, with $K^{\psi\psi}$ and $G^{\psi\psi}$ sub-matrices of the inverses of $G(\theta)$ and $K(\theta)$ with respect to ψ, respectively. The variables Z_j for $j = 1, \cdots, k_0$ are independent normal variables. It should be noted that the asymptotic null distribution of (5) depends both on the statistical model and on the parameter of interest, therefore adjustments of the profile scoring rule ratio statistic have received consideration; see Dawid et al. (2016). In particular, when ψ is scalar, an adjustment of $W_p^S(\psi)$ is given by (see Dawid et al. (2016))

$$W_p^S(\psi)_{adj} = \frac{K^{\psi\psi}(\widehat{\theta}_{S\psi})}{G^{\psi\psi}(\widehat{\theta}_{S\psi})} W_p^S(\psi) \ .$$

In a similar fashion, a simple adjustment of the profile signed scoring rule root statistic

$$r_p^S(\psi) = \mathrm{sgn}(\hat{\psi}_S - \psi)\sqrt{W_p^S(\psi)} \ ,$$

which can be shown to be asymptotically normal distributed, is given by

$$r_{p\,adj}^S(\psi) = \mu_p(\psi)^{-1/2} r_p^S(\psi) \ ,$$

with

$$\mu_p(\psi) = [G^{\psi\psi}(\hat{\theta}_{S\psi})^{-1} K^{\psi\psi}(\hat{\theta}_{S\psi})]^{-1} \ . \tag{6}$$

2.2 Higher-order inference

Several examples (see e.g. Dawid et al. (2016), Mameli and Ventura (2015), and references therein) illustrate the inaccuracy of first-order methods in the scoring rule framework, even in models with a scalar parameter, when the sample size is small or moderate. The key to unlocking the limiting behaviour of scoring rules quantities is higher-order asymptotic expansions (see Mameli and Ventura (2015)), generalizing results for likelihood quantities but allowing for the failure of the information identity, i.e., to be more explicit when $K(\theta) \neq J(\theta)$.

Asymptotic expansions are useful for deriving approximate distribution functions for appropriate pivots, such as the profile signed scoring rule root statistic. In particular, when ψ is scalar, using the asymptotic expansions of the mean and the variance of $r_p^S(\psi)$, it is possible to define the modified profile signed scoring rule root statistic

$$r_{Mp}^S(\psi) = \frac{r_p^S(\psi) - m_p(\psi)}{\sqrt{\mu_p(\psi) + v_p(\psi)}} \ , \tag{7}$$

where $m_p(\psi)$ is of order $O(n^{-1/2})$ and $v_p(\psi)$ is of order $O(n^{-1})$, and $\mu_p(\psi)$ is given in (6). The analytical expressions of the mean and variance corrections are derived in Mameli and Ventura (2015) and they involve several expected values of scoring rules derivatives. The mean and the variance corrections in (7) can improve the accuracy of the asymptotic normal approximation to the distribution of $r_{Mp}^S(\psi)$ (see Di Ciccio et al. (2001), Lee and Young

(2003), Stern (2006), and Mameli and Ventura (2015)). In particular, if third- and higher-order cumulants of $r^S_{Mp}(\psi)$ are of order $O(n^{-3/2})$ or smaller, then the distribution of the modified profile signed scoring rule root statistic $r^S_{Mp}(\psi)$ follows asymptotically a normal distribution with error $O(n^{-3/2})$. This happens in models with all higher cumulants zero and for the logarithmic score. In general if such conditions are not fulfilled, $r^S_{Mp}(\psi)$ is asymptotically normal distributed with error $O(n^{-1})$.

3 Bootstrap methods for signed scoring rule root statistics

Asymptotic formulae of the mean and variance corrections are complex expressions which involve expectations of higher-order scoring rules derivatives, whose computation may be cumbersome even in simple models. These expressions can be easily calculated through Monte Carlo simulation from the model. The objective of this paper is the investigation of a parametric bootstrap approach for computing adjustments of the scoring rules signed root statistic. The sample null distribution of $r^S_p(\psi)$ depend on the unknown parameter, therefore it is a non-pivotal statistic. The approximation of the distributions of $r^S_{Mp}(\psi)$ can be derived by resorting to the parametric bootstrap approach based on pre-pivoting, where, with pre-pivoting, we mean the estimation of a relevant sampling distribution through constrained parametric bootstrapping. This technique can be used for reducing both the error level of tests and the coverage error of confidence regions (Beran (1987, 1988)). This approach has the advantage of being easily implemented. Additionally, all the computations can be performed by Monte Carlo simulation. It only requires to simulate from the assumed model, as in the Approximate Bayesian Computation framework (Marin et al. (2012)) or in the composite likelihood approach (Cattelan and Sartori (2016)).
Consider the parametric family F_θ where $\theta = (\psi, \lambda)$. Under fairly general assumptions, Gatto and Ronchetti (1996) derived a saddlepoint approximation for the sample distribution of $j(T(\hat{F}_n)) - j(T(F_{\theta_0}))$, where $j(\cdot)$ is a real valued function, T is a statistical functional and $\hat{F}_n$ represents the empirical distribution function.
Assuming that $j(\cdot) = 2S(\cdot)$ and $\hat{\theta}_S = T(\hat{F}_n)$, we have that $W^S_p(\psi) = j(\hat{\theta}_S) - j(\hat{\theta}_{S\psi})$. Moreover, suppose that the first four cumulants of $j(\hat{\theta}_S)$ are available and that $\hat{\theta}_S$ can be expanded up to the second order term, then an approximation of the distribution function G of $W^S_p(\psi)$ is given by

$$G\left(y;\psi\right) = \Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{q}\right) + O(n^{-1}), \tag{8}$$

where $q = \alpha\left(\tilde{R}''_n(\alpha)\right)^{1/2}$, $r = sgn(\alpha)\left(2n(\alpha y - \tilde{R}_n(\alpha))\right)$, with α the saddlepoint defined by $\tilde{R}'_n(\alpha) = y$ and $\tilde{R}_n(\alpha)$ is an approximate cumulant generating function for the unbiased estimating function s_θ (see Gatto and Ronchetti (1996); La Vecchia (2016)). Then it is easy

to show that an approximation of the distribution function H of $r_{Mp}^{S}(\psi)$ is

$$H\left(y;\psi\right)=\Phi(r)+\phi(r)\left(\frac{1}{r}-\frac{1}{q}\right)+O(n^{-1}), \tag{9}$$

with $q=\alpha\left(\tilde{R}_n''(\alpha)\right)^{1/2}$, $r=sgn(\alpha)\left[2n\left(\alpha t(y;\psi)^2-\tilde{R}_n(\alpha)\right)\right]$, with α the saddlepoint defined by $\tilde{R}_n'(\alpha)=t(y;\psi)^2$. Moreover, $t(y;\psi)=y\sqrt{\tilde{v}_p(\psi)}+m_p(\psi)$ where $m_p(\psi)$ and $\tilde{v}_p(\psi)=\mu_p(\psi)+v_p(\psi)$ represent the mean and the variance corrections. In analogy, the bootstrap counterpart of (9) is

$$H^{*}\left(y;\psi\right)=\Phi(r^{*})+\phi(r^{*})\left(\frac{1}{r^{*}}-\frac{1}{q^{*}}\right)+O(n^{-1}), \tag{10}$$

with q^{*} and r^{*} indicating the bootstrap versions of q and r, respectively. Moreover, the saddlepoint α is defined by $\tilde{R^{*}}_{n}^{'}(\alpha)=t^{*}(y;\psi)^2$ with $t^{*}(y;\psi)=y\sqrt{\tilde{v}_b(\psi)}+m_b(\psi)$ where $m_b(\psi)$ and $\tilde{v}_b(\psi)$ indicate the bootstrap constrained mean and the variance corrections. Note also that $\tilde{R^{*}}_{n}(\alpha)$ is the bootstrap version of $\tilde{R}_n(\alpha)$. The difference $H\left(y;\psi\right)-H^{*}\left(y;\psi\right)$ is $O_p(n^{-1})$, then we will obtain inferences which are second order accurate.

The main steps of the constrained parametric bootstrap approach can be summarized as follows:

Step 0–Preliminaries: Let $x=(x_1,\ldots,x_n)$ be the original sample and let ψ_0 a fixed value for ψ. Evaluate the constrained scoring rule estimate $\hat{\theta}_{S\psi_0}=(\psi_0,\hat{\lambda}_{S\psi_0})$.

Step 1–Outer level: For $b=1,\ldots,B$, sample with replacement n elements from $F(x;\hat{\theta}_{S\psi_0})$ obtaining the b-th new sample x^b. Using x^b, compute the b-th bootstrap version of $r_p^S(\psi_0)$, given by $r_p^{Sb}(\psi_0)$, and store it.

Step 2–Inner level: Compute the bootstrap mean and variance

$$m^b(\psi_0)=\frac{1}{B}\sum_{i=1}^{B}r_p^{Sb}(\psi_0)$$

and

$$v^b(\psi_0)=\frac{1}{B}\sum_{i=1}^{B}(r_p^{Sb}(\psi_0)-m^b(\psi_0))^2\ .$$

Step 3–Final step: Adjust $r_p^S(\psi_0)$ using the bootstrap mean and variance, i.e.

$$r_{Mp}^{Sb}(\psi_0)=\frac{r_p^S(\psi_0)-m^b(\psi_0)}{\sqrt{v^b(\psi_0)}}\ .$$

In parametric bootstrap with nuisance parameter, the B samples could alternatively be drawn from a distribution $F(x;\tilde{\theta})$, where $\tilde{\theta}=(\tilde{\psi},\tilde{\lambda})$ is a suitable estimate of θ. Resulting

accuracy depends on the choice of $\tilde{\theta}$, provided B is large enough. In the classical likelihood approach, results in Di Ciccio et al. (2001) and Stern (2006) show that if $\tilde{\theta}$ is a $\sqrt{n}$-consistent estimator of θ, then the parametric bootstrap provides a $O(n^{-1})$ order of accuracy, while the special choice $\hat{\theta}_{S\psi}$ provides an additional order of accuracy, i.e. $O(n^{-3/2})$ order of accuracy. The resulting approximation is called constrained pre-pivoting of the signed likelihood root statistic (see Di Ciccio et al. (2001)). Other natural choices, such as $\tilde{\theta} = (\psi, \tilde{\lambda})$ or $\tilde{\theta} = \hat{\theta}_S$ are then second-order accurate. For general scoring rules, the bootstrap does not improve on the asymptotic approximation of $r^S_{Mp}(\psi)$, and thus the normal approximation of the distribution of $r^{Sb}_{Mp}(\psi)$ has an error up to the second order, i.e. to order $O(n^{-1})$. An exception could occur when third- and higher-order cumulants of $r^S_M(\psi)$ are of order $O(n^{-3/2})$ or smaller. In this case, the parametric bootstrap could provide a $O(n^{-3/2})$ order of accuracy, this happens for the logarithmic score (see Di Ciccio et al. (2001), Mameli and Ventura (2015)). The accuracy of the proposed method is clearly dependent on the number B of bootstrap simulations used. As in Di Ciccio et al. (2001), we note that means and variances require fewer simulations for precise estimation than do, for instance, tail probabilities. In particular, if confidence intervals are to be used, then it is advisable to have a small B, whereas in simulation based test at least $B = 999$ should generally be safe (see Davison and Hinkley (1997), Sections 4.2.5 and 5.2.3).

4 Simulation studies

In this section we provide simulation results to assess coverage levels of confidence intervals for a scalar parameter of interest ψ in presence of nuisance parameters based on the bootstrap modified profile signed scoring rule root statistic $r^{Sb}_{Mp}(\psi)$, both when the model is correctly specified and under model contamination. In particular, Example 4.1 focuses on the linear regression model while Example 4.2 treats the non-linear regression model. For Example 4.1 the stability of the coverage levels under model contamination is assessed by considering a contamination model of the form $Q_\epsilon = (1-\epsilon)Q_\theta + \epsilon P$, where P denotes the contaminating distribution, and ϵ is the contamination percentage. The contamination percentage ϵ is set at 10%. In order to show the instability of the classical likelihood procedures, in Example 4.2 we consider a sensitivity analysis. The parameter γ of the Tsallis score has been fixed in order to achieve approximately 90% or 95% efficiency under the assumed model (see Basu et al. (1998); (Huber and Ronchetti, 2009, Chapter 6), Mameli and Ventura (2015)).

4.1 Linear regression model

Let us consider a linear regression model as in Mameli and Ventura (2015) of the form

$$y = X\beta + \sigma\varepsilon,$$

where X is a $n \times p$ fixed matrix of explanatory variables, $\beta \in \mathbb{R}^p$ $(p \geq 1)$ an unknown regression coefficient, $\sigma > 0$ a scale parameter, and ε an n-dimensional vector of random

errors from a standard normal distribution $N(0,1)$. Let $\theta = (\beta, \sigma)$. The Hyvärinen total empirical score is

$$S_H(\theta) = \frac{2n}{\sigma^2} - \frac{1}{\sigma^4}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2,$$

where x_i^T is the i-th row of X, $i = 1, \ldots, n$.
The Tsallis total empirical score is (see Mameli and Ventura (2015), and reference therein)

$$S_T(\theta) = \frac{\gamma}{(2\pi\sigma^2)^{\frac{\gamma-1}{2}}}\sum_{i=1}^{n} e^{-\frac{(\gamma-1)}{2\sigma^2}(y_i - x_i^T\beta)^2} - \frac{n(\gamma-1)}{\sqrt{\gamma}(2\pi\sigma^2)^{(\gamma-1)/2}}.$$

The Tsallis score estimator is B-robust since the influence function is bounded (see Dawid et al. (2016) and Mameli and Ventura (2015)).
For $p = 3$, let $\psi = \beta_2$ be the scalar parameter of interest, and let $\lambda = (\beta_1, \beta_3, \sigma)$ be the nuisance parameter. In order to assess the accuracy of the parametric bootstrap modified profile signed scoring rule root statistic ($r_{Mp}^{Sb}(\beta_2)$), we consider the same simulation set-up as in Mameli and Ventura (2015). In particular, we ran a simulation experiment, for several values of n and with $\theta = (1, 2, 3, 1)$, when the model is correctly specified and when the contaminated model is $Q_{0.10} = 0.90N(0, 100) + 0.1N(0, 50^2)$. Results of the Tsallis statistics are given for $\gamma = 1.22$; see Mameli and Ventura (2015), and reference therein. Table 1 gives the results of the study based on 10,000 simulations with $B = 500$ bootstrap replications. The bootstrap statistics perform as the analytical counterparts in the case of the Tsallis and logarithm scoring rules. The same is not true for the Hyvärinen scoring rule both in the central and contaminated models. However, as the sample size increases the difference among the bootstrap and the analytical versions vanishes. The accuracy of the parametric bootstrap depends mainly on the choice of the estimate to use for generating samples. Indeed, parametric bootstrap of the Tsallis profile signed scoring rule root statistic $r_{Mp}^{T}(\psi)$ under the model $F(y; \hat{\theta}_{S\psi})$, provided B is large enough, yields an accurate parametric inference approach, bypassing any analytical computation. On the contrary, results, not shown here, indicated that the same accuracy is not retained when we sample from $F(y; \hat{\theta}_S)$. This is in agreement with the findings in the likelihood framework (Di Ciccio et al. (2001)).

Table 1 about here.

4.2 Non-Linear regression models

An useful extension of the classical regression model is obtained by replacing the linear predictor $x_i^T\beta$ by a known non-linear function $\mu(x_i, \beta)$, called the mean function. The model

$$y_i = \mu(x_i, \beta) + \sigma\varepsilon_i,$$

with $i = 1, \ldots, n$, is called a non-linear regression model, where x_i is a scalar and β is an unknown p-dimensional parameter, and ε_i are independent and identically distributed

$N(0,1)$ random variables.
The log-likelihood function for $\theta = (\beta, \sigma)$ is

$$\ell(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}[y_i - \mu_i(x_i,\beta)]^2. \tag{11}$$

The Hyvärinen score is

$$S_H(\theta) = -\frac{2n}{\sigma^2} + \frac{1}{\sigma^4}\sum_{i=1}^{n}[y_i - \mu_i(x_i,\beta)]^2. \tag{12}$$

The Tsallis score is

$$S_T(\theta) = \sum_{i=1}^{n}\left[-\gamma\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^{(\gamma-1)}\exp\left(-\frac{(\gamma-1)}{2\sigma^2}(y_i - \mu_i(x_i,\beta)^2\right) + \frac{(\gamma-1)}{\sqrt{\gamma}(2\pi\sigma^2)^{\frac{(\gamma-1)}{2}}}\right].$$

Let us consider an application to the calcium data frame (Rawlings (1988)). In this example, we consider a non-linear regression model applied to the calcium data set reported in Rawlings (1988), which is available in the `R`-package `boot`. This data set consists of $n = 27$ observations. The variable of interest is `cal`, which records the amount of radioactive calcium absorbed by cells, there is an explanatory variable (`time`) representing the time in minutes that cells were suspended in a radioactive calcium solution. As in Brazzale (2005), we fit an homoskedastic non linear model

$$y_i = \beta_0(1 - \exp(-\beta_1 x_i)) + \sigma\epsilon_i,$$

with $i = 1, \cdots, n$. In this example the mean function is $\mu(x,\beta) = \beta_0(1 - \exp(-\beta_1 x))$, and β_0 and β_1 are unknown regression coefficients. The parameter of interest is $\psi = \beta_1$ and the nuisance parameter is $\lambda = (\beta_0, \sigma)$.
To assess the reliability of the confidence intervals for the parameter ψ, we ran a simulation with 10.000 replicate data sets generated from the fitted model obtained by maximum likelihood estimation ($\hat{\beta}_0 = 4.309, \hat{\beta}_1 = 0.208, \hat{\sigma} = 0.526$). Empirical coverage levels of 95% confidence intervals for $\psi = \beta_1$ are reported in Table 2. Results of the Tsallis statistics are given for $\gamma = 1.22$. Table 2 reveals that the parametric bootstrap Tsallis modified profile signed scoring rule root performs similarly to the parametric bootstrap higher-order signed profile likelihood root. The higher-order signed profile likelihood root $r_p^*(\psi)$, which as expected gives approximately the same value of its bootstrap counterpart $r_{pb}^*(\psi)$ which is obtained by using the `nlreg` `R`-package Brazzale (2005).

Table 2 about here.

A sensitivity analysis was conducted to illustrate the instability of the maximum likelihood estimator and of the p-value associated with the parametric bootstrap higher-order

signed profile likelihood root test of the null hypothesis $H_0 : \beta_1 = 0.208$. It consists in replacing one observation of the amount of radioactive calcium absorbed by cells by an arbitrary value which varies from 0 to 10 in the case of the estimators and from 0 to 20 in the case of the p-values. The effect on the values of the estimators is shown in Fig. 1 and the effect on the p-values is shown in Fig. 2. It can be noted that the Tsallis estimator behaves like a redescending estimator. Specifically, Fig. 2 shows the p-values for the parametric bootstrap Tsallis modified profile signed scoring rule root, the parametric bootstrap Hyvärinen modified profile signed scoring rule root and the parametric bootstrap higher-order signed profile likelihood root tests. The variation in the parametric bootstrap higher-order signed profile likelihood root and Hyvärinen root tests' p-value shows their non-robustness.

Figures 1 and 2 about here.

5 Final remarks

Inferential procedures based on scoring rules offer flexibility in simplifying computations and in achieving robustness. However, these strengths may be compromised when resorting to first-order methods or to higher-order analytical approximations. Analytical adjustments of the signed scoring rules ratio statistics are easy to compute only in simple models. We face these challenges by resorting to the application of bootstrap methods. Constrained parametric bootstrap is a generally applicable and accurate inferential procedure which bypasses any analytical computation. The only requirement is the capacity to simulate from the full model. Unfortunately however, the generality of the procedure is at the expense of an increased computational burden.

Acknowledgements

We are grateful to the editor and to the two anonymous referees for many useful comments that greatly improved the paper. This work was partially supported by a grant from the University of Cagliari (Progetto di Ricerca Fondamentale o di Base 2012 L.R. 7/2007 annualità 2012), and a by grant from MIUR, 2015, Italy.

References

Aerts, M., Claeskens, G. (1999). Bootstrapping pseudolikelihood models for clustered binary data. Annals of the Institute of Statistical Mathematics 51: 515–530.

Almeida, M.P., Gidas, B. (1993). A variational method for estimating the parameters of MRF from complete or incomplete data. Annals of Applied Probability 3: 103–136.

Basu, A., Harris, I.R., Hjort, N.L., Jones, M.C. (1998). Robust and efficient estimation by minimising a density power divergence. Biometrika 85: 549–559.

Beran, R. (1987). Prepivoting to reduce error level of confidence sets. Biometrika 74: 457–468.

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. Journal of the American Statistical Association 83: 687–697.

Bernardo, J.M. (1979). Expected information as expected utility. Annals of Statistics 7: 686–690.

Brazzale, A.R. (2005). hoa: An R package bundle for higher order likelihood inference. *Rnews*, 5/1 May 2005, 20–27. ISSN 609-3631.
`https://cran.r-project.org/web/packages/cond/vignettes/Rnews-paper.pdf`

Brazzale, A., Davison, A., Reid, N. (2007). *Applied Asymptotics: Case Studies in Small-samples Statistics.* Cambridge University Press.

Brazzale, A.R., Davison, A.C. (2008). Accurate parametric inference for small samples. Statistical Science 23: 465–484.

Cattelan, M., Sartori, N. (2016). Empirical and simulated adjustments of composite likelihood ratio statistics. Journal of Statistical Computation and Simulation 86(5): 1056–1067.

Davison, A.C., Hinkley, D.V. (1997). *Bootstrap Methods and their Application.* Cambridge University Press.

Dawid, A.P., Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. The Annals of Statistics 27: 65–81.

Dawid, A.P., Lauritzen, S.L. (2005). The geometry of decision theory. In: *Proceedings of the Second International Symposium on Information Geometry and its Applications*, University of Tokyo, 22–28.

Dawid, A.P. (2007). The geometry of proper scoring rules. Annals of the Institute of Statistical Mathematics 59: 77–93.

Dawid, A.P., Musio, M. (2014). Theory and applications of proper scoring rules. Metron 72: 169–183.

Dawid, A.P., Musio, M., Ventura, L. (2016). Minimum scoring rule inference. Scandinavian Journal of Statistics 43(1): 123–138.

Di Ciccio, T.J., Martin, M.A., Stern, S.E. (2001). Simple and accurate one-sided inference from signed roots of likelihood ratios. The Canadian Journal of Statistics 29: 67–79.

Ehm, W., Gneiting, T. (2012). Local proper scoring rules of order two. The Annals of Statistics 40(1): 609–637.

Gatto, R., Ronchetti, E. (1996). General saddlepoint approximations of marginal densities and tail probabilities. Journal of the American Statistical Society 91(434): 666–673.

Godambe, V.P. (1960). An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics 31: 1208–1211.

Gneiting, T., Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102: 359–378.

Good, I.J. (1952). Rational decisions. Journal of the Royal Statistical Society, Series B 14: 107–114.

Hyvärinen, A. (2005). Estimation of non-normalized statistical models by score matching. Journal of Machine Learning Research 6: 695–709.

Huber, P.J., Ronchetti, E.M. (2009). *Robust Statistics*. John Wiley and Sons, Inc.

La Vecchia, D. (2016). Stable asymptotics for M-estimators. International Statistical Review 84(2): 267–290.

Lee, S., Young, A. (2003). Prepivoting by weighted bootstrap iteration. Biometrika 69: 19–27.

Mameli, V., Ventura, L. (2015). Higher-order asymptotics for scoring rules. Journal of Statistical Planning and Inference 165: 13–26.

Marin, J.-M., Pudlo, P., Robert, C.P., Ryder, R.J. (2012). Approximate Bayesian computational methods. Statistics and Computing 22: 1167–1180.

Parry, M., Dawid, A.P., Lauritzen, S.L. (2012). Proper local scoring rules. The Annals of Statistics 40: 561–592.

Rawlings, J.O. (1988). *Applied Regression Analysis*. Wadsworth and Brooks/Cole Statistics/Probability Series.

Stern, S.E. (2006). Simple and accurate one-sided inference based on a class of m-estimators. Biometrika 93: 973–987.

Tsallis, C. (1988). Possible generalization of Boltzmann-Gibbs statistics. Journal of Statistical Physics 52: 479–487.

Varin, C., Reid, N., Firth, D. (2011). An overview of composite likelihood methods. Statistica Sinica 21: 5–42.

Young, G.A. (2009). Routes to higher-order accuracy in parametric inference. Australian & New Zealand Journal of Statistics 51: 115–126.

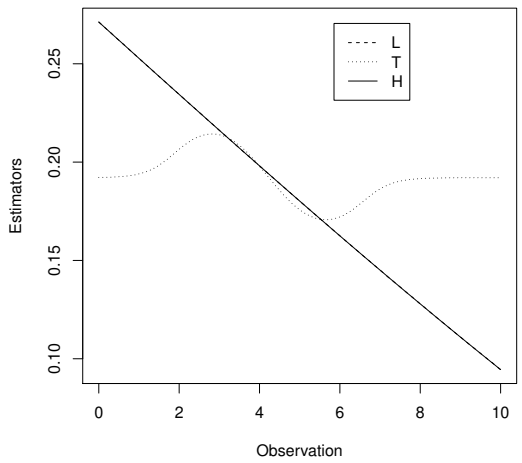

Figure 1: Sensitivity analysis of the Maximum likelihood (L), Tsallis (T) and Hyvärinen (H) estimators when an observation of the amount of radioactive calcium absorbed by cells is changed.

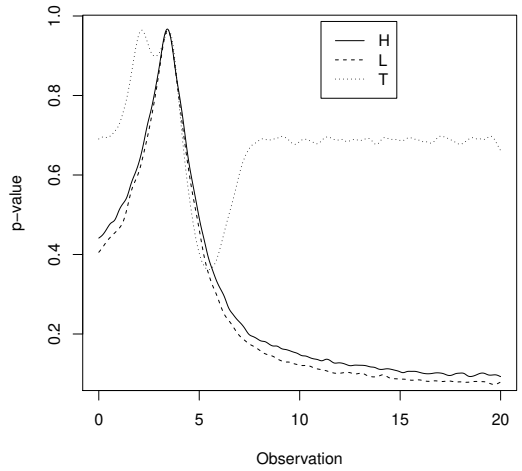

Figure 2: Sensitivity analysis of the p-values for the parametric bootstrap higher-order signed profile likelihood root (L), the bootstrap Tsallis modified profile signed scoring rule root (T) and the parametric bootstrap Hyvärinen modified profile signed scoring rule root (H) tests when an observation of the amount of radioactive calcium absorbed by cells is changed.

Table 1: Empirical coverages of 95% confidence intervals for β_2. Pivots used: higher-order signed profile likelihood root (r_p^*), parametric bootstrap higher-order signed profile likelihood root (r_{pb}^*), Tsallis modified profile signed scoring rule root (r_{Mp}^T) and parametric bootstrap Tsallis modified profile signed scoring rule root (r_{Mp}^{Tb}) with $\gamma = 1.22$. $B = 500$ bootstrap replications.

			$N(0,1)$			
n	r_p^*	r_{pb}^*	r_{Mp}^H	r_{Mp}^{Hb}	r_{Mp}^T	r_{Mp}^{Tb}
10	0.951	0.949	0.946	0.976	0.948	0.955
20	0.946	0.948	0.942	0.966	0.947	0.951
30	0.948	0.950	0.946	0.961	0.945	0.948
			$N(0,1)$ cont.			
n	r_p^*	r_{pb}^*	r_{Mp}^H	r_{Mp}^{Hb}	r_{Mp}^T	r_{Mp}^{Tb}
10	0.955	0.952	0.982	0.974	0.943	0.946
20	0.965	0.960	0.994	0.979	0.949	0.951
30	0.967	0.967	0.998	0.958	0.951	0.953

Table 2: Empirical coverages of 95% confidence intervals for β_1. Pivots used: higher-order signed profile likelihood root (r_{pb}^*), parametric bootstrap Hyvärinen modified profile signed scoring rule root (r_{Mp}^{Hb}) and parametric bootstrap Tsallis modified profile signed scoring rule root (r_{Mp}^{Tb}) with $\gamma = 1.22$. $B = 500$ bootstrap replications.

	$N(0,1)$			
n	r_p^*	r_{pb}^*	r_{Mp}^{Tb}	r_{Mp}^{Hb}
27	0.951	0.951	0.949	0.958
54	0.953	0.954	0.956	0.956