

# APPROXIMATION AND ESTIMATION OF $S$ -CONCAVE DENSITIES VIA RÉNYI DIVERGENCES

BY QIYANG HAN AND JON A. WELLNER\*

*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 [21] if and only if  $Q$  admits full-dimensional support and a first moment. We also show continuity of the divergence functional in  $Q$ : if  $Q_n \rightarrow Q$  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

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

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

---

\*Supported in part by NSF Grant DMS-1104832 and NI-AID grant R01 AI029168

*MSC 2010 subject classifications:* Primary 62G07, 62H12; secondary 62G05, 62G20

*Keywords and phrases:*  $s$ -concavity, consistency, projection, asymptotic distribution, mode estimation, nonparametric estimation, shape constraints

[12] 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  $\varphi$  if  $s > 0$ ,  $p = \exp(\varphi)$  for some concave  $\varphi$  if  $s = 0$ , and  $p = \varphi_+^{1/s}$  for some convex  $\varphi$  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  $\mathbb{R}^d$ . Nonparametric maximum likelihood estimator (MLE) of log-concave densities was studied in the univariate setting by [15, 25], and multivariate setting [10, 9], the limiting distribution theory at fixed points was studied in [1], and rate results in [13, 20]. [16] 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  $s \rightarrow -\infty$ . In fact,  $t$ -distributions with  $\nu$  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 (MLE's in the following) for general  $s$ -concave densities in [32] 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). [21] proposed an alternative to maximum likelihood based on generalized Rényi entropies. Similar procedures were also proposed in parametric settings by [3] using a family of discrepancy measures. In our setting of  $s$ -concave densities with  $s < 0$ , the methods of [21] can be formulated as follows.

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

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

where  $\mathcal{G}(\underline{X})$  denotes all non-negative closed convex functions supported on

the convex set  $\text{conv}(\underline{X})$ ,  $\mathbb{Q}_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}$  the empirical measure and  $\beta = 1 + 1/s < 0$ . As is shown by [21], the associated dual problem  $(\mathcal{D})$  is

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

where  $\mathcal{G}(\underline{X})^\circ \equiv \{G \in \mathcal{C}^*(\underline{X}) \mid \int g dG \leq 0, \text{ for all } g \in \mathcal{G}(\underline{X})\}$  is the polar cone of  $\mathcal{G}(\underline{X})$ , and  $\alpha$  is the conjugate index of  $\beta$ , i.e.  $1/\alpha + 1/\beta = 1$ . Here  $\mathcal{C}^*(\underline{X})$ , the space of signed Radon measures on  $\text{conv}(\underline{X})$ , is the topological dual of  $\mathcal{C}(\underline{X})$ , the space of continuous functions on  $\text{conv}(\underline{X})$ . 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  $f = \frac{d(\mathbb{Q}_n - G)}{dy}$  can be derived from the dual computation of  $L(\cdot, \mathbb{Q}_n)$ :

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

Here we used the notation  $\langle G, g \rangle := \int g dG$ ,  $\psi_s(\cdot) := (\cdot)^\beta / |\beta|$  and  $\Psi_s$  is the functional defined by  $\Psi_s(g) := \int \psi_s(g(x)) dx$  for clarity. Now the dual form (1.2) follows by the well known fact (e.g. [28] Corollary 4A) that the form of the above dual functional is given by

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

For the primal problem  $(\mathcal{P})$  and the dual problem  $(\mathcal{D})$ , [21] proved the following results:

**THEOREM 1.1** (Theorem 4.1, [21]).  *$(\mathcal{P})$  admits a unique solution  $g_n^*$  if  $\text{int}(\text{conv}(\underline{X})) \neq \emptyset$ , where  $g_n^*$  is a polyhedral convex function supported on  $\text{conv}(\underline{X})$ .*

**THEOREM 1.2** (Theorem 3.1, [21]). *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

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

where

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

and  $\mathcal{G}$  denotes the class of all (non-negative) closed convex functions with non-empty interior, which are coercive in the sense that  $g(x) \rightarrow \infty$ , as  $\|x\| \rightarrow \infty$ . [21] show that Fisher consistency holds at the population level: Suppose  $Q(A) := \int_A f_0 \, d\lambda$  is defined for some  $f_0 = g_0^{1/s}$  where  $g_0 \in \mathcal{G}$ ; then  $g_0$  is an optimal solution for  $(\mathcal{P}_Q)$ .

[21] 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 [22]. Koenker's package depends in turn on the MOSEK implementation of [24]; see Appendix B of [21] for further details. On the other hand, in the special case  $s = 0$ , computation of the MLE's of log-concave densities has been implemented in the R package LogConcDEAD developed in [10] in arbitrary dimensions. However, expensive search for the proper triangulation of the support  $\text{conv}(\underline{X})$  renders computation difficult in high dimensions.

In this paper, we show that the estimation procedure proposed by [21] 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 [15, 9] and [16]. 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 [32], 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 MLE's in the log-concave setting obtained by [1]. This local point of view also underlines the results on global rates of convergence considered in [13], 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. For clarity and simplicity of the presentation, proofs of the main theorems are deferred to Section 5. Proofs of a technical nature and some auxiliary results are presented in the Appendix.

*1.2. Notation.* In this paper, we denote the canonical Lebesgue measure on  $\mathbb{R}^d$  by  $\lambda$  or  $\lambda_d$  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  $\delta$  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  $\text{csupp}(Q)$  for the convex support of a measure  $Q$  defined on  $\mathbb{R}^d$ , i.e.

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

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

We write  $f_n \rightarrow_d f$  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  $s$ , i.e. answers affirmatively the question: For a given probability measure  $Q$ , does the Rényi divergence estimator converge to the log-concave projection as studied in [16] 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  $\lambda_d$ .*

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  $\text{int}(\text{dom}(\tilde{g})) \neq \emptyset$ .*

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.

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*

$$(2.1) \quad \int \|x\| \, dQ \leq \liminf_{n \rightarrow \infty} \int \|x\| \, dQ_n.$$

*If we further assume that*

$$(2.2) \quad \lim_{n \rightarrow \infty} \int \|x\| \, dQ_n = \int \|x\| \, dQ,$$

*then,*

$$(2.3) \quad L(Q) = \lim_{n \rightarrow \infty} L(Q_n).$$

*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*

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

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

$$(2.5) \quad \lim_{n \rightarrow \infty} \int (1 + \|x\|)^\kappa |f_n(x) - f(x)| \, dx = 0.$$

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

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

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

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

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

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 [29] for more details.

As a direct consequence, we have:

**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 projected measure  $P_Q$ . 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*

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

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

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

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

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

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

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

**PROOF.** It is known by Varadarajan's theorem (cf. [14] Theorem 11.4.1),  $\mathbb{Q}_n$  converges weakly to  $Q$  with probability 1. Further by the strong law of large numbers (SLLN), we know that  $\int \|x\| d\mathbb{Q}_n \rightarrow_{a.s.} \int \|x\| dQ$ . This verifies all conditions required in Theorem 2.5.  $\square$

### 2.3. Characterization of the Rényi divergence projection and estimator.

We now develop characterizations for the Rényi divergence estimator, especially in dimension 1. All proofs for this subsection can be found in the Appendix. We first show a general variational characterization in the same spirit of Theorem 2.2 in [15]. This result holds for all dimensions  $d \geq 1$ .

**THEOREM 2.8.** *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*

$$(2.12) \quad \int h \cdot g^{1/s} d\lambda \leq \int h dQ,$$

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

COROLLARY 2.9. *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*

$$\int h \, dP \leq \int h \, dQ,$$

where  $P = P_Q$  is the Rényi projection of  $Q$  to  $P_Q \in \mathcal{P}_s$ .

As a direct consequence, we have

COROLLARY 2.10 (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  $\lambda_{\max}(\Sigma_P) \leq \lambda_{\max}(\Sigma_Q)$  and  $\lambda_{\min}(\Sigma_P) \leq \lambda_{\min}(\Sigma_Q)$  where  $\Sigma_Q$  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  $\mathbb{E}_P[\|X\|^l] \leq \mathbb{E}_Q[\|X\|^l]$  holds for all  $l = 1, \dots, k$ .*

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  $X_1, \dots, X_n$  i.i.d.  $Q$  on  $\mathbb{R}$ , and let  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  be the order statistics of  $X_1, \dots, X_n$ . 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  $\hat{F}_n(t) := \int_{-\infty}^t \hat{g}_n^{1/s}(x) \, dx$ . From Theorem 4.1 in [21] 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  $i = 1, \dots, n-1$ . For a continuous piecewise linear function  $h : [X_{(1)}, X_{(n)}] \rightarrow \mathbb{R}$ , define the set of knots to be

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

THEOREM 2.11. *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  $i = 1, \dots, n-1$ . Let*

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

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

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

COROLLARY 2.12. *For  $x_0 \in \mathcal{S}_n(\hat{g}_n)$ , we have*

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

Finally we give a characterization of the Rényi divergence estimator in terms of distribution function as Theorem 2.7 in [16].

**THEOREM 2.13.** *Assume  $-1/(d+1) < s < 0$  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.  $\int_{\mathbb{R}} (F - G)(t) dt = 0$ ;
2.  $\int_{-\infty}^x (F - G)(t) dt \leq 0$  for all  $x \in \mathbb{R}$  with equality when  $x \in \tilde{\mathcal{S}}(g)$ .

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

The above theorem is useful for understanding the projected  $s$ -concave density given an arbitrary probability measure  $Q$  at hand. The following example illustrates such an example, which also gives some insight concerning the boundary properties of the class of  $s$ -concave densities.

**EXAMPLE 2.14.** 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  $q_{\tau}$  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  $q_{\tau}$  onto the class of  $s$ -concave ( $0 > s > -1/\tau$ ) distribution through  $L(\cdot, q_{\tau})$  will be given by  $q_{-1/s}$ . Let  $Q_{\tau}$  be the distribution function of  $q_{\tau}(\cdot)$ , then we can calculate

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

It is easy to check by direct calculation that  $\int_{-\infty}^x (Q_r(t) - Q_{\tau}(t)) dt \leq 0$  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 2.13 are verified. Note that, in Example 2.9 [16], the log-concave approximation of 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  $\alpha, \beta$  is a conjugate pair with  $\alpha^{-1} + \beta^{-1} = 1$  where  $\beta = 1 + 1/s$ . For  $1 - 1/(d + 1) < \alpha < 1$ , let

$$F_\alpha(f) = \frac{1}{\alpha - 1} \log \int f^\alpha(x) \, dx,$$

$$F_1(f) = \int f(x) \log f(x) \, dx.$$

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

$$(2.14) \quad (\mathcal{D}_\alpha) \quad \min_f F_\alpha(f) = \frac{1}{\alpha - 1} \log \int f^\alpha(x) \, dx$$

subject to  $f = \frac{d(\mathbb{Q}_n - G)}{dy}$  for some  $G \in \mathcal{G}(\underline{X})^\circ$

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

$$(2.15) \quad (\mathcal{D}_1) \quad \min_f F_1(f) = \int f(x) \log f(x) \, dx,$$

subject to  $f = \frac{d(\mathbb{Q}_n - G)}{dy}$  for some  $G \in \mathcal{G}(\underline{X})^\circ$ .

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

**THEOREM 2.15.** *Suppose  $d = 1$ . For all  $\kappa > 0, p \geq 1$ , we have the following weighted convergence*

$$\lim_{\alpha \nearrow 1} \int (1 + \|x\|)^\kappa |f_\alpha(x) - f_1(x)|^p \, dx = 0,$$

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

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

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  $\nu_n \rightarrow_d \nu$ . From [5, 6] and [7], we know that each  $\nu_n$  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 [27]. Furthermore, if the dimension of the support of  $\nu$  is  $d$ , then it follows from [5], Theorem 2.2 that the limit measure  $\nu$  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  $\nu$  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 its 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 [29] 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, [29]).

We first extend several results in [20] and [9] from the log-concave setting to our  $s$ -concave setting. The proofs will all be deferred to the Appendix.

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  $\nu_n \rightarrow \nu$  weakly as  $n \rightarrow \infty$ . Here  $\nu$  is a probability measure with density  $f$ . Then  $f_n \rightarrow_{a.e.} f$ , 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),  $\nu$  is absolutely continuous with respect to  $\lambda_d$ , with a version of the Radon-Nikodym derivative  $\text{cl}(\lim_n f_n)$ , 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  $\nu, \nu_1, \dots, \nu_n, \dots$  be probability measures with upper semi-continuous  $s$ -concave densities  $f, f_1, \dots, f_n, \dots$  such that  $\nu_n \rightarrow \nu$  weakly as  $n \rightarrow \infty$ . Then there exists  $a, b > 0$  such that  $f_n(x) \vee f(x) \leq (a\|x\| + b)^{1/s}$ .*

Once the existence of a suitable integrable envelope function is established, we conclude naturally by dominated convergence theorem that

THEOREM 3.6. *Assume  $-1/d < s < 0$ . 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  $\nu_n \rightarrow \nu$  weakly as  $n \rightarrow \infty$ . Then for  $\kappa < r - d$ ,*

$$(3.1) \quad \lim_{n \rightarrow \infty} \int (1 + \|x\|)^\kappa |f_n(x) - f(x)| \, dx = 0.$$

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  $\text{int}(\text{dom}(f))$ . 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 5.2 and 5.3.

**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  $\nu_n \rightarrow \nu$  weakly as  $n \rightarrow \infty$ . Then for any closed set  $S$  contained in the continuity points of  $f$  and  $\kappa < r = -1/s$ ,*

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

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 ‘converge’ 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 [29] 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  $\nu_n \rightarrow \nu$  weakly as  $n \rightarrow \infty$ . Let  $\mathcal{D}_f := \{x \in \text{int}(\text{dom}(f)) : f \text{ is differentiable at } x\}$ , and  $T \subset \text{int}(\mathcal{D}_f)$  be any compact set. Then*

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

**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 [26], [8], [33] and [17]. [18] established pointwise limit theory for the MLE’s and LSE’s of a convex decreasing density, and also treated pointwise limit theory estimation of a convex regression function. [1] 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 MLE’s established by [32] has been non-existent. [13] have shown in the case of  $d = 1$  that the MLE’s 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 MLE's of log-concave densities obtained by [1].

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  $-1/2 < s < 0$ ;
- (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  $\hat{F}_n(x) := \int_{-\infty}^x \hat{f}_n(t) dt$ ,  $\hat{H}_n(x) := \int_{-\infty}^x \hat{F}_n(t) dt$  and  $\mathbb{H}_n(x) := \int_{-\infty}^x \mathbb{F}_n(t) dt$ . 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 2.11, we define *local processes* at the level of integrated distribution functions as follows:

$$\begin{aligned} \mathbb{Y}_n^{\text{loc}}(t) &:= r_n \int_{\mathbf{l}_{n,x_0}} \left( \mathbb{F}_n(v) - \mathbb{F}_n(x_0) - \int_{x_0}^v \left( \sum_{j=0}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u - x_0)^j \right) du \right) dv; \\ \mathbb{H}_n^{\text{loc}}(t) &:= r_n \int_{\mathbf{l}_{n,x_0}} \left( \hat{F}_n(v) - \hat{F}_n(x_0) - \int_{x_0}^v \left( \sum_{j=0}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u - x_0)^j \right) du \right) dv \\ &\quad + \hat{A}_n t + \hat{B}_n, \end{aligned}$$

where  $\hat{A}_n := n^{\frac{k+1}{2k+1}} (\hat{F}_n(x_0) - \mathbb{F}_n(x_0))$  and  $\hat{B}_n := n^{\frac{k+2}{2k+1}} (\hat{H}_n(x_0) - \mathbb{H}_n(x_0))$  are defined so that  $\mathbb{Y}_n^{\text{loc}}(\cdot) \geq \mathbb{H}_n^{\text{loc}}(\cdot)$  by virtue of Theorem 2.11. Since we wish to derive asymptotic theory at the level of the underlying convex function, we modify the processes by

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

where

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

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

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

and hence

$$(4.3) \quad \begin{aligned} n^{\frac{k}{2k+1}} (\hat{g}_n(x_0 + s_n t) - g_0(x_0) - s_n t g_0'(x_0)) &= \frac{g_0(x_0)}{-r} \frac{d^2}{dt^2} \mathbb{H}_n^{\text{locmod}}(t), \\ n^{\frac{k-1}{2k+1}} (\hat{g}'_n(x_0 + s_n t) - g_0'(x_0)) &= \frac{g_0(x_0)}{-r} \frac{d^3}{dt^3} \mathbb{H}_n^{\text{locmod}}(t). \end{aligned}$$

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  $n \rightarrow \infty$ . 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  $Y_k(t) = \int_0^t W(s) \, ds - t^{k+2}$ . Careful calculation of the limit of  $\mathbb{Y}_n^{\text{loc}}$  and  $\hat{\Psi}_{k,n,2}$  reveals that

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

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*

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

and

$$(4.5) \quad \begin{pmatrix} n^{\frac{k}{2k+1}} (\hat{f}_n(x_0) - f_0(x_0)) \\ n^{\frac{k-1}{2k+1}} (\hat{f}'_n(x_0) - f'_0(x_0)) \end{pmatrix} \rightarrow_d \begin{pmatrix} \left( \frac{r f_0(x_0)^{k+1} g_0^{(k)}(x_0)}{g_0(x_0)^{(k+2)!}} \right)^{1/(2k+1)} H_k^{(2)}(0) \\ \left( \frac{r^3 f_0(x_0)^{k+2} (g_0^{(k)}(x_0))^3}{g_0(x_0)^3 [(k+2)!]^3} \right)^{1/(2k+1)} H_k^{(3)}(0) \end{pmatrix},$$

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

1.  $H_k(t) \leq Y_k(t)$  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 [32].

REMARK 4.3. Assume  $-1/2 < s < 0$  and  $k = 2$ . Let  $f_0 = \exp(\varphi_0)$  be a log-concave density where  $\varphi_0 : \mathbb{R} \rightarrow \mathbb{R}$  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

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

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

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

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 [1]. The first term is non-negative 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  $r \rightarrow \infty$ , 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*

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

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

By Theorem 2.26 in [32], 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  $T : \mathcal{P} \rightarrow \mathbb{R}$  be any functional. For an increasing convex loss function  $l(\cdot)$  on  $\mathbb{R}_+$ , we define the *minimax risk* as

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

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 [19].

THEOREM 4.5 (Theorem 1 [19]). *Let  $\{p_n\}$  be a sequence of densities in  $\mathcal{P}$  such that  $\limsup_{n \rightarrow \infty} n h^2(p_n, p) \leq \tau^2$  for some density  $p \in \mathcal{P}$ . Then*

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

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  $\epsilon > 0$ , define

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

Here  $c_\epsilon$  is chosen so that  $g_\epsilon$  is continuous at  $m_0 - \epsilon$ . This construction of perturbation class is also seen in [1, 18]. By Taylor expansion at  $m_0 - \epsilon$  we can easily see  $c_\epsilon = 3 + o(1)$  as  $\epsilon \rightarrow 0$ . Since  $\tilde{f}_\epsilon := \tilde{g}_\epsilon^{-r}$  is not a density, we

normalize it by  $f_\epsilon(x) := \frac{\tilde{f}_\epsilon(x)}{\int_{\mathbb{R}} \tilde{f}_\epsilon(y) dy}$ . Now  $f_\epsilon$  is  $s$ -concave for each  $\epsilon > 0$  with mode  $m_0 - \epsilon$ .

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

LEMMA 4.6. *Assume (A1)-(A4). Then*

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

where

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

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

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

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

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

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

PROOF. Take  $l(x) = |x|$ . Let  $\epsilon = cn^{-1/(2k+1)}$ , 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  $\limsup_{n \rightarrow \infty} nh^2(f_n, f) = \zeta_k \gamma c^{2k+1}$ . Applying Theorem 4.5, we find that

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

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

**5. Proofs.** In this section, we give proofs for main theorems presented in this paper.

5.1. *Proofs in Section 2.*

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  $n \rightarrow \infty$  for some  $M_0 > 0$ . Now we claim that, for all  $x_0 \in \text{int}(\text{csupp}(Q))$ ,

$$(5.1) \quad \sup_{n \in \mathbb{N}} g_n(x_0) < \infty.$$

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

$$\begin{aligned} L(g_n, Q) &\geq \int g_n \, dQ = \int g_n \mathbf{1}(g_n \leq g_n(x_0)) \, dQ + \int g_n \mathbf{1}(g_n > g_n(x_0)) \, dQ \\ &= \int (g_n - g_n(x_0) + g_n(x_0)) \mathbf{1}(g_n \leq g_n(x_0)) \, dQ + \int g_n \mathbf{1}(g_n > g_n(x_0)) \, dQ \\ &\geq g_n(x_0) - (g_n(x_0) - \epsilon_n) Q(\{g_n(\cdot) \leq g_n(x_0)\}). \end{aligned}$$

If  $g_n(x_0) > \epsilon_n$ , then  $x_0$  is not an interior point of the closed convex set  $\{g_n \leq g_n(x_0)\}$ , which implies  $Q(\{g_n(\cdot) \leq g_n(x_0)\}) \leq h(Q, x_0)$ , where  $h(\cdot, \cdot)$  is defined in Lemma B.2. Hence, in this case, the above term is lower bounded by

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

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

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

by the first statement of Lemma B.2. Thus we verified (5.1). Now invoking Lemma B.7, and we check conditions (A1)-(A2) as follows: (A1) follows by (5.1); (A2) follows by the choice of  $g_n$  since  $\sup_{n \in \mathbb{N}} L(g_n, Q) \leq M_0$ . By Lemma B.6 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  $\{x \in \mathbb{R}^d : \sup_{n \in \mathbb{N}} g_n(x) < \infty\} \subset \text{dom}(\tilde{g})$ , and

$$\begin{aligned} \lim_{k \rightarrow \infty, x \rightarrow y} g_{n(k)}(x) &= \tilde{g}(y), \quad \text{for all } y \in \text{int}(\text{dom}(\tilde{g})), \\ \liminf_{k \rightarrow \infty, x \rightarrow y} g_{n(k)}(x) &\geq \tilde{g}(y), \quad \text{for all } y \in \mathbb{R}^d. \end{aligned}$$

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

$$\begin{aligned} L(Q) &= \lim_{n \rightarrow \infty} \left( \int g_n \, dQ + \frac{1}{|\beta|} \int g_n^\beta \, dx \right) \\ &\geq \liminf_{n \rightarrow \infty} \int g_n \, dQ + \frac{1}{|\beta|} \liminf_{n \rightarrow \infty} \int g_n^\beta \, dx \\ &\geq \int \tilde{g} \, dQ + \frac{1}{|\beta|} \int \tilde{g}^\beta \, dx = L(\tilde{g}, Q) \geq L(Q), \end{aligned}$$

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, [23]). 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  $x_0 \in \text{Arg min}_{x \in \mathbb{R}^d} \tilde{g}(x)$ . If  $\tilde{g}(x_0) = 0$ , since  $\tilde{g}$  has domain with non-empty interior, we can choose  $x_1, \dots, x_d \in \text{dom}(\tilde{g})$  such that  $\{x_0, \dots, x_d\}$  are in general position. Then by Lemma B.8 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  $|c| \leq \delta$  with small  $\delta > 0$ , and we must necessarily have  $\frac{d}{dc}L(\tilde{g} + c, Q)|_{c=0} = 0$ . On the other hand it is easy to calculate that  $\frac{d}{dc}L(\tilde{g} + c, Q) = 1 - \int (\tilde{g}(x) + c)^{\beta-1} dx$ . This yields the desired result by noting  $\beta - 1 = 1/s$ .  $\square$

PROOF OF THEOREM 2.5. To show (2.1), we use Skorohod's theorem: since  $Q_n \rightarrow_d Q$ , 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  $X_n \rightarrow_{a.s.} X$ . Then by Fatou's lemma, we have  $\int \|x\| dQ = \mathbb{E}[\|X\|] \leq \liminf_{n \rightarrow \infty} \mathbb{E}[\|X_n\|] = \liminf_{n \rightarrow \infty} \int \|x\| dQ_n$ .

Assume (2.2). We first claim that

$$(5.2) \quad \limsup_{n \rightarrow \infty} L(Q_n) \leq L(g, Q) = L(Q).$$

Let  $g_n(\cdot), g(\cdot)$  be defined as in the statement of the theorem. Note that  $\limsup_{n \rightarrow \infty} L(g_n, Q_n) \leq \lim_{n \rightarrow \infty} L(g^{(\epsilon)}, Q_n) = L(g^{(\epsilon)}, Q)$ . Here  $g^{(\epsilon)}$  is the Lipschitz approximation of  $g$  defined in Lemma B.1, and the last equality follows from the moment convergence condition (2.2) by rewriting  $g^{(\epsilon)}(x) = \frac{g^{(\epsilon)}(x)}{1+\|x\|}(1+\|x\|)$ , and note the Lipschitz condition on  $g^{(\epsilon)}$  implies boundedness of  $\frac{g^{(\epsilon)}(x)}{1+\|x\|}$ . By construction of  $\{g^{(\epsilon)}\}_{\epsilon>0}$  we know that if  $x_0$  is a minimizer of  $g$ , then it is also a minimizer of  $g^{(\epsilon)}$ . This implies that the function class  $\{g^{(\epsilon)}\}_{\epsilon>0}$  is bounded away from zero since  $g$  is bounded away from zero by Theorem 2.2, i.e.  $\inf_{x \in \mathbb{R}^d} g^{(\epsilon)}(x) \geq \epsilon_0$  holds for all  $\epsilon > 0$  with some  $\epsilon_0 > 0$ . Now let  $\epsilon \searrow 0$ , in view of Lemma B.1, by the monotone convergence theorem applied to  $g^{(\epsilon)}$  and  $\epsilon_0^\beta - (g^{(\epsilon)})^\beta$  we have verified (5.2).

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

$$(5.3) \quad \limsup_{n \rightarrow \infty} g_n(x_0) < \infty.$$

Denote  $\epsilon_n \equiv \inf_{x \in \mathbb{R}^d} g_n(x)$ . Note by essentially the same argument as in the

proof of Theorem 2.2, we have

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

By taking lim sup as  $n \rightarrow \infty$ , (5.3) follows by virtue of Lemma B.2 and (5.2).

Now we proceed to show (2.3) and (2.4). By invoking Lemma B.7, we can easily check that all conditions are satisfied (note we also used (5.2) here). Thus we can find a subsequence  $\{g_{n(k)}\}_{k \in \mathbb{N}}$  of  $\{g_n\}_{n \in \mathbb{N}}$  with  $g_{n(k)}(x) \geq a\|x\| - b$ , holds for all  $x \in \mathbb{R}^d$  and all  $k \in \mathbb{N}$  with some  $a, b > 0$ . Hence by Lemma B.6, we can find a function  $\tilde{g} \in \mathcal{G}$  such that  $\{x \in \mathbb{R}^d : \limsup_{k \rightarrow \infty} g_{n(k)}(x) < \infty\} \subset \text{dom}(\tilde{g})$ , and that

$$\begin{aligned} \lim_{k \rightarrow \infty, x \rightarrow y} g_{n(k)}(x) &= \tilde{g}(y), \quad \text{for all } y \in \text{int}(\text{dom}(\tilde{g})), \\ \liminf_{k \rightarrow \infty, x \rightarrow y} g_{n(k)}(x) &\geq \tilde{g}(y), \quad \text{for all } y \in \mathbb{R}^d. \end{aligned}$$

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

$$\begin{aligned} \liminf_{n \rightarrow \infty} L(Q_n) &= \liminf_{n \rightarrow \infty} \left( \int g_n \, dQ_n + \frac{1}{|\beta|} \int g_n^\beta \, dx \right) \\ &\geq \liminf_{n \rightarrow \infty} \mathbb{E}[H_n + a\|X_n\| - b] + \frac{1}{|\beta|} \int \tilde{g}^\beta \, dx \\ &\geq \mathbb{E}[\liminf_{n \rightarrow \infty} H_n] + a \liminf_{n \rightarrow \infty} \int \|x\| \, dQ_n - b + \frac{1}{|\beta|} \int \tilde{g}^\beta \, dx \\ &= L(\tilde{g}, Q) + a \left( \liminf_{n \rightarrow \infty} \int \|x\| \, dQ_n - \int \|x\| \, dQ \right) \\ &\geq L(Q) + a \left( \liminf_{n \rightarrow \infty} \int \|x\| \, dQ_n - \int \|x\| \, dQ \right), \end{aligned}$$

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

$$(5.4) \quad \liminf_{n \rightarrow \infty} L(Q_n) \geq L(\tilde{g}, Q) \geq L(Q).$$

Conversely, if (2.2) does not hold true, then there exists a subsequence  $\{Q_{n(k)}\}$  such that  $\liminf_{k \rightarrow \infty} \int \|x\| \, dQ_{n(k)} > \int \|x\| \, dQ$ . However, this means that  $\liminf_{k \rightarrow \infty} L(Q_{n(k)}) > L(Q)$ , which contradicts (2.3). Hence if (2.3) holds, then (2.2) holds true. Combine (5.4) and (5.2), 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  $\{\hat{x}_n \in \text{Arg min}_{x \in \mathbb{R}^d} g_n(x)\}_{n \in \mathbb{N}}$  is bounded. If not, then we can find a subsequence such that  $\|\hat{x}_{n(k)}\| \rightarrow \infty$  as  $k \rightarrow \infty$ . However this means that  $g_{n(k)}(x) \geq g_{n(k)}(\hat{x}_{n(k)}) \geq a\|\hat{x}_{n(k)}\| - b \rightarrow \infty$  as  $k \rightarrow \infty$  for any  $x$ , a contradiction. Next we claim that there exists  $\epsilon_0 > 0$  such that  $\inf_{k \in \mathbb{N}} \epsilon_{n(k)} \geq \epsilon_0$  holds for some subsequence  $\{\epsilon_{n(k)}\}_{k \in \mathbb{N}}$  of  $\{\epsilon_n\}_{n \in \mathbb{N}}$ . This can be seen as follows: Boundedness of  $\{\hat{x}_n\}$  implies  $\hat{x}_{n(k)} \rightarrow x^*$  as  $k \rightarrow \infty$  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  $\limsup_{k \rightarrow \infty} f_{n(k)}(\hat{x}_{n(k)}) \leq f(x^*) < \infty$ , 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  $g_n(x) \geq (a\|x\| - b) \vee \epsilon_0$  holds for all  $x \in \mathbb{R}^d$ . This gives rise to

$$(5.5) \quad f_n(x) \leq \left( (a\|x\| - b) \vee \epsilon_0 \right)^{1/s}, \quad \text{for all } x \in \mathbb{R}^d.$$

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 [23]).

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

PROOF OF THEOREM 2.8. Denote  $L(\cdot) := L(\cdot, Q)$ . We first claim:

**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  $h : \mathbb{R}^d \rightarrow \mathbb{R}$  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  $t > 0$ ,  $L(g + t(\hat{g} - g)) \leq (1-t)L(g) + tL(\hat{g})$ . This implies that if we let  $h = \hat{g} - g$ , and  $t_0 = 1$ , then

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

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

$$0 \leq \lim_{t \searrow 0} \frac{1}{t} \left( L(g + th) - L(g) \right) = \int h \, dQ - \int h \cdot g^{1/s} \, d\lambda,$$

as desired.  $\square$

Before we state the proof of Theorem 2.15, we will need an elementary lemma. The proof can be found in the Appendix.

LEMMA 5.1. *Fix a sequence  $0 < \alpha_n < 1$  with  $\alpha_n \nearrow 1$ . Let  $f_{\alpha_n}$  be an  $(\alpha_n - 1)$ -concave density on  $\mathbb{R}$ . Let  $g_{\alpha_n} := f_{\alpha_n}^{\alpha_n - 1}$  be the underlying convex function. Suppose  $\{g_{\alpha_n}\}$ 's are linear on  $[a, b]$  with  $\lim_{n \rightarrow \infty} f_{\alpha_n}(a) = \gamma_a \in [0, \infty]$  and  $\lim_{n \rightarrow \infty} f_{\alpha_n}(b) = \gamma_b \in [0, \infty]$ . Then for all  $x \in [a, b]$ ,*

$$(5.6) \quad f_{\alpha_n}(x) \rightarrow \exp\left(\frac{\log \gamma_b - \log \gamma_a}{b - a}(x - a) + \log \gamma_a\right)$$

where  $\exp(-\infty) := 0$  and  $\exp(\infty) := \infty$ .

PROOF OF THEOREM 2.15. In the following, the notation  $\sup_\alpha, \inf_\alpha, \lim_\alpha$  is understood as taking corresponding operation over  $\alpha$  close to 1 unless otherwise specified. We first show almost everywhere convergence by invoking Lemma B.6. To see this, for fixed  $s_0 \in (-1/2, 0)$ , denote  $g_\alpha := f_\alpha^{\alpha - 1}$  and  $g_\alpha^{(s_0)} := (f_\alpha)^{s_0}$ . Then for  $\alpha > 1 + s_0$ , the transformed function  $g_\alpha^{(s_0)}$  is convex. We need to check two conditions in order to apply Lemma B.6 as follows:

- (C1) The set  $(X_{(1)}, X_{(n)}) \subset \{\liminf_\alpha f_\alpha(x) > 0\}$ ;
- (C2) There is a uniform lower bound function  $\tilde{g}^{s_0} \in \mathcal{G}$  such that  $g_\alpha^{(s_0)} \geq \tilde{g}^{s_0}$  holds for  $\alpha$  sufficiently close to 1.

The first assertion can be checked by using the characterization Theorem 2.11. Let  $F_\alpha$  be the distribution function of  $f_\alpha$ . Then  $\int_{X_{(1)}}^t (F_\alpha - \mathbb{F}_n)(x) dx \leq 0$  with equality attained if and only if  $t \in \mathcal{S}_n(g_\alpha)$ . For  $x \in (X_{(1)}, X_{(n)})$  close enough to  $X_{(n)}$ , we claim that  $\liminf_\alpha f_\alpha(x) > 0$ . If not, we may assume without loss of generality that  $\lim_\alpha f_\alpha(x) = 0$ . We first note that there exists some  $t \in \{1, \dots, n - 1\}$  and some subsequence  $\{\alpha(\beta)\}_{\beta \in \mathbb{N}}$  with  $\alpha(\beta) \nearrow 1$  for which (1)  $X_{(t)}$  is a knot point for  $\{g_{\alpha(\beta)}\}$ , and (2)  $X_{(u)}$  is not a knot point for any  $\{g_{\alpha(\beta)}\}$  for  $u \geq t + 1$ , i.e.  $g_{\alpha(\beta)}$ 's are linear on  $[X_{(t)}, X_{(n)}]$ . We drop  $\beta$  for notational simplicity and assume without loss of generality that both limits  $\lim_\alpha f_\alpha(X_{(n)}), \lim_\alpha f_\alpha(X_{(t)})$  exist. Now Lemma 5.1 shows that  $\min\{\lim_\alpha f_\alpha(X_{(n)}), \lim_\alpha f_\alpha(X_{(t)})\} = 0$  since we have assumed  $\lim_\alpha f_\alpha(x) = 0$  for some  $x \in (X_{(t)}, X_{(n)})$ . This in turn implies that  $\lim_\alpha f_\alpha(x) = 0$  for all  $x \in (X_{(t)}, X_{(n)})$ . Now we consider the following two cases to derive a contradiction with the fact

$$(5.7) \quad \int_{X_{(t)}}^{X_{(n)}} F_\alpha(x) dx = \int_{X_{(t)}}^{X_{(n)}} \mathbb{F}_n(x) dx$$

that follows from Theorem 2.11, thereby proving  $\liminf_\alpha f_\alpha(x) > 0$  for  $x$  close enough to  $X_{(n)}$ .

**Case 1.** If  $\lim_{\alpha} f_{\alpha}(X_{(n)}) = 0$ , then the left hand side of (5.7) converges to  $X_{(n)} - X_{(t)}$  while the right hand side is no larger than  $\frac{n-1}{n}(X_{(n)} - X_{(t)})$ .

**Case 2.** If  $\lim_{\alpha} f_{\alpha}(X_{(n)}) > 0$ , then we must necessarily have  $\lim_{\alpha} f_{\alpha}(x) = 0$  for all  $x \in [X_{(1)}, X_{(n)})$  by convexity of  $g_{\alpha}$ : If  $\lim_{\alpha} f_{\alpha}(x_0) > 0$  for some  $x_0 \in [X_{(1)}, X_{(t)}]$ , then  $\lim_{\alpha} g_{\alpha}(x_0) \vee g_{\alpha}(X_{(n)}) < \infty$  while  $\lim_{\alpha} g_{\alpha}(x) = \infty$  for all  $x \in (X_{(t)}, X_{(n)})$ , which is absurd. Note that this also forces  $\lim_{\alpha} f_{\alpha}(X_{(n)}) = \infty$ , otherwise the constraint  $\int f_{\alpha} = 1$  will be invalid eventually. Now the left hand side of (5.7) converges to 0 while the right hand side is bounded from below by  $\frac{1}{n}(X_{(n)} - X_{(t)})$ .

Similarly we can show  $\liminf_{\alpha} f_{\alpha}(x) > 0$  for  $x$  close to  $X_{(1)}$ . Now (C1) follows by convexity of  $f_{\alpha}$ .

(C2) can be seen by first noting  $M := \sup_{\alpha} \|f_{\alpha}\|_{\infty} < \infty$ . This can be verified by Lemma 3.3 combined with the first assertion proved above. This implies that the class  $\{g_{\alpha}^{(s_0)}\}_{\alpha}$  has a uniform lower bound  $M^{s_0}$ . Now (C2) follows by noting that the domain of all  $g_{\alpha}^{(s_0)}$  is  $\text{conv}(\underline{X})$ . Therefore all conditions needed for Lemma B.6 are valid, and hence we can extract a subsequence  $\{g_{\alpha_n}^{(s_0)}\}_{n \in \mathbb{N}}$  such that

$$\begin{aligned} \lim_{n \rightarrow \infty, x \rightarrow y} g_{\alpha_n}^{(s_0)}(x) &= g^{(s_0)}(y), \quad \text{for all } y \in \text{int}(\text{dom}(g^{(s_0)})); \\ \lim_{n \rightarrow \infty, x \rightarrow y} g_{\alpha_n}^{(s_0)}(x) &\geq g^{(s_0)}(y), \quad \text{for all } y \in \mathbb{R}^d, \end{aligned}$$

holds for some  $g^{(s_0)} \in \mathcal{G}$ . This implies  $f_{\alpha_n} \rightarrow_{a.e.} f^{(s_0)}$  as  $n \rightarrow \infty$  where  $f^{(s_0)} := (g^{(s_0)})^{1/s_0}$ . Now repeat the above argument with another  $s_1$  with a further extracted subsequence  $\{\alpha_{n(k)}\}$ , we see that  $f_{\alpha_{n(k)}} \rightarrow_{a.e.} f^{(s_1)}$  ( $k \rightarrow \infty$ ) for some  $s_1$ -concave  $f^{(s_1)}$  holds for the subsequence  $\{\alpha_{n(k)}\}_{k \in \mathbb{N}}$ . This implies that  $f^{(s_0)} =_{a.e.} f^{(s_1)}$ . Since a convex function is continuous in the interior of the domain, we can choose a version of upper semi-continuous  $f$  such that  $f = f^{(s)}$  a.e. for all  $\{1/2 < s < 0\} \cap \mathbb{Q}$ . This implies that  $f$  is  $s$ -concave for any rational  $1/2 < s < 0$  and hence log-concave. Next we show weighted  $L_1$  convergence: For fixed  $\kappa > 0$ , choose  $0 > s_0 > -1/(\kappa + 1)$ . Since there exists  $a, b > 0$  such that  $g_{\alpha_n}^{(s_0)} \geq g^{(s_0)} \geq a\|x\| - b$  holds for all  $n \in \mathbb{N}$ , we have an integrable envelope function:

$$(1 + \|x\|)^{\kappa} (f_{\alpha_n}(x) \vee f(x)) \leq (1 + \|x\|)^{\kappa} \left( (a\|x\| - b) \vee M \right)^{1/s_0}.$$

Now an application of the dominated convergence theorem yields the desired weighted  $L_1$  convergence. Similar arguments show weighted convergence is also valid in arbitrary  $L_p$  norms ( $p \geq 1$ ).

Finally we show that  $f = f_1$  by virtue of Theorem 2.2 in [15] and Theorem 2.8. We note that by Lemma 5.1,  $f$  must be log-linear between consecutive data points. Now since  $f_1$  and  $f$  are both log-linear between consecutive data points of  $\{X_1, \dots, X_n\}$ , we only have to consider test functions  $h$  such that  $h$  is piecewise linear on consecutive data points. Recall  $g_\alpha = f_\alpha^{\alpha-1}$  and  $g := -\log f$  are the underlying convex functions for  $f_\alpha$  and  $f$ . For any such  $h$  with the property that,  $g + th \in \mathcal{G}$  for  $t$  small enough, we wish to argue that such  $h$  is also a valid test for  $f_\alpha$  (i.e.  $g_\alpha + th \in \mathcal{G}$  for  $t > 0$  small enough), for a sequence of  $\{\alpha_k\}$  converging up to 1 as  $k \rightarrow \infty$ . Thus we only have to argue that for all  $X_{(i)} \in \mathcal{S}(g)$ ,  $X_{(i)} \in \mathcal{S}(g_\alpha)$  for a sequence of  $\{\alpha_k\}$  going up to 1 as  $k \rightarrow \infty$ . Assume the contrary that  $X_{(i)} \notin \mathcal{S}(g_\alpha)$  for all  $\alpha$  close enough to 1. Then  $\{g_\alpha\}$ 's are all linear on a closed interval  $I = [a, b]$  containing  $X_{(i)}$  for  $\alpha$  close to 1. Since  $f_\alpha \rightarrow f$  uniformly on  $I$  by Theorem 3.7, in particular  $f_\alpha(a)$  and  $f_\alpha(b)$  converges, Lemma 5.1 entails that  $f$  is log-linear over  $I$ , a contradiction to the fact  $X_{(i)} \in \mathcal{S}(g)$ . Hence we can find a subsequence  $\{\alpha_k\}$  going up to 1 as  $k \rightarrow \infty$  such that for all  $X_{(i)} \in \mathcal{S}(g)$ ,  $X_{(i)} \in \mathcal{S}(g_{\alpha_k})$ , i.e. for all feasible test function  $h$  of  $f_1$ , being linear on consecutive data points, is also valid for  $f_{\alpha_k}$ . Now combining the fact that  $f_{\alpha_k}$  converges in  $L_2$  metric to  $f$  and Theorem 2.2 in [15] we conclude  $f_1 = f$ .  $\square$

5.2. *Proofs in Section 3.* We first state some useful lemmas that give good control of tails with local information of the  $s$ -concave densities whose proofs can be found in the Appendix.

LEMMA 5.2. *Let  $x_0, \dots, x_d$  be  $d + 1$  points in  $\mathbb{R}^d$  such that its convex hull  $\Delta = \text{conv}(\{x_0, \dots, x_d\})$  is non-void. If  $f(y) \leq \min_j (\frac{1}{d} \sum_{i \neq j} f^s(x_i))^{1/s}$ , then*

$$f(y) \leq f_{\max} \left( 1 - \frac{d}{r} + \frac{d}{r} f_{\min} C (1 + \|y\|^2)^{1/2} \right)^{-r}.$$

Here the constant  $C = \lambda_d(\Delta)(d+1)^{-1/2} \sigma_{\max}(X)^{-1}$  where  $X = \begin{pmatrix} x_0 & \dots & x_d \\ 1 & \dots & 1 \end{pmatrix}$

and  $f_{\min} := \min_{0 \leq j \leq d} f(x_j)$ ,  $f_{\max} := \max_{0 \leq j \leq d} f(x_j)$ .

LEMMA 5.3. *Let  $\nu$  be a probability measure with  $s$ -concave density  $f$ . Suppose that  $B(0, \delta) \subset \text{int}(\text{dom}(f))$  for some  $\delta > 0$ . Then for any  $y \in \mathbb{R}^d$ ,*

$$\sup_{x \in B(y, \delta_t)} f(x) \leq J_0 \left( \frac{1}{t} \left( \left( \frac{\nu(B(ty, \delta_t))}{J_0 \lambda_d(B(ty, \delta_t))} \right)^{-1/r} - (1-t) \right) \right)^{-r},$$

where  $J_0 := \inf_{v \in B(0, \delta)} f(v)$  and  $\delta_t = \delta \frac{1-t}{1+t}$ .

Now we are at the position to prove Theorem 3.7.

PROOF OF THEOREM 3.7. That the sequence  $\{f_n\}_{n \in \mathbb{N}}$  converges uniformly on any compact subset in  $\text{int}(\text{dom}(f))$  follows directly from Lemma 3.2 and Theorem 10.8 [29]. Now we show that if  $f$  is continuous at  $y \in \mathbb{R}^d$  with  $f(y) = 0$ , then for any  $\eta > 0$  there exists  $\delta = \delta(y, \eta)$  such that

$$(5.8) \quad \limsup_{n \rightarrow \infty} \sup_{x \in B(y, \delta(y, \eta))} f_n(x) \leq \eta.$$

Assume without loss of generality that  $B(0, \delta_0) \subset \text{int}(\text{dom}(f))$  for some  $\delta_0 > 0$ . Let  $J_0 := \inf_{x \in B(0, \delta_0)} f(x)$ . Then uniform convergence of  $\{f_n\}$  to  $f$  over  $B(0, \delta_0)$  entails that

$$\liminf_{n \rightarrow \infty} \inf_{x \in B(0, \delta_0)} f_n(x) \geq J_0.$$

Hence with  $\delta_t = \delta_0 \frac{1-t}{1+t}$ , it follows from Lemma 5.3 that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{x \in B(y, \delta_t)} f_n(x) &\leq J_0 \left( \frac{1}{t} \left( \left( \frac{\nu(B(ty, \delta_t))}{J_0 \lambda_d(B(ty, \delta_t))} \right)^{-1/r} - (1-t) \right) \right)^{-r} \\ &\leq J_0 \left( \frac{J_0^{1/r} (\sup_{x \in B(ty, \delta_t)} f(x))^{-1/r} - (1-t)}{t} \right)^{-r} \rightarrow 0 \end{aligned}$$

as  $t \nearrow 1$ . This completes the proof for (5.8). So far we have shown that

$$\lim_{n \rightarrow \infty} \sup_{x \in S \cap B(0, \rho)} |f_n(x) - f(x)| = 0$$

holds for every  $\rho \geq 0$ , where  $S$  is the closed set contained in the continuity point of  $f$ . Our goal is to let  $\rho \rightarrow \infty$  and conclude. Let  $\Delta = \text{conv}\{x_0, \dots, x_d\}$  be a non-void simplex with  $x_0, \dots, x_d \in \text{int}(\text{dom}(f))$ . Note first by a closer look at the proof of Lemma 3.5,  $f_n(x) \vee f(x) \leq ((a\|x\| - b)_+)^{1/s}$  holds for all  $x \in \mathbb{R}^d$  with some  $a, b > 0$ . Let  $\rho_0 := \inf\{\rho \geq 0 : (a\rho - b)^{1/s} \leq f_{\min}/2\}$  where  $f_{\min} := \min_{0 \leq j \leq d} f(x_j) > 0$ . Then

$$\begin{aligned} &\{x \in \mathbb{R}^d : \|x\| \geq \rho_0\} \\ &\subset \bigcap_{n \geq 1} \{f_n \leq f_{\min}/2\} \cap \{f \leq f_{\min}/2\} \\ &\subset \bigcap_{n \geq n_0} \{f_n \leq (f_n)_{\min}\} \cap \{f \leq f_{\min}\} \\ &\subset \bigcap_{n \geq n_0} \{f_n \leq \min_j \left( \frac{1}{d} \sum_{i \neq j} f_n^s(x_i) \right)^{1/s}\} \cap \{f \leq \min_j \left( \frac{1}{d} \sum_{i \neq j} f^s(x_i) \right)^{1/s}\}, \end{aligned}$$

where  $n_0 \in \mathbb{N}$  is a large constant. The second inclusion follows from the fact that  $\lim_{n \rightarrow \infty} f_n(x_i) = f(x_i)$  holds for  $i = 0, \dots, d$ . By Lemma 5.2 we conclude that

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sup_{x: \|x\| \geq \rho \vee \rho_0} (1 + \|x\|)^\kappa (f_n(x) \vee f(x)) \\ & \leq \sup_{x: \|x\| \geq \rho \vee \rho_0} f_{\max} (1 + \|x\|)^\kappa \left( 1 - \frac{d}{r} + \frac{d}{r} f_{\min} C (1 + \|x\|^2)^{1/2} \right)^{-r} \rightarrow 0, \end{aligned}$$

as  $\rho \rightarrow \infty$ . This completes the proof.  $\square$

PROOF OF THEOREM 3.8. Since  $\nabla_\xi f_n(x) = -r g_n(x)^{1/s-1} \nabla_\xi g_n(x)$ ,

$$\begin{aligned} & |\nabla_\xi f_n(x) - \nabla_\xi f(x)| \\ & = r \left| g_n(x)^{1/s} \nabla_\xi g_n(x) - g(x)^{1/s} \nabla_\xi g(x) \right| \\ & \leq r \left( f_n(x) |\nabla_\xi g_n(x) - \nabla_\xi g(x)| + |f_n(x) - f(x)| |\nabla_\xi g(x)| \right) \\ & \leq 2r \sup_{x \in T} |f(x)| |\nabla_\xi g_n(x) - \nabla_\xi g(x)| + r \sup_{x \in T} |f_n(x) - f(x)| \sup_{x \in T} \|\nabla g(x)\|_2 \end{aligned}$$

holds for  $n$  large enough by Theorem 3.7. By Theorem 23.4 in [29],  $\nabla_\xi g_n(x) = \tau_x^T \xi$  for some  $\tau_x \in \partial g_n(x)$  since  $\partial g_n(x)$  is a closed set. Thus the first term above is further bounded by

$$2r \sup_{x \in T} |f(x)| \sup_{x \in T, \tau \in \partial g_n(x)} \|\tau - \nabla g(x)\|_2,$$

which vanishes as  $n \rightarrow \infty$  in view of Lemma 3.10 in [31]. Note that  $\nabla g(\cdot)$  is continuous on  $T$  by Corollary 25.5.1 in [29], and hence  $\sup_{x \in T} \|\nabla g(x)\|_2 < \infty$ . Now it is easy to see that the second term also vanishes as  $n \rightarrow \infty$  by virtue of Theorem 3.7.  $\square$

5.3. *Proofs in Section 4.* We first state the tightness result.

THEOREM 5.4. *We have the following conclusions.*

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

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

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

2. The localized processes satisfy

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

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 ‘envelope’ argument roughly indicated in Section 4.1. For a proof of this technical result, we refer the reader to the Appendix. Here we will provide proofs for our main results.

PROOF OF THEOREM 4.1. By the same tightness and uniqueness argument adopted in [18], [2], and [1], we only have to find out 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 5.4, we know that

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

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  $\gamma_1, \gamma_2$  yields

$$(5.9) \quad \gamma_1 = a^{\frac{2k+4}{2k+1}} b^{-\frac{3}{2k+1}}, \quad \gamma_2 = a^{-\frac{2}{2k+1}} b^{\frac{2}{2k+1}}.$$

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

$$(5.10) \quad \begin{pmatrix} n^{\frac{k}{2k+1}} (\hat{g}_n(x_0 + s_n t) - g_0(x_0) - s_n t g_0'(x_0)) \\ n^{\frac{k-1}{2k+1}} (\hat{g}'_n(x_0 + s_n t) - g_0'(x_0)) \end{pmatrix} \rightarrow_d \begin{pmatrix} \frac{g_0(x_0)}{-r} \frac{d^2}{dt^2} \mathbb{H}(t) \\ \frac{g_0(x_0)}{-r} \frac{d^3}{dt^3} \mathbb{H}(t) \end{pmatrix}$$

It is easy to see that  $\frac{d^2}{dt^2} \mathbb{H}(t) = \gamma_1 \gamma_2^2 \frac{d^2}{dt^2} H_k(\gamma_2 t)$  and  $\frac{d^3}{dt^3} \mathbb{H}(t) = \gamma_1 \gamma_2^3 \frac{d^3}{dt^3} H_k(\gamma_2 t)$ . Now plug in (5.9) we get the conclusion by direct calculation and the delta method.  $\square$

## APPENDIX A: SUPPLEMENTARY PROOFS

### A.1. Supplementary Proofs for Section 2.

PROOF OF LEMMA 2.1. Let  $Q \in \mathcal{Q}_1$ . Then by letting  $g(x) := \|x\| + 1$ , we have

$$L(Q) \leq L(g, Q) = \int (1 + \|x\|) \, dQ + \frac{1}{|\beta|} \int \frac{dx}{(1 + \|x\|)^{-\beta}} < \infty,$$

by noting  $Q \in \mathcal{Q}_1$ , and  $-\beta = -1 - 1/s > d$ . Now assume  $L(Q) < \infty$ . If  $Q \notin \mathcal{Q}_1$ , i.e.  $\int \|x\| \, dQ = \infty$ , then since for each  $g \in \mathcal{G}$ , we can find some  $a, b > 0$  such that  $g(x) \geq a\|x\| - b$ , we have

$$L(g, Q) = \int g \, dQ + \frac{1}{|\beta|} \int g^\beta \, dx \geq \int (a\|x\| - b) \, dQ = \infty,$$

a contradiction. This implies  $Q \in \mathcal{Q}_1$ .  $\square$

PROOF OF LEMMA 2.3. Let  $g, h$  be two minimizers for  $\mathcal{P}_Q$ . Since  $\psi_s(x) = \frac{1}{|\beta|}x^\beta$  is strictly convex on  $[0, \infty)$ ,  $L(t \cdot g + (1-t) \cdot h, Q)$  is strictly convex in  $t \in [0, 1]$  unless  $g = h$  a.e. with respect to the canonical Lebesgue measure. We claim if two closed functions  $g, h$  agree a.e. with respect to the canonical Lebesgue measure, then it must agree everywhere, thus closing the argument. It is easy to see  $\text{int}(\text{dom}g) = \text{int}(\text{dom}h)$ . Since  $\text{int}(\text{dom}(g)) \neq \emptyset$ , we have  $\text{ri}(\text{dom}g) = \text{int}(\text{dom}g) = \text{int}(\text{dom}h) = \text{ri}(\text{dom}h)$ . Also note that a convex function is continuous in the interior of its domain, and hence almost everywhere equality implies everywhere equality within the interior of the domain, i.e.  $g|_{\text{int}(\text{dom}g)} = h|_{\text{int}(\text{dom}h)}$ . Now by Corollary 7.3.4 in [29], and the closedness of  $g, h$ , we find that  $g = \text{cl}g = \text{cl}h = h$ .  $\square$

PROOF OF COROLLARY 2.9. Let  $g \equiv g(\cdot|Q)$ . Then by Theorem 2.2 and Lemma B.3, we find that there exists some  $a, b > 0$  such that  $g(x) \geq a\|x\| + b$ . Now take  $v \in \partial h(0)$ , i.e.  $h(x) \geq h(0) + v^T x$  holds for all  $x \in \mathbb{R}^d$ . Hence for  $t > 0$ , we have

$$g(x) + th(x) \geq a\|x\| + b + t(h(0) + v^T x) \geq (a - t\|v\|)\|x\| + (b + th(0)),$$

which implies that  $g + th \in \mathcal{G}$  for  $t > 0$  small enough. Now the conclusion follows from the Theorem 2.8.  $\square$

PROOF OF THEOREM 2.11. We first note that if  $F$  is a distribution function for a probability measure supported on  $[X_{(1)}, X_{(n)}]$ , and  $h : [X_{(1)}, X_{(n)}] \rightarrow \mathbb{R}$  an absolutely continuous function, then integration by parts (Fubini's theorem) yields

$$(A.1) \quad \int h \, dF = h(X_{(n)}) - \int_{X_{(1)}}^{X_{(n)}} h'(x)F(x) \, dx.$$

First we assume  $g_n = \hat{g}_n$ . For fixed  $t \in [X_{(1)}, X_{(n)}]$ , let  $h_1$  be a convex function whose derivative is given by  $h_1'(x) = -\mathbf{1}(x \leq t)$ . Now by Theorem 2.8 we find that  $\int h_1 dF_n = \int h_1 d\hat{F}_n \leq \int h_1 d\mathbb{F}_n$ . Plugging in (A.1) we find that  $\int_{X_{(1)}}^t F_n(x) dx \leq \int_{X_{(1)}}^t \mathbb{F}_n(x) dx$ . For  $t \in \mathcal{S}_n(g_n)$ , let  $h_2$  be the function with derivative  $h_2'(x) = \mathbf{1}(x \leq t)$ . It is easy to see  $g_n + th_2$  is convex for  $t > 0$  small enough, whence Theorem 2.8 is valid, thus giving the reverse direction of inequality. This shows the necessity.

For sufficiency, assume  $g_n$  satisfies (2.13). In view of the proof of Theorem 2.8, we only have to show (2.12) holds for all function  $h : \mathbb{R} \rightarrow \overline{\mathbb{R}}$  which is linear on  $[X_{(i)}, X_{(i+1)}]$  ( $i = 1, \dots, n-1$ ) and  $g_n + th$  convex for  $t > 0$  small enough. Since  $g_n$  is a linear function between two consecutive knots,  $h$  must be convex between consecutive knots. This implies that the derivative of such an  $h$  can be written as  $h'(x) = \sum_{j=2}^n \beta_j \mathbf{1}(x \leq X_{(j)})$ , with  $\beta_2, \dots, \beta_n$  satisfying  $\beta_j \leq 0$  if  $X_{(j)} \notin \mathcal{S}_n(g_n)$ . Now again by (A.1) we have

$$\begin{aligned} \int h d\hat{F}_n &= h(X_n) - \sum_{j=2}^n \beta_j \int_{X_{(1)}}^{X_{(j)}} \hat{F}_n(x) dx \\ &\leq h(X_n) - \sum_{j=2}^n \beta_j \int_{X_{(1)}}^{X_{(j)}} \mathbb{F}_n(x) dx = \int h d\mathbb{F}_n, \end{aligned}$$

as desired.  $\square$

PROOF OF COROLLARY 2.12. This follows directly from the Theorem 2.11 by noting for  $x_1 < x_0 < x_2$  we have

$$\frac{1}{x_2 - x_0} \int_{x_0}^{x_2} \hat{F}_n(x) dx \leq \frac{1}{x_2 - x_0} \int_{x_0}^{x_2} \mathbb{F}_n(x) dx,$$

and

$$\frac{1}{x_0 - x_1} \int_{x_1}^{x_0} \hat{F}_n(x) dx \geq \frac{1}{x_0 - x_1} \int_{x_1}^{x_0} \mathbb{F}_n(x) dx.$$

Now let  $x_1 \nearrow x_0$  and  $x_2 \searrow x_0$  we find that  $\hat{F}_n(x_0) \leq \mathbb{F}_n(x_0)$  by right continuity and  $\hat{F}_n(x_0) \geq \mathbb{F}_n(x_0-) = \mathbb{F}_n(x_0) - \frac{1}{n}$ .  $\square$

PROOF OF THEOREM 2.13. The proof closely follows the proof of Theorem 2.7 of [16]. For the reader's convenience we give a full proof here. Let  $P$  denote the probability distribution corresponding to  $F$ . We first show necessity by assuming  $g = g(\cdot|Q)$ . By Corollary 2.9 applied to  $h(x) = \pm x$ , we find by Fubini's theorem that

$$0 = \int_{\mathbb{R}} x d(Q - P)(x) = \int_{\mathbb{R}} (F - G)(t) dt$$

which proves (1). Now we turn to (2). Since the map  $s \mapsto (s-x)_+$  is convex, again by Corollary 2.9, we find

$$0 \leq \int_{\mathbb{R}} (s-x)_+ d(Q-P)(s) = - \int_{-\infty}^x (F-G)(t) dt,$$

where in the last equality we used the proved fact that  $\int_{\mathbb{R}} (F-G)d\lambda = 0$ . Now we assume  $x \in \tilde{\mathcal{S}}(g)$ , and discuss two different cases to conclude. If  $x \in \partial(\text{dom}(g))$ , then let  $h(s) = -(s-x)_+$ , it is easy to see  $g+th \in \mathcal{G}$  for  $t > 0$  small enough. Then by Theorem 2.8, we have

$$0 \leq \int h(s)d(Q-P)(s) = \int_{-\infty}^x (F-G)(t) dt.$$

If  $x \in \text{int}(\text{dom}(g))$ , then  $g'(x-\delta) < g'(x+\delta)$  for small  $\delta > 0$  by definition, and hence we define

$$H'_\delta(u) = -\frac{g'(u) - g'(x-\delta)}{g'(x+\delta) - g'(x-\delta)} \mathbf{1}_{\{u \in [x-\delta, x+\delta]\}} - \mathbf{1}_{\{u > x+\delta\}},$$

whose integral  $H_\delta(s) := \int_{-\infty}^s H'_\delta(u) du$  serves as an approximation of  $-(s-x)_+$  as  $\delta \searrow 0$ . Note that

$$(g+tH_\delta)(s) = g(s) - \frac{t}{g'(x+\delta) - g'(x-\delta)} \int_{s \wedge (x-\delta)}^{s \wedge (x+\delta)} (g'(u) - g'(x-\delta)) du - t(s-(x+\delta))_+,$$

implying  $g+tH_\delta \in \mathcal{G}$  for  $t > 0$  small enough (which may depend on  $\delta$ ). Then by Theorem 2.8,

$$0 \leq \int H_\delta(s)d(Q-P)(s) \rightarrow - \int (s-x)_+ d(Q-P)(s) = \int_{-\infty}^x (F-G)(t) dt,$$

as  $\delta \searrow 0$ , where the convergence follows easily from dominated convergence theorem. This proves (2). Now we show sufficiency by assuming (1)-(2). Consider a Lipschitz continuous function  $\Delta(\cdot)$  with Lipschitz constant  $L$ . Then

$$\begin{aligned} \int \Delta d(Q-P) &= \int \Delta'(F-G) d\lambda = - \int (L - \Delta')(F-G) d\lambda \\ &= - \int_{\mathbb{R}} \left( \int_{-L}^L \mathbf{1}_{\{s > \Delta'(t)\}} ds \right) (F-G)(t) dt \\ &= - \int_{-L}^L \int_{A(\Delta', s)} (F-G)(t) dt ds, \end{aligned}$$

where the second line follows from (1), and  $A(\Delta', s) := \{t \in \mathbb{R} : \Delta'(t) < s\}$ . Now replace the generic Lipschitz function  $\Delta$  with  $g^{(\epsilon)}$  as defined in Lemma B.1 with Lipschitz constant  $L = 1/\epsilon$ . Note in this case  $A((g^{(\epsilon)})', s) = (-\infty, a(g, \epsilon))$ , where  $a(g, s) = \min\{t \in \mathbb{R} : g'(t+) \geq s\}$  and hence  $a(g, s) \in \bar{S}(g)$ . This implies that  $\int_{A((g^{(\epsilon)})', s)} (F - G)(s) ds = 0$  for all  $s \in (-L, L)$  by (2), yielding that  $\int g^{(\epsilon)} d(Q - P) = 0$ . Similarly we have  $\int g_0^{(\epsilon)} d(Q - P) \geq 0$  where  $g_0 = g(\cdot|Q)$ . Now let  $\epsilon \searrow 0$ , by monotone convergence theorem we find that  $\int g dQ = \int g dP$  and that  $\int g_0 dQ \geq \int g_0 dP$ . This yields

$$L(g_0, Q) \geq L(g_0, P) \geq L(g, P) = L(g, Q),$$

where the second inequality follows from the Fisher consistency of functional  $L(\cdot, \cdot)$  and the fact that  $P$  is the distribution corresponding to  $g$ .  $\square$

### A.2. Supplementary Proofs for Section 3.

PROOF OF LEMMA 3.1. The proof closely follows the first part of the proof of Proposition 2 [20]. Suppose  $\dim(\text{csupp}(\nu)) = d$ , we show  $\text{csupp}(\nu) \subset \bar{C}$ . To see this, we take  $x_0 \notin \bar{C}$ , then there exists  $\delta > 0$  such that  $B(x_0, \delta) \subset C^c$ , and we claim that

$$(A.2) \quad \text{For all } x^* \in B(x_0, \delta) \subset C^c, x^* \notin \text{int}(\text{csupp}(\nu)).$$

If (A.2) holds, then  $x_0 \notin \text{csupp}(\nu)$  and hence  $\text{csupp}(\nu) \subset \bar{C}$ . Now we turn to show (A.2). Since  $x^* \notin C = \{\liminf_{n \rightarrow \infty} f_n(x) > 0\}$ , we can find a subsequence  $\{f_{n(k)}\}_{k \in \mathbb{N}}$  of  $\{f_n\}_{n \in \mathbb{N}}$  such that  $f_{n(k)}(x^*) < \frac{1}{k}$  holds for all  $k \in \mathbb{N}$ . Hence  $x^* \notin \Gamma_k := \{x \in \mathbb{R}^d : f_{n(k)}(x) \geq \frac{1}{k}\}$ . Note that  $\Gamma_k$  is a closed convex set, hence by Hyperplane Separation Theorem we can find  $b_k \in \mathbb{R}^d$  with  $\|b_k\| = 1$  such that  $\{x \in \mathbb{R}^d : \langle b_k, x \rangle \leq \langle b_k, x^* \rangle\} \subset (\Gamma_k)^c$ . Without loss of generality we may assume  $b_k \rightarrow b_{x^*}$  as  $k \rightarrow \infty$  for some  $b_{x^*} \in \mathbb{R}^d$  with  $\|b_{x^*}\| = 1$ . Now for fixed  $R > 0$  and  $\eta > 0$ , define

$$A_{R, \eta} := \{x \in \mathbb{R}^d : \langle b_{x^*}, x \rangle < \langle b_{x^*}, x^* \rangle - \eta, \|x\| \leq R\}.$$

Choose  $k_0 \in \mathbb{N}$  large enough such that  $\|b_k - b_{x^*}\| \leq \frac{\eta}{2R}$  holds for all  $k \geq k_0(x^*, \eta, R)$ . Now for  $R > \|x^*\|$  and  $x \in A_{R, \eta}$ , we have

$$\langle b_k, x - x^* \rangle = \langle b_{x^*}, x - x^* \rangle + \langle b_k - b_{x^*}, x - x^* \rangle < -\eta + \frac{\eta}{2R}(\|x\| + \|x^*\|) \leq 0$$

holds for all  $k \geq k_0(x^*, \eta, R)$ . This implies for  $R > \|x^*\|$  and  $\eta > 0$ ,

$$A_{R, \eta} \subset \{x \in \mathbb{R}^d : \langle b_k, x \rangle \leq \langle b_k, x^* \rangle\} \subset (\Gamma_k)^c = \{x \in \mathbb{R}^d : f_{n(k)}(x) < \frac{1}{k}\}.$$

Now note  $A_{R,\eta}$  is open, by Portmanteau Theorem we find that

$$\nu(A_{R,\eta}) \leq \liminf_{k \rightarrow \infty} \nu_{n(k)}(A_{R,\eta}) = \liminf_{k \rightarrow \infty} \int_{A_{R,\eta}} f_{n(k)}(x) dx \leq \liminf_{k \rightarrow \infty} \frac{\lambda_d(A_{R,\eta})}{k} = 0.$$

This implies

$$\nu(\{x \in \mathbb{R}^d : \langle b_{x^*}, x \rangle < \langle b_{x^*}, x^* \rangle\}) = \nu\left(\bigcup_{R=1}^{\infty} A_{R,1/R}\right) = \lim_{R \rightarrow \infty} \nu(A_{R,1/R}) = 0,$$

where the second equality follows from the fact  $\{A_{R,1/R}\}$  is an increasing family as  $R$  increases. By the assumption that  $\dim(\text{csupp}(\nu)) = d$ , we find  $x^* \notin \text{int}(\text{csupp}(\nu))$ , as we claimed in (A.2).

Now Suppose  $\dim C = d$ , we claim  $\overline{C} \subset \text{csupp}(\nu)$ . To see this, we only have to show  $C \subset \text{csupp}(\nu)$  by the closedness of  $\text{csupp}(\nu)$ . Suppose not, then we can find  $x_0 \in C \setminus \text{csupp}(\nu)$ . This implies that there exists  $\delta > 0$  such that  $B(x_0, \delta) \cap \text{csupp}(\nu) \neq \emptyset$ . By the assumption that  $\dim C = d$ , we can find  $x_1, \dots, x_d \in B(x_0, \delta) \cap C$  such that  $\{x_0, \dots, x_d\}$  are in general position. By definition of  $C$  we can find  $\epsilon_0 > 0, n_0 \in \mathbb{N}$  such that  $f_n(x_j) \geq \epsilon_0$  for all  $j = 0, 1, \dots, d$  and  $n \geq n_0$ . By convexity, we conclude that  $f_n(x) \geq \epsilon_0$ , for all  $x \in \text{conv}(\{x_0, \dots, x_d\})$  and  $n \geq n_0$ . This gives

$$\begin{aligned} \nu(\text{conv}(\{x_0, \dots, x_d\})) &\geq \limsup_{n \rightarrow \infty} \nu_n(\text{conv}(\{x_0, \dots, x_d\})) \\ &\geq \epsilon_0 \lambda_d(\text{conv}(\{x_0, \dots, x_d\})) > 0, \end{aligned}$$

a contradiction with  $B(x_0, \delta) \cap \text{csupp}(\nu) \neq \emptyset$ , thus completing the proof of the claim. To summarize, we have proved

1. If  $\dim(\text{csupp}(\nu)) = d$ , then  $\text{csupp}(\nu) \subset \overline{C}$ . This in turn implies  $\dim C = d$ , and hence  $\overline{C} \subset \text{csupp}(\nu)$ . Now it follows that  $\text{csupp}(\nu) = \overline{C}$ ;
2. If  $\dim C = d$ , then  $\overline{C} \subset \text{csupp}(\nu)$ . This in turn implies  $\dim(\text{csupp}(\nu)) = d$ , and hence  $\text{csupp}(\nu) \subset \overline{C}$ . Now it follows that  $\text{csupp}(\nu) = \overline{C}$ .  $\square$

**PROOF OF LEMMA 3.2.** The proof is essentially the same as the proof of Proposition 2 [9] by exploiting convexity at the level of the underlying basic convex function so we shall omit it.  $\square$

**PROOF OF LEMMA 3.3.** Set  $U_{n,t} = \{x \in \mathbb{R}^d : f_n(x) \geq t\}$ . We first claim that there exists  $n_0 \in \mathbb{N}, \epsilon_0 \in (0, 1)$  such that  $\lambda_d(U_{n,\epsilon_0}) \geq \epsilon_0$  holds for all  $n \geq n_0$ . If not, then for all  $k \in \mathbb{N}, l \in \mathbb{N}$ , there exists  $n_{k,l} \in \mathbb{N}$  such that  $\lambda_d(U_{n_{k,l}, 1/l}) \leq \frac{1}{l}$ . Note that  $\{\liminf_n f_n > 0\} = \cup_{k \in \mathbb{N}} \cup_{l \in \mathbb{N}} \cap_{n \geq k} U_{n, 1/l}$ . Since

$\lambda_d(\bigcup_{l \in \mathbb{N}} \bigcap_{n \geq k} U_{n,1/l}) = \lim_{l \rightarrow \infty} \lambda_d(\bigcap_{n \geq k} U_{n,1/l}) \leq \lim_{l \rightarrow \infty} \lambda_d(U_{n_{k,l},1/l}) = 0$ , we find that  $C = \{\liminf_n f_n > 0\}$  is a countable union of null set and hence  $\lambda_d(C) = 0$ , a contradiction to the assumption  $\dim C = d$ . This shows the claim.

Denote  $M_n := \sup_{x \in \mathbb{R}^d} f_n(x)$ ,  $\epsilon_n \in \text{Arg max } f_n(x)$ . Without loss of generality we assume  $M_n \geq \frac{\epsilon_0}{(1+\kappa_s)^{1/s}}$  where  $\kappa_s = (1/2)^s - 1 > 0$ , and we set  $\lambda_n := \frac{\kappa_s M_n^s}{\epsilon_0^s - M_n^s} \in [0, 1]$ . Now for  $x \in U_{n,\epsilon_0}$ , by convexity of  $f_n^s$  we have

$$f_n^s(\epsilon_n + \lambda_n(x - \epsilon_n)) \leq \lambda_n f_n^s(x) + (1 - \lambda_n) f_n^s(\epsilon_n) \leq \lambda_n \epsilon_0^s + (1 - \lambda_n) M_n^s = (M_n/2)^s.$$

This implies  $f_n(x) \geq M_n/2 := \Omega_n$ , for all  $x \in V_{n,\epsilon_0} := \{\epsilon_n + \lambda_n(x - \epsilon_n) : x \in U_{n,\epsilon_0}\}$ . Hence  $V_{n,\epsilon_0} \subset U_{n,\Omega_n}$  and therefore  $\lambda_d(V_{n,\epsilon_0}) = \lambda_d(U_{n,\epsilon_0}) \lambda_n^d$ , thus

$$\lambda_d(U_{n,\Omega_n}) \geq \lambda_d(V_{n,\epsilon_0}) = \lambda_d(U_{n,\epsilon_0}) \lambda_n^d \geq \epsilon_0 \lambda_n^d,$$

holds for all  $n \geq n_0$ . On the other hand,

$$1 = \int f_n \geq \Omega_n \lambda_d(U_{n,\Omega_n}) \geq \Omega_n \epsilon_0 \lambda_n^d,$$

and suppose the contrary that  $M_n \rightarrow \infty$  as  $n \rightarrow \infty$ , then

$$1 \geq \Omega_n \epsilon_0 \lambda_n^d = \frac{\epsilon_0 \kappa_s^d}{2(\epsilon_0^s - M_n^s)^d} M_n^{1+sd} \geq c M_n^{1+sd} \rightarrow \infty, \quad n \rightarrow \infty,$$

since  $1 + sd > 0$  by assumption  $-1/d < s < 0$ . Here  $c = \frac{\epsilon_0^{1-sd} \kappa_s^d}{2}$ . This gives a contradiction and the proof is complete.  $\square$

**PROOF OF THEOREM 3.4.** We only have to show  $\nu$  is absolutely continuous with respect to  $\lambda_d$ . To this end, for given  $\epsilon > 0$ , choose  $\delta = \epsilon/2M$ , where  $M := \sup_n \|f_n\|_\infty < \infty$  by virtue of Lemma 3.3. Now for Borel set  $A \subset \mathbb{R}^d$  with  $\lambda_d(A) \leq \delta$ , we can take an open  $A' \supset A$  such that  $\lambda_d(A') \leq 2\delta$  by the regularity of Lebesgue measure. Then

$$\nu(A) \leq \nu(A') \leq \liminf_{n \rightarrow \infty} \nu_n(A') = \liminf_{n \rightarrow \infty} \int_{A'} f_n \leq 2\delta M = \epsilon,$$

as desired.  $\square$

**PROOF OF LEMMA 3.5.** Let  $g_n = f_n^s$  and  $g = f^s$ . Without loss of generality we assume  $0 \in \text{int}(\text{dom}(g))$ , and choose  $\eta > 0$  small enough such that  $B_\eta := \overline{B}(0, \eta) \subset \text{int}(\text{dom}(g))$ . By the Lemma B.3, we know there exists  $a > 0, R > 0$  such that  $\frac{g(x) - g(0)}{\|x\|} \geq a$ , holds for all  $\|x\| \geq \frac{R}{2}$ . Now

we claim that there exists  $n_0 \in \mathbb{N}$  such that  $\frac{g_n(x) - g_n(0)}{\|x\|} \geq \frac{a}{8}$ , holds for all  $\|x\| \geq R$  and  $n \geq n_0$ . Note for each  $n \in \mathbb{N}$ , by convexity of  $g_n(\cdot)$ , we know that for fixed  $x \in \mathbb{R}^d$ , the quantity  $\frac{g_n(\lambda x) - g_n(0)}{\|\lambda x\|}$  is non-decreasing in  $\lambda$ , so we only have to show the claim for  $\|x\| = R$  and  $n_0 \geq n$ . Suppose the contrary, then we can find a subsequence  $\{g_{n(k)}\}$  and  $\|x_{n(k)}\| = R$  such that  $\frac{g_{n(k)}(x_{n(k)}) - g_{n(k)}(0)}{\|x_{n(k)}\|} < \frac{a}{8}$ . For simplicity of notation we think of  $\{g_n\}, \{x_n\}$  as  $\{g_{n(k)}\}, \{x_{n(k)}\}$ . Now define  $A_n := \text{conv}(\{x_n, B_\eta\})$ ;  $B_n := \{y \in \mathbb{R}^d : \|y - x_n\| \leq R/2\}$ ;  $C_n := A_n \cap B_n$ . By reducing  $\eta > 0$  if necessary, we may assume  $B_\eta \cap B_n = \emptyset$ . It is easy to see  $C_n$  is convex and  $\lambda_d(C_n) = \lambda_0$  is a constant independent of  $n \in \mathbb{N}$ . By Lemma 3.2, we know that  $g_n \rightarrow_{a.e.} g$  on  $B_\eta$ , and hence  $\sup_{x \in B_\eta} |g_n(x) - g(x)| \rightarrow 0 (n \rightarrow \infty)$  by Theorem 10.8, [29]. By further reducing  $\eta > 0$  if necessary, we may assume  $g_n(y) \leq g(0) + \frac{aR}{8}$ , holds for all  $y \in B_\eta$  and  $n \in \mathbb{N}$ . Now for any  $x^* \in C_n$ , write  $x^* = \lambda x_n + (1 - \lambda)y$ , by noting  $R/2 \leq \|x^*\| \leq R$  and convexity of  $g_n$ , we get

$$\begin{aligned} \frac{g_n(x^*) - g_n(0)}{\|x^*\|} &\leq \frac{\lambda g_n(x_n) + (1 - \lambda)g_n(y) - g_n(0)}{\|x^*\|} \\ &= \lambda \cdot \frac{g_n(x_n) - g_n(0)}{\|x_n\|} \cdot \frac{\|x_n\|}{\|x^*\|} + (1 - \lambda) \frac{g_n(y) - g_n(0)}{\|x^*\|} \\ &\leq \lambda \cdot \frac{a}{8} \frac{R}{R/2} + (1 - \lambda) \frac{aR/8}{R/2} = \frac{a}{4}. \end{aligned}$$

This gives rise to

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int_{C_n} (f_n - f) &\geq \liminf_{n \rightarrow \infty} \lambda_0((aR/4 + g_n(0))^{1/s} - (aR/2 + g(0))^{1/s}) \\ &= \lambda_0((aR/4 + g(0))^{1/s} - (aR/2 + g(0))^{1/s}) > 0, \end{aligned}$$

which is a contradiction to Lemma B.9. This establishes our claim. Now by Lemma 3.2, we find that the set  $\{\liminf_n f_n(\cdot) > 0\}$  is full-dimensional, and hence by Lemma 3.3 we conclude  $g_n(\cdot)$  is uniformly bounded away from zero. Also note by Lemma B.8 we find  $g(\cdot)$  must be bounded away from zero, which gives the desired assertion.  $\square$

**A.3. Supplementary Proofs for Section 4.** In this subsection, we mainly aim at proving Theorem 5.4. We first observe that

LEMMA A.1.  *$k$  is an even integer and  $g_0^{(k)}(x_0) > 0$ .*

PROOF OF LEMMA A.1. By Taylor expansion of  $g_0''$  around  $x_0$ , we find that locally for  $x \approx x_0$ ,

$$g_0''(x) = \frac{g_0^{(k)}(x_0)}{(k-2)!}(x-x_0)^{k-2} + o((x-x_0)^{k-2}).$$

Also note  $g_0''(x) \geq 0$  by convexity and local smoothness assumed in (A3). This gives that  $k-2$  is even and  $g_0^{(k)}(x_0) > 0$ .  $\square$

For further technical discussions, we denote throughout this subsection that for fixed  $k$ ,  $r_n := n^{\frac{k+2}{2k+1}}$ ;  $s_n := n^{-\frac{1}{2k+1}}$ ;  $x_n(t) := x_0 + s_n t$ ;  $\mathbf{l}_{n,x_0} := [x_0, x_n(t)]$ . Let  $\tau_n^+ := \inf\{t \in \mathcal{S}_n(\hat{g}_n) : t > x_0\}$ , and  $\tau_n^- := \sup\{t \in \mathcal{S}_n(\hat{g}_n) : t < x_0\}$ . The key step in establishing the limit theory, is to establish a stochastic bound for the gap  $\tau_n^+ - \tau_n^-$  as follows.

THEOREM A.2. *Assume (A1)-(A4) hold. Then*

$$\tau_n^+ - \tau_n^- = O_p(s_n).$$

PROOF. Define  $\Delta_0(x) := (\tau_n^- - x)\mathbf{1}_{[\tau_n^-, \bar{\tau}]}(x) + (x - \tau_n^+)\mathbf{1}_{[\bar{\tau}, \tau_n^+]}(x)$ , and  $\Delta_1 := \Delta_0 + \frac{\tau_n^+ - \tau_n^-}{4}\mathbf{1}_{[\tau_n^-, \tau_n^+]}$ , where  $\bar{\tau} := \frac{\tau_n^- + \tau_n^+}{2}$ . Thus we find that

$$\begin{aligned} \int \Delta_1 d(\mathbb{F}_n - F_0) &= \int \Delta_1 d(\mathbb{F}_n - \hat{F}_n) + \int \Delta_1 d(\hat{F}_n - F_0) \\ &\geq -\frac{\tau_n^+ - \tau_n^-}{4} \left| \int_{\tau_n^-}^{\tau_n^+} d(\mathbb{F}_n - \hat{F}_n) \right| + \int \Delta_1(\hat{f}_n - f_0) d\lambda \\ &\geq -\frac{\tau_n^+ - \tau_n^-}{2n} + \int \Delta_1(\hat{f}_n - f_0) d\lambda, \end{aligned}$$

where the last line follows from Corollary 2.12. Now let  $R_{1n} := \int \Delta_1(\hat{f}_n - f_0) d\lambda$ ,  $R_{2n} := \int \Delta_1 d(\mathbb{F}_n - F_0)$ . The conclusion follows directly from the following lemma.  $\square$

LEMMA A.3. *Suppose (A1)-(A4) hold. Then  $R_{1n} = O_p(\tau_n^+ - \tau_n^-)^{k+2}$  and  $R_{2n} = O_p(r_n^{-1})$ .*

PROOF OF LEMMA A.3. Define  $p_n := \hat{g}_n/g_0$  on  $[\tau_n^+, \tau_n^-]$ . It is easy to see that  $\tau_n^+ - \tau_n^- = o_p(1)$ , so with large probability, for all  $n \in \mathbb{N}$  large enough,

$\inf_{x \in [\tau_n^+, \tau_n^-]} f_0(x) > 0$  by (A2).

$$\begin{aligned} R_{1n} &= \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) (\hat{f}_n(x) - f_0(x)) \, dx = \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) f_0(x) \left( \frac{\hat{f}_n(x)}{f_0(x)} - 1 \right) \, dx \\ &= \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) f_0(x) \left( \sum_{j=1}^{k-1} \binom{-r}{j} (p_n(x) - 1)^j + \binom{-r}{k} \theta_{x,n}^{-r-k} (p_n(x) - 1)^k \right) \, dx, \end{aligned}$$

where  $\theta_{x,n} \in [1 \wedge \frac{\hat{g}_n(x)}{g_0(x)}, 1 \vee \frac{\hat{g}_n(x)}{g_0(x)}]$ . Now define

$$\begin{aligned} S_{nj} &= \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) f_0(x) \binom{-r}{j} (p_n(x) - 1)^j \, dx, \quad 1 \leq j \leq k-1, \\ S_{nk} &= \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) f_0(x) \binom{-r}{k} \theta_{x,n}^{-r-k} (p_n(x) - 1)^k \, dx. \end{aligned}$$

Expand  $f_0$  around  $\bar{\tau}$ , then we have

$$\begin{aligned} S_{nj} &= \sum_{l=0}^{k-1} \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) \frac{f_0^{(l)}(\bar{\tau})}{l!} (x - \bar{\tau})^l \binom{-r}{j} (p_n(x) - 1)^j \, dx \\ &\quad + \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) \frac{f_0^{(l)}(\eta_{n,x,k})}{k!} (x - \bar{\tau})^k \binom{-r}{k} (p_n(x) - 1)^k \, dx, \\ S_{nk} &= \sum_{l=0}^{k-1} \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) \frac{f_0^{(l)}(\bar{\tau})}{l!} \theta_{x,n}^{-r-k} (x - \bar{\tau})^l \binom{-r}{j} (p_n(x) - 1)^k \, dx \\ &\quad + \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) \frac{f_0^{(l)}(\eta_{n,x,k})}{k!} \theta_{x,n}^{-r-k} (x - \bar{\tau})^k \binom{-r}{k} (p_n(x) - 1)^k \, dx. \end{aligned}$$

Now we see the dominating term is the first term in  $S_{n1}$  since all other terms are of higher orders, and  $|\theta_{x,n} - 1| = o_p(1)$  uniformly locally in  $x$  in view of Theorem 3.7. We denote this term  $Q_{n1}$ . Note that  $1/g_0(x_0) = 1/g_0(\tau) + o_p(1)$

uniformly in  $\tau$  around  $x_0$ , and that  $\hat{g}_n$  is piecewise linear, yielding

$$\begin{aligned}
\frac{Q_{n1}}{-r f_0(\bar{\tau})} &= \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) \frac{1}{g_0(x)} (\hat{g}_n(x) - g_0(x)) \, dx \\
&= \left( \frac{1}{g_0(x_0)} + o_p(1) \right) \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) (\hat{g}_n(x) - g_0(x)) \, dx \\
&= \left( \frac{1}{g_0(x_0)} + o_p(1) \right) \left[ (\hat{g}_n(\bar{\tau}) - g_0(\bar{\tau})) \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) \, dx \right. \\
&\quad \left. + (\hat{g}'_n(\bar{\tau}) - g'_0(\bar{\tau})) \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) (x - \bar{\tau}) \, dx \right. \\
&\quad \left. - \sum_{j=2}^k \frac{g_0^{(j)}(\bar{\tau})}{j!} \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) (x - \bar{\tau})^j \, dx \right. \\
&\quad \left. - \int_{\tau_n^-}^{\tau_n^+} \epsilon_n(x) \Delta_1(x) (x - \bar{\tau})^k \, dx \right],
\end{aligned}$$

where the first two terms in the bracket is zero by construction of  $\Delta_1$ . Now note that

$$\int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) (x - \bar{\tau})^j \, dx = \begin{cases} 0 & j = 0, \text{ or } j \text{ is odd;} \\ \frac{j}{2^{j+2}(j+1)(j+2)} (\tau_n^+ - \tau_n^-)^{j+2} & j > 0, \text{ and } j \text{ is even,} \end{cases}$$

and that  $g_0^{(j)}(\bar{\tau}) = \frac{1}{(k-j)!} (g_0^{(k)}(x_0) + o_p(1)) (\bar{\tau} - x_0)^{k-j}$ . This means that for  $j \geq 2$  and  $j$  even,

$$\begin{aligned}
\frac{g_0^{(j)}(\bar{\tau})}{j!} \int_{\tau_n^-}^{\tau_n^+} \Delta_1(x) (x - \bar{\tau})^j \, dx &= \frac{j(g_0^{(k)}(x_0) + o_p(1))}{(k-j)!(j+2)!2^{j+2}} (\bar{\tau} - x_0)^{k-j} (\tau_n^+ - \tau_n^-)^{j+2} \\
&= \frac{j(g_0^{(k)}(x_0) + o_p(1))}{(k-j)!(j+2)!2^{j+2}} O_p(1) (\tau_n^+ - \tau_n^-)^{k+2}.
\end{aligned}$$

Further note that  $\|\epsilon_n\|_\infty = o_p(1)$  as  $\tau_n^+ - \tau_n^- \rightarrow_p 0$ , we get  $Q_{n1} = O_p(\tau_n^+ - \tau_n^-)^{k+2}$ . This establishes the first claim. The proof for  $R_{2n}$  follows the same line as in the proof of Lemma 4.4 [1] p1318-1319.  $\square$

LEMMA A.4. *We have the following:*

$$\begin{aligned}
f_0^{(j)}(x_0) &= j! \binom{-r}{j} g_0(x_0)^{-r-j} (g'_0(x_0))^j, \quad 1 \leq j \leq k-1; \\
f_0^{(k)}(x_0) &= k! \binom{-r}{k} g_0(x_0)^{-r-k} (g'_0(x_0))^k - r g_0(x_0)^{-r-1} g_0^{(k)}(x_0).
\end{aligned}$$

PROOF. This follows from direct calculation.  $\square$

LEMMA A.5. *For any  $M > 0$ , we have*

$$\begin{aligned} \sup_{|t| \leq M} |\hat{g}'_n(x_0 + s_n t) - \hat{g}'_0(x_0)| &= O_p(s_n^{k-1}); \\ \sup_{|t| \leq M} |\hat{g}_n(x_0 + s_n t) - g_0(x_0) - s_n t g'_0(x_0)| &= O_p(s_n^k). \end{aligned}$$

The proof is identical to Lemma 4.4 in [18] so we shall omit it.

LEMMA A.6. *Let*

$$\hat{e}_n(u) := \hat{f}_n(u) - \sum_{j=0}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u - x_0)^j - f_0(x_0) \binom{-r}{k} \left( \frac{g'_0(x_0)}{g_0(x_0)} \right)^k (u - x_0)^k.$$

*Then for any  $M > 0$ , we have  $\sup_{|t| \leq M} |\hat{e}_n(x_0 + s_n t)| = O_p(s_n^k)$ .*

PROOF. Note that

$$\begin{aligned} \text{(A.3)} \quad \hat{f}_n(u) - f_0(x_0) &= f_0(x_0) \left[ \frac{\hat{f}_n(u)}{f_0(x_0)} - 1 \right] = f_0(x_0) \left[ \left( \frac{\hat{g}_n(u)}{g_0(x_0)} \right)^{-r} - 1 \right] \\ &= f_0(x_0) \left( \sum_{j=1}^k \binom{-r}{j} \left( \frac{\hat{g}_n(u)}{g_0(x_0)} - 1 \right)^j + \underbrace{\sum_{j \geq k+1} \binom{-r}{j} \left( \frac{\hat{g}_n(u)}{g_0(x_0)} - 1 \right)^j}_{=: \hat{\Psi}_{k,n,1}(u)} \right). \end{aligned}$$

Define  $\hat{\Psi}_{k,n,1}(u) := \sum_{j \geq k+1} \binom{-r}{j} \left( \frac{\hat{g}_n(u)}{g_0(x_0)} - 1 \right)^j = \sum_{j \geq k+1} \binom{-r}{j} \frac{1}{g_0(x_0)^j} (\hat{g}_n(u) - g_0(x_0))^j$ . Note that

$$\begin{aligned} (\hat{g}_n(u) - g_0(x_0))^j &= (\hat{g}_n(u) - g_0(x_0) - (u - x_0)g'_0(x_0) + (u - x_0)g'_0(x_0))^j \\ &= \sum_{l=1}^j \binom{j}{l} [\hat{g}_n(u) - g_0(x_0) - (u - x_0)g'_0(x_0)]^l (u - x_0)^{j-l} g'_0(x_0)^{j-l} \\ &\quad + (u - x_0)^j g'_0(x_0)^j \\ &= O_p(s_n^{kl} \cdot s_n^{j-l}) + O_p(s_n^j) \\ &\quad \text{uniformly on } \{u : |u - x_0| \leq Mn^{-1/(2k+1)}\} \\ &= O_p(n^{-\frac{j}{2k+1}}), \end{aligned}$$

if  $j \geq k + 1$ . Here the third line follows from Lemma A.5. This implies  $\hat{\Psi}_{k,n,1}(u) = o_p(n^{-\frac{k}{2k+1}})$ , uniformly on  $\{u : |u - x_0| \leq Mn^{-1/(2k+1)}\}$ . Using the same expansion in the first term on the right hand side of (A.3), we arrive at

$$\begin{aligned}
& \underbrace{\hat{f}_n(u) - f_0(x_0)}_{(1)} \\
&= f_0(x_0) \underbrace{\sum_{j=1}^k \binom{-r}{j} \frac{1}{[g_0(x_0)]^j} \sum_{r=1}^j \binom{j}{r} [\hat{g}_n(u) - g_0(x_0) - (u - x_0)g_0'(x_0)]^r (u - x_0)^{j-r} g_0(x_0)^{j-r}}_{(2)} \\
&\quad + \underbrace{f_0(x_0) \sum_{j=1}^k \binom{-r}{j} \left(\frac{g_0'(x_0)}{g_0(x_0)}\right)^j (u - x_0)^j}_{(3)} + \underbrace{f_0(x_0) \hat{\Psi}_{k,n,1}(u)}_{(4)}.
\end{aligned}$$

By Lemma A.4, we see that  $\hat{e}_n(u) = (1) - (3) = (2) + (4) = O_p(s_n^k)$  uniformly on  $\{u : |u - x_0| \leq Mn^{-1/(2k+1)}\}$ . This yields the desired result.  $\square$

We are now ready for the proof of Theorem 5.4.

PROOF OF THEOREM 5.4. For the first assertion, note that

$$\begin{aligned}
& [f_0(x_0)]^{-1} \left( \hat{f}_n(u) - \sum_{j=0}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u-x_0)^j \right) \\
&= [f_0(x_0)]^{-1} \left( \hat{f}_n(u) - f_0(x_0) - \sum_{j=1}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u-x_0)^j \right) \\
&= [f_0(x_0)]^{-1} \left( f_0(x_0) \left( \sum_{j=1}^k \binom{-r}{j} \left( \frac{\hat{g}_n(u)}{g_0(x_0)} - 1 \right)^j + \hat{\Psi}_{k,n,1}(u) \right) - \sum_{j=1}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u-x_0)^j \right) \\
&\quad \text{by (A.3)} \\
&= \hat{\Psi}_{k,n,1}(u) + \sum_{j=1}^k \binom{-r}{j} \left( \frac{\hat{g}_n(u)}{g_0(x_0)} - 1 \right)^j - [f_0(x_0)]^{-1} \sum_{j=1}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u-x_0)^j \\
&= \hat{\Psi}_{k,n,1}(u) + \binom{-r}{1} \left( \frac{\hat{g}_n(u)}{g_0(x_0)} - 1 \right) - \frac{1}{f_0(x_0)} f_0'(x_0) (u-x_0) \\
&\quad + \sum_{j=2}^k \binom{-r}{j} \left( \frac{\hat{g}_n(u)}{g_0(x_0)} - 1 \right)^j - [f_0(x_0)]^{-1} \sum_{j=2}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u-x_0)^j \\
&= \hat{\Psi}_{k,n,1}(u) - \frac{r}{g_0(x_0)} \left( \hat{g}_n(u) - g_0(x_0) - g_0'(x_0)(u-x_0) \right) + \sum_{j=2}^k \binom{-r}{j} \left( \frac{\hat{g}_n(u)}{g_0(x_0)} - 1 \right)^j \\
&\quad - [f_0(x_0)]^{-1} \sum_{j=2}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u-x_0)^j \\
&= -\frac{r}{g_0(x_0)} \left( \hat{g}_n(u) - g_0(x_0) - g_0'(x_0)(u-x_0) \right) + \hat{\Psi}_{k,n,2}(u),
\end{aligned}$$

where

$$\hat{\Psi}_{k,n,2}(u) := \hat{\Psi}_{k,n,1}(u) + \sum_{j=2}^k \binom{-r}{j} \left( \frac{\hat{g}_n(u)}{g_0(x_0)} - 1 \right)^j - [f_0(x_0)]^{-1} \sum_{j=2}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u-x_0)^j.$$

Now we calculate

$$\begin{aligned}
& \int_{\mathbf{I}_{n,x_0}} \int_{x_0}^v \hat{\Psi}_{k,n,2}(u) \, dudv \\
&= \frac{1}{2} t^2 n^{-\frac{2}{2k+1}} \sup_{u \in \mathbf{I}_{n,x_0}} \left| \hat{\Psi}_{k,n,1}(u) \right| + \sum_{j=2}^k \binom{-r}{j} \int_{\mathbf{I}_{n,x_0}} \int_{x_0}^v \left( \frac{\hat{g}_n(u)}{g_0(x_0)} - 1 \right)^j \, dudv \\
&\quad - [f_0(x_0)]^{-1} \sum_{j=2}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} \int_{\mathbf{I}_{n,x_0}} \int_{x_0}^v (u-x_0)^j \, dudv \\
&= o_p(r_n^{-1}) + \sum_{j=2}^k \binom{-r}{j} \left( \frac{g'_0(x_0)}{g_0(x_0)} \right)^j \int_{\mathbf{I}_{n,x_0}} \int_{x_0}^v (u-x_0)^j \, dudv \\
&\quad - \sum_{j=2}^{k-1} \binom{-r}{j} \left( \frac{g'_0(x_0)}{g_0(x_0)} \right)^j \int_{\mathbf{I}_{n,x_0}} \int_{x_0}^v (u-x_0)^j \, dudv \\
&\quad + \left( \sum_{j=2}^k \binom{-r}{j} \frac{1}{[g_0(x_0)]^j} \right. \\
&\quad \quad \left. \times \int_{\mathbf{I}_{n,x_0}} \int_{x_0}^v \sum_{l=1}^j \binom{j}{l} (\hat{g}_n(u) - g_0(x_0) - g'_0(x_0)(u-x_0))^l (u-x_0)^{j-l} [g'_0(x_0)]^{j-l} \, dudv \right) \\
&= o_p(r_n^{-1}) + \binom{-r}{k} \left( \frac{g'_0(x_0)}{g_0(x_0)} \right)^k \int_{\mathbf{I}_{n,x_0}} \int_{x_0}^v (u-x_0)^k \, dudv \\
&\quad + \left( \sum_{j=2}^k \binom{-r}{j} \frac{1}{[g_0(x_0)]^j} \right. \\
&\quad \quad \left. \times \int_{\mathbf{I}_{n,x_0}} \int_{x_0}^v \sum_{l=1}^j \binom{j}{l} (\hat{g}_n(u) - g_0(x_0) - g'_0(x_0)(u-x_0))^l (u-x_0)^{j-l} [g'_0(x_0)]^{j-l} \, dudv \right) \\
&:= o_p(r_n^{-1}) + (2) + (1).
\end{aligned}$$

Consider (1): for each  $(j, l)$  satisfying  $1 \leq l \leq j \leq k$  and  $j \geq 2$ , we have

$$\begin{aligned}
(1) &: r_n \int_{\mathbf{I}_{n,x_0}} \int_{x_0}^v (\hat{g}_n(u) - g_0(x_0) - g'_0(x_0)(u-x_0))^l (u-x_0)^{j-l} [g'_0(x_0)]^{j-l} \, dudv \\
&= n^{\frac{k+2}{2k+1}} \cdot O(n^{-\frac{2}{2k+1}}) \cdot O_p(n^{-\frac{kl}{2k+1}}) \cdot O_p(n^{-\frac{j-l}{2k+1}}) = O_p(n^{-\frac{k(l-1)+(j-l)}{2k+1}}) = o_p(1).
\end{aligned}$$

Consider (2) as follows:

$$\begin{aligned} (2) &= \binom{-r}{k} \left( \frac{g'_0(x_0)}{g_0(x_0)} \right)^k \int_{I_{n,x_0}} \int_{x_0}^v (u-x_0)^k \, dudv \\ &= \frac{1}{(k+1)(k+2)} \binom{-r}{k} \left( \frac{g'_0(x_0)}{g_0(x_0)} \right)^k t^{k+2} r_n^{-1}. \end{aligned}$$

Hence we have

$$r_n \int_{I_{n,x_0}} \int_{x_0}^v \hat{\Psi}_{k,n,2}(u) \, dudv = \frac{1}{(k+1)(k+2)} \binom{-r}{k} \left( \frac{g'_0(x_0)}{g_0(x_0)} \right)^k t^{k+2} + o_p(1).$$

Note by definition we have

$$(A.4) \quad \mathbb{Y}_n^{\text{locmod}}(t) = \frac{\mathbb{Y}_n^{\text{loc}}(t)}{f_0(x_0)} - r_n \int_{I_{n,x_0}} \int_{x_0}^v \hat{\Psi}_{k,n,2}(u) \, dudv.$$

Let  $n \rightarrow \infty$ , by the same calculation in the proof of Theorem 6.2 [18], we have

$$\begin{aligned} \mathbb{Y}_n^{\text{locmod}}(t) &\rightarrow_d \frac{1}{\sqrt{f_0(x_0)}} \int_0^t W(s) \, ds \\ &\quad + \left[ \frac{f_0^{(k)}(x_0)}{(k+2)! f_0(x_0)} - \frac{1}{(k+1)(k+2)} \binom{-r}{k} \left( \frac{g'_0(x_0)}{g_0(x_0)} \right)^k \right] t^{k+2} \\ &= \frac{1}{\sqrt{f_0(x_0)}} \int_0^t W(s) \, ds - \frac{r g_0^{(k)}(x_0)}{g_0(x_0)(k+2)!} t^{k+2}, \end{aligned}$$

where the last line follows from Lemma A.4. Now we turn to the second assertion. It is easy to check by the definition of  $\hat{\Psi}_{k,n,2}(\cdot)$  that

$$(A.5) \quad \mathbb{H}_n^{\text{locmod}}(t) = \frac{\mathbb{H}_n^{\text{loc}}(t)}{f_0(x_0)} - r_n \int_{I_{n,x_0}} \int_{x_0}^v \hat{\Psi}_{k,n,2}(u) \, dudv.$$

On the other hand, simple calculation yields that  $\mathbb{Y}_n^{\text{loc}}(t) - \mathbb{H}_n^{\text{loc}}(t) = r_n (\mathbb{H}_n(x_0 + s_n t) - \hat{H}_n(x_0 + s_n t)) \geq 0$  where the inequality follows from Theorem 2.11. Combined with (A.4) and (A.5) we have shown the second assertion. Finally we show tightness of  $\{\hat{A}_n\}$  and  $\{\hat{B}_n\}$ . By Theorem A.2, we can find  $M > 0$  and  $\tau \in \mathcal{S}(\hat{g}_n)$  such that  $0 \leq \tau - x_0 \leq M n^{-1/(2k+1)}$  with large probability.

Now note

$$\begin{aligned}
\left| \hat{A}_n \right| &\leq r_n s_n \left| (\hat{F}_n(x_0) - \hat{F}_n(\tau)) - (\mathbb{F}_n(x_0) - \mathbb{F}_n(\tau)) \right| + \frac{r_n s_n}{n} \\
&\leq r_n s_n \left| \int_{x_0}^{\tau} \left( \hat{f}_n(u) - \sum_{j=0}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u - x_0)^j \right) du \right| \\
&\quad + r_n s_n \left| \int_{x_0}^{\tau} \left( \sum_{j=0}^{k-1} \frac{f_0^{(j)}(x_0)}{j!} (u - x_0)^j - f_0(u) \right) du \right| \\
&\quad + r_n s_n \left| \int_{x_0}^{\tau} d(\mathbb{F}_n - F_0) \right| + n^{-k/(2k+1)} \\
&=: \hat{A}_{n1} + \hat{A}_{n2} + \hat{A}_{n3} + n^{-k/(2k+1)}.
\end{aligned}$$

We calculate three terms respectively.

$$\begin{aligned}
\hat{A}_{n1} &\leq r_n s_n \left| \int_{x_0}^{\tau} \hat{\epsilon}_n(u) du \right| + r_n s_n \left| \int_{x_0}^{\tau} f_0(x_0) \binom{-r}{k} \left( \frac{g_0'(x_0)}{g_0(x_0)} \right)^k (u - x_0)^k du \right| \\
&= O_p(r_n s_n \cdot s_n^{k+1}) + o_p(r_n s_n \cdot s_n^{k+1}) = O_p(1), \quad \text{by Lemma A.6} \\
\hat{A}_{n2} &\leq r_n s_n \left| \int_{x_0}^{\tau} \frac{f_0^{(k)}(x_0)}{k!} (u - x_0)^k du \right| + r_n s_n \left| \int_{x_0}^{\tau} (u - x_0)^k \epsilon_n(u) du \right| \\
&= O_p(1), \quad \text{since } \|\epsilon_n\|_{\infty} \rightarrow_p 0 \text{ as } x_0 - \tau \rightarrow_p 0.
\end{aligned}$$

For  $\hat{A}_{n3}$ , we follow the lines of Lemma 4.1 [1] again to conclude. Fix  $R > 0$ , and consider the function class  $\mathcal{F}_{x_0, R} := \{\mathbf{1}_{[x_0, y]} : x_0 \leq y \leq x_0 + R\}$ . Then  $F_{x_0, R}(z) := \mathbf{1}_{[x_0, x_0 + R]}(z)$  is an envelop function for  $\mathcal{F}_{x_0, R}$ , and  $\mathbb{E}F_{x_0, R}^2 = \int_{x_0}^{x_0 + R} dz = R$ . Now let  $s = k, d = 1$  in Lemma 4.1 [1], we have

$$\hat{A}_{n3} = \left| \int_{x_0}^{\tau} d(\mathbb{F}_n - F_0)(z) \right| \leq |\tau - x_0|^{k+1} + O_p(1) n^{-\frac{k+1}{2k+1}} = O_p(1).$$

This completes the proof for tightness for  $\{A_n\}$ .  $\{B_n\}$  follows from similar argument so we omit the details.  $\square$

PROOF OF THEOREM 4.4. The proof is essentially the same as that of Theorem 3.6 [1].  $\square$

LEMMA A.7. *Assume (A1)-(A4). Then*

$$\int_{-\infty}^{\infty} \tilde{f}_{\epsilon}(x) dx = 1 + \pi_k \frac{r g^{(k)}(m_0)}{g(m_0)^{r+1}} \epsilon^{k+1} + o(\epsilon^{k+1}),$$

where

$$\pi_k = \frac{1}{(k+1)!} \left[ 3^{k-1}(2k^2 - 4k + 3) + 2k^2 - 1 \right].$$

PROOF OF LEMMA A.7. This is straightforward calculation by Taylor expansion. Note that

$$\begin{aligned} \int_{-\infty}^{\infty} \tilde{g}_\epsilon^{-r}(x) \, dx &= \int_{-\infty}^{\infty} (\tilde{g}_\epsilon^{-r}(x) - g^{-r}(x)) \, dx + 1 \\ &= \int_{m_0 - c_\epsilon \epsilon}^{m_0 - \epsilon} (\tilde{g}_\epsilon^{-r}(x) - g^{-r}(x)) \, dx \\ &\quad + \int_{m_0 - \epsilon}^{m_0 + \epsilon} (\tilde{g}_\epsilon^{-r}(x) - g^{-r}(x)) \, dx + 1 \\ &:= I + II + 1. \end{aligned}$$

For  $y > x$ , we have  $x^{-r} - y^{-r} = \sum_{n \geq 1}^{\infty} \binom{-r}{n} (-1)^n (y-x)^n y^{-r-n}$ . Now for the first term above, we continue our calculation of its leading term by noting

(A.6)

$$\begin{aligned} &g(x) - \tilde{g}_\epsilon(x) \\ &= g(x) - g(m_0 - c_\epsilon \epsilon) - (x - m_0 + c_\epsilon \epsilon) g'(m_0 - c_\epsilon \epsilon) \\ &= g(m_0) + \frac{g^{(k)}(m_0)}{k!} (x - m_0)^k - \left[ g(m_0) + \frac{g^{(k)}(m_0)}{k!} (-c_\epsilon \epsilon)^k \right] \\ &\quad - (x - m_0 + c_\epsilon \epsilon) \frac{g^{(k)}(m_0)}{(k-1)!} (-c_\epsilon \epsilon)^{k-1} + \text{higher order terms} \\ &= \frac{g^{(k)}(m_0)}{k!} \left[ (x - m_0)^k - c_\epsilon^k \epsilon^k + k c_\epsilon^{k-1} \epsilon^{k-1} (x - m_0 + c_\epsilon \epsilon) \right] + \text{higher order terms.} \end{aligned}$$

Here we used the fact  $k$  is an even number, as shown in Lemma A.1. Thus we have

$$\begin{aligned} &\text{leading term of I} \\ &= \int_{m_0 - c_\epsilon \epsilon}^{m_0 - \epsilon} r \left( g(x) - g(m_0 - c_\epsilon \epsilon) - (x - m_0 + c_\epsilon \epsilon) g'(m_0 - c_\epsilon \epsilon) \right) g(x)^{-r-1} \, dx \\ &= \frac{r g^{(k)}(m_0)}{k! g(m_0)^{r+1}} \int_{m_0 - c_\epsilon \epsilon}^{m_0 - \epsilon} \left[ (x - m_0)^k - c_\epsilon^k \epsilon^k + k c_\epsilon^{k-1} \epsilon^{k-1} (x - m_0 + c_\epsilon \epsilon) \right] \, dx + o(\epsilon^{k+1}) \\ &= \alpha_k \frac{r g^{(k)}(m_0)}{g(m_0)^{r+1}} \epsilon^{k+1} + o(\epsilon^{k+1}). \end{aligned}$$

Here

$$\alpha_k = \frac{1}{(k+1)!} \left[ 3^{k-1}(2k^2 - 4k + 3) - 1 \right].$$

For the second term,

$$\begin{aligned}
\text{(A.7)} \quad & g(x) - \tilde{g}_\epsilon(x) \\
&= g(x) - g(m_0 + \epsilon) - (x - m_0 - \epsilon)g'(m_0 + \epsilon) \\
&= \frac{g^{(k)}(m_0)}{k!} \left[ (x - m_0)^k - \epsilon^k - k\epsilon^{k-1}(x - m_0 - \epsilon) \right] + \text{higher order terms.}
\end{aligned}$$

Now similar calculations yield that the second term =  $\beta_k \frac{rg^{(k)}(m_0)}{g(m_0)^{r+1}} \epsilon^{k+1} + o(\epsilon^{k+1})$  with

$$\beta_k = \frac{2k^2}{(k+1)!}.$$

This gives the conclusion.  $\square$

PROOF OF LEMMA 4.6. By definition of the Hellinger metric and Lemma A.7, we have

$$\begin{aligned}
2h^2(f_\epsilon, f) &= \int_{-\infty}^{\infty} (\sqrt{f_\epsilon(x)} - \sqrt{f(x)})^2 dx \\
&= \int_{-\infty}^{\infty} \left( \tilde{g}_\epsilon^{-r/2}(x) \left( 1 - \frac{\pi_k rg^{(k)}(m_0)}{2g(m_0)^{r+1}} \epsilon^{k+1} + o(\epsilon^{k+1}) \right) - g^{-r/2}(x) \right)^2 dx \\
&\equiv \int_{-\infty}^{\infty} \left( \tilde{g}_\epsilon^{-r/2}(x)(1 + \eta_k(\epsilon)) - g^{-r/2}(x) \right)^2 dx
\end{aligned}$$

since

$$\begin{aligned}
f_\epsilon(x) &= \tilde{g}_\epsilon^{-r}(x) \left( 1 + \pi_k \frac{rg^{(k)}(m_0)}{g(m_0)^{r+1}} \epsilon^{k+1} + o(\epsilon^{k+1}) \right)^{-1} \\
&= \tilde{g}_\epsilon^{-r}(x) \left( 1 - \pi_k \frac{rg^{(k)}(m_0)}{g(m_0)^{r+1}} \epsilon^{k+1} + o(\epsilon^{k+1}) \right).
\end{aligned}$$

Here  $\eta_k(\epsilon) = O(\epsilon^{k+1})$ . Splitting two terms apart in the above integral we get

$$\begin{aligned}
2h^2(f_\epsilon, f) &= \int_{-\infty}^{\infty} \left( \tilde{g}_\epsilon^{-r/2}(x) - g^{-r/2}(x) + \eta_k(\epsilon) \tilde{g}_\epsilon^{-r/2}(x) \right)^2 dx \\
&= \int_{-\infty}^{\infty} (\tilde{g}_\epsilon^{-r/2}(x) - g^{-r/2}(x))^2 dx + (\eta_k(\epsilon))^2 \int_{-\infty}^{\infty} \tilde{g}_\epsilon^{-r}(x) dx \\
&\quad + 2\eta_k(\epsilon) \int_{-\infty}^{\infty} \tilde{g}_\epsilon^{-r/2}(x) (\tilde{g}_\epsilon^{-r/2}(x) - g^{-r/2}(x)) dx \\
&= I + II + III.
\end{aligned}$$

Now for the first term,

$$\begin{aligned}
I &= \int_{m_0 - c_\epsilon \epsilon}^{m_0 + \epsilon} \frac{r^2}{4} [g(x) - \tilde{g}_\epsilon(x)]^2 g(x)^{-r-2} dx + \text{higher order terms} \\
&= \frac{r^2}{4g(m_0)^{r+2}} \int_{m_0 - c_\epsilon \epsilon}^{m_0 + \epsilon} [g(x) - \tilde{g}_\epsilon(x)]^2 dx + \text{higher order terms} \\
&= \frac{r^2}{4g(m_0)^{r+2}} \left( \int_{m_0 - c_\epsilon \epsilon}^{m_0 - \epsilon} + \int_{m_0 - \epsilon}^{m_0 + \epsilon} \right) [g(x) - \tilde{g}_\epsilon(x)]^2 dx + \text{higher order terms} \\
&= I_1 + I_2 + \text{higher order terms.}
\end{aligned}$$

By (A.6) and (A.7) we see that for  $i = 1, 2$ ,

$$\begin{aligned}
I_i &= \frac{r^2}{4g(m_0)^{r+2}} \int_{\mathcal{I}_i} [g(x) - \tilde{g}_\epsilon(x)]^2 dx \\
&= \zeta_k^{(i)} \frac{r^2 f(m_0) g^{(k)}(m_0)^2}{g(m_0)^2} \epsilon^{2k+1} + o(\epsilon^{2k+1}).
\end{aligned}$$

Here  $\mathcal{I}_1 = [m_0 - c_\epsilon \epsilon, m_0 - \epsilon]$ ,  $\mathcal{I}_2 = [m_0 - \epsilon, m_0 + \epsilon]$ , and

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

On the other hand,  $II = O(\epsilon^{2k+2}) = o(\epsilon^{2k+1})$  and  $|III| \leq O(\epsilon^{k+1} \cdot \epsilon^{(2k+1)/2} \cdot \epsilon^{(2k+2)/2}) = o(\epsilon^{2k+1})$  by Cauchy-Schwarz. This completes the proof.  $\square$

#### A.4. Supplementary Proofs for Section 5.

PROOF OF LEMMA 5.1. First assume  $\gamma_b \neq \gamma_a$  and  $\gamma_a, \gamma_b \in (0, \infty)$ . For notational convenience we drop explicit dependence on  $n$  and the limit is taken as  $\alpha \nearrow 1$ . Let  $\gamma_{a,\alpha} = f_\alpha(a) = g_\alpha(a)^{1/(\alpha-1)}$  and  $\gamma_{b,\alpha} = f_\alpha(b) = g_\alpha(b)^{1/(\alpha-1)}$ . For any  $x \in [a, b]$ ,

(A.8)

$$\begin{aligned}
\lim_{\alpha \rightarrow 1} \log f_\alpha(x) &= \lim_{\alpha \rightarrow 1} \frac{1}{\alpha - 1} \log \left( \frac{\gamma_{b,\alpha}^{\alpha-1} - \gamma_{a,\alpha}^{\alpha-1}}{b - a} (x - a) + \gamma_{a,\alpha}^{\alpha-1} \right) \\
&= \lim_{\alpha \rightarrow 1} \frac{1}{\alpha - 1} \log \left( \frac{\gamma_b^{\alpha-1} - \gamma_a^{\alpha-1}}{b - a} (x - a) \cdot \frac{\gamma_{b,\alpha}^{\alpha-1} - \gamma_{a,\alpha}^{\alpha-1}}{\gamma_b^{\alpha-1} - \gamma_a^{\alpha-1}} + \gamma_{a,\alpha}^{\alpha-1} \right) \\
&\equiv \log \gamma_a + \lim_{\alpha \rightarrow 1} \frac{1}{\alpha - 1} \log \left( (\gamma_b^{\alpha-1} - \gamma_a^{\alpha-1}) \frac{(x - a)}{(b - a)} \cdot \frac{1}{\gamma_{a,\alpha}^{\alpha-1}} \cdot r_\alpha + 1 \right).
\end{aligned}$$

Since  $\gamma_{a,\alpha}^{\alpha-1} \rightarrow 1$ , we claim that it suffices to show that

$$(A.9) \quad r_\alpha \equiv \frac{\gamma_{b,\alpha}^{\alpha-1} - \gamma_{a,\alpha}^{\alpha-1}}{\gamma_b^{\alpha-1} - \gamma_a^{\alpha-1}} \rightarrow 1 \quad \text{as } \alpha \rightarrow 1.$$

To see this, assume without loss of generality that  $\gamma_a > \gamma_b$  and hence  $\gamma_b^{\alpha-1} - \gamma_a^{\alpha-1} > 0$ . Suppose that (A.9) holds and let  $\epsilon > 0$ . Then the second term on right hand side of (A.8) can be bounded from above by

$$\begin{aligned} & \lim_{\alpha \nearrow 1} \frac{1}{\alpha - 1} \log \left( (\gamma_b^{\alpha-1} - \gamma_a^{\alpha-1}) \frac{(x-a)}{(b-a)} (1-\epsilon) + 1 \right) \\ &= \lim_{\alpha \nearrow 1} (\log \gamma_b \cdot \gamma_b^{\alpha-1} - \log \gamma_a \cdot \gamma_a^{\alpha-1}) \frac{(x-a)}{(b-a)} (1-\epsilon) \\ &= (\log \gamma_b - \log \gamma_a) \frac{(x-a)}{(b-a)} (1-\epsilon) \end{aligned}$$

where the second line follows from L'Hospital's rule. Similarly we can derive a lower bound:

$$(\log \gamma_b - \log \gamma_a) \frac{(x-a)}{(b-a)} (1+\epsilon).$$

Thus it remains to show that (A.9) holds. But we can rewrite  $r_\alpha$  as

$$\begin{aligned} r_\alpha &= \frac{c_\alpha^{\alpha-1} - 1}{c^{\alpha-1} - 1} \\ &= \frac{c^{\alpha-1} (c_\alpha/c)^{\alpha-1} - (c_\alpha/c)^{\alpha-1} + (c_\alpha/c)^{\alpha-1} - 1}{c^{\alpha-1} - 1} \\ &= (c_\alpha/c)^{\alpha-1} + \frac{(c_\alpha/c)^{\alpha-1} - 1}{c^{\alpha-1} - 1} \\ &\rightarrow 1 + 0 \quad \text{as } \alpha \rightarrow 1 \end{aligned}$$

since  $\log((c_\alpha/c)^{\alpha-1}) = (\alpha-1) \log(c_\alpha/c) \rightarrow 0 \cdot \log 1 = 0$ , and where the second limit follows from an upper and lower bound argument using  $c_\alpha/c \rightarrow 1$ . where  $c_\alpha := \gamma_{b,\alpha}/\gamma_{a,\alpha}$  and  $c = \gamma_b/\gamma_a \neq 1$ .

This shows that (A.9) holds, thereby proving the case for  $\gamma_a \neq \gamma_b \in (0, \infty)$ . For the case  $\gamma_b = \gamma_a \in (0, \infty)$ , similarly we have

$$\lim_{\alpha \rightarrow 1} \log f_\alpha(x) = \log \gamma_a + \lim_{\alpha \rightarrow 1} \frac{1}{\alpha - 1} \log \left( \frac{c_\alpha^{\alpha-1} - 1}{b-a} (x-a) + 1 \right).$$

The second term is 0 by an argument much as above by observing  $c_\alpha = \gamma_{b,\alpha}/\gamma_{a,\alpha} \rightarrow \gamma_b/\gamma_a = 1$ . Finally, if  $\gamma_a \wedge \gamma_b = 0$ , then by the first line of (A.8) we see that  $\log f_\alpha(x) \rightarrow -\infty$ ; if  $\gamma_a \vee \gamma_b = \infty$ , then again  $\log f_\alpha(x) \rightarrow \infty$ .  $\square$

LEMMA A.8. *Let  $\nu$  be a probability measure with  $s$ -concave density  $f$ , and  $x_0, \dots, x_d \in \mathbb{R}^d$  be  $d+1$  points such that  $\Delta := \text{conv}(\{x_0, \dots, x_d\})$  is non-void. If  $f(x_0) \leq (\frac{1}{d} \sum_{i=1}^d f^s(x_i))^{1/s}$ , then*

$$f(x_0) \leq \bar{g}^{-r} \left( 1 - \frac{d}{r} + \frac{d \lambda_d(\Delta) \bar{g}^{-r}}{r \nu(\Delta)} \right)^{-r},$$

where  $\bar{g} := \frac{1}{d} \sum_{j=1}^d f^s(x_j)$ .

PROOF OF LEMMA A.8. For any point  $x \in \Delta$ , we can find some  $u = (u_1, \dots, u_d) \in \Delta_d = \{u : \sum_{i=1}^d u_i \leq 1\}$  such that  $x(u) = \sum_{i=0}^d u_i x_i$ . Here  $u_0 := 1 - \sum_{i=1}^d u_i \geq 0$ . We use the following representation of integration on the unit simplex  $\Delta_d$ : For any measurable function  $h : \Delta_d \rightarrow [0, \infty)$ , we have  $\int_{\Delta_d} h(u) \, du = \frac{1}{d!} \mathbb{E} h(B_1, \dots, B_d)$ , where  $B_i = E_i / \sum_{j=0}^d E_j$  with independent, standard exponentially distributed random variables  $E_0, \dots, E_d$ .

$$\begin{aligned} \frac{\nu(\Delta)}{\lambda_d(\Delta)} &= \frac{1}{\lambda_d(\Delta_d)} \int_{\Delta_d} g(x(u))^{-r} \, du = \mathbb{E} g\left(\sum_{j=0}^d B_j x_j\right)^{-r} \\ &\geq \mathbb{E} \left( \sum_{j=0}^d B_j g(x_j) \right)^{-r} = \mathbb{E} \left( B_0 g_0 + (1 - B_0) \sum_{i=1}^d \tilde{B}_i g(x_i) \right)^{-r}, \end{aligned}$$

where  $\tilde{B}_i := E_i / \sum_{j=1}^d E_j$  for  $1 \leq i \leq d$ . Following [11], it is known that  $B_0$  and  $\{\tilde{B}_i\}_{i=1}^d$  are independent, and  $\mathbb{E}[\tilde{B}_i] = 1/d$ . Hence it follows from Jensen's inequality that

$$\begin{aligned} \frac{\nu(\Delta)}{\lambda_d(\Delta)} &\geq \mathbb{E} \left[ \mathbb{E} \left( B_0 g_0 + (1 - B_0) \sum_{i=1}^d \tilde{B}_i g(x_i) \right)^{-r} \middle| B_0 \right] \\ &\geq \mathbb{E} \left( B_0 g_0 + (1 - B_0) \frac{1}{d} \sum_{i=1}^d g(x_i) \right)^{-r} \\ &= \mathbb{E} (B_0 g_0 + (1 - B_0) \bar{g})^{-r} \\ &= \int_0^1 d(1-t)^{d-1} (t g_0 + (1-t) \bar{g})^{-r} \, dt \\ &= \bar{g}^{-r} \int_0^1 d(1-t)^{d-1} \left( 1 - st \left( (-1/s) \left( \frac{g_0}{\bar{g}} - 1 \right) \right) \right) \, dt \\ &= \bar{g}^{-r} J_{d,s} \left( -\frac{1}{s} \left( \frac{g_0}{\bar{g}} - 1 \right) \right), \end{aligned}$$

where

$$J_{d,s}(y) = \int_0^1 d(1-t)^{d-1}(1-syt)^{1/s} dt.$$

We claim that

$$J_{d,s}(y) \geq \int_0^1 d(1-t)^{d-1}(1-t)^y dt = \frac{d}{d+y},$$

holds for  $s < 0, y > 0$ . To see this, we write  $(1-syt)^{1/s} = (1+yt/r)^{-(r/y)y}$ . Then we only have to show  $(1+yt/r)^{-r/y} \geq (1-t)$  for  $0 \leq t \leq 1$ , or equivalently  $(1+bt) \leq (1-t)^{-b}$  where we let  $b = y/r$ . Let  $g(t) := (1-t)^{-b} - (1+bt)$ . It is easy to verify that  $g(0) = 0$ ,  $g'(t) = b(1-t)^{-b-1} - b$  with  $g'(0) = 0$ , and  $g''(t) = b(b+1)(1-t)^{-b-2} \geq 0$ . Integrating  $g''$  twice yields  $g(t) \geq 0$ , and hence we have verified the claim. Now we proceed with the calculation

$$\frac{\nu(\Delta)}{\lambda_d(\Delta)} \geq \bar{g}^{-r} J_{d,s} \left( -\frac{1}{s} \left( \frac{g_0}{\bar{g}} - 1 \right) \right) \geq \bar{g}^{-r} \frac{d}{d - \frac{1}{s} \left( \frac{g_0}{\bar{g}} - 1 \right)}.$$

Solving for  $g_0$  and replacing  $-1/s = r$  proves the desired inequality.  $\square$

PROOF OF LEMMA 5.2. For fixed  $j \in \{0, \dots, d\}$ , note  $|\det(x_i - x_j) : i \neq j| = |\det X|$  where  $X = \begin{pmatrix} x_0 & \dots & x_d \\ 1 & \dots & 1 \end{pmatrix}$ . Also for each  $y \in \mathbb{R}^d$ , since  $\Delta = \text{conv}(\{x_0, \dots, x_d\})$  is non-void,  $y$  must be in the affine hull of  $\Delta$  and hence we can write  $y = \sum_{i=0}^d \lambda_i x_i$  with  $\sum_{i=0}^d \lambda_i = 1$  (not necessary non-negative), i.e.  $\lambda = X^{-1} \begin{pmatrix} y \\ 1 \end{pmatrix}$ . Let  $\Delta_j(y) := \text{conv}(\{x_i : i \neq j\} \cup \{y\})$ . Then

$$\begin{aligned} \lambda_d(\Delta_j(y)) &= \frac{1}{d!} \left| \det \begin{pmatrix} x_0 & \dots & x_{j-1} & y & x_{j+1} & \dots & x_d \\ 1 & \dots & 1 & 1 & 1 & \dots & 1 \end{pmatrix} \right| \\ &= \frac{1}{d!} |\lambda_j| |\det X| = |\lambda_j| \lambda_d(\Delta). \end{aligned}$$

Hence,

$$\begin{aligned} \max_{0 \leq j \leq d} \lambda_d(\Delta_j(y)) &\geq \lambda_d(\Delta) \max_j |\lambda_j| = \lambda_d(\Delta) \|X^{-1} \begin{pmatrix} y \\ 1 \end{pmatrix}\|_\infty \\ &\geq \lambda_d(\Delta) (d+1)^{-1/2} \|X^{-1} \begin{pmatrix} y \\ 1 \end{pmatrix}\| \\ &\geq \lambda_d(\Delta) (d+1)^{-1/2} \sigma_{\max}(X)^{-1} (1 + \|y\|^2)^{1/2} = C(1 + \|y\|^2)^{1/2}. \end{aligned}$$

Now the conclusion follows from Lemma A.8 by noting

$$f(y) \leq \bar{g}_j^{-r} \left( 1 - \frac{d}{r} + \frac{d \lambda_d(\Delta_j(y)) \bar{g}_j^{-r}}{\nu(\Delta_j(y))} \right)^{-r} \leq f_{\max} \left( 1 - \frac{d}{r} + \frac{d}{r} f_{\min} C(1 + \|y\|^2)^{1/2} \right)^{-r},$$

since  $\bar{g}_j^{-r} = \left(\frac{1}{d} \sum_{i \neq j} f^s(x_i)\right)^{1/s}$  and hence  $f_{\min} \leq \bar{g}_j^{-r} \leq f_{\max}$ , and the index  $j$  is chosen such that  $\lambda_d(\Delta_j(y))$  is maximized.  $\square$

PROOF OF LEMMA 5.3. The key point that for any point  $x \in B(y, \delta_t)$

$$B(ty, \delta_t) \subset (1-t)B(0, \delta) + tx$$

can be shown in the same way as in the proof of Lemma 4.2 [30]. Namely, pick any  $w \in B(ty, \delta_t)$ , let  $v := (1-t)^{-1}(w - tx)$ , then since

$$\|v\| = (1-t)^{-1}\|w - tx\| = (1-t)^{-1}\|w - ty + t(y - x)\| \leq (1-t)^{-1}(\delta_t + t\delta_t) = \delta,$$

and hence  $v \in B(0, \delta)$ . This implies that  $w = (1-t)v + tx \in (1-t)B(0, \delta) + tx$ , as desired. By  $s$ -concavity of  $f$ , we have

$$\begin{aligned} f(w) &\geq \left((1-t)f(v)^s + tf(x)^s\right)^{1/s} \\ &\geq \left((1-t)J_0^s + tf(x)^s\right)^{1/s} \\ &= J_0 \left(1-t + t \left(\frac{f(x)}{J_0}\right)^s\right)^{1/s}. \end{aligned}$$

Averaging over  $w \in B(ty, \delta_t)$  yields

$$\frac{\nu(B(ty, \delta_t))}{\lambda_d(B(ty, \delta_t))} \geq J_0 \left(1-t + t \left(\frac{f(x)}{J_0}\right)^s\right)^{1/s}.$$

Solving for  $f(x)$  completes the proof.  $\square$

## APPENDIX B: AUXILIARY RESULTS

LEMMA B.1 (Lemma 4.3, [16]). *For any  $\varphi(\cdot) \in \mathcal{G}$  with non-empty domain, and  $\epsilon > 0$ , define*

$$\varphi^{(\epsilon)}(x) := \sup_{(v,c)} (v^T x + c)$$

where the supremum is taken over all pairs of  $(v, c) \in \mathbb{R}^d \times \mathbb{R}$  such that

1.  $\|v\| \leq \frac{1}{\epsilon}$ ;
2.  $\varphi(y) \geq v^T y + c$  holds for all  $y \in \mathbb{R}^d$ .

Then  $\varphi^{(\epsilon)} \in \mathcal{G}$  with Lipschitz constant  $\frac{1}{\epsilon}$ . Furthermore,

$$\varphi^{(\epsilon)} \nearrow \varphi, \text{ as } \epsilon \searrow 0,$$

where the convergence is pointwise for all  $x \in \mathbb{R}^d$ .

LEMMA B.2 (Lemma 2.13, [16]). *Given  $Q \in \mathcal{Q}_0$ , a point  $x \in \mathbb{R}^d$  is an interior point of  $\text{csupp}(Q)$  if and only if*

$$h(Q, x) \equiv \sup\{Q(C) : C \subset \mathbb{R}^d \text{ closed and convex, } x \notin \text{int}(C)\} < 1.$$

Moreover, if  $\{Q_n\} \subset \mathcal{Q}$  converges weakly to  $Q$ , then

$$\limsup_{n \rightarrow \infty} h(Q_n, x) \leq h(Q, x)$$

holds for all  $x \in \mathbb{R}^d$ .

LEMMA B.3. *If  $g \in \mathcal{G}$ , then there exists  $a, b > 0$  such that for all  $x \in \mathbb{R}^d$ ,  $g(x) \geq a\|x\| - b$ .*

PROOF. The proof is essentially the same as for Lemma 1, [9], so we shall omit it.  $\square$

Consider the class of functions

$$\mathcal{G}_M := \left\{ g \in \mathcal{G} : \int g^\beta \, dx \leq M \right\}.$$

LEMMA B.4. *For a given  $g \in \mathcal{G}_M$ , denote  $D_r := D(g, r) := \{g \leq r\}$  to be the level set of  $g(\cdot)$  at level  $r$ , and  $\epsilon := \inf g$ . Then for  $r > \epsilon$ , we have*

$$\lambda(D_r) \leq \frac{M(-s)(r - \epsilon)^d}{(s + 1) \int_0^{r - \epsilon} v^d (v + \epsilon)^{1/s} \, dv},$$

where  $\beta = 1 + 1/s$ , and  $-1 < s < 0$ .

PROOF. For  $u \in [\epsilon, r]$ , by convexity of  $g(\cdot)$ , we have

$$\lambda(D_u) \geq \left( \frac{u - \epsilon}{r - \epsilon} \right)^d \lambda(D_r).$$

This can be seen as follows: Consider the epigraph  $\Gamma_g$  of  $g(\cdot)$ , where  $\Gamma_g = \{(t, x) \in \mathbb{R}^d \times \mathbb{R} : x \geq g(t)\}$ . Let  $x_0 \in \mathbb{R}^d$  be a minimizer of  $g$ . Consider the convex set  $C_r = \text{conv}(\Gamma_g \cap \{g = r\}, (x_0, \epsilon)) \subset \Gamma_g \cap \{g \leq r\}$ . where the inclusion follows from the convexity of  $\Gamma_g$  as a subset of  $\mathbb{R}^{d+1}$ . The claimed inequality follows from

$$\lambda_d(D_u) = \lambda_d(\pi_d(\Gamma_g \cap \{g = u\})) \geq \lambda_d(\pi_d(C_r \cap \{g = u\})) = \left( \frac{u - \epsilon}{r - \epsilon} \right)^d \lambda_d(D_r),$$

where  $\pi_d : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$  is the natural projection onto the first component. Now we do the calculation as follows:

$$\begin{aligned}
M &\geq \int_{D_r} (g(x)^{1/s+1} - r^{1/s+1}) \, dx \\
&= - \left( \frac{1}{s} + 1 \right) \int_{D_r} \left( \int_{\epsilon}^r \mathbf{1}(u \geq g(x)) u^{1/s} \, du \right) \, dx \\
&= - \left( \frac{1}{s} + 1 \right) \int_{\epsilon}^r u^{1/s} \, du \int_{D_r} \mathbf{1}(u \geq g(x)) \, dx \\
&= - \left( \frac{1}{s} + 1 \right) \int_{\epsilon}^r \lambda(D_u) u^{1/s} \, du \\
&\geq - \left( \frac{1}{s} + 1 \right) \int_{\epsilon}^r \left( \frac{u - \epsilon}{r - \epsilon} \right)^d \lambda(D_r) u^{1/s} \, du \\
&= \lambda(D_r) \cdot \frac{(s+1) \int_{\epsilon}^r (u - \epsilon)^d u^{1/s} \, du}{(-s)(r - \epsilon)^d}.
\end{aligned}$$

By a change of variable in the integral we get the desired inequality.  $\square$

LEMMA B.5. *Let  $G$  be a convex set in  $\mathbb{R}^d$  with non-empty interior, and a sequence  $\{y_n\}_{n \in \mathbb{N}}$  with  $\|y_n\| \rightarrow \infty$  as  $n \rightarrow \infty$ . Then there exists  $\{x_1, \dots, x_d\} \subset G$  such that*

$$\lambda_d(\text{conv}(x_1, \dots, x_d, y_{n(k)})) \rightarrow \infty,$$

as  $k \rightarrow \infty$  where  $\{y_{n(k)}\}_{k \in \mathbb{N}}$  is a suitable subsequence of  $\{y_n\}_{n \in \mathbb{N}}$ .

PROOF. Without loss of generality we assume  $0 \in \text{int}(\text{dom}(G))$ , and we first choose a convergence subsequence  $\{y_{n(k)}\}_{k \in \mathbb{N}}$  from  $\{y_n/\|y_n\|\}_{n \in \mathbb{N}}$ . Now if we let  $a := \lim_{k \rightarrow \infty} y_{n(k)}/\|y_{n(k)}\|$ , then  $\|a\| = 1$ . Since  $G$  has non-empty interior,  $\{a^T x = 0\} \cap G$  has non-empty relative interior. Thus we can choose  $x_1, \dots, x_d \subset \{a^T x = 0\} \cap G$  such that  $\lambda_{d-1}(K) \equiv \lambda_{d-1}(\text{conv}(x_1, \dots, x_d)) > 0$ . Note that

$$\text{dist}(y_{n(k)}, \text{aff}(K)) = \text{dist}(y_{n(k)}, \{a^T x = 0\}) = \langle y_{n(k)}, a \rangle = \|y_{n(k)}\| \langle y_{n(k)}/\|y_{n(k)}\|, a \rangle \rightarrow \infty,$$

as  $k \rightarrow \infty$ . Since

$$\lambda_d(\text{conv}(x_1, \dots, x_d, y_{n(k)})) = \lambda_d(\text{conv}(K, y_{n(k)})) = c \lambda_{d-1}(K) \cdot \text{dist}(y_{n(k)}, \text{aff}(K)),$$

for some constant  $c = c(d) > 0$ , the proof is complete as we let  $k \rightarrow \infty$ .  $\square$

LEMMA B.6 (Lemma 4.2, [16]). *Let  $\bar{g}$  and  $\{g_n\}_{n \in \mathbb{N}}$  be functions in  $\mathcal{G}$  such that  $g_n \geq \bar{g}$ , for all  $n \in \mathbb{N}$ . Suppose the set  $C := \{x \in \mathbb{R}^d : \limsup_{n \rightarrow \infty} g_n(x) < \infty\}$  is non-empty. Then there exist a subsequence  $\{g_{n(k)}\}_{k \in \mathbb{N}}$  of  $\{g_n\}_{n \in \mathbb{N}}$ , and a function  $g \in \mathcal{G}$  such that  $C \subset \text{dom}(g)$  and*

$$(B.1) \quad \begin{aligned} \lim_{k \rightarrow \infty, x \rightarrow y} g_{n(k)}(x) &= g(y), \quad \text{for all } y \in \text{int}(\text{dom}(g)), \\ \liminf_{k \rightarrow \infty, x \rightarrow y} g_{n(k)}(x) &\geq g(y), \quad \text{for all } y \in \mathbb{R}^d. \end{aligned}$$

LEMMA B.7. *Let  $\{g_n\}$  be a sequence of non-negative convex functions satisfying the following conditions:*

(A1). *There exists a convex set  $G$  with non-empty interior such that for all  $x_0 \in \text{int}(G)$ , we have  $\sup_{n \in \mathbb{N}} g_n(x_0) < \infty$ .*

(A2). *There exists some  $M > 0$  such that  $\sup_{n \in \mathbb{N}} \int (g_n(x))^\beta dx \leq M < \infty$ .*

*Then there exists  $a, b > 0$  such that for all  $x \in \mathbb{R}^d$  and  $k \in \mathbb{N}$*

$$g_{n(k)}(x) \geq a\|x\| - b,$$

*where  $\{g_{n(k)}\}_{k \in \mathbb{N}}$  is a suitable subsequence of  $\{g_n\}_{n \in \mathbb{N}}$ .*

PROOF. Without loss of generality we may assume  $G$  is contained in all  $\text{int}(\text{dom}(g_n))$ . We first note (A1)-(A2) implies that  $\{\hat{x}_n \in \text{Arg min}_{x \in \mathbb{R}^d} g_n(x)\}_{n=1}^\infty$  is a bounded sequence, i.e.

$$(B.2) \quad \sup_{n \in \mathbb{N}} \|\hat{x}_n\| < \infty,$$

Suppose not, then without loss of generality we may assume  $\|\hat{x}_n\| \rightarrow \infty$  as  $n \rightarrow \infty$ . By Lemma B.5, we can choose  $\{x_1, \dots, x_d\} \subset G$  such that  $\lambda_d(\text{conv}(x_1, \dots, x_d, \hat{x}_{n(k)})) \rightarrow \infty$ , as  $k \rightarrow \infty$  for some subsequence  $\{\hat{x}_{n(k)}\} \subset \{\hat{x}_n\}$ . For simplicity of notation we think of  $\{\hat{x}_n\}$  as such an appropriate subsequence. Denote  $\epsilon_n := \inf_{x \in \mathbb{R}^d} g_n(x)$ , and  $M_2 := \sup_{n \in \mathbb{N}} \epsilon_n \leq \sup_{n \in \mathbb{N}} g_n(x_0) < \infty$  by (A1). Again by (A1) and convexity we may assume that

$$\sup_{x \in \text{conv}(x_1, \dots, x_d, \hat{x}_n)} g_n(x) \leq M_1,$$

holds for some  $M_1 > 0$  and all  $n \in \mathbb{N}$ . This implies that

$$\int g_n^\beta(x) dx \geq M_1^\beta \lambda_d(\text{conv}(x_1, \dots, x_d, \hat{x}_n)) \rightarrow \infty,$$

as  $n \rightarrow \infty$ , which gives a contradiction to (A2). This shows (B.2).

Now we define  $\underline{g}(\cdot)$  be the convex hull of  $\tilde{g}(x) := \inf_{n \in \mathbb{N}} g_n(x)$ , then  $\underline{g} \leq g_n$  holds for all  $n \in \mathbb{N}$ . We claim that  $\underline{g}(x) \rightarrow \infty$  as  $\|x\| \rightarrow \infty$ . By Lemma B.4, for fixed  $\eta > 1$ , we have

$$\begin{aligned} \lambda_d(D(g_n, \eta M_2)) &\leq \frac{M(-s)(\eta M_2 - \epsilon_n)^d}{(s+1) \int_0^{\eta M_2 - \epsilon_n} v^d (v + \epsilon_n)^{1/s} dv} \\ &\leq \frac{M(-s)(\eta M_2)^d}{(s+1) \int_0^{(\eta-1)M_2} v^d (v + M_2)^{1/s} dv} < \infty, \end{aligned}$$

where  $D(g_n, \eta M_2) := \{g_n \leq \eta M_2\}$ . Hence

$$(B.3) \quad \sup_{n \in \mathbb{N}} \lambda_d(D(g_n, \eta M_2)) < \infty.$$

holds for every  $\eta > 1$ . Now combining (B.2) and (B.3), we claim that, for fixed  $\eta$  large enough, it is possible to find  $R = R(\eta) > 0$  such that

$$(B.4) \quad g_n(x) \geq \eta M_2,$$

holds for all  $x \geq R(\eta)$  and  $n \in \mathbb{N}$ . If this is not true, then for all  $k \in \mathbb{N}$ , we can find  $n(k) \in \mathbb{N}$  and  $\bar{x}_k \in \mathbb{R}^d$  with  $\|\bar{x}_k\| \geq k$  such that  $g_{n(k)}(\bar{x}_k) \leq \eta M_2$ . We consider two cases to derive a contradiction.

**[Case 1.]** If for some  $n_0 \in \mathbb{N}$  there exists infinitely many  $k \in \mathbb{N}$  with  $n(k) = n_0$ , then we may assume without loss of generality that we can find some a sequence  $\{\bar{x}_k\}_{k \in \mathbb{N}}$  with  $\|\bar{x}_k\| \rightarrow \infty$  as  $k \rightarrow \infty$ , and  $g_{n_0}(\bar{x}_k) \leq \eta M_2$ . Since the support  $g_{n_0}$  has non-empty interior, by Lemma B.5, we can find  $x_1, \dots, x_d \in \text{supp}(g_{n_0})$  such that  $\lambda_d(\text{conv}(x_1, \dots, x_d, \bar{x}_{k(j)})) \rightarrow \infty$  as  $j \rightarrow \infty$  holds for some subsequence  $\{\bar{x}_{k(j)}\}_{j \in \mathbb{N}}$  of  $\{\bar{x}_k\}_{k \in \mathbb{N}}$ . Let  $\bar{M} := \max_{1 \leq i \leq d} g_{n_0}(x_i)$ , then we find  $\lambda_d(D(g_{n_0}, \bar{M} \vee \eta M_2)) = \infty$ . This contradicts with (B.3).

**[Case 2.]** If  $\#\{k \in \mathbb{N} : n = n(k)\} < \infty$  for all  $n \in \mathbb{N}$ , then without loss of generality we may assume that for all  $k \in \mathbb{N}$ , we can find  $\bar{x}_k \in \mathbb{R}^d$  with  $\|\bar{x}_k\| \geq k$  such that  $g_k(\bar{x}_k) \leq \eta M_2$ . Recall by assumption (A1) convex set  $G$  has non-empty interior, and is contained in the support of  $g_n$  for all  $n \in \mathbb{N}$ . Again by Lemma B.5, we may take  $x_1, \dots, x_d \in G$  such that  $\lambda_d(\text{conv}(x_1, \dots, x_d, \bar{x}_{k(j)})) \rightarrow \infty$  as  $j \rightarrow \infty$  holds for some subsequence  $\{\bar{x}_{k(j)}\}_{j \in \mathbb{N}}$  of  $\{\bar{x}_k\}_{k \in \mathbb{N}}$ . In view of (A1), we conclude by convexity that  $\bar{M} := \max_{1 \leq i \leq d} \sup_{j \in \mathbb{N}} g_{k(j)}(x_i) < \infty$ . This implies

$$\lambda_d(D(g_{n_{k(j)}}, \bar{M} \vee \eta M_2)) \geq \lambda_d(\text{conv}(x_1, \dots, x_d, \bar{x}_{k(j)})) \rightarrow \infty, \quad j \rightarrow \infty,$$

which gives a contradiction.

Combining these two cases we have proved (B.4). This implies that  $\tilde{g}(x) \rightarrow \infty$  as  $\|x\| \rightarrow \infty$ , whence verifying the claim that  $\underline{g}(x) \rightarrow \infty$  as  $\|x\| \rightarrow \infty$ .

Hence in view of Lemma B.3, we find that there exists  $a, b > 0$  such that  $g_n(x) \geq a\|x\| - b$  holds for all  $x \in \mathbb{R}^d$  and  $n \in \mathbb{N}$ .  $\square$

LEMMA B.8. *Assume  $x_0, \dots, x_d \in \mathbb{R}^d$  are in general position. If  $g(\cdot)$  is a non-negative function with  $\Delta \equiv \text{conv}(x_0, \dots, x_d) \subset \text{dom}(g)$ , and  $g(x_0) = 0$ . Then for  $r \geq d$ , we have  $\int_{\Delta} (g(x))^{-r} dx = \infty$ .*

PROOF. We may assume without loss of generality that  $x_0 = 0, x_i = \mathbf{e}_i \in \mathbb{R}^d$ , where  $\mathbf{e}_i$  is the unit directional vector with 1 in its  $i$ -th coordinate and 0 otherwise. Then  $\Delta = \Delta_0 := \{x \in \mathbb{R}^d : \sum_{i=1}^d x_i \leq 1, x_i \geq 0, \forall i = 1, \dots, d\}$ . Denote  $a_i = g(x_i) \geq 0$ . We may assume there is at least one  $a_i \neq 0$ . Then by convexity of  $g$  we find  $g(x) \leq \sum_{i=1}^d a_i x_i$  for all  $x \in \Delta_0$ . This gives

$$\begin{aligned} \int_{\Delta_0} (g(x))^{-r} dx &\geq \int_{\Delta_0} \left( \sum_{i=1}^d a_i x_i \right)^{-r} dx \geq \int_{\Delta_0} \frac{1}{(\max_{i=1, \dots, d} a_i)^r \|x\|_1^r} dx \\ &\geq \frac{1}{(\max_{i=1, \dots, d} a_i)^r d^{r/2}} \int_{C_0} \frac{1}{\|x\|_2^r} dx = \infty, \end{aligned}$$

where  $C_0 := \{\|x\|_2 \leq \frac{1}{\sqrt{d}}\} \cap \{x_i \geq 0, i = 1, \dots, d\}$ . Note we used the fact that  $\|x\|_1 \leq \sqrt{d}\|x\|_2$ .  $\square$

LEMMA B.9 (Theorem 1.11, [4]). *Let  $f_n \rightarrow_d f$ , and  $\mathcal{D}$  be the class of all Borel measurable, convex subsets in  $\mathbb{R}^d$ . Then  $\lim_{n \rightarrow \infty} \sup_{D \in \mathcal{D}} \left| \int_D (f_n - f) \right| = 0$ .*

## ACKNOWLEDGEMENTS

We owe thanks to Charles Doss, Roger Koenker and Richard Samworth for helpful comments, suggestions and minor corrections.

## REFERENCES

- [1] 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 \(2010h:62290\)](#)
- [2] 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 \(2009b:62077\)](#)
- [3] 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 \(2000a:62067\)](#)
- [4] BHATTACHARYA, R. N. and RANGA RAO, R. (1976). *Normal Approximation and Asymptotic Expansions*. John Wiley & Sons, New York-London-Sydney Wiley Series in Probability and Mathematical Statistics. [MR0436272 \(55 #9219\)](#)

- [5] BORELL, C. (1974). Convex measures on locally convex spaces. *Ark. Mat.* **12** 239–252. [MR0388475 \(52 ##9311\)](#)
- [6] BORELL, C. (1975). Convex set functions in  $d$ -space. *Period. Math. Hungar.* **6** 111–136. [MR0404559 \(53 ##8359\)](#)
- [7] 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. Functional Analysis* **22** 366–389. [MR0450480 \(56 ##8774\)](#)
- [8] 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 \(43 ##2807\)](#)
- [9] 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 \(2012f:62071\)](#)
- [10] CULE, M., SAMWORTH, R. and STEWART, M. (2010). Maximum likelihood estimation of a multi-dimensional log-concave density. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **72** 545–607. [MR2758237](#)
- [11] CULE, M. L. and DÜMBGEN, L. (2008). On an auxiliary function for log-density estimation. *arXiv preprint arXiv:0807.4719*.
- [12] DHARMADHIKARI, S. and JOAG-DEV, K. (1988). *Unimodality, Convexity, and Applications. Probability and Mathematical Statistics*. Academic Press, Inc., Boston, MA. [MR954608 \(89k:60020\)](#)
- [13] DOSS, C. and WELLNER, J. A. (2015). Global rates of convergence of the MLEs of log-concave and s-concave densities. *arXiv preprint arXiv:1306.1438v2*.
- [14] DUDLEY, R. M. (2002). *Real Analysis and Probability. Cambridge Studies in Advanced Mathematics* **74**. Cambridge University Press, Cambridge Revised reprint of the 1989 original. [MR1932358\(2003h:60001\)](#)
- [15] 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 \(2011b:62096\)](#)
- [16] DÜMBGEN, L., SAMWORTH, R. and SCHUHMACHER, D. (2011). Approximation by log-concave distributions, with applications to regression. *Ann. Statist.* **39** 702–730. [MR2816336 \(2012e:62039\)](#)
- [17] 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. [MR822052 \(87i:62076\)](#)
- [18] GROENEBOOM, P., JONGBLOED, G. and WELLNER, J. A. (2001). Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist.* **29** 1653–1698. [MR1891742 \(2003a:62047\)](#)
- [19] JONGBLOED, G. (2000). Minimax lower bounds and moduli of continuity. *Statist. Probab. Lett.* **50** 279–284. [MR1792307 \(2001h:62067\)](#)
- [20] KIM, A. K. and SAMWORTH, R. J. (2014). Global rates of convergence in log-concave density estimation. *arXiv preprint arXiv:1404.2298*.
- [21] KOENKER, R. and MIZERA, I. (2010). Quasi-concave density estimation. *Ann. Statist.* **38** 2998–3027. [MR2722462 \(2011j:62108\)](#)
- [22] KOENKER, R. and MIZERA, I. (2014). Convex Optimization in R. *Journal of Statistical Software* **60**.

- [23] LANG, R. (1986). A note on the measurability of convex sets. *Arch. Math. (Basel)* **47** 90–92. [MR855142 \(87j:28002\)](#)
- [24] MOSEK APS, D. (2011). The MOSEK Optimization Tools Manual, Version 6.0. Available at <http://www.mosek.com>.
- [25] PAL, J. K., WOODROOFE, M. and MEYER, M. (2007). Estimating a Polya frequency function<sub>2</sub>. In *Complex datasets and inverse problems. IMS Lecture Notes Monogr. Ser.* **54** 239–249. Inst. Math. Statist., Beachwood, OH. [MR2459192 \(2010d:62087\)](#)
- [26] PRAKASA RAO, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* **31** 23–36. [MR0267677 \(42 ##2579\)](#)
- [27] RINOTT, Y. (1976). On convexity of measures. *Ann. Probability* **4** 1020–1026. [MR0428540 \(55 ##1561\)](#)
- [28] ROCKAFELLAR, R. T. (1971). Integrals which are convex functionals. II. *Pacific J. Math.* **39** 439–469. [MR0310612 \(46 ##9710\)](#)
- [29] ROCKAFELLAR, R. T. (1997). *Convex Analysis. Princeton Landmarks in Mathematics*. Princeton University Press, Princeton, NJ Reprint of the 1970 original, Princeton Paperbacks. [MR1451876 \(97m:49001\)](#)
- [30] SCHUHMACHER, D., HÜSLER, A. and DÜMBGEN, L. (2011). Multivariate log-concave distributions as a nearly parametric model. *Stat. Risk Model.* **28** 277–295. [MR2838319 \(2012h:62044\)](#)
- [31] SELJO, E. and SEN, B. (2011). Nonparametric least squares estimation of a multivariate convex regression function. *Ann. Statist.* **39** 1633–1657. [MR2850215 \(2012j:62119\)](#)
- [32] SEREGIN, A. and WELLNER, J. A. (2010). Nonparametric estimation of multivariate convex-transformed densities. *Ann. Statist.* **38** 3751–3781. With supplementary material available online. [MR2766867 \(2012b:62126\)](#)
- [33] WRIGHT, F. T. (1981). The asymptotic behavior of monotone regression estimates. *Ann. Statist.* **9** 443–448. [MR606630 \(82k:62047\)](#)

DEPARTMENT OF STATISTICS, BOX 354322  
UNIVERSITY OF WASHINGTON  
SEATTLE, WA 98195-4322  
E-MAIL: [royhan@uw.edu](mailto:royhan@uw.edu)

DEPARTMENT OF STATISTICS, BOX 354322  
UNIVERSITY OF WASHINGTON  
SEATTLE, WA 98195-4322  
E-MAIL: [jaw@stat.washington.edu](mailto:jaw@stat.washington.edu)  
URL: <http://www.stat.washington.edu/jaw/>