The Annals of Statistics
2016, Vol. 44, No. 3, 1332–1359
DOI: 10.1214/15-AOS1408

APPROXIMATION AND ESTIMATION OF s-CONCAVE DENSITIES VIA RÉNYI DIVERGENCES

BY QIYANG HAN AND JON A. WELLNER[1]

University of Washington

In this paper, we study the approximation and estimation of s-concave densities via Rényi divergence. We first show that the approximation of a probability measure Q by an s-concave density exists and is unique via the procedure of minimizing a divergence functional proposed by [*Ann. Statist.* **38** (2010) 2998–3027] if and only if Q admits full-dimensional support and a first moment. We also show continuity of the divergence functional in Q: if $Q_n \to Q$ in the Wasserstein metric, then the projected densities converge in weighted L_1 metrics and uniformly on closed subsets of the continuity set of the limit. Moreover, directional derivatives of the projected densities also enjoy local uniform convergence. This contains both on-the-model and off-the-model situations, and entails strong consistency of the divergence estimator of an s-concave density under mild conditions. One interesting and important feature for the Rényi divergence estimator of an s-concave density is that the estimator is intrinsically related with the estimation of log-concave densities via maximum likelihood methods. In fact, we show that for $d = 1$ at least, the Rényi divergence estimators for s-concave densities converge to the maximum likelihood estimator of a log-concave density as $s \nearrow 0$. The Rényi divergence estimator shares similar characterizations as the MLE for log-concave distributions, which allows us to develop pointwise asymptotic distribution theory assuming that the underlying density is s-concave.

1. Introduction.

1.1. *Overview.* The class of s-concave densities on $\mathbb{R}^d$ is defined by the generalized means of order s as follows. Let

$$M_s(a, b; \theta) := \begin{cases} \big((1-\theta)a^s + \theta b^s\big)^{1/s}, & s \neq 0, a, b > 0, \\ 0, & s < 0, ab = 0, \\ a^{1-\theta} b^{\theta}, & s = 0, \\ a \wedge b, & s = -\infty. \end{cases}$$

Then a density $p(\cdot)$ on $\mathbb{R}^d$ is called s-concave, that is, $p \in \mathcal{P}_s$ if and only if for all $x_0, x_1 \in \mathbb{R}^d$ and $\theta \in (0, 1)$, $p((1-\theta)x_0 + \theta x_1) \geq M_s(p(x_0), p(x_1); \theta)$.

Received June 2015; revised October 2015.

[1]Supported in part by NSF Grant DMS-11-04832 and NI-AID Grant R01 AI029168.

MSC2010 subject classifications. Primary 62G07, 62H12; secondary 62G05, 62G20.

Key words and phrases. s-concavity, consistency, projection, asymptotic distribution, mode estimation, nonparametric estimation, shape constraints.

This definition apparently goes back to Avriel (1972) with further studies by Borell (1974, 1975), Das Gupta (1976), Rinott (1976), and Uhrin (1984); see also Dharmadhikari and Joag-Dev (1988) for a nice summary. It is easy to see that the densities $p(\cdot)$ have the form $p = \varphi_+^{1/s}$ for some concave function φ if $s > 0$, $p = \exp(\varphi)$ for some concave φ if $s = 0$, and $p = \varphi_+^{1/s}$ for some convex φ if $s < 0$. The function classes $\mathcal{P}_s$ are nested in s in that for every $r > 0 > s$, we have $\mathcal{P}_r \subset \mathcal{P}_0 \subset \mathcal{P}_s \subset \mathcal{P}_{-\infty}$.

Nonparametric estimation of s-concave densities has been under intense research efforts in recent years. In particular, much attention has been paid to estimation in the special case $s = 0$ which corresponds to all log-concave densities on $\mathbb{R}^d$. The nonparametric maximum likelihood estimator (MLE) of a log-concave density was studied in the univariate setting by Walther (2002), Dümbgen and Rufibach (2009), Pal, Woodroofe and Meyer (2007); and in the multivariate setting by Cule and Samworth (2010), Cule, Samworth and Stewart (2010). The limiting distribution theory at fixed points when $d = 1$ was studied in Balabdaoui, Rufibach and Wellner (2009), and rate results in Doss and Wellner (2016), Kim and Samworth (2015). Dümbgen, Samworth and Schuhmacher (2011) also studied stability properties of the MLE projection of any probability measure onto the class of log-concave densities.

Compared with the well-studied log-concave densities (i.e., $s = 0$), much remains unknown concerning estimation and inference procedures for the larger classes $\mathcal{P}_s, s < 0$. One important feature for this larger class is that the densities in $\mathcal{P}_s (s < 0)$ are allowed to have heavier and heavier tails as $s \to -\infty$. In fact, t-distributions with ν degrees of freedom belong to $\mathcal{P}_{-1/(\nu+1)}(\mathbb{R})$ [and hence also to $\mathcal{P}_s(\mathbb{R})$ for any $s < -1/(\nu + 1)$]. The study of maximum likelihood estimators (MLEs in the following) for general s-concave densities in Seregin and Wellner (2010) shows that the MLE exists and is consistent for $s \in (-1, \infty)$. However, there is no known result about uniqueness of the MLE of s-concave densities except for $s = 0$. The difficulties in the theory of estimation via MLE lie in the fact we have still very little knowledge of "good" characterizations of the MLE in the s-concave setting. This has hindered further development of both theoretical and statistical properties of the estimation procedure.

Some alternative approaches to estimation of s-concave densities have been proposed in the literature by using divergences other than the log-likelihood functional (Kullback–Leibler divergence in some sense). Koenker and Mizera (2010) proposed an alternative to maximum likelihood based on generalized Rényi entropies. Similar procedures were also proposed in parametric settings by Basu et al. (1998) using a family of discrepancy measures. In our setting of s-concave densities with $s < 0$, the methods of Koenker and Mizera (2010) can be formulated as follows.

Given i.i.d. observations $\underline{X} = (X_1, \ldots, X_n)$, consider the primal optimization problem $(\mathcal{P})$:

$$(1.1) \qquad (\mathcal{P}) \qquad \min_{g \in \mathcal{G}(\underline{X})} L(g, \mathbb{Q}_n) \equiv \frac{1}{n} \sum_{i=1}^{n} g(X_i) + \frac{1}{|\beta|} \int_{\mathbb{R}^d} g(x)^{\beta} \, dx,$$

where $\mathcal{G}(\underline{X})$ denotes all nonnegative closed convex functions supported on the convex set $\mathrm{conv}(\underline{X})$, $\mathbb{Q}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ the empirical measure and $\beta = 1 + 1/s < 0$. As is shown by Koenker and Mizera (2010), the associated dual problem $(\mathcal{D})$ is

$$(1.2)\qquad \begin{aligned} &(\mathcal{D}) \qquad \max_f \int_{\mathbb{R}^d} \frac{(f(y))^\alpha}{\alpha}\,\mathrm{d}y, \\ &\text{subject to} \qquad f = \frac{\mathrm{d}(\mathbb{Q}_n - G)}{\mathrm{d}y} \qquad \text{for some } G \in \mathcal{G}(\underline{X})^\circ, \end{aligned}$$

where $\mathcal{G}(\underline{X})^\circ \equiv \{G \in \mathcal{C}^*(\underline{X}) | \int g\,\mathrm{d}G \le 0, \text{for all } g \in \mathcal{G}(\underline{X})\}$ is the polar cone of $\mathcal{G}(\underline{X})$, and α is the conjugate index of β, that is, $1/\alpha + 1/\beta = 1$. Here, $\mathcal{C}^*(\underline{X})$, the space of signed Radon measures on $\mathrm{conv}(\underline{X})$, is the topological dual of $\mathcal{C}(\underline{X})$, the space of continuous functions on $\mathrm{conv}(\underline{X})$. We also note that the constraint $G \in \mathcal{G}(\underline{X})^\circ$ in the dual form (1.2) comes from the "dual" of the primal constraint $g \in \mathcal{G}(\underline{X})$, and the constraint $f = \frac{\mathrm{d}(\mathbb{Q}_n - G)}{\mathrm{d}y}$ can be derived from the dual computation of $L(\cdot, \mathbb{Q}_n)$:

$$\begin{aligned} \big(L(\cdot, \mathbb{Q}_n)\big)^*(G) &= \sup_g \left(\langle G, g\rangle - \frac{1}{n}\sum_{i=1}^n g(X_i) - \frac{1}{|\beta|}\int_{\mathbb{R}^d} g(x)^\beta\,\mathrm{d}x \right) \\ &= \sup_g \left(\langle G - \mathbb{Q}_n, g\rangle - \int \psi_s\big(g(x)\big)\,\mathrm{d}x \right) = \Psi_s^*(G - \mathbb{Q}_n). \end{aligned}$$

Here, we used the notation $\langle G, g\rangle := \int g\,\mathrm{d}G$, $\psi_s(\cdot) := (\cdot)^\beta/|\beta|$ and Ψ_s is the functional defined by $\Psi_s(g) := \int \psi_s(g(x))\,\mathrm{d}x$ for clarity. Now the dual form (1.2) follows by the well-known fact [e.g., Rockafellar (1971) Corollary 4A] that the form of the above dual functional is given by

$$\Psi^*(G) = \begin{cases} \displaystyle\int \psi^*(\mathrm{d}G/\mathrm{d}x)\,\mathrm{d}x, & \text{if } G \text{ is absolute continuous with respect to} \\ & \text{Lebesgue measure}, \\ +\infty, & \text{otherwise}. \end{cases}$$

For the primal problem $(\mathcal{P})$ and the dual problem $(\mathcal{D})$, Koenker and Mizera (2010) proved the following results:

THEOREM 1.1 [Theorem 4.1, Koenker and Mizera (2010)]. *$(\mathcal{P})$ admits a unique solution g_n^* if* $\mathrm{int}(\mathrm{conv}(\underline{X})) \neq \varnothing$, *where g_n^* is a polyhedral convex function supported on* $\mathrm{conv}(\underline{X})$.

THEOREM 1.2 [Theorem 3.1, Koenker and Mizera (2010)]. *Strong duality between $(\mathcal{P})$ and $(\mathcal{D})$ holds. Any dual feasible solution is actually a density on $\mathbb{R}^d$ with respect to the canonical Lebesgue measure. The dual optimal solution f_n^* exists, and satisfies $f_n^* = (g_n^*)^{1/s}$.*

We note that the above results are all obtained in the empirical setting. At the population level, given a probability measure Q with suitable regularity conditions, consider

$$(1.3) \qquad (\mathcal{P}_Q) \qquad \min_{g \in \mathcal{G}} L_s(g, Q),$$

where

$$L(g, Q) \equiv L_s(g, Q) \equiv \int g(x)\,\mathrm{d}Q + \frac{1}{|\beta|}\int_{\mathbb{R}^d} g(x)^{\beta}\,\mathrm{d}x,$$

and $\mathcal{G}$ denotes the class of all (nonnegative) closed convex functions with non-empty interior, which are coercive in the sense that $g(x) \to \infty$, as $\|x\| \to \infty$. Koenker and Mizera (2010) show that Fisher consistency holds at the population level: Suppose $Q(A) := \int_A f_0\,\mathrm{d}\lambda$ is defined for some $f_0 = g_0^{1/s}$ where $g_0 \in \mathcal{G}$; then g_0 is an optimal solution for $(\mathcal{P}_Q)$.

Koenker and Mizera (2010) also proposed a general discretization scheme corresponding to the primal form (1.1) and the dual form (1.2) for fast computation, by which the one-dimensional problem can be solved via linear programming and the two-dimensional problem via semi-definite programming. These have been implemented in the `R` package `REBayes` by Koenker and Mizera (2014). Koenker's package depends in turn on the `MOSEK` implementation of MOSEK ApS (2011); see Appendix B of Koenker and Mizera (2010) for further details. On the other hand, in the special case $s = 0$, computation of the MLEs of log-concave densities has been implemented in the `R` package `LogConcDEAD` developed in Cule, Samworth and Stewart (2010) in arbitrary dimensions. However, expensive search for the proper triangulation of the support $\mathrm{conv}(\underline{X})$ renders computation difficult in high dimensions.

In this paper, we show that the estimation procedure proposed by Koenker and Mizera (2010) is the "natural" way to estimate s-concave densities. As a starting point, since the classes $\mathcal{P}_s$ are nested in s, it is natural to consider estimation of the extreme case $s = 0$ (the class of log-concave densities) as some kind of "limit" of estimation of the larger class $s < 0$. As we will see, estimation of s-concave distributions via Rényi divergences is intrinsically related with the estimation of log-concave distributions via maximum likelihood methods. In fact, we show that in the empirical setting in dimension 1, the Rényi divergence estimators converge to the maximum likelihood estimator for log-concave densities as $s \nearrow 0$.

We will show that the Rényi divergence estimators share characterization and stability properties similar to the analogous properties established in the log-concave setting by Cule and Samworth (2010), Dümbgen and Rufibach (2009) and Dümbgen, Samworth and Schuhmacher (2011). Once these properties are available, further theoretical and statistical considerations in estimation of s-concave densities become possible. In particular, the characterizations developed here enable us to overcome some of the difficulties of maximum likelihood estimators as proposed by Seregin and Wellner (2010), and to develop limit distribution theory at fixed points assuming that the underlying model is s-concave. The pointwise rate and limit distribution results follow a pattern similar to the corresponding results for the MLEs in the log-concave setting obtained by Balabdaoui, Rufibach and Wellner (2009). This local point of view also underlines the results on global rates of convergence considered in Doss and Wellner (2016), showing that the difficulty of estimation for such densities with tails light or heavy, comes almost solely from the shape constraints, namely, the convexity-based constraints.

The rest of the paper is organized as follows. In Section 2, we study the basic theoretical properties of the approximation/projection scheme defined by the procedure (1.3). In Section 3, we study the limit behavior of s-concave probability measures in the setting of weak convergence under dimensionality conditions on the supports of the limiting sequence. In Section 4, we develop limiting distribution theory of the divergence estimator in dimension 1 under curvature conditions with tools developed in Sections 2 and 3. Related issues and further problems are discussed in Section 5. Some proofs are given in Section 6; most of the proofs and some auxiliary results are presented in the Supplementary Material, Han and Wellner (2015).

1.2. *Notation*. In this paper, we denote the canonical Lebesgue measure on $\mathbb{R}^d$ by λ or λ_d and write $\|\cdot\|_p$ for the canonical Euclidean p-norm in $\mathbb{R}^d$, and $\|\cdot\| = \|\cdot\|_2$ unless otherwise specified. $B(x,\delta)$ stands for the open ball of radius δ centered at x in $\mathbb{R}^d$, and $\mathbf{1}_A$ for the indicator function of $A \subset \mathbb{R}^d$. We use $L_p(f) \equiv \|f\|_{L_p} \equiv \|f\|_p = (\int |f|^p \, d\lambda_d)^{1/p}$ to denote the $L_p(\lambda_d)$ norm of a measurable function f on $\mathbb{R}^d$ if no confusion arises.

We write $\operatorname{csupp}(Q)$ for the convex support of a measure Q defined on $\mathbb{R}^d$, that is,

$$\operatorname{csupp}(Q) = \bigcap \{C : C \subset \mathbb{R}^d \text{ closed and convex}, Q(C) = 1\}.$$

We let $\mathcal{Q}_0$ denote all probability measures on $\mathbb{R}^d$ whose convex support has nonvoid interior, while $\mathcal{Q}_1$ denotes the set of all probability measures Q with finite first moment: $\int \|x\| Q(dx) < \infty$.

We write $f_n \to_d f$ if P_n converges weakly to P for the corresponding probability measures $P_n(A) \equiv \int_A f_n \, d\lambda$ and $P(A) \equiv \int_A f \, d\lambda$.

We write $\alpha := 1 + s$, $\beta := 1 + 1/s$, $r := -1/s$ unless otherwise specified.

2. Theoretical properties of the divergence estimator. In this section, we study the basic theoretical properties of the proposed projection scheme via Rényi divergence (1.3). Starting from a given probability measure Q, we first show the existence and uniqueness of such projections via Rényi divergence under assumptions on the index s and Q. We will call such a projection the *Rényi divergence estimator* for the given probability measure Q in the following discussions. We next show that the projection scheme is continuous in Q in the following sense: if a sequence of probability measures Q_n, for which the projections onto the class of s-concave densities exist, converge to a limiting probability measure Q in Wasserstein distance, then the corresponding projected densities converge in weighted L_1 metrics and uniformly on closed subsets of the continuity set of the limit. The directional derivatives of such projected densities also converge uniformly in all directions in a local sense. We then turn our attention the explicit characterizations of the Rényi divergence estimators, especially in dimension 1. This helps in two ways. First, it helps to understand the continuity of the projection scheme in the index s, that is, answers affirmatively the question: For a given probability measure Q, does the Rényi divergence estimator converge to the log-concave projection as studied in Dümbgen, Samworth and Schuhmacher (2011) as $s \nearrow 0$? Second, the explicit characterizations are exploited in the development of asymptotic distribution theory presented in Section 4.

2.1. *Existence and uniqueness*. For a given probability measure Q, let $L(Q) = \inf_{g \in \mathcal{G}} L(g, Q)$.

LEMMA 2.1. *Assume* $-1/(d+1) < s < 0$ *and* $Q \in \mathcal{Q}_0$. *Then* $L(Q) < \infty$ *if and only if* $Q \in \mathcal{Q}_1$.

Now we state our main theorem for the existence of Rényi divergence projection corresponding to a general measure Q on $\mathbb{R}^d$.

THEOREM 2.2. *Assume* $-1/(d+1) < s < 0$ *and* $Q \in \mathcal{Q}_0 \cap \mathcal{Q}_1$. *Then* (1.3) *achieves its nontrivial minimum for some* $\tilde{g} \in \mathcal{G}$. *Moreover*, $\tilde{g}$ *is bounded away from zero, and* $\tilde{f} \equiv \tilde{g}^{1/s}$ *is a bounded density with respect to* λ_d.

The uniqueness of the solution follows immediately from the strict convexity of the functional $L(\cdot, Q)$.

LEMMA 2.3. $\tilde{g}$ *is the unique solution for* $(\mathcal{P}_Q)$ *if* $\operatorname{int}(\operatorname{dom}(\tilde{g})) \neq \emptyset$.

REMARK 2.4. By the above discussion, we conclude that the map $Q \mapsto \arg\min_{g \in \mathcal{G}} L(g, Q)$ is well-defined for probability measures Q with suitable regularity conditions: in particular, if $Q \in \mathcal{Q}_0$ and $-1/(d+1) < s < 0$, it is well-defined if and only if $Q \in \mathcal{Q}_1$. From now on, we denote the optimal solution as $g_s(\cdot|Q)$ or simply $g(\cdot|Q)$ if no confusion arises, and write P_Q for the corresponding s-concave distribution, and say that P_Q is the Rényi projection of Q to $P_Q \in \mathcal{P}_s$.

2.2. *Weighted global convergence in* $\|\cdot\|_{L_1}$ *and* $\|\cdot\|_\infty$.

THEOREM 2.5. *Assume* $-1/(d+1) < s < 0$. *Let* $\{Q_n\} \subset \mathcal{Q}_0$ *be a sequence of probability measures converging weakly to* $Q \subset \mathcal{Q}_0 \cap \mathcal{Q}_1$. *Then*

$$\int \|x\| \, \mathrm{d}Q \le \liminf_{n\to\infty} \int \|x\| \, \mathrm{d}Q_n. \tag{2.1}$$

If we further assume that

$$\lim_{n\to\infty} \int \|x\| \, \mathrm{d}Q_n = \int \|x\| \, \mathrm{d}Q, \tag{2.2}$$

then,

$$L(Q) = \lim_{n\to\infty} L(Q_n). \tag{2.3}$$

Conversely, if (2.3) *holds, then* (2.2) *holds true. In the former case* [*i.e.*, (2.2) *holds*], *let* $g := g(\cdot|Q)$ *and* $g_n := g(\cdot|Q_n)$, *then* $f := g^{1/s}$, $f_n := g_n^{1/s}$ *satisfy*

$$\begin{aligned} &\lim_{n\to\infty, x\to y} f_n(x) = f(y) \qquad \text{for all } y \in \mathbb{R}^d \setminus \partial\{f > 0\}, \\ &\limsup_{n\to\infty, x\to y} f_n(x) \le f(y) \qquad \text{for all } y \in \mathbb{R}^d. \end{aligned} \tag{2.4}$$

For $\kappa < r - d \equiv -1/s - d$,

$$\lim_{n\to\infty} \int (1 + \|x\|)^\kappa \big| f_n(x) - f(x) \big| \, \mathrm{d}x = 0. \tag{2.5}$$

For any closed set S *contained in the continuity points of* f *and* $\kappa < r$,

$$\lim_{n\to\infty} \sup_{x\in S} (1 + \|x\|)^\kappa \big| f_n(x) - f(x) \big| = 0. \tag{2.6}$$

Furthermore, let $\mathcal{D}_f := \{x \in \mathrm{int}(\mathrm{dom}(f)) : f$ *is differentiable at* $x\}$, *and* $T \subset \mathrm{int}(\mathcal{D}_f)$ *be any compact set. Then*

$$\lim_{n\to\infty} \sup_{x\in T, \|\xi\|_2 = 1} \big| \nabla_\xi f_n(x) - \nabla_\xi f(x) \big| = 0, \tag{2.7}$$

where $\nabla_\xi f(x) := \lim_{h\searrow 0} \frac{f(x+h\xi)-f(x)}{h}$ *denotes the* (*one-sided*) *directional derivative along* ξ.

REMARK 2.6. The one-sided directional derivative for a convex function g is well-defined and $\nabla_\xi g(x) = \inf_{h>0} \frac{g(x+h\xi)-g(x)}{h}$, hence well-defined for $f \equiv g^{1/s}$. See Section 23 in Rockafellar (1997) for more details.

As a direct consequence, we have the following result covering both on and off-the-model cases.

COROLLARY 2.7. *Assume* $-1/(d+1) < s < 0$. *Let* Q *be a probability measure such that* $Q \in \mathcal{Q}_0 \cap \mathcal{Q}_1$, *with* $f_Q := g(\cdot|Q)^{1/s}$ *the density function corresponding to the Rényi projection* P_Q (*as in Remark* 2.4). *Let* $\mathbb{Q}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ *be the empirical measure when* $X_1, \dots, X_n$ *are i.i.d. with distribution* Q *on* $\mathbb{R}^d$. *Let* $\hat{g}_n := g(\cdot|\mathbb{Q}_n)$ *and* $\hat{f}_n := \hat{g}_n^{1/s}$ *be the Rényi divergence estimator of* Q. *Then almost surely we have*

$$\begin{aligned} &\lim_{n\to\infty, x\to y} \hat{f}_n(x) = f_Q(y) \qquad \text{for all } y \in \mathbb{R}^d \setminus \partial\{f > 0\}, \\ &\limsup_{n\to\infty, x\to y} \hat{f}_n(x) \le f_Q(y) \qquad \text{for all } y \in \mathbb{R}^d. \end{aligned} \tag{2.8}$$

For $\kappa < r - d \equiv -1/s - d$,

$$\lim_{n\to\infty} \int (1 + \|x\|)^\kappa \left|\hat{f}_n(x) - f_Q(x)\right| \mathrm{d}x =_{a.s.} 0. \tag{2.9}$$

For any closed set S *contained in the continuity points of* f *and* $\kappa < r$,

$$\lim_{n\to\infty} \sup_{x\in S} (1 + \|x\|)^\kappa \left|\hat{f}_n(x) - f_Q(x)\right| =_{a.s.} 0. \tag{2.10}$$

Furthermore, for any compact set $T \subset \operatorname{int}(\mathcal{D}_{f_Q})$,

$$\lim_{n\to\infty} \sup_{x\in T, \|\xi\|_2 = 1} \left|\nabla_\xi \hat{f}_n(x) - \nabla_\xi f_Q(x)\right| =_{a.s.} 0. \tag{2.11}$$

Now we return to the correctly specified case and relax the previous assumption that $s > -1/(d+1)$ for the case of the empirical measure $Q_n \equiv \mathbb{Q}_n$ and some measure Q with finite mean and bounded density $f \in \mathcal{P}_{s'} \subset \mathcal{P}_s$ with $s' > s$.

COROLLARY 2.8. *Assume* $-1/d < s < 0$. *Let* Q *be a probability measure on* $\mathbb{R}^d$ *with density* $f \in \mathcal{P}_s$ *if* $-1/(d+1) < s$ *and* $f \in \mathcal{P}_{s'}$ *where* $s' > -1/(d+1)$ *if* $s \in (-1/d, -1/(d+1)]$. (*Thus,* f *is bounded and* f *has a finite mean.*) *Let* $\hat{f}_n \equiv \hat{f}_{n,s}$ *be defined as in Corollary* 2.7. *Then* (2.8), (2.9), (2.10) *and* (2.11) *hold with* f_Q *replaced by* f.

2.3. *Characterization of the Rényi divergence projection and estimator.* We now develop characterizations for the Rényi divergence projection, especially in dimension 1. All proofs for this subsection can be found in the Supplementary Material.

We note that the assumption $-1/(d+1) < s < 0$ is imposed only for the existence and uniqueness of the Rényi divergence projection. For the specific case of empirical measure $Q_n \equiv \mathbb{Q}_n$, this condition can be relaxed to $-1/d < s < 0$.

Now we give a variational characterization in the spirit of Theorem 2.2 in Dümbgen and Rufibach (2009). This result holds for all dimensions $d \ge 1$.

THEOREM 2.9. *Assume* $-1/(d+1) < s < 0$ *and* $Q \in \mathcal{Q}_0 \cap \mathcal{Q}_1$. *Then* $g = g(\cdot|Q)$ *if and only if*

$$\int h \cdot g^{1/s} \, \mathrm{d}\lambda \le \int h \, \mathrm{d}Q, \tag{2.12}$$

holds for all $h : \mathbb{R}^d \to \mathbb{R}$ *such that there exists* $t_0 > 0$ *with* $g + th \in \mathcal{G}$ *holds for all* $t \in (0, t_0)$.

COROLLARY 2.10. *Assume* $-1/(d+1) < s < 0$ *and* $Q \in \mathcal{Q}_0 \cap \mathcal{Q}_1$ *and let* h *be any closed convex function. Then*

$$\int h \, \mathrm{d}P \le \int h \, \mathrm{d}Q,$$

where $P = P_Q$ *is the Rényi projection of* Q *to* $P_Q \in \mathcal{P}_s$ (*as in Remark* 2.4).

As a direct consequence, we have the following.

COROLLARY 2.11 (Moment inequalities). *Assume* $-1/(d+1) < s < 0$ *and* $Q \in \mathcal{Q}_0 \cap \mathcal{Q}_1$. *Let* $\mu_Q := \mathbb{E}_Q[X]$. *Then* $\mu_P = \mu_Q$. *Furthermore, if* $-1/(d+2) < s < 0$, *we have* $\lambda_{\max}(\Sigma_P) \le \lambda_{\max}(\Sigma_Q)$ *and* $\lambda_{\min}(\Sigma_P) \le \lambda_{\min}(\Sigma_Q)$ *where* Σ_Q *is the covariance matrix defined by* $\Sigma_Q := \mathbb{E}_Q[(X - \mu_Q)(X - \mu_Q)^T]$. *Generally if* $-1/(d+k) < s < 0$ *for some* $k \in \mathbb{N}$, *then* $\mathbb{E}_P[\|X\|^l] \le \mathbb{E}_Q[\|X\|^l]$ *holds for all* $l = 1, \ldots, k$.

Now we restrict our attention to $d = 1$, and in the following we will give a full characterization of the Rényi divergence estimator. Suppose we observe $X_1, \ldots, X_n$ i.i.d. Q on $\mathbb{R}$, and let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ be the order statistics of $X_1, \ldots, X_n$. Let $\mathbb{F}_n$ be the empirical distribution function corresponding to the empirical probability measure $\mathbb{Q}_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$. Let $\hat{g}_n := g(\cdot|\mathbb{Q}_n)$ and $\hat{F}_n(t) := \int_{-\infty}^t \hat{g}_n^{1/s}(x)\,\mathrm{d}x$. From Theorem 4.1 in Koenker and Mizera (2010), it follows that $\hat{g}_n$ is a convex function supported on $[X_{(1)}, X_{(n)}]$, and linear on $[X_{(i)}, X_{(i+1)}]$ for all $i = 1, \ldots, n-1$. For a continuous piecewise linear function $h : [X_{(1)}, X_{(n)}] \to \mathbb{R}$, define the set of knots to be

$$\mathcal{S}_n(h) := \{t \in (X_{(1)}, X_{(n)}) : h'(t-) \neq h'(t+)\} \cap \{X_1, \ldots, X_n\}.$$

THEOREM 2.12. *Let* g_n *be a convex function taking the value* $+\infty$ *on* $\mathbb{R} \setminus [X_{(1)}, X_{(n)}]$ *and linear on* $[X_{(i)}, X_{(i+1)}]$ *for all* $i = 1, \ldots, n-1$. *Let*

$$F_n(t) := \int_{-\infty}^t g_n^{1/s}(x)\,\mathrm{d}x.$$

Assume $F_n(X_{(n)}) = 1$. *Then* $g_n = \hat{g}_n$ *if and only if*

$$\int_{X_{(1)}}^t \big(F_n(x) - \mathbb{F}_n(x)\big)\,\mathrm{d}x \begin{cases} = 0, & \text{if } t \in \mathcal{S}_n(g_n), \\ \le 0, & \text{otherwise}. \end{cases} \tag{2.13}$$

COROLLARY 2.13. *For $x_0 \in \mathcal{S}_n(\hat{g}_n)$, we have*

$$\mathbb{F}_n(x_0) - \frac{1}{n} \le \hat{F}_n(x_0) \le \mathbb{F}_n(x_0).$$

Finally, we give a characterization of the Rényi divergence estimator in terms of distribution function as Theorem 2.7 in Dümbgen, Samworth and Schuhmacher (2011).

THEOREM 2.14. *Assume $-1/2 < s < 0$ and $Q \in \mathcal{Q}_0 \cap \mathcal{Q}_1$ is a probability measure on $\mathbb{R}$ with distribution function $G(\cdot)$. Let $g \in \mathcal{G}$ be such that $f \equiv g^{1/s}$ is a density on $\mathbb{R}$, with distribution function $F(\cdot)$. Then $g = g(\cdot|Q)$ if and only if*:

1. $\int_{\mathbb{R}} (F - G)(t)\,\mathrm{d}t = 0$;
2. $\int_{-\infty}^{x} (F - G)(t)\,\mathrm{d}t \le 0$ *for all $x \in \mathbb{R}$ with equality when $x \in \tilde{\mathcal{S}}(g)$.*

Here, $\tilde{\mathcal{S}}(g) := \{x \in \mathbb{R} : g(x) < \frac{1}{2}(g(x+\delta) + g(x-\delta))$ holds for $\delta > 0$ small enough$\}$.

The above theorem is useful for understanding the projected s-concave density given an arbitrary probability measure $Q \in \mathcal{Q}_0 \cap \mathcal{Q}_1$. The following example illustrates these projections and also gives some insight concerning the boundary properties of the class of s-concave densities.

EXAMPLE 2.15. Consider the class of densities $\mathcal{Q}$ defined by

$$\mathcal{Q} = \left\{ q_\tau(x) = \frac{\tau - 1}{2(\tau - 2)} \left(1 + \frac{|x|}{\tau - 2}\right)^{-\tau}, \tau > 2 \right\}.$$

Note that q_τ is $-1/\tau$-concave and *not* s-concave for any $0 > s > -1/\tau$. We start from arbitrary $q_\tau \in \mathcal{Q}$ with $\tau > 2$, and we will show in the following that the projection of q_τ onto the class of s-concave ($0 > s > -1/\tau$) distribution through $L(\cdot, q_\tau)$ will be given by $q_{-1/s}$. Let Q_τ be the distribution function of $q_\tau(\cdot)$, then we can calculate

$$Q_\tau(x) = \begin{cases} \dfrac{1}{2}\left(1 - \dfrac{x}{\tau - 2}\right)^{-(\tau-1)}, & \text{if } x \le 0, \\ 1 - \dfrac{1}{2}\left(1 + \dfrac{x}{\tau - 2}\right)^{-(\tau-1)}, & \text{if } x > 0. \end{cases}$$

It is easy to check by direct calculation that $\int_{-\infty}^{x} (Q_r(t) - Q_\tau(t))\,\mathrm{d}t \le 0$ with equality attained if and only if $x = 0$. It is clear that $\tilde{\mathcal{S}}(q_\tau) = \{0\}$ and hence the conditions in Theorem 2.14 are verified. Note that, in Example 2.9 of Dümbgen, Samworth and Schuhmacher (2011), the log-concave approximation of the rescaled t_2 density is the Laplace distribution. It is easy to see from the above calculation that the log-concave projection of the whole class $\mathcal{Q}$ will be the Laplace distribution $q_\infty = \frac{1}{2}\exp(-|x|)$. Therefore, the log-concave approximation fails to distinguish densities at least amongst the class $\mathcal{Q} \cup \{t_2\}$.

2.4. *Continuity of the Rényi divergence estimator in* s. Recall that $\alpha = 1 + s$, and then α, β is a conjugate pair with $\alpha^{-1} + \beta^{-1} = 1$ where $\beta = 1 + 1/s$. For $1 - 1/d < \alpha < 1$, let

$$F_\alpha(f) = \frac{1}{\alpha - 1} \log \int f^\alpha(x)\,\mathrm{d}x,$$
$$F_1(f) = \int f(x) \log f(x)\,\mathrm{d}x.$$

For a given index $-1/d < s < 0$, and data $\underline{X} = (X_1, \ldots, X_n)$ with non-void $\mathrm{int}(\mathrm{conv}(\underline{X}))$, solving the dual problem (1.2) for the primal problem (1.1) is equivalent to solving

$$(2.14)\qquad \begin{array}{lll} (\mathcal{D}_\alpha) & \min_f F_\alpha(f) = \dfrac{1}{\alpha - 1} \log \displaystyle\int f^\alpha(x)\,\mathrm{d}x & \\ & \text{subject to} \quad f = \dfrac{\mathrm{d}(\mathbb{Q}_n - G)}{\mathrm{d}y} & \text{for some } G \in \mathcal{G}(\underline{X})^\circ, \end{array}$$

where $\mathcal{G}(\underline{X})^\circ$ is the polar cone of $\mathcal{G}(\underline{X})$ and $\mathbb{Q}_n = \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$ is the empirical measure. The maximum likelihood estimation of a log-concave density has dual form

$$(2.15)\qquad \begin{array}{lll} (\mathcal{D}_1) & \min_f F_1(f) = \displaystyle\int f(x) \log f(x)\,\mathrm{d}x, & \\ & \text{subject to} \quad f = \dfrac{\mathrm{d}(\mathbb{Q}_n - G)}{\mathrm{d}y} & \text{for some } G \in \mathcal{G}(\underline{X})^\circ. \end{array}$$

Let f_α and f_1 be the solutions of $(\mathcal{D}_\alpha)$ and $(\mathcal{D}_1)$. For simplicity, we drop the explicit notational dependence of f_α, f on n. Since $F_\alpha(f) \to F_1(f)$ as $\alpha \nearrow 1$ for f smooth enough, it is natural to expect some convergence property of f_α to f_1. The main result is summarized as follows.

THEOREM 2.16. *Suppose $d = 1$. For all $\kappa > 0$, $p \geq 1$, we have the following weighted convergence*:

$$\lim_{\alpha \nearrow 1} \int (1 + \|x\|)^\kappa \left| f_\alpha(x) - f_1(x) \right|^p \mathrm{d}x = 0.$$

Moreover, for any closed set S contained in the continuity points of f,

$$\lim_{\alpha \nearrow 1} \sup_{x \in S} (1 + \|x\|)^\kappa \left| f_\alpha(x) - f_1(x) \right| = 0$$

for all $\kappa > 0$.

3. Limit behavior of s-concave densities. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of s-concave densities with corresponding measures $d\nu_n = f_n\, d\lambda$. Suppose $\nu_n \to_d \nu$. From Borell (1974, 1975) and Brascamp and Lieb (1976), we know that each ν_n is a t-concave measure with $t = s/(1+sd)$ if $-1/d < s < \infty$, $t = -\infty$ if $s = -1/d$, and $t = 1/d$ if $s = \infty$. This result is proved via different methods by Rinott (1976). Furthermore, if the dimension of the support of ν is d, then it follows from Borell (1974), Theorem 2.2 that the limit measure ν is t-concave, and hence has a Lebesgue density with $s = t/(1-td)$. Here, we pursue this type of result in somewhat more detail. Our key dimensionality condition will be formulated in terms of the set $C := \{x \in \mathbb{R}^d : \liminf f_n(x) > 0\}$. We will show that if

(D1) Either $\dim(\text{csupp}(\nu)) = d$ or $\dim(C) = d$

holds, then the limiting probability measure ν admits an upper semi-continuous s-concave density on $\mathbb{R}^d$. Furthermore, if a sequence of s-concave densities $\{f_n\}$ converges weakly to some density f (in the sense that the corresponding probability measures converge weakly), then f is s-concave, and f_n converges to f in weighted L_1 metrics and uniformly on any closed set of continuity points of f. The directional derivatives of f_n also converge uniformly in all directions in a local sense.

In the following sections, we will not fully exploit the strength of the results we have obtained. The results obtained will be interesting in their own right, and careful readers will find them useful as technical support for Sections 2 and 4.

3.1. *Limit characterization via dimensionality condition.* Note that C is a convex set. For a general convex set K, we follow the convention [see Rockafellar (1997)] that $\dim K = \dim(\text{aff}(K))$, where $\text{aff}(K)$ is the affine hull of K. It is well known that the dimension of a convex set K is the maximum of the dimensions of the various simplices included in K [cf. Theorem 2.4, Rockafellar (1997)].

We first extend several results in Kim and Samworth (2015) and Cule and Samworth (2010) from the log-concave setting to our s-concave setting. The proofs will all be deferred to the Supplementary Material.

LEMMA 3.1. *Assume* (D1). *Then* $\text{csupp}(\nu) = \overline{C}$.

LEMMA 3.2. *Let $\{\nu_n\}_{n\in\mathbb{N}}$ be probability measures with upper semi-continuous s-concave densities $\{f_n\}_{n\in\mathbb{N}}$ such that $\nu_n \to \nu$ weakly as $n \to \infty$. Here, ν is a probability measure with density f. Then $f_n \to_{a.e.} f$, and f can be taken as $f = \text{cl}(\lim_n f_n)$ and hence upper semi-continuous s-concave.*

In many situations, uniform boundedness of a sequence of s-concave densities give rise to good stability and convergence property.

LEMMA 3.3. *Assume $-1/d < s < 0$. Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of s-concave densities on $\mathbb{R}^d$. If $\dim C = d$ where $C = \{\liminf_n f_n > 0\}$ as above, then* $\sup_{n\in\mathbb{N}} \|f_n\|_\infty < \infty$.

Now we state one limit characterization theorem.

THEOREM 3.4. *Assume $-1/d < s < 0$. Under either condition of* (D1), *ν is absolutely continuous with respect to λ_d, with a version of the Radon–Nikodym derivative* $\mathrm{cl}(\lim_n f_n)$, *which is an upper semi-continuous and an s-concave density on $\mathbb{R}^d$.*

3.2. *Modes of convergence.* It is shown above that the weak convergence of s-concave probability measures implies almost everywhere pointwise convergence at the density level. In many applications, we wish different/stronger types of convergence. This subsection is devoted to the study of the following two types of convergence:

1. Convergence in $\|\cdot\|_{L_1}$ metric;
2. Convergence in $\|\cdot\|_\infty$ metric.

We start by investigating convergence property in $\|\cdot\|_{L_1}$ metric.

LEMMA 3.5. *Assume $-1/d < s < 0$. Let $\nu, \nu_1, \ldots, \nu_n, \ldots$ be probability measures with upper semi-continuous s-concave densities $f, f_1, \ldots, f_n, \ldots$ such that $\nu_n \to \nu$ weakly as $n \to \infty$. Then there exists $a, b > 0$ such that $f_n(x) \vee f(x) \le (a\|x\| + b)^{1/s}$.*

Once the existence of a suitable integrable envelope function is established, we conclude naturally by dominated convergence theorem and have the following.

THEOREM 3.6. *Assume $-1/d < s < 0$. Let $\nu, \nu_1, \ldots, \nu_n, \ldots$ be probability measures with upper semi-continuous s-concave densities $f, f_1, \ldots, f_n, \ldots$ such that $\nu_n \to \nu$ weakly as $n \to \infty$. Then for $\kappa < r - d$,*

$$\lim_{n\to\infty} \int (1 + \|x\|)^\kappa \big| f_n(x) - f(x) \big| \, \mathrm{d}x = 0. \tag{3.1}$$

Next, we examine convergence of s-concave densities in $\|\cdot\|_\infty$ norm. We denote $g = f^s$, $g_n = f_n^s$ unless otherwise specified. Since we have established pointwise convergence in Lemma 3.2, classical convex analysis guarantees that the convergence is uniform over compact sets in $\mathrm{int}(\mathrm{dom}(f))$. To establish global uniform convergence result, we only need to control the tail behavior of the class of s-concave functions and the region near the boundary of f. This is accomplished via Lemmas B.1 and B.2.

THEOREM 3.7. *Let $\nu, \nu_1, \dots, \nu_n, \dots$ be probability measures with upper semi-continuous s-concave densities $f, f_1, \dots, f_n, \dots$ such that $\nu_n \to \nu$ weakly as $n \to \infty$. Then for any closed set S contained in the continuity points of f and $\kappa < r = -1/s$,*

$$\lim_{n\to\infty} \sup_{x\in S} (1 + \|x\|)^{\kappa} |f_n(x) - f(x)| = 0.$$

We note that no assumption on the index s is required here.

3.3. *Local convergence of directional derivatives.* It is known in convex analysis that if a sequence of convex functions g_n converges pointwise to g on an open convex set, then the subdifferential of g_n also "converges" to the subdifferential of g. If we further assume smoothness of g_n, then local uniform convergence of the derivatives automatically follows. See Theorems 24.5 and 25.7 in Rockafellar (1997) for precise statements. Here, we pursue this issue at the level of transformed densities.

THEOREM 3.8. *Let $\nu, \nu_1, \dots, \nu_n, \dots$ be probability measures with upper semi-continuous s-concave densities $f, f_1, \dots, f_n, \dots$ such that $\nu_n \to \nu$ weakly as $n \to \infty$. Let $\mathcal{D}_f := \{x \in \mathrm{int}(\mathrm{dom}(f)) : f$ is differentiable at $x\}$, and $T \subset \mathrm{int}(\mathcal{D}_f)$ be any compact set. Then*

$$\lim_{n\to\infty} \sup_{x\in T, \|\xi\|_2=1} |\nabla_\xi f_n(x) - \nabla_\xi f(x)| = 0.$$

4. Limiting distribution theory of the divergence estimator. In this section, we establish local asymptotic distribution theory of the divergence estimator $\hat{f}_n$ at a fixed point $x_0 \in \mathbb{R}$. Limit distribution theory in shape-constrained estimation was pioneered for monotone density and regression estimators by Prakasa Rao (1969), Brunk (1970), Wright (1981) and Groeneboom (1985). Groeneboom, Jongbloed and Wellner (2001) established pointwise limit theory for the MLEs and LSEs of a convex decreasing density, and also treated pointwise limit theory estimation of a convex regression function. Balabdaoui, Rufibach and Wellner (2009) established pointwise limit theorems for the MLEs of log-concave densities on $\mathbb{R}$. On the other hand, for nonparametric estimation of s-concave densities, asymptotic theory beyond the Hellinger consistency results for the MLEs established by Seregin and Wellner (2010) has been nonexistent. Doss and Wellner (2016) have shown in the case of $d = 1$ that the MLEs have Hellinger convergence rates of order $O_p(n^{-2/5})$ for each $s \in (-1, \infty)$ (which includes the log-concave case $s = 0$). However, due at least in part to the lack of explicit characterizations of the MLE for s-concave classes, no results concerning limiting distributions of the MLE at fixed points are currently available. In the remainder of this section, we formulate results of this type for the Rényi divergence estimators. These results are comparable to the pointwise limit distribution results for the MLEs of log-concave densities obtained by Balabdaoui, Rufibach and Wellner (2009).

In the following, we will see how natural and strong characterizations developed in Section 2 help us to understand the limit behavior of the Rényi divergence estimator at a fixed point. For this purpose, we assume the true density $f_0 = g_0^{-r}$ satisfies the following:

(A1) $g_0 \in \mathcal{G}$ and f_0 is an s-concave density on $\mathbb{R}$, where $-1 < s < 0$;
(A2) $f_0(x_0) > 0$;
(A3) g_0 is locally C^k around x_0 for some $k \geq 2$.
(A4) Let $k := \max\{k \in \mathbb{N} : k \geq 2, g_0^{(j)}(x_0) = 0, \text{ for all } 2 \leq j \leq k-1, g_0^{(k)}(x_0) \neq 0\}$, and $k = 2$ if the above set is empty. Assume $g_0^{(k)}$ is continuous around x_0.

4.1. *Limit distribution theory.* Before we state the main results concerning the limit distribution theory for the Rényi divergence estimator, let us sketch the route by which the theory is developed. We first denote $\hat{F}_n(x) := \int_{-\infty}^x \hat{f}_n(t)\,\mathrm{d}t$, $\hat{H}_n(x) := \int_{-\infty}^x \hat{F}_n(t)\,\mathrm{d}t$ and $\mathbb{H}_n(x) := \int_{-\infty}^x \mathbb{F}_n(t)\,\mathrm{d}t$. We also denote $r_n := n^{(k+2)/(2k+1)}$ and $\mathbf{l}_{n,x_0} = [x_0, x_0 + n^{-1/(2k+1)}t]$. Due to the form of the characterizations obtained in Theorem 2.12, we define *local processes* at the level of integrated distribution functions as follows:

$$\mathbb{Y}_n^{\mathrm{loc}}(t) := r_n \int_{\mathbf{l}_{n,x_0}} \left(\mathbb{F}_n(v) - \mathbb{F}_n(x_0) - \int_{x_0}^{v} \left(\sum_{j=0}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u - x_0)^j \right) \mathrm{d}u \right) \mathrm{d}v;$$

$$\mathbb{H}_n^{\mathrm{loc}}(t) := r_n \int_{\mathbf{l}_{n,x_0}} \left(\hat{F}_n(v) - \hat{F}(x_0) - \int_{x_0}^{v} \left(\sum_{j=0}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u - x_0)^j \right) \mathrm{d}u \right) \mathrm{d}v + \hat{A}_n t + \hat{B}_n,$$

where $\hat{A}_n := n^{(k+1)/(2k+1)}(\hat{F}_n(x_0) - \mathbb{F}_n(x_0))$ and $\hat{B}_n := n^{(k+2)/(2k+1)}(\hat{H}_n(x_0) - \mathbb{H}_n(x_0))$ are defined so that $\mathbb{Y}_n^{\mathrm{loc}}(\cdot) \geq \mathbb{H}_n^{\mathrm{loc}}(\cdot)$ by virtue of Theorem 2.12. Since we wish to derive asymptotic theory at the level of the underlying convex function, we modify the processes by

$$\begin{aligned} \mathbb{Y}_n^{\mathrm{locmod}}(t) &:= \frac{\mathbb{Y}_n^{\mathrm{loc}}(t)}{f_0(x_0)} - r_n \int_{\mathbf{l}_{n,x_0}} \int_{x_0}^{v} \hat{\Psi}_{k,n,2}(u)\,\mathrm{d}u\,\mathrm{d}v, \\ \mathbb{H}_n^{\mathrm{locmod}}(t) &:= \frac{\mathbb{H}_n^{\mathrm{loc}}(t)}{f_0(x_0)} - r_n \int_{\mathbf{l}_{n,x_0}} \int_{x_0}^{v} \hat{\Psi}_{k,n,2}(u)\,\mathrm{d}u\,\mathrm{d}v, \end{aligned} \tag{4.1}$$

where

$$
\hat{\Psi}_{k,n,2}(u) = \frac{1}{f_0(x_0)}\left(\hat{f}_n(u) - \sum_{j=0}^{k-1}\frac{f_0^{(j)}(x_0)}{j!}(u-x_0)^j\right) \tag{4.2}
$$
$$
+ \frac{r}{g_0(x_0)}\big(\hat{g}_n(u) - g_0(x_0) - g_0'(x_0)(u-x_0)\big).
$$

A direct calculation reveals that with $r = -1/s > 0$,

$$
\mathbb{H}_n^{\text{locmod}}(t) = \frac{-r \cdot r_n}{g_0(x_0)} \int_{\mathbf{I}_{n,x_0}} \int_{x_0}^{v} \big(\hat{g}_n(u) - g_0(x_0) - (u-x_0)g_0'(x_0)\big)\,\mathrm{d}u\,\mathrm{d}v
$$
$$
+ \frac{\hat{A}_n t + \hat{B}_n}{f_0(x_0)},
$$

and hence

$$
n^{k/(2k+1)}\big(\hat{g}_n(x_0 + s_n t) - g_0(x_0) - s_n t g_0'(x_0)\big) = \frac{g_0(x_0)}{-r}\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathbb{H}_n^{\text{locmod}}(t), \tag{4.3}
$$
$$
n^{(k-1)/(2k+1)}\big(\hat{g}_n'(x_0 + s_n t) - g_0'(x_0)\big) = \frac{g_0(x_0)}{-r}\frac{\mathrm{d}^3}{\mathrm{d}t^3}\mathbb{H}_n^{\text{locmod}}(t).
$$

It is clear from (4.1) that the order relationship $\mathbb{Y}_n^{\text{locmod}}(\cdot) \geq \mathbb{H}_n^{\text{locmod}}(\cdot)$ is still valid for the modified processes. Now by tightness arguments, the limit process $\mathbb{H}$ of $\mathbb{H}_n^{\text{locmod}}$, including its derivatives, exists uniquely, giving us the possibility of taking the limit in (4.3) as $n \to \infty$. Finally, we relate $\mathbb{H}$ to the canonical process H_k defined in Theorem 4.1 by looking at their respective "envelope" functions $\mathbb{Y}$ and Y_k, where $\mathbb{Y}$ denotes the limit process of $\mathbb{Y}_n^{\text{locmod}}$ and $Y_k(t) = \int_0^t W(s)\,\mathrm{d}s - t^{k+2}$. Careful calculation of the limit of $\mathbb{Y}_n^{\text{loc}}$ and $\hat{\Psi}_{k,n,2}$ reveals that

$$
\mathbb{Y}_n^{\text{locmod}}(t) \to_d \frac{1}{\sqrt{f_0(x_0)}}\int_0^t W(s)\,\mathrm{d}s - \frac{r g_0^{(k)}(x_0)}{g_0(x_0)(k+2)!}t^{k+2}.
$$

Now by the scaling property of Brownian motion, $W(at) =_d \sqrt{a}W(t)$, we get the following theorem.

THEOREM 4.1. *Under assumptions* (A1)–(A4), *we have*

$$
\begin{pmatrix} n^{k/(2k+1)}\big(\hat{g}_n(x_0) - g_0(x_0)\big) \\ n^{(k-1)/(2k+1)}\big(\hat{g}_n'(x_0) - g_0'(x_0)\big) \end{pmatrix} \tag{4.4}
$$
$$
\to_d \begin{pmatrix} -\left(\dfrac{g_0^{2k}(x_0) g_0^{(k)}(x_0)}{r^{2k} f_0(x_0)^k (k+2)!}\right)^{1/(2k+1)} H_k^{(2)}(0) \\ -\left(\dfrac{g_0^{2k-2}(x_0)[g_0^{(k)}(x_0)]^3}{r^{2k-2} f_0(x_0)^{k-1}[(k+2)!]^3}\right)^{1/(2k+1)} H_k^{(3)}(0) \end{pmatrix},
$$

and

$$
(4.5)\qquad \begin{pmatrix} n^{k/(2k+1)}\big(\hat{f}_n(x_0)-f_0(x_0)\big) \\ n^{(k-1)/(2k+1)}\big(\hat{f}_n'(x_0)-f_0'(x_0)\big) \end{pmatrix} \\ \to_d \begin{pmatrix} \Big(\dfrac{rf_0(x_0)^{k+1}g_0^{(k)}(x_0)}{g_0(x_0)(k+2)!}\Big)^{1/(2k+1)} H_k^{(2)}(0) \\ \Big(\dfrac{r^3 f_0(x_0)^{k+2}(g_0^{(k)}(x_0))^3}{g_0(x_0)^3[(k+2)!]^3}\Big)^{1/(2k+1)} H_k^{(3)}(0) \end{pmatrix},
$$

where H_k is the unique lower envelope of the process Y_k satisfying:

1. $H_k(t) \le Y_k(t)$ *for all* $t \in \mathbb{R}$;
2. $H_k^{(2)}$ *is concave*;
3. $H_k(t) = Y_k(t)$ *if the slope of* $H_k^{(2)}$ *decreases strictly at* t.

REMARK 4.2. We note that the minus sign appearing in (4.4) is due to the convexity of $\hat{g}_n$, g_0 and the concavity of the limit process $H_k^{(2)}(0)$. The dependence of the constant appearing in the limit is optimal in view of Theorem 2.23 in Seregin and Wellner (2010).

REMARK 4.3. Assume $-1 < s < 0$ and $k = 2$. Let $f_0 = \exp(\varphi_0)$ be a log-concave density where $\varphi_0 : \mathbb{R} \to \mathbb{R}$ is the underlying concave function. Then f_0 is also s-concave. Let $g_s := f_0^{-1/r} = \exp(-\varphi_0/r)$ be the underlying convex function when f_0 is viewed as an s-concave density. Then direct calculation yields that

$$
g_s^{(2)}(x_0) = \frac{1}{r^2} g_s(x_0)\big(\varphi_0'(x_0)^2 - r\varphi_0''(x_0)\big).
$$

Hence, the constant before $H_k^{(2)}(0)$ appearing in (4.5) becomes

$$
\Big(\frac{f_0(x_0)^3\varphi_0'(x_0)^2}{4!r} + \frac{f_0(x_0)^3|\varphi_0''(x_0)|}{4!}\Big)^{1/5}.
$$

Note that the second term in the above display is exactly the constant involved in the limiting distribution when $f_0(x_0)$ is estimated via the log-concave MLE; see (2.2), page 1305 in Balabdaoui, Rufibach and Wellner (2009). The first term is nonnegative, and hence illustrates the price we need to pay by estimating a true log-concave density via the Rényi divergence estimator over a larger class of s-concave densities. We also note that the additional term vanishes as $r \to \infty$, or equivalently $s \nearrow 0$.

4.2. *Estimation of the mode*. We consider the estimation of the mode of an s-concave density $f(\cdot)$ defined by $M(f) := \inf\{t \in \mathbb{R} : f(t) = \sup_{u\in\mathbb{R}} f(u)\}$.

THEOREM 4.4. *Assume* (A1)–(A4) *hold. Then*

$$n^{1/(2k+1)}(\hat{m}_n - m_0) \to_d \left(\frac{g_0(m_0)^2(k+2)!^2}{r^2 f_0(m_0) g_0^{(k)}(m_0)^2}\right)^{1/(2k+1)} M\big(H_k^{(2)}\big), \tag{4.6}$$

where $\hat{m}_n = M(\hat{f}_n), m_0 = M(f_0)$.

By Theorem 2.26 in Seregin and Wellner (2010), the dependence of the constant on local smoothness is optimal when $k = 2$. Here, we show that this dependence is also optimal for $k > 2$.

Consider a class of densities $\mathcal{P}$ dominated by the canonical Lebesgue measure on $\mathbb{R}^d$. Let $T : \mathcal{P} \to \mathbb{R}$ be any functional. For an increasing convex loss function $l(\cdot)$ on $\mathbb{R}_+$, we define the *minimax risk* as

$$R_l(n; T, \mathcal{P}) := \inf_{t_n} \sup_{p \in \mathcal{P}} \mathbb{E}_{p^{\times n}} l\big(\big|t_n(X_1, \dots, X_n) - T(p)\big|\big), \tag{4.7}$$

where the infimum is taken over all possible estimators of $T(p)$ based on $X_1, \dots, X_n$. Our basic method of deriving minimax lower bound based on the following work in Jongbloed (2000).

THEOREM 4.5 [Theorem 1 Jongbloed (2000)]. *Let* $\{p_n\}$ *be a sequence of densities in* $\mathcal{P}$ *such that* $\limsup_{n\to\infty} nh^2(p_n, p) \le \tau^2$ *for some density* $p \in \mathcal{P}$. *Then*

$$\liminf_{n\to\infty} \frac{R_l(n; T, \{p, p_n\})}{l(\exp(-2\tau^2)/4 \cdot |T(p_n) - T(p)|)} \ge 1. \tag{4.8}$$

For fixed $g \in \mathcal{G}$ and $f := g^{1/s} = g^{-r}$, let $m_0 := M(g)$ be the mode of g. Consider a class of local perturbations of g: For every $\varepsilon > 0$, define

$$\tilde{g}_\varepsilon(x) = \begin{cases} g(m_0 - \varepsilon c_\varepsilon) + (x - m_0 + \varepsilon c_\varepsilon) g'(m_0 - \varepsilon c_\varepsilon), \\ \qquad x \in [m_0 - \varepsilon c_\varepsilon, m_0 - \varepsilon), \\ g(m_0 + \varepsilon) + (x - m_0 - \varepsilon) g'(m_0 + \varepsilon), \\ \qquad x \in [m_0 - \varepsilon, m_0 + \varepsilon), \\ g(x), \qquad \text{otherwise.} \end{cases}$$

Here, c_ε is chosen so that g_ε is continuous at $m_0 - \varepsilon$. This construction of a perturbation class is also seen in Balabdaoui, Rufibach and Wellner (2009), Groeneboom, Jongbloed and Wellner (2001). By Taylor expansion at $m_0 - \varepsilon$, we can easily see $c_\varepsilon = 3 + o(1)$ as $\varepsilon \to 0$. Since $\tilde{f}_\varepsilon := \tilde{g}_\varepsilon^{-r}$ is not a density, we normalize it by $f_\varepsilon(x) := \frac{\tilde{f}_\varepsilon(x)}{\int_{\mathbb{R}} \tilde{f}_\varepsilon(y)\,dy}$. Now f_ε is s-concave for each $\varepsilon > 0$ with mode $m_0 - \varepsilon$.

The following result follows from direct calculation. For a proof, we refer to the Supplemental Material.

LEMMA 4.6. *Assume* (A1)–(A4). *Then*

$$h^2(f_\varepsilon, f) = \zeta_k \frac{r^2 f(m_0)(g^{(k)}(m_0))^2}{g(m_0)^2}\varepsilon^{2k+1} + o(\varepsilon^{2k+1}),$$

where

$$\begin{aligned}\zeta_k = {} & \frac{1}{108(k!)^2(k+1)(k+2)(2k+1)}\big[-4\cdot 3^{k+2}(2k+1)(3^{k+2}+k^2+k-3) \\ & + (k+1)(k+2)(27(3^{2k+1}-1) + 2\cdot 3^{2k}(2k+1)(2k(2k-9)+27))\big] \\ & + \frac{2k^2(2k^2+1)}{3(k!)^2(k+1)(2k+1)}.\end{aligned}$$

THEOREM 4.7. *For an s-concave density f_0, let $\mathcal{SC}_{n,\tau}(f_0)$ be defined by*

$$\mathcal{SC}_{n,\tau}(f_0) := \left\{ f : s\text{-concave density}, h^2(f, f_0) \le \frac{\tau^2}{n} \right\}.$$

Let $m_0 = M(f_0)$ be the mode of f_0. Suppose (A1)–(A4) *hold. Then*

$$\begin{aligned}&\sup_{\tau>0} \liminf_{n\to\infty} n^{1/(2k+1)} \inf_{t_n} \sup_{f\in\mathcal{SC}_{n,\tau}} \mathbb{E}_f |T_n - M(f)| \\ &\qquad \ge \rho_k \left(\frac{g_0(m_0)^2}{r^2 f_0(m_0) g_0^{(k)}(m_0)^2} \right)^{1/(2k+1)},\end{aligned}$$

where $\rho_k = (2(2k+1)\zeta_k)^{-1/(2k+1)}/4$.

PROOF. Take $l(x) = |x|$. Let $\varepsilon = cn^{-1/(2k+1)}$, and let $\gamma = \frac{r^2 f(m_0)(g^{(k)}(m_0))^2}{g(m_0)^2}$, $f_n := f_{cn^{-1/(2k+1)}}$. Then $\limsup_{n\to\infty} nh^2(f_n, f) = \zeta_k \gamma c^{2k+1}$. Applying Theorem 4.5, we find that

$$\liminf_{n\to\infty} n^{1/(2k+1)} R_l(n; T, \{f, f_n\}) \ge \frac{1}{4} c \exp(-2\zeta_k \gamma c^{(2k+1)}).$$

Now we choose $c = (2(2k+1)\zeta_k\gamma)^{-1/(2k+1)}$ to conclude. □

5. Discussion. We show in this paper that the class of s-concave densities can be approximated and estimated via Rényi divergences in a robust and stable way. We also develop local asymptotic distribution theory for the divergence estimator, which suggests that the convexity constraint is the main complexity within the class of s-concave densities regardless heavy tails. In the rest of this section, we will sketch some related problems and future research directions.

5.1. *Behavior of Rényi projection for generic measures Q when $s < -1/(d+1)$.* We have considered in this paper two regions for the index s: (1) $-1/(d+1) < s < 0$ and (2) $-1/d < s \le -1/(d+1)$. In case (1), we showed that starting from a generic measure Q with the interior of its convex support non-void and a first moment, the Rényi projection through (1.3) exists and enjoys nice continuity properties that cover both on and off-the-model situations. In case (2), we showed that the Rényi projection for the empirical measure still enjoys such continuity properties when Q is a probability measure corresponding to a true s-concave density with a finite first moment.

It remains open to investigate the behavior of the Rényi projection in the region (2) for a generic measure Q. If Q does not admit a first moment, that is, $\int \|x\| \, \mathrm{d}Q(x) = \infty$, then the first term in the functional (1.3) diverges for any candidate convex function. We conjecture that the Rényi divergence projection fails to exist in this case. We do not know if the Rényi projection exists when $-1/d < s \le -1/(d+1)$ and $Q \notin \mathcal{P}_s$ but $\int \|x\| \, \mathrm{d}Q(x) < \infty$.

It should be mentioned that the MLEs for the classes $\mathcal{P}_s$ exist (for an increasingly large sample size n as $s \searrow -1/d$), and are Hellinger consistent for $-1/d < s < 0$ [cf. Seregin and Wellner (2010)]. Moreover, it is known from Doss and Wellner (2016) that the MLE does not exist for $s < -1/d$. But we do not yet know any continuity properties of the Maximum Likelihood projection "off the model". This leaves the interval $-1/d < s \le -1/(d+1)$ presently without a nicely stable nonparametric estimation procedure. See Koenker and Mizera (2010) pages 3008 and 3016 for some further discussion.

5.2. *Global rates of convergence for Rényi divergence estimators.* Classical empirical process theory relates the maximum likelihood estimators with Hellinger loss via "basic inequalities" as coined in van de Geer (2000) and van der Vaart and Wellner (1996). This reduces the problem of global rates of convergence to the study of modulus of continuity of empirical process indexed by a suitable transformation of the function class of interest. We expect that similar "basic inequalities" can be exploited to relate the Rényi divergence estimators to some divergence (not necessarily Hellinger distance). We also expect some uniformity in the rates of convergence for the Rényi divergence estimators as observed by Kim and Samworth (2015) in the case of the MLEs for log-concave densities.

5.3. *Conjectures about the global rates in higher dimensions.* It is now well understood from the work of Doss and Wellner (2016) that the MLEs for s-concave densities ($-1 < s < 0$) and log-concave densities in dimension 1 converge at rates no worse than $O_p(n^{-2/5})$ in Hellinger loss. In higher dimensions, Kim and Samworth (2015) provide an important lower bound on the bracketing entropy for a subclass of log-concave densities on the order of $O(\varepsilon^{-(d/2)\vee(d-1)})$ in Hellinger distance, and a matching upper bound up to logarithmic factors for $d \le 3$. Lack of corresponding results in discrete convex geometry precludes further upper bounds beyond $d = 3$. If a matching upper bound can be achieved for $d \geq 4$ (with possible logarithmic losses), the rates of convergence r_n^2 in squared Hellinger distances become

$$r_n^2 = O\big(n^{-1/(d-1)}\big), \qquad d \geq 4$$

(up to logarithmic factors). It is also worth mentioning that minimum contrast estimator may well be rate inefficient in higher dimensions, as observed by Birgé and Massart (1993) in another context with "trans-Donsker" class of functions. Therefore, it is also interesting to design sieved/regularized estimator to achieve the efficient rates.

5.4. *Adaptive estimation of concave-transformed class of functions*. The rates conjectured above are conservative in that they are derived from the *global* point of view. From a local perspective, adaptive estimation may be possible when the underlying function/density exhibits special structures. In fact, it is shown by Guntuboyina and Sen (2015) that in the univariate convex regression setting, if the underlying convex function is piecewise linear, then the rate of convergence for the global risk in the discrete l_2 norm adapts to nearly parametric rate $n^{-1/2}$ (up to logarithmic factors). It would be interesting to examine if same phenomenon can be observed for the MLEs/Rényi divergence estimators, and more generally for minimum contrast estimators of concave-transformed classes of functions.

6. Proofs.

In this section, we give proofs for Theorem 2.2, Theorem 2.5, Theorem 2.9 and Theorem 4.1.

PROOF OF THEOREM 2.2. We note that $L(Q) < \infty$ by Lemma 2.1. Hence, we can take a sequence $\{g_n\}_{n\in\mathbb{N}} \subset \mathcal{G}$ such that $\infty > M_0 \geq L(g_n, Q) \searrow L(Q)$ as $n \to \infty$ for some $M_0 > 0$. Now we claim that, for all $x_0 \in \operatorname{int}(\operatorname{csupp}(Q))$,

$$\sup_{n\in\mathbb{N}} g_n(x_0) < \infty. \tag{6.1}$$

Denote $\varepsilon_n \equiv \inf_{x\in\mathbb{R}^d} g_n(x)$. First, note

$$\begin{aligned} L(g_n, Q) &\geq \int g_n \,\mathrm{d}Q = \int g_n \mathbf{1}\big(g_n \leq g_n(x_0)\big)\,\mathrm{d}Q + \int g_n \mathbf{1}\big(g_n > g_n(x_0)\big)\,\mathrm{d}Q \\ &= \int \big(g_n - g_n(x_0) + g_n(x_0)\big)\mathbf{1}\big(g_n \leq g_n(x_0)\big)\,\mathrm{d}Q \\ &\quad + \int g_n \mathbf{1}\big(g_n > g_n(x_0)\big)\,\mathrm{d}Q \\ &\geq g_n(x_0) - \big(g_n(x_0) - \varepsilon_n\big) Q\big(\{g_n(\cdot) \leq g_n(x_0)\}\big). \end{aligned}$$

If $g_n(x_0) > \varepsilon_n$, then x_0 is not an interior point of the closed convex set $\{g_n \leq g_n(x_0)\}$, which implies $Q(\{g_n(\cdot) \leq g_n(x_0)\}) \leq h(Q, x)$, where $h(\cdot, \cdot)$ is defined in Lemma E.3. Hence, in this case, the above term is lower bounded by

$$L(g_n, Q) \geq g_n(x_0) - \big(g_n(x_0) - \varepsilon_n\big)h(Q, x_0) \geq g_n(x_0)\big(1 - h(Q, x_0)\big).$$

This inequality also holds for $g_n(x_0) = \varepsilon_n$, which implies that

$$g_n(x_0) \le \frac{L(g_n, Q)}{1 - h(Q, x_0)} \le \frac{M_0}{1 - h(Q, x_0)}$$

by the first statement of Lemma E.3. Thus, we verified (6.1). Now invoking Lemma E.8, and we check conditions (A1)–(A2) as follows: (A1) follows by (6.1); (A2) follows by the choice of g_n since $\sup_{n\in\mathbb{N}} L(g_n, Q) \le M_0$. By Lemma E.7, we can find a subsequence $\{g_{n(k)}\}_{k\in\mathbb{N}}$ of $\{g_n\}_{n\in\mathbb{N}}$, and a function $\tilde{g} \in \mathcal{G}$ such that $\{x \in \mathbb{R}^d : \sup_{n\in\mathbb{N}} g_n(x) < \infty\} \subset \operatorname{dom}(\tilde{g})$, and

$$\lim_{k\to\infty, x\to y} g_{n(k)}(x) = \tilde{g}(y) \qquad \text{for all } y \in \operatorname{int}\bigl(\operatorname{dom}(\tilde{g})\bigr),$$

$$\liminf_{k\to\infty, x\to y} g_{n(k)}(x) \ge \tilde{g}(y) \qquad \text{for all } y \in \mathbb{R}^d.$$

Again for simplicity, we assume that $\{g_n\}$ satisfies the above properties. We note that

$$\begin{aligned} L(Q) &= \lim_{n\to\infty}\left(\int g_n \,\mathrm{d}Q + \frac{1}{|\beta|}\int g_n^{\beta}\,\mathrm{d}x\right) \\ &\ge \liminf_{n\to\infty}\int g_n\,\mathrm{d}Q + \frac{1}{|\beta|}\liminf_{n\to\infty}\int g_n^{\beta}\,\mathrm{d}x \\ &\ge \int \tilde{g}\,\mathrm{d}Q + \frac{1}{|\beta|}\int \tilde{g}^{\beta}\,\mathrm{d}x = L(\tilde{g}, Q) \ge L(Q), \end{aligned}$$

where the third line follows from Fatou's lemma for the first term, and Fatou's lemma and the fact that the boundary of a convex set has Lebesgue measure zero for the second term [Theorem 1.1, Lang (1986)]. This establishes $L(\tilde{g}, Q) = L(Q)$, and hence $\tilde{g}$ is the desired minimizer. Since $\tilde{g} \in \mathcal{G}$ achieves its minimum, we may assume $x_0 \in \operatorname{Arg\,min}_{x\in\mathbb{R}^d} \tilde{g}(x)$. If $\tilde{g}(x_0) = 0$, since $\tilde{g}$ has domain with nonempty interior, we can choose $x_1, \ldots, x_d \in \operatorname{dom}(\tilde{g})$ such that $\{x_0, \ldots, x_d\}$ are in general position. Then by Lemma E.9 we find $L(\tilde{g}, Q) = \infty$, a contradiction. This implies $\tilde{g}$ must be bounded away from zero.

For the last statement, since $\tilde{g}$ is a minimizer of (1.3), and the fact that $\tilde{g}$ is bounded away from zero, then $L(\tilde{g} + c, Q)$ is well-defined for all $|c| \le \delta$ with small $\delta > 0$, and we must necessarily have $\frac{\mathrm{d}}{\mathrm{d}c} L(\tilde{g} + c, Q)|_{c=0} = 0$. On the other hand it is easy to calculate that $\frac{\mathrm{d}}{\mathrm{d}c} L(\tilde{g} + c, Q) = 1 - \int (\tilde{g}(x) + c)^{\beta-1}\,\mathrm{d}x$. This yields the desired result by noting $\beta - 1 = 1/s$. □

PROOF OF THEOREM 2.5. To show (2.1), we use Skorohod's theorem: since $Q_n \to_d Q$, there exist random vectors $X_n \sim Q_n$ and $X \sim Q$ defined on a common probability space $(\Omega, \mathcal{B}, \mathbb{P})$ satisfying $X_n \to_{a.s.} X$. Then by Fatou's lemma, we have $\int \|x\|\,\mathrm{d}Q = \mathbb{E}[\|X\|] \le \liminf_{n\to\infty} \mathbb{E}[\|X_n\|] = \liminf_{n\to\infty} \int \|x\|\,\mathrm{d}Q_n$.

Assume (2.2). We first claim that

$$\limsup_{n\to\infty} L(Q_n) \le L(g, Q) = L(Q). \tag{6.2}$$

Let $g_n(\cdot), g(\cdot)$ be defined as in the statement of the theorem. Note that $\limsup_{n\to\infty} L(g_n, Q_n) \le \lim_{n\to\infty} L(g^{(\varepsilon)}, Q_n) = L(g^{(\varepsilon)}, Q)$. Here, $g^{(\varepsilon)}$ is the Lipschitz approximation of g defined in Lemma E.2, and the last equality follows from the moment convergence condition (2.2) by rewriting $g^{(\varepsilon)}(x) = \frac{g^{(\varepsilon)}(x)}{1+\|x\|}(1 + \|x\|)$, and note the Lipschitz condition on $g^{(\varepsilon)}$ implies boundedness of $\frac{g^{(\varepsilon)}(x)}{1+\|x\|}$. By construction of $\{g^{(\varepsilon)}\}_{\varepsilon>0}$, we know that if x_0 is a minimizer of g, then it is also a minimizer of $g^{(\varepsilon)}$. This implies that the function class $\{g^{(\varepsilon)}\}_{\varepsilon>0}$ is bounded away from zero since g is bounded away from zero by Theorem 2.2, that is, $\inf_{x\in\mathbb{R}^d} g^{(\varepsilon)}(x) \ge \varepsilon_0$ holds for all $\varepsilon > 0$ with some $\varepsilon_0 > 0$. Now let $\varepsilon \searrow 0$, in view of Lemma E.2, by the monotone convergence theorem applied to $g^{(\varepsilon)}$ and $\varepsilon_0^\beta - (g^{(\varepsilon)})^\beta$ we have verified (6.2).

Next, we claim that, for all $x_0 \in \operatorname{int}(\operatorname{dom}(Q))$,

$$\limsup_{n\to\infty} g_n(x_0) < \infty. \tag{6.3}$$

Denote $\varepsilon_n \equiv \inf_{x\in\mathbb{R}^d} g_n(x)$. Note by essentially the same argument as in the proof of Theorem 2.2, we have

$$g_n(x_0) \le \frac{L(Q_n)}{1 - h(Q_n, x_0)}.$$

By taking lim sup as $n \to \infty$, (6.3) follows by virtue of Lemma E.3 and (6.2).

Now we proceed to show (2.3) and (2.4). By invoking Lemma E.8, we can easily check that all conditions are satisfied [note we also used (6.2) here]. Thus we can find a subsequence $\{g_{n(k)}\}_{k\in\mathbb{N}}$ of $\{g_n\}_{n\in\mathbb{N}}$ with $g_{n(k)}(x) \ge a\|x\| - b$, holds for all $x \in \mathbb{R}^d$ and all $k \in \mathbb{N}$ with some $a, b > 0$. Hence, by Lemma E.7, we can find a function $\tilde{g} \in \mathcal{G}$ such that $\{x \in \mathbb{R}^d : \limsup_{k\to\infty} g_{n(k)}(x) < \infty\} \subset \operatorname{dom}(\tilde{g})$, and that

$$\lim_{k\to\infty, x\to y} g_{n(k)}(x) = \tilde{g}(y) \qquad \text{for all } y \in \operatorname{int}\bigl(\operatorname{dom}(\tilde{g})\bigr),$$

$$\liminf_{k\to\infty, x\to y} g_{n(k)}(x) \ge \tilde{g}(y) \qquad \text{for all } y \in \mathbb{R}^d.$$

Again for simplicity, we assume $\{g_n\}$ admit the above properties. Now define random variables $H_n \equiv g_n(X_n) - (a\|X_n\| - b)$. Then by the same reasoning as in the proof of Theorem 2.2, we have

$$\begin{aligned}\liminf_{n\to\infty} L(Q_n) &= \liminf_{n\to\infty}\biggl(\int g_n \,\mathrm{d}Q_n + \frac{1}{|\beta|}\int g_n^\beta \,\mathrm{d}x\biggr) \\ &\ge \liminf_{n\to\infty} \mathbb{E}\bigl[H_n + a(X_n) - b\bigr] + \frac{1}{|\beta|}\int \tilde{g}^\beta \,\mathrm{d}x\end{aligned}$$

$$\begin{aligned}&\geq \mathbb{E}\Big[\liminf_{n\to\infty} H_n\Big] + a \liminf_{n\to\infty} \int \|x\| \, \mathrm{d}Q_n - b + \frac{1}{|\beta|} \int \tilde{g}^{\beta} \, \mathrm{d}x \\ &= L(\tilde{g}, Q) + a\Big(\liminf_{n\to\infty} \int \|x\| \, \mathrm{d}Q_n - \int \|x\| \, \mathrm{d}Q\Big) \\ &\geq L(Q) + a\Big(\liminf_{n\to\infty} \int \|x\| \, \mathrm{d}Q_n - \int \|x\| \, \mathrm{d}Q\Big).\end{aligned}$$

Note the expectation is taken with respect to the probability space $(\Omega, \mathcal{B}, \mathbb{P})$ defined above. This establishes that if (2.2) holds true, then

$$\liminf_{n\to\infty} L(Q_n) \geq L(\tilde{g}, Q) \geq L(Q). \tag{6.4}$$

Conversely, if (2.2) does not hold true, then there exists a subsequence $\{Q_{n(k)}\}$ such that $\liminf_{k\to\infty} \int \|x\| \, \mathrm{d}Q_{n(k)} > \int \|x\| \, \mathrm{d}Q$. However, this means that $\liminf_{k\to\infty} L(Q_{n(k)}) > L(Q)$, which contradicts (2.3). Hence, if (2.3) holds, then (2.2) holds true. Combine (6.4) and (6.2), and by virtue of Lemma 2.3, we find $\tilde{g} \equiv g$. This completes the proof for (2.3) and (2.4).

We show (2.5). First, we claim that $\{\hat{x}_n \in \operatorname{Arg\,min}_{x\in\mathbb{R}^d} g_n(x)\}_{n\in\mathbb{N}}$ is bounded. If not, then we can find a subsequence such that $\|\hat{x}_{n(k)}\| \to \infty$ as $k \to \infty$. However, this means that $g_{n(k)}(x) \geq g_{n(k)}(\hat{x}_{n(k)}) \geq a\|\hat{x}_{n(k)}\| - b \to \infty$ as $k \to \infty$ for any x, a contradiction. Next, we claim that there exists $\varepsilon_0 > 0$ such that $\inf_{k\in\mathbb{N}} \varepsilon_{n(k)} \geq \varepsilon_0$ holds for some subsequence $\{\varepsilon_{n(k)}\}_{k\in\mathbb{N}}$ of $\{\varepsilon_n\}_{n\in\mathbb{N}}$. This can be seen as follows: Boundedness of $\{\hat{x}_n\}$ implies $\hat{x}_{n(k)} \to x^*$ as $k \to \infty$ for some subsequence $\{\hat{x}_{n(k)}\}_{k\in\mathbb{N}} \subset \{\hat{x}_n\}_{n\in\mathbb{N}}$ and some $x^* \in \mathbb{R}$. Hence, by (2.4) we have $\limsup_{k\to\infty} f_{n(k)}(\hat{x}_{n(k)}) \leq f(x^*) < \infty$, since $f(\cdot)$ is bounded. This implies that $\sup_{k\in\mathbb{N}} \|f_{n(k)}\|_\infty < \infty$, which is equivalent to the claim. As before, we will understand the notation for whole sequence as a suitable subsequence. Now we have $g_n(x) \geq (a\|x\| - b) \vee \varepsilon_0$ holds for all $x \in \mathbb{R}^d$. This gives rise to

$$f_n(x) \leq \big((a\|x\| - b) \vee \varepsilon_0\big)^{1/s} \qquad \text{for all } x \in \mathbb{R}^d. \tag{6.5}$$

Note that $-1/(d+1) < s < 0$ implies $1/s < -(d+1)$, whence we get an integrable envelope. Now a simple application of dominated convergence theorem yields the desired result (2.5), in view of the fact that the boundary of a convex set has Lebesgue measure zero [cf. Theorem 1.1 in Lang (1986)].

Finally, (2.6) and (2.7) are direct results of Theorems 3.7 and 3.8 by noting that (2.5) entails $f_n \to_d f$ (in the sense that the corresponding probability measures converge weakly). □

PROOF OF THEOREM 2.9. Denote $L(\cdot) := L(\cdot, Q)$. We first claim the following.

CLAIM. $g = \arg\min_{g\in\mathcal{G}} L(g)$ if and only if $\lim_{t\searrow 0} \frac{L(g+th)-L(g)}{t} \geq 0$, holds for all $h : \mathbb{R}^d \to \mathbb{R}$ such that there exists $t_0 > 0$ with $g + th \in \mathcal{G}$ holds for all $t \in (0, t_0)$.

To see this, we only have to show sufficiency. Now suppose g is not a minimizer of $L(\cdot)$. By Theorem 2.2, we know there exists $\hat{g} \in \mathcal{G}$ such that $\hat{g} = g(\cdot|Q)$. By convexity, we have that for any $t > 0$, $L(g + t(\hat{g} - g)) \le (1-t)L(g) + tL(\hat{g})$. This implies that if we let $h = \hat{g} - g$, and $t_0 = 1$, then

$$\frac{L(g+th) - L(g)}{t} \le \frac{1}{t}\big((1-t)L(g) + tL(\hat{g}) - L(g)\big) = -t\big(L(g) - L(\hat{g})\big),$$

and thus $\lim_{t\searrow 0} \frac{L(g+th)-L(g)}{t} \le -(L(g) - L(\hat{g})) < 0$, where the strict inequality follows from Lemma 2.3. This proves our claim. Now the theorem follows from simple calculation:

$$0 \le \lim_{t\searrow 0} \frac{1}{t}\big(L(g+th) - L(g)\big) = \int h \, \mathrm{d}Q - \int h \cdot g^{1/s} \, \mathrm{d}\lambda,$$

as desired. □

Proofs of Theorems 2.16 and all the results in Section 3 are given in the Supplementary Material, Han and Wellner (2015).

Before we prove Theorem 4.1, we will need the following tightness result.

THEOREM 6.1. *We have the following conclusions*:

1. *For fixed $K > 0$, the modified local process $\mathbb{Y}_n^{\mathrm{locmod}}(\cdot)$ converges weakly to a drifted integrated Gaussian process on $C[-K, K]$*:

$$\mathbb{Y}_n^{\mathrm{locmod}}(t) \to_d \frac{1}{\sqrt{f_0(x_0)}} \int_0^t W(s)\,\mathrm{d}s - \frac{r g_0^{(k)}(x_0)}{g_0(x_0)(k+2)!} t^{k+2},$$

   where $W(\cdot)$ is the standard two-sided Brownian motion starting from 0 *on* $\mathbb{R}$.
2. *The localized processes satisfy*

$$\mathbb{Y}_n^{\mathrm{locmod}}(t) - \mathbb{H}_n^{\mathrm{locmod}}(t) \ge 0,$$

   with equality attained for all t such that $x_0 + tn^{-1/(2k+1)} \in \mathcal{S}(\hat{g}_n)$.
3. *The sequences $\{\hat{A}_n\}$ and $\{\hat{B}_n\}$ are tight.*

The above theorem includes everything necessary in order to apply the "invelope" argument roughly indicated in Section 4.1. For a proof of this technical result, we refer the reader to the Supplementary Material. Here, we will provide proofs for our main results.

PROOF OF THEOREM 4.1. By the same tightness and uniqueness argument adopted in Groeneboom, Jongbloed and Wellner (2001), Balabdaoui and Wellner (2007) and Balabdaoui, Rufibach and Wellner (2009), we only have to find the rescaling constants. To this end, we denote $\mathbb{H}(\cdot), \mathbb{Y}(\cdot)$ the corresponding limit of

$\mathbb{H}_n^{\text{locmod}}(\cdot)$ and $\mathbb{Y}_n^{\text{locmod}}(\cdot)$ in the uniform topology on the space $C[-K, K]$, and let $\mathbb{Y}(t) = \gamma_1 Y_k(\gamma_2 t)$, where by Theorem 6.1, we know that

$$\mathbb{Y}(t) = \frac{1}{\sqrt{f_0(x_0)}} \int_0^t W(s)\,\mathrm{d}s - \frac{r g_0^{(k)}(x_0)}{g_0(x_0)(k+2)!} t^{k+2}.$$

Let $a := (f_0(x_0))^{-1/2}$ and $b := \frac{r g_0^{(k)}(x_0)}{g_0(x_0)(k+2)!}$, then by rescaling property of Brownian motion, we find that $\gamma_1 \gamma_2^{3/2} = a$, $\gamma_1 \gamma_2^{k+2} = b$. Solving for γ_1, γ_2 yields

$$\gamma_1 = a^{(2k+4)/(2k+1)} b^{-3/(2k+1)}, \qquad \gamma_2 = a^{-2/(2k+1)} b^{2/(2k+1)}. \tag{6.6}$$

On the other hand, by (4.3), let $n \to \infty$, we find that

$$\begin{pmatrix} n^{k/(2k+1)}(\hat{g}_n(x_0 + s_n t) - g_0(x_0) - s_n t g_0'(x_0)) \\ n^{(k-1)/(2k+1)}(\hat{g}_n'(x_0 + s_n t) - g_0'(x_0)) \end{pmatrix} \to_d \begin{pmatrix} \dfrac{g_0(x_0)}{-r} \dfrac{\mathrm{d}^2}{\mathrm{d}t^2} \mathbb{H}(t) \\ \dfrac{g_0(x_0)}{-r} \dfrac{\mathrm{d}^3}{\mathrm{d}t^3} \mathbb{H}(t) \end{pmatrix}. \tag{6.7}$$

It is easy to see that $\frac{\mathrm{d}^2}{\mathrm{d}t^2}\mathbb{H}(t) = \gamma_1 \gamma_2^2 \frac{\mathrm{d}^2}{\mathrm{d}t^2} H_k(\gamma_2 t)$ and $\frac{\mathrm{d}^3}{\mathrm{d}t^3}\mathbb{H}(t) = \gamma_1 \gamma_2^3 \frac{\mathrm{d}^3}{\mathrm{d}t^3} H_k(\gamma_2 t)$. Now by substitution in (6.6) we get the conclusion by direct calculation and the delta method. □

Acknowledgements. The authors owe thanks to Charles Doss, Roger Koenker and Richard Samworth, as well as two referees and an Associate Editor for helpful comments, suggestions and minor corrections.

SUPPLEMENTARY MATERIAL

Supplement to "Approximation and estimation of s-concave densities via Rényi divergences" (DOI: 10.1214/15-AOS1408SUPP; .pdf). In the supplement Han and Wellner (2015), we provide details of the omitted proofs for Sections 2, 3, 4 and 6 and some auxiliary results from convex analysis used in the main paper.

REFERENCES

AVRIEL, M. (1972). r-convex functions. *Math. Program.* **2** 309–323. MR0301151

BALABDAOUI, F., RUFIBACH, K. and WELLNER, J. A. (2009). Limit distribution theory for maximum likelihood estimation of a log-concave density. *Ann. Statist.* **37** 1299–1331. MR2509075

BALABDAOUI, F. and WELLNER, J. A. (2007). Estimation of a k-monotone density: Limit distribution theory and the spline connection. *Ann. Statist.* **35** 2536–2564. MR2382657

BASU, A., HARRIS, I. R., HJORT, N. L. and JONES, M. C. (1998). Robust and efficient estimation by minimising a density power divergence. *Biometrika* **85** 549–559. MR1665873

BIRGÉ, L. and MASSART, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Related Fields* **97** 113–150. MR1240719

BORELL, C. (1974). Convex measures on locally convex spaces. *Ark. Mat.* **12** 239–252. MR0388475

BORELL, C. (1975). Convex set functions in d-space. *Period. Math. Hungar.* **6** 111–136. MR0404559

BRASCAMP, H. J. and LIEB, E. H. (1976). On extensions of the Brunn–Minkowski and Prékopa–Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Funct. Anal.* **22** 366–389. MR0450480

BRUNK, H. D. (1970). Estimation of isotonic regression. In *Nonparametric Techniques in Statistical Inference* (*Proc. Sympos.*, *Indiana Univ.*, *Bloomington*, *Ind.*, 1969) 177–197. Cambridge Univ. Press, London. MR0277070

CULE, M. and SAMWORTH, R. (2010). Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density. *Electron. J. Stat.* **4** 254–270. MR2645484

CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multidimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 545–607. MR2758237

DAS GUPTA, S. (1976). S-unimodal function: Related inequalities and statistical applications. *Sankhyā Ser. B* **38** 301–314. MR0651042

DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality*, *Convexity*, *and Applications*. Academic Press, Boston, MA. MR0954608

DOSS, C. and WELLNER, J. A. (2016). Global rates of convergence of the MLEs of log-concave and s-concave densities. *Ann. Statist.* **44** 954–981.

DÜMBGEN, L. and RUFIBACH, K. (2009). Maximum likelihood estimation of a log-concave density and its distribution function: Basic properties and uniform consistency. *Bernoulli* **15** 40–68. MR2546798

DÜMBGEN, L., SAMWORTH, R. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.* **39** 702–730. MR2816336

GROENEBOOM, P. (1985). Estimating a monotone density. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer*, *Vol. II* (*Berkeley*, *Calif.*, 1983). *Wadsworth Statist./Probab. Ser.* 539–555. Wadsworth, Belmont, CA. MR0822052

GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: Characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. MR1891742

GUNTUBOYINA, A. and SEN, B. (2015). Global risk bounds and adaptation in univariate convex regression. *Probab. Theory Related Fields* **163** 379–411. MR3405621

HAN, Q. and WELLNER, J. A. (2015). Supplement to "Approximation and estimation of s-concave densities via Rényi divergences." DOI:10.1214/15-AOS1408SUPP.

JONGBLOED, G. (2000). Minimax lower bounds and moduli of continuity. *Statist. Probab. Lett.* **50** 279–284. MR1792307

KIM, A. K. and SAMWORTH, R. J. (2015). Global rates of convergence in log-concave density estimation. Preprint. Available at arXiv:1404.2298v2.

KOENKER, R. and MIZERA, I. (2010). Quasi-concave density estimation. *Ann. Statist.* **38** 2998–3027. MR2722462

KOENKER, R. and MIZERA, I. (2014). Convex optimization in R. *J. Stat. Softw.* **60**(5) 1–23.

LANG, R. (1986). A note on the measurability of convex sets. *Arch. Math.* (*Basel*) **47** 90–92. MR0855142

MOSEK APS DENMARK (2011). The MOSEK Optimization Tools Manual, Version 6.0. Available at http://www.mosek.com.

PAL, J. K., WOODROOFE, M. and MEYER, M. (2007). Estimating a Polya frequency function$_2$. In *Complex Datasets and Inverse Problems*. *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **54** 239–249. IMS, Beachwood, OH. MR2459192

PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. MR0267677

RINOTT, Y. (1976). On convexity of measures. *Ann. Probab.* **4** 1020–1026. MR0428540

ROCKAFELLAR, R. T. (1971). Integrals which are convex functionals. II. *Pacific J. Math.* **39** 439–469. MR0310612

ROCKAFELLAR, R. T. (1997). *Convex Analysis*. *Princeton Landmarks in Mathematics*. Princeton Univ. Press, Princeton, NJ. MR1451876

SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.* **38** 3751–3781. MR2766867

UHRIN, B. (1984). Some remarks about the convolution of unimodal functions. *Ann. Probab.* **12** 640–645. MR0735860

VAN DE GEER, S. A. (2000). *Applications of Empirical Process Theory*. *Cambridge Series in Statistical and Probabilistic Mathematics* **6**. Cambridge Univ. Press, Cambridge. MR1739079

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York. MR1385671

WALTHER, G. (2002). Detecting the presence of mixing with multiscale maximum likelihood. *J. Amer. Statist. Assoc.* **97** 508–513. MR1941467

WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. MR0606630

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
BOX 354322
SEATTLE, WASHINGTON 98195-4322
USA
E-MAIL: royhan@uw.edu
jaw@stat.washington.edu
URL: http://www.stat.washington.edu/jaw/