Noname manuscript No.
(will be inserted by the editor)

A class of asymptotically efficient estimators based on sample spacings

M. Ekström · S. M. Mirakhmedov · S. Rao Jammalamadaka

the date of receipt and acceptance should be inserted later

Abstract In this paper we consider general classes of estimators based on higher-order sample spacings, called the Generalized Spacings Estimators. Such classes of estimators are obtained by minimizing the Csiszár divergence between the empirical and true distributions for various convex functions, and include the "maximum spacing estimators" as well as the maximum likelihood estimators (MLEs) as special cases, and are especially useful when the latter do not exist. These results generalize several earlier studies on spacings-based estimation, by utilizing non-overlapping spacings that are of an order which increases with the sample size. These estimators are shown to be consistent as well as asymptotically normal under a fairly general set of regularity conditions. When the step size and the the number of spacings grow with the sample size, an asymptotically efficient class of estimators, called the "Minimum Power Divergence Estimators," are shown to exist. Simulation studies give further support to the performance of these asymptotically efficient estimators in finite samples, and compare well relative to the MLEs. Unlike the MLEs, some of these estimators are also shown to be quite robust under heavy contamination.

Keywords Sample spacings · Estimation · Asymptotic efficiency · Robustness

Mathematics Subject Classification (2000) 62F10 · 62Fl2

M. Ekström
Department of Statistics, USBE, Umeå University, Umeå, Sweden, and Department of Forest Resource Management, Swedish University of Agricultural Sciences, Umeå, Sweden
E-mail: Magnus.Ekstrom@umu.se

S. M. Mirakhmedov
Department of Probability and Statistics, Institute of Mathematics, National University of Uzbekistan, Tashkent, Uzbekistan

S. Rao Jammalamadaka
Department of Statistics and Applied Probability, University of California, Santa Barbara, USA

1 Introduction

Let $\{F_\theta, \theta \in \Theta\}$ be a family of absolutely continuous distribution functions on the real line, and denote the corresponding densities by $\{f_\theta, \theta \in \Theta\}$. For any convex function ϕ on the positive half real line, the quantity

$$S_\phi(\theta) = \int_{-\infty}^{\infty} \phi\left(\frac{f_\theta(x)}{f_{\theta_0}(x)}\right) f_{\theta_0}(x)dx$$

is called the ϕ-divergence between the distributions F_θ and F_{θ_0}. The ϕ-divergences, introduced by Csiszár (1963) as information-type measures, have several statistical applications including in estimation. Although Csiszár (1977) describes how this measure can also be used for discrete distributions, we are concerned with the case of absolutely continuous distributions in the present paper.

Let $\xi_1, ..., \xi_{n-1}$ be a sequence of independent and identically distributed (i.i.d.) random variables (r.v.'s) from F_{θ_0}, $\theta_0 \in \Theta$. The goal is to estimate the unknown parameter θ_0. If $\phi(x) = -\log x$, then $S_\phi(\theta)$ is known as the Kullback-Leibler divergence between F_θ and F_{θ_0}. In this case, a consistent estimator of this $S_\phi(\theta)$ is given by

$$\frac{1}{n-1}\sum_{i=1}^{n-1} \phi\left(\frac{f_\theta(\xi_i)}{f_{\theta_0}(\xi_i)}\right) = -\frac{1}{n-1}\sum_{i=1}^{n-1} \log \frac{f_\theta(\xi_i)}{f_{\theta_0}(\xi_i)}. \tag{1}$$

Minimization of this statistic with respect to θ is equivalent to maximization of the log-likelihood function, $\sum_{i=1}^{n} \log f_\theta(\xi_i)$. Thus, by finding a value of $\theta \in \Theta$ that minimizes (1), we obtain the well-known maximum likelihood estimator (MLE) of θ_0. Note, in order to minimize the right-hand side of (1) with respect to θ, we do not need to know the value of θ_0. On the other hand, for convex functions other than $\phi(x) = -\log x$, a minimization of the left-hand side of (1) with respect to θ would require the knowledge of θ_0, the parameter that is to be estimated. Thus, for general convex ϕ functions, it is not obvious how to approximate $S_\phi(\theta)$ in order to obtain a statistic that can be used for estimating the parameters.

One solution to this problem was provided by Beran (1977), who proposed that f_{θ_0} could be estimated by a suitable nonparametric estimator $\hat{f}$ (e.g. a kernel estimator) in the first stage, and in the second stage the estimator of θ_0 should be chosen as any parameter value $\theta \in \Theta$ that minimizes the approximation

$$\hat{S}_\phi(\theta) = \int_{-\infty}^{\infty} \hat{f}(x)\phi\left(\frac{f_\theta(x)}{\hat{f}(x)}\right) dx \tag{2}$$

of $S_\phi(\theta)$. In the estimation method suggested by Beran (1977), the function $\phi(x) = \frac{1}{2}|1-\sqrt{x}|^2$ was used, and this particular case of ϕ-divergence is known as the squared Hellinger distance. Read & Cressie (1988, p. 124) put forward a similar idea based on power divergences, and it should be noted that the

family of power divergences is a sub-class of the family of ϕ-divergences. The general approach of estimating parameters by minimizing a distance (or a divergence) between a nonparametric density estimate and the model density over the parameter space have been further extended by subsequent authors, and many of these procedures combine strong robustness features with asymptotic efficiency. See Basu et al. (2011) and the references therein for details.

Here we propose an alternate approach obtained by approximating $S_\phi(\theta)$, using the sample spacings. Let $\xi_{(1)} \leq ... \leq \xi_{(n-1)}$ denote the ordered sample of $\xi_1, ..., \xi_{n-1}$, and let $\xi_{(0)} = -\infty$ and $\xi_{(n)} = \infty$. For an integer $m = m(n)$, sufficiently smaller than n, we put $k = n/m$. Without loss of generality, when stating asymptotic results, we may assume that $k = k(n)$ is an integer, and define non-overlapping mth order spacings as

$$D_{j,m}(\theta) = F_\theta(\xi_{(jm)}) - F_\theta(\xi_{((j-1)m)}), \;\; j = 1, ..., k.$$

Let

$$S_{\phi,n}(\theta) = \frac{1}{k} \sum_{j=1}^{k} \phi(kD_{j,m}(\theta)), \;\; \theta \in \Theta. \tag{3}$$

In Equation (3), the reciprocal of the argument of ϕ is related to a nonparametric histogram density estimator considered in Prakasa Rao (1983, Section 2.4). More precisely, $g_n(x) = (kD_{j,m}(\theta))^{-1}$, for $F_\theta(\xi_{((j-1)m)}) \leq x < F_\theta(\xi_{(jm)})$, is an estimator of the density of the r.v. $F_\theta(\xi_1)$, i.e., of $g(x) = f_{\theta_0}(F_\theta^{-1}(x))/f_\theta(F_\theta^{-1}(x))$, where $F_\theta^{-1}(x) = \inf\{u : F_\theta(u) \geq x\}$. When both k and m increase with the sample size, then, for large n,

$$kD_{j,m}(\theta) \approx \frac{f_\theta(\xi_{(j-1)m+\lfloor m/2 \rfloor})}{f_{\theta_0}(\xi_{(j-1)m+\lfloor m/2 \rfloor})}, \;\; j = 1, ..., k, \tag{4}$$

where $\lfloor m/2 \rfloor$ is the largest integer smaller than or equal to $m/2$. Thus, intuitively, as $k, m \to \infty$ as $n \to \infty$, $S_{\phi,n}(\theta)$ should converge in probability to $S_\phi(\theta)$. Furthermore, since ϕ is a convex function, by Jensen's inequality, we have $S_\phi(\theta) \geq S_\phi(\theta_0) = \phi(1)$. This suggests that if the distribution F_θ is a smooth function in θ, an argument minimizing $S_{\phi,n}(\theta)$, $\theta \in \Theta$, should be close to the true value of θ_0, and hence be a reasonable estimator.

An argument $\theta = \hat{\theta}_{\phi,n}$ which minimizes $S_{\phi,n}(\theta)$, $\theta \in \Theta$, will be referred to as a Generalized Spacings Estimator (GSE) of θ_0. When convenient, a root of the equation $(d/d\theta)S_{\phi,n}(\theta) = 0$ will also be referred to as a GSE.

By using different functions ϕ and different values of m, we get various criteria for statistical estimation. The ideas behind this proposed family of estimation methods generalizes the ideas behind the maximum spacing (MSP) method, as introduced by Ranneby (1984); the same method was introduced from a different point of view by Cheng and Amin (1983). Ranneby derived the MSP method from an approximation of the Kullback-Leibler divergence between F_θ and F_{θ_0}, i.e., $S_\phi(\theta)$ with $\phi(x) = -\log x$, and defined the MSP estimator as any parameter value in Θ that minimizes (3) with $\phi(x) = -\log x$ and

$m = 1$. This estimator has been shown, under general conditions, to be consistent (Ranneby 1984; Shao & Hahn 1996; Ekström 1996, 1998) and asymptotically efficient (Shao & Hahn 1994; Ghosh & Jammalamadaka 2001). Based on the maximum entropy principle, Kapur & Kesavan (1992) proposed to estimate θ_0 by selecting the value of $\theta \in \Theta$ that minimizes (3) with $\phi(x) = x \log x$ and $m = 1$. With this particular choice of ϕ-function, $S_\phi(\theta)$ becomes the Kullback-Leibler divergence between F_{θ_0} and F_θ (rather than F_θ and F_{θ_0}). We refer to Ekström (2008) for a survey of estimation methods based on spacings and the Kullback-Leibler divergence. Ekström (1997, 2001) and Ghosh & Jammalamadaka (2001) considered a subclass of the estimation methods proposed in the current paper, namely the GSEs with $m = 1$. Under general regularity conditions it turns out that this subclass produces estimators that are consistent and asymptotically normal, and that the MSP estimator, which corresponds to the special case when $\phi(x) = -\log x$, has the smallest asymptotic variance in this subclass (Ekström 1997; Ghosh & Jammalamadaka 2001). Estimators based on overlapping (rather than non-overlapping) mth order spacings was considered in Ekström (1997, 2008), where small Monte Carlo studies indicated, in an asymptotic sense, that larger orders of spacings are always better (when $\phi(x)$ is a convex function other than the negative log function). Menéndez et al. (1997, 2001a, b) and Mayoral et al. (2003) consider minimum divergence estimators based on spacings that are related to our GSEs. In the asymptotics they use only a fixed number of spacings, and their results suggests that GSEs will not be asymptotically fully efficient when k in (3) is held fixed.

In the present paper it is shown that GSEs are consistent and asymptotically normal under general conditions. In contrast to the aforementioned papers, we allow both the number of spacings, k, and the order of spacings, m, to increase to infinity with the sample size. We show that if both of them do tend to infinity, then there exists a class of asymptotically efficient GSEs that we call the Minimum Power Divergence Estimators (MPDEs). In contrast, if m is held fixed, then the only asymptotically optimal GSE is the one based on $\phi(x) = -\log x$. The main results are stated in the next section, followed by a simulation study assessing (i) the performance of these estimators for different m and n and (ii) the robustness of these estimators under contamination. In the latter case we also assess a suggested data-driven rule for choosing the order of spacings, m. Detailed proofs are to be found in the Appendix and the online Supplementary Material.

2 Main results

In this section, we state the main results and the assumptions that are needed. Unless otherwise stated, it will henceforth be assumed that $\xi_1, ..., \xi_{n-1}$ are i.i.d. r.v.'s from F_{θ_0}, $\theta_0 \in \Theta$.

We will prove the consistency of the GSEs under the following assumptions:

Assumption 1 *The family of distributions* $\{F_\theta(\cdot),\ \theta \in \Theta\}$ *have common support, with continuous densities* $\{f_\theta(\cdot),\ \theta \in \Theta\}$, *and* $F_\theta(x) \neq F_{\theta_0}(x)$ *for at least one* x *when* $\theta \neq \theta_0$.

Assumption 2 *The parameter space* $\Theta \subset R$ *contains an open interval of which* θ_0 *is an interior point, and* $F_\theta(\cdot)$ *is differentiable with respect to* θ.

Assumption 3 *The function* $\phi(t)$, $t > 0$, *satisfies the following conditions:*
(a) it is strictly convex;
(b) $\min\{0, \phi(t)\}/t \to 0$ *as* $t \to \infty$;
(c) it is bounded from above by $\psi(t) = c_1(t^{-c_2} + t^{c_3})$ *for all* $t > 0$, *where* c_1, c_2, *and* c_3 *are some nonnegative constants;*
(d) it is twice differentiable.

Assumption 3 is valid for a wide class of convex functions including the following,

$$\phi(x) = \phi_\lambda(x) = \begin{cases} \lambda^{-1}(1+\lambda)^{-1}(x^{\lambda+1} - 1), & \text{if } \lambda \neq -1, 0, \\ -\log x, & \text{if } \lambda = -1, \\ x \log x, & \text{if } \lambda = 0, \end{cases} \tag{5}$$

where the cases $\lambda = -1$ and $\lambda = 0$ are given by continuity, i.e., by noting that $\lim_{\lambda \to 0}(x^\lambda - 1)/\lambda = \log x$.

Theorem 1 *Under the Assumptions 1-3, when* $m > c_2$ *is fixed, or* $m \to \infty$ *such that* $m = o(n)$, *the equation* $(d/d\theta)S_{\phi,n}(\theta) = 0$ *has a root* $\hat{\theta}_{\phi,n}$ *with a probability tending to 1 as* $n \to \infty$, *such that*

$$\hat{\theta}_{\phi,n} = \hat{\theta}_{\phi,n}(\xi_1, ..., \xi_{n-1}) \xrightarrow{p} \theta_0.$$

For the purpose of the next theorem, we will use the notation $f(x, \theta)$ for the density $f_\theta(x)$, and we denote its partial derivatives by

$$f_{ij}(x, \theta) = \frac{\partial^{i+j}}{\partial x^i \partial \theta^j} f(x, \theta).$$

Let $W_1, W_2, ...$ be a sequence of independent standard exponentially distributed r.v.'s, $G_m = W_1 + ... + W_m$, and $\bar{G}_m = m^{-1}G_m$.

We then have the following important result:

Theorem 2 *Let* $m = o(n)$. *In addition to Assumptions 1-2, assume the following conditions:*
(i) The function ϕ *is a strictly convex function and thrice continuously differentiable.*
(ii) The quantities $Var(W_1\phi'(\bar{G}_m))$, $E(W_1^2\phi''(\bar{G}_m))$, *and* $E(W_1^3\phi'''(\bar{G}_m))$ *exist and are bounded away from zero.*
(iii) The density function $f(x, \theta)$, *the inverse* $F_\theta^{-1}(x)$, *and the partial derivatives* f_{10} *and* f_{11} *are continuous in* x *at* $\theta = \theta_0$, *and* f_{01}, f_{02}, *and* f_{03} *are continuous in* x *and* θ *in an open neighborhood of* θ_0.

(iv) The Fisher information,

$$I(\theta) = \int_{-\infty}^{\infty} \left[\frac{f_{01}(x,\theta)}{f(x,\theta)} \right]^2 f(x,\theta)dx,$$

takes values in the interval $(0,\infty)$ *for* θ *in a neighborhood of* θ_0.

Then, for any consistent root $\hat{\theta}_{\phi,n}$ *of* $(d/d\theta)S_{\phi,n}(\theta) = 0$, *we have*

$$\lim_{n\to\infty} \sup_x \left| P\Big(\sqrt{n}(\hat{\theta}_{\phi,n} - \theta_0) \le x\Big) - \Phi\left(x\sqrt{\frac{I(\theta_0)}{e_m(\phi)}} \right) \right| = 0,$$

where Φ *is the standard normal cumulative distribution function,*

$$\begin{aligned} e_m(\phi) &= \sigma_\phi^2 \left(E\left(\bar{G}_m^2 \phi''(\bar{G}_m) \right) \right)^{-2}, \\ \sigma_\phi^2 &= mVar\left(\bar{G}_m \phi'(\bar{G}_m) \right) + (2m+1)\mu_m^2 - 2m\mu_m E\left(\bar{G}_m^2 \phi'(\bar{G}_m) \right), \end{aligned} \quad (6)$$

and

$$\mu_m = E\left(W_1 \phi'(\bar{G}_m) \right) = E\left(\bar{G}_m \phi'(\bar{G}_m) \right).$$

The theorems will be proved using proof methods related to those of, e.g., Lehmann & Casella (1998) for the MLE. A generalization to the multiparameter case is possible, much like in the case of maximum likelihood estimation. However, as in Lehmann & Casella (1998) for the MLE, this will require somewhat more complex assumptions and proofs and we refrain from attempting it here.

For the case $m = 1$ and by assuming that $\lim_{x\to 0} \phi'(x)x^2e^{-x} = \lim_{x\to\infty} \phi'(x)x^2e^{-x} = 0$, Ekström (1997) and Ghosh & Jammalamadaka (2001) showed that $e_m(\phi) \ge 1$, with equality if and only if $\phi(x) = a\log x + bx + c$, for some constants a, b, and c. That this inequality holds true for general m can be seen by integrating by parts, i.e., assuming that $\lim_{x\to 0} \phi'(x/m)x^{m+1}e^{-x} = \lim_{x\to\infty} \phi'(x/m)x^{m+1}e^{-x} = 0$, we get

$$\begin{aligned} &\left(E(\bar{G}_m^2 \phi''(\bar{G}_m)) \right)^2 = \left(mE(\bar{G}_m^2 \phi'(\bar{G}_m)) - (m+1)\mu_m \right)^2 \\ &= \left(m\, Cov\big(\bar{G}_m \phi'(\bar{G}_m), \bar{G}_m\big) - \mu_m \right)^2 \\ &= m^2 \left(Cov\big(\bar{G}_m \phi'(\bar{G}_m), \bar{G}_m\big) \right)^2 + (2m+1)\mu_m^2 - 2m\mu_m E(\bar{G}_m^2 \phi'(\bar{G}_m)) \\ &\le \sigma_\phi^2, \end{aligned}$$

where the inequality on the right-hand side follows by noting that

$$\left(Cov\big(\bar{G}_m \phi'(\bar{G}_m), \bar{G}_m\big) \right)^2 \le Var\big(\bar{G}_m \phi'(\bar{G}_m)\big)\, Var(\bar{G}_m)$$

and $Var(\bar{G}_m) = m^{-1}$. Hence, $e_m(\phi) = \sigma_\phi^2 \left(E(\bar{G}_m^2 \phi''(\bar{G}_m)) \right)^{-2} \ge 1$, with equality if and only if $x\phi'(x) = a + bx$ or, equivalently, if and only if $\phi(x) = a\log x + bx + c$, where $a < 0$. It should be observed that the corresponding estimator $\hat{\theta}_{\phi,n}$ does not depend on the chosen values of $a < 0$, b,

and c. Thus, without loss of generality we may choose $a = -1$ and $b = c = 0$, i.e., for m fixed, the asymptotically optimal estimator $\hat{\theta}_{\phi,n}$ is based on the function $\phi(x) = -\log x$. If however, we let $m \to \infty$, then the asymptotically optimal estimator is no longer unique. For example, let us consider the family of power divergences (Read & Cressie 1988; Pardo 2006) given by

$$T_\lambda(\theta) = \frac{1}{\lambda(\lambda+1)} \int_{-\infty}^{\infty} \left(\left(\frac{f_\theta(x)}{f_{\theta_0}(x)} \right)^{\lambda+1} - 1 \right) f_{\theta_0}(x) dx,$$

where the cases $\lambda = -1$ and $\lambda = 0$ are given by continuity, i.e.,

$$T_{-1}(\theta) = \lim_{\lambda \to -1} T_\lambda(\theta) = \int_{-\infty}^{\infty} \left(-\log \left(\frac{f_\theta(x)}{f_{\theta_0}(x)} \right) \right) f_{\theta_0}(x) dx$$

and

$$T_0(\theta) = \lim_{\lambda \to 0} T_\lambda(\theta) = \int_{-\infty}^{\infty} \left(\frac{f_\theta(x)}{f_{\theta_0}(x)} \log \left(\frac{f_\theta(x)}{f_{\theta_0}(x)} \right) \right) f_{\theta_0}(x) dx,$$

where $T_{-1}(\theta)$ is the Kullback-Leibler divergence and $T_0(\theta)$ is the reversed Kullback-Leibler divergence (cf. Pardo & Pardo 2000).

If we set $\phi(x) = \phi_\lambda(x)$, which is defined in Equation (5), then note that $T_\lambda(\theta) = S_{\phi_\lambda}(\theta)$, i.e., the family of power divergences is a sub-class of the family of ϕ-divergences. The divergence $T_\lambda(\theta) = S_{\phi_\lambda}(\theta)$ is estimated by $S_{\phi_\lambda,n}(\theta)$, and Theorem 1 establishes the existence of a consistent root of the equation $(d/d\theta)S_{\phi_\lambda,n}(\theta) = 0$. The next result asserts that any such sequence is asymptotically normal and efficient when $k, m \to \infty$ as $n \to \infty$.

Corollary 1 *Let $m \to \infty$ such that $m = o(n)$. Suppose that Assumptions 1-2 and conditions (iii) and (iv) of Theorem 2 hold true, and that the function ϕ_λ is defined by (5) for some $\lambda \in (-\infty, \infty)$. Then, for any consistent root $\hat{\theta}_{\lambda,n}$ of $(d/d\theta)S_{\phi_\lambda,n}(\theta) = 0$, we have*

$$\sqrt{n}(\hat{\theta}_{\lambda,n} - \theta_0) \xrightarrow{d} N\left(0, \frac{1}{I(\theta_0)}\right) \text{ as } n \to \infty.$$

Such a sequence $\hat{\theta}_{\lambda,n}$ of roots is typically provided by $\arg\min_{\theta \in \Theta} S_{\phi_\lambda,n}(\theta)$, and in this case the estimator may be referred to as a Minimum Power Divergence Estimator (MPDE). Another family of divergence measures was provided by Rényi (1961), and extended in Liese & Vajda (1987). These are given by

$$R_\alpha(\theta) = \frac{1}{\alpha(\alpha-1)} \log \int_{-\infty}^{\infty} f_\theta(x)^\alpha f_{\theta_0}(x)^{1-\alpha} dx,$$

where the cases $\alpha = 0$ and $\alpha = 1$ are given by continuity, i.e., $R_0(\theta) = \lim_{\alpha \to 0} R_\alpha(\theta) = T_{-1}(\theta)$ and $R_1(\theta) = \lim_{\alpha \to 1} R_\alpha(\theta) = T_0(\theta)$. Thus, $R_0(\theta)$ and $R_1(\theta)$ are Kullback-Leibler divergences and belong to the family of power divergences. When $\alpha \neq 0, 1$, $R_\alpha(\theta)$ may be estimated by $\alpha^{-1}(\alpha - 1)^{-1} \log S_{\phi_\alpha,n}(\theta)$, where $\phi_\alpha(x) = x^\alpha$, and if $\lambda = \alpha - 1$ then note that

$\arg\min_{\theta\in\Theta} \alpha^{-1}(\alpha - 1)^{-1} \log S_{\phi_\alpha,n}(\theta) = \arg\min_{\theta\in\Theta} S_{\phi_\lambda,n}(\theta)$. Thus, each MPDE may also be regarded as a minimum Rényi divergence estimator.

The Hellinger distance between F_θ and F_{θ_0} is given by

$$H(\theta) = \left(\frac{1}{2} \int_{-\infty}^{\infty} \left(f_\theta(x)^{1/2} - f_{\theta_0}(x)^{1/2} \right)^2 dx \right)^{1/2},$$

and may be estimated by $(1+S_{\phi,n}(\theta))^{1/2}$, where $\phi(x) = -x^{1/2}$. In this case, we have, for $\lambda = -1/2$, $\arg\min_{\theta\in\Theta}(1 + S_{\phi,n}(\theta))^{1/2} = \arg\min_{\theta\in\Theta} S_{\phi_\lambda,n}(\theta)$. Thus the MPDE with $\lambda = -1/2$ may also be referred to as a minimum Hellinger distance estimator.

It is desired to have statistical estimation procedures that perform well when an assumed parametric model is correctly specified, thereby attaining high efficiency at the assumed model. A problem is that assumed parametric assumptions are almost never literally true. Thus, in addition, it is desired to have estimation procedures that are relatively insensitive to small departures from the model assumptions, and that somewhat larger deviations from the model assumptions do not cause a "catastrophe" (Huber & Ronchetti 2009). Procedures satisfying these features are called robust. Due to its relationship with the estimator suggested by Beran (1977), we conjecture that our minimum Hellinger distance estimator, i.e. the MPDE with $\lambda = -1/2$, is robust (in the sense of Beran (1977) and Lindsay (1994)). In addition, by arguments put forward in Lindsay (1994), we conjecture that GSEs based on bounded ϕ functions are robust with respect to contaminations of the original data (cf. Mayoral et al. 2003). In the next section, we consider the MPDE with $\lambda = -1/2$ and apply Monte Carlo simulations to compare its performance with those of the MLE and the MPDEs with $\lambda = -1$ and -0.9.

3 A simulation study

In this section we explore the finite sample properties of the Minimum Power Divergence Estimators (MPDEs) $\hat{\theta}_{\lambda,n} = \arg\min_{\theta\in\Theta} S_{\phi_\lambda,n}(\theta)$, where ϕ_λ is defined in Equation (5).

First we consider estimating the mean θ of a $N(\theta, 1)$ distribution and compare the root mean square errors (RMSEs) of various MPDEs, with the RMSE of the MLE. These are shown in Fig. 1. MPDEs are computed for $\lambda = -1, -0.9$, and -0.5, and for all values of m, the order of the spacings, that are divisors of n. We define a relative RMSE of an MPDE to be its RMSE divided by the RMSE of the MLE. Each RMSE was estimated from 1,000 Monte Carlo samples with $n = 840$ from the $N(\theta, 1)$ distribution with $\theta = 0$. We present relative RMSEs for m up to 150, because when m is larger than that the relative RMSEs tend to be quite large in comparison. From Fig. 1, we see that MPDEs with λ equal to -1 or -0.9 are about as good as the MLE for comparatively small values of m (and less well for larger m). The MPDE with $\lambda = -0.5$ is not quite as good in terms of RMSE. For example, with an optimally chosen m, it

had an RMSE about 0.5% larger than that of the MLE. The simulation results indicate that the optimal choice of m is 15, 15, and 14 for $\lambda = -1, -0.9$, and -0.5, respectively (although for $\lambda = -1$ and -0.9 there are about ten other candidates, respectively, that perform almost as good).

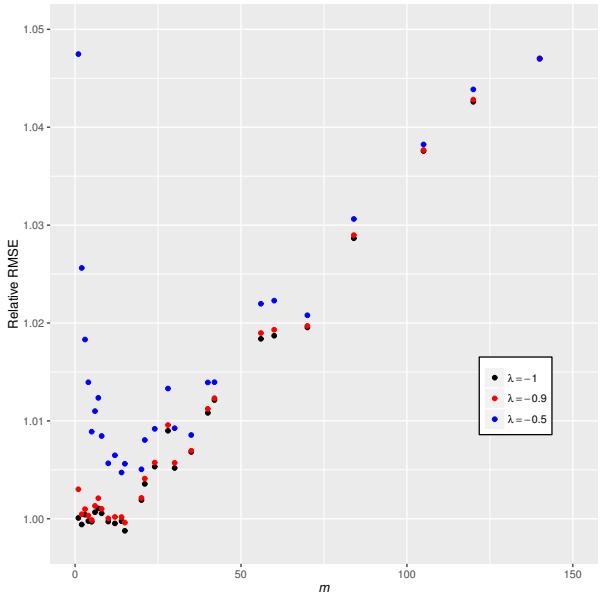

Fig. 1 RMSEs for different MPDEs relative the RMSE of the MLE when estimating the mean of a normal distribution. The relative RMSE of an MPDE is its RMSE divided by the RMSE of the MLE.

Next we consider the issue of stability/robustness of these estimators. It is known that the MSP estimator (Cheng & Amin 1983; Ranneby 1984), i.e., the MPDE with $\lambda = -1$ and $m = 1$, much like the MLE, suffers from lack of stability under even small deviations from the underlying model i.e. the distributions of the MSP and ML estimators can be greatly perturbed if the assigned model is only approximately true. This is demonstrated in a simulation study by Nordahl (1992), and by Fujisawa & Eguchi (2008) in a numerical study on the MLE. As in Nordahl (1992) we will assume a proportion $1 - \varepsilon$ of the data is generated from a normal distribution, while a proportion ε is generated by some unknown mechanism that produces "outliers." For example, measurements are made, which are 95% of the time correct, while 5% of the time operator reading/writing errors are made or the recording instrument malfunctions. Therefore, we assume that a random sample $\xi_1, ..., \xi_{n-1}$ is generated from an ε-contaminated normal distribution $G(x - \theta_0)$, where

$$G(x) = (1 - \varepsilon)\Phi(x) + \varepsilon H(x),$$

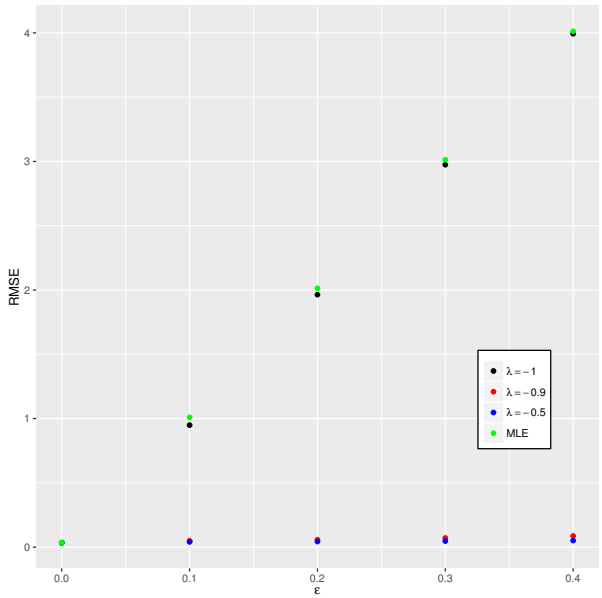

Fig. 2 RMSEs for the MLE and for different MPDEs when estimating the mean of a normal distribution under contamination, where ε denotes the level of contamination. Note, black plot symbols are often (partly) hidden behind green plot symbols, and red plot symbols behind blue ones.

in which $H(x)$ denotes an arbitrary distribution that models the outliers and ε is the level of contamination. Of interest is to estimate θ_0, the mean of the observations in the case when no recording errors occur. If $n\varepsilon$ is rather small, we may have few observations from H, making it difficult to assess the model for H. In such a case, instead of modelling the mixture distribution G, one may (wrongly) assume that all observations come from $\Phi(x-\theta)$ with $\theta = \theta_0$, and then use an estimation method that provides a good estimate of θ_0 even in presence of outliers coming from H. That is, robust estimation aims at finding an estimator $\hat{\theta}$ that efficiently estimates θ_0 even though the data is contaminated by an outlier distribution H (Fujisawa & Eguchi 2008).

In our Monte Carlo simulation, we used $H(x) = \Phi((x-\rho)/\tau)$, $\tau > 0$, with $\rho = 10$ and $\tau = 1$. For each $\varepsilon = 0.0, 0.1, ..., 0.4$, we generated 1,000 Monte Carlo samples with $n = 840$ from $G(x-\theta_0)$ with $\theta_0 = 0$, and for every sample, the MLE of θ_0 was computed using the model $F_\theta(x) = \Phi(x-\theta)$. MPDEs for this case were also computed for $\lambda = -1, -0.9$, and -0.5, and for the previously found optimal values of m for the respective values of λ. For each level of contamination, we used the 1,000 samples for computing estimated RMSEs for the respective estimators of θ_0. The resulting RMSEs are shown in Fig. 2. In case of contamination, we see that the MPDEs with λ equal to -0.9 or -0.5 are superior to the MLE and the MPDE with $\lambda = -1$. In other words, the MLE and MPDEs such as the MSP estimator, which all can be derived

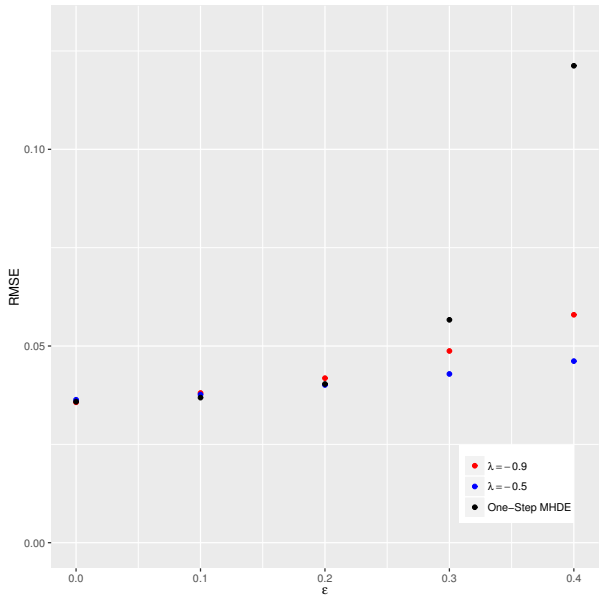

Fig. 3 RMSEs for the one-step MHDE and for two MPDEs, when estimating the mean of a normal distribution under contamination and where ε is the level of contamination. Note, red and blue plot symbols are sometimes (partly) hidden behind black plot symbols.

from the Kullback-Leibler divergence, perform poorly under contamination, and other MPDEs are to be preferred.

Looking at Fig. 1, it is clear that the choice of m, the order of spacings, is important for the quality of MPDE estimators. We now propose a data-based approach for choosing m for MPDEs (and more generally for GSEs). No asymptotic optimality is claimed for the approach. The main purpose is rather to provide sensible answers for finite sample sizes. For a given λ (or ϕ-function), let $\hat{\theta}_m$ denote the MPDE (or GSE) using the order of spacings m. The suggested approach is given by the following algorithm:

Step 1: Compute $\hat{\theta}_1$.

Step 2: For r in $1, ..., R$: Draw a bootstrap sample $x^\star_{r,1}, ..., x^\star_{r,n-1}$ from $F_{\hat{\theta}_1}$. For some set of positive integers, $\mathcal{M}$, compute $\hat{\theta}^\star_{r,m}$ for each $m \in \mathcal{M}$, where $\hat{\theta}^\star_{r,m}$ denotes the rth bootstrap replicate of $\hat{\theta}_m$.

Step 3: Choose $m_{opt} = \arg\min_{m\in\mathcal{M}} \frac{1}{R}\sum_{r=1}^{R}\left(\hat{\theta}^\star_{r,m} - \hat{\theta}_1\right)^2$.

Under the same settings as in Fig. 2 and with m chosen according to the above algorithm, with $\mathcal{M}$ defined as the set of divisors of n, we consider two MPDEs, with $\lambda = -0.9$ and -0.5, respectively. We compare these with Karunamuni & Wu's (2011) one-step minimum Hellinger distance estimator (MHDE), obtained from a one-step Newton-Raphson approximation to the

solution of the equation $\hat{S}'_\phi(\theta) = 0$, where $\hat{S}_\phi(\theta)$ is defined as in (2), with $\phi(x) = \frac{1}{2}|1 - \sqrt{x}|^2$ and $\hat{f}$ a kernel density density estimator. Karunamuni & Wu (ibid.) show that their one-step MHDE has the same asymptotic behavior as Beran's (1977) MHDE, as long as the initial estimator in the Newton-Raphson algorithm is reasonably good, and that it retains excellent robustness properties of the MHDEs. In our simulations we used the median as the initial estimator of θ_0, and the kernel estimator was based on the Epanechnikov kernel with bandwidth chosen to be $(15e)^{1/5}(\pi/32)^{1/10}\hat{\sigma}n^{-1/5}$, where $\hat{\sigma} = \text{median}\{|\xi_i - \text{median}\{\xi_j\}|\}/\Phi(3/4)$ (cf. Basu et al. 2011, pp. 108-109). The resulting RMSEs are shown in Fig. 3. When $\varepsilon = 0$, the MPDE with $\lambda = -0.9$ is the winner in terms of RMSE (in comparison with the MLE, the MPDE estimators with $\lambda = -0.9$ and -0.5 and the one-step MHDE had RMSEs that were about 0.0, 1.8, and 0.4 percent larger than that of the MLE, respectively). For $\varepsilon = 0.1$, the one-step MHDE performs somewhat better than the two MPDEs, but for $\varepsilon = 0.2, 0.3$, and 0.4, the most efficient estimator is the MPDE with $\lambda = -0.5$. Under heavy contamination, i.e. for levels of contamination equal to or larger than 0.3, both MPDEs are clearly more robust than the one-step MHDE.

By Corollary 1, MPDEs are asymptotically normal and efficient under a general set of regularity conditions. When applying an MPDE, a particular value of λ needs to be chosen. In our simulations, we considered three choices, $\lambda = -1, -0.9$, and -0.5. Much like the MLE, the MPDE with $\lambda = -1$ can be greatly perturbed if the assigned model is only approximately true. In the choice between $\lambda = -0.9$ and $\lambda = -0.5$, the former appears to provide better estimates if there is no contamination, while the latter seems to give more robust estimators when some contamination is suspected.

4 Concluding remarks

In this paper we propose classes of estimators, called Generalized Spacings Estimators or GSEs, based on non-overlapping higher order spacings and show that under some regularity conditions, they are consistent and asymptotically normal. Within these classes, we demonstrate the existence of asymptotically efficient estimators, called MPDEs. Through simulations, we demonstrate that they perform well also in moderate sample sizes relative to the MLEs. However, unlike the MLEs, some of these spacings estimators are quite robust under contamination. In this article we also propose a data-driven choice for the order of spacings, m, based on bootstrapping, and the Monte Carlo studies indicate that this practical way of choosing m leads to MPDEs which perform comparatively well and even much better at higher levels of contamination, than the one-step MHDEs proposed in the literature. Moreover, the GSEs suggested here can be suitably extended, and used in more general situations. For example, by using mth nearest neighbor balls as a multidimensional analogue to univariate mth order spacings, our proposed classes of estimators can be extended to multivariate observations (Kuljus & Ranneby (2015) studied this

problem for $m = 1$), but specifics need further exploration. Another possibility is to define GSEs based on *overlapping* spacings of order m. The derivation of the asymptotic distribution of such estimators is an open problem.

Appendix

To simplify the notation in the proofs, we will write $S(\theta)$, $S_n(\theta)$, and $\hat{\theta}_n$ rather than $S_\phi(\theta)$, $S_{\phi,n}(\theta)$, and $\hat{\theta}_{\phi,n}$, respectively, and when $\theta = \theta_0$, we use the simplified notations $S_n = S_n(\theta_0)$ and $D_{j,m} = D_{j,m}(\theta_0)$. It should be noted that $D_{j,m}$, $j = 1, ..., k$, are distributed as non-overlapping mth order spacings from a uniform distribution.

Recall that $W_1, W_2, ...$ are independent standard exponentially distributed r.v.'s. Let $G_{j,m} = W_{(j-1)m+1} + ... + W_{jm}$ and $\bar{G}_{j,m} = m^{-1}G_{j,m}$, $j = 1, ..., k$. Note that $G_{1,m}, ..., G_{k,m}$ are i.i.d. gamma r.v.'s. To keep the notation simple, we denote $G_m = G_{1,m}$ and $\bar{G}_m = m^{-1}G_m$.

Lemma 1 *(Holst & Rao 1980) Let $\varphi(u)$, defined on $(0, 1)$, be continuous, except possibly for finitely many u, and bounded in absolute value by an integrable function. Then,*

$$\frac{1}{n}\sum_{i=1}^{n}\varphi\left(\frac{i}{n+1}\right) \to \int_0^1 \varphi(u)du, \text{ as } n \to \infty.$$

Proposition 1 *Assume that $m > c_2$ is fixed or that $m \to \infty$ such that $m = o(n)$. Then, under Assumptions 1 and 3, for any fixed $\theta \neq \theta_0$,*

$$P(S_{\phi,n}(\theta) > S_{\phi,n}(\theta_0)) \to 1 \text{ as } n \to \infty.$$

Proof See the online Supplementary Material. □

Proof of Theorem 1 The proof is very similar to that of Theorem 3.7 in Lehmann & Casella (1998, p. 447), and therefore we omit the details. Note, however, that in Lehmann & Casella (1998), the log-likelihood needs to be replaced with $S_n(\theta)$, and the reference to Theorem 3.2 needs to be replaced with Proposition 1 of the current paper. □

Proof of Theorem 2 To keep notation simple we write $D_j(\theta)$ instead of $D_{j,1}(\theta)$. Note that

$$D_{j,m}(\theta) = D_{(j-1)m+1}(\theta) + D_{(j-1)m+2}(\theta) + ... + D_{jm}(\theta). \tag{7}$$

Under condition (i) we have

$$0 = S_n'(\hat{\theta}_n) = S_n'(\theta_0) + (\hat{\theta}_n - \theta_0)S_n''(\theta_0) + 2^{-1}(\hat{\theta}_n - \theta_0)^2 S_n'''(\tilde{\theta}_0),$$

where $\tilde{\theta}_0$ lies between $\hat{\theta}_n$ and θ_0. Hence,

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{\sqrt{n}S_n'(\theta_0)}{-S_n''(\theta_0) - 2^{-1}(\hat{\theta}_n - \theta_0)S_n'''(\tilde{\theta}_0)} \tag{8}$$

Set $g_\theta(u) = f_{01}(F_{\theta_0}^{-1}(u),\theta)/f\left(F_{\theta_0}^{-1}(u),\theta\right)$, and let $\tilde{U}_j$ denote a value in the interval $I_j = (F_{\theta_0}(X_{(j-1)}), F_{\theta_0}(X_{(j)}))$. We have,

$$\begin{aligned}\sqrt{n}S_n'(\theta_0) &= \frac{\sqrt{m}}{\sqrt{k}}\sum_{j=1}^{k}\sum_{l=1}^{m} kD_{(j-1)m+l}\phi'(kD_{j,m})\frac{D'_{(j-1)m+l}(\theta_0)}{D_{(j-1)m+l}} \\ &= \frac{\sqrt{m}}{\sqrt{k}}\sum_{j=1}^{k}\sum_{l=1}^{m} kD_{(j-1)m+l}\phi'(kD_{j,m})g_{\theta_0}(\tilde{U}_{(j-1)m+l}),\end{aligned}$$

where the last equality follows by the mean value theorem. Set $u_j = (j - 1/2)/n$. From the existence of the limiting distribution of the Kolmogorov-Smirnov statistic, we have

$$|\tilde{U}_j - u_j| = O_p(n^{-1/2}) \text{ as } n \to \infty, \text{ uniformly in } j. \tag{9}$$

Keeping this in the mind and that $k = n/m$, we write

$$\begin{aligned}\sqrt{n}S_n'(\theta_0) = \sqrt{m}\Bigg\{&\frac{1}{m\sqrt{k}}\sum_{j=1}^{k}\sum_{l=1}^{m}\left[nD_{(j-1)m+l}\phi'(kD_{j,m}) - \mu_{j,l,m}\right]g_{\theta_0}(u_{(j-1)m+l}) \\ &+\frac{1}{m\sqrt{k}}\sum_{j=1}^{k}\sum_{l=1}^{m}\mu_{j,l,m}g_{\theta_0}(\tilde{U}_{(j-1)m+l}) \\ &+\frac{1}{m\sqrt{k}}\sum_{j=1}^{k}\sum_{l=1}^{m}\left[nD_{(j-1)m+l}\phi'(kD_{j,m}) - \mu_{j,l,m}\right] \\ &\qquad\times\left[g_{\theta_0}(\tilde{U}_{(j-1)m+l}) - g_{\theta_0}(u_{(j-1)m+l})\right]\Bigg\} \\ = \sqrt{m}\left\{A_k + B_k + C_k\right\},&\end{aligned} \tag{10}$$

where $\mu_{j,l,m} = E\big(W_{(j-1)m+l}\phi'(\bar{G}_{j,m})\big) = \mu_m$. The summands A_k and C_k are sums of functions of m-step non-overlapping uniform spacings. Such statistics are well-studied in the literature; the results needed are given by Mirakhmedov (2005) and Mirakhmedov & Jammalamadaka (2013). We will use these facts below. Note,

$$\int_0^1 g_{\theta_0}(u)du = 0 \quad \text{and} \quad \int_0^1 g_{\theta_0}^2(u)du = I(\theta_0), \tag{11}$$

the Fisher information in a single observation. Taking this into account, and by using Corollary 1 and Remark 1 of Mirakhmedov (2005), and Lemma 1, we obtain that the limiting distribution of A_k is $N(0,\tilde{\sigma}_A^2)$, which we denote as

$$A_k \overset{d}{\sim} N(0,\tilde{\sigma}_A^2), \tag{12}$$

where

$$\tilde{\sigma}_A^2 = \frac{1}{k}\sum_{j=1}^{k} Var\left(\frac{1}{m}\sum_{l=1}^{m} W_l\phi'(\bar{G}_{j,m})g_{\theta_0}(u_{(j-1)m+l})\right) - \frac{1}{km}\left\{\frac{1}{\sqrt{k}}\sum_{j=1}^{k} Cov\left(\frac{1}{m}\sum_{l=1}^{m} W_l\phi'(\bar{G}_{j,m})g_{\theta_0}(u_{(j-1)m+l}), G_{j,m}\right)\right\}^2 = \tilde{\sigma}_{1A}^2 - \tilde{\sigma}_{2A}^2. \quad (13)$$

We have

$$\tilde{\sigma}_{1A}^2 = \frac{1}{k}\sum_{j=1}^{k} E\left(\frac{1}{m}\sum_{l=1}^{m} W_{(j-1)m+l}\phi'(\bar{G}_{j,m})g_{\theta_0}(u_{(j-1)m+l})\right)^2 - \frac{1}{k}\sum_{j=1}^{k}\left\{E\left(\frac{1}{m}\sum_{l=1}^{m} W_{(j-1)m+l}\phi'(\bar{G}_{j,m})g_{\theta_0}(u_{(j-1)m+l})\right)\right\}^2 = \tilde{\sigma}_{1,1A}^2 - \tilde{\sigma}_{2,1A}^2. \quad (14)$$

Because $m = o(n)$ and $|u_{(j-1)m+l} - j/(k+1)| = o(1)$ as $n \to \infty$, uniformly in j and l, the continuity of g_{θ_0}, the equalities (11) and Lemma 1 imply that both $k^{-1}\sum_{j=1}^{k} g_{\theta_0}^2(u_{(j-1)m+l})$ and $k^{-1}\sum_{j=1}^{k} g_{\theta_0}(u_{(j-1)m+s})g_{\theta_0}(u_{(j-1)m+l})$ converge to the same limit $I(\theta_0)$ for all $s, l = 1, ..., m$. Therefore, we obtain

$$\begin{aligned}
\tilde{\sigma}_{1,1A}^2 &= \frac{1}{k}\sum_{j=1}^{k}\frac{1}{m^2}\sum_{l=1}^{m} E\left(W_{(j-1)m+l}\phi'(\bar{G}_{j,m})\right)^2 g_{\theta_0}^2(u_{(j-1)m+l}) \\
&\quad + \frac{1}{k}\sum_{j=1}^{k}\frac{1}{m^2}\sum_{\substack{s=1\\ s\neq l}}^{k}\sum_{l=1}^{k} g_{\theta_0}(u_{(j-1)m+s})g_{\theta_0}(u_{(j-1)m+l}) \\
&\qquad \times E\left(W_{(j-1)m+s}W_{(j-1)m+l}\phi'^2(\bar{G}_{j,m})\right) \\
&= \frac{1}{m^2}\sum_{l=1}^{m}\left(E\left(W_l\phi'(\bar{G}_m)\right)^2\frac{1}{k}\sum_{j=1}^{k} g_{\theta_0}^2(u_{(j-1)m+l})\right) \\
&\quad + \frac{1}{m^2}\sum_{\substack{s=1\\ s\neq l}}^{m}\sum_{l=1}^{m} E\left(W_kW_l\phi'^2(\bar{G}_m)\right)\frac{1}{k}\sum_{j=1}^{k} g_{\theta_0}(u_{(j-1)m+k})g_{\theta_0}(u_{(j-1)m+l}) \\
&= I(\theta_0)\frac{1}{m^2}\left\{\sum_{s=1}^{m} E\left(W_s\phi'(\bar{G}_m)\right)^2 + \sum_{\substack{s=1\\ s\neq l}}^{m}\sum_{l=1}^{m} E\left(W_sW_l\phi'^2(\bar{G}_m)\right)\right\}(1+o(1)) \\
&= I(\theta_0)\frac{1}{m^2}E\left\{\phi'^2(\bar{G}_m)\left(\sum_{s=1}^{m} W_s^2 + \sum_{\substack{s=1\\ s\neq l}}^{m}\sum_{l=1}^{m} W_sW_l\right)\right\}(1+o(1)) \\
&= I(\theta_0)E\left(\bar{G}_m\phi'(\bar{G}_m)\right)^2(1+o(1)). \quad (15)
\end{aligned}$$

Similar arguments give

$$\sigma_{2,1,A}^2 = I(\theta_0)\left(\frac{1}{m}\sum_{k=1}^{m} E(W_k\phi'(\bar{G}_m))\right)^2 (1+o(1))$$
$$= I(\theta_0)\left(E(\bar{G}_m\phi'(\bar{G}_m))\right)^2 (1+o(1)). \tag{16}$$

By substituting (15) and (16) into (14) we get

$$\sigma_{1A}^2 = I(\theta_0)Var\big(\bar{G}_m\phi'(\bar{G}_m)\big)\,(1+o(1)). \tag{17}$$

Next, we have

$$\sigma_{2A}^2 = \frac{1}{m}\left\{\frac{1}{k}\sum_{j=1}^{k} Cov\left(\frac{1}{m}\sum_{l=1}^{m} W_l\phi'(\bar{G}_m)g_{\theta_0}(u_{(j-1)m+k}), G_{j,m}\right)\right\}^2$$
$$= \frac{1}{m}\left(\frac{1}{m}\sum_{l=1}^{m} Cov\left(W_l\phi'(\bar{G}_m), G_m\right)\frac{1}{k}\sum_{j=1}^{k} g_{\theta_0}(u_{(j-1)m+l})\right)^2 = o(1),$$

by Lemma 1 and the first equality in (11). This together with (17) and (13) conclude that $\tilde{\sigma}_A^2 = I(\theta_0)Var\left(\bar{G}_m\phi'(\bar{G}_m)\right)$. Thus, by (12),

$$A_k \overset{d}{\sim} N\left(0, \sigma_A^2 I(\theta_0)\right), \tag{18}$$

where $\sigma_A^2 = Var\left(\bar{G}_m\phi'(\bar{G}_m)\right)$. Let $U_{(j)} = F_{\theta_0}(\xi_{(j)})$ be the order statistics of $U_j = F_{\theta_0}(\xi_j)$. As in (9), $|U_{(j)} - u_j| = O_p(n^{-1/2})$ as $n \to \infty$, uniformly in j, implying that $|\tilde{U}_{(j)} - U_{(j)}| = O_p(n^{-1/2})$ as $n \to \infty$, uniformly in j. This, together with the continuity of g_{θ_0} and the fact that $\mu_{j,l,m} = \mu_m$ for all j and l, imply that B_k has the same asymptotic distribution as

$$\frac{1}{m\sqrt{k}}\sum_{j=1}^{k}\sum_{l=1}^{m}\mu_{j,l,m}g_{\theta_0}(U_{((j-1)m+l)}) = \frac{\mu_m}{m\sqrt{k}}\sum_{j=1}^{k}\sum_{l=1}^{m} g_{\theta_0}(U_{(j-1)m+l}),$$

which is a sum of independent r.v.'s. Hence, by (11) and the central limit theorem,

$$B_k \overset{d}{\sim} N\big(0, m^{-1}I(\theta_0)\mu_m^2\big). \tag{19}$$

By the asymptotic normality of a sum of functions of uniform m-spacings (Corollary 1 in Mirakhmedov (2005)), the continuity of g_{θ_0}, and (9), we obtain

$$C_k \xrightarrow{p} 0. \tag{20}$$

Next, consider $Cov\,(A_k, B_k)$. We shall use arguments like those on p. 39 in Ghosh (1997). That is, we use a two term Taylor expansion for $g_{\theta_0}(\tilde{U}_{(j-1)m+s})$ at $u_{(j-1)m+s}$, Theorem 2.2(1) of Mirakhmedov & Jammalamadaka (2013), and Lemma 1. Also, note that the r.v.'s $\tilde{U}_{(j-1)m+s}$ and $\tilde{U}_{(i-1)m+s}$, $s = 1, ..., m$, with $j \neq i$, are asymptotically independent, because the intervals $I_{(j-1)m+s}$ and

$I_{(i-1)m+s}$ are mutually exclusive. Then, by taking (9) and (11) into account, and after some long and tedious algebra,

$$
\begin{aligned}
&Cov(A_k, B_k)\\
&= \frac{1}{m^2 k}\sum_{j=1}^{k}\sum_{i=1}^{k}\sum_{l=1}^{m}\sum_{s=1}^{m} Cov\Big(nD_{(j-1)m+l}\phi'\Big(\frac{n}{m}D_{j,m}\Big)g_{\theta_0}(u_{(j-1)m+l}),\\
&\qquad\qquad \mu_{j,s,m}g_{\theta_0}(\tilde{U}_{(i-1)m+s})\Big)\\
&= \frac{\mu_m}{m^2 k}\sum_{j=1}^{k}\sum_{l=1}^{m}\sum_{s=1}^{m}\Big\{Cov\Big(nD_{(j-1)m+l}\phi'\Big(\frac{n}{m}D_{j,m}\Big)g_{\theta_0}(u_{(j-1)m+l}), \tilde{U}_{(j-1)m+s}\Big)\\
&\qquad\qquad + o(1)\Big\}\\
&= -I(\theta_0)\frac{\mu_m}{m^2}\sum_{l=1}^{m}\sum_{s=1}^{m} Cov(W_l\phi'(\bar{G}_m), W_s) + o(1)\\
&= -I(\theta_0)\mu_m Cov(\bar{G}_m\phi'(\bar{G}_m), \bar{G}_m) + o(1)\\
&= -I(\theta_0)\mu_m E\big(\bar{G}_m^2\phi'(\bar{G}_m)\big) + I(\theta_0)\mu_m^2 + o(1), \qquad (21)
\end{aligned}
$$

since $E\bar{G}_m = 1$. Thus, from (10), (13), (18), (19), (20) and (21) we obtain

$$\sqrt{n}S_n'(\theta_0) \overset{d}{\sim} N(0, I(\theta_0)\sigma_\phi^2), \qquad (22)$$

where σ_ϕ^2 is defined in (6). Let us consider the denominator of (8). Write

$$
\begin{aligned}
S_n''(\theta_0) = \frac{1}{k}\sum_{j=1}^{k}\phi''(kD_{j,m})\left(\sum_{l=1}^{m} kD'_{(j-1)m+l}(\theta_0)\right)^2\\
+\frac{1}{k}\sum_{j=1}^{k}\phi'(kD_{j,m})\sum_{l=1}^{m} kD''_{(j-1)m+l}(\theta_0) = \Delta_k + \nabla_k. \qquad (23)
\end{aligned}
$$

We have

$$
\begin{aligned}
\Delta_k = &\frac{1}{k}\sum_{j=1}^{k}\phi''(kD_{j,m})\sum_{s=1}^{m}\big(kD_{(j-1)m+s}\big)^2\left[\frac{D'_{(j-1)m+s}(\theta_0)}{D_{(j-1)m+s}}\right]^2 + \frac{1}{k}\sum_{j=1}^{k}\phi''(kD_{j,m})\\
&\times\sum_{\substack{s=1\\ s\neq l}}^{m}\sum_{l=1}^{m} kD_{(j-1)m+s}kD_{(j-1)m+l}\frac{D'_{(j-1)m+s}(\theta_0)}{D_{(j-1)m+s}}\frac{D'_{(j-1)m+l}(\theta_0)}{D_{(j-1)m+l}}\\
= &\Delta_{1k} + \Delta_{2k}. \qquad (24)
\end{aligned}
$$

By the mean value theorem,

$$\begin{aligned}\Delta_{1k} &= \frac{1}{k}\sum_{j=1}^{k}\phi''(kD_{j,m})\sum_{s=1}^{m}k^2D_{(j-1)m+s}^2 g_{\theta_0}^2(\tilde{U}_{(j-1)m+s}) \\ &= \frac{1}{k}\sum_{j=1}^{k}\phi''(kD_{j,m})\sum_{s=1}^{m}k^2D_{(j-1)m+s}^2 g_{\theta_0}^2(u_{(j-1)m+s}) \\ &\quad + \frac{1}{k}\sum_{j=1}^{k}\phi''(kD_{j,m})\sum_{s=1}^{m}k^2D_{(j-1)m+s}^2\Big(g_{\theta_0}^2(\tilde{U}_{(j-1)m+s}) - g_{\theta_0}^2(u_{(j-1)m+s})\Big).\end{aligned}$$

The second term here tends to zero in probability due to continuity of the function g_{θ_0} and (9). For the first term the central limit theorem (Corollary 1 and Remark 1 of Mirakhmedov (2005)) is valid with asymptotical expectation

$$\begin{aligned}&\frac{1}{k}\sum_{j=1}^{k}E\left(\phi''(\bar{G}_m)\frac{1}{m^2}\sum_{s=1}^{m}W_s^2 g_{\theta_0}^2(u_{(j-1)m+s})\right) \\ &= \frac{1}{m^2}\sum_{k=1}^{m}E(W_k^2\phi''(\bar{G}_m))\frac{1}{k}\sum_{j=1}^{k}g_{\theta_0}^2(u_{(j-1)m+k}) \\ &= I(\theta_0)\frac{1}{m^2}\sum_{s=1}^{m}E(W_s^2\phi''(\bar{G}_m))\,(1+o(1))\,,\end{aligned}$$

because of Lemma 1 and (11). Hence,

$$\Delta_{1k} \overset{p}{\sim} I(\theta_0)E\left(\phi''(\bar{G}_m)\frac{1}{m^2}\sum_{k=1}^{m}E(W_k^2)\right).$$

By using the same arguments one can show that

$$\Delta_{2k} \overset{p}{\sim} I(\theta_0)E\left(\phi''(\bar{G}_m)\frac{1}{m^2}\sum_{\substack{s=1\\ s\neq l}}^{m}\sum_{l=1}^{m}W_sW_l\right).$$

Thus, putting last two relations into (24) we obtain

$$\Delta_k \overset{p}{\sim} I(\theta_0)E\left(\bar{G}_m^2\phi''(\bar{G}_m)\right). \tag{25}$$

Consider ∇_k. By noting that

$$D''_{(j-1)m+l}(\theta_0) = D_{(j-1)m+l}\left\{\left(\frac{D'_{(j-1)m+l}(\theta_0)}{D_{(j-1)m+l}}\right)' + \left(\frac{D'_{(j-1)m+l}(\theta_0)}{D_{(j-1)m+l}}\right)^2\right\},$$

and by using Lemma 1, we see that

$$\nabla_k \overset{p}{\sim} E\bar{G}_m\phi'\left(\bar{G}_m\right)\left(\int_0^1 g'_{\theta_0}(u)du + \int_0^1 g_{\theta_0}^2(u)du\right).$$

On the other hand, it is easy to see that $\int_0^1 g'_{\theta_0}(u)du = -I(\theta_0)$, and therefore, by (11), $\nabla_k \xrightarrow{p} 0$. This fact, together with (23) and (25), imply that

$$S''_n(\theta_0) \overset{p}{\sim} I(\theta_0)E\left(\bar{G}_m^2\phi''(\bar{G}_m)\right). \tag{26}$$

Similar arguments show that

$$(\hat{\theta}_n - \theta_0)S'''_n(\tilde{\theta}_0) \xrightarrow{p} 0. \tag{27}$$

Theorem 2 follows from (8), (22), (26), and (27). □

Proof of Corollary 1 By straightforward algebra we find that

$$e_m(\phi_\lambda) = \frac{m}{\lambda^2}\left(\frac{\Gamma(m)\Gamma(m+2(1+\lambda))}{\Gamma^2(m+1+\lambda)} - 1 - \frac{1+2\lambda}{m}\right).$$

By Stirling's approximation formula, $\Gamma(x+1) = \sqrt{2\pi x}(x/e)^x(1+O(x^{-1}))$, and we find for large enough m that

$$e_m(\phi) = 1 + (1+\lambda)^2 m^{-1} + c_\lambda m^{-2},$$

where c_λ is a constant depending on λ only. Hence, the corollary follows from Theorem 2. □

Acknowledgements The authors thank three anonymous reviewers for their valuable constructive comments that lead to an improved version of the paper.

References

Basu A, Shioya H, Park C (2011) Statistical inference: The minimum distance approach. Chapman & Hall, Boca Raton.

Beran R (1977) Minimum Hellinger distance estimates for parametric models. Ann Statist 5:445–463.

Cheng RCH, Amin NAK (1983) Estimating parameters in continuous univariate distributions with a shifted origin. J R Statist Soc B 45:394–403.

Csiszár I (1963) Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweiz der Ergodizität von Markoffschen Ketten. Magyar Tud Akad Mat Kutató Int Közl 8:85–108.

Csiszár I (1977) Information measures: a critical survey. Transactions of the seventh Prague conference on information theory, statistical decision functions, random processes, 73–86.

Ekström M (1996) Strong consistency of the maximum spacing estimate. Theory Probab Math Statist 55:55–72.

Ekström M (1997) Maximum spacing methods and limit theorems for statistics based on spacings. Dissertation, Umeå University.

Ekström M (1998) On the consistency of the maximum spacing method. J Statist Plann Inference 70:209–224.

Ekström M (2001) Consistency of generalized maximum spacing estimates. Scand J Statist 28:343–354.

Ekström M (2008) Alternatives to maximum likelihood estimation based on spacings and the Kullback-Leibler divergence. J Statist Plann Inference 138:1778–1791.

Fujisawa H, Eguchi S (2008) Robust parameter estimation with a small bias against heavy contamination. J Multivar Anal 99:2053–2081.

Ghosh K (1997) Some contributions to inference using spacings. Dissertation, University of California, Santa Barbara.

Ghosh K, Jammalamadaka S Rao (2001) A general estimation method using spacings. J Statist Plann Inference 93:71–82.
Holst L, Rao JS (1980) Asymptotic theory for some families of two-sample nonparametric statistics. Sankhyā Ser A 42:1–28.
Huber PJ, Ronchetti EM (2009) Robust statistics, 2nd edn. Wiley, Hoboken.
Kapur JN, Kesavan HK (1992) Entropy optimization principles with applications. Academic Press, New York.
Karunamuni RJ, Wu J (2011) Minimum Hellinger distance estimation in a nonparametric mixture model. J Stat Plan Inference 139:1118–1133.
Kuljus K, Ranneby B (2015) Generalized maximum spacing estimation for multivariate observations. Scand J Stat 42:1092–1108.
Lehmann EL, Casella G (1998) Theory of point estimation, 2nd edn. Springer, New York.
Liese F, Vajda I (1987) Convex statistical distances. Teubner, Leipzig.
Lindsay, BG (1994) Efficiency versus robustness: the case for minimum Hellinger distance and related methods. Ann Stat 22:1081–1114.
Mayoral AM, Morales D, Morales J, Vajda I (2003) On efficiency of estimation and testing with data quantized to fixed number of cells. Metrika 57:1–27.
Menéndez ML, Morales D, Pardo L (1997) Maximum entropy principle and statistical inference on condensed ordered data. Statist Probab Lett 34:85–93.
Menéndez ML, Morales D, Pardo L, Vajda I (2001a) Minimum divergence estimators based on grouped data. Ann Inst Statist Math 53:277–288.
Menéndez ML, Pardo L, Vajda I (2001b) Minimum disparity estimators for discrete and continuous models. Appl Math 46:439–466.
Mirakhmedov SA (2005) Lower estimation of the remainder term in the CLT for a sum of the functions of k-Spacings. Stat Probab Lett 73:411–424.
Mirakhmedov SM, Jammalamadaka S Rao (2013) Higher-order expansions and efficiencies of tests based on spacings. J Nonparametr Stat 25:339–359.
Nordahl G (1992) A comparison of different maximum spacing estimators. Statistical research report 1992-1, University of Umeå.
Pardo L (2006) Statistical inference based on divergence measures. Chapman & Hall–CRC, Boca Raton.
Pardo MC, Pardo JA (2000) Use of Rènyi's divergence to test for the equality of the coefficient of variation. J Comput Appl Math 116:93–104.
Prakasa Rao BLS (1983) Nonparametric functional estimation. Academic Press, Orlando.
Ranneby B (1984) The maximum spacing method. An estimation method related to the maximum likelihood method. Scand J Stat 11:93–112.
Read TRC, Cressie NAC (1988) Goodness-of-fit statistics for discrete multivariate data. Springer, New York.
Rényi A (1961) On measures of entropy and information. Proceedings of the fourth Berkeley symposium on mathematical statistics and probability 1:547–561.
Shao Y, Hahn MG (1994) Maximum spacing estimates: a generalization and improvement of maximum likelihood estimates I. Probability in Banach Spaces 9:417–431.
Shao Y, Hahn MG (1999) Strong consistency of the maximum product of spacings estimates with applications in nonparametrics and in estimation of unimodal densities. Ann Inst Statist Math 51:31–49.