

On efficiency of the plug-in principle for estimating smooth integrated functionals of a nonincreasing density

Rajarshi Mukherjee and Bodhisattva Sen*

*Department of Biostatistics
Harvard University
655 Huntington Avenue
Boston, MA 02115
e-mail: rmukherj@hsph.harvard.edu*

*Department of Statistics
Columbia University
1255 Amsterdam Avenue
New York, NY 10027
e-mail: bodhi@stat.columbia.edu*

Abstract: We consider the problem of estimating smooth integrated functionals of a monotone nonincreasing density f on $[0, \infty)$ using the non-parametric maximum likelihood based plug-in estimator. We find the exact asymptotic distribution of this natural (tuning parameter-free) plug-in estimator, properly normalized. In particular, we show that the simple plug-in estimator is always $\sqrt{n}$ -consistent, and is additionally asymptotically normal with zero mean and the semiparametric efficient variance for estimating a subclass of integrated functionals. To the best of our knowledge, this is the first time that such a large class of integrated functionals based on the Grenander estimator is treated in a unified fashion. Compared to the previous results on this topic (see e.g., Nickl (2008), Giné and Nickl (2016, Chapter 7), Jankowski (2014), and Söhl (2015)) our results hold for a much larger class of functionals (which include linear and non-linear functionals) under less restrictive assumptions on the underlying f — we do not require f to be (i) smooth, (ii) bounded away from 0, or (iii) compactly supported. Further, when f is the uniform distribution on a compact interval we explicitly characterize the asymptotic distribution of the plug-in estimator — which now converges at a non-standard rate — thereby extending the results in Groeneboom and Pyke (1983) for the case of the quadratic functional.

Keywords and phrases: Efficient estimation, empirical distribution function, Hadamard directionally differentiable, least concave majorant, non-parametric maximum likelihood estimator, shape-constrained density estimation.

Received May 2019.

*Supported by NSF Grants DMS-17-12822 and AST-16-14743.

1. Introduction

Estimation of functionals of data generating distributions is a fundamental research problem in nonparametric statistics. Consequently, substantial effort has gone into understanding estimation of nonparametric functionals (e.g., linear, quadratic, and other “smooth” functionals) both under density and regression (specifically Gaussian white noise models) settings — a comprehensive snapshot of this huge literature can be found in Hall and Marron (1987), Bickel and Ritov (1988), Donoho et al. (1990), Donoho and Nussbaum (1990), Fan (1991), Kerkycharian and Picard (1996), Nemirovski (2000), Laurent (1996), Cai and Low (2003, 2004, 2005) and the references therein. However, most of the above papers have focused on smoothness restrictions on the underlying (nonparametric) function class. In contrast, parallel results for purely shape-restricted function classes (e.g., monotonicity/convexity/log-concavity) are limited.

Estimation of a shape-restricted density is somewhat special among nonparametric density estimation problems. In shape-constrained problems classical maximum likelihood and/or least squares estimation strategies (over the class of all such shape-restricted densities) lead to consistent (see e.g., Grenander (1956), Groeneboom et al. (2001), Dümbgen and Rufibach (2009)) and adaptive estimators (see e.g., Birgé (1987), Kim et al. (2018)) of the underlying density. Consequently, the resulting procedures are tuning parameter-free; compare this with density estimation techniques under smoothness constraints that usually rely on tuning parameter(s) (e.g., bandwidth for kernel based procedures or resolution level for wavelet projections), the choice of which requires delicate care for the purposes of data adaptive implementations.

The simplest and the most well-studied shape-restricted density estimation problem is that of estimating a nonincreasing density on $[0, \infty)$. In this case, the most natural estimator of the underlying density is the Grenander estimator — the nonparametric maximum likelihood estimator (NPMLE) which also turns out to be the same as the least squares estimator; see e.g., Grenander (1956), Prakasa Rao (1969), Groeneboom (1985), Groeneboom and Jongbloed (2014). Even for this problem only a few papers exist that study nonparametric functional estimation; see e.g., Nickl (2008), Giné and Nickl (2016, Chapter 7), Jankowski (2014), and Söhl (2015). Moreover, these papers often assume additional constraints such as: (i) smoothness, (ii) compact support, and (iii) restrictive lower bounds on the underlying true density.

In this paper we study estimation of smooth integrated functionals of a nonincreasing density on $[0, \infty)$ using simple (tuning parameter-free) plug-in estimators. We characterize the limiting distribution of the plug-in estimator in this problem, under minimal assumptions. Moreover, we show that the limiting distribution (of the plug-in estimator) is asymptotically normal with the efficient variance for a large class of functionals, under minimal assumptions on the underlying density. To the best of our knowledge, this is the first time that such a large class of integrated functionals based on the Grenander estimator is treated in a unified fashion.

1.1. Problem formulation and a summary of our contributions

Suppose that $X_1, \dots, X_n$ are independent and identically distributed (i.i.d.) random variables from a nonincreasing density f on $[0, \infty)$. Let P denote the distribution of each X_i and F the corresponding distribution function. The goal is to study estimation and uncertainty quantification of the functional τ defined as

$$\tau(g, f) := \int_0^\infty g(f(x), x) dx, \quad (1)$$

where $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ is a known (smooth) function. Functionals of the form (1) belong to the class of problems considered by many authors including Levit (1978), Bickel and Ritov (1988), Ibragimov and Khas'minskiĭ (1991), Birgé and Massart (1995), Kerkycharian and Picard (1996), Laurent (1996), Laurent (1997), Nemirovski (2000) etc. and include the monomial functionals

$$\int_0^\infty f^p(x) dx \quad \text{for } p \geq 2.$$

Of particular significance among them is the quadratic functional corresponding to $p = 2$ — being related to other inferential problems in density models such as goodness-of-fit testing and construction of confidence balls (see e.g., Giné and Nickl (2016, Chapter 8.3) and the references therein for more details).

Let $\hat{f}_n$ be the Grenander estimator in this problem (see Section 2 for its characterization). Then, a natural estimator of $\tau(g, f)$ is the simple *plug-in estimator*

$$\tau(g, \hat{f}_n) := \int_0^\infty g(\hat{f}_n(x), x) dx. \quad (2)$$

In this paper we characterize the asymptotic distribution of $\tau(g, \hat{f}_n)$ for estimating $\tau(g, f)$, when g is assumed to be smooth; see Section 2.2 for the exact assumptions on g . Specifically, in Theorem 2.2, we show that the simple plug-in estimator $\tau(g, \hat{f}_n)$ is $\sqrt{n}$ -consistent, and we characterize its limiting distribution. Further, when F is strictly concave we show that $\tau(g, \hat{f}_n)$ (properly normalized) is asymptotically normal with the (semiparametric) efficient variance (see Theorem 2.3). Moreover, we can consistently estimate this limiting variance which readily yields asymptotically valid confidence intervals for $\tau(g, f)$. Compared to the previous results on this topic (see e.g., Nickl (2007), Nickl (2008), Jankowski (2014), Giné and Nickl (2016, Chapter 7) and Söhl (2015)) our results hold under less restrictive assumptions on the underlying density (see conditions (A1)-(A3) in Section 2.2) — in particular, we do not assume that f is smooth, bounded away from 0, or compactly supported.

An important subclass of the above considered functional class is:

$$\mu(h, f) := \int_0^\infty h(f(x)) dx, \quad (3)$$

where $h : \mathbb{R} \rightarrow \mathbb{R}$ is a known (smooth) function. In Section 3 we study estimation of this functional in detail. The simple plug-in estimator for estimating $\mu(h, f)$ is

$$\mu(h, \hat{f}_n) := \int_0^\infty h(\hat{f}_n(x)) dx.$$

For the class of functionals $\mu(h, f)$, under certain assumptions on the function $h(\cdot)$, we prove in Theorem 3.1 that the plug-in estimator is asymptotically normal with the semiparametric efficient variance, even when F is not strictly concave. In fact, the asymptotic normality even holds when f is piecewise constant. Finally, in the special case when f is constant (i.e., f is uniform on a compact interval) we explicitly characterize the asymptotic distribution of the plug-in estimator $\mu(h, \hat{f}_n)$, which now converges at a non-standard rate; see Theorem 3.2. This theorem extends the results in Groeneboom and Pyke (1983) beyond the quadratic functional.

Finally, in Section 4 we focus on a special case of $\mu(h, f)$:

$$\nu(h, f) := \int_0^\infty h(f(x)) f(x) dx, \quad (4)$$

where $h : \mathbb{R} \rightarrow \mathbb{R}$ is a known (smooth) function. As before, the simple plug-in estimator for $\nu(h, f)$ is $\nu(h, \hat{f}_n) := \int h(\hat{f}_n(x)) \hat{f}_n(x) dx$. Another intuitive estimator in this scenario would be $\frac{1}{n} \sum_{i=1}^n h(\hat{f}_n(X_i))$. It is easy to show that in this case both the above plug-in estimators are exactly the same (see Lemma 4.1). Further, we show in Remark 4.1 that both these estimators are also equivalent to the standard one-step bias-corrected estimator used in nonparametric statistics. As a consequence, we show that all these three estimators are now asymptotically normal with the (semiparametric) efficient variance.

1.2. Comparison with the literature

Estimation of functionals of the form (1) (or (3) and (4)) in the density model, for f belonging to smoothness classes (e.g., Sobolev and Besov balls), has been studied in detail in the literature. Historically, the case of “smooth” g while estimating (1) (or “smooth” h while estimating (3) and (4)) is relatively well-understood and we direct the interested reader to Hall and Marron (1987), Bickel and Ritov (1988), Donoho et al. (1990), Donoho and Nussbaum (1990), Fan (1991), Birgé and Massart (1995), Kerkycharian and Picard (1996), Laurent (1996), Nemirovski (2000), Cai and Low (2003), Cai and Low (2004), Cai and Low (2005), Tchetgen et al. (2008) and the references therein.

A general feature of the results for estimating “smooth” integrated nonparametric functionals is an elbow effect in the rate of estimation based on the smoothness of the underlying function class. For example, when estimating the quadratic functional, $\sqrt{n}$ -efficient estimation can be achieved as soon as the underlying density has Hölder smoothness index $\beta \geq 1/4$, whereas the optimal rate of estimation is $n^{-4\beta/(4\beta+1)}$ (in root mean squared sense) for $\beta < 1/4$.

Since monotone densities have one weak derivative one can therefore expect to have $\sqrt{n}$ -consistent efficient estimation of $\tau(g, f)$ (under suitable assumptions on g). However, standard methods of estimation under smoothness restrictions involve expanding the infinite dimensional function in a suitable orthonormal basis of L_2 and estimating an approximate functional created by truncating the basis expansion at a suitable point. The point of truncation decides the approximation error of the truncated functional as a surrogate of the actual functional and depends on the smoothness of the functional of interest and approximation properties of the orthonormal basis used. This point of truncation is then balanced with the bias and variance of the resulting estimator and therefore directly depends on the smoothness of the function. Consequently, most optimal estimators proposed in the literature depend explicitly on the knowledge of the smoothness indices. In this regard, our main result in the paper shows the optimality of tuning parameter-free plug-in procedure based on the Grenander estimator (under less restrictive assumptions) for estimating integral functionals of monotone densities.

We also mention that the results of this paper are directly motivated by the general questions tackled in Bickel and Ritov (2003) and Nickl (2007). In particular, Bickel and Ritov (2003) considered nonparametric function estimators that satisfy the desirable “plug-in property”. An estimator is said to have the plug-in property if it is simultaneously minimax optimal in a function space norm (e.g., in l_2 -norm) and can also be “plugged-in” to estimate specific classes of functionals efficiently (and/or at $\sqrt{n}$ -rate). Subsequently, Bickel and Ritov (2003) demonstrated general principles for constructing such nonparametric estimators pertaining to linear functionals (allowing for certain extensions) with examples arising in various problems such as nonparametric regression, survival analysis, and density estimation. Indeed, in this paper we show that the Grenander estimator satisfies such a plug-in property for estimating smooth integrated functionals of monotone densities.

Note that such a plug-in property of the Grenander based estimator is also illustrated in the papers Söhl (2015) and Nickl (2007) where the authors study linear functionals and the entropy functional crucially using a Kiefer-Wolfowitz type uniform central limit theorem (CLT) for the Grenander estimator. Such a proof strategy however comes with the price that more restrictive assumptions need to be made on the density f to derive such a CLT. In particular, the results in Söhl (2015) and Nickl (2007) assume a bounded domain and a uniform positive lower bound on the true density f — assumptions we forgo in this paper.

Jankowski (2014) also considers estimation of linear functionals (and the entropy functional) based on the Grenander estimator. In Jankowski (2014, Theorem 3.1), the limiting distribution of the Grenander based plug-in estimator is derived for estimating a linear functional (under possible misspecification, i.e., f need not be nonincreasing). Jankowski (2014, Theorem 4.1) gives the limiting distribution of the plug-in estimator when estimating the entropy functional. However, all these results assume that the support of f is bounded, an assumption that we do not make in this paper.

1.3. Organization of the paper

The rest of the paper is organized as follows. In Section 2.1 we collect useful notation, definitions, and results which will help the presentation of the rest of the paper. In Section 2.2, we define the class of integrated functionals of interest along with our main assumptions. Section 2.3 discusses the construction of the plug-in estimator and the main results of this paper (Theorems 2.2 and 2.3) which gives $\sqrt{n}$ -consistency and a characterization of the asymptotic distribution of the estimator. Subsequently, in Sections 3 and 4 we focus on the special classes of functionals — $\mu(h, f)$ and $\nu(h, f)$ — and provide explicit results (Theorem 3.1 and Corollary 4.2) to show that our plug-in estimator is always $\sqrt{n}$ -consistent, asymptotically normal and semiparametric efficient. Since the case of the uniform density on a compact interval deserves a finer asymptotic expansion, we devote Section 3.1 to this purpose. Section 5 presents some numerical results that validate and illustrate our theoretical findings. In Section 6, we discuss some potential future research problems. Finally, all the technical proofs are relegated to Section 7.

2. Estimation of the integrated functional $\tau(g, f)$

2.1. Preliminaries

In this subsection we introduce some notation and definitions to be used in the rest of the paper. Throughout we let $\mathbb{R}_+$ denote the (compact) nonnegative real line $[0, \infty]$. The underlying probability space on which all random elements are defined is $(\Omega, \mathcal{F}, \mathbb{P})$. Suppose that we have nonnegative random variables

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} P$$

having a *nonincreasing* density f and (concave) distribution function F . In the rest of the paper, we use the operator notation, i.e., for any function $\psi : \mathbb{R} \rightarrow \mathbb{R}$, $P[\psi] := \int \psi(x) dP(x)$ denotes the expectation of $\psi(X)$ under the distribution $X \sim P$. For any two functions $\gamma_1, \gamma_2 : [0, \infty) \rightarrow \mathbb{R}$ with $\int |\gamma_1(x)| dx < +\infty$, we view $\int \gamma_2(x) d(\gamma_1(x))$ as a Lebesgue-Stieltjes integral.

Let $\mathbb{P}_n$ denote the empirical measure of the X_i 's and let $\mathbb{F}_n$ denote the corresponding empirical distribution function, i.e., for $x \in \mathbb{R}$,

$$\mathbb{F}_n(x) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{(-\infty, x]}(X_i).$$

For a nonempty set T , we let $\ell^\infty(T)$ denote the set of uniformly bounded, real-valued functions on T . Of particular importance is the space $\ell^\infty(\mathbb{R}_+)$, which we equip with the uniform metric $\|\cdot\|_\infty$ and the ball σ -field; see Pollard (1984, Chapters IV and V) for background.

Following Beare and Fang (2017), we now define the notion of the *least concave majorant* (LCM) operator. Given a nonempty convex set $T \subset \mathbb{R}_+$, the LCM

over T is the operator $\mathcal{M}_T : \ell^\infty(\mathbb{R}_+) \rightarrow \ell^\infty(T)$ that maps each $\xi \in \ell^\infty(\mathbb{R}_+)$ to the function $\mathcal{M}_T \xi$ where

$$\mathcal{M}_T \xi(x) := \inf \left\{ \eta(x) : \eta \in \ell^\infty(T), \eta \text{ is concave, and } \xi \leq \eta \text{ on } T \right\}, \quad x \in T.$$

We write $\mathcal{M}$ as shorthand for $\mathcal{M}_{\mathbb{R}_+}$ and refer to $\mathcal{M}$ as the LCM operator. Beare and Fang (2017, Proposition 2.1) shows that the LCM operator $\mathcal{M}$ is *Hadamard directionally differentiable* (see e.g., Beare and Fang (2017, Definition 2.2)) at any concave function $\theta \in \ell^\infty(\mathbb{R}_+)$ tangentially to $C(\mathbb{R}_+)$ (here $C(\mathbb{R}_+)$ denotes the collection of all continuous real-valued functions on $\mathbb{R}_+$ vanishing at infinity). Let us denote the Hadamard directional derivative of $\mathcal{M}$ at the concave function $\theta \in \ell^\infty(\mathbb{R}_+)$ by $\mathcal{M}'_\theta$. The following result, due to Beare and Fang (2017, Proposition 2.1), characterizes the Hadamard directional derivative of $\mathcal{M}'_\theta$, for $\theta \in \ell^\infty(\mathbb{R}_+)$, tangentially to $C(\mathbb{R}_+)$ and is crucial to our subsequent analysis.

Proposition 2.1 (Proposition 2.1 of Beare and Fang (2017)). *The LCM operator $\mathcal{M} : \ell^\infty(\mathbb{R}_+) \rightarrow \ell^\infty(\mathbb{R}_+)$ is Hadamard directionally differentiable at any concave $\theta \in \ell^\infty(\mathbb{R}_+)$ tangentially to $C(\mathbb{R}_+)$. Its directional derivative $\mathcal{M}'_\theta : C(\mathbb{R}_+) \rightarrow \ell^\infty(\mathbb{R}_+)$ is uniquely determined as follows: for any $\xi \in \ell^\infty(\mathbb{R}_+)$ and $x \geq 0$, we have*

$$\mathcal{M}'_\theta \xi(x) = \mathcal{M}_{T_{\theta,x}} \xi(x),$$

where $T_{\theta,x} = \{x\} \cup U_{\theta,x}$, and $U_{\theta,x}$ is the union of all open intervals $A \subset \mathbb{R}_+$ such that (i) $x \in A$, and (ii) θ is affine over A .

Let $\hat{F}_n$ be the LCM of $\mathbb{F}_n$, i.e., $\hat{F}_n$ is the smallest concave function that sits above $\mathbb{F}_n$. Let $\mathbb{B}(\cdot)$ be the Brownian bridge process on $[0, 1]$. By Donsker's theorem (see e.g., van der Vaart (1998, Theorem 19.3)) we know that the empirical process $\mathbb{D}_n := \sqrt{n}(\mathbb{F}_n - F)$ converges in distribution to $\mathbb{G} \equiv \mathbb{B} \circ F$ in $\ell^\infty(\mathbb{R}_+)$, i.e.,

$$\mathbb{D}_n := \sqrt{n}(\mathbb{F}_n - F) \xrightarrow{d} \mathbb{G}.$$

As the delta method is valid for Hadamard directionally differentiable functions (see Shapiro (1991)), as a consequence of the above discussion it follows that (see Beare and Fang (2017, Theorem 2.1)), for any concave F , the stochastic processes $\hat{\mathbb{D}}_n := \sqrt{n}(\hat{F}_n - F)$ converges in distribution to $\hat{\mathbb{G}}$,

$$\hat{\mathbb{D}}_n := \sqrt{n}(\hat{F}_n - F) \xrightarrow{d} \hat{\mathbb{G}},$$

where

$$\hat{\mathbb{G}} := \mathcal{M}'_F \mathbb{G}. \quad (5)$$

To give more intuition about the stochastic process $\hat{\mathbb{G}}$ and the operator $\mathcal{M}'_F$, let us consider two simple scenarios. First, suppose that F is strictly concave. Then $\hat{\mathbb{G}} = \mathcal{M}'_F \mathbb{G} = \mathbb{G}$ as by Proposition 2.1, for every $x > 0$,

$$\mathcal{M}'_F \mathbb{G}(x) = \mathcal{M}_{\{x\}} \mathbb{G}(x) = \mathbb{G}(x);$$

also see Beare and Fang (2017, Proposition 2.2). Thus, $\mathbb{D}_n$ and $\hat{\mathbb{D}}_n$ converge to the same limiting object in this case.

Next, let us suppose that F is piecewise affine. In this case f is piecewise constant with jumps (say) at $0 < t_1 < t_2 < \dots < t_k < \infty$ and values $v_1 > \dots > v_k$, for some $k \geq 2$ integer, i.e., for $x > 0$,

$$f(x) = \sum_{i=1}^k v_i \mathbf{1}_{(t_{i-1}, t_i]}(x), \quad (6)$$

where $t_0 \equiv 0$. Then,

$$\mathcal{M}'_F \mathbb{G}(x) = \begin{cases} \mathcal{M}_{(t_{i-1}, t_i)} \mathbb{G}(x), & \text{if } x \in (t_{i-1}, t_i) \\ \mathbb{G}(x), & \text{if } x = t_i \text{ for } i = 0, 1, \dots, k. \end{cases} \quad (7)$$

2.2. Assumptions

In this subsection we state the assumptions needed for our main results.

(A1) The true underlying density $f : [0, \infty) \rightarrow [0, \infty)$ is nonincreasing with

$$f(0+) := \lim_{x \downarrow 0} f(x) < +\infty.$$

(A2) We assume that for some $\alpha \in (0, 1]$,

$$\int_{f < 1} f^{1-\alpha}(x) dx < +\infty.$$

(A3) Let $\mathcal{S}$ denote the support of f . The function g in (1) satisfies the following conditions:

- (i) The functional (1) is well-defined, i.e., $|\tau(g, f)| < +\infty$.
- (ii) $g(\cdot, x) \in C^2([0, \infty))$, for every $x \in [0, \infty)$, where $C^2([0, \infty))$ denotes the class of all twice continuously differentiable functions on $[0, \infty)$.
- (iii) Let $\dot{g}$ and $\ddot{g}$ denote the derivatives of $g(\cdot, \cdot)$ only with respect to the first coordinate, i.e., for $z, x \in [0, \infty)$,

$$\dot{g}(z, x) = \frac{\partial}{\partial z} g(z, x) \quad \text{and} \quad \ddot{g}(z, x) = \frac{\partial^2}{\partial z^2} g(z, x).$$

We assume that the following hold:

(a) For any $B > 0$ there exists a constant $K_B < \infty$ such that

$$\sup_{x \in \mathcal{S}, z \in [0, B]} |\ddot{g}(z, x)| \leq K_B.$$

(b) $\dot{g}(f(\cdot), \cdot)$ is of bounded total variation on $[0, \infty)$, i.e.,

$$\int_{\mathcal{S}} |d[\dot{g}(f(x), x)]| < +\infty.$$

Let us briefly discuss assumptions (A1)-(A3). Conditions (A1) and (A2) on the true density f imply that f is bounded and satisfies a tail decay condition. Note that (A2) is satisfied if for some $m > 1$, $\limsup_{x \rightarrow \infty} x^m f(x) < \infty$; see e.g., van de Geer (2000, Example 7.4.2). In particular, if f is nonincreasing, bounded and compactly supported, conditions (A1) and (A2) immediately hold.

Condition (A3)-(i) is unavoidable since it pertains to the existence of the functional we are interested in estimating. Parts (ii) and (iii,a) of (A3) pertain to the function $g(\cdot, \cdot)$ and quantify the notion of the smooth functional we consider. Finally, part (iii,b) of (A3) is slightly stronger than the requirement that $\int \dot{g}(f(x), x) f(x) dx$ exists — which in turn appears in the information bound for estimating $\tau(g, f)$ and is therefore needed to be finite for efficient estimators to exist. More precisely, the finite total variation of $\dot{g}(f(x), x)$ helps us define several Lebesgue-Stieltjes integrals with respect to $\dot{g}(f(x), x)$ without further conditions.

Observe that our assumption (A3) implies twice (right) differentiability of $g(\cdot, x)$ (for every x) at 0 and hence does not include the entropy functional corresponding to $g(z, x) = -z \log z$, unless f is assumed to be bounded away from 0.

2.3. $\sqrt{n}$ -Consistency of $\tau(g, \hat{f}_n)$

As mentioned before, a natural estimator of f in this situation is $\hat{f}_n$, the Grenander estimator — the (nonparametric) maximum likelihood estimator of a nonincreasing density on $[0, \infty)$; see Grenander (1956). The Grenander estimator $\hat{f}_n$ is defined as the left-hand slope of $\hat{F}_n$, the LCM of the empirical distribution function $\mathbb{F}_n$. Note that $\hat{F}_n$ is piecewise affine and a valid *concave* distribution function by itself. Further, $\hat{f}_n$ is a piecewise constant (nonincreasing) density with possible jumps only at the data points X_i 's. Moreover, $\hat{f}_n(x) = 0$ for $x > \max_{i=1, \dots, n} X_i$. Consequently, a natural plug-in estimator of $\tau(g, f)$ is

$$\tau(g, \hat{f}_n) := \int_0^\infty g(\hat{f}_n(x), x) dx.$$

We study estimation and uncertainty quantification of the functional $\tau(g, \hat{f}_n)$. The following result shows that the plug-in estimator $\tau(g, \hat{f}_n)$ is $\sqrt{n}$ -consistent and explicitly characterizes its asymptotic distribution.

Theorem 2.2. *Assume that conditions (A1)-(A3) hold. Then,*

$$\sqrt{n} \left(\tau(g, \hat{f}_n) - \tau(g, f) \right) \xrightarrow{d} - \int_0^\infty \hat{\mathbb{G}}(x) d[\dot{g}(f(x), x)] =: Y \quad (8)$$

where $\hat{\mathbb{G}}(\cdot)$ is defined in (5).

We defer the proof of Theorem 2.2 to Section 7.1 and before proceeding further, make a few comments regarding the implications of Theorem 2.2.

Remark 2.1 (Random variable Y). *The limiting random variable Y in the statement of Theorem 2.2 is well-defined. Note that $Y = -\int_{\mathcal{S}} \hat{\mathbb{G}}(x) d[\dot{g}(f(x), x)]$ as $\hat{\mathbb{G}}$ is zero outside $\mathcal{S}$, the support of the distribution P . That Y is well-defined follows from the following facts: (i) $\hat{\mathbb{G}}$ is a (measurable) random element in $\ell^\infty(\mathbb{R}_+)$ (see e.g., Beare and Fang (2017)); (ii) by assumption (A3)-(iii,b), for any $\phi \in \ell^\infty(\mathbb{R}_+)$, the map $\phi \mapsto \int_{\mathcal{S}} \phi(x) d[\dot{g}(f(x), x)]$ is continuous and measurable (from $\ell^\infty(\mathbb{R}_+)$ to $\mathbb{R}$).*

Remark 2.2 (Assumptions in Theorem 2.2). *Note that Theorem 2.2 is valid under much weaker assumptions on f than previously studied in the literature. Although, similar results have been derived in Nickl (2008) and Giné and Nickl (2016, Chapter 7), the derivations relied on further smoothness, lower bound, and compact support assumptions on f . In contrast, our only technical assumptions (A1) and (A2) are significantly less restrictive and only demand f to be bounded and have a polynomial-type tail decay. The key idea in our analysis, that allows us to forgo the above mentioned restrictive assumptions, is to first use integration-by-parts to express the functional $\tau(g, f)$, and its estimate $\tau(g, \hat{f}_n)$, in terms of F and $\hat{F}_n$ respectively. This representation turns out to be crucial and useful given the recent result in Beare and Fang (2017, Theorem 2.1) where the authors derive a process convergence result for $\sqrt{n}(\hat{F}_n - F)$, for any F (as long as it is concave); see (5). As this weak convergence result circumvents the conditions typically assumed in semiparametric theory (see e.g., Söhl (2015), Nickl (2007), Nickl (2008), Giné and Nickl (2016)), we are also able to do without these restrictive assumptions.*

Remark 2.3 (Connection to estimation in smoothness classes). *For smoothness classes of functions it is well-known that $\sqrt{n}$ -consistent estimation of $\tau(g, f)$ is possible if f has more than $1/4$ derivatives (see e.g., Birgé and Massart (1995)) and a typical construction of such an estimator proceeds via a bias-corrected one-step estimator; see e.g., Bickel and Ritov (1988), Robins et al. (2008), Robins et al. (2017). Since monotone densities have one weak derivative, one can expect $\sqrt{n}$ -consistent estimation of $\tau(g, f)$. Theorem 2.2 shows that in fact the plug-in principle based on the Grenander estimator is $\sqrt{n}$ -consistent with a distributional limit without any further assumptions on f . It will be further shown in Section 4 that for estimating the special functional $\nu(h, f)$, defined in (4), the one-step bias-corrected estimator is equivalent to the plug-in estimator. Consequently, this will provide more intuition on $\sqrt{n}$ -consistent efficient estimation of $\nu(h, f)$, without further assumptions.*

Remark 2.4 (Tuning parameter-free estimation). *Being based on the Grenander estimator and the plug-in principle, the estimator $\tau(g, \hat{f}_n)$ is completely tuning parameter-free unlike those considered in Robins et al. (2008), Mukherjee et al. (2016), Mukherjee et al. (2017).*

In the following result (proved in Section 7.2), we show that when F is strictly

concave, the plug-in estimator $\tau(g, \hat{f}_n)$ is $\sqrt{n}$ -consistent and asymptotically normal with the (semiparametric) efficient variance.

Theorem 2.3. *Suppose that F is strictly concave on $[0, \infty)$. Then, under the assumptions of Theorem 2.2, we have*

$$Y = - \int_0^\infty \mathbb{G}(x) d[\dot{g}(f(x), x)] \stackrel{d}{=} N(0, \sigma_{\text{eff}}^2(g, f)), \quad (9)$$

where, for $X \sim P$,

$$\sigma_{\text{eff}}^2(g, f) := \text{Var}(\dot{g}(f(X), X)) < +\infty. \quad (10)$$

In general, when F is not strictly concave, we cannot say if Y (described by (8)) admits a simpler description. Indeed, the integral $\int \hat{\mathbb{G}}(x) d[\dot{g}(f(x), x)]$ can be a highly non-linear functional of the Gaussian process $\mathbb{G}$ and consequently there is no immediate reason to believe that the limiting distribution is normal. Surprisingly though, for a special class of functionals, namely $\mu(h, f)$ introduced in (3), the distribution of $\int \hat{\mathbb{G}}(x) d[\dot{g}(f(x), x)]$ is always Gaussian. The next section is therefore devoted to the complete understanding of this special case.

Remark 2.5 (Linear functional). *Theorem 2.2 covers the case of estimating an integrated linear functional $\int w(x)f(x)dx$ where $w(\cdot)$ is any weight function with finite variation on $\mathcal{S}$. Indeed, this follows by taking $g(z, x) = w(x)z$ in Theorem 2.2.*

When F is strictly concave, Theorem 2.3 shows that the limiting distribution of the plug-in estimator is asymptotically normal. When F has a density f which is piecewise constant and has the form (6) and $w(\cdot)$ is continuous, Theorem 2.2 along with (7) yields

$$\sqrt{n} \left\{ \int w(x) \hat{f}_n(x) dx - \int w(x) f(x) dx \right\} \xrightarrow{d} - \sum_{i=1}^k \int_{t_{i-1}}^{t_i} \mathcal{M}_{(t_{i-1}, t_i)} \mathbb{G}(x) dw(x);$$

cf. Jankowski (2014) where the author derives the limiting distribution for linear functionals (under model-misspecification) and expresses the limit in a different form.

3. Plug-in efficient estimation of $\mu(h, f)$

In this section we focus on the special case of estimating $\mu(h, f)$ (as defined in (3)) where $h : [0, \infty) \rightarrow \mathbb{R}$ is assumed to be a known function satisfying:

$$(A4) \quad h \in C^2([0, \infty)),$$

where $C^2([0, \infty))$ denotes the class of all twice continuously differentiable functions on $[0, \infty)$ and

$$\int |d[h'(f(x))]| < +\infty.$$

The natural plug-in estimator of $\mu(h, f)$ is

$$\mu(h, \hat{f}_n) := \int_0^\infty h(\hat{f}_n(x)) dx.$$

Theorem 2.2 can be used to obtain the $\sqrt{n}$ -consistency and the asymptotic distribution of $\mu(h, \hat{f}_n)$ by considering $g(z, x) = h(z)$. Note that for $h(\cdot)$ satisfying (A4), the corresponding g automatically satisfies (A3) defined in Section 2.2. However, in this case we can simplify the form of the limiting distribution Y (see (8)) and prove that it is indeed a mean zero normal random variable with the efficient variance. Theorem 3.1 (proved in Section 7.3) formalizes this.

Theorem 3.1. *Assume that conditions (A1)-(A2) and (A4) hold. Then,*

$$\sqrt{n} \left(\mu(h, \hat{f}_n) - \mu(h, f) \right) \xrightarrow{d} N(0, \sigma_{\text{eff}}^2(h, f)), \quad (11)$$

where, for $X \sim P$, $\sigma_{\text{eff}}^2(h, f) := \text{Var} \left(h'(f(X)) \right) < +\infty$.

A few remarks about the implications of Theorem 3.1 are in order now.

Remark 3.1 (Assumptions in Theorem 3.1). *Assumption (A4) in Theorem 3.1 is implied by assumption (A3) where one considers $g(z, x) = h(z)$ with $h \in C^2([0, \infty))$. Thus, the $\sqrt{n}$ -consistency and existence of asymptotic distribution for the plug-in principle is automatic from Theorem 2.2. However, unlike Theorem 2.3, the requirement of F being strictly concave is no longer necessary for the validity of Theorem 3.1.*

Remark 3.2 (Asymptotic efficiency). *Observe that the limiting variance in Theorem 3.1 (see (11)) matches the nonparametric efficiency bound in this problem (Bickel et al. (1993), van der Vaart (1998), van der Vaart (2002), Laurent (1996)). Indeed, this efficiency bound corresponds to the case when the maximal tangent space in the model is the whole of $L_2(F)$. Although such a bound can be shown to hold for models where f has a lower bound on its slope (in absolute value) and is compactly supported, it is not immediate for more general monotone densities. Consequently, the statement that the “plug-in estimator $\mu(h, \hat{f}_n)$ is asymptotically efficient for any concave F (under assumptions (A1), (A2), and (A4))” should be understood in the sense of simply achieving the efficiency bound in the nonparametric model. When the tangent space is $L_2(F)$, this is indeed the semiparametric efficiency bound.*

Remark 3.3 (Monomial functionals). *Of special interest is the case when $h(x) = x^p$, where $p \geq 2$ is an integer. For example, the quadratic functional $\int f^2$ is obtained when $p = 2$. In this case, Theorem 3.1 yields an asymptotically efficient variance of*

$$\sigma_{\text{eff}}^2(h, f) = \text{Var}_P \left(p f(X)^{p-1} \right) = p^2 \left(\int_0^\infty f^{2p-1}(x) dx - \left(\int_0^\infty f^p(x) dx \right)^2 \right), \quad (12)$$

for any concave F .

Remark 3.4 (Construction of confidence intervals). *As an important consequence of Theorem 3.1 one can construct asymptotically valid confidence intervals for the functional $\mu(h, f)$. In particular, one can estimate the efficient asymptotic variance σ_{eff}^2 as follows. Note that,*

$$\sigma_{\text{eff}}^2(h, f) = \text{Var}\left(h'(f(X))\right) = \mu(h_1, f) - (\mu(h_2, f))^2$$

where $h_1(z) := z(h'(z))^2$ and $h_2(z) := zh'(z)$ for $z \geq 0$. As a result, if $h \in C^3([0, \infty))$ then $h_1, h_2 \in C^2([0, \infty))$ and Theorem 3.1 implies that $\hat{\sigma}_n^2 := \mu(h_1, \hat{f}_n) - (\mu(h_2, \hat{f}_n))^2$ is a consistent estimator of $\sigma_{\text{eff}}^2(h, f)$. Consequently $\mu(h, \hat{f}_n) \pm z_{\alpha/2} \hat{\sigma}_n$ (with $z_{\alpha/2}$ being the $1 - \alpha/2$ quantile of the standard normal distribution) is an asymptotically valid $100 \times (1 - \alpha)\%$ confidence interval for $\mu(h, f)$. Indeed, the additional assumption of $h \in C^3([0, \infty))$ compared to $h \in C^2([0, \infty))$ in Theorem 3.1 can be reduced since we only demand consistency of $\hat{\sigma}_n$, instead of asymptotic normality. We omit the details for the sake of avoiding repetitions. Finally, the validity of this confidence interval is point-wise (true for every fixed underlying f) and not in a uniform sense.

Remark 3.5 (When $\sigma_{\text{eff}}^2(h, f) = 0$). *If $\sigma_{\text{eff}}^2 \equiv \text{Var}(h'(f(X))) = 0$ then $h'(u) = c$ a.e. u (with respect to the distribution of $f(X)$), where $c \in \mathbb{R}$ is a constant. Two scenarios can lead to the above case; they are: (i) $h(u) = cu$, in which case the functional under consideration is $\mu(h, f) := \int h(f(x))dx = c \int f(x)dx = c$ and, moreover, $\mu(h, \hat{f}_n) = c$, and there is no estimation error; (ii) $f(\cdot)$ is a constant. In the latter case, f is the uniform distribution on $[0, c]$, for some $c > 0$. In Section 3.1 below we deal with this scenario (i.e., when f is uniform) and obtain a non-degenerate distributional limit, after proper normalization.*

3.1. When f is uniformly distributed

As mentioned in Remark 3.5, Theorem 3.1 does not give a non-degenerate limit when f is the uniform distribution on the interval $[0, c]$, for any $c > 0$. Indeed, as we will see, in such a case $\mu(h, \hat{f}_n)$ has a faster rate of convergence. In this subsection we focus our attention to the case when f is the uniform distribution on $[0, 1]$; an obvious scaling argument can then be used to generalize the result to the case when f is uniform on $[0, c]$, for any $c > 0$.

Suppose that P is the uniform distribution on $[0, 1]$, i.e., $f(x) = \mathbf{1}_{[0,1]}(x)$, $x \in \mathbb{R}$. It has been shown in Groeneboom and Pyke (1983) (also see Groeneboom (1983)) that in this case

$$\frac{n \int_0^1 (\hat{f}_n^2(x) - 1)dx - \log n}{\sqrt{3 \log n}} \xrightarrow{d} N(0, 1). \quad (13)$$

The above result yields the asymptotic distribution for the plug-in estimator of the quadratic functional (as $\int f^2(x)dx = 1$), properly normalized. In the following theorem we extend the above result to general smooth functionals of the Grenander estimator.

Theorem 3.2. *Let $X_1, \dots, X_n$ be i.i.d. $\text{Uniform}([0, 1])$. Suppose that $h \in C^4([0, \infty))$, where $C^4([0, \infty))$ is the space of all four times continuously differentiable functions on $[0, \infty)$. If $h''(1) \neq 0$, then*

$$\frac{n \left\{ \mu(h, \hat{f}_n) - \mu(h, f) \right\} - \frac{1}{2} h''(1) \log n}{\sqrt{3 \left[\frac{1}{2} h''(1) \right]^2 \log n}} \xrightarrow{d} N(0, 1). \quad (14)$$

The proof of the above result is given in Section 7.5 and follows closely the line of argument presented in Groeneboom and Pyke (1983). Indeed, as is somewhat apparent from the statement of the theorem, the proof relies on a Taylor expansion of h and thereafter controlling the error terms. In particular, we prove that

$$\frac{n \int_0^1 (\hat{f}_n(x) - 1)^k dx}{\sqrt{\log n}} = o_{\mathbb{P}}(1), \quad k = 3, 4,$$

and then a simple three term Taylor expansion yields the desired result.

Remark 3.6 (If $h''(1) = 0$). *In the setup of Theorem 3.2 if $h''(1) = 0$ then it is easy to check from the proof that $\frac{n \{ \mu(h, \hat{f}_n) - \mu(h, f) \}}{\sqrt{\log n}} \xrightarrow{\mathbb{P}} 0$. We do not have a further guess of the rescaling which will yield a non-degenerate limiting distribution.*

Remark 3.7 (Comparison with Groeneboom (2019)). *Recently, in Groeneboom (2019, Corollary 4.1) the author proved the same result (as in Theorem 3.2) under slightly different assumptions. The following are the main differences in the two approaches: (i) Groeneboom (2019, Corollary 4.1) relies on analyticity of the function $h(z)$ on $\{z \in \mathbb{C} : \text{Re}(z) > 0\}$ whereas we require four bounded continuous derivatives on the positive real line (including 0); (ii) the proof technique in Groeneboom (2019) can handle the entropy functional whereas it is not immediate whether our proof technique can be modified to include this case.*

4. Other efficient estimators of $\nu(h, f)$

In this section we focus on the special case of estimating the functional $\nu(h, f)$ (as defined in (4)) where $h : [0, \infty) \rightarrow \mathbb{R}$ is assumed to be a known function and consider other (efficient) estimators of $\nu(h, f)$. Our plug-in estimator of $\nu(h, f)$ is

$$\nu(h, \hat{f}_n) := \int h(\hat{f}_n(x)) \hat{f}_n(x) dx = \hat{P}_n[h \circ \hat{f}_n], \quad (15)$$

where $\hat{P}_n$ denotes the probability measure associated with the Grenander estimator (i.e., $\hat{P}_n$ has distribution function $\hat{F}_n$).

We first contrast the above plug-in estimator with two other natural estimators and argue that the special structure of the Grenander estimator implies their equivalence. Observe that another natural estimator of $\nu(h, f) \equiv P[h \circ f]$ would be

$$\mathbb{P}_n[h \circ \hat{f}_n] := \frac{1}{n} \sum_{i=1}^n h(\hat{f}_n(X_i)).$$

Indeed, Lemma 4.1 below shows that this natural estimator is exactly the same as the plug-in estimator $\hat{P}_n[h \circ \hat{f}_n]$.

Lemma 4.1. *For any function $h : \mathbb{R} \rightarrow \mathbb{R}$, $\hat{P}_n[h \circ \hat{f}_n] = \mathbb{P}_n[h \circ \hat{f}_n]$.*

Remark 4.1 (One-step estimator). *There is however another interesting by-product of Lemma 4.1: The plug-in estimator $\hat{f}_n$ is also equivalent to the classically studied one-step estimator which is traditionally efficient for estimating integrated functionals, such as $\nu(h, f)$, over smoothness classes for f ; see e.g., Bickel et al. (1993), van der Vaart (1998), van der Vaart (2002). In particular, a first order influence function of $\nu(h, f)$ at P is $P[h \circ f + f(h' \circ f)]$ and consequently a one-step estimator obtained as a bias-corrected version of a $\tilde{f}$ -based plug-in estimator (here $\tilde{f}$ is any estimator of f) is given by*

$$\tilde{\nu} := \nu(h, \tilde{f}) + \mathbb{P}_n \left[h \circ \tilde{f} + \tilde{f}(h' \circ \tilde{f}) - \tilde{P}[h \circ \tilde{f} + \tilde{f}(h' \circ \tilde{f})] \right]$$

where $\tilde{P}$ denotes the probability measure associated with $\tilde{f}$. However, Lemma 4.1 implies that when $\tilde{f} = \hat{f}_n$ (the Grenander estimator)

$$\mathbb{P}_n \left[h \circ \tilde{f} + \tilde{f}(h' \circ \tilde{f}) - \tilde{P}[h \circ \tilde{f} + \tilde{f}(h' \circ \tilde{f})] \right] = 0 \quad \Rightarrow \quad \tilde{\nu} = \nu(h, \hat{f}_n).$$

Consequently, in this case of estimating a monotone density, the first order influence function based one-step estimator, obtained as a bias-corrected version of a Grenander-based plug-in estimator, coincides with the simple plug-in estimator (15).

Since all the three intuitive estimators in this problem turn out to be equivalent, one can expect them to be efficient as well (in a semiparametric sense). Indeed this is the case. The following result, an immediate consequence of Theorem 3.1, shows this.

Corollary 4.2. *Assume that conditions (A1)-(A2) and (A4) hold. Then,*

$$\sqrt{n} \left(\nu(h, \hat{f}_n) - \nu(h, f) \right) \xrightarrow{d} N(0, \sigma_{\text{eff}}^2(h, f)), \quad (16)$$

where, for $X \sim P$, $\sigma_{\text{eff}}^2(h, f) := \text{Var} \left(h'(f(X))f(X) + h(f(X)) \right) < \infty$.

Remark 4.2. *If $\sigma_{\text{eff}}^2(h, f) = 0$ then the above limit is degenerate. This can happen in the following two ways: (i) $f(\cdot)$ is a constant — this case has been dealt in detail in Section 3.1; and (ii) $h'(u)u + h(u) = c$ (for some constant $c \in \mathbb{R}$) which leads to $h(u) = C/u + c$ for some constant C . However, case (ii) is ruled out as because of assumption (A4) — the above $h(\cdot)$ does not satisfy (A4).*

5. Simulation study

In this section we illustrate the distributional convergence of our plug-in estimator $\tau(g, \hat{f}_n)$ (see (2)) for estimating $\tau(g, f)$. Let us consider the case of

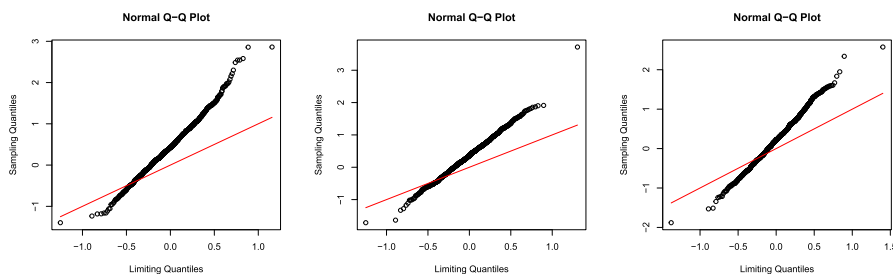


FIG 1. The Q-Q plots compare the sampling distribution of $\sqrt{n}\{\tau(g, \hat{f}_n) - \tau(g, f)\}$, for $g(z, x) = z^2$, against the limiting normal distribution $N(0, \sigma_{\text{eff}}^2(f))$ (see Theorem 2.3) when $F = \text{Exponential}(1)$. The three plots correspond to sample sizes $n = 5000, 20000$, and 100000 . Drawn in red is the $y = x$ line, which illustrates the difference between the finite sample distribution and the limiting distribution.

estimating the quadratic functional, i.e., $\tau(g, f) = \int f^2(x)dx$ which corresponds to $g(z, x) = z^2$.

We first consider the case when the true distribution function F is strictly concave and thus, the asymptotic distribution of $\tau(g, \hat{f}_n)$ is given by (9) (in Theorem 2.3) where $\sigma_{\text{eff}}^2(f)$ is given in (12) (with $p = 2$). For the Q-Q plots in Figure 1 we took f to be the exponential density with parameter 1 for which $\tau(g, f) = 1/2$ and $\sigma_{\text{eff}}^2(f) = 4/12$. The plots show the sampling distribution of $\sqrt{n}\{\tau(g, \hat{f}_n) - \tau(g, f)\}$ as the sample size increases ($n = 5000, 20000, 100000$). The sampling distributions is approximated from 1000 independent replications. The sampling distribution of $\sqrt{n}\{\tau(g, \hat{f}_n) - \tau(g, f)\}$, as n increases, seems to converge to $N(0, \sigma_{\text{eff}}^2(f))$, but even for moderate sample sizes there is a non-negligible bias. Although the Q-Q plots in Figure 1 show a visible deviation between the sampling quantiles and the limiting quantiles, the quantiles lie approximately on a straight line suggesting that the sampling distribution is well-approximated by a normal (with a non-zero mean). Although these findings broadly corroborate the theoretical result in Theorem 2.3, it is indeed true that the asymptotic regime seems to kick in quite slowly. We do not know if it is possible to approximate the finite sample distribution accurately (maybe using a resampling/bootstrap scheme). We leave this as an open research question.

Let us now consider the case when f is piecewise constant. To fix ideas, as in Beare and Fang (2017), let us take (with $c := 1 - 1/\sqrt{2}$)

$$F(x) := \begin{cases} \frac{x}{\sqrt{2}-1}, & x \in [0, c] \\ 2 - \sqrt{2} + (\sqrt{2} - 1)x, & x \in [c, 1]. \end{cases} \quad (17)$$

Then, from Beare and Fang (2017, Proposition 2.1), it follows that, for any $\xi \in C(\mathbb{R}_+)$,

$$\mathcal{M}'_F \xi(x) = \begin{cases} \mathcal{M}_{[0, c]} \xi(x), & x \in [0, c] \\ \mathcal{M}_{[c, 1]} \xi(x), & x \in [c, 1]. \end{cases}$$

Theorem 3.1 gives the asymptotic distribution of $\tau(g, \hat{f}_n) = \mu(h, \hat{f}_n) = \int f^2(x)dx$

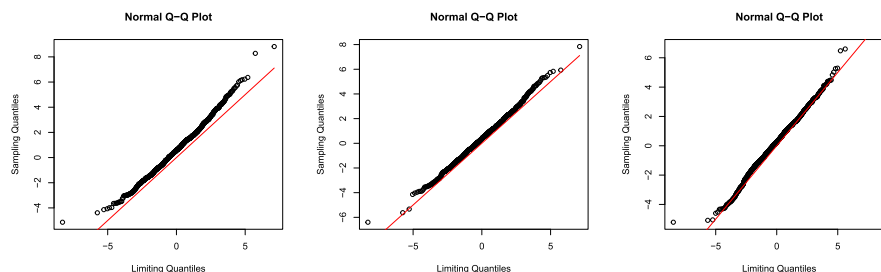


FIG 2. The same plots as in Figure 1 for the case when F is given in (17).

(where $h(x) = x^2$) in this case. Figure 2 shows the corresponding Q-Q plots (for sample sizes $n = 5000, 20000, 100000$) when F is defined in (17) for which $\tau(g, f) \approx 1.828$ and $\sigma_{\text{eff}}^2(f) \approx 3.314$. We can see that the sampling distribution of $\sqrt{n}\{\tau(g, \hat{f}_n) - \tau(g, f)\}$ converges to the desired limiting normal distribution. In this case, the normal approximation seems quite good even for moderately small sample sizes.

Let us now provide some numerical evidence to illustrate that the limiting distribution in (8) (in Theorem 2.2) need not always be normal when estimating a general functional of the form $\tau(g, f)$ (see (1)). In this simulation study we consider the case when f is piecewise constant; in particular, we take F as defined in (17). Consider estimating the functional $\tau(g, f) = \int x f^2(x) dx$ (i.e., $g(z, x) = xz^2$). The Q-Q plots in Figure 3 show the sampling distribution of $\sqrt{n}\{\tau(g, \hat{f}_n) - \tau(g, f)\}$ against the quantiles of a mean zero normal distribution with the efficient variance (which can be computed using (10)) as the sample size varies ($n = 5000, 20000$ and 100000). The Q-Q plots indicate several interesting features: (i) the sampling quantiles show a non-linear trend and is quite different from the theoretical quantiles (assuming the normal limit); (ii) the sampling distribution does not seem to change with n , suggesting that the sampling distribution is already close to its asymptotic limit (which is given by (8)). Thus, we conjecture that in this scenario, the sampling distribution of $\sqrt{n}\{\tau(g, \hat{f}_n) - \tau(g, f)\}$ does not converge to a normal limit with (mean 0 and) the efficient variance.

6. Discussion

In this paper we characterize the asymptotic distribution of the nonparametric maximum likelihood (NPML) based plug-in estimator (see Theorem 2.2) for estimating smooth integrated functionals of a monotone nonincreasing density f , under less restrictive assumptions on f . We also show that for a large class of functionals the asymptotic limit is normal with mean zero and the semiparametric efficient variance (see Theorem 3.1), under minimal assumptions on f . In particular, we do not assume that the underlying true density f is (i) smooth, or (ii) compactly supported, or (iii) lower bounded away from zero.

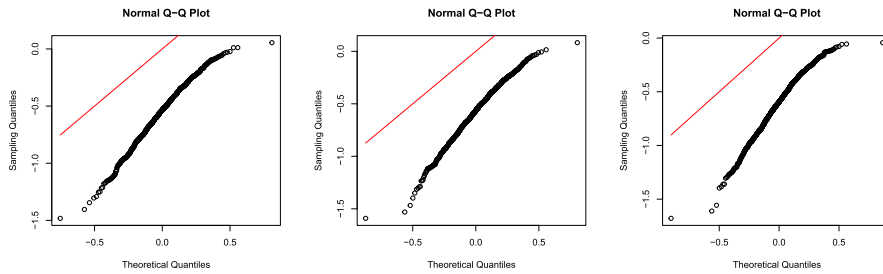


FIG 3. The same plots as in Figure 1 for the case when F is given in (17) and the functional of interest is $\tau(g, f) = \int x f^2(x) dx$ (i.e., $g(z, x) = xz^2$).

Finally, we believe that our basic proof idea can be used to study the asymptotic behavior of estimated functionals (based on the NPMLE) in other shape-restricted problems (e.g., decreasing convex densities or log-concave densities). A key open question in this direction is to express the NPMLE of the underlying density as a “nice” (e.g., Hadamard directionally differentiable) functional of the empirical distribution function of the data and to derive a corresponding weak convergence result analogous to Beare and Fang (2017, Theorem 2.1). As future research, we plan to explore this direction for other shape-restricted problems.

7. Proofs

7.1. Proof of Theorem 2.2

We first give an overview of the main ideas and steps involved in the proof of Theorem 2.2. Observe that,

$$\begin{aligned} \tau(g, \hat{f}_n) - \tau(g, f) &= \int_0^\infty \{g(\hat{f}_n(x), x) - g(f(x), x)\} dx \\ &= \int_0^\infty (\hat{f}_n(x) - f(x)) \dot{g}(f(x), x) dx + w_n, \end{aligned} \quad (18)$$

where

$$w_n := \frac{1}{2} \int_0^\infty (\hat{f}_n(x) - f(x))^2 \ddot{g}(\xi_n(x), x) dx,$$

with $|\xi_n(x) - f(x)| \leq |\hat{f}_n(x) - f(x)|$, for every $x \geq 0$.

Step 1: We claim that

$$n^{1/2} w_n = o_{\mathbb{P}}(1). \quad (19)$$

We prove (19) in Section 7.1.1. Below we complete the proof of Theorem 2.2 assuming the validity of (19).

Step 2: The first term in (18) can be handled as follows:

$$\begin{aligned} & \int_0^\infty (\hat{f}_n(x) - f(x)) \dot{g}(f(x), x) dx \\ &= \left[(\hat{F}_n(x) - F(x)) \dot{g}(f(x), x) \right]_0^\infty - \int_S (\hat{F}_n(x) - F(x)) d[\dot{g}(f(x), x)]. \end{aligned}$$

As $\dot{g}(f(\cdot), \cdot)$ is of finite total variation (by (A3)-(iii,b)) we have $(1 - F(x)) \dot{g}(f(x), x) \rightarrow 0$ as $x \rightarrow +\infty$. Thus,

$$\begin{aligned} \sqrt{n} \int_0^\infty (\hat{f}_n(x) - f(x)) \dot{g}(f(x), x) dx &= -\sqrt{n} \int_S (\hat{F}_n(x) - F(x)) d[\dot{g}(f(x), x)] \\ &= -\int_S \hat{\mathbb{D}}_n(x) d[\dot{g}(f(x), x)] \\ &\xrightarrow{d} -\int_S \hat{\mathbb{G}}(x) d[\dot{g}(f(x), x)] \end{aligned}$$

where the last step follows from Beare and Fang (2017, Theorem 2.1) and a version of the continuous mapping theorem: Observe that (i) $\hat{\mathbb{D}}_n \xrightarrow{d} \hat{\mathbb{G}}$ in the space $\ell^\infty(\mathbb{R}_+)$; (ii) for any $\psi : \mathbb{R}_+ \rightarrow \mathbb{R}$ uniformly bounded, $\psi \mapsto \int_S \psi(x) d[\dot{g}(f(x), x)]$ is a continuous function. This proves that (8) holds.

7.1.1. Proof of (19)

Fix $\epsilon > 0$ and $\eta > 0$. For $M^* > 0$ a suitable constant (to be chosen later) we define

$$A_n := \{\|\hat{f}_n\|_\infty \leq M^* f(0+)\}, \quad \text{for } n = 1, 2, \dots$$

We have to study the probability

$$\begin{aligned} \mathbb{P}\left(n^{1/2}|w_n| > \eta\right) &\leq \mathbb{P}\left(\sup_{x \in \mathcal{S}} |\ddot{g}(\xi_n(x), x)| \|\hat{f}_n - f\|_2^2 > 2\eta n^{-1/2}\right) \\ &\leq \mathbb{P}\left(\sup_{x \in \mathcal{S}, z \in [0, (M^*+1)f(0+)]} |\ddot{g}(z, x)| \|\hat{f}_n - f\|_2^2 > 2\eta n^{-1/2}, A_n\right) + \mathbb{P}(A_n^c) \\ &\leq \mathbb{P}\left(\|\hat{f}_n - f\|_2^2 > 2K^{-1}\eta n^{-1/2}, A_n\right) + \mathbb{P}(A_n^c), \end{aligned} \tag{20}$$

where we have used assumption (A3) which implies that

$$\sup_{x \in \mathcal{S}, z \in [0, (M^*+1)f(0+)]} |\ddot{g}(z, x)| \leq K,$$

and for any function $\psi : \mathbb{R} \rightarrow \mathbb{R}$, $\|\psi\|_2^2 := \int \psi^2(x) dx$.

First, observe that by Woodroffe and Sun (1993, Theorem 2 and Remark 3), the sequence of random variables $\frac{\hat{f}_n(0+)}{f(0+)}$ is asymptotically P -tight and consequently, given any $\epsilon > 0$ there exists $M^* > 0$ (depending on $\epsilon > 0$) such that for n large enough (depending on $\epsilon > 0$)

$$\mathbb{P}\left(\|\hat{f}_n\|_\infty > M^* f(0+)\right) \leq \frac{\epsilon}{2}. \tag{21}$$

Thus, the second term in (20) is bounded from above by $\epsilon/2$.

Next, note that $\|\hat{f}_n\|_\infty \leq M^* f(0+)$ implies that

$$\begin{aligned} \|\hat{f}_n - f\|_2^2 &\leq 2(\sqrt{M^*} + 1)^2 f(0+) \frac{1}{2} \int \left(\sqrt{\hat{f}_n(x)} - \sqrt{f(x)} \right)^2 dx \\ &= 2(\sqrt{M^*} + 1)^2 f(0+) H^2(\hat{P}_n, P), \end{aligned}$$

where

$$H^2(\hat{P}_n, P) := \frac{1}{2} \int \left(\sqrt{\hat{f}_n(x)} - \sqrt{f(x)} \right)^2 dx$$

is the Hellinger divergence between $\hat{P}_n$ and P . As assumptions (A1)-(A2) are satisfied, by van de Geer (2000, Theorem 7.12), there exists $c > 0$ such that for all $\tilde{C} \geq c$,

$$\mathbb{P}\left(H^2(\hat{P}_n, P) > \tilde{C}^2 n^{-2/3}\right) \leq c \exp\left(-\frac{\tilde{C}^2 n^{1/3}}{c^2}\right).$$

Consequently, for an appropriate C^* (depending on M^*) and $\tilde{c}$, we have

$$\mathbb{P}\left(\|\hat{f}_n - f\|_2^2 > C^* n^{-2/3}, \|\hat{f}_n\|_\infty \leq M^* f(0+)\right) \leq c \exp\left(-\tilde{c} n^{1/3}\right) \leq \frac{\epsilon}{2}, \quad (22)$$

for n large enough (depending on ϵ). Consequently, using (22) and (21), for n sufficiently large (which, in particular, implies that $C^* n^{-2/3} \leq 2K^{-1}\eta n^{-1/2}$), (20) yields

$$\mathbb{P}\left(n^{1/2}|w_n| > \eta\right) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2}.$$

As ϵ and η are arbitrary, we conclude $w_n = o_{\mathbb{P}}(n^{-1/2})$ as $n \rightarrow \infty$. $\square$

7.2. Proof of Theorem 2.3

We first note that $\sigma_{\text{eff}}^2(g, f) < +\infty$. This follows from the fact that $\dot{g}^2(f(\cdot), \cdot)$ is a bounded function on $[0, \infty)$ (as $\dot{g}(f(\cdot), \cdot)$ is a function of bounded total variation on $[0, \infty)$).

Now, to show (9), first note that when F is strictly concave on $\mathbb{R}_+$, it follows from Proposition 2.1 that, $\mathcal{M}'_F \xi = \xi$ for all $\xi \in C(\mathbb{R}_+)$. Thus, if F is strictly concave then $\hat{\mathbb{G}} = \mathbb{G}$ and consequently,

$$Y = - \int_{\mathcal{S}} \mathbb{G}(x) d\psi(x),$$

where

$$\psi(x) := \dot{g}(f(x), x) \quad x \in \mathcal{S}.$$

Let $\mathcal{S} = [0, s]$ where $s \in \mathbb{R}$ or $s = +\infty$. As $\psi(\cdot)$ is of finite total variation on $\mathcal{S}$ (by (A3)-(iii,b)), we shall assume without loss of generality that $\psi(s) := \lim_{x \rightarrow s} \psi(x) = 0$ (otherwise we can work with $\psi(\cdot) - \psi(s)$ instead of $\psi(\cdot)$ and that would not change the value of Y).

If $s \in \mathbb{R}$ then it is easy to see that Y is a centered Gaussian random variable. In the following we show that even when $s = +\infty$, Y is a centered Gaussian random variable. We write $Y = -\int_0^N \mathbb{G}(x) d\psi(x) - \int_N^\infty \mathbb{G}(x) d\psi(x)$. Now $\mathbb{E}|\int_N^\infty \mathbb{G}(x) d\psi(x)| \leq \int_N^\infty \mathbb{E}|\mathbb{G}(x)| d\psi(x)$ which converges to 0 (by (A3)-(iii,b)) as $N \rightarrow \infty$. Consequently Y is the in probability (and hence in distribution) limit of $-\int_0^N \mathbb{G}(x) d\psi(x)$ as $N \rightarrow \infty$. Now for every N , $-\int_0^N \mathbb{G}(x) d\psi(x)$ is a Gaussian random variable. Therefore to show (9), we only need to show that the mean and variance of $-\int_0^N \mathbb{G}(x) d\psi(x)$ converge to 0 and $\sigma_{\text{eff}}^2(g, f)$ respectively as $N \rightarrow \infty$. Indeed by Fubini's Theorem $-\int_0^N \mathbb{G}(x) d\psi(x)$ has mean 0 and this completes the mean part of the convergence.

We next show that $\text{Var}(Y)$ matches $\sigma_{\text{eff}}^2(g, f)$. If $s = +\infty$, the same proof will also establish the validity of the convergence of the variance of $-\int_0^N \mathbb{G}(x) d\psi(x)$ to $\sigma_{\text{eff}}^2(g, f)$ (by simply interpreting integrals appearing in the proof with respect to ψ on $[0, \infty)$ as limits of integrals with respect to ψ on $[0, N]$).

To operationalize our argument, we begin with the following simplification:

$$\begin{aligned} \int_0^\infty \mathbb{G}(x) d\psi(x) &= \int_0^\infty \mathbb{B} \circ F(x) d\psi(x) \\ &\stackrel{d}{=} \int_0^s \mathbb{W} \circ F(x) d\psi(x) - \mathbb{W}(1) \int_0^s F(x) d\psi(x), \end{aligned}$$

where $\mathbb{W}$ is the standard Brownian motion on $[0, 1]$ and $\stackrel{d}{=}$ stands for equality in distribution. To simplify notation, in the following every integral is taken on the set $[0, s]$. Therefore,

$$\begin{aligned} \text{Var}(Y) &= \text{Var}\left(\int \mathbb{W} \circ F(x) d\psi(x)\right) + \text{Var}\left(\mathbb{W}(1) \int F(x) d\psi(x)\right) \\ &\quad - 2 \text{Cov}\left(\int \mathbb{W} \circ F(x) d\psi(x), \mathbb{W}(1) \int F(x) d\psi(x)\right) \\ &= \mathbb{E}\left(\int \mathbb{W} \circ F(x) d\psi(x)\right)^2 + \left(\int F(x) d\psi(x)\right)^2 \\ &\quad - 2 \int F(x) d\psi(x) \mathbb{E}\left(\mathbb{W}(1) \int \mathbb{W} \circ F(x) d\psi(x)\right). \end{aligned}$$

Now, by simple integration by parts, since $\lim_{x \rightarrow \infty} \psi(x) = 0$,

$$\int F(x) d\psi(x) = - \int \psi(x) dF(x),$$

and consequently it is enough to prove the following lemma to complete the proof of Theorem 2.3 (as then $\text{Var}(Y)$ reduces to $\text{Var}(\psi(X)) = \text{Var}(\dot{g}(f(X), X))$).

Lemma 7.1. *The following identities hold under assumptions (A1)-(A3):*

$$\mathbb{E}\left(\mathbb{W}(1) \int \mathbb{W} \circ F(x) d\psi(x)\right) = \int F(x) d\psi(x), \quad \text{and}$$

$$\mathbb{E} \left(\int \mathbb{W} \circ F(x) d\psi(x) \right)^2 = \int \psi^2(x) dF(x).$$

Proof. The first identity is obvious since under our assumptions one can interchange the integral and expectation (in particular Fubini's Theorem is applicable since $\int_0^s \mathbb{E}(|\mathbb{G}(x)|) d(\dot{g}(f(x), x)) < +\infty$ by part (iii,b) of (A3)) and observe that

$$\mathbb{E}(\mathbb{W}(1)\mathbb{W}(F(x))) = \text{Cov}(\mathbb{W}(1), \mathbb{W}(F(x))) = \min\{1, F(x)\} = F(x).$$

For the second identity (once again we freely interchange integrals and expectation by Fubini's Theorem which is applicable by part (iii,b) of (A3)), note that

$$\begin{aligned} \mathbb{E} \left(\int_0^s \mathbb{W} \circ F(x) d\psi(x) \right)^2 &= \mathbb{E} \left(\int_0^s \int_0^s \mathbb{W} \circ F(x) \mathbb{W} \circ F(x') d\psi(x) d\psi(x') \right) \\ &= 2 \int_0^s \int_0^{x' \wedge s} F(x) d\psi(x) d\psi(x'). \end{aligned} \quad (23)$$

Now, let $v(x') := \int_0^{x'} F(x) d\psi(x)$, for $x' \geq 0$. Then by repeated integration by parts and noting that $\lim_{x \rightarrow s} \psi(x) = 0$ by our assumption,

$$\begin{aligned} &\int_0^s \int_0^{x' \wedge s} F(x) d\psi(x) d\psi(x') \\ &= \lim_{x' \rightarrow s} v(x') \psi(x') - v(0) \psi(0) - \int_0^s \psi(x') dv(x') \\ &= - \int_0^s \psi(x') dv(x') \\ &= - \left[\int_0^s \psi(x') F(x') d\psi(x') \right] \\ &= - \left[\lim_{x' \rightarrow s} F(x') \frac{1}{2} \psi^2(x') - F(0) \frac{1}{2} \psi^2(0) - \frac{1}{2} \int_0^s \psi^2(x') dF(x') \right] \\ &= \frac{1}{2} \int_0^s \psi^2(x') dF(x'). \end{aligned} \quad (24)$$

Plugging back (24) into (23) completes the proof of Lemma 7.1. $\square$

7.3. Proof of Theorem 3.1

Proposition 2.1 will help us demonstrate that indeed Y (as defined in (16)) has a Gaussian distribution with the desired efficient variance. However, to give more intuition to the reader, we begin by discussing two simple scenarios.

Case (i): First, let F be strictly concave on $\mathbb{R}_+$. From Proposition 2.1 it follows that $\mathcal{M}'_\theta$ is linear if and only if θ is strictly concave on $\mathbb{R}_+$ (Beare and

Fang (2017)). Thus, $\mathcal{M}'_F \xi = \xi$ for all $\xi \in C(\mathbb{R}_+)$. Thus, if F is strictly concave then $\hat{\mathbb{G}} = \mathbb{G}$ and consequently

$$Y = - \int \mathbb{G}(x) d[h'(f(x))].$$

In this form, it is easy to see that Y is a centered Gaussian random variable. The fact that its variance matches the efficiency bound can be demonstrated as in the proof of Theorem 2.3 in Section 7.2 (with $\psi(x) = h'(f(x))$).

Case (ii): Next consider the other extreme case, i.e., F is piecewise affine. In this case f is piecewise constant with jumps (say) at $0 < t_1 < t_2 < \dots < t_k < \infty$ and values $v_1 > \dots > v_k$, for some $k \geq 2$ integer, i.e.,

$$f(x) = \sum_{i=1}^k v_i \mathbf{1}_{(t_{i-1}, t_i]}(x),$$

where $t_0 \equiv 0$. As f only takes $k+1$ values (namely, $\{v_1, \dots, v_k, 0\}$) we have that

$$Y = - \sum_{i=1}^k a_i \hat{\mathbb{G}}(t_i)$$

where

$$a_i := \{h'(f(t_i-)) - h'(f(t_i+))\}.$$

Further, from Proposition 2.1, we have, for every $i = 1, \dots, k$, $\hat{\mathbb{G}}(t_i) \equiv \mathcal{M}'_F \mathbb{G}(t_i) = \mathbb{G}(t_i)$. As $\mathbb{G}(t_k) = 0$ a.s., we have

$$Y = \sum_{i=1}^{k-1} a_i \mathbb{G}(t_i) = \int \mathbb{G}(x) d[h'(f(x))]$$

which is the same distribution as in the case when F was strictly concave.

Since in both the above cases we have $Y = \int \mathbb{G}(x) d[h'(f(x))]$, one can conjecture that the result should be generally true for any concave F . This intuition indeed turns out to be correct as shown below.

Let $\psi(x) := h'(x)$, for $x \geq 0$. Note that by assumption (A3) applied to $g(z, x) = h(z)$, we have that $\psi(f(x))$ has bounded total variation. Consequently, let us assume that the function $\psi \circ f$ is monotone (otherwise we can split $\psi \circ f$ as the difference of two monotone functions and apply the following to each part). Note that it is enough to show that for any $\xi \in C(\mathbb{R}_+)$,

$$\int \mathcal{M}'_F \xi(x) d[\psi(f(x))] = \int \xi(x) d[\psi(f(x))].$$

Then we can take ξ to be $\mathbb{G} \equiv \mathbb{B} \circ F$ (which is almost surely continuous) and we will obtain the desired result.

Let

$$I := \{x \in (0, \infty) : F \text{ is affine in an open interval around } x\}.$$

It is easy to show that I is a Borel measurable set. Consider the collection $\{T_{F,x} : x \in I\}$, where $T_{F,x}$ is defined in Proposition 2.1 (also see Beare and Fang (2017, Proposition 2.1)). First note that $x \in T_{F,x}$, for every $x \in I$. We now claim that there are at most countably many such distinct $T_{F,x}$'s, as x varies in I . This follows from the fact that for any x in the support of F , F has a strictly positive slope on $T_{F,x}$ (i.e., F cannot be affine with slope 0 on $T_{F,x}$, as F is concave and nondecreasing and x is in the support of F) which implies that we can associate a unique rational number in the range of F to this interval $T_{F,x}$. Let $\{T_{F,x_i}\}_{i \geq 1}$ (for $x_i \in (0, \infty)$) be an enumeration of this countable collection $\{T_{F,x} : x \in I\}$. Obviously, $I \subset \cup_{i=1}^{\infty} T_{F,x_i}$.

Let $\xi \in C(\mathbb{R}_+)$. Then,

$$\begin{aligned} & \left| \int \mathcal{M}'_F \xi(x) d[\psi(f(x))] - \int \xi(x) d[\psi(f(x))] \right| \\ & \leq \int_I |\mathcal{M}'_F \xi(x) - \xi(x)| d[\psi(f(x))] + \int_{[0, \infty) \setminus I} |\mathcal{M}'_F \xi(x) - \xi(x)| d[\psi(f(x))] \\ & \leq \sum_{i=1}^{\infty} \int_{T_{F,x_i}} |\mathcal{M}'_F \xi(x) - \xi(x)| d[\psi(f(x))] \end{aligned}$$

where the last inequality follows from the fact that for $x \in [0, \infty) \setminus I$, $\mathcal{M}'_F \xi(x) = \xi(x)$ (see Beare and Fang (2017, Remark 2.2)) and the above integrals are viewed as a Lebesgue-Stieltjes. Notice now that on the open interval T_{F,x_i} , f is constant (as F is affine on T_{F,x_i}). Hence each of the integrals in the last display equals 0. This completes the proof. $\square$

7.4. Proof of Lemma 4.1

Let $X_{(1)} < X_{(2)} < \dots < X_{(n)}$ be the ordered data points and let

$$\hat{\theta}_i = \hat{f}_n(X_{(i)}), \quad \text{for } i = 1, \dots, n,$$

be the fitted values of the Grenander estimator at the ordered data points. A simple characterization of $\hat{\theta} := (\hat{\theta}_1, \dots, \hat{\theta}_n)$ is given by

$$\hat{\theta} = \arg \min_{\theta \in \mathcal{C}} \sum_{i=1}^n \left(\frac{1}{n(X_{(i)} - X_{(i-1)})} - \theta_i \right)^2 (X_{(i)} - X_{(i-1)}),$$

where

$$\mathcal{C} = \{\theta \in \mathbb{R}^n : \theta_1 \geq \dots \geq \theta_n\},$$

and $X_{(0)} \equiv 0$; see e.g., Groeneboom and Jongbloed (2014, Exercise 2.5). From the characterization of projection onto the closed convex cone $\mathcal{C}$ it follows that

for any function $h : \mathbb{R} \rightarrow \mathbb{R}$ (see e.g., Robertson et al. (1988, Theorem 1.3.6))

$$\begin{aligned} & \sum_{i=1}^n \left(\frac{1}{n(X_{(i)} - X_{(i-1)})} - \hat{\theta}_i \right) (X_{(i)} - X_{(i-1)}) h(\hat{\theta}_i) = 0 \\ \Leftrightarrow & \frac{1}{n} \sum_{i=1}^n h(\hat{\theta}_i) = \sum_{i=1}^n h(\hat{\theta}_i) \hat{\theta}_i (X_{(i)} - X_{(i-1)}) \\ \Leftrightarrow & \mathbb{P}_n[h \circ \hat{f}_n] = \hat{P}_n[h \circ \hat{f}_n]. \end{aligned}$$

The last step follows from the fact that

$$\hat{P}_n[h \circ \hat{f}_n] = \int h(\hat{f}_n(x)) \hat{f}_n(x) dx = \sum_{i=1}^n h(\hat{\theta}_i) \hat{\theta}_i (X_{(i)} - X_{(i-1)}),$$

as $\hat{f}_n(\cdot)$ is constant on the interval $(X_{(i-1)}, X_{(i)}]$, for every $i = 1, \dots, n$. $\square$

7.5. Proof of Theorem 3.2

Note that,

$$\mu(h, \hat{f}_n) - \mu(h, f) = \int_0^1 [h(\hat{f}_n(x)) - h(f(x))] dx.$$

For $x \in [0, 1]$, note that for some $\xi_n(x)$ lying between $f(x) \equiv 1$ and $\hat{f}_n(x)$ one has by a four term Taylor expansion (allowed by the assumption that $h \in C^4([0, \infty))$)

$$\begin{aligned} & \frac{n \int_0^1 [h(\hat{f}_n(x)) - h(1)] dx - \frac{1}{2} h''(1) \log n}{\sqrt{3[\frac{1}{2} h''(1)]^2 \log n}} \\ &= \frac{n \int_0^1 (\hat{f}_n(x) - 1)^2 dx - \log n}{\sqrt{3 \log n}} \\ &+ \frac{1}{6} \frac{n \int_0^1 h^{(3)}(1) (\hat{f}_n(x) - 1)^3 dx}{\sqrt{3[\frac{1}{2} h''(1)]^2 \log n}} + \frac{1}{24} \frac{n \int_0^1 h^{(4)}(\xi_n(x)) (\hat{f}_n(x) - 1)^4 dx}{\sqrt{3[\frac{1}{2} h''(1)]^2 \log n}}. \end{aligned} \quad (25)$$

Therefore, if we prove that

$$\frac{n \int_0^1 (\hat{f}_n(x) - 1)^k dx}{\sqrt{\log n}} = o_{\mathbb{P}}(1), \quad \text{for } k = 3, 4,$$

then

$$\begin{aligned} \left| \frac{n \int_0^1 h^{(4)}(\xi_n(x)) (\hat{f}_n(x) - 1)^4 dx}{\sqrt{3 \log n}} \right| &\leq \sup_{z \in [0, \|\hat{f}_n\|_{\infty} \vee 1]} |h^{(4)}(z)| \frac{n \int_0^1 (\hat{f}_n(x) - 1)^4 dx}{\sqrt{3 \log n}} \\ &= o_{\mathbb{P}}(1), \end{aligned}$$

since $\|\hat{f}_n\|_\infty = O_{\mathbb{P}}(1)$ by (Woodroffe and Sun, 1993, Equation (2)) and $h^{(4)}$ is bounded on compact intervals. Therefore, modulo the claim above, one readily obtains the desired result (14) from (25) and (13).

The next two subsections are devoted to understanding $\frac{n \int_0^1 (\hat{f}_n(x)-1)^k dx}{\sqrt{\log n}}$ for $k = 3, 4$. In the analysis, we crucially use Groeneboom and Pyke (1983, Theorem 2.1) and consequently need the following notation to proceed. Let

$$T_n := \sum_{j=1}^n j N_j, \quad S_n := \sum_{j=1}^n \sum_{i=1}^{N_j} S_{ji},$$

where $\{S_{ji} : i \geq 1, j \geq 0\}$ are independent random variables with $S_{ji} \sim \text{Gamma}(j, 1)$, for $j \geq 1$, and $S_{0i} \sim \text{Gamma}(1, 1)$, and $N_j \sim \text{Poisson}(1/j)$ for $j \geq 1$ and $N_0 = 1$ a.s.; further, $\{N_j\}$ and $\{S_{ji}\}$ are independent. Also let,

$$W_n = \frac{1}{n} \sum_{j=1}^n j N_j, \quad V_n = n^{-1/2} \sum_{j=1}^n \sum_{i=1}^{N_j} (S_{ji} - j).$$

With this notation we are ready to analyze $\frac{n \int_0^1 (\hat{f}_n(x)-1)^k dx}{\sqrt{\log n}}$ for $k = 3, 4$.

7.5.1. Proof of $\frac{n \int_0^1 (\hat{f}_n(x)-1)^3 dx}{\sqrt{\log n}} = o_{\mathbb{P}}(1)$

Let us start by considering the distribution of

$$L_n := \int_0^1 (\hat{f}_n(x) - 1)^3 dx = \int \hat{f}_n^3(x) dx - 3 \int \hat{f}_n^2(x) dx + 2.$$

By an argument similar to that in Groeneboom and Pyke (1983, Section 3), it can be seen that L_n has the same distribution as that of

$$L_n^* := \frac{1}{n} \sum_{j=1}^n j^3 \sum_{i=1}^{N_j} \frac{1}{S_{ji}^2} - \frac{3}{n} \sum_{j=1}^n j^2 \sum_{i=1}^{N_j} \frac{1}{S_{ji}} + 2 \Big| \{T_n = n, S_n = n\}.$$

For constants c_0, c_n, γ_n , we define the sequence of random variables

$$\begin{aligned} U_n(c_0, c_n, \gamma_n) := & \frac{1}{c_n} \left[\sum_{j=1}^n j^3 \sum_{i=1}^{N_j} \left(\frac{1}{S_{ji}^2} - \frac{1}{j^2} \right) - 3 \sum_{j=1}^n j^2 \sum_{i=1}^{N_j} \left(\frac{1}{S_{ji}} - \frac{1}{j} \right) \right. \\ & \left. + c_0 \sum_{j=1}^n \sum_{i=1}^{N_j} (S_{ji} - j) - \gamma_n \right]. \end{aligned}$$

The particular choice of c_0, c_n, γ_n is crucial in the analysis that follows. In particular, a specific choice for the constant c_0 implies certain fine-tuned cancellations

which are necessary for controlling the asymptotic variance of U_n at a desired level. It is worth noting that such a definition and the eventual choice of $c_0 = 1$ is also present in Groeneboom and Pyke (1983, Equation (3.2)). Our choice of c_0, c_n, γ_n is more tailored to the current problem. Finally, conditional on the joint event that $W_n = 1$ and $V_n = 0$ (i.e., $T_n = n$ and $S_n = n$) one has

$$U_n(c_0, c_n, \gamma_n) | \{W_n = 1, V_n = 0\} \stackrel{d}{=} c_n^{-1}(nL_n^* - \gamma_n) | \{T_n = n, S_n = n\}. \quad (26)$$

The following lemma, crucial to our proof, is the analogue of Groeneboom and Pyke (1983, Lemma 3.1); a correction to the characteristic function formula in Groeneboom and Pyke (1983, Lemma 3.1) was recently pointed out in Groeneboom (2019, Lemma 3.1).

Lemma 7.2. *Let $c_n = \sqrt{\log n}$, $\gamma_n = 0$ and $c_0 = -1$. Then $(U_n, V_n, W_n) \xrightarrow{d} (U, V, W)$ where $U = \delta_0$, the Dirac measure at 0, and (V, W) has the infinitely divisible characteristic function $\phi_{V,W}(s, t) = \exp\left(\int_0^1 (e^{itu - s^2 u/2} - 1)u^{-1} du\right)$.*

Proof. With the choice of c_0, c_n, γ_n provided in the statement of the theorem, denote $U_n \equiv U_n(c_0, c_n, \gamma_n)$. We only prove the fact that $U_n \xrightarrow{\mathbb{P}} 0$. The subsequent conclusion follows along similar lines as in the proof of Groeneboom and Pyke (1983, Lemma 3.1). In the following analysis we repeatedly use the fact that $\mathbb{E}(S_{ji}^{-k}) = \frac{\Gamma(j-k)}{\Gamma(j)}$ for $j > k$.

First, define U_n^* to be U_n without its first 3 summands in j , i.e.,

$$U_n^* := \frac{1}{c_n} \left[\sum_{j=4}^n j^3 \sum_{i=1}^{N_j} \left(\frac{1}{S_{ji}^2} - \frac{1}{j^2} \right) - 3 \sum_{j=4}^n j^2 \sum_{i=1}^{N_j} \left(\frac{1}{S_{ji}} - \frac{1}{j} \right) - \sum_{j=4}^n \sum_{i=1}^{N_j} (S_{ji} - j) \right].$$

We will show that $U_n^* \xrightarrow{\mathbb{P}} 0$ and as $U_n - U_n^* \xrightarrow{\mathbb{P}} 0$ (since $c_n \rightarrow \infty$), this will establish the desired result. Note that,

$$\begin{aligned} c_n \mathbb{E}(U_n^*) &= \sum_{j=4}^n j^3 \mathbb{E}(N_j) \left[\frac{\Gamma(j-2)}{\Gamma(j)} - \frac{1}{j^2} \right] - 3 \sum_{j=4}^n j^2 \mathbb{E}(N_j) \left[\frac{\Gamma(j-1)}{\Gamma(j)} - \frac{1}{j} \right] \\ &= \sum_{j=4}^n \frac{4}{(j-1)(j-2)}. \end{aligned}$$

As $\limsup_{n \rightarrow \infty} \left| \sum_{j=4}^n \frac{4}{(j-1)(j-2)} \right| < \infty$, one has $\mathbb{E}(U_n^*) \rightarrow 0$ (as $c_n \rightarrow \infty$). Subsequently it is enough to prove that $\text{Var}(U_n^*) \rightarrow 0$ which in turn is implied by proving that $\limsup_{n \rightarrow \infty} \text{Var}(c_n U_n^*) < \infty$. First note that

$$\begin{aligned} \text{Var}(\mathbb{E}(c_n U_n^* | N_1, \dots, N_n)) &= \sum_{j=4}^n \left(\frac{4j}{(j-1)(j-2)} \right)^2 \text{Var}(N_j) \\ &= \sum_{j=4}^n \left(\frac{4j}{(j-1)(j-2)} \right)^2 \frac{1}{j}, \end{aligned}$$

and consequently, $\limsup_{n \rightarrow \infty} \text{Var}(\mathbb{E}(c_n U_n^* | N_1, \dots, N_n)) < \infty$. Next, note that

$$\begin{aligned} & \text{Var}(c_n U_n^* | N_1, \dots, N_n) \\ &= \sum_{j=4}^n N_j \left[j^6 \text{Var}\left(\frac{1}{S_{j1}^2}\right) + 9j^4 \text{Var}\left(\frac{1}{S_{j1}}\right) + \text{Var}(S_{j1}) \right. \\ & \quad \left. - 6j^5 \text{Cov}\left(\frac{1}{S_{j1}^2}, \frac{1}{S_{j1}}\right) - 2j^3 \text{Cov}\left(\frac{1}{S_{j1}^2}, S_{j1}\right) + 6 \text{Cov}\left(\frac{1}{S_{j1}}, S_{j1}\right) \right] \\ &= \sum_{j=4}^n N_j \frac{15j^5 + 129j^4 - 404j^3 + 116j^2 + 48j}{(j-1)^2(j-2)^2(j-3)(j-4)}. \end{aligned}$$

Consequently,

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \mathbb{E}(\text{Var}(c_n U_n^* | N_1, \dots, N_n)) \\ &= \limsup_{n \rightarrow \infty} \sum_{j=4}^n \frac{15j^4 + 129j^3 - 404j^2 + 116j + 48}{(j-1)^2(j-2)^2(j-3)(j-4)} < \infty. \end{aligned}$$

This completes the proof of the lemma. $\square$

We finally note that Lemma 7.2 immediately implies that U_n is asymptotically independent of (W_n, V_n) and that $U_n - U_m \xrightarrow{\mathbb{P}} 0$ for any $n, m \rightarrow \infty$. Consequently, the analysis provided in the proof of Groeneboom and Pyke (1983, Theorem 3.1) goes through verbatim yielding that

$$U_n | \{W_n = 1, V_n = 0\} \xrightarrow{d} 0$$

which in turn proves that (using (26)) $c_n^{-1}(nL_n^* - \gamma_n) | \{T_n = n, S_n = n\} \xrightarrow{d} 0$ for our choice of $\gamma_n = 0$. This yields the desired result.

7.5.2. Proof of $\frac{n \int_0^1 (\hat{f}_n(x) - 1)^4 dx}{\sqrt{\log n}} = o_{\mathbb{P}}(1)$

The proof technique is similar to that of the case $\frac{\int_0^1 (\hat{f}_n(x) - 1)^3 dx}{\sqrt{\log n}}$ albeit with much more cumbersome details and a different choice of c_0 . As before, let us start by considering the distribution of

$$L_n := \int_0^1 (\hat{f}_n(x) - 1)^4 dx = \int \hat{f}_n^4(x) dx - 4 \int \hat{f}_n^3(x) dx + 6 \int \hat{f}_n^2(x) dx - 3.$$

As argued in Groeneboom and Pyke (1983, Section 3), it is equivalent to studying the (conditional) distribution of

$$L_n^* := \frac{1}{n} \sum_{j=1}^n j^4 \sum_{i=1}^{N_j} \frac{1}{S_{ji}^3} - \frac{4}{n} \sum_{j=1}^n j^3 \sum_{i=1}^{N_j} \frac{1}{S_{ji}^2} + \frac{6}{n} \sum_{j=1}^n j^2 \sum_{i=1}^{N_j} \frac{1}{S_{ji}} - 3 \Big| \{T_n = n, S_n = n\}.$$

We once again introduce a sequence of random variables

$$U_n(c_0, c_n, \gamma_n) := \frac{1}{c_n} \left[\sum_{j=1}^n j^4 \sum_{i=1}^{N_j} \left(\frac{1}{S_{ji}^3} - \frac{1}{j^3} \right) - 4 \sum_{j=1}^n j^3 \sum_{i=1}^{N_j} \left(\frac{1}{S_{ji}^2} - \frac{1}{j^2} \right) \right. \\ \left. + 6 \sum_{j=1}^n j^2 \sum_{i=1}^{N_j} \left(\frac{1}{S_{ji}} - \frac{1}{j} \right) + c_0 \sum_{j=1}^n \sum_{i=1}^{N_j} (S_{ji} - j) - \gamma_n \right].$$

Then, conditional on the joint event that $W_n = 1$ and $V_n = 0$ (i.e., $T_n = n$ and $S_n = n$) one has

$$U_n(c_0, c_n, \gamma_n) | \{W_n = 1, V_n = 0\} \stackrel{d}{=} c_n^{-1} (nL_n^* - \gamma_n) | \{T_n = n, S_n = n\}.$$

Once again we have a crucial lemma which is the analogue of Groeneboom and Pyke (1983, Lemma 3.1); a correction to the characteristic function formula in Groeneboom and Pyke (1983, Lemma 3.1) was recently pointed out in Groeneboom (2019).

Lemma 7.3. *Let $c_n = \sqrt{\log n}$, $\gamma_n = 0$ and $c_0 = 1$. Then $(U_n, V_n, W_n) \xrightarrow{d} (U, V, W)$ where $U = \delta_0$ and (V, W) has the infinitely divisible characteristic function $\phi_{V,W}(s, t) = \exp \left(\int_0^1 (e^{itu - s^2 u/2} - 1) u^{-1} du \right)$.*

Proof. With the above choice of c_0, c_n, γ_n , denote by $U_n \equiv U_n(c_0, c_n, \gamma_n)$. We only prove the fact that $U_n \xrightarrow{\mathbb{P}} 0$. The subsequent conclusion follows along similar lines as in the proof of Groeneboom and Pyke (1983, Lemma 3.1) and Lemma 7.2.

First, define U_n^* to be U_n without its first 4 summands in j , i.e.,

$$U_n^* = c_n^{-1} \left[\sum_{j=5}^n j^4 \sum_{i=1}^{N_j} \left(\frac{1}{S_{ji}^3} - \frac{1}{j^3} \right) - 4 \sum_{j=5}^n j^3 \sum_{i=1}^{N_j} \left(\frac{1}{S_{ji}^2} - \frac{1}{j^2} \right) \right. \\ \left. + 6 \sum_{j=5}^n j^2 \sum_{i=1}^{N_j} \left(\frac{1}{S_{ji}} - \frac{1}{j} \right) + \sum_{j=5}^n \sum_{i=1}^{N_j} (S_{ji} - j) \right].$$

We show that $U_n^* \xrightarrow{\mathbb{P}} 0$ and indeed by $U_n - U_n^* \xrightarrow{\mathbb{P}} 0$ (since $c_n \rightarrow \infty$) we establish the desired claim. Note that,

$$c_n \mathbb{E}(U_n^*) = \sum_{j=5}^n j^4 \mathbb{E}(N_j) \left[\frac{\Gamma(j-3)}{\Gamma(j)} - \frac{1}{j^3} \right] - 4 \sum_{j=5}^n j^3 \mathbb{E}(N_j) \left[\frac{\Gamma(j-2)}{\Gamma(j)} - \frac{1}{j^2} \right] \\ + 6 \sum_{j=5}^n j^2 \mathbb{E}(N_j) \left[\frac{\Gamma(j-1)}{\Gamma(j)} - \frac{1}{j} \right] \\ = \sum_{j=5}^n \frac{2(j-9)}{(j-1)(j-2)(j-3)}.$$

Now $\limsup_{n \rightarrow \infty} |\sum_{j=5}^n \frac{2(j-9)}{(j-1)(j-2)(j-3)}| < \infty$ by standard ratio test. Consequently, using $c_n \rightarrow \infty$ one has $\mathbb{E}(U_n^*) \rightarrow 0$.

Subsequently it is enough to prove that $\text{Var}(U_n^*) \rightarrow 0$ which in turn is implied by proving that $\limsup_{n \rightarrow \infty} \text{Var}(c_n U_n^*) < \infty$. First note that

$$\begin{aligned} \text{Var}(\mathbb{E}(c_n U_n^* | N_1, \dots, N_n)) &= \sum_{j=5}^n \left(\frac{2j(j-9)}{(j-1)(j-2)(j-3)} \right)^2 \text{Var}(N_j) \\ &= \sum_{j=5}^n \left(\frac{2j(j-9)}{(j-1)(j-2)(j-3)} \right)^2 \frac{1}{j}, \end{aligned}$$

and consequently, $\limsup_{n \rightarrow \infty} \text{Var}(\mathbb{E}(c_n U_n^* | N_1, \dots, N_n)) < \infty$. Next note that

$$\begin{aligned} &\text{Var}(c_n U_n^* | N_1, \dots, N_n) \\ &= \sum_{j=5}^n N_j \left[j^8 \text{Var}\left(\frac{1}{S_{j1}^3}\right) + 16j^6 \text{Var}\left(\frac{1}{S_{j1}^2}\right) + 36j^4 \text{Var}\left(\frac{1}{S_{j1}}\right) + \text{Var}(S_{j1}) \right. \\ &\quad - 8j^7 \text{Cov}\left(\frac{1}{S_{j1}^3}, \frac{1}{S_{j1}^2}\right) + 12j^6 \text{Cov}\left(\frac{1}{S_{j1}^3}, \frac{1}{S_{j1}}\right) + 2j^4 \text{Cov}\left(\frac{1}{S_{j1}^3}, S_{j1}\right) \\ &\quad \left. - 48j^5 \text{Cov}\left(\frac{1}{S_{j1}^2}, \frac{1}{S_{j1}}\right) - 8j^3 \text{Cov}\left(\frac{1}{S_{j1}^2}, S_{j1}\right) + 12j^2 \text{Cov}\left(\frac{1}{S_{j1}}, S_{j1}\right) \right] \\ &= \sum_{j=5}^n N_j \frac{96j^7 + 4297j^6 - 14259j^5 - 28198j^4 + 102180j^3 - 33336j^2 - 4320j}{(j-1)^2(j-2)^2(j-3)^2(j-4)(j-5)(j-6)}. \end{aligned}$$

Consequently,

$$\begin{aligned} &\limsup_{n \rightarrow \infty} \mathbb{E}(\text{Var}(c_n U_n^* | N_1, \dots, N_n)) \\ &= \limsup_{n \rightarrow \infty} \sum_{j=5}^n \frac{96j^6 + 4297j^5 - 14259j^4 - 28198j^3 + 102180j^2 - 33336j - 4320}{(j-1)^2(j-2)^2(j-3)^2(j-4)(j-5)(j-6)} < \infty. \end{aligned}$$

This completes the proof of the lemma. $\square$

We finally note that as before Lemma 7.3 immediately implies that in the asymptotic distributional limit, U_n is independent of W_n, V_n and that $U_n - U_m \xrightarrow{\mathbb{P}} 0$ for any $n, m \rightarrow \infty$. Consequently, the analysis provided in proving Groeneboom and Pyke (1983, Theorem 3.1) goes through verbatim yielding that $U_n | \{W_n = 1, V_n = 0\} \xrightarrow{d} 0$ which in turn proves that

$$c_n^{-1}(nL_n^* - \gamma_n) | \{T_n = n, S_n = n\} \xrightarrow{d} 0$$

for our choice of $\gamma_n = 0$. Consequently, $\frac{\int_0^1 (\hat{f}_n(x) - 1)^4 dx}{\sqrt{\log n}} \xrightarrow{\mathbb{P}} 0$.

Acknowledgements

The authors would like to thank Jon Wellner for many insightful and helpful comments that led to major improvements in some of the results in this manuscript. The authors would also like to thank Piet Groeneboom for helpful conversations and for sharing his working manuscript.

References

- Beare, B. K. and Z. Fang (2017). Weak convergence of the least concave majorant of estimators for a concave distribution function. *Electron. J. Stat.* 11(2), 3841–3870. [MR3714300](#)
- Bickel, P. J., C. A. J. Klaassen, Y. Ritov, and J. A. Wellner (1993). *Efficient and adaptive estimation for semiparametric models*. Johns Hopkins Series in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD. [MR1245941](#)
- Bickel, P. J. and Y. Ritov (1988). Estimating integrated squared density derivatives: sharp best order of convergence estimates. *Sankhyā Ser. A* 50(3), 381–393. [MR1065550](#)
- Bickel, P. J. and Y. Ritov (2003). Nonparametric estimators which can be “plugged-in”. *Ann. Statist.* 31(4), 1033–1053. [MR2001641](#)
- Birgé, L. (1987). On the risk of histograms for estimating decreasing densities. *Ann. Statist.* 15(3), 1013–1022. [MR0902242](#)
- Birgé, L. and P. Massart (1995). Estimation of integral functionals of a density. *Ann. Statist.* 23(1), 11–29. [MR1331653](#)
- Cai, T. T. and M. G. Low (2003). A note on nonparametric estimation of linear functionals. *Ann. Statist.* 31(4), 1140–1153. [MR2001645](#)
- Cai, T. T. and M. G. Low (2004). Minimax estimation of linear functionals over nonconvex parameter spaces. *Ann. Statist.* 32(2), 552–576. [MR2060169](#)
- Cai, T. T. and M. G. Low (2005). Nonquadratic estimators of a quadratic functional. *Ann. Statist.* 33(6), 2930–2956. [MR2253108](#)
- Donoho, D. L., R. C. Liu, and B. MacGibbon (1990). Minimax risk over hyperrectangles, and implications. *Ann. Statist.* 18(3), 1416–1437. [MR1062717](#)
- Donoho, D. L. and M. Nussbaum (1990). Minimax quadratic estimation of a quadratic functional. *Journal of Complexity* 6(3), 290–323. [MR1081043](#)
- Dümbgen, L. and K. Rufibach (2009). Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency. *Bernoulli* 15(1), 40–68. [MR2546798](#)
- Fan, J. (1991). On the estimation of quadratic functionals. *Ann. Statist.* 19(3), 1273–1294. [MR1126325](#)
- Giné, E. and R. Nickl (2016). *Mathematical foundations of infinite-dimensional statistical models*. Cambridge Series in Statistical and Probabilistic Mathematics, [40]. Cambridge University Press, New York. [MR3588285](#)
- Grenander, U. (1956). On the theory of mortality measurement. II. *Skand. Aktuarietidskr.* 39, 125–153 (1957). [MR0093415](#)

- Groeneboom, P. (1983). The concave majorant of Brownian motion. *Ann. Probab.* 11(4), 1016–1027. [MR0714964](#)
- 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., Belmont, CA, pp. 539–555. Wadsworth. [MR0822052](#)
- Groeneboom, P. (2019). Grenander functionals and Cauchy’s formula. *arXiv preprint* [arXiv:1902.08806](#).
- Groeneboom, P. and G. Jongbloed (2014). *Nonparametric estimation under shape constraints*, Volume 38 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, New York. Estimators, algorithms and asymptotics. [MR3445293](#)
- Groeneboom, P., G. Jongbloed, and J. A. Wellner (2001). Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist.* 29(6), 1653–1698. [MR1891742](#)
- Groeneboom, P. and R. Pyke (1983). Asymptotic normality of statistics based on the convex minorants of empirical distribution functions. *Ann. Probab.* 11(2), 328–345. [MR0690131](#)
- Hall, P. and J. S. Marron (1987). Estimation of integrated squared density derivatives. *Statistics & Probability Letters* 6(2), 109–115. [MR0907270](#)
- Ibragimov, I. A. and R. Z. Khas’minskiĭ (1991). Asymptotically normal families of distributions and efficient estimation. *Ann. Statist.* 19(4), 1681–1724. [MR1135145](#)
- Jankowski, H. (2014). Convergence of linear functionals of the Grenander estimator under misspecification. *Ann. Statist.* 42(2), 625–653. [MR3210981](#)
- Kerkycharian, G. and D. Picard (1996). Estimating nonquadratic functionals of a density using Haar wavelets. *Ann. Statist.* 24(2), 485–507. [MR1394973](#)
- Kim, A. K. H., A. Guntuboyina, and R. J. Samworth (2018). Adaptation in log-concave density estimation. *Ann. Statist.* 46(5), 2279–2306. [MR3845018](#)
- Laurent, B. (1996). Efficient estimation of integral functionals of a density. *Ann. Statist.* 24(2), 659–681. [MR1394981](#)
- Laurent, B. (1997). Estimation of integral functionals of a density and its derivatives. *Bernoulli* 3(2), 181–211. [MR1466306](#)
- Levit, B. Y. (1978). Asymptotically efficient estimation of nonlinear functionals. *Problemy Peredachi Informatsii* 14(3), 65–72. [MR0533450](#)
- Mukherjee, R., W. K. Newey, and J. M. Robins (2017). Semiparametric efficient empirical higher order influence function estimators. *arXiv preprint* [arXiv:1705.07577](#).
- Mukherjee, R., E. T. Tchetgen, and J. Robins (2016). On adaptive estimation of nonparametric functionals. *arXiv preprint* [arXiv:1608.01364](#).
- Nemirovski, A. (2000). Topics in non-parametric statistics. In *Lectures on probability theory and statistics (Saint-Flour, 1998)*, Volume 1738 of Lecture Notes in Math., pp. 85–277. Springer, Berlin. [MR1775640](#)
- Nickl, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields* 138(3-4), 411–449. [MR2299714](#)

- Nickl, R. (2008). Uniform central limit theorems for the Grenander estimator. (unpublished manuscript).
- Pollard, D. (1984). *Convergence of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York. [MR0762984](#)
- Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā Ser. A* 31, 23–36. [MR0267677](#)
- Robertson, T., F. T. Wright, and R. L. Dykstra (1988). *Order restricted statistical inference*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. Chichester: John Wiley & Sons Ltd. [MR0961262](#)
- Robins, J., L. Li, E. Tchetgen, and A. van der Vaart (2008). Higher order influence functions and minimax estimation of nonlinear functionals. In *Probability and statistics: essays in honor of David A. Freedman*, Volume 2 of Inst. Math. Stat. (IMS) Collect., pp. 335–421. Inst. Math. Statist., Beachwood, OH. [MR2459958](#)
- Robins, J. M., L. Li, R. Mukherjee, E. T. Tchetgen, and A. van der Vaart (2017). Minimax estimation of a functional on a structured high-dimensional model. *Ann. Statist.* 45(5), 1951–1987. [MR3718158](#)
- Shapiro, A. (1991). Asymptotic analysis of stochastic programs. *Ann. Oper. Res.* 30(1-4), 169–186. Stochastic programming, Part I (Ann Arbor, MI, 1989). [MR1118896](#)
- Söhl, J. (2015). Uniform central limit theorems for the Grenander estimator. *Electron. J. Stat.* 9(1), 1404–1423. [MR3360732](#)
- Tchetgen, E., L. Li, J. Robins, and A. van der Vaart (2008). Minimax estimation of the integral of a power of a density. *Statist. Probab. Lett.* 78(18), 3307–3311. [MR2479495](#)
- van de Geer, S. A. (2000). *Applications of empirical process theory*, Volume 6 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. [MR1739079](#)
- van der Vaart, A. (2002). Semiparametric statistics. In *Lectures on probability theory and statistics (Saint-Flour, 1999)*, Volume 1781 of Lecture Notes in Math., pp. 331–457. Springer, Berlin. [MR1915446](#)
- van der Vaart, A. W. (1998). *Asymptotic statistics*, Volume 3 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge. [MR1652247](#)
- Woodroffe, M. and J. Sun (1993). A penalized maximum likelihood estimate of $f(0+)$ when f is nonincreasing. *Statist. Sinica* 3(2), 501–515. [MR1243398](#)