Simulated adjustment of the signed scoring rule root statistic

V. Mameli\textsuperscript{1}, M. Musio\textsuperscript{2}, L. Ventura\textsuperscript{3}

\textsuperscript{1} Dept. of Environmental Sciences, Informatics and Statistics, Ca’ Foscari University of Venice, Italy
\textsuperscript{2} Dept. of Mathematics and Computer Science, University of Cagliari, Italy
\textsuperscript{3} Dept. of Statistical Sciences, University of Padova, Italy

E-mail for correspondence: mmusio@unica.it

Abstract: We focus on adjustments of the signed scoring rule root statistic generalizing results for likelihood quantities. In particular, for a scalar parameter of interest, we investigate a bootstrap adjustment of the signed scoring rule root statistic. An example is discussed.

Keywords: Asymptotic expansions; Bootstrap; Higher-order inference; Tsallis scoring rule.

1 Introduction

A scoring rule is a loss function \( S(x, Q) \) measuring the quality of a quoted probability distribution \( Q \) for the random variable \( X \), in the light of the realized outcome \( x \) of \( X \). It is proper if, for any distribution \( P \) for \( X \), the expected score \( S(P, Q) := E_{X \sim P} S(X, Q) \) is minimized by quoting \( Q = P \). There is a wide variety of proper scoring rules (see, e.g., Dawid and Musio, 2014, and references therein). A prominent example is the log-score \( S(x, Q) = -\log q(x) \), with \( q(\cdot) \) the density (or the probability mass function) of \( X \). Another useful example is the Tsallis score, given by

\[
S(x, Q) = (\gamma - 1) \int q(y) \gamma \, d\mu(y) - \gamma q(x)^{\gamma - 1}, \quad \text{with } \gamma > 1, \quad (1)
\]

also called the density power score (Basu et al., 1998). Other examples of special proper scoring rules are given in Dawid and Musio (2014). Proper scoring rules, different from the log-score, can be used as an alternative to the full likelihood, when the aim is to increase the robustness or to simplify
computations. For instance, Dawid et al. (2014) give sufficient conditions that guarantee the robustness of the estimator based on (1).

Proper scoring rule inference is usually based on the first-order approximations to the distribution of the scoring rule estimator or of the scoring rule ratio test statistic. However, several examples (see e.g. Dawid et al., 2014, Mameli and Ventura, 2015, and reference therein) illustrate the inaccuracy of first-order methods, even in models with a scalar parameter, when the sample size is small or moderate. For more accurate inference refinements can be considered to improve the first-order approximations.

Analytical higher-order asymptotic expansions for proper scoring rules, generalizing results for likelihood quantities but allowing for the failure of the information identity, have been discussed in Mameli and Ventura (2015). However, the calculation of the quantities involved in the analytical adjustments of the signed and signed profile scoring rule root statistic is cumbersome when the dimension of the parameter (or of the nuisance parameter) is large, even for simple models.

Paralleling results for likelihood statistics (see, e.g., Young, 2009), the aim of this paper is to discuss the alternative approach to higher-order adjustments, based on a parametric bootstrap. In particular, focus is on the signed profile scoring rule root statistic.

2 Background on first-order inference

Suppose that we wish to fit a parametric statistical model \( F_\theta = F(x; \theta) \), with \( \theta \in \Theta \subseteq \mathbb{R}^k \), based on the random sample \((x_1, \ldots, x_n)\).

To estimate \( \theta \), we might consider the goodness-of-fit by the total empirical score \( S(\theta) = \sum_{i=1}^n S(x_i, F_\theta) \). Asymptotic arguments indicate that \( \hat{\theta}_S = \arg \min_\theta S(\theta) \to \theta_0 \) as \( n \to \infty \), where \( \theta_0 \) is the true parameter value.

Maximum likelihood estimation, as well as composite likelihood estimation, are special cases of score estimation when \( S(\theta) \) is the negative log-likelihood (Dawid and Musio, 2014).

Score estimation forms a special case of M-estimation (Huber and Ronchetti, 2009). Indeed, let \( s(x, \theta) = \partial S(x, F_\theta)/\partial \theta \). Under suitable regularity conditions on the scoring rule and on the statistical model, \( \theta \) can be estimated by \( \hat{\theta}_S \), the root of the estimating equation \( s(\theta) = \sum_{i=1}^n s(x_i, \theta) = 0 \), which is an unbiased estimating function. In particular, \( \hat{\theta}_S \) is consistent and asymptotically normal with mean \( \theta \) and variance \( V(\theta) = K(\theta)^{-1}J(\theta)K(\theta)^{-1} \), with \( J(\theta) = E_\theta(s(\theta)s(\theta)^T) \) and \( K(\theta) = E_\theta(\partial s(\theta)/\partial \theta^T) \). The form of \( V(\theta) \) is due to the failure of the information identity since, in general \( K(\theta) \neq J(\theta) \). In view of this, the asymptotic distribution of the scoring rule ratio statistic \( W^S(\theta) = 2\{S(\theta) - S(\hat{\theta}_S)\} \) departs from the familiar likelihood result, and involves a linear combination of independent chi-square random variables whose coefficients are the eigenvalues of \( J(\theta)K(\theta)^{-1} \) (Dawid et al., 2014).

When \( \theta \) is partitioned as \( \theta = (\psi, \lambda) \), where \( \psi \) is a scalar parameter of interest and \( \lambda \) is a \((k - 1)\)-dimensional nuisance parameter, the profile scoring
rule ratio statistic for $\psi$ is given by $W^S_p(\psi) = 2(S(\hat{\theta}_{S\psi}) - S(\hat{\theta}_S))$, where $\hat{\theta}_{S\psi} = (\psi, \hat{\lambda}_{S\psi})$ represents the constrained score estimate. The asymptotic distribution of $W^S_p(\psi)$ is a linear combination of independent chi-square random variables (see Dawid et al., 2014). Also the asymptotic distribution of the profile signed scoring rule root statistic

$$r^S_p(\psi) = \text{sgn}(\hat{\psi}_S - \psi) \sqrt{W^S_p(\psi)}$$

(2)

departs from the familiar likelihood result. A simple adjustment of (2), which recovers normality, is $r^S_{p,adj}(\psi) = \mu_1(\psi)^{-1/2} r^S_p(\psi)$, with $\mu_1(\psi) = [G^{\psi \psi}(\hat{\theta}_{S\psi})^{-1} H^{\psi \psi}(\hat{\theta}_{S\psi})]^{-1}$, where $G^{\psi \psi}(\theta)$ and $H^{\psi \psi}(\theta)$ are sub-matrices of the inverses of $G(\theta)$ and $H(\theta)$, respectively.

3 Adjustments of $r^S_p(\psi)$

Let us focus on the profile signed scoring rule root statistic (2) for a scalar parameter of interest. As for likelihood quantities, theory in Mameli and Ventura (2015) shows that a suitable centering and scaling of $r^S_p(\psi)$, i.e.

$$r^S_{p,M}(\psi) = \frac{r^S_p(\psi) - m(\psi)}{\sqrt{\mu_1(\psi) + v(\psi)}}$$

(3)

improves the accuracy of the asymptotic normal approximation to the distribution of $r^S_{p,adj}(\psi)$. In (3) we have that $m(\psi)$ is of order $O(n^{-1/2})$ and $v(\psi)$ is of order $O(n^{-1})$. The analytical expressions of $m(\psi)$ and $v(\psi)$ are derived in Mameli and Ventura (2015) and they involve several expected values of scoring rules derivatives. However, the analytical calculations of $m(\psi)$ and $v(\psi)$ are cumbersome as the dimension of the nuisance parameter is large, even for simple models.

Here we exploit a parametric bootstrap approach in order to compute (3). The idea is, for a fixed $\psi$, to draw $B$ samples from the distribution $F(x; \hat{\theta}_{S\psi})$ and compute (3) using the bootstrap mean and variance of $r^S_p(\psi)$. In the classical likelihood approach, the parametric bootstrap provides a $O(n^{-3/2})$ order of accuracy, and the resulting approximation is sometimes called constrained pre-pivoting of the signed likelihood root statistic (see DiCiccio et al., 2001).

Example: Let us consider the linear regression model $y = X\beta + \sigma\epsilon$, where $X$ is a $n \times p$ fixed matrix, $\beta \in \mathbb{R}^p$ ($p \geq 1$), $\sigma = 1$, and $\epsilon$ an $n$-dimensional Gaussian vector. Let $\psi = \beta_2$ be the scalar parameter of interest, and let $\lambda = (\beta_1, \beta_3)$ be the nuisance parameter. We ran a simulation experiment, for $n = 10, 20$ and $\beta = (1, 2, 3)$, in order to assess the accuracy of the parametric bootstrap Tsallis modified profile signed scoring rule root statistic
Simulated adjustment of signed scoring rule root statistic \( (r^T_{pMb}(\beta_2)) \). Note that the Tsallis score estimator is \( B \)-robust since the influence function is bounded (see Dawid et al., 2014 and Mameli and Ventura, 2015). Table 1 gives the results of the study based on 10,000 simulations with \( B = 500 \) bootstrap replications. We note that 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 \( r^T_{pM}(\psi) \) under the model \( F(y; \theta_{S\psi}) \), 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 \( F(y; \hat{\theta}_S) \). Note that also in the likelihood framework \( \theta_{S\psi} \) is the best choice to providing accurate inference (Di Ciccio et al., 2001).

TABLE 1. Empirical coverages of 95\% confidence intervals for \( \beta_2 \). Pivots used: parametric bootstrap higher-order signed profile likelihood root \( (r^*_{pb}) \) and parametric bootstrap Tsallis modified profile signed scoring rule root \( (r^T_{pMb}) \) with \( \gamma = 1.2 \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( r^*_{pb} )</th>
<th>( r^T_{pMb} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.9498</td>
<td>0.9500</td>
</tr>
<tr>
<td>20</td>
<td>0.9536</td>
<td>0.9547</td>
</tr>
</tbody>
</table>

References


