A central limit theorem for the Hellinger loss of Grenander-type estimators

Lopuhaä, Hendrik P.; Musta, Eni

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

ORIGINAL ARTICLE

WILEY

A central limit theorem for the Hellinger loss of Grenander-type estimators

Hendrik P. Lopuhaä | Eni Musta

Delft Institute of Applied Mathematics, Delft University of Technology, Delft, The Netherlands

Correspondence
Eni Musta, Delft Institute of Applied Mathematics, Delft University of Technology, 2628 XE Delft, The Netherlands.
Email: E.Musta@tudelft.nl

We consider Grenander-type estimators for a monotone function $\lambda : [0, 1] \to \mathbb{R}^+$, obtained as the slope of a concave (convex) estimate of the primitive of λ. Our main result is a central limit theorem for the Hellinger loss, which applies to estimation of a probability density, a regression function or a failure rate. In the case of density estimation, the limiting variance of the Hellinger loss turns out to be independent of λ.

KEYWORDS
central limit theorem, Grenander estimator, Hellinger distance, isotonic estimation

1 | INTRODUCTION

One of the problems in shape-constrained nonparametric statistics is to estimate a real-valued function under monotonicity constraints. Early references for this type of problem can be found in Grenander (1956), Brunk (1958), and Marshall and Proschan (1965), concerning the estimation of a probability density, a regression function, and a failure rate under monotonicity constraints. The asymptotic distribution of these types of estimators was first obtained by Prakasa Rao (1969, 1970) and reproved by Groeneboom (1985), who introduced a more accessible approach based on inverses. The latter approach initiated a stream of research on isotonic estimators, for example, see Groeneboom and Wellner (1992), Huang and Zhang (1994), Huang and Wellner (1995), and Lopuhaä and Nane (2013). Typically, the pointwise asymptotic behavior of isotonic estimators is characterized by a cube-root n rate of convergence and a nonnormal limit distribution.

The situation is different for global distances. In Groeneboom (1985), a central limit theorem was obtained for the L_1-error of the Grenander estimator of a monotone density (Groeneboom, Hooghiemstra, & Lopuhaä, 1999), and a similar result was established in Durot (2002) for the regression context. Extensions to general L_p-errors can be found in Kulikov and Lopuhaä (2005) and Durot (2007), where the latter provides a unified approach that applies to a variety of statistical

models. For the same general setup, an extremal limit theorem for the supremum distance was obtained in Durot, Kulikov, and Lopuhaä (2012).

Another widely used global measure of departure from the true parameter of interest is the Hellinger distance. It is a convenient metric in maximum likelihood problems, which goes back to the works of LeCam (1970, 1973), and it has nice connections with Bernstein norms and empirical process theory methods to obtain rates of convergence, due fundamentally to the works of Birgé and Massart (1993), Wong and Shen (1995), and others; see section 3.4 of van der Vaart and Wellner (1996) or chapter 4 in van de Geer (2000) for a more detailed overview. Consistency in Hellinger distance of shape-constrained maximum likelihood estimators was investigated by Pal, Woodroofe, and Meyer (2007), Seregin and Wellner (2010), and Doss and Wellner (2016), whereas rates on Hellinger risk measures were obtained in Seregin and Wellner (2010), Kim and Samworth (2016), and Kim, Guntuboyina, and Samworth (2016).

In contrast with L_p-distances or the supremum distance, there is no distribution theory available for the Hellinger loss of shape-constrained nonparametric estimators. In this paper, we present a first result in this direction, that is, a central limit theorem for the Hellinger loss of Grenander-type estimators for a monotone function λ. This type of isotonic estimator was also considered by Durot (2007) and is defined as the left-hand slope of a concave (or convex) estimate of the primitive of λ, based on n observations. We will establish our results under the same general setup of Durot (2007), which includes estimation of a probability density, a regression function, or a failure rate under monotonicity constraints. In fact, after approximating the squared Hellinger distance by a weighted L_2-distance, a central limit theorem can be obtained by mimicking the approach introduced in the work of Durot (2007). An interesting feature of our main result is that, in the monotone density model, the variance of the limiting normal distribution for the Hellinger distance does not depend on the underlying density. This phenomena was also encountered for the L_1-distance in Groeneboom (1985) and Groeneboom et al. (1999).

In Section 2, we define the setup and approximate the squared Hellinger loss by a weighted L_2-distance. A central limit theorem for the Hellinger distance is established in Section 3. We end this paper by a short discussion on the consequences for particular statistical models and a simulation study on testing exponentiality against a nonincreasing density.

2 | DEFINITIONS AND PREPARATORY RESULTS

Consider the problem of estimating a nonincreasing (or nondecreasing) function $\lambda : [0, 1] \to \mathbb{R}^+$ on the basis of n observations. Suppose that we have at hand a cadlag step estimator Λ_n for

$$\Lambda(t) = \int_0^t \lambda(u) \, \mathrm{d}u, \qquad t \in [0, 1].$$

If λ is nonincreasing, then the Grenander-type estimator $\hat{\lambda}_n$ for λ is defined as the left-hand slope of the least concave majorant of Λ_n, with $\hat{\lambda}_n(0) = \lim_{t \downarrow 0} \hat{\lambda}_n(t)$. If λ is nondecreasing, then the Grenander-type estimator $\hat{\lambda}_n$ for λ is defined as the left-hand slope of the greatest convex minorant of Λ_n, with $\hat{\lambda}_n(0) = \lim_{t \downarrow 0} \hat{\lambda}_n(t)$. We aim at proving the asymptotic normality of the Hellinger distance between $\hat{\lambda}_n$ and λ defined by

$$H(\hat{\lambda}_n, \lambda) = \left(\frac{1}{2} \int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)} \right)^2 \mathrm{d}t \right)^{1/2}. \tag{1}$$

We will consider the same general setup as in the work of Durot (2007), that is, we will assume the following conditions:

(A1) λ is monotone and differentiable on $[0, 1]$ with $0 < \inf_t |\lambda'(t)| \leq \sup_t |\lambda'(t)| < \infty$.

(A2') Let $M_n = \Lambda_n - \Lambda$. There exist $C > 0$ such that, for all $x > 0$ and $t = 0, 1$,

$$\mathbb{E}\left[\sup_{u \in [0,1], x/2 \leq |t-u| \leq x} (M_n(u) - M_n(t))^2\right] \leq \frac{Cx}{n}. \tag{2}$$

Durot (2007) also considered an additional condition (A2) in order to obtain bounds on pth moments; see theorem 1 and corollary 1 in Durot (2007). However, we only need condition (A2') for our purposes.

(A3) $\hat{\lambda}_n(0)$ and $\hat{\lambda}_n(1)$ are stochastically bounded.

(A4) Let B_n be either a Brownian bridge or a Brownian motion. There exists $q > 12$, $C_q > 0$, $L : [0, 1] \mapsto \mathbb{R}$, and versions of $M_n = \Lambda_n - \Lambda$ and B_n, such that

$$P\left(n^{1-1/q} \sup_{t \in [0,1]} \left|M_n(t) - n^{-1/2} B_n \circ L(t)\right| > x\right) \leq C_q x^{-q}$$

for $x \in (0, n]$. Moreover, L is increasing and twice differentiable on $[0, 1]$ with $\sup_t |L''(t)| < \infty$ and $\inf_t L'(t) > 0$.

In Durot (2007), a variety of statistical models are discussed for which the above assumptions are satisfied, such as estimation of a monotone probability density, a monotone regression function, and a monotone failure rate under right censoring. In Section 4, we briefly discuss the consequence of our main result for these models. We restrict ourselves to the case of a nonincreasing function λ. The case of nondecreasing λ can be treated similarly. Note that, even if this may not be a natural assumption, for example, in the regression setting, we need to assume that λ is positive for the Hellinger distance to be well defined.

The reason that one can expect a central limit theorem for the Hellinger distance is the fact that the squared Hellinger distance can be approximated by a weighted squared L_2-distance. This can be seen as follows:

$$\int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)}\right)^2 dt = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 \left(\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}\right)^{-2} dt$$

$$\approx \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1} dt. \tag{3}$$

Because L_2-distances for Grenander-type estimators obey a central limit theorem (e.g., Durot, 2007; Kulikov & Lopuhaä, 2005), similar behavior might be expected for the squared Hellinger distance. An application of the delta method will then do the rest.

The next lemma makes the approximation in (3) precise.

Lemma 1. *Assume (A1), (A2'), (A3), and (A4). Moreover, suppose that there are $C' > 0$ and $s > 3/4$ with*

$$\left|\lambda'(t) - \lambda'(x)\right| \leq C'|t-x|^s, \quad \text{for all } t, x \in [0, 1]. \tag{4}$$

If λ is strictly positive, we have that

$$\int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)}\right)^2 dt = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1} dt + o_p(n^{-5/6}).$$

In order to prove Lemma 1, we need the preparatory lemma below. To this end, we introduce the inverse of $\hat{\lambda}_n$, defined by

$$\hat{U}_n(a) = \underset{u \in [0,1]}{\operatorname{argmax}} \left\{ \Lambda_n^+(u) - au \right\}, \quad \text{for all } a \in \mathbb{R}, \tag{5}$$

where

$$\Lambda_n^+(t) = \max \left\{ \Lambda_n(t), \lim_{u \uparrow t} \Lambda_n(u) \right\}.$$

Note that

$$\hat{\lambda}_n(t) \geq a \Rightarrow \hat{U}_n(a) \geq t. \tag{6}$$

Furthermore, let g denote the inverse of λ. We then have the following result.

Lemma 2. *Under the conditions of Lemma 1, it holds*

$$\int_0^1 \left| \hat{\lambda}_n(t) - \lambda(t) \right|^3 dt = o_P \left(n^{-5/6} \right).$$

Proof. We follow the line of reasoning in the first step of the proof of theorem 2 in Durot (2007) with $p = 3$. For completeness, we briefly sketch the main steps. We will first show that

$$\int_0^1 \left| \hat{\lambda}_n(t) - \lambda(t) \right|^3 dt = \int_{\lambda(0)}^{\lambda(1)} \left| \hat{U}_n(b) - g(b) \right|^3 \lambda'(g(b))^2 db + o_P \left(n^{-5/6} \right).$$

To this end, consider

$$I_1 = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t) \right)_+^3 dt, \qquad I_2 = \int_0^1 \left(\lambda(t) - \hat{\lambda}_n(t) \right)_+^3 dt,$$

where $x_+ = \max\{x, 0\}$. We approximate I_1 by

$$J_1 = \int_0^1 \int_0^{(\lambda(0)-\lambda(t))^3} \mathbb{1}_{\left\{ \hat{\lambda}_n(t) \geq \lambda(t) + a^{1/3} \right\}} da\, dt.$$

From the reasoning on page 1,092 of Durot (2007), we deduce that

$$0 \leq I_1 - J_1 \leq \int_0^{n^{-1/3} \log n} \left(\hat{\lambda}_n(t) - \lambda(t) \right)_+^3 dt + \left| \hat{\lambda}_n(0) - \lambda(1) \right|^3 \mathbb{1}_{\left\{ n^{1/3} \hat{U}_n(\lambda(0)) > \log n \right\}}.$$

Because the $\hat{\lambda}_n(0)$ is stochastically bounded and $\lambda(1)$ is bounded, together with lemma 4 in Durot (2007), the second term is of the order $o_p(n^{-5/6})$. Furthermore, for the first term, we can choose $p' \in [1, 2)$ such that the first term on the right-hand side is bounded by

$$\left| \hat{\lambda}_n(0) - \lambda(1) \right|^{3-p'} \int_0^{n^{-1/3} \log n} \left| \hat{\lambda}_n(t) - \lambda(t) \right|^{p'} dt.$$

As in Durot (2007), we get

$$\mathbb{E} \left[\int_0^{n^{-1/3} \log n} \left| \hat{\lambda}_n(t) - \lambda(t) \right|^{p'} dt \right] \leq K n^{-(1+p')/3} \log n = o \left(n^{-5/6} \right),$$

by choosing $p' \in (3/2, 2)$. It follows that $I_1 = J_1 + o_P(n^{-5/6})$. By a change of variable $b = \lambda(t) + a^{1/3}$, we find

$$I_1 = \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} 3(b - \lambda(t))^2 \mathbb{1}_{\left\{ g(b) < \hat{U}_n(b) \right\}} dt\, db + o_p \left(n^{-5/6} \right).$$

Then, by a Taylor expansion, (A1), and (4), there exists a $K > 0$, such that

$$\left| (b - \lambda(t))^2 - \left\{ (g(b) - t) \lambda'(g(b)) \right\}^2 \right| \leq K(t - g(b))^{2+s}, \tag{7}$$

for all $b \in (\lambda(1), \lambda(0))$ and $t \in (g(b), 1]$. We find

$$I_1 = \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} 3(t - g(b))^2 \lambda'(g(b))^2 \mathbb{1}_{\{g(b) < \hat{U}_n(b)\}} \, dt \, db + R_n + o_p(n^{-5/6}), \tag{8}$$

where

$$|R_n| \leq \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} 3K(t - g(b))^{2+s} \mathbb{1}_{\{g(b) < \hat{U}_n(b)\}} \, dt \, db$$

$$\leq \frac{3K}{3+s} \int_{\lambda(1)}^{\lambda(0)} \left| \hat{U}_n(b) - g(b) \right|^{3+s} \, db = O_p \left(n^{-(3+s)/3} \right) = o_p \left(n^{-5/6} \right),$$

by using (23) in Durot (2007), that is, for every $q' < 3(q - 1)$, there exists $K_{q'} > 0$ such that

$$\mathbb{E}\left[\left(n^{1/3} \left| \hat{U}_n(a) - g(a) \right| \right)^{q'} \right] \leq K_{q'}, \quad \text{for all } a \in \mathbb{R}. \tag{9}$$

It follows that

$$I_1 = \int_{\lambda(1)}^{\lambda(0)} \left(\hat{U}_n(b) - g(b) \right)^3 \lambda'(g(b))^2 \mathbb{1}_{\{g(b) < \hat{U}_n(b)\}} \, db + o_p \left(n^{-5/6} \right).$$

In the same way, one finds

$$I_2 = \int_{\lambda(1)}^{\lambda(0)} \left(g(b) - \hat{U}_n(b) \right)^3 \lambda'(g(b))^2 \mathbb{1}_{\{g(b) > \hat{U}_n(b)\}} \, db + o_p \left(n^{-5/6} \right),$$

and it follows that

$$\int_0^1 \left| \hat{\lambda}_n(t) - \lambda(t) \right|^3 \, dt = I_1 + I_2 = \int_{\lambda(1)}^{\lambda(0)} \left| \hat{U}_n(b) - g(b) \right|^3 \lambda'(g(b))^2 \, db + o_p \left(n^{-5/6} \right).$$

Now, because λ' is bounded, by Markov's inequality, for each $\epsilon > 0$, we can write

$$\mathbb{P}\left(n^{5/6} \int_{\lambda(0)}^{\lambda(1)} \left| \hat{U}_n(b) - g(b) \right|^3 \lambda'(g(b))^2 \, db > \epsilon \right)$$

$$\leq \frac{1}{c \epsilon n^{1/6}} \int_{\lambda(0)}^{\lambda(1)} \mathbb{E}\left[n \left| \hat{U}_n(b) - g(b) \right|^3 \right] \, db \leq K n^{-1/6} \to 0.$$

For the last inequality, we again used (9) with $q' = 3$. It follows that

$$\int_{\lambda(0)}^{\lambda(1)} \left| \hat{U}_n(b) - g(b) \right|^3 \lambda'(g(b))^2 \, db = o_P \left(n^{-5/6} \right), \tag{10}$$

which finishes the proof. □

Proof of Lemma 1. Similar to (3), we write

$$\int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)} \right)^2 \, dt = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t) \right)^2 (4\lambda(t))^{-1} \, dt + R_n,$$

where

$$R_n = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t) \right)^2 \left\{ \left(\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)} \right)^{-2} - (4\lambda(t))^{-1} \right\} \, dt.$$

Write

$$4\lambda(t) - \left(\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}\right)^2 = \lambda(t) - \hat{\lambda}_n(t) - 2\sqrt{\lambda(t)}\left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)}\right)$$

$$= \left(\lambda(t) - \hat{\lambda}_n(t)\right)\left(1 + \frac{2\sqrt{\lambda(t)}}{\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}}\right).$$

Because $0 < \lambda(1) \le \lambda(t) \le \lambda(0) < \infty$, this implies that

$$|R_n| \le \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 \frac{\left|4\lambda(t) - \left(\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}\right)^2\right|}{4\lambda(t)\left(\sqrt{\hat{\lambda}_n(t)} + \sqrt{\lambda(t)}\right)^2} \, dt \le C \int_0^1 \left|\hat{\lambda}_n(t) - \lambda(t)\right|^3 dt$$

for some positive constant C only depending on $\lambda(0)$ and $\lambda(1)$. Then, from Lemma 2, it follows that $n^{5/6} R_n = o_P(1)$. □

3 | MAIN RESULT

In order to formulate the central limit theorem for the Hellinger distance, we introduce the process X, defined as

$$X(a) = \operatorname*{argmax}_{u \in \mathbb{R}} \left\{ W(u) - (u - a)^2 \right\}, \qquad a \in \mathbb{R}, \tag{11}$$

with W being a standard two-sided Brownian motion. This process was introduced and investigated by Groeneboom (1985, 1989) and plays a key role in the asymptotic behavior of isotonic estimators. The distribution of the random variable $X(0)$ is the pointwise limiting distribution of several isotonic estimators, and the constant

$$k_2 = \int_0^\infty \operatorname{cov}\left(|X(0)|^2, |X(a) - a|^2\right) \, da \tag{12}$$

appears in the limit variance of the L_p-error of isotonic estimators (e.g., Durot, 2002, 2007; Groeneboom, 1985; Groeneboom et al., 1999; Kulikov & Lopuhaä, 2005). We then have the following central limit theorem for the squared Hellinger loss.

Theorem 1. *Assume (A1), (A2'), (A3), (A4), and (4). Moreover, suppose that λ is strictly positive. Then, the following holds:*

$$n^{1/6} \left\{ n^{2/3} \int_0^1 \left(\sqrt{\hat{\lambda}_n(t)} - \sqrt{\lambda(t)}\right)^2 dt - \mu^2 \right\} \to N(0, \sigma^2),$$

where

$$\mu^2 = \mathbb{E}\left[|X(0)|^2\right] \int_0^1 \frac{|\lambda'(t)L'(t)|^{2/3}}{2^{2/3}\lambda(t)} \, dt, \qquad \sigma^2 = 2^{1/3} k_2 \int_0^1 \frac{|\lambda'(t)L'(t)|^{2/3}L'(t)}{\lambda(t)^2} \, dt,$$

where k_2 is defined in (12).

Proof. According to Lemma 1, it is sufficient to show that $n^{1/6}(n^{2/3}I_n - \mu^2) \to N(0, \sigma^2)$, with

$$I_n = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)^2 (4\lambda(t))^{-1} \, dt.$$

Again, we follow the same line of reasoning as in the proof of theorem 2 in Durot (2007). We briefly sketch the main steps of the proof. We first express I_n in terms of the inverse process $\hat{U}_n$, defined in (5). To this end, similar to the proof of Lemma 2, consider

$$\tilde{I}_1 = \int_0^1 \left(\hat{\lambda}_n(t) - \lambda(t)\right)_+^2 (4\lambda(t))^{-1} \, dt, \qquad \tilde{I}_2 = \int_0^1 \left(\lambda(t) - \hat{\lambda}_n(t)\right)_+^2 (4\lambda(t))^{-1} \, dt.$$

For the first integral, we can now write

$$\tilde{I}_1 = \int_0^1 \int_0^\infty \mathbb{1}_{\left\{\hat{\lambda}_n(t) \geq \lambda(t) + \sqrt{4a\lambda(t)}\right\}} \, da \, dt.$$

Then, if we introduce

$$\tilde{J}_1 = \int_0^1 \int_0^{(\lambda(0)-\lambda(t))^2/4\lambda(t)} \mathbb{1}_{\left\{\hat{\lambda}_n(t) \geq \lambda(t) + \sqrt{4a\lambda(t)}\right\}} \, da \, dt, \tag{13}$$

we obtain

$$0 \leq \tilde{I}_1 - \tilde{J}_1 \leq \int_0^{\hat{U}_n(\lambda(0))} \int_{(\lambda(0)-\lambda(t))^2/4a\lambda(t)}^\infty \mathbb{1}_{\left\{\hat{\lambda}_n(t) \geq \lambda(t) + \sqrt{4a\lambda(t)}\right\}} \, da \, dt$$

$$\leq \frac{1}{4\lambda(1)} \int_0^{\hat{U}_n(\lambda(0))} \left(\hat{\lambda}_n(t) - \lambda(t)\right)_+^2 \, dt.$$

Similar to the reasoning in the proof of Lemma 2, we conclude that $\tilde{I}_1 = \tilde{J}_1 + o_p(n^{-5/6})$. Next, the change of variable $b = \lambda(t) + \sqrt{4a\lambda(t)}$ yields

$$\tilde{J}_1 = \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} \frac{b - \lambda(t)}{2\lambda(t)} \mathbb{1}_{\{\hat{U}_n(b) > g(b)\}} \, dt \, db$$

$$= \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} \frac{b - \lambda(t)}{2b} \mathbb{1}_{\{\hat{U}_n(b) > g(b)\}} \, dt \, db \tag{14}$$

$$+ \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} \frac{(b - \lambda(t))^2}{2b\lambda(t)} \mathbb{1}_{\{\hat{U}_n(b) > g(b)\}} \, dt \, db.$$

Let us first consider the second integral on the right-hand side of (14). We then have

$$\int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} \frac{(b - \lambda(t))^2}{2b\lambda(t)} \mathbb{1}_{\{\hat{U}_n(b) > g(b)\}} \, dt \, db$$

$$\leq \frac{1}{2\lambda(1)^2} \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} (b - \lambda(t))^2 \mathbb{1}_{\{\hat{U}_n(b) > g(b)\}} \, dt \, db$$

$$\leq \frac{1}{2\lambda(1)^2} \sup_{x \in [0,1]} |\lambda'(x)| \int_{\lambda(1)}^{\lambda(0)} \mathbb{1}_{\{\hat{U}_n(b) > g(b)\}} \int_{g(b)}^{\hat{U}_n(b)} (t - g(b))^2 \, dt \, db$$

$$= \frac{1}{6\lambda(1)^2} \sup_{x \in [0,1]} |\lambda'(x)| \int_{\lambda(1)}^{\lambda(0)} \mathbb{1}_{\{\hat{U}_n(b) > g(b)\}} \left(\hat{U}_n(b) - g(b)\right)^3 \, db = o_P\left(n^{-5/6}\right),$$

again by using (9) with $q' = 3$. Then, consider the first integral on the right-hand side of (14). Similar to (7), there exists $K > 0$ such that

$$\left| (b - \lambda(t) - (g(b) - t)\lambda'(g(b))) \right| \leq K(t - g(b))^{1+s},$$

for all $b \in (\lambda(1), \lambda(0))$ and $t \in (g(b), 1]$. Taking into account that $\lambda'(g(b)) < 0$, similar to (8), it follows that

$$\tilde{I}_1 = \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} \frac{|\lambda'(g(b))|}{2b}(t - g(b))\mathbb{1}_{\{\hat{U}_n(b) > g(b)\}} \, dt \, db + \tilde{R}_n + o_p\left(n^{-5/6}\right),$$

where

$$|\tilde{R}_n| \leq \int_{\lambda(1)}^{\lambda(0)} \int_{g(b)}^{\hat{U}_n(b)} \frac{K}{2\lambda(1)}(t - g(b))^{1+s}\mathbb{1}_{\{g(b) < \hat{U}_n(b)\}} \, dt \, db$$

$$\leq \frac{K}{2\lambda(1)(2 + s)} \int_{\lambda(1)}^{\lambda(0)} \left| \hat{U}_n(b) - g(b) \right|^{2+s} db = O_p\left(n^{-(2+s)/3}\right) = o_p\left(n^{-5/6}\right),$$

by using (9) once more, and the fact that $s > 3/4$. It follows that

$$\tilde{I}_1 = \int_{\lambda(1)}^{\lambda(0)} \frac{|\lambda'(g(b))|}{4b}\left(\hat{U}_n(b) - g(b)\right)^2\mathbb{1}_{\{\hat{U}_n(b) > g(b)\}} \, db + o_p\left(n^{-5/6}\right).$$

In the same way,

$$\tilde{I}_2 = \int_{\lambda(1)}^{\lambda(0)} \frac{|\lambda'(g(b))|}{4b}\left(\hat{U}_n(b) - g(b)\right)^2\mathbb{1}_{\{\hat{U}_n(b) < g(b)\}} \, db + o_p\left(n^{-5/6}\right),$$

so that

$$I_n = \tilde{I}_1 + \tilde{I}_2 = \int_{\lambda(1)}^{\lambda(0)} \left(\hat{U}_n(b) - g(b)\right)^2 \frac{|\lambda'(g(b))|}{4b} \, db + o_P\left(n^{-5/6}\right).$$

We then mimic step 2 in the proof of theorem 2 in Durot (2007). Consider the representation

$$B_n(t) = W_n(t) - \xi_n t,$$

where W_n is a standard Brownian motion, $\xi_n = 0$ if B_n is a Brownian motion, and ξ_n is a standard normal random variable independent of B_n if B_n is a Brownian bridge. Then, define

$$\mathbb{W}_t(u) = n^{1/6}\left\{ W_n\left(L(t) + n^{-1/3}u\right) - W_n(L(t)) \right\}, \quad \text{for } t \in [0, 1],$$

which has the same distribution as a standard Brownian motion. Now, for $t \in [0, 1]$, let $d(t) = |\lambda'(t)|/(2L'(t)^2)$ and define

$$\tilde{V}(t) = \underset{|u| \leq \log n}{\mathrm{argmax}} \left\{ \mathbb{W}_t(u) - d(t)u^2 \right\}. \tag{15}$$

Then, similar to (26) in Durot (2007), we will obtain

$$n^{2/3}I_n = \int_0^1 \left| \tilde{V}(t) - n^{-1/6}\frac{\xi_n}{2d(t)} \right|^2 \left| \frac{\lambda'(t)}{L'(t)} \right|^2 \frac{1}{4\lambda(t)} \, dt + o_P\left(n^{-1/6}\right). \tag{16}$$

To prove (16), by using the approximation

$$\hat{U}_n(a) - g(a) \approx \frac{L(\hat{U}_n(a)) - L(g(a))}{L'(g(a))}$$

and a change of variable $a^\xi = a - n^{1/2}\xi_n L'(g(a))$, we first obtain

$$n^{2/3}I_n = n^{2/3} \int_{\lambda(1)+\delta_n}^{\lambda(0)-\delta_n} \left|L(\hat{U}_n(a^\xi)) - L(g(a^\xi))\right|^2 \frac{|\lambda'(g(a))|}{(L'(g(a)))^2} \frac{1}{4a} \, da + o_p\left(n^{-1/6}\right),$$

where $\delta_n = n^{-1/6}/\log n$. Apart from the factor $1/4a$, the integral on the right-hand side is the same as in the proof of theorem 2 in Durot (2007) for $p = 2$. This means that we can apply the same series of succeeding approximations for $L(\hat{U}_n(a^\xi)) - L(g(a^\xi))$ as in Durot (2007), which yields

$$n^{2/3}I_n = n^{2/3} \int_{\lambda(1)+\delta_n}^{\lambda(0)-\delta_n} \left|\tilde{V}(g(a)) - n^{-1/6}\frac{\xi_n}{2d(g(a))}\right|^2 \frac{|\lambda'(g(a))|}{(L'(g(a)))^2} \frac{1}{4a} \, da + o_p\left(n^{-1/6}\right).$$

Finally, because the integrals over $[\lambda(1), \lambda(1)+\delta_n]$ and $[\lambda(0)-\delta_n, \lambda(0)]$ are of the order $o_p(n^{-1/6})$, this yields (16) by a change of variables $t = g(a)$.

The next step is to show that the term with ξ_n can be removed from (16). This can be done exactly as in Durot (2007), because the only difference with the corresponding integral in Durot (2007) is the factor $1/4\lambda(t)$, which is bounded and does not influence the argument in Durot (2007). We find that

$$n^{2/3}I_n = \int_0^1 |\tilde{V}(t)|^2 \left|\frac{\lambda'(t)}{L'(t)}\right|^2 \frac{1}{4\lambda(t)} \, dt + o_P\left(n^{-1/6}\right).$$

Then, define

$$Y_n(t) = \left(|\tilde{V}(t)|^2 - \mathbb{E}\left[|\tilde{V}(t)|^2\right]\right) \left|\frac{\lambda'(t)}{L'(t)}\right|^2 \frac{1}{4\lambda(t)}. \tag{17}$$

By approximating $\tilde{V}(t)$ by

$$V(t) = \underset{u\in\mathbb{R}}{\text{argmax}} \left\{\mathbb{W}_t(u) - d(t)u^2\right\},$$

and using that, by Brownian scaling, $d(t)^{2/3}V(t)$ has the same distribution as $X(0)$ (see Durot, 2007, for details), we have that

$$\int_0^1 \mathbb{E}\left[|\tilde{V}(t)|^2\right] \left|\frac{\lambda'(t)}{L'(t)}\right|^2 \frac{1}{4\lambda(t)} \, dt = \mathbb{E}\left[|X(0)|^2\right] \int_0^1 d(t)^{-4/3} \left|\frac{\lambda'(t)}{L'(t)}\right|^2 \frac{1}{4\lambda(t)} \, dt + o\left(n^{-1/6}\right)$$
$$= \mu^2 + o\left(n^{-1/6}\right).$$

It follows that

$$n^{1/6}(I_n - \mu^2) = n^{1/6} \int_0^1 Y_n(t) \, dt + o_P(1).$$

We then first show that

$$\text{Var}\left(n^{1/6} \int_0^1 Y_n(t) \, dt\right) \to \sigma^2. \tag{18}$$

Once more, following the proof in Durot (2007), we have

$$v_n = \text{Var}\left(\int_0^1 Y_n(t) \, dt\right)$$
$$= 2 \int_0^1 \int_s^1 \left|\frac{\lambda'(t)}{L'(t)} \frac{\lambda'(s)}{L'(s)}\right|^2 \frac{1}{4\lambda(t)} \frac{1}{4\lambda(s)} \text{cov}\left(|\tilde{V}(t)|^2, |\tilde{V}(s)|^2\right) \, dt \, ds.$$

After the same sort of approximations as in Durot (2007), we get

$$v_n = 2 \int_0^1 \int_s^{\min(1, s+c_n)} \left| \frac{\lambda'(s)}{L'(s)} \right|^4 \frac{1}{(4\lambda(s))^2} \mathrm{cov}\left(|V_t(s)|^2, |V_s(s)|^2 \right) \mathrm{d}t \, \mathrm{d}s + o\left(n^{-1/3}\right),$$

where $c_n = 2n^{-1/3} \log n / \inf_t L'(t)$ and where, for all s and t,

$$V_t(s) = \underset{u \in \mathbb{R}}{\mathrm{argmax}} \left\{ \mathbb{W}_t(u) - d(s)u^2 \right\}.$$

Then, use that $d(s)^{2/3} V_t(s)$ has the same distribution as

$$X\left(n^{1/3} d(s)^{2/3} (L(t) - L(s)) \right) - n^{1/3} d(s)(L(t) - L(s))$$

so that the change of variable $a = n^{1/3} d(s)^{2/3}(L(t) - L(s))$ in v_n leads to

$$n^{1/3} v_n \to 2 \int_0^1 \int_0^\infty \left| \frac{\lambda'(s)}{L'(s)} \right|^4 \frac{1}{(4\lambda(s))^2} \frac{1}{d(s)^{10/3} L'(s)} \mathrm{cov}\left(|X(a)|^2, |X(0)|^2 \right) \mathrm{d}a \, \mathrm{d}s$$

$$\to 2k_2 \int_0^1 \left| \frac{\lambda'(s)}{L'(s)} \right|^4 \frac{1}{(4\lambda(s))^2} \frac{2^{10/3} |L'(s)|^{17/3}}{|\lambda'(s)|^{10/3}} \mathrm{d}s = \sigma^2,$$

which proves (18).

Finally, asymptotic normality of $n^{1/6} \int_0^1 Y_n(t) \mathrm{d}t$ follows by Bernstein's method of big blocks and small blocks in the same way as in step 6 of the proof of theorem 2 in Durot (2007). □

Corollary 1. *Assume* $(A1), (A2'), (A3), (A4)$, *and* (4) *and let* $H(\hat{\lambda}_n, \lambda)$ *be the Hellinger distance defined in (1). Moreover, suppose that* λ *is strictly positive. Then,*

$$n^{1/6} \left\{ n^{1/3} H(\hat{\lambda}_n, \lambda) - \tilde{\mu} \right\} \to N\left(0, \tilde{\sigma}^2\right).$$

$\tilde{\mu} = 2^{-1/2} \mu$ *and* $\tilde{\sigma}^2 = \sigma^2/8\mu^2$ *, where* μ^2 *and* σ^2 *are defined in Theorem 1.*

Proof. This follows immediately by applying the delta method with $\phi(x) = 2^{-1/2}\sqrt{x}$ to the result in Theorem 1. □

4 | EXAMPLES

The type of scaling for the Hellinger distance in Corollary 1 is similar to that in the central limit theorem for L_p-distances. This could be expected in view of the approximation in terms of a weighted squared L_2-distance (see Lemma 1), and the results, for example, in Kulikov and Lopuhaä (2005) and Durot (2007). Actually, this is not always the case. The phenomenon of observing different speeds of convergence for the Hellinger distance from those for the L_1 and L_2 norms was considered by Birgé (1986). In fact, this is related to the existence of a lower bound for the function we are estimating. If the function of interest is bounded from below, which is the case considered in this paper, then the approximation (3) holds; see Birgé (1986) for an explanation.

When we insert the expressions for μ^2 and σ^2 from Theorem 1, then we get

$$\tilde{\sigma}^2 = \frac{k_2}{4\mathbb{E}\left[|X(0)|^2\right]} \frac{\int_0^1 |\lambda'(t)L'(t)|^{2/3} L'(t)\lambda(t)^{-2} \mathrm{d}t}{\int_0^1 |\lambda'(t)L'(t)|^{2/3} \lambda(t)^{-1} \mathrm{d}t},$$

where k_2 is defined in (12). This means that, in statistical models where $L = \Lambda$ in condition (A4) and, hence, $L' = \lambda$, the limiting variance $\tilde{\sigma}^2 = k_2/(4\mathbb{E}[|X(0)|^2])$ does not depend on λ.

One such a model is estimation of the common monotone density λ on $[0, 1]$ of independent random variables $X_1, \dots, X_n$. Then, Λ_n is the empirical distribution function of $X_1, \dots, X_n$, and $\hat{\lambda}_n$ is Grenander's estimator (Grenander, 1956). In that case, if $\inf_t \lambda(t) > 0$, the conditions of Corollary 1 are satisfied with $L = \Lambda$ (see theorem 6 in Durot, 2007), so that the limiting variance of the Hellinger loss for the Grenander estimator does not depend on the underlying density. This behavior was conjectured in Wellner (2015) and coincides with that of the limiting variance in the central limit theorem for the L_1-error for the Grenander estimator, first discovered by Groeneboom (1985); see also Durot (2002, 2007), Groeneboom et al. (1999), and Kulikov and Lopuhaä (2005).

Another example is when we observe independent identically distributed inhomogeneous Poisson processes $N_1, \dots, N_n$ with common mean function Λ on $[0, 1]$ with derivative λ, for which $\Lambda(1) < \infty$. Then, Λ_n is the restriction of $(N_1 + \dots + N_n)/n$ to $[0, 1]$. Also in that case, the conditions of Corollary 1 are satisfied with $L = \Lambda$ (see theorem 4 in Durot, 2007), so that the limiting variance of the Hellinger loss for $\hat{\lambda}_n$ does not depend on the common underlying intensity λ. However, note that, for this model, the L_1-loss for $\hat{\lambda}_n$ is asymptotically normal according to theorem 2 in Durot (2007) but with limiting variance depending on the value $\Lambda(1) - \Lambda(0)$.

Consider the monotone regression model $y_{i,n} = \lambda(i/n) + \epsilon_{i,n}$, for $i = 1, \dots, n$, where the $\epsilon_{i,n}$'s are i.i.d. random variables with mean zero and variance $\sigma^2 > 0$. Let

$$\Lambda_n(t) = \frac{1}{n} \sum_{i \leq nt} y_{i,n}, \quad t \in [0, 1]$$

be the empirical distribution function. Then, $\hat{\lambda}_n$ is (a slight modification of) Brunk's (1958) estimator. Under appropriate moment conditions on the $\epsilon_{i,n}$, the conditions of Corollary 1 are satisfied with $L(t) = t\sigma^2$ (see theorem 5 in Durot, 2007). In this case, the limiting variance of the Hellinger loss for $\hat{\lambda}_n$ depends on both λ and σ^2, whereas the L_1-loss for $\hat{\lambda}_n$ is asymptotically normal according to theorem 2 in Durot (2007) but with limiting variance only depending on σ^2.

Suppose we observe a right-censored sample $(X_1, \Delta_1), \dots, (X_n, \Delta_n)$, where $X_i = \min(T_i, Y_i)$ and $\Delta_i = \mathbb{1}_{\{T_i \leq Y_i\}}$, with the T_i's being nonnegative i.i.d. failure times and the Y_i's are i.i.d. censoring times independent of the T_i's. Let F be the distribution function of the T_i's with density f and let G be the distribution function of the Y_i's. The parameter of interest is the monotone failure rate $\lambda = f/(1 - F)$ on $[0, 1]$. In this case, Λ_n is the restriction of the Nelson–Aalen estimator to $[0, 1]$. If we assume (A1) and $\inf_t \lambda(t) > 0$, then, under suitable assumptions on F and G, the conditions of Corollary 1 hold with

$$L(t) = \int_0^t \frac{\lambda(u)}{(1 - F(u))(1 - G(u))} \, \mathrm{d}u, \quad t \in [0, 1];$$

see theorem 3 in Durot (2007). This means that the limiting variance of the Hellinger loss depends on λ, F, and G, whereas the limiting variance of the L_1-loss depends only on their values at 0 and 1. In particular, in the case of nonrandom censoring times, $L = (1 - F)^{-1} - 1$, the limiting variance of the Hellinger loss depends on λ and F, whereas the limiting variance of the L_1-loss depends only on the value $F(1)$.

5 | TESTING EXPONENTIALITY AGAINST A NONDECREASING DENSITY

In this section, we investigate a possible application of Theorem 1, that is, testing for an exponential density against a nonincreasing alternative by means of the Hellinger loss. The exponential distribution is one of the most used and well-known distributions. It plays a very important role in reliability, survival analysis, and renewal process theory, when modeling random times until some event. As a result, a lot of attention has been given in the literature to testing for exponentiality against a wide variety of alternatives, by making use of different properties and characterizations of the exponential distribution (Alizadeh Noughabi & Arghami, 2011; Haywood & Khmaladze, 2008; Jammalamadaka & Taufer, 2003; Meintanis, 2007). In this section, we consider a test for exponentiality, assuming that data come from a decreasing density. The test is based on the Hellinger distance between the parametric estimator of the exponential density and the Grenander-type estimator of a general decreasing density. In order to be able to apply the result of Corollary 1, we first investigate a test whether the data is exponentially distributed with a fixed parameter $\lambda_0 > 0$. Because such a test may not be very interesting from a practical point of view, we also investigate testing exponentiality, leaving the parameter $\lambda > 0$ unspecified.

5.1 | Testing a simple null hypothesis of exponentiality

Let $f_\lambda(x) = \lambda e^{-\lambda x} \mathbb{1}_{\{x \geq 0\}}$ be the exponential density with parameter $\lambda > 0$. Assume we have a sample of i.i.d. observations $X_1, \ldots, X_n$ from some distribution with density f and for $\lambda_0 > 0$ fixed, we want to test

$$H_0 : f = f_{\lambda_0} \quad \text{against} \quad H_1 : f \text{ is nonincreasing.}$$

Under the alternative hypothesis, we can estimate f on an interval $[0, \tau]$ by the Grenander-type estimator $\hat{f}_n$ from Section 2. Then, as a test statistic, we take $T_n = H(\hat{f}_n, f_{\lambda_0})$, the Hellinger distance on $[0, \tau]$ between $\hat{f}_n$ and f_{λ_0}, and at level α, we reject the null hypothesis if $T_n > c_{n,\alpha,\lambda_0}$, for some critical value $c_{n,\alpha,\lambda_0} > 0$.

According to Corollary 1, it follows that T_n is asymptotically normally distributed, but the mean and the variance depend on the constant k_2 defined in (12). To avoid computation of k_2, we estimate the mean and the variance of T_n empirically. We generate $B = 10,000$ samples from f_{λ_0}. For each of these samples, we compute the Grenander estimator $\hat{f}_{n,i}$ and the Hellinger distance $T_{n,i} = H(\hat{f}_{n,i}, f_{\lambda_0})$, for $i = 1, 2, \ldots, B$. Finally, we compute the mean $\bar{T}$ and the variance s_T of the values $T_{n,1}, \ldots, T_{n,B}$. For the critical value of the test, we take $c_{n,\alpha,\lambda_0} = \bar{T} + q_{1-\alpha} s_T$, where $q_{1-\alpha}$ is the $100(1 - \alpha)\%$ quantile of the standard normal distribution. Note that, even if in the density model the asymptotic variance is independent of the underlying distribution, the asymptotic mean does depend on λ_0, that is, the test is not distribution free. Another possibility, instead of the normal approximation, is to take as a critical value $\tilde{c}_{n,\alpha,\lambda_0}$ the empirical $100(1 - \alpha)\%$ quantile of the values $T_{n,1}, \ldots, T_{n,B}$.

To investigate the performance of the test, we generate $N = 10,000$ samples from f_{λ_0}. For each sample, we compute the value of the test statistic $T_n = H(\hat{f}_n, f_{\lambda_0})$ and we reject the null hypothesis if $T_n > c_{n,\alpha,\lambda_0}$ (or if $T_n > \tilde{c}_{n,\alpha,\lambda_0}$). The percentage of rejections gives an approximation of the level of the test. Table 1 shows the results of the simulations for different sample sizes n and two values of λ_0 and $\alpha = 0.01, 0.05, 0.10$. Here, we take $\tau = 5$ because the mass of the exponential distribution with parameter one or five outside the interval $[0, 5]$ is negligible. We observe that the percentage of rejections is close to the nominal level if we use $\tilde{c}_{n,\alpha,\lambda_0}$ as a critical value for the

TABLE 1 Simulated levels of T_n using (top) c_{n,α,λ_0} and (bottom) $\tilde{c}_{n,\alpha,\lambda_0}$, with $\alpha = 0.01, 0.05, 0.10$, under the null hypothesis varying the sample size n and the parameter λ_0

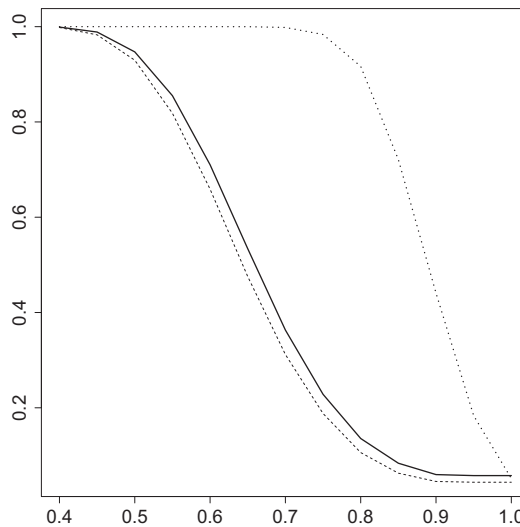|	$\lambda_0 = 1$			$\lambda_0 = 5$		
	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
$n = 20$	0.0229	0.0680	0.1016	0.0310	0.0791	0.1127
	0.0118	0.0498	0.0971	0.0117	0.0533	0.1058
$n = 50$	0.0244	0.0684	0.1123	0.0243	0.0659	0.1086
	0.0106	0.0469	0.0923	0.0103	0.0494	0.0964
$n = 100$	0.0190	0.0589	0.1021	0.0236	0.0673	0.1126
	0.0106	0.0531	0.1063	0.0091	0.0453	0.0951

FIGURE 1 Simulated powers using (solid) c_{n,α,λ_0} and (dashed) $\tilde{c}_{n,\alpha,\lambda_0}$, with $\alpha = 0.05$, of T_n and the power of (dotted) the likelihood ratio test for $\lambda = 1$, $\nu = 0.4, 0.45, \ldots, 1$, and $n = 100$

test, but it is a bit higher if we use c_{n,α,λ_0}. This is due to the fact that, for small sample sizes, the normal approximation of Corollary 1 is not very precise.

Moreover, to investigate the power, we generate a sample from the Weibull distribution with shape parameter ν and scale parameter λ_0^{-1}. Recall that Weibull$(1, \lambda_0^{-1})$ corresponds to the exponential distribution with parameter λ_0 and that a Weibull distribution with $\nu < 1$ has a decreasing density. We compute the Hellinger distance $T_n = H(\hat{f}_n, f_{\lambda_0})$ and we reject the null hypothesis if $T_n > c_{n,\alpha,\lambda_0}$ (or if $T_n > \tilde{c}_{n,\alpha,\lambda_0}$). After repeating the procedure $N = 10,000$ times, we compute the percentage of times that we reject the null hypothesis, which gives an approximation of the power of the test.

The results of the simulations, done with $n = 100$, $\lambda_0 = 1$, $\alpha = 0.05$, and alternatives for which ν varies between 0.4 and 1 by steps of 0.05, are shown in Figure 1. As a benchmark, we compute the power of the likelihood ratio (LR) test statistic for each ν. As expected, our test is less powerful with respect to the LR test, which is designed to test against a particular alternative. However, as the sample size increases, the performance improves significantly and the difference of the results when using c_{n,α,λ_0} or $\tilde{c}_{n,\alpha,\lambda_0}$ becomes smaller.

TABLE 2 Simulated levels of R_n under the null hypothesis varying the sample size n and the parameter λ

λ	0.05	0.1	0.5	1	2	3	4	5
$n = 50$	0.051	0.052	0.049	0.049	0.05	0.053	0.051	0.054
$n = 100$	0.049	0.047	0.050	0.052	0.054	0.047	0.049	0.050
$n = 500$	0.052	0.049	0.049	0.049	0.053	0.052	0.053	0.048
$n = 1,000$	0.053	0.046	0.049	0.051	0.049	0.048	0.048	0.052

5.2 | Testing a composite null hypothesis of exponentiality

Assume we have a sample of i.i.d. observations $X_1, \ldots, X_n$ from some distribution with density f and we want to test

$$H_0 : f = f_\lambda, \text{ for some } \lambda > 0 \quad \text{against} \quad H_1 : f \text{ is nonincreasing.}$$

Under the null hypothesis, we can construct a parametric estimator of the density that is given by $f_{\hat{\lambda}_n}$, where $\hat{\lambda}_n = n/\sum_{i=1}^n X_i$ is the maximum likelihood estimator of λ. On the other hand, under the alternative hypothesis, we can estimate f on an interval $[0, \tau]$ by the Grenander-type estimator $\hat{f}_n$ from Section 2. Then, as a test statistic, we take $R_n = H(\hat{f}_n, f_{\hat{\lambda}_n})$, the Hellinger distance on $[0, \tau]$ between the two estimators, and at level α, we reject the null hypothesis if $R_n > d_{n,\alpha}$ for some critical value $d_{n,\alpha} > 0$. Because the limit distribution of the test statistic is not known, we use a bootstrap procedure to calibrate the test. We generate $B = 1,000$ bootstrap samples of size n from $f_{\hat{\lambda}_n}$, and for each of them, we compute the estimators $f_{\hat{\lambda}_{n,i}^*}$, $\hat{f}_{n,i}^*$ and the test statistic $R_{n,i}^* = H(f_{\hat{\lambda}_{n,i}^*}, \hat{f}_{n,i}^*)$, for $i = 1, 2, \ldots, B$. Then, we determine the 100αth upper percentile $d_{n,\alpha}^*$ of the values $R_{n,1}^*, \ldots, R_{n,B}^*$. Finally, we reject the null hypothesis if $R_n > d_{n,\alpha}^*$.

To investigate the level of the test, for $\alpha = 0.05$ and $\lambda > 0$ fixed, we start with a sample from an exponential distribution with parameter λ and repeat the above procedure $N = 10,000$ times. We count the number of times we reject the null hypothesis, that is, the number of times the value of the test statistics exceeds the corresponding 5th upper percentile. Dividing this number by N gives an approximation of the level. Table 2 shows the results of the simulations for different sample sizes n and different values of λ. The rejection probabilities are close to 0.05 for all the values of λ, which shows that the test performs well in the different scenarios (slightly and strongly decreasing densities).

To investigate the power, for $\alpha = 0.05$ and fixed $0 < \nu < 1$ and $\lambda > 0$, we now start with a sample from a Weibull distribution with shape parameter ν and scale parameter λ^{-1} and compute the value $R_n = H(f_{\hat{\lambda}_n}, \hat{f}_n)$. In order to calibrate the test, we treat this sample as if it were an exponential sample and estimate λ by $\hat{\lambda}_n = n/\sum_{i=1}^n X_i$. Next, we generate $B = 1,000$ bootstrap samples of size n from the exponential density with parameter $\hat{\lambda}_n$. For each bootstrap sample, we compute the test statistic $R_{n,i}^* = H(f_{\hat{\lambda}_{n,i}^*}, \hat{f}_{n,i}^*)$, for $i = 1, 2, \ldots, B$, and we determine the 5th upper percentile $d_{n,0.05}^*$ of the values $R_{n,1}^*, \ldots, R_{n,B}^*$. Finally, we reject the null hypothesis if $R_n > d_{n,0.05}^*$. After repeating the above procedure $N = 10,000$ times, each time starting with a Weibull sample, we compute the percentage of times that we reject the null hypothesis, which gives an approximation of the power of the test.

We compare the Hellinger distance test to some of the tests from Alizadeh Noughabi and Arghami (2011), which are designed to test exponentiality against all the possible alternatives, that is, not only against decreasing densities. These tests are all distribution free, which means that their critical values can be computed independently of λ. Then, for each of the Weibull samples generated before, we count the percentage of times that the tests
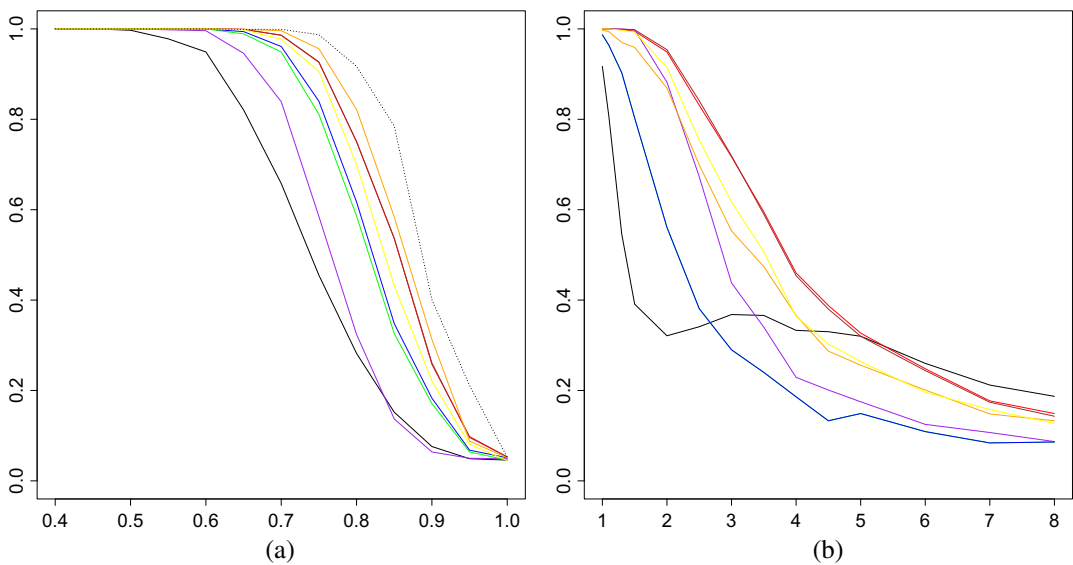
FIGURE 2 Simulated powers of (black solid) the Hellinger distance test and some other competitor tests, that is, (blue) T_1, (green) T_2, (yellow) ω_n^2, (brown) S_n, (red) EP_n, (purple) KL_{mn}, and (orange) CO_n, and the power of (black dotted) the likelihood ratio test for (left) $n = 100$, $\lambda = 1$, $0.4 \leq \nu \leq 1$ and (right) $1 \leq \beta \leq 8$. (a) Weibull. (b) Beta

$T_1, T_2, \omega_n^2, S_n, EP_n, KL_{mn}$, and CO_n (see Alizadeh Noughabi & Arghami, 2011, for a precise definition) reject the null hypothesis. Finally, we also compare the power of our test with the LR test for each ν.

The results of the simulations, done with $n = 100$, $\lambda = 1$, and alternatives for which ν varies between 0.4 and 1, are shown in the left panel in Figure 2. Actually, we also investigated the power for different choices of λ, and we observed similar behavior as for $\lambda = 1$. The figure shows that the test based on the Hellinger distance performs worse than the other tests. In this case, the test of Cox and Oakes CO_n has greater power. However, Alizadeh Noughabi and Arghami (2011) concluded that none of the tests is uniformly most powerful with respect to the others.

We repeated the experiment taking, instead of the Weibull distribution, the beta distribution with parameters $\alpha = 1$ and $1 \leq \beta \leq 8$ as alternative. Note that it has a nonincreasing density on $[0, 1]$ proportional to $(1 - x)^{\beta - 1}$, and the extreme case $\beta = 1$ corresponds to the uniform distribution. Results are shown in the right panel in Figure 2. We observe that, for small values of β, the Hellinger distance test again behaves worse than the others, and in this case, R_n and EP_n have greater power. However, for larger β, the Hellinger distance test outperforms all the others.

ORCID

Eni Musta http://orcid.org/0000-0003-3356-4307

REFERENCES

Alizadeh Noughabi, H., & Arghami, N. R. (2011). Testing exponentiality based on characterizations of the exponential distribution. *Journal of Statistical Computation and Simulation*, *81*(11), 1641–1651.

Birgé, L. (1986). On estimating a density using Hellinger distance and some other strange facts. *Probability Theory and Related Fields*, *71*(2), 271–291.

Birgé, L., & Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probability Theory and Related Fields*, *97*(1–2), 113–150.

Brunk, H. D. (1958). On the estimation of parameters restricted by inequalities. *Annals of Mathematical Statistics*, *29*, 437–454.

Doss, C. R., & Wellner, J. A. (2016). Global rates of convergence of the MLEs of logconcave and s-concave densities. *The Annals of Statistics*, *44*(3), 954–981.

Durot, C. (2002). Sharp asymptotics for isotonic regression. *Probability Theory and Related Fields*, *122*(2), 222–240.

Durot, C. (2007). On the $\mathbb{L}_p$-error of monotonicity constrained estimators. *The Annals of Statistics*, *35*(3), 1080–1104.

Durot, C., Kulikov, V. N., & Lopuhaä, H. P. (2012). The limit distribution of the L_∞-error of Grenander-type estimators. *The Annals of Statistics*, *40*(3), 1578–1608.

van de Geer, S. A. (2000). Applications of empirical process theory. Vol. 6. *Cambridge Series in Statistical and Probabilistic Mathematics* (pp. xii–286). Cambridge, UK: Cambridge University Press.

Grenander, U. (1956). On the theory of mortality measurement II. *Skand Aktuarietidskr*, *39*, 125–153.

Groeneboom, P. (1985). Estimating a monotone density. In *Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II*. *The Wadsworth & Brooks/Cole Statistics/Probability Series* (pp. 539–555).

Groeneboom, P. (1989). Brownian motion with a parabolic drift and airy functions. *Probability Theory Related Fields*, *81*(1), 79–109.

Groeneboom, P., Hooghiemstra, G., & Lopuhaä, H. P. (1999). Asymptotic normality of the L_1 error of the Grenander estimator. *The Annals of Statistics*, *27*(4), 1316–1347.

Groeneboom, P., & Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*. (Vol. 19). Basel, Switzerland: Birkhäuser Verlag. *DMV Seminar*.

Haywood, J., & Khmaladze, E. (2008). On distribution-free goodness-of-fit testing of exponentiality. *Journal of Economics*, *143*(1), 5–18.

Huang, J., & Wellner, J. A. (1995). Estimation of a monotone density or monotone hazard under random censoring. *Scandinavian Journal of Statistics Theory and Applications*, *22*(1), 3–33.

Huang, Y., & Zhang, C. H. (1994). Estimating a monotone density from censored observations. *The Annals of Statistics*, *22*(3), 1256–1274.

Jammalamadaka, S. R., & Taufer, E. (2003). Testing exponentiality by comparing the empirical distribution function of the normalized spacings with that of the original data. *Journal of Nonparametric Statistics*, *15*(6), 719–729.

Kim, A. K. H., Guntuboyina, A., & Samworth, R. J. (2016). Adaptation in log-concave density estimation. arXiv:1609.00861.

Kim, A. K. H., & Samworth, R. J. (2016). Global rates of convergence in log-concave density estimation. *The Annals of Statistics*, *44*(6), 2756–2779.

Kulikov, V. N., & Lopuhaä, H. P. (2005). Asymptotic normality of the L_k-error of the Grenander estimator. *The Annals of Statistics*, *33*(5), 2228–2255.

LeCam, L. (1970). On the weak convergence of probability measures. *Annals of Mathematical Statistics*, *41*, 621–625.

LeCam, L. (1973). Convergence of estimates under dimensionality restrictions. *The Annals of Statistics*, *1*, 38–53.

Lopuhaä, H. P., & Nane, G. F. (2013). Shape constrained non-parametric estimators of the baseline distribution in Cox proportional hazards model. *Scandinavian Journal of Statistics Theory and Applications*, *40*(3), 619–646.

Marshall, A. W., & Proschan, F. (1965). Maximum likelihood estimation for distributions with monotone failure rate. *Annals of Mathematical Statistics*, *36*, 69–77.

Meintanis, S. G. (2007). Test for exponentiality against Weibull and gamma decreasing hazard rate alternatives. *Kybernetika (Prague)*, *43*(3), 307–314.

Pal, J. K., Woodroofe, M., & Meyer, M. (2007). Estimating a Polya frequency function$_2$. Complex datasets and inverse problems. *IMS Lecture Notes Monograph Series*, *54*, 239–249.

Prakasa Rao, B. L. S. (1969). Estimation of a unimodal density. *Sankhyā (Statistics) The Indian Journal of Statistics Series A*, *31*, 23–36.

Prakasa Rao, B. L. S. (1970). Estimation for distributions with monotone failure rate. *Annals of Mathematical Statistics*, *41*, 507–519.

Seregin, A., & Wellner, J. A. (2010). Nonparametric estimation of multivariate convextransformed densities. *The Annals of Statistics*, *38*(6), 3751–3781. With supplementary material available online.

van der Vaart, A. W., & Wellner, J. A. (1996). *Weak convergence and empirical processes: With applications to statistics. Springer Series in Statistics*, pp. xvi–508. New York, NY: Springer-Verlag.

Wellner, J. A. (2015). Musings about shape constrained estimation and inference: Some problems. Paper presented at the workshop of Shape Constrained Inference: Open Problems and New Directions, Lorentz Centre, Leiden, The Netherlands.

Wong, W. H., & Shen, X. (1995). Probability inequalities for likelihood ratios and convergence rates of sieve MLEs. *The Annals of Statistics*, *23*(2), 339–362.