Kybernetika

Dietmar Ferger; John Venz

Density estimation via best L^2-approximation on classes of step functions

Kybernetika, Vol. 53 (2017), No. 2, 198–219

Persistent URL: http://dml.cz/dmlcz/146801

KYBERNETIKA — VOLUME *53* (2017), NUMBER 2, PAGES 198–219

DENSITY ESTIMATION VIA BEST L^2-APPROXIMATION ON CLASSES OF STEP FUNCTIONS

Dietmar Ferger and John Venz

We establish consistent estimators of jump positions and jump altitudes of a multi-level step function that is the best L^2-approximation of a probability density function f. If f itself is a step-function the number of jumps may be unknown.

Keywords: argmin-theorem, density estimation, step functions, martingale inequalities, multivariate cadlag stochastic processes

Classification: 62F10, 62G07, 60G44

1. INTRODUCTION

Let $X_1, \dots, X_n$ be i.i.d. random variables defined on a probability space $(\Omega, \mathfrak{A}, P)$ with values in a compact interval $[l, r]$ and with bounded probability density function (pdf) $f : [l, r] \to \mathbb{R}$ and cumulative distribution function (cdf) F. Furthermore, let there exist a unique step function with $d + 1$ levels and domain $[l, r]$, that has minimal L^2-distance to the pdf f. We call this step function a *best approximation* of the pdf f. The goal is to estimate the positions and altitudes of the jump discontinuities of the best approximating step function. As main result consistency of our estimators will be established.

In the following, we define the problem more formally and thereby introduce necessary notation. In general, a $(d+1)$-levels step function on the domain $[l, r]$ with jump positions $t_1, \dots, t_d$ and levels $a_0, a_1, \dots, a_d$ can be written as

$$f_{t,a}(x) = a_0 \mathbf{1}_{[t_0,t_1]}(x) + \sum_{i=1}^{d} a_i \mathbf{1}_{(t_i,t_{i+1}]}(x), \quad x \in [l, r], \quad t_0 := l,\ t_{d+1} := r,$$

$$t = (t_1, \dots, t_d) \in \Delta_d := \{(t_1, \dots, t_d) \in (l, r)^d : t_1 < \dots < t_d\},$$

$$a = (a_0, \dots, a_d) \in R_{d+1} := \{(a_0, \dots, a_d) \in \mathbb{R}^{d+1} : a_i \neq a_{i+1},\ i = 0, \dots, d-1\}.$$

We define

$$D(t, a) := \int_l^r (f(x) - f_{t,a}(x))^2 \, \mathrm{d}x \tag{1}$$

DOI: 10.14736/kyb-2017-2-0198

and note that $\sqrt{D(t,a))}$ is the usual L^2-distance between f and $f_{t,a}$. Furthermore, let there exist a unique minimizer (τ, α) of $D(t, a)$, i.e.

$$(\tau, \alpha) = \text{argmin}\{D(t, a) \ : \ t \in \Delta_d, \ a \in R_{d+1}\}. \tag{2}$$

In other words, $f_{\tau,\alpha}$ is the unique best approximation of f w.r.t. the L^2-distance. The function $f_{\tau,\alpha}$ has jump positions $\tau = (\tau_1, \ldots, \tau_d)$ and levels $\alpha = (\alpha_0, \ldots, \alpha_d)$. The main issue of the present paper is to construct consistent estimators (τ_n, α_n) for (τ, α). This immediately leads to a density estimator f_{τ_n,α_n} of the step function $f_{\tau,\alpha}$. We later show that both f_{τ_n,α_n} and $f_{\tau,\alpha}$ are indeed density functions. In particular, f_{τ_n,α_n} is a histogram, where the cells and the pertaining cell-heights both are random.

Suppose the pdf f a priori is known to be a step function with d jumps. Then the best L^2-approximation $f_{\tau,\alpha}$ of f coincides with f and consequently our estimate f_{τ_n,α_n} of $f_{\tau,\alpha}$ is an estimate of f. In fact, it is a tailor-made solution for such types of underlying densities f. Moreover, even if the pdf f is not a step function, the function f_{τ_n,α_n} estimates its best approximation, which in turn gives the low, middle and high density regions. This classification can be used as an initial step in the usual kernel density estimation, which makes the statistician to adapt the bandwidth in the corresponding regions.

As explained above, our method will be particularly useful if the unknown pdf f itself is a step function. It is to be noted that the usual kernel density estimator is continuous and therefore performs poorly in that situation. Of course, it may happen that we do not know the number d of steps. Then we are able to present an estimator for d.

In the next section 2 we derive appropriate estimators (τ_n, α_n) of (τ, α). Weak and strong consistency of these estimators for (τ, α) is the main result in section 3. Our proofs rely on an argmin-Theorem for multivariate cadlag processes recently published in [6]. Moreover, we prove and apply a generalization of an inequality for sub-martingales in continuous time which originally can be traced back to Birnbaum and Marshall [2]. Here, martingale properties of the empirical process and probability bounds for its oscillation modulus are of great importance.

Section 4 focuses on the case that f is a step-function with unknown number d of jumps. We introduce an estimator for d and prove strong consistency. In section 5 we report on a simulation study.

2. ESTIMATION OF JUMP POSITIONS AND JUMP ALTITUDES OF THE BEST APPROXIMATING STEP FUNCTION

In this section, we derive estimators for the values of (τ, α) given by (2) under the assumption that $[l, r]$ is known. The next proposition is our starting point.

Proposition 2.1. The solution $(\tau, \alpha) \in \Delta_d \times R_{d+1}$ of the minimization (2) is

$$\tau = (\tau_1, \ldots, \tau_d) = \underset{(t_1,\ldots,t_d)\in\Delta_d}{\text{argmax}} \ M(t_1, \ldots, t_d) \tag{3}$$

with

$$M(t_1, \ldots, t_d) = \sqrt{\sum_{i=0}^{d} \frac{\{F(t_{i+1}) - F(t_i)\}^2}{t_{i+1} - t_i}},$$

and

$$\alpha = (\alpha_0, \dots, \alpha_d) \quad \text{with} \quad \alpha_j = \frac{F(\tau_{j+1}) - F(\tau_j)}{\tau_{j+1} - \tau_j}, \quad j = 0, 1, \dots, d, \tag{4}$$

where $\tau_0 := l,\ \tau_{d+1} := r$.

Proof. We rewrite

$$\begin{aligned} D(t, a) &= \sum_{i=0}^{d} \int_{t_i}^{t_{i+1}} (f(x) - a_i)^2 \, \mathrm{d}x \\ &= \sum_{i=0}^{d} \left(\int_{t_i}^{t_{i+1}} f^2(x) \, \mathrm{d}x - 2a_i \{F(t_{i+1}) - F(t_i)\} + {a_i}^2 (t_{i+1} - t_i) \right) \\ &= \int_l^r f^2(x) \, \mathrm{d}x - 2 \sum_{i=0}^{d} a_i \{F(t_{i+1}) - F(t_i)\} + \sum_{i=0}^{d} a_i^2 (t_{i+1} - t_i). \end{aligned}$$

Consequently, for fixed $t \in \Delta_d$, we have

$$\frac{dD}{da_j}(t, a) = -2\{F(t_{j+1}) - F(t_j)\} + 2a_j(t_{j+1} - t_j), \quad j = 0, 1, \dots, d.$$

By equating $\frac{dD}{da_j}(t, a)$ to zero for every $j = 0, 1, \dots, d$, one obtains

$$a_j = a_j(t) = \frac{F(t_{j+1}) - F(t_j)}{t_{j+1} - t_j}, \quad j = 0, 1, \dots, d. \tag{5}$$

Since the Hessian is positive definite it follows that $(a_0(t), \dots, a_d(t))$ minimizes $D(t, \cdot)$ for every fixed $t \in \Delta_d$. Thus from Proposition 1.35 in [12],

$$\tau = \operatorname*{argmin}_{t \in \Delta_d} D(t, a_0(t), \dots, a_d(t)),$$

and α is given by (4) through inserting τ into (5). From (5), we have

$$D(t, a_0(t), \dots, a_d(t)) = \int_l^r f^2(x) \, \mathrm{d}x - \sum_{i=0}^{d} \frac{\{F(t_{i+1}) - F(t_i)\}^2}{t_{i+1} - t_i}. \tag{6}$$

Because $\int_l^r f^2(x) \, \mathrm{d}x$ is constant w.r.t. t and the square root function is strictly monotone increasing, the minimization of $D(t, a_0(t), \dots, a_d(t))$ takes place at τ given by (3). □

As a consequence of Proposition 2.1, we can easily verify that the step function $f_{\tau,\alpha}$ is indeed a density function:

$$\int_l^r f_{\tau,\alpha}(u) \, \mathrm{d}u = \int_l^r \sum_{j=0}^{d} \alpha_j \mathbb{1}_{(\tau_j, \tau_{j+1}]}(u) \, \mathrm{d}u = \sum_{j=0}^{d} \frac{F(\tau_{j+1}) - F(\tau_j)}{\tau_{j+1} - \tau_j} (\tau_{j+1} - \tau_j) = 1. \tag{7}$$

Let

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le x\}}, \quad x \in \mathbb{R},$$

be the empirical df pertaining to $X_1, \ldots, X_n$. By replacing the true cdf F in (3) with F_n, we obtain the empirical analogue M_n of M:

$$M_n(t_1, \ldots, t_d) := \sqrt{\sum_{i=0}^{d} \frac{\{F_n(t_{i+1}) - F_n(t_i)\}^2}{t_{i+1} - t_i}}. \tag{8}$$

Moreover, we restrict the domain Δ_d of maximization to certain subsets

$$\Delta_{n,d} \subseteq \Delta_d \text{ such that } \Delta_{n,d} \uparrow \Delta_d \text{ as } n \to \infty. \tag{9}$$

Thus every *supremizer* (as defined in [6]) of M_n,

$$\tau_n = (\tau_{1,n}, \ldots, \tau_{d,n}) \in \underset{(t_1,\ldots,t_d)\in\Delta_{n,d}}{\operatorname{argsup}} M_n(t_1, \ldots, t_d), \tag{10}$$

is a reasonable estimator for τ.

Additionally, a consistent estimator for the vector α is needed. We propose the following one that is induced by (4).

$$\alpha_n = (\alpha_{0,n}, \ldots, \alpha_{d,n}) \quad \text{with} \quad \alpha_{j,n} := \frac{F_n(\tau_{j+1,n}) - F_n(\tau_{j,n})}{\tau_{j+1,n} - \tau_{j,n}}, \quad j = 0, 1, \ldots, d, \tag{11}$$

where $\tau_{0,n} := l$, $\tau_{d+1,n} := r$. We remark that the step function f_{τ_n,α_n} is a density function because the calculations in (7) remain valid if τ and α are replaced with their respective estimators τ_n and α_n.

Observe that M_n and M can be considered as random elements in the multivariate Skorokhod space $D(\Delta_d)$ as defined in [6]. This can be seen directly from the definition upon noticing that every cdf and every empirical df is right-continuous with left-hand limits (rcll), and continuous transformations of rcll functions are again rcll.

3. WEAK AND STRONG CONSISTENCY

There are two basic assumptions. Firstly, it is required that

$$\tau = \underset{t\in\Delta_d}{\operatorname{argmax}}\, M(t)$$

is not only unique (as a consequence of (2)), but actually is *well-separated* in the sense that

$$\sup_{t\in\partial\Delta_d} M(t) < M(\tau),$$

where $\partial\Delta_d$ denotes the boundary of Δ_d. Secondly, we need that f is bounded away from zero and infinity, that is

$$\rho := \inf_{x\in[l,r]} f(x) > 0 \quad \text{and} \quad ||f|| := \sup_{x\in[r,l]} |f(x)| < \infty. \tag{12}$$

(If further information on τ is available one can drop this assumption as we will see later on.) The following theorem is the main result in this chapter. Here, we deal with two types of subsets $\Delta_{n,d}$, namely for every positive β let

$$\Delta_{n,d} := \{t \in (l,r)^d : t_{i+1} - t_i > n^{-2\beta} \; \forall\, 1 \leq i \leq d-1\} \tag{13}$$

or alternatively,

$$\Delta_{n,d} := \{t \in (l,r)^d : t_{i+1} - t_i > n^{-2\beta} \; \forall\, 0 \leq i \leq d\} \tag{14}$$

Recall that by definition (10)

$$\tau_n = (\tau_{1,n}, \ldots, \tau_{d,n}) \in \underset{t \in \Delta_{n,d}}{\mathrm{argsup}} M_n(t),$$

Theorem 3.1. Assume that τ is the well-separated maximizing point of M and that f is bounded away from zero and infinity.

If $\Delta_{n,d}$ is of type (13), then

$$\tau_n \xrightarrow{P} \tau, \quad n \to \infty, \qquad \forall\, 0 < \beta < 1. \tag{15}$$

If $\Delta_{n,d}$ is of type (14), then

$$\tau_n \to \tau \;\text{ a.s.}\quad n \to \infty, \qquad \forall\, 0 < \beta < 1. \tag{16}$$

The proof of Theorem 3.1 includes several steps that will be taken subsequently. The main idea of the proof is to make use of an argmax-Theorem recently published in [6, Theorem 3.3]. In short, this theorem guarantees that if

$$\sup_{t \in \Delta_{n,d}} |M_n(t) - M(t)| \to 0, \; n \to \infty, \tag{17}$$

in probability or almost surely then every supremizer $\tau_n \in \Delta_{n,d}$ of the process M_n converges to the true vector τ in probability or almost surely, respectively. Therefore, the remainder of this section deals with checking the condition (17). As a start, the following lemma gives an upper bound for $|M_n(t) - M(t)|$.

Lemma 3.2. There exists a constant $\gamma > 0$ such that for all $t \in \Delta_d$

$$|M_n(t) - M(t)| \leq \gamma \cdot n^{-1/2} \sum_{i=0}^{d} \left\{ \frac{|\alpha_n(t_{i+1}) - \alpha_n(t_i)|}{\sqrt{t_{i+1} - t_i}} \right\}, \tag{18}$$

where

$$\alpha_n(s) := \sqrt{n}(F_n(s) - F(s)), \; s \in [l,r],$$

denotes the empirical process.

P r o o f. Because of the norm equivalence in $\mathbb{R}^{d+1}$ there exists a $\gamma > 0$ such that $||x||_2 \le \gamma \cdot ||x||_1$ for all $x \in \mathbb{R}^{d+1}$.

Now for $t \in \Delta_d$, let $x = (x_0, \ldots, x_d)$ and $y = (y_0, \ldots, y_d)$ be defined by

$$x_i := \frac{F_n(t_{i+1}) - F_n(t_i)}{\sqrt{t_{i+1} - t_i}}, \quad 0 \le i \le d,$$

and

$$y_i := \frac{F(t_{i+1}) - F(t_i)}{\sqrt{t_{i+1} - t_i}}, \quad 0 \le i \le d.$$

Then

$$|\ M_n(t) - M(t)\ | \ = \ \big|\ ||x||_2 - ||y||_2\ \big| \ \le \ ||x-y||_2 \le \gamma \cdot ||x-y||_1$$

and

$$||x-y||_1 = \sum_{i=0}^{d} \frac{|F_n(t_{i+1}) - F(t_{i+1}) - (F_n(t_i) - F(t_i))|}{\sqrt{t_{i+1} - t_i}},$$

resulting in the right-hand side of (18). □

In what follows, we consider the sum on the right-hand side of (18) a little more in detail. To begin with, we confine ourselves to considerations for $t \in \Delta_{n,d}$ defined in (13). Now the following upper estimates hold for the ith summands of the sum on the right-hand side of (18), where for convenience we put

$$\delta_n := n^{-2\beta}.$$

Recall that $t_0 = l$, $t_{d+1} = r$. Further notice that $F_n(l) = F(l) = 0$ and $F_n(r) = F(r) = 1$.

- $i = 0$: $\frac{|\alpha_n(t_1) - \alpha_n(t_0)|}{\sqrt{t_1 - t_0}} = \frac{|\alpha_n(t_1)|}{\sqrt{t_1 - l}} \le \sup_{l<u<r} \frac{|\alpha_n(u)|}{\sqrt{u-l}}$.
- $0 < i < d$: $\frac{|\alpha_n(t_{i+1}) - \alpha_n(t_i)|}{\sqrt{t_{i+1} - t_i}} \le \sup_{\substack{l<u<v<r \\ v-u>\delta_n}} \frac{|\alpha_n(v) - \alpha_n(u)|}{\sqrt{v-u}}$.
- $i = d$: $\frac{|\alpha_n(t_{d+1}) - \alpha_n(t_d)|}{\sqrt{r - t_d}} = \frac{|\alpha_n(t_d)|}{\sqrt{r - t_d}} \le \sup_{l<u<r} \frac{|\alpha_n(u)|}{\sqrt{r-u}}$.

Thus by Lemma 3.2 we obtain for $\Delta_{n,d}$ of type (13):

$$\frac{1}{\gamma} \sup_{t \in \Delta_{n,d}} |M_n(t) - M(t)| \le A_n + (d-1) \cdot B_n + C_n, \tag{19}$$

where

$$A_n := n^{-1/2} \sup_{l<u<r} \frac{|\alpha_n(u)|}{\sqrt{u-l}}, \tag{20}$$

$$B_n := n^{-1/2} \sup_{\substack{l<u<v<r \\ v-u>\delta_n}} \frac{|\alpha_n(v) - \alpha_n(u)|}{\sqrt{v-u}}, \tag{21}$$

$$C_n := n^{-1/2} \sup_{l<u<r} \frac{|\alpha_n(u)|}{\sqrt{r-u}}. \tag{22}$$

Similarly, if $\Delta_{n,d}$ is of type (14) then

$$\frac{1}{\gamma} \sup_{t \in \Delta_{n,d}} |M_n(t) - M(t)| \leq (d+1) \cdot B_n. \tag{23}$$

Consequently, by (19) and (23), it suffices to prove that each of the terms $A_n,\ B_n,\ C_n$ converges to zero in probability or a.s., respectively, in order to show condition (17).

For this, we need the following lemma, which generalizes the Birnbaum–Marshall inequality, see [13, p. 873, Inequality 4]. The proof of the lemma is based on the Chow Inequality [7, Theorem 6.6.1].

Lemma 3.3. Let $(S(u), \mathfrak{F}(u))$, $u \in [a,b)$, $0 < a < b < \infty$, be a submartingale with trajectories that are right-continuous with existing left-handed limits (rcll). Let $S(u)^+ := \max\{S(u), 0\}$ and $H(u) := \mathbb{E}(S(u)^+) < \infty$, $u \in [a,b)$. Furthermore, let $w : [a,b) \to (0,\infty)$ be rcll and monotone decreasing. Then for all $\lambda > 0$

$$P\left(\sup_{a \leq u < b} w(u)S(u) > \lambda\right) \leq \lambda^{-1}\left(\int_a^b H(u)(-w)(\mathrm{d}u) + \lim_{u \nearrow b} w(u)H(u)\right).$$

Proof. We define

$$I_m := \{u_k := a + (b-a)k2^{-m} : 0 \leq k \leq 2^m - 1\}.$$

Note that $u_0 = a$, $u_1 \searrow a$, and $u_{2^m-1} \nearrow b$ as $m \to \infty$. Moreover, put

$$S_k := S(u_k), \quad \mathfrak{F}_k := \mathfrak{F}(u_k), \quad 1 \leq k \leq 2^m - 1.$$

For every function $f : [a,b) \to \mathbb{R}$ rcll, we have that

$$\sup_{u \in [a,b)} f(u) = \sup_{m \geq 1} \sup_{u \in I_m} f(u).$$

Conclude that

$$\begin{aligned} p(\lambda) := P\left(\sup_{a \leq u < b} w(u)S(u) > \lambda\right) &= P\left(\bigcup_{m \geq 1}\left\{\sup_{u \in I_m} w(u)S(u) > \lambda\right\}\right) \\ &= \lim_{m \to \infty} P\left(\max_{1 \leq k < 2^m} w(u_k)S_k > \lambda\right). \end{aligned} \tag{24}$$

By the Chow Inequality, we obtain

$$\begin{aligned} &\lim_{m \to \infty} P\left(\max_{1 \leq k \leq 2^m - 1} w(u_k)S_k > \lambda\right) \\ &\quad \leq \lambda^{-1} \lim_{m \to \infty} \sum_{1 \leq k \leq 2^m - 2} (w(u_k) - w(u_{k+1}))\,\mathbb{E}(S_k^+) + w(u_{2^m-1})\mathbb{E}(S_{2^m-1}^+). \end{aligned} \tag{25}$$

Note that $H(u_k) = \mathbb{E}(S_k^+)$. Define $v(u) := -w(u),\ u \in [a,b)$. Then

$$\begin{aligned}
&\sum_{1\le k\le 2^m-2} (w(u_k) - w(u_{k+1}))\, H(u_k) \\
&= \sum_{2\le j\le 2^m-1} (w(u_{j-1}) - w(u_j))\, H(u_{j-1}) \\
&= \sum_{1\le j\le 2^m-1} (v(u_j) - v(u_{j-1}))\, H(u_{j-1}) - (v(u_1) - v(u_0))H(u_0).
\end{aligned} \tag{26}$$

Since $u_1 \searrow a$ as $m \to \infty$, and $u_0 = a$, it follows that

$$(v(u_1) - v(u_0))H(u_0) \ \to\ 0,\ m \to \infty. \tag{27}$$

On combining (24), (25), (26), and (27), we arrive at

$$p(\lambda) \le \lambda^{-1} \left(\lim_{m\to\infty} \sum_{1\le j\le 2^m-1} ((-w)(u_j) - (-w)(u_{j-1}))\, H(u_{j-1}) + \lim_{u\nearrow b} w(u)H(u) \right).$$

The first summand in brackets is in fact the integral $\int_a^b H(u)(-w)(\mathrm{d}u)$. □

By the quantile transformation we can w.l.o.g. assume that

$$X_i = F^{-1}(U_i),\ i \ge 1,$$

where F^{-1} denotes the quantile function of F and $U_i, i \ge 1$, are i.i.d. random variables uniformly distributed on $(0,1)$. Let

$$G_n(x) := n^{-1} \sum_{i=1}^{n} \mathbf{1}_{\{U_i \le x\}}, \quad x \in (0,1),$$

be the *uniform empirical distribution function.* Then the following simple relation

$$F_n(x) = G_n(F(x)), \quad x \in \mathbb{R}, \tag{28}$$

will be very useful in our proofs below. Similarly, we will benefit a lot from the inequalities

$$\rho(v-u) \le F(v) - F(u) \le ||f||(v-u) \quad \forall\, l \le u \le v \le r, \tag{29}$$

which follow from (12).

Lemma 3.4. Assume that (12) holds, that is f is bounded away from zero and infinity. Then:

(1) $A_n = n^{-1/2} \sup\limits_{l<u<r} \frac{|\alpha_n(u)|}{\sqrt{u-l}} \xrightarrow{P} 0,\ n \to \infty,$

(2) $C_n = n^{-1/2} \sup\limits_{l<u<r} \frac{|\alpha_n(u)|}{\sqrt{r-u}} \xrightarrow{P} 0,\ n \to \infty.$

P r o o f . According to (29) it is $u-l \geq ||f||^{-1}(F(u)-F(l)) = ||f||^{-1}F(u) \quad \forall\, u \in (l,r)$, whence in combination with (28)

$$A_n = \sup_{l<u<r} \frac{|F_n(u)-F(u)|}{\sqrt{u-l}} \leq \sqrt{||f||} \sup_{l<u<r} \frac{|G_n(F(u))-F(u)|}{\sqrt{F(u)}} \leq \sqrt{||f||} \sup_{0<u<1} \frac{|G_n(u)-u|}{\sqrt{u}} \tag{30}$$

Now

$$\sup_{0<u<1} \frac{|G_n(u)-u|}{\sqrt{u}} \leq \max\{a_n,\ b_n\}, \tag{31}$$

where

$$a_n := \sup_{0<u<U_{1:n}} \frac{|G_n(u)-u|}{\sqrt{u}}, \quad b_n := \sup_{U_{1:n}\leq u<1} \frac{|G_n(u)-u|}{\sqrt{u}},$$

with $U_{1:n} := \min_{1\leq i\leq n} U_i$ the first order statistic in the sample $U_1,\ldots,U_n$. From $G_n(u)=0$ for all $u<U_{1:n}$ infer that

$$0 \leq a_n \leq \sqrt{U_{1:n}} \to 0 \quad \text{a.s.}, \tag{32}$$

where the convergence follows from the First Borel–Cantelli Lemma observing that $\mathbb{P}(|U_{1:n}|>\epsilon) = (1-\epsilon)^n$ for all $\epsilon \in (0,1]$ and $n \in \mathbb{N}$.

As to the second supremum b_n observe that for every $\epsilon>0$ and $c>0$:

$$P(b_n>\epsilon) \leq \gamma_n + \zeta_n, \tag{33}$$

where

$$\gamma_n := P\left(b_n>\epsilon,\ U_{1:n}>\frac{1}{nc}\right), \quad \zeta_n := P\left(U_{1:n}\leq \frac{1}{nc}\right).$$

Now

$$\gamma_n \leq P\left(\sup_{1/(nc)\leq u<1} \frac{|G_n(u)-u)|}{\sqrt{u}} > \epsilon\right) \tag{34}$$

and further

$$\begin{aligned} P\left(\sup_{1/(nc)\leq u<1} \frac{|G_n(u)-u)|}{\sqrt{u}} > \epsilon\right) &= P\left(\sup_{1/(nc)\leq u<1} \frac{|G_n(u)-u|}{1-u}\cdot\frac{1-u}{\sqrt{u}} > \epsilon\right) \\ &\leq P\left(\sup_{1/(nc)\leq u<1} \left(\frac{|G_n(u)-u|}{1-u}\right)^2 \cdot \frac{(1-u)^2}{u} > \epsilon^2\right). \end{aligned} \tag{35}$$

Setting $S_n(u) := \left(\frac{G_n(u)-u}{1-u}\right)^2$ and $w(u) := \frac{(1-u)^2}{u}$ gives

$$\gamma_n \leq P\left(\sup_{1/(nc)\leq u<1} S_n(u)w(u) > \epsilon^2\right).$$

It is well-known, confer Koul [10, section 2.4.3], that $V_n(u) := \frac{G_n(u)-u}{1-u}$, $u \in [0,1)$, is a martingale with rcll trajectories, hence $S_n(u) = V_n(u)^2$ is a non-negative submartingale with rcll trajectories. Let $H_n(u) := \mathbb{E}(S_n(u)^+) = \mathbb{E}(S_n(u))$ and note that $u \mapsto w(u)$ is rcll and monotone decreasing. Then Lemma 3.3 yields

$$\gamma_n \leq \epsilon^{-2} \left(\int_{1/(nc)}^{1} H_n(u)(-w)(\mathrm{d}u) + \lim_{u \nearrow 1} w(u) H_n(u) \right). \tag{36}$$

Since $\mathbb{E}G_n(u) = u$ and $Var(G_n(u)) = n^{-1}u(1-u)$ it follows that $H_n(u) = \mathbb{E}(S_n(u)) = \frac{1}{n}\frac{u}{1-u}$, whence $w(u)H_n(u) = n^{-1}(1-u)$ and thus

$$\lim_{u \nearrow 1} w(u) H_n(u) = 0. \tag{37}$$

Integration by parts yields

$$\int_{1/(nc)}^{1} H_n(u)(-w)(\mathrm{d}u) = H_n(u)(-w(u))\Big|_{1/(nc)}^{1} - \int_{1/(nc)}^{1} (-w(u)) H_n(\mathrm{d}u). \tag{38}$$

Obviously,

$$H_n(u)(-w(u))\Big|_{1/(nc)}^{1} = n^{-1}(1 - 1/(nc)) \leq n^{-1}.$$

Furthermore, since $H_n'(u) = n^{-1}(1-u)^{-2}$, we have that

$$-\int_{1/(nc)}^{1} (-w(u)) H_n(\mathrm{d}u) = n^{-1} \int_{1/(nc)}^{1} u^{-1}\,\mathrm{d}u = n^{-1}\log(nc). \tag{39}$$

Finally, looking back over (36) and (37) - (39), we can conclude that

$$\gamma_n \leq \epsilon^{-2} \left(\frac{1}{n}\log(nc) + \frac{1}{n} \right) \;\rightarrow\; 0, \quad n \to \infty, \quad \forall c > 0. \tag{40}$$

In view of ζ_n recall that $nU_{1:n}$ converges in distribution to the Exponential with parameter 1. Thus

$$\lim_{n\to\infty} \zeta_n = \lim_{n\to\infty} P(nU_{1:n} \leq \frac{1}{c}) = 1 - e^{-1/c} \quad \forall\, c > 0. \tag{41}$$

Since (33), (40) and (41) hold for all $c > 0$ taking the limit $c \to \infty$ yields that

$$b_n \xrightarrow{P} 0, \quad n \to \infty.$$

Consequently, by (32) and (31) we obtain that

$$\sup_{0<u<1} \frac{|G_n(u) - u|}{\sqrt{u}} \xrightarrow{P} 0, \quad n \to \infty, \tag{42}$$

which finally by (30) yields part (1) of the lemma.

Similarly as in the derivation of (30) we obtain that

$$C_n \le \sqrt{||f||} \sup_{0<u<1} \frac{|G_n(u) - u|}{\sqrt{1-u}} = \sqrt{||f||} \sup_{0<u<1} \frac{|G_n(1-u) - (1-u)|}{\sqrt{u}}. \tag{43}$$

Since $(U_i : 1 \le i \le n) \stackrel{\mathcal{L}}{=} (1 - U_i : 1 \le i \le n)$ it follows that

$$(G_n(1-u_k) : 1 \le k \le q) \stackrel{\mathcal{L}}{=} (1 - G_n(u_k-) : 1 \le k \le q) \quad \forall\, 0 < u_1 < \cdots < u_q < 1 \quad \forall\, q \in \mathbb{N}. \tag{44}$$

If $u_k := u_{k,m} := k2^{-m}, 1 \le k < 2^m$, then for every $x \in \mathbb{R}$

$$\begin{aligned}
&P\Big(\sup_{0<u<1} \frac{|G_n(1-u) - (1-u)|}{\sqrt{u}} > x\Big) \\
&= \lim_{m\to\infty} P\Big(\max_{1\le k<2^m} \frac{|G_n(1-u_k) - (1-u_k)|}{\sqrt{u_k}} > x\Big) \\
&= \lim_{m\to\infty} P\Big(\max_{1\le k<2^m} \frac{|G_n(u_k-) - u_k|}{\sqrt{u_k}} > x\Big) \qquad \text{by (44)} \\
&= P\Big(\sup_{0<u<1} \frac{|G_n(u-) - u)|}{\sqrt{u}} > x\Big),
\end{aligned}$$

whence

$$\sup_{0<u<1} \frac{|G_n(1-u) - (1-u)|}{\sqrt{u}} \stackrel{\mathcal{L}}{=} \sup_{0<u<1} \frac{|G_n(u-) - u)|}{\sqrt{u}} = \sup_{0<u<1} \frac{|G_n(u) - u)|}{\sqrt{u}},$$

where the last equality holds because G_n is rcll and so is the ratio as well. Thus (42) and (43) immediately yield the second part (2) of the lemma. □

Lemma 3.5. Assume that is f is bounded away from zero and infinity. If $\delta_n = n^{-2\beta}$ with $0 < \beta < 1$, then:

$$B_n = n^{-1/2} \sup_{\substack{l<u<v<r \\ v-u>\delta_n}} \frac{|\alpha_n(v) - \alpha_n(u)|}{\sqrt{v-u}} \to 0 \quad \text{a.s. as } n \to \infty.$$

Proof. Let $\bar{\alpha}_n$ be the uniform empirical process, i. e.,

$$\bar{\alpha}_n(t) := \sqrt{n}(G_n(t) - t), \quad t \in [0,1].$$

It follows from the inequalities (29) that

$$B_n = n^{-1/2} \sup_{\substack{l<u<v<r \\ v-u>\delta_n}} \frac{|\alpha_n(v) - \alpha_n(u)|}{\sqrt{v-u}} \le \sqrt{||f||}\, n^{-1/2} \sup_{\substack{0<u<v<1 \\ v-u>\bar{\delta}_n}} \frac{|\bar{\alpha}_n(v) - \bar{\alpha}_n(u)|}{\sqrt{v-u}}, \tag{45}$$

where

$$\bar{\delta}_n := \rho\delta_n = \rho n^{-2\beta}.$$

Observe that

$$\sup_{\substack{0<u<v<1\\ v-u>\bar{\delta}_n}} \frac{|\bar{\alpha}_n(v)-\bar{\alpha}_n(u)|}{\sqrt{v-u}} \leq \sup_{0<s<1} \sup_{\bar{\delta}_n<u<1-s} \frac{|\bar{\alpha}_n(s+u)-\bar{\alpha}_n(s)|}{\sqrt{u}}. \tag{46}$$

Now let $k_n := \lfloor n^\beta \rfloor$, $s \in (0,1)$, $u \in (\bar{\delta}_n, 1-s)$.

Case A: $s = j/k_n$ for some $j = 0,1,\ldots,k_n-1$. Then

$$\frac{|\bar{\alpha}_n(s+u)-\bar{\alpha}_n(s)|}{\sqrt{u}} \leq \max_{0\leq j<k_n} \sup_{\bar{\delta}_n<u<1-\frac{j}{k_n}} \frac{|\bar{\alpha}_n(\frac{j}{k_n}+u)-\bar{\alpha}_n(\frac{j}{k_n})|}{\sqrt{u}}.$$

Case B: $s \in (\frac{j}{k_n}, \frac{j+1}{k_n})$ for some $j = 0,1,\ldots,k_n-1$. Then

$$\begin{aligned}
&|\bar{\alpha}_n(s+u)-\bar{\alpha}_n(s)| \\
&= \left|\bar{\alpha}_n(\frac{j}{k_n}+u)-\bar{\alpha}_n(\frac{j}{k_n})+\bar{\alpha}_n(\frac{j}{k_n})-\bar{\alpha}_n(s)+\bar{\alpha}_n(s+u)-\bar{\alpha}_n(\frac{j}{k_n}+u)\right| \\
&\leq \left|\bar{\alpha}_n(\frac{j}{k_n}+u)-\bar{\alpha}_n(\frac{j}{k_n})\right| + \left|\bar{\alpha}_n(\frac{j}{k_n})-\bar{\alpha}_n(s)\right| + \left|\bar{\alpha}_n(s+u)-\bar{\alpha}_n(\frac{j}{k_n}+u)\right|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{|\bar{\alpha}_n(s+u)-\bar{\alpha}_n(s)|}{\sqrt{u}} \\
&\qquad\leq \max_{0\leq j<k_n} \sup_{\bar{\delta}_n<u<1-\frac{j}{k_n}} \left\{\frac{|\bar{\alpha}_n(\frac{j}{k_n}+u)-\bar{\alpha}_n(\frac{j}{k_n})|}{\sqrt{u}}\right\} + \frac{2}{\sqrt{\bar{\delta}_n}}\omega_n(\frac{1}{k_n}),
\end{aligned} \tag{47}$$

where

$$\omega_n(a) := \sup_{|u-v|\leq a} |\bar{\alpha}_n(u)-\bar{\alpha}_n(v)|, \quad a \in (0,\infty),$$

denotes the oscillation modulus of $\bar{\alpha}_n$. For verifying inequality (47), note that $|s-\frac{j}{k_n}| < \frac{1}{k_n}$ and $|\frac{j}{k_n}+u-(s+u)| < \frac{1}{k_n}$.

Taking Case A and Case B together, we arrive with (45) and (46) at

$$B_n \leq \sqrt{||f||}(c_n+d_n), \tag{48}$$

where

$$c_n := n^{-\frac{1}{2}} \max_{0\leq j<k_n} \sup_{\bar{\delta}_n<u<1-\frac{j}{k_n}} \left\{\frac{|\bar{\alpha}_n(\frac{j}{k_n}+u)-\bar{\alpha}_n(\frac{j}{k_n})|}{\sqrt{u}}\right\},$$

$$d_n := n^{-\frac{1}{2}} \frac{2}{\sqrt{\bar{\delta}_n}}\omega_n(\frac{1}{k_n}).$$

First, we prove $c_n \to 0$ a.s. as $n \to \infty$. For this, let $\epsilon > 0$. Now

$$P(c_n > \epsilon) \leq \sum_{0 \leq j < k_n} p_{j,n} \tag{49}$$

with

$$p_{j,n} := P\left(n^{-\frac{1}{2}} \sup_{\bar{\delta}_n < u < 1 - \frac{j}{k_n}} \frac{|\bar{\alpha}_n(\frac{j}{k_n} + u) - \bar{\alpha}_n(\frac{j}{k_n})|}{\sqrt{u}} > \epsilon \right). \tag{50}$$

Next we use the *differential property* of the uniform empirical process:

$$\{\bar{\alpha}_n(y+u) - \bar{\alpha}_n(y) : 0 \leq u \leq 1-y\} \stackrel{\mathcal{L}}{=} \{\bar{\alpha}_n(u) : 0 \leq u \leq 1-y\} \in D[0, 1-y]$$

for every fixed $y \in [0,1]$. This is a simple consequence of the stationarity of the increments of $\bar{\alpha}_n$ (confer, e.g., Dudley [4], Lemma 1.14(b)) and of Theorem 12.5 (iii) in Billingsley [1]. Infer from (50) that

$$\begin{aligned} p_{j,n} &= P\left(n^{-\frac{1}{2}} \sup_{\bar{\delta}_n < u < 1 - \frac{j}{k_n}} \frac{|\bar{\alpha}_n(u)|}{\sqrt{u}} > \epsilon \right) \\ &\leq P\left(\sup_{\bar{\delta}_n \leq u \leq 1/2} \frac{|\bar{\alpha}_n(u)|}{\sqrt{u}} > \epsilon\sqrt{n} \right) + P\left(\sup_{1/2 \leq u < 1} \frac{|\bar{\alpha}_n(u)|}{\sqrt{u}} > \epsilon\sqrt{n} \right) \\ &=: r_n(\epsilon) + s_n(\epsilon). \end{aligned} \tag{51}$$

Since

$$\begin{aligned} r_n(\epsilon) &\leq P\left(\sup_{\bar{\delta}_n \leq u \leq 1/2} \frac{|\bar{\alpha}_n^+(u)|}{\sqrt{u}} > 1/2\epsilon\sqrt{n} \right) + P\left(\sup_{\bar{\delta}_n \leq u \leq 1/2} \frac{|\bar{\alpha}_n^-(u)|}{\sqrt{u}} > 1/2\epsilon\sqrt{n} \right) \\ &=: r_n^+(\epsilon) + r_n^-(\epsilon), \end{aligned} \tag{52}$$

we can apply Corollary 1 in Shorack and Wellner [13], p. 446 (with $b := \delta := 1/2$ there). It says when $0 < a \leq 1/4$, and $\lambda > 0$ one has the inequality

$$P(\sup_{a \leq u \leq 1/2} \frac{|\bar{\alpha}_n^\pm(u)|}{\sqrt{u}} > \lambda) \leq 6 \log(1/(2a)) \exp\{-\frac{1}{8}\gamma^\pm \lambda^2\}, \tag{53}$$

where $\gamma^- = 1$ for all $\lambda > 0$ and $\gamma^+ \geq 3/4\sqrt{na}/\lambda$, if $\lambda \geq 3/2\sqrt{na}$. Put $\lambda := 1/2\epsilon\sqrt{n}$ and $a := \bar{\delta}_n = \rho n^{-2\beta} \leq 1/4$ for some $n_0 \in \mathbb{N}$. Then by (53) we have that

$$r_n^-(\epsilon) \leq 6 \log(1/(2\rho) n^{2\beta}) \exp\{-1/32 \epsilon^2 n\} \quad \forall\, n \geq n_0. \tag{54}$$

As to $r_n^+(\epsilon)$ observe that $\lambda \geq 3/2\sqrt{na}$ iff $n^{-\beta} \leq 1/(3\sqrt{\rho})\epsilon$. Consequently, there exists some integer $n_1 = n_1(\epsilon)$ such that

$$r_n^+(\epsilon) \leq 6 \log(1/(2\rho) n^{2\beta}) \exp\{-3/64 \sqrt{\rho} \epsilon n^{1-\beta}\} \quad \forall\, n \geq n_1. \tag{55}$$

Combining (54) and (55) results in

$$r_n(\epsilon) \le 6\log(1/(2\rho)n^{2\beta})\left(\exp\{-1/32\epsilon^2 n\} + \exp\{-3/64\sqrt{\rho}\epsilon n^{1-\beta}\}\right) \quad \forall\, n \ge n_2, \quad (56)$$

with $n_2 := n_0 \vee n_1 \in \mathbb{N}$. Since $\sup_{1/2\le u<1} \frac{|\bar{\alpha}_n(u)|}{\sqrt{u}} \le \sqrt{2}\sup_{0\le u\le 1} |\bar{\alpha}_n(u)|$ Massart's [11] inequality yields that

$$s_n(\epsilon) \le 2\exp\{-n\epsilon^2\} \quad \forall\, n \ge 1. \quad (57)$$

From (49) and (51) it follows that

$$P(c_n > \epsilon) \le k_n(r_n(\epsilon) + s_n(\epsilon)) \le n^\beta(r_n(\epsilon) + s_n(\epsilon)),$$

whence (56) and (57) guarantee that

$$\sum_{n\ge 1} P(c_n > \epsilon) < \infty \quad \forall \epsilon > 0$$

form which in turn we can conclude with the First Borel–Cantelli Lemma that

$$c_n \longrightarrow 0 \quad \text{a.s. as } n \to \infty. \quad (58)$$

We now prove that $d_n \to 0$ a.s. as $n \to \infty$. For every $\epsilon > 0$ we obtain:

$$P\left(d_n \ge 2\epsilon\right) = P\left(\omega_n(k_n^{-1}) \ge \epsilon\sqrt{n\bar{\delta}_n}\right) = P\left(\omega_n(a) \ge \lambda\sqrt{a}\right), \quad (59)$$

where $a := k_n^{-1} \in (0, 1/2]$ for all n larger than some $n_3 \in \mathbb{N}$ and $\lambda := \epsilon\sqrt{n\bar{\delta}_n k_n} > 0$. Therefore we can apply the inequality of Mason, Shorack and Wellner [13, p. 545] (with $\delta := 1/2$ there). It gives

$$P\left(\omega_n(a) \ge \lambda\sqrt{a}\right) \le 160\; a^{-1}\exp\{-1/32\;\lambda^2\;\psi(\lambda/\sqrt{na})\} \quad (60)$$

with function $\psi(u) := \frac{2}{u^2}\left[(1+u)\log(1+u) - u\right], u > 0$. Inserting a and λ in (60) yields

$$P\left(\omega_n(a) \ge \lambda\sqrt{a}\right) \le 160\; k_n \exp\{-1/32\;\epsilon^2 n\bar{\delta}_n k_n \psi(\epsilon\; k_n\sqrt{\bar{\delta}_n})\}. \quad (61)$$

By Proposition 1 on p. 441 in [13] ψ is strictly positive and decreasing on $(0, \infty)$. Since $k_n\sqrt{\bar{\delta}_n} \le \sqrt{\rho}$ the factor $\psi\left(\epsilon k_n\sqrt{\bar{\delta}_n}\right)$ is greater or equal to $\psi(\epsilon\sqrt{\rho}) > 0$. Moreover, $k_n \ge (1/2)n^\beta$ for all n larger some $n_4 \in \mathbb{N}$ and so $n\bar{\delta}_n k_n \ge (1/2)\rho n^{1-\beta}$. Consequently, we can infer from (59) and (61) that the following inequality holds:

$$P\left(d_n \ge 2\epsilon\right) \le 160 n^\beta \exp\{-L\; n^{1-\beta}\} \quad \forall\, n \ge n_5$$

with positive and finite constant $L = -\frac{1}{64}\rho\psi(\epsilon\sqrt{\rho})$ and natural number $n_5 = \max\{n_3, n_4\}$. Thus, with the First Borel–Cantelli Lemma we arrive at

$$d_n \to 0 \quad \text{a.s. as } n \to \infty, \quad (62)$$

which in conclusion with (48) and (58) gives the desired result. □

Now we are able to prove Theorem 3.1.

Proof. [Theorem 3.1] We make use of Theorem 3.3 and Remark 3.1 in [6]. In the sequel, the validity of the assumptions there are verified. Firstly, $\Delta_{n,d} \uparrow \Delta_d$ for $\Delta_{n,d}$ of type (13) and (14). Thus in particular in both cases

$$\liminf_{n\to\infty} \Delta_{n,d} = \Delta_d.$$

Secondly, as a result of (19) and Lemmas 3.4 and 3.5 , we see that

$$\sup_{t\in\Delta_{d,n}} |M_n(t) - M(t)| \xrightarrow{P} 0,\ n \to \infty, \quad \text{if } \Delta_{n,d} \text{ is of type (13)},$$

and by (23) and Lemma 3.5

$$\sup_{t\in\Delta_{d,n}} |M_n(t) - M(t)| \to 0 \ \text{ a.s. as } n \to \infty, \quad \text{if } \Delta_{n,d} \text{ is of type (14)}.$$

Thirdly, for every $n \in \mathbb{N}$ the estimator τ_n is a supremizing point of the restriction of M_n on $\Delta_{n,d}$, and τ is the well-separated maximizer of M on Δ_d. Herewith, all requirements of Theorem 3.3 in [6] are fulfilled, which yields the weak (15) and strong (16) consistency of τ_n. □

Assume the statistician has knowledge about the minimal distance between the jump-positions in the sense that

$$\min\{\tau_{i+1} - \tau_i : 0 \le i \le d\} > m, \tag{63}$$

where $m > 0$ is known. Then τ lies in the region

$$\widehat{\Delta} := \{t = (t_1, \dots, t_d) \in (l, r)^d : t_{i+1} - t_i > m\},$$

whence a reasonable estimator is now given by

$$\widehat{\tau}_n = (\hat{\tau}_{1,n}, \dots, \hat{\tau}_{d,n}) \in \underset{t\in\widehat{\Delta}}{\operatorname{argsup}} M_n(t).$$

In this case we need no boundedness condition on f. Moreover, the mathematical treatment is very easy.

Theorem 3.6. If τ is the well-separated maximizing point of M, then

$$\widehat{\tau}_n \to \tau \ \text{ a.s. as } n \to \infty.$$

Proof. It follows from Lemma 3.2 that

$$\sup_{t\in\widehat{\Delta}} |M_n(t) - M(t)| \le \gamma \frac{1}{\sqrt{m}} 2(d+1) \sup_{u\in\mathbb{R}} |F_n(u) - F(u)|,$$

where the right-hand side converges to zero (for every df F) with probability one by the Glivenko-Cantelli Theorem. Thus, an application of Theorem 3.3 in [6] finishes the proof. □

Recall the definition of α_n in (11). Its natural counterpart is given by

$$\hat{\alpha}_n = (\hat{\alpha}_{0,n}, \dots, \hat{\alpha}_{d,n}) \quad \text{with} \quad \hat{\alpha}_{j,n} := \frac{F_n(\hat{\tau}_{j+1,n}) - F_n(\hat{\tau}_{j,n})}{\hat{\tau}_{j+1,n} - \hat{\tau}_{j,n}}, \quad j = 0, 1, \dots, d, \tag{64}$$

where $\hat{\tau}_{0,n} := l, \ \hat{\tau}_{d+1,n} := r$.

Corollary 3.7. Assume that τ is the well-separated maximizing point of M and that f is bounded away from zero and infinity.

If $\Delta_{n,d}$ is of type (13), then

$$(\tau_n, \alpha_n) \xrightarrow{P} (\tau, \alpha) \quad n \to \infty, \qquad \forall\, 0 < \beta < 1. \tag{65}$$

If $\Delta_{n,d}$ is of type (14), then

$$(\tau_n, \alpha_n) \to (\tau, \alpha) \ \text{ a.s.} \quad n \to \infty, \qquad \forall\, 0 < \beta < 1. \tag{66}$$

Proof.

For the proof of (65) it suffices by (15) of Theorem 3.1 to show that $\alpha_n \xrightarrow{P} \alpha$. Since

$$\begin{aligned} |F_n(\tau_{j,n}) - F(\tau_j)| &\le |F_n(\tau_{j,n}) - F(\tau_{j,n})| + |F(\tau_{j,n}) - F(\tau_j)| \\ &\le \sup_{s\in\mathbb{R}} |F_n(s) - F(s)| + |F(\tau_{j,n}) - F(\tau_j)| \\ &\xrightarrow{P} 0, \ n \to \infty, \end{aligned}$$

one can deduce from the Glivenk-Cantelli Theorem and (15) of Theorem 3.1 that $F_n(\tau_{j,n}) \xrightarrow{P} F(\tau_j), \ n \to \infty, \ j = 0, 1, \dots, d+1$, upon noticing that F is continuous. Therefore, by the continuity theorem for convergence in probability

$$\alpha_{j,n} \xrightarrow{P} \frac{F(\tau_{j+1}) - F(\tau_j)}{\tau_{j+1} - \tau_j}, \quad n \to \infty, \quad j = 0, 1, \dots, d,$$

From Proposition 2.1, we know that $\alpha_j = \frac{F(\tau_{j+1}) - F(\tau_j)}{\tau_{j+1} - \tau_j}, \ j = 0, 1, \dots, d$. To sum up it follows that $(\tau_n, \alpha_n) \xrightarrow{P} (\tau, \alpha), \ n \to \infty$ as desired. The proof for (66) follows in the same line. □

Similarly, for $\widehat{\Delta}$ with known m given in (63) we obtain analogously as above that the following corollary is true.

Corollary 3.8. If τ is the well-separated maximizing point of M, then

$$(\hat{\tau}_n, \hat{\alpha}_n) \to (\tau, \alpha) \ \text{ a.s.} \quad n \to \infty. \tag{67}$$

Remark 3.9. Kanazawa [8] uses the Hellinger-distance in place of the L_2-distance, i. e., he considers

$$\tilde{D}(t,a) := \int_l^r (\sqrt{f(x)} - \sqrt{f_{t,a}(x)})^2 \, \mathrm{d}x.$$

and

$$(\tilde{\tau}, \tilde{\alpha}) = \operatorname{argmin}\{\tilde{D}(t,a) \ : \ t \in \Delta_d, \ a \in R_{d+1}, \sum_{i=0}^{d} a_i(t_{i+1} - t_i) = 1\}.$$

The corresponding minimization problem can be solved by the method of Lagrange multipliers and gives

$$\tilde{\tau} = (\tilde{\tau}_1, \dots, \tilde{\tau}_d) = \operatorname*{argmax}_{(t_1,\dots,t_d) \in \Delta_d} \tilde{M}(t_1, \dots, t_d),$$

where

$$\tilde{M}(t_1, \dots, t_d) = \sqrt{\sum_{i=0}^{d} \frac{(\int_{t_i}^{t_{i+1}} \sqrt{f(x)} \, \mathrm{d}x)^2}{t_{i+1} - t_i}}. \tag{68}$$

Comparing $\tilde{M}$ with our M in (3) we see that f is simply replaced by $\sqrt{f}$. Now, Kanazawa [8] makes a further transformation based on

$$\int_{t_i}^{t_{i+1}} \sqrt{f(x)} \, \mathrm{d}x = \int_{t_i}^{t_{i+1}} (f(x))^{-1/2} F(\mathrm{d}x) = \int_{F(t_i)}^{F(t_{i+1})} (f(F^{-1}(u))^{-1/2} \, \mathrm{d}u$$

and similarly

$$t_{i+1} - t_i = \int_{F(t_i)}^{F(t_{i+1})} (f(F^{-1}(u))^{-1} \, \mathrm{d}u.$$

Therefore,

$$\tilde{M}(t_1, \dots, t_d)^2 = C(F(t_1), \dots, F(t_d))$$

with

$$C(p_1, \dots, p_d) = \sum_{i=0}^{d} \frac{(\int_{p_i}^{p_{i+1}} (f(F^{-1}(u))^{-1/2} \, \mathrm{d}u}{\int_{p_i}^{p_{i+1}} (f(F^{-1}(u))^{-1} \, \mathrm{d}u}.$$

Let

$$\pi = (\pi_1, \dots, \pi_d) := \operatorname*{argmax}_{0 < p_1 < \dots < p_d < 1} C(p_1, \dots, p_d).$$

Since F is continuous and strictly increasing the map $(t_1, \dots, t_d) \mapsto (F(t_1), \dots, F(t_d))$ is a bijection, whence

$$\pi_j = F(\tilde{\tau}_j), \quad 1 \le j \le d,$$

or equivalently

$$\tilde{\tau}_j = F^{-1}(\pi_j) \quad 1 \le j \le d. \tag{69}$$

Kanazawa [8] motivates an empirical analogue C_n of C which is based on the spacings of the observations X_i. This leads to the estimator

$$p_n = (p_{1,n}, \dots, p_{d,n}) = \operatorname*{argmax}_{0 < p_1 < \dots < p_d < 1} C_n(p_1, \dots, p_d).$$

He proves that

$$p_n \xrightarrow{P} \pi,\ n \to \infty. \tag{70}$$

In view of (69) a reasonable estimator $\tilde{\tau}_n = (\tilde{\tau}_{1,n}, \ldots, \tilde{\tau}_{d,n})$ for $\tilde{\tau}$ is given by

$$\tilde{\tau}_{j,n} = F_n^{-1}(p_{j,n}), \quad 1 \le j \le d.$$

For the proof of the consistency (70) he requires (amongst others) that F^{-1} is twice continuously differentiable with first derivative such that $0 < m \le (F^{-1})' \le M$ for all $u \in [0,1]$ with constants m and M, confer A.1-A.4 in [8]. In this case F_n^{-1} converges to F^{-1} uniformly on $(0,1)$ with probability one. Consequently, one obtains from (70) that

$$\tilde{\tau}_n \xrightarrow{P} \tilde{\tau},\ n \to \infty.$$

Note that it is cannot be taken for granted that $\tilde{\tau} = \tau$. In fact, this is not true in general. To see this consider $d = 1$ and the density

$$f(x) := (5 - x^2)/12, \quad x \in [-1, 2].$$

Here, $\tau_1 = 1.25$, whereas $\tilde{\tau}_1 = 1.34087$. (The pertaining levels are $\alpha_0 = 0.380208, \alpha_1 = 0.192708$ and $\tilde{\alpha}_0 = 0.3751114, \tilde{\alpha}_1 = 0.177895$, respectively.)

On the other hand, if $f = f_{\tau,\alpha}$ is a d-step function, then $(\tau, \alpha) = (\tilde{\tau}, \tilde{\alpha})$, because then

$$0 = D(\tau, \alpha) < D(t, a) \quad \forall\ (t, a) \in \Delta_d \times R_{d+1}$$

and

$$0 = \tilde{D}(\tau, \alpha) < \tilde{D}(t, a) \quad \forall\ (t, a) \in \Delta_d \times R_{d+1}$$

upon noticing that the L_2-distance between two step-functions is zero if and only if both coincide and the same holds for the Hellinger-distance.

However, we would like to point out that the case $f = f_{\tau,\alpha}$ is excluded by Kanazawa's differentiability assumption on F^{-1}.

If f is continuous there is no *correct* number of steps as Kanazawa [9] points out. As a consequence to obtain a consistent estimator d should depend on the sample size n. Kanazawa [9] suggests a sample-based $d = \tilde{d}_n$ and shows that $\tilde{d}_n \sim \lambda(f) n^{1/3}$ in probability and gives an explicit formula for the functional $\lambda(f)$.

Remark 3.10. Recall that our estimators are specifically designed for densities $f = f_{\tau,\alpha}$ that are d-step functions with known number d of jumps. Our approach is global in the sense that we do not estimate f locally at each point $x \in [l, r]$ as it is done in kernel-density estimation. In particular, no smoothing parameter is involved in contrast to the problem of bandwidth-selection for kernel-density estimators. On the other hand, here d may be unknown, confer, e. g., Chu and Cheng [3]. We treat this case for our global approach in the next section.

4. ESTIMATION OF THE NUMBER OF STEPS

Throughout this section f is a step-function with unknown number d of steps. We wish to estimate the true value d_0 of d under the additional information that $d_0 \leq \bar{d}$, where the upper bound $\bar{d}$ is known. Notice that our involved functions depend on d, that is $D = D^{(d)}, M = M^{(d)}, \tau = \tau^{(d)}, \alpha = \alpha^{(d)}, M_n = M_n^{(d)}, \tau_n = \tau_n^{(d)}, \alpha_n = \alpha_n^{(d)}$. Define

$$\hat{d}_n := \operatorname*{argmax}_{1 \leq d \leq \bar{d}} M_n^{(d)}(\tau_n^{(d)}).$$

Theorem 4.1. Assume for every $d \in \{1, \ldots, \bar{d}\}$ that $\tau^{(d)}$ is the well-separated maximizing point of $M^{(d)}$ and that $\alpha^{(d)} > 0$ (componentwise) as well as $\Delta_{n,d}$ is of type (14). Then

$$\mathbb{P}(\hat{d}_n = d_0 \text{ for eventually all } n \in \mathbb{N}) = 1. \tag{71}$$

Proof. Define $Q_n(d) := M_n^{(d)}(\tau_n^{(d)})$ and $Q(d) := M^{(d)}(\tau^{(d)}),\ 1 \leq d \leq \bar{d}$. Observe for each such d we have that:

$$|Q_n(d) - Q(d)| \leq \sup_{t \in \Delta_{n,d}} |M_n^{(d)}(t) - M^{(d)}(t)| + |M^{(d)}(\tau_n^{(d)}) - M^{(d)}(\tau^{(d)})|.$$

By (23) and Lemma 3.5

$$\sup_{t \in \Delta_{d,n}} |M_n^{(d)}(t) - M^{(d)}(t)| \to 0 \text{ a.s. as } n \to \infty.$$

An application of Theorem 3.1 yields that $\tau_n^{(d)} \to \tau^{(d)}$ a.s., whence by continuity of $M^{(d)}$ we see that $Q_n(d) \to Q(d)$ a.s. Since there are only finitely many d it follows that

$$\sup_{1 \leq d \leq \bar{d}} |Q_n(d) - Q(d)| \to 0 \text{ a.s.} \tag{72}$$

Let $||\cdot||$ denote the L_2-Norm. Notice that $f = f_{\tau^{(d_0)},\alpha^{(d_0)}}$, since d_0 is the true number of steps. By definition (1) of $D = D^{(d)}$ it is

$$D^{(d)}(\tau^{(d)}, \alpha^{(d)}) = ||f_{\tau^{(d_0)},\alpha^{(d_0)}} - f_{\tau^{(d)},\alpha^{(d)}}||^2 \begin{cases} = 0, & d = d_0 \\ > 0, & d \neq d_0, \end{cases} \tag{73}$$

where the second relation holds, because the L_2-distance between two step-functions is zero if and only if both coincide as already stated above. On the other hand, it follows from (6) with $t := \tau^{(d)}$ that

$$D^{(d)}(\tau^{(d)}, \alpha^{(d)}) = ||f||^2 - M^{(d)}(\tau^{(d)})^2$$

and thus (73) guarantees that

$$Q(d) \begin{cases} = ||f||, & d = d_0 \\ < ||f||, & d \neq d_0. \end{cases} \tag{74}$$

So, d_0 is the unique maximizing point of Q and (72) in combination with Corollary 2.3 of Ferger [5] ensures that $\hat{d}_n \to d_0$ a.s. In fact, we arrive at the desired result (71), for $\hat{d}_n$ and d_0 are natural numbers. □

The following result yields a consistent estimate for the positions of the jumps and the pertaining levels if the number of jumps is unknown, but with known upper bound.

Corollary 4.2. Under the assumptions of Theorem 4.1 it follows that

$$(\tau_n^{(\hat{d}_n)}, \alpha_n^{(\hat{d}_n)}) \to (\tau^{(d_0)}, \alpha^{(d_0)}) \text{ a.s.}$$

Proof. Since

$$\begin{aligned} &\{(\tau_n^{(d_0)}, \alpha_n^{(d_0)}) \to (\tau^{(d_0)}, \alpha^{(d_0)}), n \to \infty\} \cap \{\hat{d}_n = d_0 \text{ for eventually all } n \in \mathbb{N}\} \\ &\subseteq \{(\tau_n^{(\hat{d}_n)}, \alpha_n^{(\hat{d}_n)}) \to (\tau^{(d_0)}, \alpha^{(d_0)}), n \to \infty\} \end{aligned}$$

the assertion follows immediately from Theorem 3.1 and Theorem 4.1. □

5. SIMULATION

In this section, we present results of a small simulation study for the estimator τ_n that is based on an explicit 'true' pdf with two jump points ($d = 2$). Here we confine ourselves to a simple and specific case. Considerations for a broad simulation study are given at the end of this section. The underlying true pdf f of a virtual set of datapoints shall be

$$f(x) = \begin{cases} 0.5 & 0 \le x \le 0.25 \\ 2 & 0.25 < x \le 0.5 \\ 0.75 & 0.5 < x \le 1. \end{cases} \tag{75}$$

In the notation of this article, that is $(\tau_1, \tau_2) = (0.25, 0.5)$ and $(\alpha_0, \alpha_1, \alpha_2) = (0.5, 2, 0.75)$. The simulation procedure is as follows. For given sample sizes, we draw samples of random numbers that are distributed according to the above pdf. Specifically, we draw 10000 replications for each sample size. Then for each sample, by assuming $d = 2$ is known, an estimate $(\tau_{1,n}, \tau_{2,n})$ is calculated by maximizing the function M_n from (8) on a triangular grid $\tilde{\Delta}$ that is defined as follows:

$$\tilde{\Delta} := \left\{(t_1, t_2) \in E^2 : 0 < t_1 < t_2 < 1\right\}, \text{ where } E := \{0.01, 0.02, 0.03, \ldots, 0.99\}.$$

The grid is chosen to achieve computational tractability but also ensures that $\tilde{\Delta}$ is a true subset of $\Delta_{n,2}$ of type (14) even for $n = 50$, since for $(t_1, t_2) \in \tilde{\Delta}$ holds $t_{i+1} - t_i \ge 0.1 > 50^{-2\cdot 0.8}$, $i = 0, 1, 2$. The use of a grid instead of a continuous domain induces some error which is a limitation of this simulation study. Nevertheless, aggregated results over 10000 replications will give a reasonable impression of the order of empirical performance measures for the estimator $(\tau_{1,n}, \tau_{2,n})$. All simulations and calculations were done in Matlab R2014a [Ref: MATLAB and Statistics Toolbox Release 2014a, The MathWorks, Inc., Natick, Massachusetts, United States.].

n	$\tau_{1,n}$ Mean	$\tau_{1,n}$ Bias	$\tau_{1,n}$ MSE	$\tau_{2,n}$ Mean	$\tau_{2,n}$ Bias	$\tau_{2,n}$ MSE
50	0.3290	0.0790	0.0223	0.4476	-0.0524	0.0172
100	0.2813	0.0313	0.0048	0.4732	-0.0268	0.0050
150	0.2644	0.0144	0.0015	0.4850	-0.0150	0.0021
200	0.2581	0.0081	0.0006	0.4818	-0.0182	0.0010
300	0.2538	0.0038	0.0002	0.4949	-0.0051	0.0003
500	0.2518	0.0018	<0.0001	0.4978	-0.0022	0.0001

Tab. 1. Simulation results for estimates $(\tau_{1,n}, \tau_{2,n})$ of $(\tau_1, \tau_2) = (0.25, 0.5)$ from pdf (75); 10000 replications in each simulation; n denotes sample size; MSE = mean squared error.

The results of the simulations are presented in Table 1. We calculated the empirical mean, bias, and empirical mean squared error (MSE) of $(\tau_{1,n}, \tau_{2,n})$ over all 10000 replications for each given sample size. For sample sizes of 150 and above, the MSE is of order 10^{-3} or below for both $\tau_{1,n}$ and $\tau_{2,n}$. However, for sample sizes 50 or 100, we observe quite notable bias. For all sample sizes, the bias is positive for the smaller $\tau_{1,n}$, and negative for the greater $\tau_{2,n}$.

A specific example estimate f_{τ_n,α_n} of f along with a normal kernel density estimate can be found in Figure 1. (The normal kernel bandwidth is the bandwidth that is theoretically optimal for estimating normal densities. Other choices of bandwidth couldn't improve the fit significantly.)

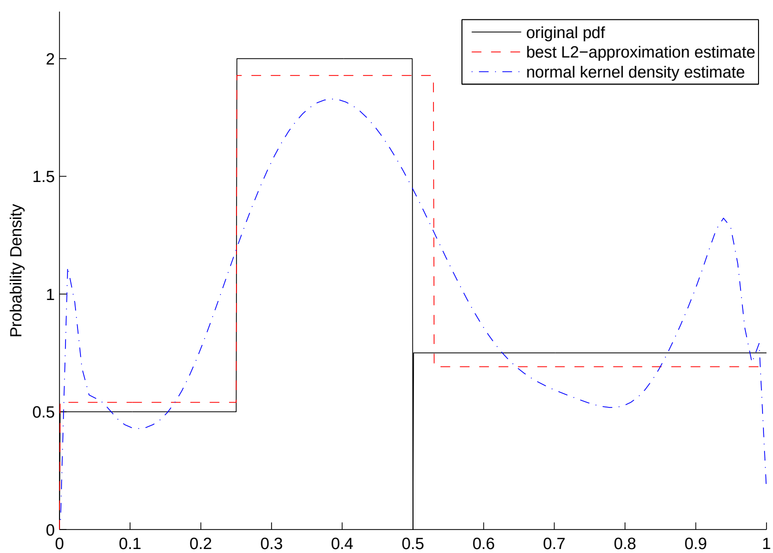

Fig. 1. example density estimates of pdf (75), sample size n=200.

A full comprehensive simulation study including comparisons with histogram estimators is beyond the scope of this article. The performance of the estimators will likely depend on the true vs. assumed number of jump points, the distance between jump points and the jump heights. All these influences need to be considered at the same time in a thoroughly conducted simulation study that may be included in future work.

(Received April 1, 2016)

REFERENCES

[1] P. Billingsley: Convergence of Probability Measures. Second Edition. John Wiley and Sons, Inc., New York 1999. DOI:10.1002/9780470316962

[2] Z. Birnbaum and A. Marshall: Some multivariate Chebyshev inequalities with extensions to continuous parameter processes. Ann. Math. Statist. *32* (1961), 687–703. DOI:10.1214/aoms/1177704964

[3] C. K. Chu and P. E. Cheng: Estimation of jump points and jump values of a density function. Statist. Sinica *6* (1996), 79–95.

[4] R. .M. Dudley: Uniform Central Limit Theorems. Cambridge University Press, New York 1999. DOI:10.1017/cbo9780511665622

[5] D. Ferger: Minimum distance estimation in normed linear spaces with Donsker-classes. Math. Methods Statist. *19* (2010), 246–266. DOI:10.3103/s1066530710030038

[6] D. Ferger: Arginf-sets of multivariate cadlag processes and their convergence in hyperspace topologies. Theory Stoch. Process. *20* (2015), 36, 13–41.

[7] P. Gaenssler and W. Stute: Wahrscheinlichkeitstheorie. Springer-Verlag, Berlin, Heidelberg, New York 1977.

[8] Y. Kanazawa: An optimal variable cell histogram. Comm. Statist. Theory Methods *17* (1988), 1401–1422. DOI:10.1080/03610928808829688

[9] Y. Kanazawa: An optimal variable cell histogram based on the sample spacings. Ann. Statist. *20* (1992), 291–304. DOI:10.1214/aos/1176348523

[10] H. L. Koul: Weighted Empirical Processes in Dynamic Nonlinear Models. Second Edition. Springer-Verlag, New York 2002. DOI:10.1007/978-1-4613-0055-7

[11] P. Massart: The tight constant in the Dvoretzky–Kiefer–Wolfowitz inequality. Ann. Probab. *18* (1990), 1269–1283. DOI:10.1214/aop/1176990746

[12] R. T. Rockefellar and R. J.-B. Wets: Variational Analysis. Springer-Verlag, Berlin, Heidelberg 1998.

[13] G. R. Shorack and J. A. Wellner: Empirical Processes With Applications to Statistics. John Wiley and Sons, New York, Chichester, Brisbane, Toronto, Singapore 1986.

Dietmar Ferger, Institut für Math. Stochastik, FR Mathematik, Fakultät Mathematik und Naturwissenschaften, TU Dresden, 01062 Dresden. Deutschland.
e-mail: dietmar.ferger@tu-dresden.de

John Venz, Behaviorale Epidemiologie, FR Psychologie, Fakultät Mathematik und Naturwissenschaften, TU Dresden, 01062 Dresden. Deutschland.
e-mail: john.venz@tu-dresden.de