ELSEVIER

Journal of Statistical Planning and Inference 93 (2001) 71–82

journal of statistical planning and inference

www.elsevier.com/locate/jspi

A general estimation method using spacings

Kaushik Ghosh[a,*], S. Rao Jammalamadaka[b,1]

[a] *Department of Statistics, George Washington University, Washington, DC 20052, USA*

[b] *Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA*

Received 2 December 1999; received in revised form 1 June 2000; accepted 16 June 2000

Abstract

A general parametric estimation method which makes use of the coverage probabilities or spacings is proposed. Under some regularity conditions, it is shown that such estimators are asymptotically normal. This method generalizes the maximum spacing method of estimation that has been discussed in the literature. Furthermore, it is shown that the maximum spacing estimator is asymptotically most efficient within the subclass of spacings-based estimators under consideration. © 2001 Elsevier Science B.V. All rights reserved.

MSC: 62F10; 62F12; 62E20

Keywords: Spacings; Estimation; Maximum likelihood; Entropy; Kullback–Leibler information; Hellinger distance

1. Introduction

Let $X_1, X_2, \ldots, X_{n-1}$ be a random sample from some continuous distribution function F_θ, $\theta \in \Theta$ with support on $\mathbb{R}$. Here, the unknown parameter θ maybe a vector. In this paper, we propose a general method of estimating θ based on spacings – the gaps between successive order statistics. This spacings-based estimation procedure provides an alternative to the traditional parametric estimation methods like the method of moments, minimum χ^2, maximum likelihood (ML), etc. The estimation method that we propose here generalizes the idea contained in the maximum spacings estimator (MSPE) introduced by Cheng and Amin (1979, 1983) and independently discussed by Ranneby (1984) and enjoys similar advantages. Cheng and Amin (1983) show that in such situations as a three-parameter Gamma, Lognormal or Weibull distribution where the ML method breaks down due to unboundedness of the likelihood, the maximum

* Corresponding author. Tel.: +1-202-994-6889; fax: +1-202-994-6917.

E-mail address: ghosh@gwis2.circ.gwu.edu (K. Ghosh).

[1] Research partially supported by NSF grant DMS-9803600.

0378-3758/01/$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0378-3758(00)00160-9

spacings estimation (MSPE) method produces consistent and asymptotically efficient estimators. In situations like mixtures of normals where the MLE is known to produce inconsistent estimators, the MSP estimators are consistent (see Ranneby, 1984). Our study focuses on the properties of this general class of estimators. We discuss asymptotic normality of such spacings-based estimators in general and show that the MSPE, which corresponds to a special case, has the smallest asymptotic variance in this class.

Section 2 describes the estimation procedure with motivation and examples. Section 3 deals with the asymptotic normality of this class of estimators. In Section 4, the results of a simulation study are presented. The long proofs of Theorems 2.1 and 3.1 are given in the appendix.

Throughout this article, we use the symbols $\stackrel{\text{def}}{=}, \stackrel{\text{d}}{=}, \stackrel{\text{d}}{\to}$ and $\stackrel{\text{P}}{\to}$ to denote definition, equality in distribution, convergence in distribution and convergence in probability, respectively.

2. Generalized spacings estimator

2.1. The estimator

Suppose we have a random sample i.e., independent identically distributed (i.i.d.) observations $X_1, X_2, \dots, X_{n-1}$ from a continuous distribution with distribution function F_θ, $\theta \in \Theta \subset \mathbb{R}^d$. Just like in rank-theory, the continuity assumption eliminates the possibility of two observations being equal. Let the order statistics be denoted by $X_{(1)} < X_{(2)} < \cdots < X_{(n-1)}$. First, we construct the following "1-step" spacings:

$$D_i(\theta) = F_\theta(X_{(i)}) - F_\theta(X_{(i-1)}), \quad i = 1, \dots, n, \tag{1}$$

where $F_\theta(X_{(0)}) \stackrel{\text{def}}{=} 0$ and $F_\theta(X_{(n)}) \stackrel{\text{def}}{=} 1$. The generalized spacings estimator (GSE) of θ is defined to be the argument $\hat{\theta}$ which minimizes

$$T(\theta) \stackrel{\text{def}}{=} \sum_{i=1}^{n} h(nD_i(\theta)), \tag{2}$$

where $h : (0, \infty) \to \mathbb{R}$ is a strictly convex function.

Some standard choices of $h(\cdot)$ that have been used in the context of goodness-of-fit testing, are: $h(x) = -\log x$, $x \log x$, x^2, $-\sqrt{x}$, $1/x$ and $|x-1|$ (see Pyke, 1965; Rao and Sethuraman, 1975; Wells et al., 1993).

The maximum spacings estimator (MSPE) discussed by Cheng and Amin (1979, 1983) and Ranneby (1984) corresponds to estimating θ by maximizing the product

$$\prod_{i=1}^{n} \{F_\theta(X_{(i)}) - F_\theta(X_{(i-1)})\}$$

which corresponds to the special case of using $h(x) = -\log(x)$ in (2). See Shao and Hahn (1994) for a detailed discussion of the Maximum Spacing Method. The generalized spacings estimator (GSE) that we propose thus generalizes the idea behind the MSPE and hence the name.

The estimators are not always explicitly obtainable analytically but can always be computed through numerical methods (like the EM algorithm or Newton–Raphson method) under broad conditions.

2.2. Motivation

Let $U_1, U_2, \ldots, U_{n-1}$ be i.i.d. U$(0,1)$ and let

$$T_i = U_{(i)} - U_{(i-1)}, \quad i = 1, \ldots, n$$

with $0 \equiv U_{(0)} < U_{(1)} < \cdots < U_{(n)} \equiv 1$. These $(T_1, \ldots, T_n)$ are referred to as "uniform spacings" and form an exchangeable sequence of random variables. From well known properties of uniform spacings (see, e.g. Pyke, 1965, p. 398), we have

$$E(T_i) = \frac{1}{n}, \quad i = 1, \ldots, n. \tag{3}$$

Note that by the probability integral transformation, $\{D_i(\theta_0)\}_{i=1}^n$ defined in (1) have the same joint distribution as the uniform spacings $\{T_i\}_{i=1}^n$ where θ_0 is the true (but unknown) value of the parameter. Eq. (3) motivates us to find the value of θ such that the vector $\{D_i(\theta)\}_{i=1}^n$ is as close as possible to the vector of its expectations, viz., $\{1/n\}_{i=1}^n$.

Csiszár (1963) introduced the following class of divergence measures, called "h-divergence", between two probability distributions $F_1(\cdot)$ and $F_2(\cdot)$:

$$I_h(F_1, F_2) = \int_{\mathbb{R}} h\left(\frac{\mathrm{d}F_1(x)}{\mathrm{d}F_2(x)}\right) \mathrm{d}F_2(x),$$

where $h : (0, \infty) \to \mathbb{R}$ is a convex function with $h(1) = 0$. Corresponding to the two discrete probability distributions on n points, $F_1 = \{D_i(\theta)\}_{i=1}^n$ and $F_2 = \{1/n\}_{i=1}^n$, Csiszár's divergence measure reduces to our Eq. (2) as the function to be minimized to estimate θ. Hence, our proposed method may also be viewed as a minimum divergence method of estimation. In particular,

$$h(x) = -\log(x) \tag{4}$$

minimizes the Kullback–Leibler divergence, while

$$h(x) = x \log(x) \tag{5}$$

maximizes the entropy. One may also choose

$$h(x) = \begin{cases} x^\alpha & \text{if } \alpha > 1, \\ -x^\alpha & \text{if } 0 < \alpha < 1, \\ x^\alpha & \text{if } -\frac{1}{2} < \alpha < 0 \end{cases} \tag{6}$$

and when $\alpha = \frac{1}{2}$, this corresponds to minimizing the Hellinger distance.

Note that in (6), although $h(x) = x^\alpha$ for any $\alpha < 0$ gives rise to a valid estimation procedure, the regularity conditions of Theorems 2.1 and 3.1 are valid only when

$\alpha > -\frac{1}{2}$. This ensures that $E\{Wh'(W)\}^2$ in (9) exists, where W is an exponential r.v. with mean 1.

The following theorem, whose proof is given in the appendix, analytically justifies why minimizing (2) should bring us close to θ_0.

Theorem 2.1. *Let $T(\theta)$ be defined as in* (2) *where $h(\cdot)$ is a continuously differentiable* (*non-linear*) *convex function with $h(1) = 0$.*

Assume that $F_\theta(\cdot)$ has a continuous density $f_\theta(\cdot)$ and for $u \in (0,1)$, define

$$F_{\theta_0}^{-1}(u) = \inf\{x\colon F_{\theta_0}(x) \geqslant u\}.$$

For any $\theta \in \Theta$, assume that $l_\theta(\cdot)$ defined by

$$l_\theta(u) = \frac{f_\theta F_{\theta_0}^{-1}(u)}{f_{\theta_0} F_{\theta_0}^{-1}(u)}$$

satisfies

$$\int_0^1 E[h(Wl_\theta(u))]^2\,\mathrm{d}u < \infty \quad \textit{and} \quad \int_0^1 E[Wl_\theta(u)h'(Wl_\theta(u))]^2\,\mathrm{d}u < \infty$$

where W is an Exponential r.v. with mean 1. *Then,*

$$P_{\theta_0}(T(\theta_0) \leqslant T(\theta)) \to 1 \quad \textit{as } n \to \infty.$$

Heuristically, this theorem says that in large samples, when θ_0 is the true value, $T(\theta_0)$ tends to take the smallest value with a very high probability. Hence, it makes sense to minimize $T(\theta)$ to estimate such a θ_0.

2.3. Some examples

Example 2.1. Suppose $X_1, X_2, \ldots, X_{n-1}$ are i.i.d. observations from $\mathrm{U}(\theta - \frac{1}{2},\ \theta + \frac{1}{2})$, $\theta \in \mathbb{R}$. In this case, we know that any $\hat{\theta}$ in the interval $(X_{(n-1)} - \frac{1}{2}, X_{(1)} + \frac{1}{2})$ is an MLE, and hence is not unique. However, it can be shown after some simplification, that the GSE corresponding to any of the choices (4) to (6) is given by the midrange, $\hat{\theta} = (X_{(1)} + X_{(n-1)})/2$, which is also the UMVUE for θ.

Example 2.2. Let $X_1, X_2, \ldots, X_{n-1}$ be a random sample from $U(0,\theta), \theta \in (0,\infty)$. It is well known that the MLE is $X_{(n-1)}$. It can be checked that the estimator corresponding to $h(x) = x \log x$ as in (5) is

$$X_{(n-1)} + \exp\left\{\frac{X_{(1)}\log X_{(1)} + \sum_{i=2}^{n-1}(X_{(i)} - X_{(i-1)})\log(X_{(i)} - X_{(i-1)})}{X_{(n-1)}}\right\}$$

while that corresponding to (6) is

$$X_{(n-1)} + \left\{\frac{X_{(1)}^{\alpha} + \sum_{i=2}^{n-1}(X_{(i)} - X_{(i-1)})^{\alpha}}{X_{(n-1)}}\right\}^{1/(\alpha-1)}.$$

Finally if we take $h(\cdot) = -\log x$ as in (4), the resulting estimator (which is the MSPE) is given by

$$\frac{n}{n-1} X_{(n-1)}.$$

This is again the UMVUE for θ.

Example 2.3. Suppose we have a random sample $X_1, X_2, \ldots, X_{n-1}$ from the double exponential distribution given by the following density:

$$f_\theta(x) = \tfrac{1}{2}\exp^{-|x-\theta|}, \quad x \in \mathbb{R}; \ \theta \in \mathbb{R}.$$

For odd sample size (say, $n-1 = 2k+1$), (4) gives the same estimator as the MLE: $X_{(k)}$. For $n-1 = 2k$ (even sample size), the MLE is not unique; it is any value in the interval $(X_{(k)}, X_{(k+1)})$. The estimator corresponding to (4) in this case is the midpoint of the previous interval, i.e. $(X_{(k)} + X_{(k+1)})/2$ which again corresponds to the UMVUE.

3. Properties of GSE

For simplicity, we will assume from now on that θ is a scalar. The results given below, however, apply to the vector parameter case with obvious modifications.

3.1. Consistency

One can prove consistency of the GSEs under assumptions of existence of continuous derivative of the density function with respect to θ, along the lines of Lehmann (1983, pp. 413–414). See Ghosh (1997) for a detailed proof. The above, however, gives local consistency, since taking derivatives does not guarantee the presence of a global minimum. The same method of estimation has been independently proposed by Ranneby and Ekström (1997) in which a proof of consistency under global conditions is given. See also Ekström (1997) for further discussions.

3.2. Asymptotic normality

Here we show that under regularity conditions, the estimators obtained by using the GS method are asymptotically normal with the correct center. As should be expected, the asymptotic variance depends not only on the parent distribution of the sample, but also on the function $h(\cdot)$.

For the purpose of this theorem, we will use the notation $f(x,\theta)$ for the density $f_\theta(x)$. We will also use the notation $f_{ij}(x,\theta)$ to denote

$$\frac{\partial^{i+j}}{\partial x^i \partial \theta^j} f(x,\theta), \quad i,\ j = 0, 1, 2, \ldots .$$

Theorem 3.1. *Let $h(\cdot)$ be a convex function (except, possibly, a straight line) which is thrice continuously differentiable. Let W denote an exponential random variable with mean 1. Suppose further that the following assumptions hold:*

- *For almost all x, f_{03} is continuous in θ in an open neighborhood of θ_0.*
- *$f, f_{01}, f_{10}, f_{11}, f_{02}$ are continuous in x for $\theta=\theta_0$.*
- $$\int_0^1 \left\{ \frac{f_{02}(F_{\theta_0}^{-1}(u),\theta_0)}{f(F_{\theta_0}^{-1}(u),\theta_0)} \right\}^2 \mathrm{d}u < \infty.$$
- $E[Wh'(W)]^2<\infty$, $E[W^2h''(W)]^2<\infty, E[W^3h'''(W)]^2<\infty$.
- *The distributions have common support with finite Fisher information*

$$I(\theta)=\int_{\mathbb{R}} \left[\frac{f_{01}(x,\theta)}{f(x,\theta)}\right]^2 f(x,\theta)\,\mathrm{d}x. \tag{7}$$

Then, for any consistent root $\hat{\theta}_n$ of $T'(\theta)=0$, we have

$$\sqrt{n}(\hat{\theta}_n-\theta_0)\xrightarrow{\mathrm{d}} \mathrm{N}\left(0,\frac{\sigma_h^2}{I(\theta_0)}\right) \tag{8}$$

where

$$\sigma_h^2=\frac{E\{Wh'(W)\}^2-2E\{Wh'(W)\}\mathrm{Cov}\{Wh'(W),W\}}{[E\{W^2h''(W)\}]^2}. \tag{9}$$

Since we have a class of estimators to choose from, corresponding to different $h(\cdot)$ functions, one obvious question is: "Is there any $h(\cdot)$ that gives rise to an estimator with the smallest asymptotic variance σ_h^2 and if so, what is it?" The following theorem provides an answer:

Theorem 3.2. *Assuming that*

$$\lim_{w\to 0} w^2h'(w)\mathrm{e}^{-w}=0 \quad \textit{and} \quad \lim_{w\to\infty} w^2h'(w)\mathrm{e}^{-w}=0,$$

σ_h^2 given by (9) is minimized iff $h(x)=a\log(x)+bx+c$ where a,b and c are constants.

Proof. By integration by parts, we have

$$\begin{aligned}E\{W^2h''(W)\}&=\int_0^\infty w^2h''(w)\mathrm{e}^{-w}\,\mathrm{d}w\\&=w^2\mathrm{e}^{-w}h'(w)\Big|_0^\infty-\int_0^\infty(2w-w^2)h'(w)\mathrm{e}^{-w}\,\mathrm{d}w\\&=E\{W^2h'(W)\}-2E\{Wh'(W)\}.\end{aligned}$$

Let us write

$$N_h = \mathrm{Var}\{Wh'(W)\} + 3[E\{Wh'(W)\}]^2 - 2E\{Wh'(W)\}E\{W^2h'(W)\}$$

and

$$\begin{aligned} D_h &= [E\{W^2h'(W)\} - 2E\{Wh'(W)\}]^2 \\ &= [\mathrm{Cov}\{Wh'(W), W\}]^2 + 3[E\{Wh'(W)\}]^2 \\ &\quad -2E\{Wh'(W)\}E\{W^2h'(W)\}. \end{aligned}$$

Then, $\sigma_h^2 = N_h/D_h$.

By Cauchy–Schwarz inequality,

$$[\mathrm{Cov}\{Wh'(W), W\}]^2 \leqslant \mathrm{Var}\{Wh'(W)\}.$$

Hence, $\sigma_h^2 \geqslant 1$, equality holding if and only if

$$\begin{aligned} & wh'(w) = a + bw \quad \text{(for some constants } a \text{ and } b) \\ & \Rightarrow h(w) = a\log(w) + bw + c \quad \text{(where } c \text{ is a constant).} \qquad \square \end{aligned}$$

Since we know that $\sum_{i=1}^{n} D_i(\theta) = 1$ and that $h(\cdot)$ has to be convex, we can, without loss of generality, choose $a = -1,\ b = 0,\ c = 0$. Hence, asymptotically, the MSPE has the smallest variance among this class. This, incidentally, coincides with the Cramer–Rao lower bound, which is the asymptotic variance of the MLE when the latter exists. Hence, MSPE is asymptotically equivalent to MLE when the latter exists.

Remark. The smoothness assumptions on $h(\cdot)$ that we require, eliminate some important special cases including the estimator corresponding to $h(x) = |x-1|$. An extension of the present results would be needed and is under investigation.

4. Simulation studies

We performed a simulation study of the proposed estimators for the three-parameter Weibull distribution:

$$f(x) = \gamma\beta^{-\gamma}(x-\alpha)^{\gamma-1}\exp[-\{(x-\alpha)/\beta\}^{\gamma}]; \quad x > \alpha,\ \alpha \in \mathbb{R},\ \beta, \gamma > 0.$$

The true parameter value was chosen to be ($\alpha_0 = 0,\ \beta_0 = 1,\ \gamma_0 = 0.5$). We considered samples of size 10(10)50 with 1000 simulations in each case. Since the functions were not differentiable, the minimizations were performed using the downhill simplex method in multidimensions due to Nelder and Mead (1965). All the simulations were done using `C` routines in Press et al. (1997). The results are presented in Table 1.

It is apparent that the GSE with $h(x) = -\log(x)$ outperforms in terms of mean squared error, followed by $h(x) = x\log(x)$ and then closely by $h(x) = |x-1|$. A striking feature

Table 1
Mean-squared error (bias) of the estimators based on 1000 simulations

$h(x)$	α		β		γ	
$n=10$						
$-\log(x)$	0.0075	(−0.0021)	0.7086	(0.1976)	0.0442	(−0.0195)
x^2	275.5675	(−1.5787)	287.1084	(1.9721)	77.8794	(0.8739)
$\|x-1\|$	1.5547	(−0.1439)	3.1738	(0.5259)	3.0737	(0.2359)
$x\log(x)$	1.8297	(−0.1416)	3.5465	(0.4265)	0.8127	(0.0731)
$n=20$						
$-\log(x)$	0.0001	(0.0013)	0.2968	(0.1121)	0.0118	(−0.0301)
x^2	3.4908	(−0.1089)	4.9633	(0.3326)	0.5612	(0.0362)
$\|x-1\|$	0.0099	(−0.0127)	0.5839	(0.2395)	0.0633	(0.0513)
$x\log(x)$	0.0045	(−0.0069)	0.4374	(0.1652)	0.0331	(−0.0171)
$n=30$						
$-\log(x)$	0.0000	(0.0006)	0.1982	(0.0978)	0.0091	(−0.0170)
x^2	0.1127	(−0.0336)	0.7384	(0.1903)	0.0712	(0.0039)
$\|x-1\|$	0.0004	(−0.0030)	0.3684	(0.1709)	0.0211	(0.0298)
$x\log(x)$	0.0001	(−0.0016)	0.2581	(0.1224)	0.0132	(−0.0032)
$n=40$						
$-\log(x)$	0.0000	(0.0002)	0.1628	(0.0879)	0.0107	(−0.0063)
x^2	0.1771	(−0.0205)	0.5853	(0.1242)	0.0747	(−0.0024)
$\|x-1\|$	0.0000	(−0.0011)	0.2684	(0.1427)	0.0138	(0.0248)
$x\log(x)$	0.0031	(−0.0024)	0.2276	(0.1226)	0.0129	(0.0074)
$n=50$						
$-\log(x)$	0.0000	(0.0001)	0.1253	(0.0787)	0.0093	(−0.0058)
x^2	0.0334	(−0.0120)	0.3570	(0.1097)	0.0261	(−0.0033)
$\|x-1\|$	0.0000	(−0.0006)	0.2132	(0.1167)	0.0128	(0.0252)
$x\log(x)$	0.0000	(−0.0005)	0.1650	(0.1094)	0.0104	(0.0188)

from the simulations is the poor performance of $h(x) = x^2$ for small n. We should mention here that the ML method does not work here since $\gamma_0 < 1$.

5. Concluding remarks

A general method of estimation based on spacings which works via the cumulative distribution functions (unlike the MLEs which work through the densities) is proposed and discussed. The asymptotic properties of such estimators such as consistency and asymptotic normality are established. Some members of this class, i.e. the MSPEs have asymptotically the same efficiency as the MLEs when the latter exist.

Since the spacings depend on the observations through their CDF and not the PDF, modifications of the PDF at a countable number of points does not affect the resulting GSE, unlike the MLE.

The results that we obtained here can be generalized to the case of higher order (or m-step) spacings. It will then be of interest to decide the optimum "step-size" that provide the best estimators in such a case. This will be considered in a separate article.

Appendix A.

Lemma A.1. *Let* $L:(0,\infty)\times(0,1)\to\mathbb{R}$ *satisfy the following*:

1. *For each* y, $L(x,y)$ *is differentiable w.r.t.* x *on* $(0,\infty)$ *and the derivative* $L_{1,0}(x,y)$ *is continuous in* x *where* $L_{1,0}(x,y)=(\partial/\partial x)L(x,y)$.
2. *For each* x, $L(x,y)$ *and* $L_{1,0}(x,y)$ *are continuous in* y.
3. $\int_0^1 E\{L(W,u)\}\,\mathrm{d}u<\infty$ *and* $\int_0^1 E\{WL_{1,0}(W,u)\}\,\mathrm{d}u<\infty$.

   Then,

$$\frac{1}{n}\sum_{k=1}^{n}L(nT_k,\xi_k)\xrightarrow{\mathrm{P}}\int_0^1 E\{L(W,u)\}\,\mathrm{d}u$$

where $\xi_k\stackrel{\mathrm{def}}{=}(k-\frac{1}{2})/n;\ k=1,\ldots,n.$

Proof. Since $\{nT_k\}_{k=1}^n$ have the same distribution as independent exponential random variables divided by their mean (see Pyke, 1965), on a common probability space we can construct i.i.d. $\{W_k\}_{k=1}^n$ such that $W_k\sim\mathrm{Exp}(1)$ and $\{W_k/\bar{W}\}_{k=1}^n\stackrel{\mathrm{d}}{=}\{nT_k\}_{k=1}^n$. Then, by a Taylor expansion of $\bar{W}$ around its mean 1, we obtain

$$\begin{aligned}\frac{1}{n}\sum_{k=1}^{n}L(nT_k,\xi_k)&\stackrel{\mathrm{d}}{=}\frac{1}{n}\sum_{k=1}^{n}L\left(\frac{W_k}{\bar{W}},\xi_k\right)\\&=\frac{1}{n}\sum_{k=1}^{n}\Bigg[L(W_k,\xi_k)\\&\quad-\frac{(\bar{W}-1)W_k}{\{1+\theta_{kn}(\bar{W}-1)\}^2}L_{1,0}\left(\frac{W_k}{1+\theta_{kn}(\bar{W}-1)},\xi_k\right)\Bigg],\end{aligned}\tag{A.1}$$

where $0\leqslant\theta_{kn}\leqslant 1$. We show that the first term converges to the desired limit while the second term goes to zero in probability. Now,

$$\begin{aligned}&\left|\frac{1}{n}\sum_{k=1}^{n}L(W_k,\xi_k)-\int_0^1 E\{L(W,u)\}\,\mathrm{d}u\right|\\&\quad\leqslant\left|\frac{1}{n}\sum_{k=1}^{n}[L(W_k,\xi_k)-E\{L(W,\xi_k)\}]\right|\\&\qquad+\left|\frac{1}{n}\sum_{k=1}^{n}E\{L(W,\xi_k)\}-\int_0^1 E\{L(W,u)\}\,\mathrm{d}u\right|.\end{aligned}$$

Since W_k are i.i.d., the first of the terms on the RHS of the inequality converge to zero in probability by Kolomogorov's SLLN (see Shiryayev, 1984, p. 366) and the second term converges to zero by the continuity of $L(x,y)$ in y. As for the second term in (A.1), note that $L_{1,0}(x,y)$ is continuous in x and that $(\bar{W}-1)=\mathrm{O_p}(n^{-1/2})$, so that it is equal to

$$=\frac{1}{n}\sum_{k=1}^{n}[(\bar{W}-1)W_kL_{1,0}(W_k,\xi_k)]+o_p(1).$$

By the same argument as before, it can be shown that

$$\frac{1}{n}\sum_{k=1}^{n}W_kL_{1,0}(W_k,\xi_k)\xrightarrow{P}\int_0^1 E\{WL_{1,0}(W,u)\}\,du.$$

But since this term is multiplied by $(\bar{W}-1)=o_p(1)$, it gives us the desired result. □

Proof of Theorem 2.1. Fix any $\theta\in\Theta$. Then, by Mean Value Theorem,

$$\frac{1}{n}[T(\theta)-T(\theta_0)]\stackrel{d}{=}\frac{1}{n}\sum_{i=1}^{n}\left[h\left\{nT_i\frac{f_\theta F_{\theta_0}^{-1}(\tilde{U}_i)}{f_{\theta_0}F_{\theta_0}^{-1}(\tilde{U}_i)}\right\}-h(nT_i)\right], \tag{A.2}$$

where $\tilde{U}_i\in(U_{(i-1)},U_{(i)})$.

From the existence of the limiting distribution of the Kolmogorov–Smirnov statistic, we have

$$|\tilde{U}_i-\xi_i|\xrightarrow{P}0\quad\text{as } n\to\infty \text{ uniformly in } i. \tag{A.3}$$

Since $h(\cdot)$ is continous, we have

$$h\left\{nT_i\frac{f_\theta F_{\theta_0}^{-1}(\tilde{U}_i)}{f_{\theta_0}F_{\theta_0}^{-1}(\tilde{U}_i)}\right\}=h\left\{nT_i\frac{f_\theta F_{\theta_0}^{-1}(\xi_i)}{f_{\theta_0}F_{\theta_0}^{-1}(\xi_i)}\right\}+o_p(1), \tag{A.4}$$

where $o_p(1)$ is uniform in i. We should remark here that the behavior near 0 and 1 is not crucial to such uniformity arguments for the following reason: It is known that the empirical process of the "normalized spacings", $\{nD_j\}_{j=1}^n$ converges to a certain Gaussian Process (see Rao and Sethuraman, 1975). By invoking the law of iterated logarithm for this process, it can be concluded that it suffices if the boundedness and continuity requirements hold in an interval of values of u away from 0 and 1, so that (A.4) and later on (A.7) are valid. See Sethuraman and Rao (1970) for more details.

Hence, applying Lemma A.1, we have from (A.2)

$$\begin{aligned}\frac{1}{n}[T(\theta)-T(\theta_0)]&\xrightarrow{P}\int_0^1\left[Eh\left(W\frac{f_\theta F_{\theta_0}^{-1}(u)}{f_{\theta_0}F_{\theta_0}^{-1}(u)}\right)-Eh(W)\right]du\\&\geqslant 0\quad\text{(by Jensen's inequality)}\end{aligned}$$

with the equality holding if $\theta=\theta_0$. □

Proof of Theorem 3.1. By the assumptions, $T(\theta)$ is thrice differentiable in a neighborhood of θ_0. Applying a Taylor expansion, we have

$$\sqrt{n}(\hat{\theta}_n-\theta_0)=\frac{(1/\sqrt{n})T'(\theta_0)}{-(1/n)T''(\theta_0)-[(\hat{\theta}_n-\theta_0)/2n]T'''(\theta_n^*)} \tag{A.5}$$

where θ_n^* lies between $\hat{\theta}_n$ and θ_0. Writing $g_\theta(u) = f_{01}(F_{\theta_0}^{-1}(u), \theta)/f(F_{\theta_0}^{-1}(u), \theta)$ and $\tilde{U}_j \in (F_{\theta_0}(X_{(j-1)}), F_{\theta_0}(X_{(j)}))$, we have

$$\begin{aligned}\frac{1}{\sqrt{n}} T'(\theta_0) &= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [nD_j(\theta_0) h'\{nD_j(\theta_0)\} - \mu] g_{\theta_0}(\xi_j) + \frac{\mu}{\sqrt{n}} \sum_{j=1}^{n} g_{\theta_0}(\tilde{U}_j) \\ &\quad + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} [nD_j(\theta_0) h'\{nD_j(\theta_0)\} - \mu] \times \{g_{\theta_0}(\tilde{U}_j) - g_{\theta_0}(\xi_j)\} \\ &= A_n + B_n + C_n \quad \text{(say)} \end{aligned} \tag{A.6}$$

where $\mu = E[Wh'(W)]$. It can be checked that A_n and B_n can both be represented as functionals of the same empirical process, either of the uniform random variables (see Shorack and Wellner, 1986) or of the normalized spacings (see Rao and Sethuraman, 1975). While this is useful to establish the asymptotic joint normality of A_n and B_n, computation of the asymptotic variances and covariances is easier done directly as we do below. Indeed, by using arguments similar to those in Corollary 2.1 of Holst and Rao (1981), we can show that $A_n \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma_A^2)$ where

$$\sigma_A^2 = \mathrm{Var}\{Wh'(W)\} I(\theta_0)$$

and $I(\theta_0)$ defined in (7) is the "Fisher Information" in a single observation.

From (A.3) and the Central Limit Theorem,

$$B_n \xrightarrow{d} \mathrm{N}(0, \mu^2 I(\theta_0)).$$

By continuity of g_{θ_0} and from (A.3),

$$C_n \xrightarrow{\mathrm{P}} 0. \tag{A.7}$$

Following the steps as in p. 39 of Ghosh (1997), we can also show

$$\lim_{n\to\infty} \mathrm{Cov}(A_n, B_n) = -\mu\, \mathrm{Cov}(Wh'(W), W) I(\theta_0).$$

Hence,

$$A_n + B_n + C_n \xrightarrow{\mathrm{d}} \mathrm{N}(0, \sigma_1^2)$$

where

$$\sigma_1^2 = [\mathrm{Var}\{Wh'(W)\} + [E\{Wh'(W)\}]^2 - 2E\{Wh'(W)\}\mathrm{Cov}\{Wh'(W), W\}] I(\theta_0).$$

We consider next the denominator of (A.5). Using arguments similar to those in the proofs of Theorem 2.1 and Lemma A.1, it can be shown that

$$\frac{1}{n} T''(\theta_0) \xrightarrow{\mathrm{P}} \mu_1 I(\theta_0),$$

where $\mu_1 = E\{W^2 h''(W)\}$. Finally, the second term in the denominator

$$\frac{1}{n}(\hat{\theta}_n - \theta_0) T'''(\theta_n^\star) \xrightarrow{\mathrm{P}} 0$$

by the continuity of f_{03} with respect to θ in a neighborhood of θ_0 and the consistency of $\hat{\theta}_n$ and the use of Lemma A.1. Combining all these results, we get

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathrm{d}} \mathrm{N}\left(0, \frac{\sigma_h^2}{I(\theta_0)}\right). \qquad \square$$

References

Cheng, R.C.H., Amin, N.A.K., 1979. Maximum product of spacings estimation with application to the lognormal distribution. Math Report 79-1, Department of Mathematics, UWIST, Cardiff.

Cheng, R.C.H., Amin, N.A.K., 1983. Estimating parameters in continuous univariate distributions with a shifted origin. J. Roy. Statist. Soc. Ser. B 45, 394–403.

Csiszár, I., 1963. Eine Informationstheoretische Ungleichung und ihre Anwendung auf den Beweiz der Ergodizitāt von Markoffschen Ketten. Magyar Tud. Akad. Mat. Kutató Int. Kōzl 8, 85–108.

Ekström, M., 1997. Maximum spacing methods and limit theorems for statistics based on spacings. Ph.D. Thesis, Department of Mathematical Statistics, Umeå University, Sweden.

Ghosh, K., 1997. Some contributions to inference using spacings, Ph.D. Thesis, Department of Statistics and Applied Probability, University of California, Santa Barbara, USA.

Holst, L., Rao, J.S., 1981. Asymptotic spacings theory with applications to the two sample problem. Canad. J. Statist. 9 (1), 79–89.

Lehmann, E.L., 1983. Theory of Point Estimation. Wiley, New York.

Nelder, J.A., Mead, R., 1965. Comput. J. 7 308–313.

Press, W.H. et al., 1997. Numerical Recipes in C, 2nd Edition. Cambridge University Press, Cambridge.

Pyke, R., 1965. Spacings (with discussion). J. Roy. Statist. Soc. Ser. B 27, 395–449.

Ranneby, B., 1984. The maximum spacing method: an estimation method related to the maximum likelihood method. Scand. J. Statist. 11, 93–112.

Ranneby, B., Ekström, M., 1997. Maximum spacing estimates based on different metrics. Preprint.

Rao, J.S., Sethuraman, J., 1975. Weak convergence of empirical distribution functions of random variables subject to perturbations and scale factors. Ann. Statist. 3, 299–313.

Sethuraman, J., Rao, J.S., 1970. Pitman efficiencies of tests based on spacings. In: Puri, M.L. (Ed.), Nonparametric Techniques in Statistical Inference. Cambridge University Press, Cambridge, pp. 405–415.

Shao, Y., Hahn, M.G., 1994. Maximum spacing estimates: a generalization and improvement of maximum likelihood estimates. Probab. Banach Spaces 9, 417–431.

Shiryayev, A.N., 1984. Probability. Springer, New York.

Shorack, G.R., Wellner, J.A., 1986. Empirical Processes with Applications to Statistics. Wiley, New York.

Wells, M.T. et al., 1993. Large sample theory of spacings statistics for tests of fit for composite hypothesis. J. Roy. Statist. Soc. Ser. B 55 (1), 189–203.