Journal of Multivariate Analysis 107 (2012) 71–89

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Nonparametric estimation of multivariate scale mixtures of uniform densities

Marios G. Pavlides [a,*], Jon A. Wellner [b]

[a] Centre for Statistical Science and Operational Research, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, UK
[b] Department of Statistics, University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

Article history:
Received 7 May 2010
Available online 10 January 2012

AMS 2000 subject classifications:
62G05
62G07
62G20
62F20
62H12

Keywords:
Nonparametric estimation
Monotonicity
Multivariate
Minimax
Consistency
Uniform
Mixture

ABSTRACT

Suppose that $\boldsymbol{U} = (U_1, \ldots, U_d)$ has a Uniform$([0, 1]^d)$ distribution, that $\boldsymbol{Y} = (Y_1, \ldots, Y_d)$ has the distribution G on $\mathbb{R}_+^d$, and let $\boldsymbol{X} = (X_1, \ldots, X_d) = (U_1 Y_1, \ldots, U_d Y_d)$. The resulting class of distributions of $\boldsymbol{X}$ (as G varies over all distributions on $\mathbb{R}_+^d$) is called the *Scale Mixture of Uniforms* class of distributions, and the corresponding class of densities on $\mathbb{R}_+^d$ is denoted by $\mathcal{F}_{\mathrm{SMU}}(d)$. We study maximum likelihood estimation in the family $\mathcal{F}_{\mathrm{SMU}}(d)$. We prove existence of the MLE, establish Fenchel characterizations, and prove strong consistency of the almost surely unique maximum likelihood estimator (MLE) in $\mathcal{F}_{\mathrm{SMU}}(d)$. We also provide an asymptotic minimax lower bound for estimating the functional $f \mapsto f(\boldsymbol{x})$ under reasonable differentiability assumptions on $f \in \mathcal{F}_{\mathrm{SMU}}(d)$ in a neighborhood of $\boldsymbol{x}$. We conclude the paper with discussion, conjectures and open problems pertaining to global and local rates of convergence of the MLE.

1. Introduction and summary

Fix a non-negative integer k, and suppose that $X_1, \ldots, X_n$ are i.i.d. random variables distributed according to a density in the convex family of *k-monotone* densities (with respect to Lebesgue measure) on $(0, \infty)$:

$$\mathcal{F}_k := \left\{ f_{k,G}(\cdot) \equiv \int_0^\infty k \frac{(y - \cdot)_+^{k-1}}{y^k} \, dG(y) \;\middle|\; G \in \mathcal{G}_1 \right\}, \tag{1.1}$$

where $\mathcal{G}_1$ will denote the set of all distribution functions on $(0, \infty)$ grounded at 0. Here, we use the notation $x_+ \equiv x \cdot 1_{[x \geq 0]}$ for any $x \in \mathbb{R}$. It has been shown by Williamson [59] that the family $\mathcal{F}_k$ is identifiably indexed by $\mathcal{G}_1$. In other words, if G_1, G_2 are distinct elements in $\mathcal{G}_1$, then $f_{k,G_1}(\cdot)$ and $f_{k,G_2}(\cdot)$ differ on a Lebesgue non-null set. Note that $\mathcal{F}_k$ is exactly the collection of all scale mixtures of Beta $(1, k)$ densities.

The Beta $(1, 1)$ distribution is the standard uniform distribution, $U(0, 1)$. Therefore, the class $\mathcal{F}_1$ coincides with the class of all scale mixtures of uniform densities on $(0, \infty)$. A well-known theorem by Khintchine (see, e.g., [16, p.158])

* Corresponding author.
E-mail addresses: marios.pavlides@gmail.com (M.G. Pavlides), jaw@stat.washington.edu (J.A. Wellner).

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.01.001

asserts that the class of densities on $(0, \infty)$ with concave distribution functions is one and the same with our class $\mathcal{F}_1$. It can be seen that $\mathcal{F}_1$ is also the class of all upper semi-continuous, non-increasing densities on $(0, \infty)$. This class is induced by order restrictions, a term we use to explicitly mean that there exists a partial ordering $(\ll)$ on the common support $\mathcal{X}$ of the densities in $\mathcal{F}_1$ such that $f \in \mathcal{F}_1$ if and only if f is isotone with respect to this ordering: i.e., $f \in \mathcal{F}_1$ if and only if $f(x) \le f(y)$ whenever $x, y \in \mathcal{X}$ such that $x \ll y$. In this case, $(\ll)$ is the natural partial ordering, $(\ge)$, on $(0, \infty)$.

Non-increasing, upper semi-continuous densities (in short, *monotone densities*) arise naturally via connections with renewal theory and uniform mixing (see, e.g., [60]). Maximum likelihood estimation of monotone densities on $(0, \infty)$ was initiated by Grenander [18,19], with related work by Ayer et al. [3], Brunk [11], van Eeden [51–55]. Asymptotic theory of the MLE in $\mathcal{F}_1$ (the Grenander estimator) was developed by Prakasa Rao [44] with later contributions by [20,21,8,9,30]. See [4] for descriptions of the behavior of the Grenander estimator at zero.

Nonparametric estimation in families of densities described by order restrictions goes back at least to the work of [18,19,11,12,45], with further development by Wegman [56–58], Sager [48,49]. Also see the books by Barlow et al. [5] and Robertson et al. [46]. [40–43] addressed estimation in various order restricted classes of multivariate densities from the perspective of the excess mass approach studied previously by e.g., [48,49,36]. Polonik shows that (under reasonable assumptions) the MLE in such classes exists and coincides with an estimator he constructs and calls *the silhouette*. Forcing the elements of the class to be upper semi-continuous, the MLE is seen to be unique. Brunk [11] also gives a graphical construction of the maximum likelihood estimator, and establishes L_1-consistency of the MLE.

In this paper, our goal is to extend the notion of "monotone densities" to higher dimensions; i.e., to densities on $(0, \infty)^d$ with $d > 1$. Such an extension is not unique: for example, we may consider the family, $\mathcal{F}_{\mathrm{BDD}}(d)$, of "block-decreasing densities" (a term coined by Biau and Devroye [6]) that contains all upper-semicontinuous densities on $(0, \infty)^d$ that are non-increasing in each coordinate, while keeping all other coordinates fixed. This class was perhaps first introduced by Robertson [45]. The particular proper subclass of $\mathcal{F}_{\mathrm{BDD}}(d)$ studied here is the family $\mathcal{F}_{\mathrm{SMU}}(d)$ of all multivariate *scale mixtures of uniform densities*; i.e., the family of upper semi-continuous densities on $(0, \infty)^d$ of the form

$$f_G(\boldsymbol{x}) = \int_{(0,\infty)^d} \left(\frac{1}{|\boldsymbol{y}|} 1_{(\boldsymbol{0},\boldsymbol{y}]}(\boldsymbol{x}) \right) dG(\boldsymbol{y}), \quad \boldsymbol{x} \in (0, \infty)^d \tag{1.2}$$

for some $G \in \mathcal{G}_d$, the set of all distribution functions on $(0, \infty)^d$ that are grounded (zero) at $\mathbf{0}$; here we use the notation $|\boldsymbol{y}| \equiv \prod_{i=1}^d y_i$ for $\boldsymbol{y} = (y_1, \ldots, y_d)' \in (0, \infty)^d$. For any fixed $G \in \mathcal{G}_d$, it is clear that if $\boldsymbol{Y} = (Y_1, \ldots, Y_d)'$ is distributed according to G on $(0, \infty)^d$ and if $U_1, \ldots, U_d$ are i.i.d. $U(0, 1)$ (and independent of $\boldsymbol{Y}$), then the vector $\boldsymbol{X} := (U_1Y_1, \ldots, U_dY_d)$ is distributed according to $f_G(\cdot)$ on $(0, \infty)^d$.

Whereas the family $\mathcal{F}_{\mathrm{BDD}}(d)$ is characterized by order restrictions (and thus the results by Polonik apply), its subclass $\mathcal{F}_{\mathrm{SMU}}$ is not; as will be made more explicit in Section 2, densities in the class $\mathcal{F}_{\mathrm{SMU}}$ also satisfy non-negativity restrictions on their d-dimensional differences around all rectangles. Because of this additional *shape restriction*, estimation in this family requires separate treatment.

A univariate parallelism to the latter point would be to consider the family $\mathcal{F}_2$ in (1.1), induced by mixtures of triangular densities; this class can easily be seen to be exactly the class of all non-increasing, convex (and hence continuous) densities on $(0, \infty)$. Thus $\mathcal{F}_2 \subset \mathcal{F}_1$ is not an order-constrained class of densities, in contrast to its superclass $\mathcal{F}_1$. Convex densities arise in connection with Poisson process models for bird migration and scale mixtures of triangular densities (see, e.g., [26,2,32]). Estimation of non-increasing, convex densities on $(0, \infty)$ was apparently initiated by Anevski [1] and was further pursued by Anevski [2] and Jongbloed [28]. The asymptotic distribution theory and further characterizations of the nonparametric MLE of such a density and its first derivative at a fixed point (both under reasonable assumptions) was obtained by Groeneboom et al. [24,25]. These authors show that the local rate of convergence of the MLE of the functional $f \mapsto f(x)$ is of the order $n^{2/5}$, whereas the Grenander estimator (the MLE in $\mathcal{F}_1$) converges locally at the rate of only $n^{1/3}$.

The developments here have several motivations. One of these is to provide a multivariate family of shape-constrained densities with convergence rates for reasonable estimators which are (nearly) independent of the dimension d of the underlying space. As will be seen from the lower bound calculations in Section 4, it seems that the SMU class studied here may provide such a class. Another motivation comes from problems concerning multivariate analogues of interval censored data; see e.g. [27,61,62]. These apparently quite different models involve very similar mathematical considerations, and it might be helpful to develop methods for multivariate interval censored data problems by first studying the somewhat simpler SMU model.

Here is an outline of the remainder of the present paper. In Section 2, we provide characterizations of the family $\mathcal{F}_{\mathrm{SMU}}(d)$ that will prove useful in the sequel. Section 3 addresses existence, strong, pointwise consistency as well as L_1 and Hellinger consistency of a sequence of maximum likelihood estimators in $\mathcal{F}_{\mathrm{SMU}}(d)$. In Section 4, we derive a local asymptotic minimax lower bound for estimation of $f(\boldsymbol{x})$ at a fixed point $\boldsymbol{x}$ under for which f satisfies $\partial^d f(\boldsymbol{x})/(\partial x_1 \cdots \partial x_d) \neq 0$. The lower bound entails a rate of convergence of $n^{1/3}$ for all dimensions d and yields a constant depending on f which reduces to the known lower bound constant for $d = 1$. The paper concludes in Section 5 with a discussion of conjectures and open problems related with both the local (pointwise) and the global (L_1 and Hellinger) rates of convergence of the MLE in $\mathcal{F}_{\mathrm{SMU}}(d)$.

2. Properties of the Scale Mixtures of Uniform family of densities

2.1. Properties of $\mathcal{F}_{SMU}(d)$

A density function, f, on $(0,\infty)^d$ will be called a (multivariate) *Scale Mixture of Uniform densities* if there exists a distribution function, G, on $(0,\infty)^d$ such that

$$f(\boldsymbol{x}) = f_G(\boldsymbol{x}) = \int_{(0,\infty)^d} \frac{1}{|\boldsymbol{v}|} 1_{(\boldsymbol{0},\boldsymbol{v}]}(\boldsymbol{x})\, dG(\boldsymbol{v}) \tag{2.1}$$

$$= \int_{\boldsymbol{v}\geq\boldsymbol{x}} \frac{1}{|\boldsymbol{v}|}\, dG(\boldsymbol{v}) \quad \text{for all } \boldsymbol{x} \in (0,\infty)^d. \tag{2.2}$$

It is clear from (2.2) that a SMU density is also a block-decreasing density: $f_G(\cdot)$ is non-increasing in each coordinate, while keeping all other coordinates fixed. Also, the map $G \mapsto f_G$ is identifiable in the following sense: if $G_1 \neq G_2$, then $f_{G_1} \neq f_{G_2}$ on a set of positive Lebesgue measure; also see Theorem 2.3 below. The following lemma gives a formal statement of a slightly more general result. The proof is standard.

Lemma 2.1. *Two upper semi-continuous and block-decreasing functions f and g on $\mathbb{R}^d$ differ nowhere in the interior of their support or else on a Lebesgue non-negligible set.*

The distribution function F_G corresponding to $\boldsymbol{X} \sim f_G$ is given by

$$F_G(\boldsymbol{x}) = \int_{(0,\infty)^d} \frac{|\boldsymbol{x}\wedge\boldsymbol{v}|}{|\boldsymbol{v}|}\, dG(\boldsymbol{v}), \tag{2.3}$$

where $\leq$ denotes the natural partial ordering on $\mathbb{R}^d$, while

$$\boldsymbol{x}\wedge\boldsymbol{v} \equiv (x_1,\dots,x_d)\wedge(v_1,\dots,v_d) = (\min\{x_1,v_1\},\dots,\min\{x_d,v_d\}),$$

and $\boldsymbol{x}\vee\boldsymbol{v} \equiv (x_1,\dots,x_d)\vee(v_1,\dots,v_d) = (\max\{x_1,v_1\},\dots,\max\{x_d,v_d\})$. The distribution function F_G of $\boldsymbol{X}$ is generally not concave when $d > 1$, unlike the case when $d = 1$. An SMU density (and a block-decreasing density, in general) can possibly diverge at the origin, whereas the pointwise bound $f(\boldsymbol{x}) \leq 1/|\boldsymbol{x}|$ holds since, for $\boldsymbol{x} \in (0,\infty)^d$ we have

$$1 = \int_{(0,\infty)^d} f(\boldsymbol{y})\, d\boldsymbol{y} \geq \int_{(\boldsymbol{0},\boldsymbol{x}]} f(\boldsymbol{y})\, d\boldsymbol{y} \geq |\boldsymbol{x}| f(\boldsymbol{x}).$$

Further, a d-dimensional analogue of the proof of [13, Theorem 6.2, p. 173] can be used to show that

$$\lim_{|\boldsymbol{x}|\to\infty}\{|\boldsymbol{x}| f(\boldsymbol{x})\} = \lim_{\boldsymbol{x}\downarrow\boldsymbol{0}}\{|\boldsymbol{x}| f(\boldsymbol{x})\} = 0, \tag{2.4}$$

whenever f is a block-decreasing density on $(0,\infty)^d$.

For any two points $\boldsymbol{x},\boldsymbol{y} \in [0,\infty)^d$, such that $\boldsymbol{x} \leq \boldsymbol{y}$, we write $[\boldsymbol{x},\boldsymbol{y}] \equiv [x_1,y_1]\times\cdots\times[x_d,y_d]$, $[\boldsymbol{x},\boldsymbol{y}) \equiv [x_1,y_1)\times\cdots\times[x_d,y_d)$, $(\boldsymbol{x},\boldsymbol{y}] \equiv (x_1,y_1]\times\cdots\times(x_d,y_d]$, $(\boldsymbol{x},\boldsymbol{y}) \equiv (x_1,y_1)\times\cdots\times(x_d,y_d)$ for the natural closed, lower-closed upper open, lower open upper closed, and open rectangles respectively. Note that the closed rectangle $[\boldsymbol{x},\boldsymbol{y}]$ has (at most) 2^d vertices, the points $\boldsymbol{u} = (u_1,\dots,u_d)$ where each u_i is either x_i or y_i. Following [7], we write $\operatorname{sgn}_{[\boldsymbol{x},\boldsymbol{y}]}(\boldsymbol{u}) \in \{-1,1\}$, the signum of the vertex $\boldsymbol{u}$, according as the number of i, $1 \leq i \leq d$, satisfying $u_i = x_i$ is odd or even respectively.

Thus any two vertices defining an edge of the rectangle have alternating signs. Then, if $\boldsymbol{u} = (u_1,\dots,u_d)$ is some vertex of $[\boldsymbol{x},\boldsymbol{y}]$ and $\delta \in \{-1,+1\}$ is its signum, then $(\delta,\boldsymbol{u})$ is an element of the set

$$\Delta_d[\boldsymbol{x},\boldsymbol{y}] = \left\{ \left((-1)^{\sum_{i=1}^{d}\left\{\mathbb{1}_{[u_i=x_i]}\right\}}, \boldsymbol{u} \right) \;\middle|\; \boldsymbol{u} \in \{x_1,y_1\}\times\cdots\times\{x_d,y_d\} \right\}.$$

Definition 2.1. For an upper semicontinuous and coordinatewise decreasing function $g\colon (0,\infty)^d \to [0,\infty)$ define the g-volume of a (possibly degenerate) rectangle $[\boldsymbol{x},\boldsymbol{y})$ by:

$$V_g[\boldsymbol{x},\boldsymbol{y}) = \sum_{(\delta,\boldsymbol{u})\in\Delta_d[\boldsymbol{x},\boldsymbol{y}]} \{\delta g(\boldsymbol{u})\}, \tag{2.5}$$

provided that g is defined and is finite for all $\boldsymbol{u}$ in the summand. Correspondingly, for an upper semicontinuous and coordinatewise increasing function $g\colon (0,\infty)^d \to [0,\infty)$, we define the g-volume of a rectangle $(\boldsymbol{x},\boldsymbol{y}]$ by the sum on the right side of (2.5).

It is easily seen that for an SMU density, f_G, the f_G-volume of any rectangle $[\boldsymbol{x}, \boldsymbol{y})$ is always of the sign $(-1)^d$: indeed, consider (2.2) and observe that

$$(-1)^d V_{f_G}[\boldsymbol{x}, \boldsymbol{y}) = \int_{[\boldsymbol{x},\boldsymbol{y})} \frac{1}{|\boldsymbol{v}|}\, dG(\boldsymbol{v}) \geq 0. \tag{2.6}$$

From (2.6), or, alternatively, from the fact that the class of sets $[\boldsymbol{x}, \boldsymbol{y})$ is a π-system which generates the Borel σ-field of subsets of $[0, \infty)^d$ and then extending as in [7], it is clear that $(-1)^d V_f$ extends uniquely to a (non-negative) measure on the Borel σ-field $\mathcal{B}_+^d = \mathcal{B}^d \cap [0, \infty)^d$ given by

$$(-1)^d V_f(A) = \int_A \frac{1}{|\boldsymbol{v}|} dG(\boldsymbol{v}) \quad \text{for } A \in \mathcal{B}_+^d;$$

in particular,

$$(-1)^d V_f(\boldsymbol{x}, \boldsymbol{y}] = \int_{(\boldsymbol{x},\boldsymbol{y}]} \frac{1}{|\boldsymbol{v}|} dG(\boldsymbol{v}).$$

This argument extends easily to an arbitrary upper semicontinuous function g with the $(-1)^d g$-volumes of all rectangles $[\boldsymbol{x}, \boldsymbol{y})$ non-negative.

Lemma 2.2. *Suppose that g is a non-negative, upper semi-continuous function satisfying $(-1)^d V_g[\boldsymbol{x}, \boldsymbol{y}) \geq 0$ for all lower-closed upper open rectangles $[\boldsymbol{x}, \boldsymbol{y})$, and vanishing if any coordinate tends to ∞. Then $(-1)^d V_g$ can be extended to a countably additive measure on $\mathcal{B}_+^d$.*

Of course it is easy to exhibit a block-decreasing density that is not an SMU density: consider the uniform density on the closed triangle in $\mathbb{R}_+^2$ with vertices (0, 0), (0, 1) and (1, 0). Then,

$$(-1)^2 V_f[(1/8, 1/8), (1/2, 3/4)) = -2 < 0,$$

showing that this density is not an SMU density, even though it is block-decreasing.

The following theorem establishes identifiability of the mixing distribution G as well as providing a useful characterization of SMU densities.

Theorem 2.3.

(a) *For the class of SMU densities $\mathcal{F}_{SMU}(d) = \{f_G : G \in \mathcal{G}_d\}$ with f_G as given in (2.1), $f \in \mathcal{F}_{SMU}(d)$ if and only if $f \equiv f_G$, where $G \in \mathcal{G}_d$ is given by*

$$G(\boldsymbol{x}) = \int_{(0,\infty)^d} (-1)^d V_f(\boldsymbol{u}, \boldsymbol{x}] \cdot \mathbb{1}_{[\boldsymbol{u} \leq \boldsymbol{x}]}\, d\boldsymbol{u}. \tag{2.7}$$

Thus there is a one-to-one correspondence between $G \in \mathcal{G}_d$ and $f_G \in \mathcal{F}_{SMU}(d)$.

(b) *Suppose that the Lebesgue density f on $(0, \infty)^d$ is such that it converges to zero in each coordinate, while keeping all other coordinates fixed. Then, f is an SMU density if and only if $(-1)^d V_f[\boldsymbol{x}, \boldsymbol{y}) \geq 0$ for all $\boldsymbol{0} \leq \boldsymbol{x} \leq \boldsymbol{y}$.*

Proof. (a) Suppose that $f \equiv f_G$, for $G \in \mathcal{G}_d$ (recall that this implies that $G(\boldsymbol{0}) = 0$), is an SMU density evaluated at an arbitrary $\boldsymbol{x} \in (0, \infty)^d$ as:

$$f(\boldsymbol{x}) = \int_{(0,\infty)^d} \frac{1}{|\boldsymbol{y}|} \mathbb{1}_{(\boldsymbol{0},\boldsymbol{x}]}\, dG(\boldsymbol{y}) = \int_{y_1 \geq x_1} \cdots \int_{y_d \geq x_d} \frac{1}{|\boldsymbol{y}|}\, dG(\boldsymbol{y}), \tag{2.8}$$

so that $df(\boldsymbol{x}) = (-1)^d |\boldsymbol{x}|^{-1}\, dG(\boldsymbol{x})$ and thus,

$$\begin{aligned} G(\boldsymbol{x}) &= \int_{(0,\infty)^d} \mathbb{1}_{(\boldsymbol{0},\boldsymbol{x}]}(\boldsymbol{y}) |\boldsymbol{y}|\, d\{(-1)^d f(\boldsymbol{y})\} \\ &= \int_{(\boldsymbol{0},\boldsymbol{x}]} \int_{(\boldsymbol{0},\boldsymbol{x}]} \mathbb{1}_{(\boldsymbol{0},\boldsymbol{y}]}(\boldsymbol{u})\, d\boldsymbol{u}\, d\{(-1)^d f(\boldsymbol{y})\} \\ &= \int_{(\boldsymbol{0},\boldsymbol{x}]} \left\{ \int_{\boldsymbol{y} \in (\boldsymbol{u},\boldsymbol{x}]} d\{(-1)^d f(\boldsymbol{y})\} \right\} d\boldsymbol{u} \\ &= \int_{(\boldsymbol{0},\boldsymbol{x}]} (-1)^d V_f(\boldsymbol{u}, \boldsymbol{x}]\, d\boldsymbol{u}, \end{aligned}$$

where the second to last equality follows by Fubini–Tonelli.

We will now show that G is unique: suppose that (2.8) above holds for $G = G_i \in \mathcal{G}_d$ and $i = 1, 2$. Recall that this implies that $G_1(\mathbf{0}) = G_2(\mathbf{0}) = 0$ and, thus, $G_0(\cdot) \equiv G_1(\cdot) - G_2(\cdot)$ is such that $G_0(\mathbf{0}) = 0$, $\int_{(0,\infty)^d} G_0(\boldsymbol{x})\, d\boldsymbol{x} = 0$ and

$$0 = \int_{(0,\infty)^d} \frac{1}{|\boldsymbol{y}|} \mathbb{1}_{(\mathbf{0},\boldsymbol{x}]}\, dG_0(\boldsymbol{y}) = \int_{(\mathbf{0},\boldsymbol{x}]} \frac{1}{|\boldsymbol{y}|} dG_0(\boldsymbol{y}) \tag{2.9}$$

holds for all $\boldsymbol{x} \in (0,\infty)^d$ and, thus, necessarily $G_0(\boldsymbol{x})$ has to be independent of $\boldsymbol{x}$ and therefore everywhere equal to its value at $\mathbf{0}$: $G_0(\mathbf{0}) = 0$. This completes the assertion of uniqueness, since $G_1 \equiv G_2$.

(b) If f is in $\mathcal{F}_{\mathrm{SMU}}$, there exists $G \in \mathcal{G}_d$ such that

$$f(\boldsymbol{x}) = \int_{(0,\infty)^d} \frac{1}{|\boldsymbol{y}|} \mathbb{1}_{(\mathbf{0},\boldsymbol{y}]}(\boldsymbol{x})\, dG(\boldsymbol{y}) = \int_{\boldsymbol{y} \geq \boldsymbol{x}} \frac{1}{|\boldsymbol{y}|}\, dG(\boldsymbol{y}),$$

so that it is easily seen that $(-1)^d V_f[\boldsymbol{x},\boldsymbol{y}) = \int_{[\boldsymbol{x},\boldsymbol{y})} |\boldsymbol{y}|^{-1}\, dG(\boldsymbol{y}) \geq 0$ holds true for all $\mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{y}$.

On the other hand, assume that the Lebesgue density f is such that it converges to zero in each coordinate, while keeping all other coordinates fixed, and satisfies $(-1)^d V_f[\boldsymbol{x},\boldsymbol{y}] \geq 0$ for all $\mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{y}$. By Lemma 2.2, this implies that for $\boldsymbol{x}_1 \leq \boldsymbol{x}_2 \leq \boldsymbol{x}$, elements of $(0,\infty)^d$, we have $(-1)^d V_f[\boldsymbol{x}_1,\boldsymbol{x}) \geq (-1)^d V_f[\boldsymbol{x}_2,\boldsymbol{x})$ and, letting $\boldsymbol{x} \to \infty$, this yields $f(\boldsymbol{x}_1) \geq f(\boldsymbol{x}_2)$ because we assumed that f vanishes as any one of its coordinates diverges to infinity, so that $V_f[\boldsymbol{x}_i,\boldsymbol{x}) \to (-1)^d f(\boldsymbol{x}_i)$ for $i \in \{1,2\}$. Thus, f is block-decreasing.

Hence, by appealing to part (a), it thus suffices to show that G, as defined on $(0,\infty)^d$ by (2.7) is a valid distribution function. Indeed, this is easily shown along the lines of the following sketch. In particular, (i) G is grounded at $\mathbf{0}$ trivially by inspection: $G(\mathbf{0}) = 0$. (ii) By virtue of the fact that f is block-decreasing, $0 \leq \lim_{|\boldsymbol{x}|\to\infty} f(\boldsymbol{x}) \leq \lim_{|\boldsymbol{x}|\to\infty}\{1/|\boldsymbol{x}|\} = 0$ is true and this can be used to show straightforwardly that $\lim_{x_1 \wedge \cdots \wedge x_d \to \infty} G(x_1,\ldots,x_d) = 1$. (iii) Similarly, it is an easy task to show that $V_G(\boldsymbol{x},\boldsymbol{y}] \geq 0$ for all $\mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{y}$. Conditions (i)–(iii) are necessary and sufficient for G to be a bona-fide distribution function. This completes the proof. $\square$

2.2. Lebesgue measurability of block-decreasing functions

Now we note a technical fact concerning the (Lebesgue) measurability of block-decreasing functions which will be needed in our proofs in Section 3.2.

Proposition 2.4. *Let f be a real-valued, non-negative function on $(0,\infty)^d$ that is non-increasing and convergent to zero in each coordinate x_j, keeping all other coordinates fixed, as x_j coordinate tends to ∞. Then:*

(a) *f is Lebesgue-measurable.*
(b) *There exists such a function f that is not Borel-measurable. Such an f exists with f also satisfying* $\sup\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in (0,\infty)^d\} < \infty$.

Proof. Proposition 2.4 (a) follows from Theorem 3 of [31]. Proposition 2.4 (b) is standard and follows from Proposition 1.2.2 in [50]. $\square$

3. Existence and consistency of the MLE

Let $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$ be i.i.d. random vectors distributed according to some density $f_0 = f_{G_0} \in \mathcal{F}_{\mathrm{SMU}}(d)$ where f_0 is unknown. Our goal is to estimate the unknown SMU density, f_0, based on $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$. We will be interested in maximizing the likelihood function $f \mapsto \prod_{i=1}^n f(\boldsymbol{X}_i)$ or, equivalently, the log-likelihood function $f \mapsto n\mathbb{P}_n \log\{f(\boldsymbol{X})\}$ over $f \in \mathcal{F}_{\mathrm{SMU}}(d)$ where $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{\boldsymbol{X}_i}$ is the empirical measure of the data. Any such maximizer, $\widehat{f}_n \in \mathcal{F}_{\mathrm{SMU}}(d)$, should one exist, will be called a (nonparametric) *maximum likelihood estimator* of f_0, based on $\boldsymbol{X}_1,\ldots,\boldsymbol{X}_n$. Since $f_0 = f_{G_0}$ is given by (2.1) it follows from Theorem 2.3 that estimation of $f_0 \in \mathcal{F}_{\mathrm{SMU}}$ is equivalent to estimation of G_0.

3.1. On existence and uniqueness of an MLE

We begin with a definition followed by the main theorem of this subsection.

Definition 3.1 (*Rectangular Grid Generated by Data*)**.** Suppose that $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$ are (fixed or random) elements in $(0,\infty)^d$ and suppose that $\boldsymbol{x}_i = (x_{i1},\ldots,x_{id})'$ where $i = 1,2,\ldots,n$. Define the matrix $A = [x_{ij}] \in M_{n\times d}((0,\infty))$ whose ith row is exactly $\boldsymbol{x}_i'$, for $i \in \{1,2,\ldots,n\}$. Also let $A^\sharp = \{\, (x_{(i_1),1}, x_{(i_2),2},\ldots,x_{(i_d),d}) \mid i_1,\ldots,i_d \in \{1,2,\ldots,n\}\}$ denote *the rectangular grid generated by* A, where $x_{(i),j}$ denotes the ith smallest element among $x_{1j},\ldots,x_{nj}$ where $i \in \{1,2,\ldots,n\}$ and $j \in \{1,2,\ldots,d\}$. In particular, $\boldsymbol{x}_* = (x_{(1),1}, x_{(1),2},\ldots,x_{(1),d})$ and $\boldsymbol{x}^* = (x_{(n),1}, x_{(n),2},\ldots,x_{(n),d})$ denote the element-wise minimum and maximum of $\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n$, respectively. For each fixed $j \in \{1,2,\ldots,d\}$, let $n_j(A) := \mathrm{card}(\{x_{i,j} \mid i = 1,2,\ldots,n\})$, and notice that we have: $\mathrm{card}(A^\sharp) = \prod_{j=1}^d n_j(A) \equiv N \leq n^d$.

Theorem 3.1 (*Existence and Characterization of an MLE in $\mathcal{F}_{SMU}(d)$*)**.**

(a) *A maximum likelihood estimator (MLE), $\widehat{f}_n \equiv f_{\widehat{G}_n} \in \mathcal{F}_{SMU}(d)$ of $f_0 \equiv f_{G_0} \in \mathcal{F}_{SMU}(d)$ almost surely exists, where $\widehat{G}_n \in \mathcal{G}_d$ is a purely-atomic probability measure, with at most n atoms, all of which are concentrated on $A^\sharp$—the rectangular grid generated by the data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$.*

(b) *For almost all ω, the unique MLE, $\widehat{f}_n \equiv f_{\widehat{G}_n} \in \mathcal{F}_{SMU}(d)$, is completely characterized by the following Fenchel conditions:*

$$\mathbb{P}_n\left\{\frac{\mathbb{1}_{[\boldsymbol{X}\le \boldsymbol{x}]}}{\widehat{f}_n(\boldsymbol{X})}\right\} \le |\boldsymbol{x}|; \quad \textit{for all } \boldsymbol{x} \in (0,\infty)^d, \tag{3.1}$$

$$\textit{and} \quad \mathbb{P}_n\left\{\frac{\mathbb{1}_{[\boldsymbol{X}\le \boldsymbol{y}]}}{\widehat{f}_n(\boldsymbol{X})}\right\} = |\boldsymbol{y}|; \quad \textit{if and only if} \tag{3.2}$$

$$\boldsymbol{y} \in (0,\infty)^d \quad \textit{satisfies } \widehat{G}_n(\{\boldsymbol{y}\}) > 0; \ \textit{or, equivalently,}$$
$$(-1)^d \lim_{\epsilon\downarrow 0}\left\{V_{\widehat{f}_n}[\boldsymbol{y}, \boldsymbol{y}+\epsilon\boldsymbol{1})\right\} > 0.$$

Maximum likelihood estimation in mixture models has been studied in general by Lindsay [34], and this material is nicely summarized in [35, Chapter 5]. To prove the present theorem, we will therefore appeal to the results in [35, Chapter 5] and [47]. We begin with three lemmas.

Lemma 3.2. *The support set $\mathcal{Y} \equiv \mathrm{supp}(\widehat{G}_n)$ of the mixing measure $\widehat{G}_n$ of any MLE $\widehat{f}_n$ is contained in the grid $A^\# \subset (0,\infty)^d$ generated by the observed data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$; i.e., $\mathcal{Y} \subset A^\#$.*

Proof. First we show that $\mathcal{Y} \subset (\boldsymbol{0}, \boldsymbol{X}^*]$ where $\boldsymbol{X}^* \equiv \boldsymbol{X}_1 \vee \cdots \vee \boldsymbol{X}_n$ and the maximums are taken coordinatewise. If $\widehat{f}_n$ maximizes $L_n(f) = n\mathbb{P}_n \log f(X)$ over $f \in \mathcal{F}_{\mathrm{SMU}}(d)$ and there is some $y \in (0,\infty)^d \setminus (\boldsymbol{0}, \boldsymbol{X}^*]$ with $y \in \mathcal{Y}$, then $\widehat{f}_n(y) > 0$. Since $\widehat{f}_n$ is block decreasing, this implies that $0 < \int_{(\boldsymbol{0},\boldsymbol{X}^*]}\widehat{f}_n(\boldsymbol{x})d\boldsymbol{x} \equiv \beta < 1$. Then consider $\tilde{f}(\boldsymbol{x}) \equiv (\widehat{f}_n(\boldsymbol{x})/\beta)1_{(\boldsymbol{0},\boldsymbol{X}^*]}(\boldsymbol{x})$; it is easily seen that $\tilde{f} \in \mathcal{F}_{\mathrm{SMU}}(d)$ and has greater likelihood than $\widehat{f}_n$, contradicting the assumption that $\widehat{f}_n$ maximizes the likelihood. Thus $\mathcal{Y} \subset (\boldsymbol{0}, \boldsymbol{X}^*]$, and we may restrict attention to the class of estimators with support contained in $(\boldsymbol{0}, \boldsymbol{X}^*]$, say $\mathcal{K}^*(d)$. Suppose that $\widehat{f}_n \in \mathcal{K}^*(d)$. Consider the mixing measure $\tilde{G}_n$ defined by

$$\tilde{G}_n \equiv \sum_{j:\,\boldsymbol{W}_j\in A^\#}\pi_j\delta_{\boldsymbol{W}_j}\Big/\sum_{j:\,\boldsymbol{W}_j\in A^\#}\pi_j \equiv C\sum_{j:\,\boldsymbol{W}_j\in A^\#}\pi_j\delta_{\boldsymbol{W}_j}$$

where

$$\pi_j \equiv (-1)^d V_{\widehat{f}_n}[\boldsymbol{W}_j, \boldsymbol{W}_j^+)\cdot|\boldsymbol{W}_j|, \quad \text{for } \boldsymbol{W}_j \in A^\#$$

where $\boldsymbol{W}_j^+ \in A^\#$ defines the smallest rectangle above and right of $\boldsymbol{W}_j$ in the partition of $[\boldsymbol{0}, \boldsymbol{X}^*]$ defined by the data. Then it is easy to see that

$$\tilde{f}(\boldsymbol{x}) = \int_{(0,\infty)^d}\frac{1}{|\boldsymbol{u}|}1_{(\boldsymbol{0},\boldsymbol{u}]}(\boldsymbol{x})d\tilde{G}_n(\boldsymbol{u})$$

satisfies

$$\begin{aligned}\tilde{f}(\boldsymbol{W}_j) &= C\sum_{k:\,\boldsymbol{W}_k\ge\boldsymbol{W}_j}\frac{\pi_j}{|\boldsymbol{W}_j|}\\ &= C\sum_{k:\,\boldsymbol{W}_k\ge\boldsymbol{W}_j}(-1)^d V_{\widehat{f}_n}[\boldsymbol{W}_j, \boldsymbol{W}_k)\\ &= C(-1)^d V_{\widehat{f}_n}[\boldsymbol{W}_j, 2\boldsymbol{X}^*) = C\widehat{f}_n(\boldsymbol{X}_j),\end{aligned}$$

and this implies that

$$\tilde{f}(\boldsymbol{x}) = C\sum_{j:\,\boldsymbol{W}_j\in A^\#}1_{(\boldsymbol{W}_j^-,\boldsymbol{W}_j]}(\boldsymbol{x})$$

where $\boldsymbol{W}_i^-$ defines the smallest rectangle below and to the left of $\boldsymbol{W}_j$ in the partition of $[\boldsymbol{0}, \boldsymbol{X}^*]$ defined by the data. If $\widehat{f}_n \ne \tilde{f}$, then there exists $\boldsymbol{y} \in (\boldsymbol{W}_j^-, \boldsymbol{W}_j]$ for some $\boldsymbol{W}_j \in A^\#$ such that $\widehat{f}_n(\boldsymbol{y}) \ne \tilde{f}(\boldsymbol{y})$, and then necessarily $\widehat{f}_n(\boldsymbol{y}) > \tilde{f}(\boldsymbol{y}) = \tilde{f}(\boldsymbol{W}_j)$.

This yields, since $\tilde{f}_n \in \mathcal{K}^*(d)$,

$$
\begin{aligned}
1 &= \int_{(\mathbf{0},\boldsymbol{X}^*]} \tilde{f}(\boldsymbol{x})d\boldsymbol{x} = C \sum_{j:\ \boldsymbol{W}_j \in A^\#} \left\{ \widehat{f}_n(\boldsymbol{W}_j) \int_{(\boldsymbol{W}_j^-,\boldsymbol{W}_j]} d\boldsymbol{x} \right\} \\
&< C \sum_{j:\ \boldsymbol{W}_j \in A^\#} \widehat{f}_n(\boldsymbol{W}_j) \int_{(\boldsymbol{W}_j^-,\boldsymbol{W}_j]} \widehat{f}_n(\boldsymbol{x})d\boldsymbol{x} = C \int_{(\mathbf{0},\boldsymbol{X}^*]} \widehat{f}_n(\boldsymbol{x})d\boldsymbol{x} = C
\end{aligned}
$$

since $f \in \mathcal{K}^*(d)$. Thus $\tilde{f}$ has a greater log-likelihood than $\widehat{f}_n$, and it follows that $\text{supp}(\widehat{G}_n) \subset A^\#$. $\square$

Now we can prove uniqueness of the MLEs $\widehat{f}_n$ and $\widehat{G}_n$.

Lemma 3.3. *There exists a set of points $\mathcal{Y} = \{\boldsymbol{y}_1, \ldots, \boldsymbol{y}_m\} \subset (0,\infty)^d$ with $m \le n$ such that a $\mathcal{F}_{SMU}(d)$ density $\widehat{f}_n$ with corresponding mixing measure $\widehat{G}_n$ is the MLE only if $\text{supp}(\widehat{G}_n) \subset \mathcal{Y}$. Thus any MLE has the form*

$$
\widehat{f}_n(\boldsymbol{x}) = \sum_{j=1}^m \pi_j \frac{1}{|\boldsymbol{y}_j|} 1_{(\mathbf{0},\boldsymbol{y}_j]}(\boldsymbol{x}) \tag{3.3}
$$

where $\pi_j \ge 0$, $\sum_{j=1}^m \pi_j = 1$. Moreover, the vector $(\widehat{f}_n(X_i))_{i=1}^n$ is unique.

Proof. As in [34,35], define $\Gamma(\boldsymbol{u}) \in (0,\infty)^n$ by

$$
\Gamma(\boldsymbol{u}) := \left(\frac{1}{|\boldsymbol{u}|} \mathbb{1}_{(\mathbf{0},\boldsymbol{u}]}(\boldsymbol{X}_1), \ldots, \frac{1}{|\boldsymbol{u}|} \mathbb{1}_{(\mathbf{0},\boldsymbol{u}]}(\boldsymbol{X}_n) \right),
$$

and define the set $\Gamma \equiv \{\Gamma(\boldsymbol{u}) \mid \boldsymbol{u} \in (0,\infty)^d\}$. Then Γ is a closed and bounded, hence compact, subset of $[0,\infty)^n$. Thus by Rockafellar [47, Theorem 17.2] $\overline{\text{conv}(\Gamma)} = \text{conv}(\overline{\Gamma}) = \text{conv}(\Gamma)$ is also a compact subset of $[0,\infty)^n$. Thus the continuous function $\prod_{i=1}^n z_i$ attains its supremum on $\text{conv}(\Gamma)$. Let $S = \text{argmax}_{\boldsymbol{z}\in\text{conv}(\Gamma)} \sum_{i=1}^n \log z_i$. Since the intersection of Γ and the interior $(0,\infty)^n$ of $[0,\infty)^n$ is not empty, we have $S \subset (0,\infty)^n$. Since $\sum_{i=1}^n \log z_i$ is strictly concave, S consists of a single point, $\hat{\boldsymbol{f}} = (\hat{f}_i)_{i=1}^n > \mathbf{0}$. Therefore for any MLE $\widehat{f}_n$ it follows that the vector $(\widehat{f}_n(X_i))_{i=1}^n$ is unique. Note that the gradient of $\sum_{i=1}^n \log z_i$ at $\hat{\boldsymbol{f}}$ is proportional to $1/\hat{\boldsymbol{f}} \equiv (1/\hat{f}_i)_{i=1}^n$.

Now $\dim(\text{conv}(\Gamma)) = n$; if we consider the n points $\boldsymbol{u}_i = \boldsymbol{X_i}$, then the n vectors $\Gamma(\boldsymbol{u}_i) = (1_{(0,\boldsymbol{X}_i]}(\boldsymbol{X}_1), \ldots, 1_{(0,\boldsymbol{X}_i]}(\boldsymbol{X}_n))/|\boldsymbol{X}_i|$, $i = 1, \ldots, n$, are almost surely linearly independent. (In fact, the matrix M with rows $|\boldsymbol{X}_i|\Gamma(\boldsymbol{X}_i)$, $i = 1, \ldots, n$ has $\det(M) = 1$ a.s. if the $\boldsymbol{X}_i$'s are i.i.d. with any density f.) By Rockafellar [47, Theorem 27.4] the vector $1/\hat{\boldsymbol{f}}$ belongs to the normal cone of $\text{conv}(\Gamma)$ at $\hat{\boldsymbol{f}}$. Since $1/\hat{\boldsymbol{f}} > 0$ we have $\hat{\boldsymbol{f}} \in \partial(\text{conv}(\Gamma))$ and the plane τ defined by $\sum_{i=1}^n z_i/\hat{f}_i = n$ is a support plane of $\text{conv}(\Gamma)$ at $\hat{\boldsymbol{f}}$. Thus for $v_i = 1/(n\hat{f}_i)$, $i = 1, \ldots, n$, it follows that

$$
q(\boldsymbol{u}) \equiv |\boldsymbol{u}| - \sum_{i=1}^n v_i 1_{(\mathbf{0},\boldsymbol{u}]}(\boldsymbol{X}_i) \ge 0
$$

for all $\boldsymbol{u} \in [0,\infty)^d$ and $q(\boldsymbol{u}) = 0$ if $\boldsymbol{u} = \mathbf{0}$ or $\Gamma(\boldsymbol{u}) \in \tau$. We let $\mathcal{Y}$ denote the set of vectors $\boldsymbol{u}$ such that $\Gamma(\boldsymbol{u}) \in \tau$; i.e., $\Gamma(\mathcal{Y}) = \tau \cap \Gamma$.

The intersection $\tau \cap \text{conv}(\Gamma)$ is an exposed face of $\text{conv}(\Gamma)$; see e.g. [47, p. 162]. By Rockafellar [47, Theorem 18.3], $\tau \cap \text{conv}(\Gamma) = \text{conv}(\Gamma(\mathcal{Y}))$, and by Theorem 18.1, $\text{supp}(\widehat{G}_n) \subset \mathcal{Y}$. This implies that for any MLE $\widehat{f}_n$, the support of the corresponding mixing measure $\widehat{G}_n$ is a subset of $\mathcal{Y}$, and thus any MLE has form (3.3) with $\boldsymbol{y}_j \in \mathcal{Y}$ for $j = 1, \ldots, m$. To see that $m \le n$, note that $\boldsymbol{y}_j \in \mathcal{Y} \subset A^\#$ satisfy

$$
|\boldsymbol{y}_j| = \sum_{i=1}^n v_i 1_{(\mathbf{0},\boldsymbol{y}_j]}(\boldsymbol{X}_i) = \langle \boldsymbol{v}, |\boldsymbol{y}_j|\Gamma(\boldsymbol{y}_j)\rangle, \quad j = 1, \ldots, m. \tag{3.4}
$$

Suppose that the vectors $\{|\boldsymbol{y}_j|\Gamma(\boldsymbol{y}_j)\}_{j=1}^m$ are linearly dependent; i.e.,

$$
\sum_{j=1}^m b_j |\boldsymbol{y}_j| \Gamma(\boldsymbol{y}_j) = \mathbf{0}
$$

in $\mathbb{R}^n$ for some $b_j, j = 1, \ldots, m$. Since all the coordinates of the $|\boldsymbol{y}_j|\Gamma(\boldsymbol{y}_j)$ vectors take values in $\{0, 1\}$, this system of equations is algebraically equivalent to the same system in which all the b_j's take only integer values, i.e., $b_j \in \mathbb{Z}$ for $j = 1, \ldots, m$.

Then it follows on the one hand that

$$\sum_{j=1}^{m} b_j \langle \boldsymbol{v}, |\boldsymbol{y}_j| \Gamma(\boldsymbol{y}_j)\rangle = \sum_{j=1}^{m} b_j \sum_{i=1}^{n} v_i 1_{(\mathbf{0},\boldsymbol{y}_j]}(\boldsymbol{X}_i)$$
$$= \left\langle \boldsymbol{v}, \sum_{j=1}^{m} b_j |\boldsymbol{y}_j| \Gamma(\boldsymbol{y}_j)\right\rangle = \langle \boldsymbol{v}, \mathbf{0}\rangle = 0,$$

and hence, by (3.4), $\sum_{j=1}^{m} b_j|\boldsymbol{y}_j| = 0$, or, since $\boldsymbol{y}_j = \boldsymbol{W}_{i_j} \in A^{\#}$ for some i_j,

$$\sum_{j=1}^{m} b_j |\boldsymbol{W}_{i_j}| = 0$$

with all $b_j \in \mathbb{Z}$. But this equation has at most countably many solutions $\{|W_{i_j}|, j = 1, \ldots, m\}$, and hence occurs with P_0^n-probability 0. That is, for any fixed vector $\boldsymbol{b} = (b_j)_{j=1}^k$ with all $b_j \in \mathbb{Z}$, the function $f_{\boldsymbol{b}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \sum_{j=1}^{k} b_j|\boldsymbol{W}_{i_j}|$ has at most a finite number of zeros, so $P_0^n(f_{\boldsymbol{b}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = 0) = 0$, and since $\mathbb{Z}$ is countable $P_0^n(\cup_{\boldsymbol{b}\in\mathbb{Z}^k}\{f_{\boldsymbol{b}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = 0\}) = 0$. Thus $P_0^n(\cap_{\boldsymbol{b}\in\mathbb{Z}^k}\{f_{\boldsymbol{b}}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) \neq 0\}) = 1$. Hence it follows that the linear dependence condition only holds on an event with probability 0.

Thus the vectors $|\boldsymbol{y}_j|\Gamma(\boldsymbol{y}_j)$, $j = 1, \ldots, m$ are linearly independent almost surely P_0^n, and hence $m \le n$ (P_0^n-almost surely). □

Lemma 3.4. *The discrete mixing measure $\widehat{G}_n$ which defines an MLE is P_0^n-almost surely unique.*

Proof. Suppose that there exist two different MLE's $\widehat{f}_n^1$ and $\widehat{f}_n^2$. then

$$\widehat{f}_n^l(x) = \sum_{j=1}^{m} \pi_j^l \frac{1}{|\boldsymbol{y}_j|} 1_{(\mathbf{0},\boldsymbol{y}_j]}(\boldsymbol{x}), \quad l = 1, 2,$$

where $\pi_j^l \ge 0$ and $\sum_{j=1}^{m} \pi_j^l = 1$ for $l = 1, 2$. Therefore

$$\delta_n(\boldsymbol{x}) \equiv \widehat{f}_n^1(\boldsymbol{x}) - \widehat{f}_n^2(\boldsymbol{x}) = \sum_{j=1}^{m} r_j \frac{1}{|\boldsymbol{y}_j|} 1_{(\mathbf{0},\boldsymbol{y}_j]}(\boldsymbol{x})$$

where $r_j \equiv \pi_j^1 - \pi_j^2$ has at least n zeros (since we know that

$$(\widehat{f}_n^1(\boldsymbol{X}_i))_{i=1}^n = (\widehat{f}_n^2(\boldsymbol{X}_i))_{i=1}^n = (\widehat{f}_n(\boldsymbol{X}_i))_{i=1}^n$$

is unique). So, uniqueness holds if the vectors

$$(1_{(\mathbf{0},\boldsymbol{y}_j]}(\boldsymbol{X}_{\boldsymbol{i}}))_{i=1}^n \in \{0, 1\}^n, \quad \text{for } j = 1, \ldots, m \le n$$

are (almost surely) linearly independent. But this follows from the proof of Lemma 3.3. □

Theorem 3.1 does not assert that the MLE is always unique. An MLE is P_0^n almost surely unique, but we now present an example in which there exist an infinite number of MLE's.

Example 3.1 (*A MLE in $\mathcal{F}_{SMU}$ is Not Always Unique*)**.** To be able to graphically illustrate the set Γ, in the proof of Theorem 3.1, we need to restrict consideration to $n = 2$ and in order that we be able to graphically illustrate the MLE(s) we need to restrict consideration to $d = 2$. Suppose that $\boldsymbol{X}_1 = (1, 3)$ and $\boldsymbol{X}_2 = (3, 2)$ are the observation points. The set

$$\Gamma \equiv \left\{ \frac{1}{u_1 u_2} \left(\mathbb{1}_{(\mathbf{0},\boldsymbol{u}]}(\boldsymbol{X}_1), \mathbb{1}_{(\mathbf{0},\boldsymbol{u}]}(\boldsymbol{X}_2)\right) \;\middle|\; \boldsymbol{u} = (u_1, u_2) \in (0, \infty)^2 \right\}$$

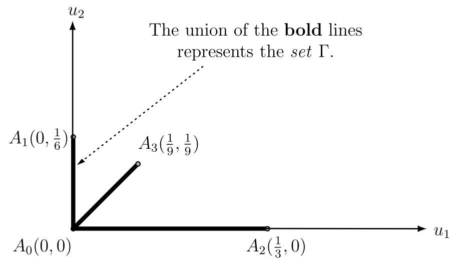
and its convex hull, Conv(Γ), are illustrated in Fig. 1.

Using [35, Theorem 22, p. 118], it follows that any MLE, $\widehat{f}_2$, will have a unique value for $\widehat{\boldsymbol{f}} \equiv (\widehat{f}_2(\boldsymbol{X}_1), \hat{f}_2(\boldsymbol{X}_2))$ that is given by $\widehat{\boldsymbol{f}} = (\tilde{w}_1^{-1}, \tilde{w}_2^{-1})$ where $\tilde{\boldsymbol{w}} = (\tilde{w}_1, \tilde{w}_2)$ maximizes the function $(w_1, w_2) \mapsto \log(w_1 w_2)$ on the set

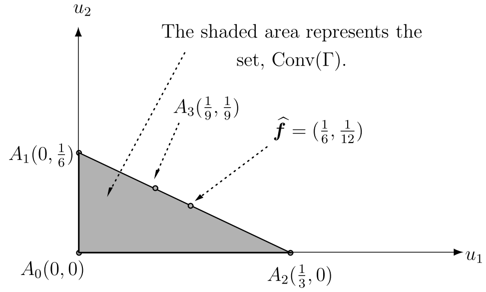
$$\left\{ (w_1, w_2) \in (0, \infty)^2 \;\middle|\; \frac{w_1}{3} \le 2 \text{ and } \frac{w_2}{6} \le 2 \right\}.$$

It is immediate that $\tilde{\boldsymbol{w}} = (6, 12)$ from which we conclude that $\tilde{\boldsymbol{f}} = (1/6, 1/12)$ has exactly two representations as convex combinations in terms of pairs of the points $\{A_1, A_2, A_3\}$ (see Fig. 1(a) again):

$$\left(\frac{1}{6}, \frac{1}{12}\right) = \frac{1}{2}\left(0, \frac{1}{6}\right) + \frac{1}{2}\left(\frac{1}{3}, 0\right), \quad \text{and} \quad \left(\frac{1}{6}, \frac{1}{12}\right) = \frac{1}{4}\left(\frac{1}{3}, 0\right) + \frac{3}{4}\left(\frac{1}{9}, \frac{1}{9}\right).$$



(a) Γ.



(b) Conv (Γ).

Fig. 1. The sets Γ and Conv(Γ) based on two observations: $\boldsymbol{X}_1 = (1, 3)$ and $\boldsymbol{X}_2 = (3, 2)$.

These two convex combinations yield two different maximum likelihood estimators, as shown in Fig. 2(a) and (b).

It should be noted, however, that infinitely many maximum likelihood estimators exist in this case since each convex combination of these two MLEs is again an MLE, by virtue of linearity of f_G (recall (2.1)) as a function of the mixing distribution, G. □

3.2. Strong pointwise consistency of the MLE

Let $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n, \ldots$ be the coordinate random elements on the (completed) infinite product space $(\Omega^\infty, \mathcal{A}^\infty, P^\infty)$ such that these coordinates are i.i.d. according to $f_0 \equiv f_{G_0}$ on $(0, \infty)^d$. Let $A \in \mathcal{A}^\infty$ be the event (with P^∞-probability one) that for each $n \in \mathbb{N}$ there exists a unique SMU density, $\widehat{f}_n \equiv f_{\hat{G}_n}$, maximizing the log-likelihood.

From Theorem 2.3 we have that for each $n \in \mathbb{N}$ and a fixed $\omega \in A$, there exists a unique Borel probability measure, $\widehat{G}_n$ on $((0, \infty)^d, \|\cdot\|_2)$, such that

$$\widehat{f}_n(\boldsymbol{x}) = \int_{(0,\infty)^d} \frac{1}{|\boldsymbol{u}|} \mathbb{1}_{(\boldsymbol{0},\boldsymbol{u}]}(\boldsymbol{x})\, d\widehat{G}_n(\boldsymbol{u}) = \int_{\boldsymbol{u}\geq\boldsymbol{x}} \frac{1}{|\boldsymbol{u}|}\, d\widehat{G}_n(\boldsymbol{u}) \tag{3.5}$$

holds true for *all* $\boldsymbol{x} \in (0, \infty)^d$. We are ready to formulate and prove the following proposition.

Proposition 3.5 (*Strong Consistency of the MLE in $\mathcal{F}_{SMU}$*).

(a) (i) *The sequence of maximum likelihood mixing distributions $\{\widehat{G}_n\}_{n=1}^\infty$ converges weakly to G_0 as $n \to \infty$, P^∞-almost surely.*

(ii) *In addition, for Lebesgue almost all $\boldsymbol{x} \in (0, \infty)^d$, $\widehat{f}_n(\boldsymbol{x}) \to_{a.s.} f_0(\boldsymbol{x})$ as $n \to \infty$. In particular, if f_0 is continuous at $\boldsymbol{x} \in (0, \infty)^d$, then*

$$\left|\widehat{f}_n(\boldsymbol{x}) - f_0(\boldsymbol{x})\right| \to_{a.s.} 0 \quad \text{as } n \to \infty.$$

(b) *The sequence of maximum likelihood estimators, $\{\widehat{f}_n\}_{n=1}^\infty$, is strongly consistent in the total variation (or L_1) and in the Hellinger metrics. That is,*

$$\int_{(0,\infty)^d} \left|\hat{f}_n(\boldsymbol{x}) - f_0(\boldsymbol{x})\right| d\boldsymbol{x} \to_{a.s.} 0 \quad \text{as } n \to \infty,$$

and, with $h^2(p, q) = (1/2)\int\{\sqrt{p(\boldsymbol{x})} - \sqrt{q(\boldsymbol{x})}\}^2 d\boldsymbol{x}$,

$$h\left(\widehat{f}_n, f_0\right) \to_{a.s.} 0 \quad \text{as } n \to \infty.$$

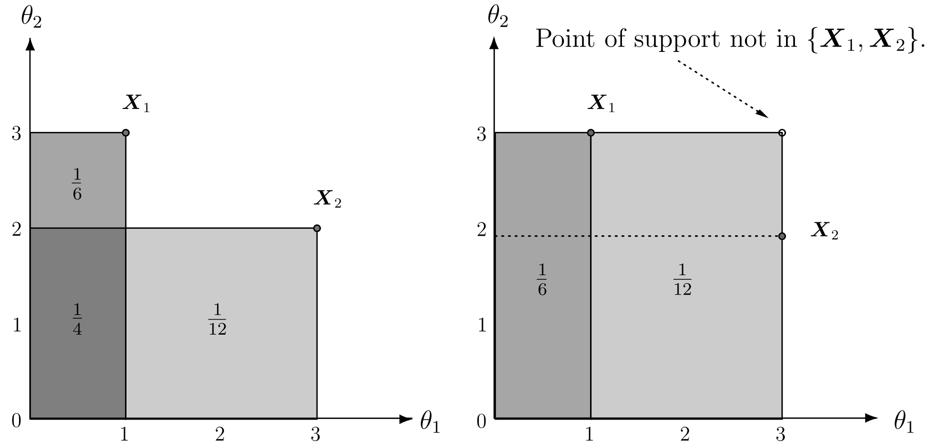
Proof. (a) (i) To be able to apply Theorems 3.4, 3.5 and 3.7 of [39], with the refinement on page 143 of the same article, we need to provide the relevant setup as well as establish the assumptions of Pfanzagl's theorems. We do this below.

Let $\mathcal{C}_0\left((0, \infty)^d, \|\cdot\|_2\right)$ denote the set of all real-valued, continuous functions on $(0, \infty)^d$ that vanish at ∞. Let Θ_* denote the set of all Borel sub-probability measures on $(0, \infty)^d$, equipped with the vague topology, τ, which makes the space a compact, metrizable, topological space, and thus with a countable base. It is also a convex subset of the linear space of all finite, signed, Borel measures on $((0, \infty)^d, \|\cdot\|_2)$. For clarity, the vague topology is the smallest topology that makes the functions

$$\mu \mapsto \int_{(0,\infty)^d} g(\boldsymbol{x})\, d\mu(\boldsymbol{x})$$

continuous, for each $g \in \mathcal{C}_0\left((0, \infty)^d, \|\cdot\|_2\right)$. By metrizability, the topology τ is completely characterized by convergent sequences, $\theta_n \overset{v}{\Rightarrow} \theta$ as $n \to \infty$, on (Θ_*, τ).

Let also $\Theta \subseteq \Theta_*$ be the set of all Borel probability measures on $(0, \infty)^d$, and notice that $\mu \in \Theta$. Also, for each $\theta_* \in \Theta_*$ there exists a unique $c \in [0, 1]$ and a unique $\theta \in \Theta$, such that $\theta_* = c\theta$. Further, notice that letting $m(\nu, \cdot) \equiv f_\nu(\cdot)$, for each



(a) Example 3.1: MLE 1. (b) Example 3.1: MLE 2.

Fig. 2. Two maximum likelihood estimators in $\mathcal{F}_{\mathrm{SMU}}(2)$, supported on the grid generated by the data: $\boldsymbol{X}_1 = (1, 3)$ and $\boldsymbol{X}_2 = (3, 2)$. The two figures show the *contour/level plots* of the respective maximum likelihood densities.

$\nu \in \Theta_*$, and $M_n(\cdot) \equiv \mathbb{P}_n \log\{m(\cdot, \boldsymbol{X})\}$, we have

$$M_n(\theta_*) = \log\{c\} + M_n(\theta) \le M_n(\theta), \quad \text{since } c \in [0, 1],$$

whence, $\sup_{\theta\in\Theta_*} (M_n(\theta)) = \sup_{\theta\in\Theta} (M_n(\theta))$.

With reference measure the Lebesgue measure $\lambda \equiv Q$ and for each $\nu \in \Theta_*$, let $P_\nu \in \Theta_*$ be the sub-probability, Borel measure on $((0, \infty)^d, \|\cdot\|_2)$ with Radon–Nikodym derivative with respect to λ being f_ν, Lebesgue almost surely. Then by virtue of Fubini–Tonelli, $P_\nu \in \Theta$ when and only when $\nu \in \Theta$. Also, notice that for each fixed $\boldsymbol{x} \in (0, \infty)^d$, the functional $\nu \mapsto f_\nu(\boldsymbol{x})$ is not vaguely continuous at any $\nu \in \Theta_*$ with a discontinuity point on the boundary of $[\boldsymbol{x}, \infty)$. However, since for a fixed $\boldsymbol{x} \in (0, \infty)^d$, the function $\boldsymbol{y} \mapsto \mathbb{1}_{[\boldsymbol{x},\infty)}(\boldsymbol{y})/|\boldsymbol{y}|$ is easily seen to be an upper semi-continuous function on $(0, \infty)^d$—vanishing at ∞, Doob [15, Theorem 10, p. 138], applies and asserts that the function $\nu \mapsto f_\nu(\boldsymbol{x})$ on (Θ_*, τ) is itself (vaguely) upper semi-continuous. Since this holds for all $\boldsymbol{x} \in (0, \infty)^d$, it holds almost-surely. Also, the mapping $\nu \mapsto f_\nu(\boldsymbol{x})$ is affine on Θ_* (and hence concave also).

It remains to establish that for each fixed τ-open subset U of Θ_*, the real-valued function $T_U(\cdot)$ on $(0, \infty)^d$ defined by

$$T_U(\boldsymbol{x}) = \sup_{\nu\in U} \left\{ \int_{(0,\infty)^d} \frac{1}{|\boldsymbol{u}|} \mathbf{1}_{(\mathbf{0},\boldsymbol{u}]}(\boldsymbol{x})\, \mathrm{d}\nu(\boldsymbol{u}) \right\}$$

is a $\mathcal{A}$-measurable function. We can choose to take $\mathcal{A}$ to be the Lebesgue σ-field, in which case measurability follows by observing that $T_U(\cdot)$ is a block-decreasing function and appeal to Proposition 2.4.

We now apply Theorem 3.4 of [39] to our setting and further appeal to the fact that a vaguely convergent sequence of probability measures with limit a probability measure, is, in fact, weakly convergent. This gives the desired conclusion: the random sequence of maximum likelihood mixing probability measures $\{\hat{G}_n\}_{n=1}^{\infty}$ converges weakly to G_0 as $n \to \infty$, P^∞-almost surely.

(ii) Combining the fact that, for each fixed $\boldsymbol{x} \in (0, \infty)^d$, $\nu \mapsto f_\nu(\boldsymbol{x})$ is vaguely upper semi-continuous on Θ_* with the conclusion of part (a)(i), we get

$$\overline{\lim_{n\to\infty}} \left\{ f_{\widehat{G}_n}(\boldsymbol{x}) \right\} \le f_0(\boldsymbol{x});\ P^\infty\text{-a.s.} \quad \text{for all } \boldsymbol{x} \in (0, \infty)^d. \tag{3.6}$$

Let

$$F_{G_0}(\cdot) = \int_{(0,\infty)^d} \frac{|\cdot \wedge \boldsymbol{u}|}{|\boldsymbol{u}|}\, dG_0(\boldsymbol{u})$$

and

$$F_{\widehat{G}_n}(\cdot) = \int_{(0,\infty)^d} \frac{|\cdot \wedge \boldsymbol{u}|}{|\boldsymbol{u}|}\, d\widehat{G}_n(\boldsymbol{u})$$

be the distribution functions corresponding to the densities $f_0(\cdot)$ and $\widehat{f}_n(\cdot)$, respectively, $n \in \mathbb{N}$. These distribution functions are everywhere continuous on the Euclidean set $(0, \infty)^d$. In fact, since for each fixed $\boldsymbol{x} \in (0, \infty)^d$, the function $\boldsymbol{u} \mapsto |\boldsymbol{x} \wedge \boldsymbol{u}| / |\boldsymbol{u}|$ is bounded (by 1) and continuous on $(0, \infty)^d$,we then have that

$$F_{\widehat{G}_n}(\boldsymbol{x}) \to_{a.s.} F_{G_0}(\boldsymbol{x}) \quad \text{for all } \boldsymbol{x} \in (0, \infty)^d \tag{3.7}$$

follows directly by the definition of almost sure weak convergence of the mixing random measures $\{\widehat{G}_n\}_{n=1}^{\infty}$ to G_0, established in part (a)(i).

Let B be the set of points on $(0, \infty)^d$ at which f_0 is continuous. Then B^c has Lebesgue measure zero, $\lambda(B^c) = 0$, exactly because f_0 is discontinuous on the boundary $\partial[\boldsymbol{x}_0, \infty)$ for a (possibly non-existent) $\boldsymbol{x}_0 \in (0, \infty)^d$ where P_0 is discontinuous (i.e., such that $P_0(\{\boldsymbol{x}_0\}) > 0$). Since P_0 can have at most countably many discontinuity points $\boldsymbol{x}_0 \in (0, \infty)^d$ and since $\lambda(\partial[\boldsymbol{x}_0, \infty)) = 0$, we get by countable subadditivity of λ that indeed $\lambda(B^c) = 0$.

Fix arbitrary $\boldsymbol{x} \in B$ and $\epsilon > 0$. Then, since f_0 is lower semi-continuous at $\boldsymbol{x}$, there exists an open neighborhood $U_{\boldsymbol{x},\epsilon}$ of $\boldsymbol{x}$ such that for every $\boldsymbol{y} \in U_{\boldsymbol{x},\epsilon}$ we have that $f_0(\boldsymbol{y}) > f_0(\boldsymbol{x}) - \epsilon$. In particular, there exists an $U_{\boldsymbol{x},\epsilon} \ni \boldsymbol{x}_\epsilon > \boldsymbol{x}$ satisfying $f_0(\boldsymbol{x}_\epsilon) > f_0(\boldsymbol{x}) - \epsilon$. Since f_0 is block-decreasing, we have:

$$\frac{V_{F_{G_0}}(\boldsymbol{x}, \boldsymbol{x}_\epsilon]}{\lambda\left((\boldsymbol{x}, \boldsymbol{x}_\epsilon]\right)} = \frac{\int_{(\boldsymbol{x},\boldsymbol{x}_\epsilon]} \{f_0(\boldsymbol{y})\}\, d\boldsymbol{y}}{\lambda\left((\boldsymbol{x}, \boldsymbol{x}_\epsilon]\right)} \geq f_0(\boldsymbol{x}_\epsilon) > f_0(\boldsymbol{x}) - \epsilon. \tag{3.8}$$

Further, for each fixed $n \in \mathbb{N}$, since $\widehat{f}_n(\cdot)$ is block-decreasing (as a SMU density), we have

$$f_{\widehat{G}_n}(\boldsymbol{x}) \geq \frac{\int_{(\boldsymbol{x},\boldsymbol{x}_\epsilon]} \left\{f_{\widehat{G}_n}(\boldsymbol{y})\right\}\, d\boldsymbol{y}}{\lambda\left((\boldsymbol{x}, \boldsymbol{x}_\epsilon]\right)} \tag{3.9}$$

$$= \frac{V_{F_{\widehat{G}_n}}(\boldsymbol{x}, \boldsymbol{x}_\epsilon]}{\lambda\left((\boldsymbol{x}, \boldsymbol{x}_\epsilon]\right)}. \tag{3.10}$$

Eq. (3.7) further implies that

$$V_{F_{\widehat{G}_n}}(\boldsymbol{x}, \boldsymbol{x}_\epsilon] \to V_{F_{G_0}}(\boldsymbol{x}, \boldsymbol{x}_\epsilon], \quad \text{as } n \to \infty. \tag{3.11}$$

Combining Eqs. (3.8)–(3.11) and the fact that $\epsilon > 0$ was arbitrary, we get

$$\varliminf_{n\to\infty} \left\{f_{\widehat{G}_n}(\boldsymbol{x})\right\} \geq f_0(\boldsymbol{x});\ P^{\infty}\text{-a.s.} \qquad \text{for } \boldsymbol{x} \in B. \tag{3.12}$$

Eqs. (3.6) and (3.12) yield the assertion: for Lebesgue almost all $\boldsymbol{x} \in (0, \infty)^d$ (and, in particular, at the points of continuity of f), $f_{\widehat{G}_n}(\boldsymbol{x}) \to_{a.s.} f_0(\boldsymbol{x})$ as $n \to \infty$ holds.

(b) Showing consistency in the L_1 (total-variation) norm is a direct consequence of part (a) (ii) and Glick's Theorem, [17]; see also [14, p. 25].

Convergence in the Hellinger metric follows from the following well-known inequalities of [33, p.46]:

$$h^2(P, Q) \leq \frac{1}{2}\|P - Q\|_{L_1} \leq h(P, Q)\left\{2 - h^2(P, Q)\right\}^{\frac{1}{2}},$$

where $h^2(P, Q) = 2^{-1}\int\left(\sqrt{dP} - \sqrt{dQ}\right)^2$ is the squared Hellinger metric and $\|\cdot\|_{L_1}$ is the L_1-norm. □

4. A local asymptotic minimax lower bound

Let $\boldsymbol{X}_i := (X_{i,1}, \ldots, X_{i,d})'$ for $i = 1, 2, \ldots, n$ be i.i.d. random vectors from density $f \in \mathcal{F}_{\mathrm{SMU}}(d)$. For a fixed $\boldsymbol{x}_0 \equiv (x_{0,1}, \ldots, x_{0,d})' \in (0, \infty)^d$, we want to estimate the functional $T(f) := f(\boldsymbol{x}_0)$ on the basis of $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$. We shall make the following assumption:

Assumption 4.1. Suppose that $f \in \mathcal{F}_{\mathrm{SMU}}$ is continuously differentiable at $\boldsymbol{x}_0$, $f(\boldsymbol{x}_0) > 0$, and, in particular, there exists an open ball $A(\boldsymbol{x}_0)$ around $\boldsymbol{x}_0$ such that f is everywhere strictly positive on $A(\boldsymbol{x}_0)$ and where $(\partial/\partial x_j) f(\boldsymbol{x}_0) < 0$ exist for all $j \in \{1, 2, \ldots, d\}$ and are continuous on $A(\boldsymbol{x}_0) \subseteq (0, \infty)^d$. Further, we assume that the full mixed derivative of f exists, is continuous on $A(\boldsymbol{x}_0)$, and satisfies

$$(-1)^d \left.\frac{\partial^d f}{\partial x_1 \cdots \partial x_d}(\boldsymbol{x})\right|_{\boldsymbol{x}=\boldsymbol{y}} > 0 \quad \text{for all } \boldsymbol{y} \in A(\boldsymbol{x}_0).$$

Proposition 4.1. *Suppose that $f \in \mathcal{F}_{\mathrm{SMU}}$ satisfies Assumption 4.1 at the fixed point $\boldsymbol{x}_0 \in (0, \infty)^d$. Then there is a sequence $\{f_n\} \subset \mathcal{F}_{\mathrm{SMU}}$ such that any estimator sequence $\{T_n\}$ of $f(x_0)$ satisfies*

$$\varliminf_{n\to\infty} \max\left\{\boldsymbol{E}_{f_n}\left\{n^{\frac{1}{3}}\,|T_n - f_n(\boldsymbol{x}_0)|\right\}, \boldsymbol{E}_f\left\{n^{\frac{1}{3}}\,|T_n - f(\boldsymbol{x}_0)|\right\}\right\}$$

$$\geq \frac{e^{-\frac{1}{3}}}{2^d}\left\{3^{d-1}\right\}^{\frac{1}{3}}\left\{(-1)^d \left.\frac{\partial^d f(\boldsymbol{x})}{\partial x_1 \cdots \partial x_d}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} \cdot f(\boldsymbol{x}_0)\right\}^{\frac{1}{3}}. \tag{4.1}$$

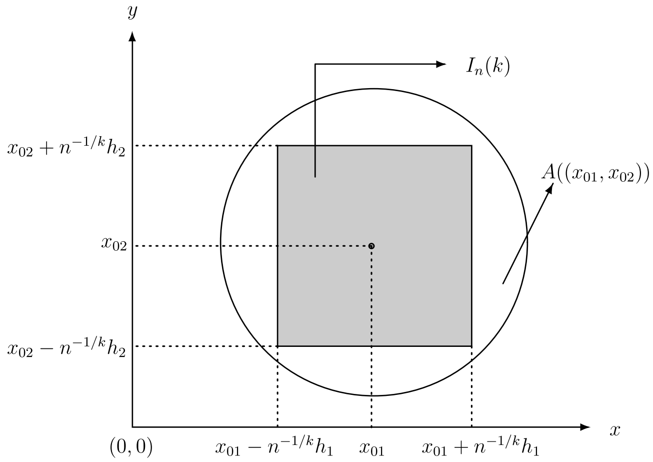


Fig. 3. Perturbation rectangle $I_n(k)$, for the case $d = 2$, with center $\boldsymbol{x}_0 = (x_{01}, x_{02})$ and $\boldsymbol{h} = (h_1, h_2)$.

Remark. The lower bound in Proposition 4.1 should be contrasted to a similar lower bound for estimation of $f(\boldsymbol{x}_0)$ for $f \in \mathcal{F}_{\mathrm{BDD}}$ which is derived by Pavlides [38]. In that case the natural hypothesis is $\partial f(\boldsymbol{x}_0)/\partial x_i < 0$ for $i = 1, \ldots, d$, and the resulting rate of convergence is $n^{1/(d+2)}$.

To prove Proposition 4.1 we will make use of the following lemma. It was established in the form presented here by [23]; see also Groeneboom and Jongbloed [22,29].

Lemma 4.2. *Let $\mathcal{F}$ be a class of densities on a measurable space $(\mathcal{X}, \mathcal{A})$ and f a fixed element of $\mathcal{F}$. Let $\mathcal{F}_f$ denote any open Hellinger ball with center $f \in \mathcal{F}$. Assume that there exists a sequence $\{f_n\}_{n=1}^{\infty} \subseteq \mathcal{F}$ such that*

$$\lim_{n\to\infty} \left\{\sqrt{n}h(f_n, f)\right\} = \alpha \tag{4.2}$$

and

$$\lim_{n\to\infty} |T(f_n) - T(f)| = \beta \tag{4.3}$$

both hold for some constants $0 < \alpha, \beta < \infty$, and where T is a functional on $\mathcal{F}$. Here, $h^2(f_n, f) \equiv 2^{-1}\int\{\sqrt{f_n(x)} - \sqrt{f(x)}\}^2\, d\mu(x)$, is the Hellinger distance between the μ-densities f_n and f. Let $l(\cdot)$ be a convex function, symmetric about zero, which is non-decreasing on $[0, \infty)$.

Then, it holds that

$$\varliminf_{n\to\infty} \left\{R_{n,l}(\mathcal{F}_f)\right\} \geq l\left(\frac{1}{4}\beta e^{-2\alpha^2}\right) \tag{4.4}$$

where $R_{n,l}(\mathcal{F}) \equiv \inf_{T_n} \sup_{g\in\mathcal{F}} \boldsymbol{E}_{g^{\otimes n}}\{l(T_n - T(g))\}$ is the minimax risk for estimating the functional $T(f)$ based on n i.i.d observations from $\mathcal{F}$.

In particular, for the loss $l(x) = |x|$ on we have

$$\varliminf_{n\to\infty} \left\{R_{n,|\cdot|}(\mathcal{F}_f)\right\} \geq \frac{1}{4}\beta e^{-2\alpha^2}. \tag{4.5}$$

Hereafter, fix an otherwise arbitrary vector $\boldsymbol{h} := (h_1, \ldots, h_d) \in (0, \infty)^d$, and define $\boldsymbol{H} := \mathrm{diag}(\boldsymbol{h}) \in M_{d\times d}\left((0, \infty)\right)$. For each $k \in \mathbb{N}$, consider the perturbation rectangle

$$I_n(k) := \bigotimes_{i=1}^{d} \left[x_{0,i} - n^{-\frac{1}{k}}h_i, x_{0,i} + n^{-\frac{1}{k}}h_i\right],$$

only for those positive integers $n \geq n_0(k, \boldsymbol{x}_0, \boldsymbol{h})$ for which $I_n(k) \subseteq A(\boldsymbol{x}_0)$ for all $n \geq n_0$. The two-dimensional case, $d = 2$, is illustrated in Fig. 3.

Recall Assumption 4.1. Let $b := (\partial^d/\partial x_1 \cdots \partial x_d)f(\boldsymbol{x})\big|_{\boldsymbol{x}=\boldsymbol{x}_0}$ and observe that $(-1)^d b > 0$. Finally, define the functions h_n on $I_n(3d)$ as follows:

$$h_n(y_1, \ldots, y_d) := (-1)^d \prod_{i=1}^{d} \left\{\mathbb{1}_{\left(x_{0,i}, x_{0,i}+n^{-\frac{1}{3d}}h_i\right]}(y_i) - \mathbb{1}_{\left[x_{0,i}-n^{-\frac{1}{3d}}h_i, x_{0,i}\right]}(y_i)\right\},$$

and

$$g_n(\boldsymbol{y}) := b \int_{\boldsymbol{u} \succeq \boldsymbol{y}} \left\{ \mathbb{1}_{I_n(3d)}(\boldsymbol{u}) \cdot h_n(\boldsymbol{u}) \right\} d\boldsymbol{u},$$

where we observe that $g_n(\boldsymbol{y}) \geq 0$ for all $\boldsymbol{y} \in I_n(3d)$, since $\boldsymbol{x}_0$ is the center of the rectangle $I_n(3d)$. In fact, consideration of the geometry of the definition of $g_n(\cdot)$ reveals that, for $\boldsymbol{y} \in I_n$, $g_n(\boldsymbol{y})$ is equal to $(-1)^d b > 0$ times the volume of the rectangle $[\boldsymbol{v}_n(\boldsymbol{y}) \wedge \boldsymbol{y}, \boldsymbol{v}_n(\boldsymbol{y}) \vee \boldsymbol{y}]$, where $\boldsymbol{v}_n(\boldsymbol{y})$ is defined as that vertex of I_n that is closest in L_2-distance from $\boldsymbol{y} \in I_n$. Since I_n is a decreasing sequence of compact sets, it is then immediately clear that $g_n(\boldsymbol{y})$ is (pointwise) non-increasing in $n \in \mathbb{N}$, for each fixed $\boldsymbol{y} \in (0, \infty)^d$.

Assume that $f \in \mathcal{F}_{\mathrm{SMU}}$, and for fixed vectors $\boldsymbol{x}_0, \boldsymbol{h} \in (0, \infty)^d$ we further assume that f satisfies Assumption 4.1. For $n \geq n_0(3d, \boldsymbol{x}_0, \boldsymbol{h})$, define the perturbed density, f_n of f at $\boldsymbol{x}_0$, by

$$f_n(\boldsymbol{x}) = \begin{cases} \dfrac{f(\boldsymbol{x}) + \theta g_n(\boldsymbol{x})}{d_n} : & \text{if } \boldsymbol{x} \in I_n(3d) \\ \dfrac{f(\boldsymbol{x})}{d_n} : & \text{if } \boldsymbol{x} \in I_n^c(3d) \end{cases} \tag{4.6}$$

for some arbitrary but fixed $\theta \in (0, 1)$ and where d_n is the normalizing constant for f_n, uniquely determined by $\int_{(0,\infty)^d} f_n(\boldsymbol{x})\, d\boldsymbol{x} = 1$. We will see the importance of the value of b and the fact that $0 < \theta < 1$ in the following proposition that establishes that $\{f_n\}_{n \geq n_1} \subseteq \mathcal{F}_{\mathrm{SMU}}(d)$ for a sufficiently large $n_1 \in \mathbb{N}$.

Proposition 4.3. *There exists a positive integer $n_1 := n_1(d, \boldsymbol{x}_0, \boldsymbol{h}) \geq n_0(3d, \boldsymbol{x}_0, \boldsymbol{h})$ such that $f_n \in \mathcal{F}_{SMU}$ for all $n \geq n_1$.*

Proof. Since $f \in \mathcal{F}_{\mathrm{SMU}}(d)$, we get from Theorem 2.3 that

$$V_f[\boldsymbol{x}, \boldsymbol{y}] \geq 0, \quad \text{for all } d\text{-boxes } [\boldsymbol{x}, \boldsymbol{y}]. \tag{4.7}$$

From the definition of $g_n(\cdot)$, we see that its full, mixed partial derivative exists in a neighborhood of $\boldsymbol{x}_0$. Hence, by definition and the fact that $(-1)^d b > 0$ and $\theta \in (0, 1)$, we have that

$$\begin{aligned} (-1)^d \frac{\partial^d f_n}{\partial x_1 \cdots \partial x_d}(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\boldsymbol{y}} &\geq (-1)^d \frac{\partial^d f}{\partial x_1 \cdots \partial x_d}(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\boldsymbol{y}} - (-1)^d b\theta \\ &= \left[(-1)^d \frac{\partial^d f}{\partial x_1 \cdots \partial x_d}(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\boldsymbol{y}} - (-1)^d b \right] + (1-\theta)(-1)^d b \\ &\geq 2^{-1}(1-\theta)(-1)^d b > 0, \end{aligned} \tag{4.8}$$

where the second to last inequality follows from Assumption 4.1 that the full mixed partial derivative of f exists and is continuous at $\boldsymbol{x}_0$ from which we get, by definition of continuity, that there exists a large enough positive integer $n_1 := n_1(d, \boldsymbol{x}_0, \boldsymbol{h}) \geq n_0(3d, \boldsymbol{x}_0, \boldsymbol{h})$ such that

$$(-1)^d \frac{\partial^d f}{\partial x_1 \cdots \partial x_d}(\boldsymbol{x})\bigg|_{\boldsymbol{x}=\boldsymbol{y}} - (-1)^d b \geq -2^{-1}(1-\theta)(-1)^d b$$

holds true for all $\boldsymbol{y} \in I_n(3d)$ and $n \geq n_1$. The result in (4.8) suggests that

$$(-1)^d V_{f_n}[\boldsymbol{x}, \boldsymbol{y}] \equiv (-1)^d \int_{(\boldsymbol{x}, \boldsymbol{y}]} \left\{ \frac{\partial^d f_n}{\partial w_1 \cdots \partial w_n}(\boldsymbol{w})\bigg|_{\boldsymbol{w}=\boldsymbol{u}} \right\} d\boldsymbol{u} \geq 0$$

holds true for all d-boxes $(\boldsymbol{x}, \boldsymbol{y}]$ with $\boldsymbol{x}, \boldsymbol{y} \in I_n(3d)$ and $n \geq n_1$.

The last case not considered is the one that is exactly one between $\boldsymbol{x}$ and $\boldsymbol{y}$, in the d-box $[\boldsymbol{x}, \boldsymbol{y}]$, is an element of $I_n(3d)$. See also Fig. 4. For this case, we can appeal to Lemma 2.2 by setting $[\boldsymbol{x}_0, \boldsymbol{y}_0] := [\boldsymbol{x}, \boldsymbol{y}] \cap I_n(3d)$—the latter being well-defined as the intersection of two rectangles is itself an rectangle. Then, from Lemma 2.2 and (4.7), we have,

$$(-1)^d V_{f_n}[\boldsymbol{x}, \boldsymbol{y}] = (-1)^d V_{f_n}[\boldsymbol{x}_0, \boldsymbol{y}_0] + (-1)^d \sum_{i=1}^{m} \left\{ V_{f_n}[\boldsymbol{x}_i, \boldsymbol{y}_i] \right\} \geq 0 + 0 = 0,$$

exactly since $[\boldsymbol{x}_i, \boldsymbol{y}_i] \subseteq I_n^c(3d)$ for all $i \in \{1, 2, \ldots, m\}$ (where m is as defined in Lemma 2.2). For completeness, notice that we were not concerned above with end-point discontinuities of f (or f_n) on the entailed rectangle, subsets of $I_n(3d)$, as, in fact, f (and f_n) is (are) continuous there for $n \geq n_1$, by Assumption 4.1.

All these observations finally yield that $(-1)^d V_{f_n}[\boldsymbol{x}, \boldsymbol{y}] \geq 0$ holds true for all d-boxes $[\boldsymbol{x}, \boldsymbol{y}]$ and thus Theorem 2.3 asserts that $f_n \in \mathcal{F}_{\mathrm{SMU}}$ for all $n \geq n_1$. □

We are ready to prove the main proposition of this section.

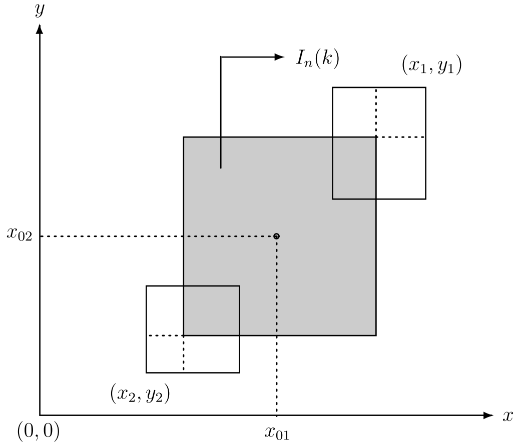


Fig. 4. Perturbation rectangle $I_n(k)$, for the case $d = 2$, with two rectangles intersecting $I_n(k)$ but otherwise not subsets of it.

Proof. Recall Proposition 4.3. First, we establish that

$$\int_{I_n} g_n(\boldsymbol{x})\, d\boldsymbol{x} = (-1)^d b \prod_{i=1}^{d} \left\{h_i^2\right\} \cdot n^{-\frac{2}{3}}, \tag{4.9}$$

where, hereafter, I_n will be the short-hand form for $I_n(3d)$. By definition, notice that,

$$\begin{aligned}
\frac{1}{b}\int_{I_n} g_n(\boldsymbol{x})\, d\boldsymbol{x} &= \int_{I_n}\int_{I_n}\prod_{i=1}^{d}\left\{\mathbb{1}_{[x_i \le u_i]}\right\} h_n(\boldsymbol{u})\, d\boldsymbol{u} d\boldsymbol{x} \\
&= \int_{I_n} h_n(\boldsymbol{u}) \left\{\int_{I_n} \mathbb{1}_{(\boldsymbol{0},\boldsymbol{u}]}(\boldsymbol{x})\, d\boldsymbol{x}\right\} d\boldsymbol{u} \\
&= \int_{I_n}\prod_{i=1}^{d}\left\{u_i - \left(x_{0i} - h_i n^{-\frac{1}{3d}}\right)\right\} h_n(\boldsymbol{u})\, d\boldsymbol{u} \\
&= \prod_{i=i}^{d}\left\{\int_{x_{0i}-h_i n^{-\frac{1}{3d}}}^{x_{0i}+h_i n^{-\frac{1}{3d}}} \left([u_i - (x_{0i} - h_i n^{-\frac{1}{3d}})]\right.\right. \\
&\quad \left.\left. \times \left[\mathbb{1}_{[x_{0i}-h_i n^{-\frac{1}{3d}}, x_{0i}]}(u_i) - \mathbb{1}_{\left(x_{0i}, x_{0i}+h_i n^{-\frac{1}{3d}}\right]}(u_i)\right]\right) du_i\right\} \\
&= \prod_{i=1}^{d}\left\{\int_{x_{0i}-h_i n^{-\frac{1}{3d}}}^{x_{0i}} [u_i - (x_{0i} - h_i n^{-\frac{1}{3d}})]\, du_i\right. \\
&\quad \left. - \int_{x_{0i}}^{x_{0i}+h_i n^{-\frac{1}{3d}}} [u_i - (x_{0i} - h_i n^{-\frac{1}{3d}})]\, du_i\right\} \\
&= \prod_{i=1}^{d}\left\{\int_0^{h_i n^{-\frac{1}{3d}}} [-y + h_i n^{-\frac{1}{3d}}]\, dy - \int_0^{h_i n^{-\frac{1}{3d}}} [w + h_i n^{-\frac{1}{3d}}]\, dw\right\} \\
&= \prod_{i=1}^{d}\left\{\int_0^{h_i n^{-\frac{1}{3d}}} (-2y)\; dy\right\} = (-1)^d \prod_{i=1}^{d}\left\{h_i^2 n^{-\frac{2}{3d}}\right\} = (-1)^d \prod_{i=1}^{d}\left\{h_i^2\right\} \cdot n^{-\frac{2}{3}},
\end{aligned}$$

thus yielding (4.9).

We next derive another equality, the most important fact about it being the factor n^{-1} on the right hand side:

$$\int_{I_n} g_n^2(\boldsymbol{x})\, d\boldsymbol{x} = \left(\frac{8}{3}\right)^d b^2 \prod_{i=1}^{d}\left\{h_i^3\right\} \cdot n^{-1}. \tag{4.10}$$

Before we start deriving (4.10), let us first define four rectangles R_j^i with $j = 1, 2, 3, 4$ for each $i \in \{1, 2, \ldots, d\}$:

(i) $R_1^i = \left[x_{0i} - h_i n^{-\frac{1}{3d}}, x_{0i}\right] \times \left[x_{0i} - h_i n^{-\frac{1}{3d}}, x_{0i}\right]$,
(ii) $R_2^i = \left[x_{0i} - h_i n^{-\frac{1}{3d}}, x_{0i}\right] \times \left(x_{0i}, x_{0i} + h_i n^{-\frac{1}{3d}}\right]$,
(iii) $R_3^i = \left(x_{0i}, x_{0i} + h_i n^{-\frac{1}{3d}}\right] \times \left[x_{0i} - h_i n^{-\frac{1}{3d}}, x_{0i}\right]$,
(iv) $R_4^i = \left(x_{0i}, x_{0i} + h_i n^{-\frac{1}{3d}}\right] \times \left(x_{0i}, x_{0i} + h_i n^{-\frac{1}{3d}}\right]$.

Then, by definition:

$$\begin{aligned}
\frac{1}{b^2}\int_{I_n} g_n^2(\boldsymbol{x})\, d\boldsymbol{x} &= \int_{I_n}\left\{\int_{I_n} h_n(\boldsymbol{u})\mathbb{1}_{[\boldsymbol{x}\le\boldsymbol{u}]}\, d\boldsymbol{u}\right\}^2 d\boldsymbol{x} \\
&= \int_{I_n}\int_{I_n}\int_{I_n} h_n(\boldsymbol{u})h_n(\boldsymbol{v})\mathbb{1}_{[\boldsymbol{x}\le\boldsymbol{u}\wedge\boldsymbol{v}]}\, d\boldsymbol{v}d\boldsymbol{u}d\boldsymbol{x} \\
&= \int_{I_n}\int_{I_n}\left\{\prod_{i=1}^d\left[(u_i\wedge v_i) - (x_{0i} - h_i n^{-\frac{1}{3d}})\right] \times h_n(\boldsymbol{u})h_n(\boldsymbol{v})\right\} d\boldsymbol{v}d\boldsymbol{u} \\
&= \prod_{i=1}^d\left\{\int_{R_1^i+R_3^i}\left[(u\wedge v) - (x_{0i} - h_i n^{-\frac{1}{3d}})\right] dvdu\right. \\
&\quad \left. -\ 2\int_{R_2^i}\left[(u\wedge v) - (x_{0i} - h_i n^{-\frac{1}{3d}})\right] dvdu\right\} \\
&= 2^d\prod_{i=1}^d\{S_{1i} + S_{2i} - S_{3i}\},
\end{aligned} \tag{4.11}$$

where the last equality follows by symmetry and Fubini–Tonelli and the integrals in the braces are to be evaluated below:

$$\begin{aligned}
S_{1i} &\equiv \int_{x_{0i}-h_i n^{-\frac{1}{3d}}}^{x_{0i}}\int_v^{x_{0i}}\left\{v - \left(x_{0i} - h_i n^{-\frac{1}{3d}}\right)\right\} dudv \\
&= \int_{x_{0i}-h_i n^{-\frac{1}{3d}}}^{x_{0i}}\left\{(x_{0i} - v)\left(v - x_{0i} + h_i n^{-\frac{1}{3d}}\right)\right\} dv \\
&= \int_{x_{0i}+h_i n^{-\frac{1}{3d}}}^{h_i n^{-\frac{1}{3d}}}\left\{y\left(-y + h_i n^{-\frac{1}{3d}}\right)\right\} dy \quad \text{[change of variable]}
\end{aligned}$$

while, again, by a change of variable argument:

$$\begin{aligned}
S_{2i} &\equiv \int_{x_{0i}}^{x_{0i}+h_i n^{-\frac{1}{3d}}}\int_v^{x_{0i}+h_i n^{-\frac{1}{3d}}}\left\{v - \left(x_{0i} - h_i n^{-\frac{1}{3d}}\right)\right\} dudv \\
&= \int_{x_{0i}}^{x_{0i}+h_i n^{-\frac{1}{3d}}}\left\{\left[(x_{0i} - v) + h_i n^{-\frac{1}{3d}}\right]\left[(v - x_{0i}) + h_i n^{-\frac{1}{3d}}\right]\right\} dv \\
&= \int_0^{h_i n^{-\frac{1}{3d}}}\left\{\left(-y + h_i n^{-\frac{1}{3d}}\right)\left(y + h_i n^{-\frac{1}{3d}}\right)\right\} dy,
\end{aligned}$$

and similarly:

$$\begin{aligned}
S_{3i} &\equiv \int_{x_{0i}-h_i n^{-\frac{1}{3d}}}^{x_{0i}}\left\{h_i n^{-\frac{1}{3d}}\left(v - x_{0i} + h_i n^{-\frac{1}{3d}}\right)\right\} dv \\
&= h_i n^{-\frac{1}{3d}}\int_0^{h_i n^{-\frac{1}{3d}}}\left\{h_i n^{-\frac{1}{3d}} - y\right\} dy.
\end{aligned}$$

Let now $q_i := h_i n^{-1/3d}$, for $i \in \{1, 2, \ldots, d\}$, and observe that

$$S_{1i} + S_{2i} - S_{3i} = \int_0^{q_i}\left\{y(q_i - y) + q_i^2 - y^2 + q_i^2 - q_i y\right\} dy = \cdots = \frac{4}{3}h_i^3 n^{-\frac{1}{d}},$$

so that plugging all these in (4.11) yields the desired (4.10).

Now, recall from the definition of f_n that $\theta \in (0, 1)$ was arbitrary but fixed. Also, from $\int_{(0,\infty)^d} f_n(\boldsymbol{x})\, d\boldsymbol{x} = 1$ we can get an explicit expression for the normalizing constant d_n:

$$\begin{aligned} d_n &= \int_{I_n} f(\boldsymbol{x})\, d\boldsymbol{x} + \int_{I_n^c} f(\boldsymbol{x})\, d\boldsymbol{x} + \theta \int_{I_n} g_n(\boldsymbol{x})\, d\boldsymbol{x} \\ &= 1 + \theta \int_{I_n} g_n(\boldsymbol{x})\, d\boldsymbol{x} = 1 + (-1)^d \theta b \prod_{i=1}^{d} \left\{h_i^2\right\} \cdot n^{-\frac{2}{3}}, \end{aligned} \tag{4.12}$$

where the second to last equality follows from $\int_{(0,\infty)^d} f(\boldsymbol{x})\, d\boldsymbol{x} = 1$, while the last equality follows from (4.9). Notice from (4.12) that $d_n \downarrow 1$ as $n \uparrow \infty$. Also, from the easily verifiable identity $g_n(\boldsymbol{x}_0) = (-1)^d b \prod_{i=1}^{d} \{h_i\}\, n^{-1/3}$, we have

$$\begin{aligned} n^{\frac{1}{3}} |f_n(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)| &= n^{\frac{1}{3}} \left| \frac{f(\boldsymbol{x}_0) + (-1)^d b \prod_{i=1}^{d} \{h_i\}\, n^{-\frac{1}{3}}}{d_n} - f(\boldsymbol{x}_0) \right| \\ &= \left| n^{\frac{1}{3}} \left\{ \frac{1}{d_n} - 1 \right\} f(\boldsymbol{x}_0) + \frac{(-1)^d b\theta \prod_{i=1}^{d} \{h_i\}}{d_n} \right| \\ &\longrightarrow (-1)^d b\theta \prod_{i=1}^{d} \{h_i\} \;\; (> 0), \quad \text{as } n \to \infty. \end{aligned} \tag{4.13}$$

Also,

$$\begin{aligned} 2nh^2(f_n, f) &= n \int_{I_n} \left\{ \sqrt{f_n(\boldsymbol{x})} - \sqrt{f(\boldsymbol{x})} \right\}^2 d\boldsymbol{x} + n \int_{I_n^c} \left\{ \sqrt{f_n(\boldsymbol{x})} - \sqrt{f(\boldsymbol{x})} \right\}^2 d\boldsymbol{x} \\ &= n \int_{I_n} \left\{ \frac{f_n(\boldsymbol{x}) - f(\boldsymbol{x})}{\sqrt{f_n(\boldsymbol{x})} + \sqrt{f(\boldsymbol{x})}} \right\}^2 d\boldsymbol{x} + \delta_n^2 \int_{I_n^c} f(\boldsymbol{x})\, d\boldsymbol{x}, \end{aligned} \tag{4.14}$$

where,

$$\begin{aligned} \delta_n &\equiv \sqrt{n} \left\{ 1 - \frac{1}{\sqrt{d_n}} \right\} = \sqrt{n} \left\{ \frac{\sqrt{d_n} - 1}{\sqrt{d_n}} \right\} \\ &= \frac{\sqrt{n} \left\{ \sqrt{1 + \mathcal{O}\left(n^{-\frac{2}{3}}\right)} - 1 \right\}}{\sqrt{d_n}} \to 0, \quad \text{as } n \to \infty, \end{aligned}$$

with the convergence on the last display following from (4.12). Applying this to (4.14), we have:

$$2nh^2(f_n, f) = n \int_{I_n} \left\{ \frac{f_n(\boldsymbol{x}) - f(\boldsymbol{x})}{\sqrt{f_n(\boldsymbol{x})} + \sqrt{f(\boldsymbol{x})}} \right\}^2 d\boldsymbol{x} + o(1) \tag{4.15}$$

as $n \to \infty$, because $0 \le \int_{I_n^c} f(\boldsymbol{x})\, d\boldsymbol{x} \le 1$.

For fixed $n \in \mathbb{N}$, such that f and g_n be continuous and strictly positive on I_n, let $\boldsymbol{x}_{(n)}$ and $\boldsymbol{x}^{(n)}$ denote, respectively, a minimizer and a maximizer of f on the compact set I_n. Let also $\boldsymbol{y}_{(n)}$ and $\boldsymbol{y}^{(n)}$ denote, respectively, a minimizer and a maximizer of g_n on the compact set I_n. Observe that, since I_n is a decreasing sequence of compact sets converging to $\{\boldsymbol{x}_0\}$, all of $\boldsymbol{x}_{(n)}, \boldsymbol{x}^{(n)}$, $\boldsymbol{y}_{(n)}$ and $\boldsymbol{y}^{(n)}$ converge to $\boldsymbol{x}_0$ as $n \to \infty$. Also,

$$\begin{aligned} \sup_{\boldsymbol{x} \in I_n} \left| \frac{f_n(\boldsymbol{x}) - f(\boldsymbol{x})}{f(\boldsymbol{x})} \right| &= \sup_{\boldsymbol{x} \in I_n} \left| \left(\frac{1}{d_n} - 1 \right) + \frac{\theta g_n(\boldsymbol{x})}{d_n f(\boldsymbol{x})} \right| \\ &\le \left(1 - \frac{1}{d_n} \right) + \frac{\theta \sup_{\boldsymbol{x} \in I_n} \{g_n(\boldsymbol{x})\}}{d_n \inf_{\boldsymbol{x} \in I_n} \{f(\boldsymbol{x})\}} \\ &\to 0, \quad \text{as } n \to \infty, \end{aligned} \tag{4.16}$$

because g_n is pointwise non-increasing in $n \in \mathbb{N}$, $g_n(\boldsymbol{x}_0) = \mathcal{O}\left(n^{-1/3}\right)$ and $f(\boldsymbol{x}_0) > 0$.

Also,

$$D_1(n) \equiv \int_{I_n} \{f_n(\boldsymbol{x}) - f(\boldsymbol{x})\}^2 \, d\boldsymbol{x}$$
$$= \frac{1}{d_n^2} \int_{I_n} \left\{ \theta^2 g_n^2(\boldsymbol{x}) - \mathcal{O}\left(n^{-\frac{2}{3}}\right) f(\boldsymbol{x}) g_n(\boldsymbol{x}) + \mathcal{O}\left(n^{-\frac{4}{3}}\right) f^2(\boldsymbol{x}) \right\} \, d\boldsymbol{x}$$

and noticing that

$$0 \le \int_{I_n} \{g_n(\boldsymbol{x}) f(\boldsymbol{x})\} \, d\boldsymbol{x} \le f\left(\boldsymbol{x}^{(n)}\right) \int_{I_n} \{g_n(\boldsymbol{x})\} \, d\boldsymbol{x} = \mathcal{O}\left(n^{-\frac{2}{3}}\right),$$

so that,

$$nD_1(n) = \frac{n}{d_n^2} \left\{ \left(\frac{8}{3}\right)^d \theta^2 b^2 \prod_{i=1}^{d} \left\{h_i^3\right\} \cdot n^{-1} + o\left(n^{-\frac{4}{3}}\right) \right\}$$
$$\longrightarrow \left(\frac{8}{3}\right)^d \theta^2 b^2 \prod_{i=1}^{d} \left\{h_i^3\right\}, \quad \text{as } n \to \infty. \tag{4.17}$$

Now, since f is block-decreasing, we have,

$$0 < f\left(\boldsymbol{x}_0 + n^{-\frac{1}{3d}} \boldsymbol{I}_d \boldsymbol{h}\right) \le f(\boldsymbol{x}) \le f\left(\boldsymbol{x}_0 - n^{-\frac{1}{3d}} \boldsymbol{I}_d \boldsymbol{h}\right)$$

for all $\boldsymbol{x} \in I_n$ and $n \ge n_1$. Hence,

$$\frac{nD_1(n)}{f\left(\boldsymbol{x}_0 - n^{-\frac{1}{3d}} \boldsymbol{I}_d \boldsymbol{h}\right)} \le n \int_{I_n} \frac{\{f_n(\boldsymbol{x}) - f(\boldsymbol{x})\}^2}{f(\boldsymbol{x})} \, d\boldsymbol{x} \le \frac{nD_1(n)}{f\left(\boldsymbol{x}_0 + n^{-\frac{1}{3d}} \boldsymbol{I}_d \boldsymbol{h}\right)}$$

which, ahead with Eq. (4.17) and sandwich, yields

$$n \int_{I_n} \frac{\{f_n(\boldsymbol{x}) - f(\boldsymbol{x})\}^2}{f(\boldsymbol{x})} \, d\boldsymbol{x} \longrightarrow \left(\frac{8}{3}\right)^d \theta^2 b^2 \cdot \frac{\prod_{i=1}^{d} \left\{h_i^3\right\}}{f(\boldsymbol{x}_0)}, \quad \text{as } n \to \infty.$$

Applying all of the above to (4.15), and appealing to Lemma 2 of [29], we get

$$nh^2(f_n, f) = \frac{1}{8} \int_{I_n} \frac{\{f_n(\boldsymbol{x}) - f(\boldsymbol{x})\}^2}{f(\boldsymbol{x})} \, d\boldsymbol{x} + o(1) \tag{4.18}$$
$$\to \frac{8^{d-1}}{3^d f(\boldsymbol{x}_0)} \theta^2 b^2 \prod_{i=1}^{d} \left\{h_i^3\right\} \tag{4.19}$$

as $n \to \infty$, so that by applying (4.13) and (4.19) to Lemma 4.2, we get

$$\underline{\lim}_{n \to \infty} \inf_{T_n} \max \left\{ \boldsymbol{E}_{f_n} \left\{ n^{\frac{1}{3}} \left|T_n - f_n(\boldsymbol{x}_0)\right| \right\}, \boldsymbol{E}_f \left\{ n^{\frac{1}{3}} \left|T_n - f(\boldsymbol{x}_0)\right| \right\} \right\}$$
$$\ge \frac{1}{4} \left\{(-1)^d b\right\} \theta c \exp\left\{ -\frac{2^{3d-2}}{3^d f(\boldsymbol{x}_0)} \theta^2 b^2 c^3 \right\} =: G_{f,\boldsymbol{x}_0}(c, \theta)$$

where $c \equiv \prod_{i=1}^{d} \{h_i\}$. For a fixed $\theta \in (0, 1)$ the maximum of $G_{f,\boldsymbol{x}_0}(c, \theta)$ is attained at

$$c(\theta) = \left\{ \frac{3^{d-1} f(\boldsymbol{x}_0)}{2^{3d-2} \theta^2 b^2} \right\}^{\frac{1}{3}}$$

and is equal to

$$G_f\left(c(\theta), \theta\right) = \frac{e^{-\frac{1}{3}}}{2^d} \left\{3^{d-1} \theta\right\}^{\frac{1}{3}} \left\{ (-1)^d \left. \frac{\partial^d f(\boldsymbol{x})}{\partial x_1 \cdots \partial x_d} \right|_{\boldsymbol{x}=\boldsymbol{x}_0} f(\boldsymbol{x}_0) \right\}^{\frac{1}{3}},$$

the latter being an increasing function of $\theta \in (0, 1)$.

This implies that

$$\varliminf_{n\to\infty} \inf_{T_n} \max\left\{ \boldsymbol{E}_{f_n}\left\{ n^{\frac{1}{3}}\,|T_n - f_n(\boldsymbol{x}_0)| \right\}, \boldsymbol{E}_f\left\{ n^{\frac{1}{3}}\,|T_n - f(\boldsymbol{x}_0)| \right\} \right\}$$

$$\geq \frac{e^{-\frac{1}{3}}}{2^d}\left\{\theta\cdot 3^{d-1}\right\}^{\frac{1}{3}}\left\{ (-1)^d \left.\frac{\partial^d f(\boldsymbol{x})}{\partial x_1\cdots\partial x_d}\right|_{\boldsymbol{x}=\boldsymbol{x}_0}\cdot f(\boldsymbol{x}_0)\right\}^{\frac{1}{3}}.$$

Overall, we are allowed to take $\theta \uparrow 1$ in the above display, even if $\theta = 1$ is not a valid configuration, yielding the lower bound in the wording of the proposition. The proof is thus complete. □

5. Discussion and open problems

Once consistency has been established, interest focuses on rates of convergence of the MLE and other properties, including the behavior of $\widehat{f}_n$ at zero and pointwise limiting distributions. We have the following conjectures concerning the MLE $\widehat{f}_n$ for the class $\mathcal{F}_{\mathrm{SMU}}(d)$. Work is currently underway on all of these further problems.

Conjecture 1. *If $f_0(0) < \infty$, then we conjecture that $P_0(\widehat{f}_n(0) \leq M(\log n)^{d-1}) \to 1$ for some $M > 0$.*

Conjecture 2. *If $f_0(0) < \infty$ and f_0 is concentrated on $[\boldsymbol{0}, M\boldsymbol{1}]$ for some $0 < M < \infty$, then $h(\widehat{f}_n, f_0) = O_p(n^{-1/3}(\log n)^{\gamma})$ for some γ depending only on d.*

Concerning rates of convergence of the estimators at a fixed point, we do not yet have any upper bound results to accompany the lower bound results of Proposition 4.1. Thus there remain the following two possibilities: (a) the pointwise rate of convergence under Assumption 4.1 is $n^{1/3}$, and we expect convergence in distribution with the rate $n^{1/3}$, or, (b) the lower bound given in Proposition 4.1 is not yet sharp, and we should expect log terms in the rate (as might be expected from the covering number results of [10]). Our corresponding conjectures for these two possible scenarios are given below as Conjectures 3a and 3b respectively.

Conjecture 3a. *Suppose that f_0 has $\partial^d f_0(\boldsymbol{x})/\partial x_1\cdots\partial x_d$ continuous in a neighborhood of $\boldsymbol{x}_0$ with*

$$\partial^d f_0(\boldsymbol{x}_0) \equiv \left.\frac{\partial^d f_0(\boldsymbol{x})}{\partial x_1\cdots\partial x_d}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} \neq 0.$$

Let $\{W(\boldsymbol{t}):\ \boldsymbol{t}\in\mathbb{R}^d\}$ be a 2^d-sided Brownian sheet process on $\mathbb{R}^d$ and let

$$\mathbb{Y}(\boldsymbol{t}) \equiv \sqrt{f_0(\boldsymbol{x}_0)}W(\boldsymbol{t}) + \frac{(-1)^d}{2^d}(-1)^d\partial^d f_0(\boldsymbol{x}_0)|\boldsymbol{t}|^2.$$

Then, in keeping with our lower bound results of Section 4, we conjecture that

$$n^{1/3}(\widehat{f}_n(\boldsymbol{x}_0) - f_0(\boldsymbol{x}_0)) \to_d \partial^d\mathbb{H}(\boldsymbol{t})|_{\boldsymbol{t}=\boldsymbol{0}}$$

where the process $\mathbb{H}$ is determined by

(i) $\mathbb{H}(\boldsymbol{t}) \geq \mathbb{Y}(\boldsymbol{t})$ *for all* $\boldsymbol{t}\in\mathbb{R}^d$,

(ii) $\int_{\mathbb{R}^d}(\mathbb{H}(\boldsymbol{t}) - \mathbb{Y}(\boldsymbol{t}))d(\partial^d\mathbb{H}(\boldsymbol{t})) = 0,$ *and*

(iii) $V_{\partial^d\mathbb{H}}[\boldsymbol{u},\boldsymbol{v}) \geq 0$ *for all* $\boldsymbol{u}\leq\boldsymbol{v}\in\mathbb{R}^d$.

Partial results concerning Conjecture 3a were obtained in [37].

Conjecture 3b. *As suggested in part by the covering number results of [10], the pointwise rate of convergence is $(n/(\log n)^{d-1/2})^{1/3}$. This would entail an improved version of Proposition 4.1. In this case, we do not yet have conjectures concerning the limiting distribution.*

Acknowledgments

We owe thanks to Marina Meila, Fritz Scholz, and Arseni Seregin for helpful discussions concerning the proof of uniqueness, and especially Lemmas 3.3 and 3.4. We also thank the referees for several helpful suggestions and for catching a slip in a proof in the first version of the paper. The first author's research was supported by NSF grant DMS-0503822. The second author's research was supported by NSF grants DMS-0503822 and DMS-0804587 and NIH/NIAID grants 2R01 AI029168 and 4 R37 AI029168.

References

[1] D. Anevski, Estimating the derivative of a convex density. Technical Report, dept. of Math. Statistics, Univ. of Lund, 1994.
[2] D. Anevski, Estimating the derivative of a convex density, Stat. Neerl. 57 (2) (2003) 245–257.
[3] M. Ayer, H.D. Brunk, G.M. Ewing, W.T. Reid, E. Silverman, An empirical distribution function for sampling with incomplete information, Ann. Math. Statist. 26 (1955) 641–647.
[4] F. Balabdaoui, H. Jankowski, M. Pavlides, A. Seregin, J.A. Wellner, On the Grenander estimator at zero, Statist. Sinica 21 (2011) 873–879.
[5] R.E. Barlow, D.J. Bartholomew, J.M. Bremner, H.D. Brunk, Statistical inference under order restrictions, in: The Theory and Application of Isotonic Regression, in: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, London-New York-Sydney, 1972.
[6] G. Biau, L. Devroye, On the risk of estimates for block decreasing densities, J. Multivariate Anal. 86 (1) (2003) 143–165.
[7] P. Billingsley, Probability and measure, third ed. in: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, a Wiley-Interscience Publication, 1995.
[8] L. Birgé, Estimating a density under order restrictions: nonasymptotic minimax risk, Ann. Statist. 15 (3) (1987) 995–1012.
[9] L. Birgé, The Grenander estimator: a nonasymptotic approach, Ann. Statist. 17 (4) (1989) 1532–1549.
[10] R. Blei, F. Gao, W.V. Li, Metric entropy of high dimensional distributions, Proc. Amer. Math. Soc. 135 (12) (2007) 4009–4018. (electronic).
[11] H.D. Brunk, On the estimation of parameters restricted by inequalities, Ann. Math. Statist. 29 (1958) 437–454.
[12] H.D. Brunk, Estimation of isotonic regression, in: Nonparametric Techniques in Statistical Inference (Proc. Sympos., Indiana Univ., Bloomington, Ind., 1969), Cambridge Univ. Press, London, 1970, pp. 177–197.
[13] L. Devroye, Nonuniform Random Variate Generation, Springer-Verlag, New York, 1986.
[14] L. Devroye, A Course in Density Estimation, in: Progress in Probability and Statistics., vol. 14, Birkhäuser Boston Inc, Boston, MA, 1987.
[15] J.L. Doob, Measure Theory, in: Graduate Texts in Mathematics, vol. 143, Springer-Verlag, New York, 1994.
[16] W. Feller, An Introduction to Probability Theory and Its Applications, second ed. vol. II, John Wiley & Sons Inc., New York, 1971.
[17] N. Glick, Consistency conditions for probability estimators and integrals of density estimators, Util. Math. 6 (1974) 61–74.
[18] U. Grenander, On the theory of mortality measurement, I. Skand. Aktuarietidskr. 39 (1956) 70–96.
[19] U. Grenander, On the theory of mortality measurement. II, Skand. Aktuarietidskr. 39 (1957) 125–153.
[20] P. Groeneboom, Estimating a monotone density, in: Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, in: Wadsworth Statist./Probab. Ser. Wadsworth, vol. II, Berkeley, Calif, 1983, pp. 539–555. Belmont, CA.
[21] P. Groeneboom, Brownian motion with a parabolic drift and Airy functions, Probab. Theory Related Fields 81 (1) (1989) 79–109.
[22] P. Groeneboom, Lectures on inverse problems, in: Lectures on Probability Theory and Statistics, in: Lecture Notes in Math., vol. 1648, Springer, Berlin, 1996, pp. 67–164. Saint-Flour, 1994.
[23] P. Groeneboom, G. Jongbloed, Isotonic estimation and rates of convergence in Wicksell's problem, Ann. Statist. 23 (5) (1995) 1518–1542.
[24] P. Groeneboom, G. Jongbloed, J.A. Wellner, A canonical process for estimation of convex functions: the invelope of integrated Brownian motion $+t^4$, Ann. Statist. 29 (6) (2001) 1620–1652.
[25] P. Groeneboom, G. Jongbloed, J.A. Wellner, Estimation of a convex function: characterizations and asymptotic theory, Ann. Statist. 29 (6) (2001) 1653–1698.
[26] F.R. Hampel, Design, modelling, and analysis of some biological data sets, in: Design, Data & Analysis, John Wiley & Sons, Inc., New York, NY, USA, 1987, pp. 93–128.
[27] Jewell, P. Nicholas, van der Laan, Mark, Current status data: review, recent developments and open problems, in: Handbook of Statist., in: Advances in Survival Analysis, vol. 23, Elsevier, Amsterdam, 2004, pp. 625–642.
[28] G. Jongbloed, Three statistical inverse problems. Ph.D. Thesis, Delft University, 1995.
[29] G. Jongbloed, Minimax lower bounds and moduli of continuity, Statist. Probab. Lett. 50 (3) (2000) 279–284.
[30] J. Kim, D. Pollard, Cube root asymptotics, Ann. Statist. 18 (1) (1990) 191–219.
[31] R. Lang, A note on the measurability of convex sets, Arch. Math. (Basel) 47 (1) (1986) 90–92.
[32] D. Lavee, U.N. Safrie, I. Meilijson, For how long do trans-saharan migrants stop over at an oasis? Ornis Scandinavica 22 (1991) 33–44.
[33] L. Le Cam, Asymptotic Methods in Statistical Decision Theory, in: Springer Series in Statistics., Springer-Verlag, New York, 1986.
[34] B.G. Lindsay, The geometry of mixture likelihoods: a general theory, Ann. Statist. 11 (1) (1983) 86–94.
[35] B.G. Lindsay, Mixture Models: Theory, Geometry and Applications, in: NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 5, IMS, Hayward CA, 1995.
[36] D.W. Müller, G. Sawitzki, Excess mass estimates and tests for multimodality, J. Amer. Statist. Assoc. 86 (415) (1991) 738–746.
[37] M. Pavlides, Nonparametric estimation of multivariate monotone densities. Ph.D. Thesis, University of Washington, 2008.
[38] M. Pavlides, Local asymptotic minimax theory for block-decreasing densities. Tech. rep., Frederick University, Nicosia, Cyprus, 2009.
[39] J. Pfanzagl, Consistency of maximum likelihood estimators for certain nonparametric families, in particular: mixtures, J. Statist. Plann. Inference 19 (2) (1988) 137–158.
[40] W. Polonik, Density estimation under qualitative assumptions in higher dimensions, J. Multivariate Anal. 55 (1) (1995) 61–81.
[41] W. Polonik, Measuring mass concentrations and estimating density contour clusters—an excess mass approach, Ann. Statist. 23 (3) (1995) 855–881.
[42] W. Polonik, Minimum volume sets and generalized quantile processes, Stochastic Process. Appl. 69 (1) (1997) 1–24.
[43] W. Polonik, The silhouette, concentration functions and ML-density estimation under order restrictions, Ann. Statist. 26 (5) (1998) 1857–1877.
[44] B.L.S. Prakasa Rao, Estimation of a unimodal density, Sankhyā Ser. A 31 (1969) 23–36.
[45] T. Robertson, On estimating a density which is measurable with respect to a σ-lattice, Ann. Math. Statist. 38 (1967) 482–493.
[46] T. Robertson, F.T. Wright, R.L. Dykstra, Order restricted statistical inference, in: Probability and Mathematical Statistics, in: Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons Ltd, Chichester, 1988.
[47] R.T. Rockafellar, Convex Analysis. Princeton Mathematical Series, Princeton University Press, Princeton, N.J, 1970, No. 28.
[48] T.W. Sager, An iterative method for estimating a multivariate mode and isopleth, J. Amer. Statist. Assoc. 74 (1979) 329–339. 366, part 1.
[49] T.W. Sager, Nonparametric maximum likelihood estimation of spatial patterns, Ann. Statist. 10 (4) (1982) 1125–1136.
[50] G.R. Shorack, Probability for statisticians, in: Springer Texts in Statistics, Springer-Verlag, New York, 2000.
[51] C. van Eeden, Maximum likelihood estimation of ordered probabilities. Statist. Afdeling S 188 (VP 5). Math. Centrum Amsterdam, 1956.
[52] C. van Eeden, Maximum likelihood estimation of ordered probabilities, Nederl. Akad. Wetensch. Proc. Ser. A. **59** = Indag. Math. 18 (1956) 444–455.
[53] C. van Eeden, 1956. Maximum likelihood estimation of ordered probabilities. II. Statist. Afdeling Rep. S 196 (VP7). Math. Centrum Amsterdam.
[54] C. van Eeden, Maximum likelihood estimation of partially or completely ordered parameters. I, Nederl. Akad. Wetensch. Proc. Ser. A. **60** = Indag. Math. 19 (1957) 128–136.
[55] C. van Eeden, Maximum likelihood estimation of partially or completely ordered parameters. II, Nederl. Akad. Wetensch. Proc. Ser. A. **60** = Indag. Math. 19 (1957) 201–211.
[56] E.J. Wegman, A note on estimating a unimodal density, Ann. Math. Statist. 40 (1969) 1661–1667.
[57] E.J. Wegman, Maximum likelihood estimation of a unimodal density function, Ann. Math. Statist. 41 (1970) 457–471.
[58] E.J. Wegman, Maximum likelihood estimation of a unimodal density. II, Ann. Math. Statist. 41 (1970) 2169–2174.
[59] R.E. Williamson, Multiply monotone functions and their Laplace transforms, Duke Math. J. 23 (1956) 189–207.
[60] M. Woodroofe, J. Sun, A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing, Statist. Sinica 3 (2) (1993) 501–515.
[61] G.Y.C. Wong, Q. Yu, Generalized MLE of a joint distribution function with multivariate interval-censored data, J. Multivariate Anal. 69 (1999) 155–166.
[62] Yu Shaohua, Yu Qiqing, Y.C. Wong, George, Consistency of the generalized MLE of a joint distribution function with multivariate interval-censored data, J. Multivariate Anal. 97 (2005) 720–732.