The Version of Record of this Manuscript has been published and is available in Journal of Statistical Planning and Inference (ISSN: 0378-3758) from the 21 March 2015 on https://www.sciencedirect.com/science/article/abs/pii/S0378375815000567.

https://doi.org/10.1016/j.jspi.2015.03.005

Higher-order asymptotics for scoring rules

Valentina Mameli[1,*], Laura Ventura[1]

[a]*Department of Mathematics and Computer Science, University of Cagliari, Italy*
[b]*Department of Statistical Sciences, University of Padova, Italy*

Abstract

In this paper we discuss higher-order asymptotic expansions for proper scoring rules generalizing results for likelihood quantities, but meanwhile bring in the difficulty caused by the failure of the information identity. In particular, we derive higher-order approximations to the distribution of the scoring rule estimator, of the scoring rule ratio test statistic and, for a scalar parameter of interest, of the signed scoring rule root statistic. From these expansions, a modified signed scoring rule root statistic is proposed. Examples are given illustrating the accuracy of the modified signed scoring rule root statistic with respect to first-order methods.

Keywords: Asymptotic expansions, Hyvärinen scoring rule, Information identity, Likelihood asymptotics, Robustness, Third-order inference, Tsallis scoring rule.

---

1. Introduction

The theory of higher-order asymptotics for likelihood-based procedures provides very accurate inferences in a variety of parametric statistical problems (see, e.g., [**?** **?**], and references therein). In this paper we address the issue of higher-order asymptotics for proper scoring rules.

A scoring rule (see, for instance, the recent overviews by [**?**] and [**?**], and references therein) is a special kind of loss function designed to measure the quality of a probability distribution for a random variable, given its observed

*Corresponding author

value. It is proper if it encourages honesty in the probability evaluation. Proper scoring rules supply unbiased estimating equations for any statistical model, which can be chosen to increase robustness or for ease of computation. The appeal of scoring rule inference lies in the potential adaptation of the scoring rule to the particular application [**?**], and it forms a special case of M-estimation (see e.g. [**?**]). Among the most important proper scoring rules are the Brier score [**?**], the logarithmic score [**?**], the Tsallis score [**?**], and the Hyvärinen score [**?**]. In particular, when using the logarithmic score, the full likelihood and the composite likelihood [**?**] are obtained as special cases of proper scoring rules (see for instance **?**]).

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 illustrate the inaccuracy of first-order methods, even in models with a scalar parameter, when the sample size is small or moderate. In particular, numerical studies in [**?**] show that the Wald-type statistic based on the Tsallis scoring rule provides unsatisfactory results in terms of actual coverage of confidence intervals. Simulation studies in **?**], conducted to compare the Hyvärinen estimator [**?**] to the full and the pairwise maximum likelihood estimators in the first order autoregressive model, reveal that these estimates are, on average, close to the true values. However, a loss of efficiency is present especially near to boundaries of the parameter space. For more accurate inference refinements can be considered to improve the first-order approximations.

In this paper we discuss higher-order asymptotic expansions for proper scoring rules generalizing results for likelihood quantities, but allowing for the failure of the information identity. In particular, we derive higher-order approximations to the distribution of the scoring rule estimator, of the scoring rule ratio test statistic, and of the signed scoring rule root statistic for a scalar parameter. From these expansions, a modified signed scoring rule root statistic is proposed. Simulation studies are conducted to compare different procedures based on the Tsallis and the Hyvärinen scoring rules. While the former gives rise to estimators that may be robust [**?**], the later appears particularly attractive for inference in the natural exponential family [**?**]. The examples presented show that the proposed modified signed scoring rule root statistic performs extremely well even for small sample sizes and outperforms the first-order scoring rule statistics.

The paper unfolds as follows. In Section 2 we introduce scoring rules and we review first-order inference for proper scoring rules. We develop higher-order expansions in Section 3. Simulation studies are presented in Section 4. We conclude with a discussion in Section 5.

2. Background on scoring rules

For notations and background on scoring rules we refer to [**?**]. Consider a random variable X on the sample space $\mathcal{X}$ and a class of distributions $\mathcal{P}$ on $\mathcal{X}$. A scoring rule $S^*(x,Q)$ is a loss function, from $\mathcal{X} \times \mathcal{P}$ to $(-\infty, \infty]$, which measures the quality of a probability distribution Q for the random variable X given its observed result x. The expected value of $S^*(x,Q)$ under P is denoted by $S^*(P,Q)$.

A scoring rule is proper relative to the class $\mathcal{P}$ if

$$S^*(P,Q) \geq S^*(P,P), \quad \text{for any distribution} \quad P, Q \in \mathcal{P}. \tag{1}$$

It is strictly proper when the equality is achieved only at $Q = P$. The validity of this property depends on the class of distributions $\mathcal{P}$ under consideration.

There exists a great variety of scoring rules (for examples see, e.g., [**?**], [**?**] and [**?**], and references therein). One of the well-known scoring rules is the logarithmic scoring rule $S^*(x,Q) = -\log q(x)$ [**?**], with $q(\cdot)$ the density function of Q with respect to the measure μ. This scoring rule is just the negative log-likelihood. Another prominent example is the Hyvärinen score [**?**], which is defined as

$$S_H^*(x,Q) = 2\frac{\partial^2 \log q(x)}{\partial x^2} + \left(\frac{\partial \log q(x)}{\partial x}\right)^2 \;, \; x \in \mathbb{R}. \tag{2}$$

The Hyvärinen score is a special case of homogeneous local proper scoring rule [**?**]. It was subsequently generalised to the case of a Riemannian manifold in

[**?**] and to the case of non-negative and positive real domains, and for binary variables, in [**?**]. Another scoring rule of particular interest is the Tsallis score [**?**], which is a special case of the Bregman score [**?**]. The Tsallis scoring rule is given by

$$S^*_T(x,Q) = (\gamma-1)\int q(y)^\gamma d\mu(y) - \gamma q(x)^{\gamma-1}\ ,\ x \in \mathbb{R},\ \gamma > 1. \tag{3}$$

Estimators based on (**??**) are minimum density power divergence estimators [**?**]. Values of the parameter γ near 1 can give rise to robust estimators with negligible loss of efficiency with respect to the maximum likelihood estimator (see [**?**]).

Proper scoring rules can also be extended to the case of a random vector. Let $\{X_k\}$ be a set of marginal or conditional variables with associated proper scoring rule S^*_k. A proper scoring rule for the random vector X is defined as

$$S^*(\mathbf{x},Q) = \sum_k S^*_k(\mathbf{x}_k, Q_k)\ , \tag{4}$$

where $X_k \sim Q_k$ when $X \sim Q$, and $\mathbf{x}$ and $\mathbf{x}_k$ are the values assumed by X and X_k, respectively. Scoring rules of the form (**??**) are called composite scoring rules; see **?**] and [**?**]. Note that when each S^*_k is the logarithmic score, equation (**??**) is a negative composite log-likelihood; see **?**].

2.1. First-order inference for scoring rules

Let X be a random variable taking values in a sample space $\mathcal{X}$ and $\mathcal{Q} = \{Q_\theta : \theta \in \Theta \subseteq \mathbb{R}^k\} \subseteq \mathcal{P}$, with $k \geq 1$, be a parametric family of distributions on $\mathcal{X}$. Let $q(x;\theta)$ denote the probability density function of Q_θ. Given a proper scoring rule $S^*(x,\theta)$ on $\mathcal{X}$, hereafter we will implicitly identify $S^*(x,Q_\theta)$ with $S^*(x,\theta)$. Since any positive affine transformation of a proper scoring rule is again a proper scoring rule, in the rest of the paper we consider $S(x,\theta) = -S^*(x,\theta)$. This choice is adopted to be consistent with the results for likelihood quantities.

Let $(X_1,...,X_n)$ be a sequence of independent and identical distributed random variables with distribution function Q_θ, with $\theta \in \Theta$. Let θ_0 denote the true value of the parameter θ. To estimate θ we can consider the value $\widehat{\theta}_S$ which maximizes the total empirical score

$$S(\theta) = \sum_{i=1}^{n} S(x_i, \theta), \tag{5}$$

where $(x_1, ..., x_n)$ is the observed sample. In what follows, we assume the following conditions (for further details see **?**], **?**], **?**], **?**]).

Condition 1 Let Θ be an open set and suppose that the support of X does not depend on θ. Furthermore suppose that the operations of integration and differentiation with respect to θ may be exchanged and all the components of θ are identifiable.

Condition 2 $S(x, \theta)$ is a function absolutely continuous in θ and continuously differentiable in a neighborhood of θ_0. Suppose that $E_{\theta_0}(S(X, \theta))$ exists for all $\theta \in \Theta$ and has a unique maximum at $\theta = \theta_0$, with θ_0 inner point of Θ.

Condition 3 $s_\theta(x, \theta) = \frac{\partial S(x,\theta)}{\partial \theta^T}$ is assumed to be absolutely continuous in θ with derivative $s_{\theta\theta}(x, \theta) = \frac{\partial^2 S(x,\theta)}{\partial\theta\partial\theta^T}$ such that $E_{\theta_0}(s_{\theta\theta}(X, \theta_0))^2$ is finite. Let $s_{\theta\theta\theta}(x, \theta)$ be the third derivative of $S(x, \theta)$. Suppose, further, that the absolute value of $s_{\theta\theta\theta}(X, \theta_0)$ is dominated by an integrable function for every θ in a neighborhood of θ_0.

Condition 4 The matrices $J(\theta_0) = E_{\theta_0}\left(s_\theta(X, \theta_0)s_\theta(X, \theta_0)^T\right)$ and $K(\theta_0) = -E_{\theta_0}(s_{\theta\theta}(X, \theta_0))$ are positive definite and non singular, respectively.

Note that *Condition 2* ensures that $\widehat{\theta}_S$ is the solution of the scoring rule estimating equation

$$s_\theta(\theta) = \sum_{i=1}^{n} s_\theta(x_i, \theta) = 0. \tag{6}$$

It can be proved that the scoring rule estimating function (**??**) is unbiased for any proper scoring rule (see [**?**], [**?**], [**?**]). When $S(x, \theta)$ is minus the logarithmic score, the scoring rule estimating equation (**??**) is just the likelihood equation, and the scoring rule estimator $\widehat{\theta}_S$ is just the maximum likelihood estimator (MLE).

Under *Conditions 1-4*, the scoring rule estimator $\widehat{\theta}_S$ is asymptotically normal, with mean θ_0 and variance $\frac{1}{n}V(\theta_0)$, where

$$V(\theta_0) = K(\theta_0)^{-1}J(\theta_0)(K(\theta_0)^{-1})^T \ .$$

The matrix $G(\theta) = V(\theta)^{-1}$ is known as the Godambe information matrix [**?**]. The form of $V(\theta)$ is due to the failure of the second Bartlett identity since, in general, $K(\theta) \neq J(\theta)$. In the special case of the logarithmic score, we have that $G(\theta) = K(\theta) = J(\theta)$ is the Fisher information matrix for a single observation. From the general theory of M-estimators, the influence function (IF) of the estimator $\widehat{\theta}_S$ is given by

$$IF_S(x;\theta) = K(\theta)^{-1}s_\theta(x,\theta) \ , \tag{7}$$

and it measures the effect on the estimator of an infinitesimal contamination at the point x, standardized by the mass of the contamination. The estimator $\widehat{\theta}_S$ is B-robust if and only if $s_\theta(x,\theta)$ is bounded in x, for each θ. Robustness properties of the Tsallis score are discussed in [**?**] and [**?**].

Hypothesis testing and confidence regions for θ can be formed in the usual way by using a consistent estimate of the asymptotic variance $\frac{1}{n}V(\theta)$. In particular, inference for θ can be based on the scoring rule Wald-type statistic

$$W_w^S(\theta) = n(\widehat{\theta}_S - \theta)^T V(\widehat{\theta}_S)^{-1}(\widehat{\theta}_S - \theta) \ , \tag{8}$$

which has an asymptotic χ^2_k distribution; see [**?**]. Standard error evaluation requires consistent estimation of the matrices $J(\theta)$ and $K(\theta)$. If n is large, these matrices may be empirically estimated by

$$\hat{J} = \frac{1}{n}\sum_{i=1}^{n} s(x_i,\widehat{\theta}_S)s(x_i,\widehat{\theta}_S)^T , \qquad \widehat{K} = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial s(x_i,\theta)}{\partial\theta}\Big|_{\theta=\widehat{\theta}_S} \ .$$

Moreover, when it is possible to simulate from the model, the matrices $J(\theta)$ and $K(\theta)$ can be estimated through Monte Carlo samples; see, for instance, **?**] and **?**] for a detailed account on the estimation of the two matrices under the composite likelihood setting. The asymptotic χ^2_k distributional result holds also for the scoring rule score-type statistic $W^S_s(\theta) = n^{-1} s_\theta(\theta)^T J(\widehat{\theta}_S)^{-1} s_\theta(\theta)$; see **?**].

On the contrary, the asymptotic distribution of the scoring rule ratio statistic

$$W^S(\theta) = 2\left\{S(\widehat{\theta}_S) - S(\theta)\right\} \tag{9}$$

departs from the familiar likelihood result, and involves a linear combination of independent chi-square random variables with coefficients given by the eigenvalues of a matrix related to Godambe information [**?**]. More precisely,

$$W^S(\theta) \xrightarrow{L} \sum_{j=1}^{k} \mu_j Z_j^2 \ ,$$

where $\mu_1, \ldots, \mu_k$ are the eigenvalues of $J(\theta)K(\theta)^{-1}$ and $Z_1, \ldots, Z_k$ are independent standard normal variates.

Since the asymptotic null distribution of (**??**) depends both on the statistical model and on the parameter of interest, adjustments to $W^S(\theta)$ are of interest aiming at an asymptotic null distribution that depends only on the dimension of the parameter of interest. For instance, in the scalar parameter case, i.e. for $k = 1$, the adjusted scoring rule root statistic satisfies

$$W^S_{adj}(\theta) = \frac{W^S(\theta)}{\mu_1} \xrightarrow{L} \chi^2_1 \ ,$$

where

$$\mu_1 = \mu_1(\theta) = J(\theta)K(\theta)^{-1}. \tag{10}$$

Similarly, the adjusted signed scoring rule root statistic

$$r^S_{adj}(\theta) = \text{sgn}(\widehat{\theta}_S - \theta)\sqrt{W^S_{adj}(\theta)} \xrightarrow{L} N(0, 1) \ . \tag{11}$$

For $k > 1$, adjustments of $W^S(\theta)$ are discussed in [**?**].

Note that analogous limiting results can be shown to hold when θ is 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 (see, e.g., [**?**]). For instance, the Wald-type statistic for ψ is given by

$$W^S_{wp}(\psi) = n(\widehat{\psi}_S - \psi)^T (G^{\psi\psi}(\widehat{\theta}_S))^{-1} (\widehat{\psi}_S - \psi),$$

where $\widehat{\theta}_S = (\widehat{\psi}_S, \widehat{\lambda}_S)$ and its asymptotic null distribution is $\chi^2_{k_0}$. Moreover, we can define the profile scoring rule ratio statistic for ψ as

$$W_p^S(\psi) = 2\left(S(\widehat{\theta}_S) - S(\widehat{\theta}_{S\psi})\right), \tag{12}$$

where $\widehat{\theta}_{S\psi} = (\psi, \widehat{\lambda}_{S\psi})$ represents the constrained score estimate. The asymptotic distribution of (**??**) is $\sum_{j=1}^{k_0} \nu_j Z_j^2$, where $\nu_1, \ldots, \nu_{k_0}$ are the eigenvalues of the matrix $(K^{\psi\psi})^{-1}G^{\psi\psi}$, with $K^{\psi\psi}$ and $G^{\psi\psi}$ sub-matrices of K^{-1} and G^{-1} with respect to ψ, respectively. Moreover, $Z_1, \ldots, Z_{k_0}$ are independent standard normal variates. Note that when $k_0 = 1$, an adjustment of $W_p^S(\psi)$, which recovers the usual χ^2_1 asymptotic distribution, is given by (see [**?**])

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

3. Higher-order expansions for scoring rules

In this section, we discuss higher-order asymptotic expansions for $(\widehat{\theta}_S - \theta)$, for $W^S(\theta)$ and, when $k = 1$, for the signed scoring rule root statistic

$$r^S(\theta) = \text{sgn}(\widehat{\theta}_S - \theta)\sqrt{W^S(\theta)}.$$

As for likelihood quantities, these expansions are the basis for the study of higher-order asymptotic properties. In particular, in this paper it is of interest to compute the modified signed scoring rule root statistic, given by

$$r_M^S(\theta) = \frac{r^S(\theta) - m(\theta)}{\sqrt{\mu_1(\theta) + v(\theta)}}, \tag{13}$$

where $m(\theta)$ is of order $O(n^{-1/2})$, $v(\theta)$ is of order $O(n^{-1})$ and $\mu_1(\theta)$ is given in (**??**). The mean and the variance corrections in (**??**) can improve the accuracy of the asymptotic normal approximation to the distribution of $r^S(\theta)$ (see [**?**], [**?**], [**?**]).

The expansions of the mean and the variance of $r^S(\theta)$ obtained in this section can be similarly derived for the signed profile scoring rule root statistic

$$r_p^S(\psi) = \text{sgn}(\widehat{\psi}_S - \psi)\sqrt{W_p^S(\psi)}, \tag{14}$$

where ψ is the scalar parameter of interest. In particular, an expansion for $W_p^S(\psi)$ can be obtained by writing $S(\widehat{\theta}_S) - S(\widehat{\theta}_{S\psi})$ as the difference between $S(\widehat{\theta}_S) - S(\theta)$ and $S(\widehat{\theta}_{S\psi}) - S(\theta)$, and expanding both these terms. Examples of (**??**) for a scalar parameter of interest and of the higher-order modification of (**??**), denoted by $r_{pM}^S(\psi)$, are discussed in Section **??**.

Let assume that the first four derivatives of the total empirical score in (**??**), $s_\theta(\theta)$, $s_{\theta\theta}(\theta)$, $s_{\theta\theta\theta}(\theta)$, $s_{\theta\theta\theta\theta}(\theta)$, have joint cumulants up to the fourth order which are of order $O(n)$ (see **?** , Section 3.4]).

In the following, it is convenient to suppress the dependence on θ for the derivatives of $s_\theta(\theta)$ and, when θ is multidimensional, to use index notation and Einstein's summation convention. The components of θ are denoted by θ^r, the corresponding components of s_θ are s_r, and the components of the derivatives of s_θ are denoted by s_{rs}, s_{rst}, ..., where the indices $r, s, t, \ldots$ range over $1, \ldots, k$. The expected values of these derivatives are $\nu_{rs} = E_\theta(s_{rs})$, $\nu_{rst} = E_\theta(s_{rst})$, ..., and these quantities are of order $O(n)$. Further, the zero-mean variables s_r, $H_{rs} = s_{rs} - \nu_{rs}$, ..., are of order $O_p(n^{1/2})$. These assumptions are satisfied since s_θ is a sum of n independent random variables. Finally, we use the notation $\nu_{r,s} = E_\theta(s_r s_s)$, $\nu_{rs,t} = E_\theta(s_{rs} s_t)$, ..., and we denote with $u_{rs} = -\nu_{rs}$ and with u^{rs} the inverse matrix of u_{rs}.

3.1. Expansion of $(\widehat{\theta}_S - \theta)$

Under regularity conditions, $s_r(\widehat{\theta}_S)$ can be Taylor expanded around θ. This gives

$$
\begin{aligned}
0 = s_r(\widehat{\theta}_S) &= s_r + (\widehat{\theta}_S - \theta)^s s_{rs} + \frac{1}{2}(\widehat{\theta}_S - \theta)^{st} s_{rst} \\
&+ \frac{1}{6}(\widehat{\theta}_S - \theta)^{stu} s_{rstu} + O_p(n^{-1}) \,, \qquad (15)
\end{aligned}
$$

where $(\widehat{\theta}_S - \theta)^{st} = (\widehat{\theta}_S - \theta)^s (\widehat{\theta}_S - \theta)^t$, and so on.

Let us denote the generic partial derivative of order m of $S(\theta)$ by $s_{Rm} = s_{r_1 \ldots\ldots r_m}$, and consider the decomposition $s_{Rm} = \nu_{Rm} + H_{Rm}$, where $\nu_{Rm} = O(n)$ and the centered variables $H_{Rm} = s_{Rm} - \nu_{Rm}$ are of order $O_p(n^{1/2})$, for

$m > 1$. Remembering that $\nu_{rs} = -u_{rs}$, then (**??**) can be re-written as

$$\begin{aligned} 0 &= s_r + (\widehat{\theta}_S - \theta)^s \, (-u_{rs} + H_{rs}) + \frac{1}{2}(\widehat{\theta}_S - \theta)^{st} \, (\nu_{rst} + H_{rst}) \\ &+ \frac{1}{6}(\widehat{\theta}_S - \theta)^{stu} \, (\nu_{rstu} + H_{rstu}) + O_p(n^{-1}) \ , \end{aligned}$$

which gives

$$\begin{aligned} u_{rs}(\widehat{\theta}_S - \theta)^s &= s_r + (\widehat{\theta}_S - \theta)^s H_{rs} + \frac{1}{2}(\widehat{\theta}_S - \theta)^{st} \, (\nu_{rst} + H_{rst}) \\ &+ \frac{1}{6}(\widehat{\theta}_S - \theta)^{stu} \, (\nu_{rstu} + H_{rstu}) + O_p(n^{-1}) \ . \end{aligned}$$

Now, it is possible to isolate $(\widehat{\theta}_S - \theta)^h$ on the left-hand side by multiplying both sides by u^{hr}. This gives

$$\begin{aligned} (\widehat{\theta}_S - \theta)^h &= s^h + H^h_s(\widehat{\theta}_S - \theta)^s + \frac{1}{2}\nu^h_{st}(\widehat{\theta}_S - \theta)^{st} + \frac{1}{2}H^h_{st}(\widehat{\theta}_S - \theta)^{st} \\ &+ \frac{1}{6}\nu^h_{stu}(\widehat{\theta}_S - \theta)^{stu} + O_p(n^{-2}) \ , \end{aligned} \qquad (16)$$

where $H^h_s = u^{hr}H_{rs}, \ldots, H^h_{S_n} = u^{hr}H_{rS_n}$, $\nu^h_s = u^{hr}\nu_{rs}, \ldots, \nu^h_{S_n} = u^{hr}\nu_{rS_n}$, and $s^h = u^{hr}s_r$.

Using the iterative substitution method (see [**?**], Section 9.3.2), after some simple algebra we obtain

$$\begin{aligned} (\widehat{\theta}_S - \theta)^h &= s^h + \frac{1}{2}\nu^r_{st}s^{st} + H^r_s s^s + \frac{1}{6}(\nu^h_{stu} + 3\nu^h_{sv}\nu^v_{tu})s^{stu} \\ &+ \frac{1}{2}(H^h_{st} + 2\nu^h_{sv}H^v_t + H^r_v\nu^v_{st})s^{st} + H^h_v H^v_s s^s + O_p(n^{-2}) \ , \end{aligned} \qquad (17)$$

where $s^{st} = s^s s^t$, $s^{stu} = s^s s^t s^u$, and so on.

3.2. Asymptotic bias of $\widehat{\theta}_S$

Expansion (**??**) makes it possible to obtain an asymptotic expansion for the bias of $\widehat{\theta}_S$. We have

$$E_\theta(\widehat{\theta}_S - \theta)^h = \frac{1}{2}\nu^h_{st}E_\theta(s^{st}) + E_\theta(H^h_s s^s) + O(n^{-2}) \ , \qquad (18)$$

where the expectation of terms of order $O_p(n^{-3/2})$ in (**??**) is of order $O(n^{-2})$. Indeed, the expectations $E_\theta(s_r s_s s_t)$, $E_\theta(s_r s_s H_{tu})$ and $E_\theta(s_r H_{st} H_{uv})$ are of order $O(n)$.

The two expected values in (**??**) are, respectively,

$$E_\theta(s^{st}) = u^{sv}u^{tw}E_\theta(s_v s_w) = u^{sv}u^{tw}\nu_{v,w}$$

and

$$E_\theta(H^h_s s^s) = u^{hu}u^{sv}E_\theta(H_{us}s_v) = u^{hu}u^{sv}E_\theta((s_{us}-\nu_{us})\,s_v) = u^{hu}u^{sv}\nu_{us,v} \ .$$

Then, we obtain

$$\begin{aligned} E_\theta(\widehat{\theta}_S-\theta)^h &= \frac{1}{2}u^{hu}\nu_{ust}E_\theta(s^{st}) + E_\theta(H^h_s s^s) + O(n^{-2}) \\ &= \frac{1}{2}u^{hu}\nu_{ust}u^{sv}u^{tw}\nu_{v,w} + u^{hu}u^{sv}\nu_{us,v} + O(n^{-2}) \\ &= u^{hs}u^{uv}\left(\nu_{su,v} + \frac{1}{2}u^{tw}\nu_{sut}\nu_{v,w}\right) + O(n^{-2}) \ . \end{aligned} \tag{19}$$

Remark 1. Note that when the information identity holds, i.e. when $\nu_{r,s} = u_{rs}$ or, equivalently, $u^{tw}\nu_{v,w} = \delta^t_v$, where δ^t_v is the Kronecker's delta, equation (**??**) reduces to

$$\begin{aligned} E_\theta(\widehat{\theta}_S-\theta)^r &= u^{rs}u^{uv}\left(\nu_{su,v} + \frac{1}{2}\delta^t_v\nu_{sut}\right) + O(n^{-2}) \\ &= \frac{1}{2}u^{rs}u^{uv}\left(2\nu_{su,v} + \nu_{suv}\right) + O(n^{-2}) \ , \end{aligned} \tag{20}$$

recovering the result for the logarithmic score, that is for the MLE; see, e.g., **?** , Section 5.3], **?** , Section 9.4.2] and **?** , Section 5.3].

Remark 2. A first-order bias corrected scoring rule estimator can be defined as

$$\widehat{\theta}^c_S = \widehat{\theta}_S - d(\widehat{\theta}_S) \ ,$$

where $d(\theta)$ has components $d^r(\theta) = u^{rs}u^{uv}\left(\nu_{su,v} + \frac{1}{2}u^{tw}\nu_{sut}\nu_{v,w}\right)$. The asymptotic expansion (**??**) could also be used to obtain approximations of moments of $\widehat{\theta}_S$ of order higher than one; see, e.g., **?** , Section 9.4.3] and **?** , Section 5.3].

Remark 3. In the scalar case, i.e. for $k=1$, equation (**??**) reduces to

$$E_\theta(\widehat{\theta}_S-\theta) = \nu_{\theta\theta}^{-2}\left(\nu_{\theta\theta,\theta} - \frac{1}{2}\nu_{\theta\theta}^{-1}\nu_{\theta\theta\theta}\nu_{\theta,\theta}\right) + O(n^{-2}) \ , \tag{21}$$

where $\nu_{\theta\theta} = E_\theta(s_{\theta\theta})$, $\nu_{\theta,\theta} = E_\theta(s_\theta^2)$, $\nu_{\theta\theta,\theta} = E_\theta(s_{\theta\theta}s_\theta)$ and $\nu_{\theta\theta\theta} = E_\theta(s_{\theta\theta\theta})$. Moreover, when the second order Bartlett identity holds, i.e. when $\nu_{\theta,\theta} = -\nu_{\theta\theta} = u_{\theta\theta}$, equation (**??**) simplifies to

$$E_\theta(\widehat{\theta}_S - \theta) = u_{\theta\theta}^{-2}\left(\nu_{\theta\theta,\theta} + \frac{1}{2}\nu_{\theta\theta\theta}\right) + O(n^{-2}) \ . \tag{22}$$

3.3. Expansion of $W^S(\theta)$

Direct application of stochastic Taylor formula gives

$$\begin{aligned} S(\widehat{\theta}_S) - S(\theta) &= (\widehat{\theta}_S - \theta)^r s_r + \frac{1}{2}(\widehat{\theta}_S - \theta)^{rs} s_{rs} + \frac{1}{6}(\widehat{\theta}_S - \theta)^{rst} s_{rst} \\ &+ \frac{1}{24}(\widehat{\theta}_S - \theta)^{rstu} s_{rstu} + O_p(n^{-3/2}) \ . \end{aligned}$$

Using the quantities H_{R_n} and substituting (**??**) in the above expansion, we obtain

$$\begin{aligned} W^S(\theta) &= 2\left[(\widehat{\theta}_S - \theta)^r s_r + \frac{1}{2}(\widehat{\theta}_S - \theta)^{rs}\left(H_{rs} + \nu_{rs}\right)\right. \\ &+ \frac{1}{6}(\widehat{\theta}_S - \theta)^{rst}\left(H_{rst} + \nu_{rst}\right) \\ &+ \left.\frac{1}{24}(\widehat{\theta}_S - \theta)^{rstu}\left(H_{rstu} + \nu_{rstu}\right)\right] + O_p(n^{-3/2}) \\ &= u_{rs}s^{rs} + \frac{1}{3}(\nu_{rst}s^t + 3H_{rs})s^{rs} + \frac{1}{12}(\nu_{rstu} + 3\nu_{rsv}\nu_{tu}^v)s^{rstu} \\ &+ \frac{1}{3}(H_{rst} + 3\nu_{rsv}H_t^v)s^{rst} + H_{rv}H_s^v s^{rs} + O_p(n^{-3/2}) \ . \end{aligned} \tag{23}$$

From (**??**) it is possible to obtain an asymptotic expansion for $E_\theta(W^S(\theta))$ of the form

$$E_\theta(W^S(\theta)) = u^{rs}\nu_{r,s} + R(\theta) + O(n^{-2}) \ , \tag{24}$$

where $R(\theta)$ is of order $O(n^{-1})$ and is given by

$$R(\theta) = \frac{1}{12}\left(B_{rstu}u^{rs}u^{tu} + B_{rstuvw}u^{rs}u^{tu}u^{vw}\right) \ ,$$

with

$$\begin{aligned} B_{rstu} &= 3\nu_{ysux}u^{yw}u^{xl}\nu_{w,r}\nu_{t,l} \\ &+ 12\left(\nu_{yt,us}u^{yw}\nu_{w,r} + \nu_{r,t,us} + \nu_{t,ysu}u^{yw}\nu_{w,r}\right) , \end{aligned} \tag{25}$$

and

$$\begin{aligned}
B_{rstuvw} &= (3\nu_{rsz}\nu_{uyw} + 6\nu_{rzy}\nu_{suw})u^{xy}u^{zp}\nu_{x,v}\nu_{t,p} \\
&+ 12(\nu_{r,st}\nu_{uv,w} + \nu_{r,tv}\nu_{su,w} + \nu_{x,st}\nu_{uvy}u^{xy}\nu_{w,y}) \\
&+ 24\nu_{r,yv}\nu_{suw}u^{xy}\nu_{x,t} + 4\nu_{r,t,v}\nu_{suw}\ . \qquad (26)
\end{aligned}$$

In order to obtain expression (**??**) the following approximations have been used

$$\begin{aligned}
E_\theta(s_l s_k s_x s_z) &= \nu_{l,k,x,z} = \nu_{l,k}\nu_{x,z} + \nu_{l,x}\nu_{k,z} + \nu_{l,z}\nu_{k,x} + O(n), \\
E_\theta(H_{rst}s_x s_l s_k) &= \nu_{rst,x}\nu_{l,k} + \nu_{rst,l}\nu_{x,k} + \nu_{rst,k}\nu_{z,l} + O(n), \\
E_\theta(H_{qt}s_x s_l s_k) &= \nu_{qt,x}\nu_{l,k} + \nu_{qt,l}\nu_{x,k} + \nu_{qt,k}\nu_{x,l} + O(n), \\
E_\theta(H_{rs}H_{tu}s_v s_w) &= \nu_{v,w}(\nu_{rs,tu} - \nu_{rs}\nu_{tu}) + \nu_{rs,v}\nu_{tu,w} + \nu_{rs,w}\nu_{tu,v} + O(n).
\end{aligned}$$

as well as the following identity in (**??**)

$$E_\theta(H_{rs}s_v s_w) = \nu_{rs,v,w} - \nu_{rs}\nu_{v,w}.$$

Note that we are also implicitly using the identity $u^{vl}\nu_{ls} = -\delta_s^v$.

Remark 4. If the information identity holds, equations (**??**) and (**??**) reduce to

$$\begin{aligned}
B_{rstu} &= 3\nu_{ysux}\delta_r^y\delta_t^x + 12\left(\nu_{yt,us}\delta_r^y + \nu_{r,t,us} + \nu_{t,ysu}\delta_r^y\right) \\
&= 3\nu_{rsut} + 12\left(\nu_{rt,us} + \nu_{r,t,us} + \nu_{t,rsu}\right),
\end{aligned}$$

and

$$\begin{aligned}
B_{rstuvw} &= (3\nu_{rsz}\nu_{uyw} + 6\nu_{rzy}\nu_{suw})\delta_v^y\delta_t^z \\
&+ 12(\nu_{r,st}\nu_{uv,w} + \nu_{r,tv}\nu_{su,w} + \nu_{r,st}\nu_{uvy}\delta_w^y) \\
&+ 24\nu_{r,yv}\nu_{suw}\delta_t^y + 4\nu_{r,t,v}\nu_{suw} \\
&= (3\nu_{rst}\nu_{uvw} + 6\nu_{rtv}\nu_{suw}) \\
&+ 12(\nu_{r,st}\nu_{uv,w} + \nu_{r,tv}\nu_{su,w} + \nu_{r,st}\nu_{uvw}) \\
&+ 24\nu_{r,tv}\nu_{suw} + 4\nu_{r,t,v}\nu_{suw}\ .
\end{aligned}$$

In this case, $R(\theta)$ is consistent with the result for the likelihood ratio test; see, e.g., **?** , Section 5.3], **?** , Section 9.4.5] and **?** , Section 5.4].

3.4. Asymptotics for scoring rule signed square root

As in **?** , Chapter 7], we define the vector with components

$$\begin{aligned} W_t^S &= s_t + \frac{1}{6}u^{rs}(3H_{rt} + u^{vw}\nu_{rtv}s_w)s_s \\ &+ \frac{1}{72}u^{rs}u^{vw}s_w\,(27H_{rt}H_{sv} + 8u^{xy}u^{zp}\nu_{rtv}\nu_{sxz}s_y s_p \\ &+ 30u^{xy}\nu_{rtv}H_{sx}s_y + 12H_{rtv}s_s + 3u^{xy}\nu_{rtvx}s_s s_y) + O_p(n^{-1})\ , \qquad (27) \end{aligned}$$

which is such that $W^S(\theta) = W_t^S W_u^S u^{tu} + O_p(n^{-3/2})$.

From (**??**), it is possible to obtain an asymptotic expansion for the expectation of W_t^S, with error of order $O(n^{-1})$. We have

$$E_\theta(W_t^S) = \frac{1}{6}u^{rs}(3\nu_{rt,s} + u^{vw}\nu_{rtv}\nu_{w,s}) + O(n^{-1})\ . \qquad (28)$$

If the information identity holds, the above expansion reduces to

$$\begin{aligned} E_\theta(W_t^S) &= \frac{1}{6}u^{rs}(3\nu_{rt,s} + \delta_s^v\nu_{rtv}) + O(n^{-1}) \\ &= \frac{1}{6}u^{rs}(3\nu_{rt,s} + \nu_{rts}) + O(n^{-1})\ . \qquad (29) \end{aligned}$$

Using (**??**), (**??**) and (**??**), we can obtain the expression for $Cov_\theta(W_t^S, W_s^S)$. We have

$$\begin{aligned} Cov_\theta(W_t^S, W_s^S) &= E_\theta(W_t^S W_s^S) - E_\theta(W_t^S)E_\theta(W_s^S) \\ &= E_\theta((W^S)^2)u_{ts} - E_\theta(W_t^S)E_\theta(W_s^S) \\ &= \nu_{t,s} + \frac{1}{12}(B_{rstu}u^{ru} + B_{rstuvw}u^{ru}u^{vw}) \\ &- E_\theta(W_t^S)E_\theta(W_s^S) + O(n^{-1})\ . \end{aligned}$$

Inserting (**??**) in the above formula, with straightforward but rather lengthy calculations, we obtain

$$\begin{aligned} Cov_\theta(W_t^S, W_s^S) &= \nu_{t,s} + \frac{1}{12}\{B_{rstu}u^{ru} + u^{ru}u^{vw}[B_{rstuvw} - 3\nu_{rt,u}\nu_{vs,w} \\ &- u^{lm}\nu_{m,u}\,(\nu_{vt,w}\nu_{rsl} + \nu_{vs,w}\nu_{rtl}) - \frac{1}{3}\nu_{rtx}\nu_{vsz}u^{xy}u^{zp}\nu_{y,u}\nu_{p,w}]\} \\ &+ O(n^{-1})\ . \end{aligned}$$

Let $r^S = W_t^S i_t^{-1/2}$, where $i_t^{-1/2}$ denotes an arbitrary square root matrix of u^{rs}. Then

$$E_\theta(r^S) = m(\theta) + O(n^{-3/2}) \ , \tag{30}$$

with

$$m(\theta) = \frac{1}{6} u^{rs} i_t^{-1/2} (3\nu_{rt,s} + u^{vw} \nu_{rtv} \nu_{w,s})$$

of order $O(n^{-1/2})$. Moreover,

$$Var_\theta(r^S) \quad = \quad Cov_\theta(W_t^S, W_s^S) u^{ts} = u^{ts} \nu_{t,s} + v(\theta) + O(n^{-2}) \ ,$$

with

$$\begin{aligned} v(\theta) \quad &= \quad R(\theta) - \frac{1}{12} \left[3\nu_{rt,u} \nu_{vs,w} + \frac{1}{3} \nu_{rtx} \nu_{vsz} u^{xy} u^{zp} \nu_{y,u} \nu_{p,w} \right. \\ &+ \quad \left. u^{lm} \nu_{m,u} \left(\nu_{vt,w} \nu_{rsl} + \nu_{vs,w} \nu_{rtl} \right) \right] u^{rs} u^{tu} u^{vw} \end{aligned}$$

of order $O(n^{-1})$. The first term $u^{ts}\nu_{t,s}$ in $Var_\theta(r^S)$ indicates that r^S has not a standard normal distribution.

Remark 5. In the scalar case, we have

$$m(\theta) = \frac{1}{6} u_{\theta\theta}^{-3/2} (3\nu_{\theta\theta,\theta} + u_{\theta\theta}^{-1} \nu_{\theta\theta\theta} \nu_{\theta,\theta}) \tag{31}$$

and

$$v(\theta) = R(\theta) - m(\theta)^2 \ , \tag{32}$$

where $R(\theta)$ is given by

$$R(\theta) = \frac{1}{12} (B_4 u_{\theta\theta}^{-2} + B_6 u_{\theta\theta}^{-3}) \ , \tag{33}$$

with

$$B_4 = 3\nu_{\theta\theta\theta\theta} u_{\theta\theta}^{-2} \nu_{\theta,\theta}^2 + 12(\nu_{\theta\theta,\theta\theta} u_{\theta\theta}^{-1} \nu_{\theta,\theta} + \nu_{\theta,\theta,\theta\theta} + \nu_{\theta,\theta\theta\theta} u_{\theta\theta}^{-1} \nu_{\theta,\theta}) \tag{34}$$

and

$$B_6 = 9\nu_{\theta\theta\theta}^2 u_{\theta\theta}^{-2} \nu_{\theta,\theta}^2 + 24\nu_{\theta,\theta\theta}^2 + 36\nu_{\theta,\theta\theta} \nu_{\theta\theta\theta} u_{\theta\theta}^{-1} \nu_{\theta,\theta} + 4\nu_{\theta,\theta,\theta} \nu_{\theta\theta\theta} \ , \tag{35}$$

where the symbol $\nu = \nu(\theta)$ is used to indicate moments of scoring rule derivatives, i.e. $\nu_{\theta\theta\theta} = E_\theta(s_{\theta\theta\theta})$, $\nu_{\theta,\theta\theta} = E_\theta(s_\theta s_{\theta\theta})$, and so on.

Note that in the scalar parameter case for the scoring rule ratio statistic we have

$$\begin{aligned}
W^S(\theta) &= u_{\theta\theta} s^{\theta\theta} + \frac{1}{3}(\nu_{\theta\theta\theta} s^{\theta} + 3H_{\theta\theta}) s^{\theta\theta} + \frac{1}{12}(\nu_{\theta\theta\theta\theta} + 3\nu_{\theta\theta\theta}\nu^{\theta}_{\theta\theta}) s^{\theta\theta\theta\theta} \\
&+ \frac{1}{3}(H_{\theta\theta\theta} + 3\nu_{\theta\theta\theta} H^{\theta}_{\theta}) s^{\theta\theta\theta} + H_{\theta\theta} H^{\theta}_{\theta} s^{\theta\theta} + O_p(n^{-3/2}) \\
&= u_{\theta\theta} u^{\theta\theta} u^{\theta\theta} s_\theta^2 + \frac{1}{3}(\nu_{\theta\theta\theta} u^{\theta\theta} s_\theta + 3H_{\theta\theta}) \left(u^{\theta\theta}\right)^2 s_\theta^2 \\
&+ \frac{1}{12}(\nu_{\theta\theta\theta\theta} + 3\nu_{\theta\theta\theta}\nu_{\theta\theta\theta} u^{\theta\theta}) s_\theta^4 \left(u^{\theta\theta}\right)^4 + \frac{1}{3}(H_{\theta\theta\theta} + 3\nu_{\theta\theta\theta} H_{\theta\theta} u^{\theta\theta}) s_\theta^3 \left(u^{\theta\theta}\right)^3 \\
&+ H_{\theta\theta} H_{\theta\theta} u^{\theta\theta} s_\theta^2 + O_p(n^{-3/2}) \ , \qquad (36)
\end{aligned}$$

and its expected value is

$$\begin{aligned}
E(W^S(\theta)) &= u^{\theta\theta}\nu_{\theta,\theta} + \frac{1}{3}(\nu_{\theta\theta\theta} u^{\theta\theta} \nu_{\theta,\theta,\theta} + 3\nu_{\theta\theta,\theta,\theta}) \left(u^{\theta\theta}\right)^2 \\
&+ \frac{3}{12}(\nu_{\theta\theta\theta\theta} + 3\nu_{\theta\theta\theta}\nu_{\theta\theta\theta} u^{\theta\theta}) \nu^2_{\theta,\theta} \left(u^{\theta\theta}\right)^4 \\
&+ \frac{1}{3}(3\nu_{\theta\theta\theta,\theta}\nu_{\theta,\theta} + 9\nu_{\theta\theta\theta}\nu_{\theta\theta,\theta}\nu_{\theta,\theta} u^{\theta\theta}) \left(u^{\theta\theta}\right)^3 \\
&+ \left(\nu_{\theta\theta,\theta\theta}\nu_{\theta,\theta} + 2\nu^2_{\theta\theta,\theta}\right) \left(u^{\theta\theta}\right)^3 + O(n^{-3/2}) \\
&= u^{\theta\theta}\nu_{\theta,\theta} + \frac{1}{3}(\nu_{\theta\theta\theta} u^{\theta\theta} \nu_{\theta,\theta,\theta} + 3\nu_{\theta\theta,\theta,\theta}) \left(u^{\theta\theta}\right)^2 \\
&+ \frac{3}{12}(\nu_{\theta\theta\theta\theta} + 3\nu_{\theta\theta\theta}\nu_{\theta\theta\theta} u^{\theta\theta}) \nu^2_{\theta,\theta} \left(u^{\theta\theta}\right)^4 \\
&+ \frac{1}{3}(3\nu_{\theta\theta\theta,\theta}\nu_{\theta,\theta} + 9\nu_{\theta\theta\theta}\nu_{\theta\theta,\theta}\nu_{\theta,\theta} u^{\theta\theta}) \left(u^{\theta\theta}\right)^3 \\
&+ \left(\nu_{\theta\theta,\theta\theta}\nu_{\theta,\theta} + 2\nu^2_{\theta\theta,\theta}\right) \left(u^{\theta\theta}\right)^3 + O(n^{-3/2}) \ . \qquad (37)
\end{aligned}$$

Given the expressions (**??**) and (**??**) for $m(\theta)$ and $v(\theta)$, the proposed modified signed scoring rule root statistic is

$$r^S_M(\theta) = \frac{r^S(\theta) - m(\theta)}{\sqrt{\mu_1(\theta) + v(\theta)}} \ ,$$

with $\mu_1(\theta)$ in (**??**).

The mean and the variance corrections introduced can improve the accuracy of the normal approximation of the distribution of $r^S_M(\theta)$. In particular, if third- and higher-order cumulants of $r^S_M(\theta)$ are of order $O(n^{-3/2})$ or smaller, then the normal approximation to the distribution of the modified signed scoring rule root statistic $r_M^S(\theta)$ has an error of order $O(n^{-3/2})$. This holds in particular in models with all higher cumulants zero and for the logarithmic score. If cumulants do not fulfilled such conditions, the normal approximation of the distribution of $r_M^S(\theta)$ has an error up to the second order, i.e. to order $O(n^{-1})$. Appendix A gives the expansions for the third cumulant of $r_M^S(\theta)$.

4. Simulations

In this section we provide simulation results to assess coverage probabilities of confidence intervals for a scalar parameter of interest based on the modified signed scoring rule root statistic $r_M^S(\theta)$ and on $r_{pM}^S(\psi)$, for a scalar parameter of interest ψ in presence of nuisance parameters. We discuss three examples which are classical and historical examples of applications of higher-order asymptotics in the likelihood framework (see [**?**], [**?**], **?**], [**?**]). Example 4.1 refers to the situation of a normal model with known coefficient of variation. Examples 4.2 and 4.3 deal with nuisance parameters. In particular, Example 4.2 focuses on the log-gamma density for which the Hyvärinen scoring rule is particularly attractive, since the estimators can be found in closed form. Examples 4.3 considers robust scoring rules for the linear regression model. In all the examples the Tsallis scoring rule is robust. In order to assess the stability of the coverage levels under small, arbitrary departures from the assumed model, we consider a contamination model of the form $Q_\epsilon = (1-\epsilon)Q_\theta + \epsilon G$, where Q_θ and G denote the true and the contaminating distributions, respectively. The parameter γ of the Tsallis score is the trade-off between robustness and efficiency. Its value is fixed in order to ensure a given asymptotic variance - hence a given asymptotic efficiency - at the assumed model. It is generally chosen to achieve approximately 90% or 95% efficiency under the assumed model distribution (see [**?**]; **?** , Chapter 6]).

4.1. Normal parabola

Let $x_1, \ldots, x_n$ be a sample from a normal distribution with mean θ and variance θ^2. This is a $(2, 1)$ curved exponential family, known as the normal parabola ([**?**]). This example has been discussed, among others, in **?** , Chapter 7].

The log-likelihood function is

$$\ell(\theta) = -\frac{n}{2}\log 2\pi - n\log\theta - \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\theta)^2, \tag{38}$$

the Hyvärinen total empirical score is

$$S_H(\theta) = \frac{2n}{\theta^2} - \frac{1}{\theta^4}\sum_{i=1}^{n}(x_i-\theta)^2, \tag{39}$$

and the Tsallis total empirical score is

$$S_T(\theta) = -\frac{n(\gamma-1)(2\pi\theta^2)^{-(\gamma-1)/2}}{\sqrt{\gamma}} + \gamma(2\pi\theta^2)^{-(\gamma-1)/2}\sum_{i=1}^{n} e^{-\frac{(\gamma-1)(x_i-\theta)^2}{2\theta^2}}. \tag{40}$$

Both the MLE and the Hyvärinen score estimator are roots of a quadratic equation, while the Tsallis score estimator cannot be found in closed form. Note that the Tsallis score estimator is B-robust since the corresponding $s_\theta(x, \theta)$ is bounded.

We ran a simulation experiment, for several values of n and $\theta = 2$, in order to assess the accuracy of $r_M^S(\theta)$, when data are generated both from the central model and from the normal model contaminated by a normal distribution with mean 2 and variance 4^2. The contamination percentage ϵ is set at 5%. For comparison, we consider the scoring rule Wald-type statistic $W_w^S(\theta)$ and the first-order adjustment $r_{adj}^S(\theta)$, both for the Hyvärinen and the Tsallis scoring rules, and the corresponding likelihood-type counterparts. Results of the Tsallis statistics are given for $\gamma = 1.2$, which gives approximately 90% efficiency under the central model (see [**?**]; [**?** , Chapter 6]). Table **??** gives the results of the simulation study based on $10,000$ Monte Carlo trials.

Note that, under the central model, the Hyvärinen and the Tsallis modified signed scoring rule roots (r_M^H and r_M^T) perform better than the first-order adjustments (r_{adj}^H and r_{adj}^T). Moreover, they outperform also the corresponding

Wald-type statistics (W_w^H and W_w^T). The third-order pivots are highly accurate even for rather small sample sizes and behave quite similarly with the corresponding likelihood-type statistic counterpart. The differences among the pivots considered vanish as the sample size increases. Under the contaminated model, the Tsallis procedures show a reasonable behavior, while both likelihood and Hyvärinen statistics exhibit poor coverage. Finally, we remark that similar empirical coverages, not reported here, were obtained also for different values of θ.

Table 1: *Normal parabola.* Empirical coverages of 95% confidence intervals, for $\theta = 2$ and $n = 5, 10, 20$. Pivots used: Wald (w), signed likelihood root (r), higher-order signed likelihood root (r^*), Hyvärinen and Tsallis scoring rules Wald-type (W_w^H and W_w^T), Hyvärinen and Tsallis adjusted signed scoring rule roots (r_{adj}^H and r_{adj}^T) and Hyvärinen and Tsallis modified signed scoring rule roots (r_M^H and r_M^T). The Tsallis statistics are given for $\gamma = 1.2$.

					$N(2,4)$				
n	r	r^*	w	r_{adj}^H	r_M^H	W_w^H	r_{adj}^T	r_M^T	W_w^T
5	0.938	0.946	0.869	0.904	0.942	0.867	0.938	0.948	0.924
10	0.950	0.951	0.916	0.936	0.952	0.914	0.950	0.952	0.925
20	0.946	0.949	0.928	0.940	0.949	0.928	0.947	0.948	0.949
					$N(2,4)$ cont.				
n	r	r^*	w	r_{adj}^H	r_M^H	W_w^H	r_{adj}^T	r_M^T	W_w^T
5	0.915	0.918	0.868	0.890	0.918	0.871	0.937	0.939	0.801
10	0.914	0.914	0.895	0.899	0.913	0.901	0.940	0.941	0.889
20	0.901	0.898	0.928	0.892	0.890	0.913	0.941	0.940	0.935

4.2. The log-gamma

Consider a random sample $x_1, \ldots, x_n$ of size n from a log-gamma distribution with scale parameter $\alpha > 0$ and shape parameter $\beta > 0$. Let $\theta = (\alpha, \beta)$. The log-likelihood function is

$$\ell(\theta) = -n\beta \log \alpha - n \log(\Gamma(\beta)) + \sum_{i=1}^{n} \left(\beta x_i - \frac{1}{\alpha} e^{x_i} \right), \tag{41}$$

and the Hyvärinen total empirical score is

$$S_H(\theta) = \frac{2}{\alpha}\sum_{i=1}^{n} e^{x_i} - \sum_{i=1}^{n}\left(\beta - \frac{1}{\alpha}e^{x_i}\right)^2 . \tag{42}$$

The Hyvärinen scoring rule estimates are $\widehat{\beta}_H = \frac{\left(\sum_{i=1}^n e^{x_i}\right)^2}{n\sum_{i=1}^n e^{2x_i} - \left(\sum_{i=1}^n e^{x_i}\right)^2}$ and $\widehat{\alpha}_H = \frac{n\sum_{i=1}^n e^{2x_i} - \left(\sum_{i=1}^n e^{x_i}\right)^2}{n\sum_{i=1}^n e^{x_i}}$. The MLE $\widehat{\theta}$ is the solution of the likelihood equations $\psi(\beta) = \bar{x} - \log\alpha$ and $\alpha = \frac{1}{n\beta}\sum_{i=1}^n e^{x_i}$, where $\psi(\cdot)$ is the di-gamma function. The Tsallis total empirical score is

$$S_T(\theta) = \frac{\gamma}{(\alpha^{\beta}\Gamma(\beta))^{\gamma-1}}\sum_{i=1}^{n} e^{(\gamma-1)(x_i\beta - \frac{1}{\alpha}e^{x_i})} - n(\gamma-1)\frac{\Gamma(\beta\gamma)}{\gamma^{\beta\gamma}\Gamma(\beta)^{\gamma}}. \tag{43}$$

The Tsallis scoring rule is robust for this family of distributions, since the density $p(x;\theta)$ and its derivative with respect to θ are bounded in x, for each θ; see Condition 5.2 in [**?**].

Note that in the case of α known, the log-gamma density belongs to the one-parameter natural exponential family with natural parameter β. For this family, the Hyvärinen modified signed scoring rule root test statistic $r_M^H(\beta)$ coincides with the adjusted signed scoring rule root test statistic $r_{adj}^H(\beta)$ and with the scoring rule score-type statistic. Details on the equivalence of these three statistics in the one-parameter exponential family are summarized in Appendix B.

Let β be the parameter of interest and let α be the nuisance parameter. In order to assess the quality of the proposed Hyvärinen and Tsallis modified signed scoring rule roots ($r_{pM}^H(\beta)$ and $r_{pM}^T(\beta)$), we run a simulation experiment with $n = 10, 15, 20$, $\beta = 2$ and $\alpha = 1$, both under the central model and a contaminated model, i.e. the log-gamma model contaminated by a log-gamma distribution with parameters $\alpha = e$ and $\beta = 1$. The contamination percentage ϵ is set at 10%. For comparison, we also consider the scoring rule Wald-type statistic $W_{wp}^S(\beta)$ and the first-order adjustments $r_p^S(\beta)_{adj}$, both for the Hyvärinen and the Tsallis scoring rules, and the corresponding likelihood-type counterparts. Results of the Tsallis statistics are given for $\gamma = 1.25$, which gives approximately 95% efficiency under the central model (see [**?**]; **?** , Chapter 6]). Table **??** gives the results of the simulation study based on 10,000 Monte Carlo trials. Under the correct model, the modified signed profile scoring rule root statistics r^H_{pM} and r^T_{pM} perform better with respect to the corresponding first-order counterpart $r^H_{p\,adj}$ and $r^T_{p\,adj}$. Larger sample sizes would show, as one would expect, rather little differences between the results of all the procedures. Under the contaminated model, the Tsallis modified signed scoring rule root statistic show a reasonable performance in terms of coverage, while both the likelihood and Hyvärinen modified signed scoring rule root statistics exhibit poor coverage, as well as the first order counterparts.

As it is well known, the log-score is invariant with respect to transformations of the data. Contrariwise, it is easy to show that the Tsallis score [**?**] and the Hyvärinen score are invariant only with respect to linear transformations of the data. Results not reported here, shown that the Tsallis modified signed scoring rule root statistic performs the first order counterparts also in the case of the exponential distribution (i.e. for $\beta = 1$ and $Y = \exp(X)$, where X is a log-gamma random variable). The unreported results indicated also that the modified signed scoring rule root statistic based on a version of the Hyvärinen scoring rule [**?**] for non-negative data outperforms the corresponding first order statistics. Note that, the Hyvärinen scoring rule in (**??**) is not proper for the exponential distribution [**?**] leading to a biased estimating function.

4.3. Linear regression model

Let us consider a linear regression model as in [**?**]

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

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.

Let $\theta = (\beta, \sigma)$. The Hyvärinen total empirical score is

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

Table 2: *Log-gamma*. Empirical coverages of 95% confidence intervals, for $\beta = 2$ and $\alpha = 1$ and $n = 10, 15, 20$. Pivots used: profile Wald (w_p), signed profile likelihood root (r_p), higher-order signed profile likelihood root (r_p^*), Hyvärinen and Tsallis profile scoring rules Wald-type (W_{wp}^H and W_{wp}^T), Hyvärinen and Tsallis adjusted signed profile scoring rule roots ($r_{p\ adj}^H$ and $r_{p\ adj}^T$) and Hyvärinen and Tsallis modified signed profile scoring rule roots (r_{pM}^H and r_{pM}^T). The Tsallis statistics are given for $\gamma = 1.25$.

					$Log-Gamma(1,2)$				
n	r_p	r_p^*	w_p	$r_{p\ adj}^H$	r_{pM}^H	W_{wp}^H	$r_{p\ adj}^T$	r_{pM}^T	W_{wp}^T
10	0.925	0.945	0.963	0.988	0.946	0.974	0.988	0.948	0.950
15	0.934	0.951	0.963	0.997	0.947	0.975	0.997	0.948	0.955
20	0.944	0.953	0.960	0.999	0.953	0.975	0.999	0.954	0.957
					$Log-Gamma(1,2)$ cont.				
n	r_p	r_p^*	w_p	$r_{p\ adj}^H$	r_{pM}^H	W_{wp}^H	$r_{p\ adj}^T$	r_{pM}^T	W_{wp}^T
10	0.910	0.920	0.932	0.981	0.931	0.950	0.981	0.936	0.927
15	0.918	0.908	0.919	0.994	0.923	0.938	0.994	0.933	0.927
20	0.919	0.904	0.912	0.998	0.916	0.930	0.998	0.929	0.919

where x_i^T is the i-th row of X. When the parameter σ is known and equal to one, the Hyvärinen test statistics coincide with classical likelihood statistics; the Hyvärinen estimator coincides with the maximum likelihood estimator also in the case with σ unknown (see [**?**]). This is an example where the matrix $J(\theta)$ is different from the matrix $K(\theta)$, but the Godambe information matrix reduces to the Fisher information one. Consequently, Wald and score type statistics based on the Hyvärinen scoring rule coincides with the full-likelihood counterparts. The Tsallis total empirical score is [**?**]

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

Let $p = 3$, $\psi = \beta_2$ be the scalar parameter of interest, and let $\lambda = (\beta_1, \beta_3, \sigma)$ be the nuisance parameter. We ran a simulation experiment, for several values of n, with $\beta = (1, 2, 3)$ and $\sigma = 1$, in order to assess the accuracy of $r_{pM}^S(\beta_2)$ when the model is correctly specified and when the contaminated model is $Q_{0.1} = 0.90N(0, 1) + 0.1N(100, 30^2)$. The entries of the first column of the matrix X are 1, those of the second column and third column are generated as independent standard normal variables. For comparison, we consider also the Wald-type statistics, $W^S_{wp}(\beta_2)$, the first-order adjustments, $r^S_p(\beta_2)_{adj}$, and the corresponding likelihood-type counterparts. Results of the Tsallis statistics are given for $\gamma = 1.22$ [**?** **?** **?**], the Tsallis scoring rule estimator is robust for the linear regression model, and gives approximately 95% efficiency under the central model (see [**?**]; [**?**]; **?** , Chapter 6]). Table **??** gives the results of the simulation study based on $10,000$ Monte Carlo trials.

Under the central model, $r^T_{pM}(\beta_2)$ and $r^H_{pM}(\beta_2)$ perform better than the corresponding first-order adjustments $r^T_{p\,adj}(\beta_2)$ and $r^H_{p\,adj}(\beta_2)$, respectively. The higher-order Tsallis signed root statistic show a more stable behavior in the presence of small deviations from the model.

Table 3: *Linear regression model.* Empirical coverages of 95% confidence intervals, for $\beta_2 = 2$ when $\lambda = (1, 3, 1)$ and $n = 10, 15, 20$. Pivots used: profile Wald (w_p), signed profile likelihood root (r_p), higher-order signed profile likelihood root (r^*_p), Hyvärinen and Tsallis profile scoring rules Wald-type (W^H_{wp} and W^T_{wp}), Hyvärinen and Tsallis adjusted signed profile scoring rule roots ($r^H_{p\,adj}$ and $r^T_{p\,adj}$) and Hyvärinen and Tsallis modified signed profile scoring rule roots (r^H_{pM} and r^T_{pM}). The Tsallis statistics are given for $\gamma = 1.22$.

					$N(0,1)$				
n	r_p	r^*_p	w_p	$r^H_{p\,adj}$	r^H_{pM}	W^H_{wp}	$r^T_{p\,adj}$	r^T_{pM}	W^T_{wp}
10	0.887	0.953	0.855	0.855	0.944	0.855	0.862	0.951	0.834
15	0.912	0.951	0.893	0.893	0.947	0.893	0.898	0.949	0.886
20	0.927	0.951	0.915	0.915	0.948	0.915	0.915	0.953	0.910
					$N(0,1)$ cont.				
n	r_p	r^*_p	w_p	$r^H_{p\,adj}$	r^H_{pM}	W^H_{wp}	$r^T_{p\,adj}$	r^T_{pM}	W^T_{wp}
10	0.885	0.971	0.842	0.653	0.981	0.842	0.857	0.940	0.999
15	0.925	0.969	0.890	0.645	0.989	0.890	0.911	0.951	1
20	0.936	0.958	0.921	0.598	0.993	0.921	0.922	0.950	1

5. Final remarks

Asymptotic results have been widely discussed and illustrated for likelihood procedures in the statistical literature; see, e.g., [? ? ?], and references therein. This paper discusses higher-order asymptotics to the scoring rule setting extending results for likelihood procedures; see for example [?]. The results discussed in this paper include also, as a special instance, the results obtained by [?] for composite likelihood, which is a composite logarithmic scoring rule (see Section 2).

Modern likelihood theory based on higher-order asymptotic expansions is very accurate, even for small sample sizes, when the model is correct (see, among others, [?] and [?], and references therein). However, [?] show that small deviations from the assumed model can wipe out the improvements of the accuracy obtained at the model by higher-order likelihood asymptotics. Under small deviations from the assumed model, the behaviour of robust procedures - such as the Tsallis score - in terms of second-order accuracy is generally more stable and reliable than that of their classical counterparts [?]. The Hyvärinen score is not robust and, under model misspecification, its behavior is similar to classical likelihood procedures. However, it has been considered in the examples in order to illustrate also for this scoring rule the improvement due to higher-order asymptotics under the central model. Moreover, in the case of the log-gamma distribution it gives rise to closed form estimators, which can be suitable as a starting point for iterative procedures (see [?]). In the linear regression model setting, inference based on the Wald-type Hyvärinen scoring rule statistic coincides with the classical likelihood counterpart. Although examples illustrated in Section 4 focus on simple models, they are classical and historical examples of application of higher-order asymptotics in the likelihood framework ([?], [?], ?], ?]) which highlight the importance of proper scoring rule. For more complex models, paralleling likelihood procedures, a promising line of research appears to be the investigation of the performance of a bootstrap approximation when analytical calculation of the expected values involved in $r^S_M(\theta)$ (or

$r^S_{pM}(\psi)$) is cumbersome as the dimension of the parameter θ (and in particular of the nuisance parameter λ) is large.

Acknowledgements

We are grateful to the associate editor and anonymous referees for many useful comments that greatly improved the paper. This work was partially supported by grants from the University of Cagliari (Progetto di Ricerca Fondamentale o di Base 2012 L.R. 7/2007 annualità 2012) and from the University of Padova grant no. CPDA131553.

Appendix A: third order cumulant of r^S in the scalar parameter case

In this section we derive the third cumulant and the skewness of r^S for $k = 1$. The evaluation of the third cumulant

$$k_3(r^S) = E_\theta((r^S)^3) - 3E_\theta((r^S)^2)E_\theta(r^S) + 2(E_\theta(r^S))^3,$$

requires the calculation of the following quantities $E_\theta(r^S)$, $E_\theta((r^S)^2)$, $E_\theta((r^S)^3)$ and $(E_\theta(r^S))^3$. By using the expected values of r^S and $(r^S)^2 = W^S$ in equations (**??**) and (**??**), respectively, it follows easily that

$$E_\theta(r^S)E_\theta((r^S)^2) = \frac{1}{6}u_{\theta\theta}^{-5/2}\left(3\nu_{\theta\theta,\theta} + u_{\theta\theta}^{-1}\nu_{\theta\theta\theta}\nu_{\theta,\theta}\right) + O(n^{-3/2}).$$

Moreover, by noting that the leading term of the mean of r^S in (**??**) is of order $n^{-1/2}$, we can deduce that $(E_\theta(r^S))^3 = O(n^{-3/2})$. In order to find the expectation of $E_\theta((r^S)^3)$, we need the following expansion

$$(r^S)^3 = u_{\theta\theta}^{-3/2}s_\theta^3 + \frac{1}{2}u_{\theta\theta}^{-5/2}s_\theta^3\left(u_{\theta\theta}^{-1}\nu_{\theta\theta\theta}s_\theta + 3H_{\theta\theta}\right) + O_p(n^{-3/2}).$$

After taking account that $E_\theta(s_\theta^4) = 3\nu_{\theta,\theta}^2 + O(n)$ and $E_\theta(s_\theta^3 H_{\theta\theta}) = 3\nu_{\theta,\theta}\nu_{\theta,\theta\theta} + O(n)$, the expected value of $(r^S)^3$ reduces to

$$E_\theta((r^S)^3) = u_{\theta\theta}^{-3/2}\nu_{\theta,\theta,\theta} + \frac{3}{2}u_{\theta\theta}^{-5/2}\left(u_{\theta\theta}^{-1}\nu_{\theta\theta\theta}\nu_{\theta,\theta}^2 + 3\nu_{\theta\theta,\theta}\nu_{\theta,\theta}\right) + O(n^{-3/2}).$$

Finally, now we are in the right position to find the third order cumulant and the skewness of r^S. In particular,

$$k_3(r^S) = u_{\theta\theta}^{-3/2}\left[\nu_{\theta,\theta,\theta} + u_{\theta\theta}^{-1}\nu_{\theta,\theta}\left(u_{\theta\theta}^{-1}\nu_{\theta\theta\theta}\nu_{\theta,\theta} + 3\nu_{\theta\theta,\theta}\right)\right] + O(n^{-3/2}), \qquad (44)$$

$$\rho_3(r^S) = \nu_{\theta,\theta}^{-3/2}\left[\nu_{\theta,\theta,\theta} + u_{\theta\theta}^{-1}\nu_{\theta,\theta}\left(u_{\theta\theta}^{-1}\nu_{\theta\theta\theta}\nu_{\theta,\theta} + 3\nu_{\theta\theta,\theta}\right)\right] + O(n^{-3/2}). \qquad (45)$$

By using properties of cumulants we note that the third cumulant of r_M^S is $k_3(r_M^S) = var(r^S)^{-3/2}k_3(r^S)$.

In order to achieve the third order normal approximation for the distribution of r_M^S, we require that the third- and higher-order cumulants of r^S are of order $O(n^{-3/2})$ or smaller. Otherwise, if cumulants do not satisfy these conditions, the mean and the variance corrections improve the accuracy of this approximation only up to the second order, i.e. to the order $O(n^{-1})$.

Appendix B: the one-parameter natural exponential family

Let X be a non-degenerate random variable belonging to the one-parameter natural exponential family

$$p(x;\theta) \quad = \quad \exp\{\theta x - k(\theta) + a(x)\},\ x \in \mathbb{R}\ . \qquad (46)$$

Consider $X_1, \ldots, X_n$ independent and identically distributed random variables with density (**??**). The Hyvarinen total empirical score in this case reduces to (see for example [**?**], **?**])

$$S_H(\theta) = -\left\{2\sum_{i=1}^{n} a''(x_i) + \sum_{i=1}^{n}\left[\theta + a'(x_i)\right]^2\right\}. \qquad (47)$$

The Hyvärinen score estimator for this family is [**?**]

$$\widehat{\theta}_H = -\frac{\sum_{i=1}^{n} a'(X_i)}{n}\ ,$$

which can be computed without knowledge of $k(\theta)$; see also [**?**]. The Hyvärinen score estimator is an unbiased estimator for the parameter θ and its variance coincides with the inverse of the Godambe function $G^{-1} = \frac{1}{n} Var\,(a'(X))$. Moreover, it is easy to show that in this situation the modified signed scoring rule (**??**), the adjusted signed scoring rule (**??**), and the scoring rule score-type statistics are asymptotically equivalent up to order $O(n^{-3/2})$. In order to calculate the pivot $r^H_M(\theta)$ we need the first four derivatives of the total empirical score in (**??**), i.e.

$$s_\theta = -2 \sum_{i=1}^{n} \left[\theta + a'(x_i)\right], \qquad s_{\theta\theta} = -2n, \qquad s_{\theta\theta\theta} = s_{\theta\theta\theta\theta} = 0,$$

and the moments of these scoring rule derivatives

$$\nu_{\theta,\theta} = 4nVar\,(a'(X)) = J(\theta), \quad \nu_{\theta,\theta,\theta\theta} = -2nJ(\theta) \qquad \nu_{\theta\theta,\theta\theta} = 4n^2,$$
$$\nu_{\theta\theta} = -2n = -u_{\theta\theta} = -K(\theta), \quad \nu_{\theta\theta,\theta} = \nu_{\theta,\theta\theta} = \nu_{\theta,\theta\theta\theta} = 0, \quad \nu_{\theta\theta\theta} = \nu_{\theta\theta\theta\theta} = 0.$$

The quantities R, B_4 and B_6 at (**??**), (**??**) and (**??**), respectively, are null, as well as the quantities $H_{\theta\theta}$ and $H_{\theta\theta\theta}$. Then the signed scoring rule root statistic reduces to

$$r^H(\theta) = u_{\theta\theta}^{-1/2} s_\theta(\theta) + O_p(n^{-3/2}) \tag{48}$$

with $m(\theta) = 0$ and $Var_\theta(r^H) = \mu_1 + O_p(n^{-2})$, with $\mu_1 = \nu_{\theta,\theta}/u_{\theta\theta} = 2Var\,(a'(X))$. The modified signed scoring rule root statistic is

$$r^H_M(\theta) = J(\theta)^{-1/2} s_\theta(\theta) + O_p(n^{-3/2}), \tag{49}$$

which coincides with the scoring rule score-type statistic up to the order $O_p(n^{-3/2})$. Note also that the adjusted scoring rule root statistic $\frac{r^H}{\mu_1^{1/2}}$ coincides with the small-sample signed likelihood root statistic r^H_M.

References

[] Barndorff-Nielsen, O.E., 1976. Plausibility inference. Journal of the Royal Statistical Society, Series B 38, 103–131.

[] Barndorff-Nielsen, O.E., Cox, D.R., 1994. Inference and asymptotics. volume 52 of *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.

[] 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. doi:http://dx.doi.org/10.1093/biomet/85.3.549.

[] Brazzale, A., Davison, A., 2008. Accurate parametric inference for small samples. Statistical Science 23, 465–484. doi:10.1214/08-STS273.

[] Brazzale, A., Davison, A., Reid, N., 2007. Applied asymptotics: case studies in small-sample statistics. Cambridge University Press.

[] Brier, G.W., 1950. Verification of forecasts expressed in terms of probability. Monthly Weather Review 78, 1–3. doi:10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2.

[] Dawid, A.P., 1986. Probability forecasting, in: Kotz, S., Johnson, N.L., Read, C.B. (Eds.), In Encyclopedia of Statistical Sciences, Wiley-Interscience. pp. 210–218.

[] Dawid, A.P., 2007. The geometry of proper scoring rules. Ann. Instit. Statist. Math. 59, 77–93. doi:10.1007/s10463-006-0099-8.

[] 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 Tokio. pp. 22–28.

[] Dawid, A.P., Musio, M., 2014. Theory and applications of proper scoring rules. Metron 72, 169–183. doi:10.1007/s40300-014-0039-y.

[] Dawid, A.P., Musio, M., Ventura, L., 2014. Minimum scoring rule inference. ArXiv:1403.3920.

[] DiCiccio, 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, pp. 67–76. doi:10.2307/3316051.

[] Forbes, P., Lauritzen, S.L., 2013. Linear estimating equations for exponential families with application to gaussian linear concentration models. Manuscript.

[] Ghosh, A., Basu, A., 2013. Robust estimation for independent non-homogeneous observations using density power divergence with applications to linear regression. Electronic Journal of Statistics 7, 2420–2456. doi:10.1214/13-EJS847.

[] Gneiting, T., Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102, 359–378. doi:10.1198/016214506000001437.

[] Godambe, V.P., 1960. An optimum property of regular maximum likelihood estimation. The Annals of Mathematical Statistics 31, 1208–1211. doi:10.1214/aoms/1177705693.

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

[] Huber, P.J., 1981. Robust Statistics. John Wiley and Sons.

[] Huber, P.J., Ronchetti, E.M., 2009. Robust Statistics. John Wiley and Sons, Inc.

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

[] Hyvärinen, A., 2007. Some extensions of score matching. Computational Statistics and Data Analysis 51, 2499 – 2512. doi:10.1016/j.csda.2006.09.003.

[] Jin, Z., 2010. Aspects of Composite Likelihood Inference. Ph.D. thesis. Department of Statistics, University of Toronto.

[] Jurečková, J., Sen, P.K., 1996. Robust Statistical Procedures, Asymptotics and Interrelations. John Wiley and Sons, Inc.

[] Lee, S., Young, G., 2005. Parametric bootstraping with nuisance parameters. Statistics and Probability Letters 71, 143–153. doi:10.1016/j.spl.2004.10.026.

[] Machete, R., 2013. Contrasting probabilistic scoring rules. Journal of Statistical Planning and Inference 143, 1781–1790. doi:10.1016/j.jspi.2013.05.012.

[] Mameli, V., Musio, M., Dawid, A.P., 2014. Comparison of approaches to inference in stationary ar(1) models, in: Proceedings of the 47-th Scientific meeting of the Italian Statistical Society, University of Cagliari. pp. 1–9.

[] McCullagh, P., 1987. Tensor methods in statistics. Chapman and Hall, London.

[] Pace, L., Salvan, A., 1997. Principles of statistical inference from a Neo-Fisherian Perspective. World Scientific, Singapore.

[] Parry, M., Dawid, A.P., Lauritzen, S.L., 2012. Proper local scoring rules. The Annals of Statistics 40, 561–592. doi:10.1214/12-AOS971.

[] Reid, N., 2003. Asymptotics and the theory of inference. The Annals of Statistics 31, 1695–2095. doi:10.1214/aos/1074290325.

[] Ronchetti, E., Ventura, L., 2001. Between stability and higher-order asymptotics. Statistics and Computing 11, 67–73. doi:10.1023/A:1026562000322.

[] Severini, T., 2000. Likelihood Methods in Statistics. Oxford University Press, Oxford.

[] Stern, S.E., 2006. Simple and accurate one-sided inference based on a class of m-estimators. Biometrika 93(4), 973–987.

[] Tsallis, C., 1988. Possible generalization of boltzmann-gibbs statistics. Journal of Statistical Physics 52, 479–487. doi:10.1007/BF01016429.

[] van der Vaart, A.W., 1998. Asymptotic Statistics. Cambridge Unversity Press.

[] Varin, C., 2008. On composite marginal likelihoods. AStA Adv. Statist. Anal. 92, 1–28. doi:10.1007/s10182-008-0060-7.

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