END-POINT ESTIMATION FOR DECREASING DENSITIES: ASYMPTOTIC BEHAVIOR OF THE PENALIZED LIKELIHOOD RATIO

Jayanta Kumar Pal

Technical Report #2007-1
February 2, 2007

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

Submitted to the Annals of Statistics

END-POINT ESTIMATION FOR DECREASING DENSITIES: ASYMPTOTIC BEHAVIOR OF THE PENALIZED LIKELIHOOD RATIO

By JAYANTA KUMAR PAL*

University of Michigan and SAMSI

We consider the problem of estimating the mode of a decreasing density on the positive real line. This has application in several interesting phenomena arising from examples in renewal theory, biased and distance samplings. We use a penalized likelihood ratio based approach and derive the scale-free universal large sample distribution of the log-likelihood ratio under the null, using a suitably chosen penalty parameter, to use it for inference. We present simulation results to corroborate our findings, and compare the performance of the confidence sets with existing results.

1. Introduction. Density estimation has gained importance in statistics and became useful as regression curve estimation in several problems. In the presence of a monotonicity constraint, the non-parametric maximum likelihood estimator of the density can be estimated, as shown by Grenander, [2]. This does not get affected by the subjectivity of a bandwidth , as in the kernel density estimation. However, the Grenander estimator is non-regular in nature and the usual likelihood properties do not apply. In particular, its behavior near the end-points is not desirable, as observed by Woodroofe and Sun [11]. However, in several problems, the parameter of interest is the

*Research supported by NSF-SAMSI cooperative agreement No. DMS-0112069

AMS 2000 subject classifications: Primary 62G07,62G20; secondary 62F25

Keywords and phrases: monotone density estimation, penalized likelihood ratio, brownian motion

end-point value itself, as manifested in the following examples.

Let f denote a left-continuous density for which $f(x) = 0$ for $x \leq 0$ and is non-increasing in $x > 0$. We consider the problem of estimating $f(0+) = \lim_{x \searrow 0} f(x)$, the mode. To start with, let X and U be independent positive random variables. Assuming X has density g and U is uniform $[0, 1]$, the density of $Y = XU$ is given by $f(y) = \int_y^\infty g(x)/x \, dx$, decreasing in y. Here, $f(0+) = E(1/X)$ is the parameter of interest, based on the observations $Y_1, \ldots, Y_n$. This has been mentioned by Vardi, in [9].

Next, let G and μ be the distribution function and mean of the inter-arrival times of a renewal process. The observations are the backward recurrence times X, the time elapsed since the last occurrence of the renewal. In the equilibrium, X has density $f(x) = (1 - G(x))/\mu$, decreasing in x. We seek to estimate the mode $f(0+) = 1/\mu$. as an application, consider the case of estimating *natural fecundity of human populations*. We suppose a fertility clinic observes the check up duration for a cross-section of couples who are attempting to become pregnant. The goal is to estimate the mean waiting time for pregnancy. We assume that the couples terminate their check-up once they succeed in their attempt, or they discontinue the follow-up for whatever reason. This problem can be formulated as the same renewal theory setting, with observations being the experience waiting time for different couples. See Keiding et al, [4] for a discussion. This has another application in the problem of estimating the mean breakdown time of a machine under investigation, referred as *Inspection and failure problem* by [11].

The last two examples are from different sampling scenarios, where the chances of including a particular observation in the sample depends on its value. First, we revisit an astronomical model discussed by Lynden-Bell, [6], where the chances of observing a galaxy increases with its normalized angular diameter Y. Let f be the density of $X = 1/Y^3$. Then, f is decreasing, and a natural parameter of interest, the proportion of galaxies included in the observed sample, equals $1/f(0+)$. This has been referred as a *biased sampling* problem by [11] and further discussed by Woodroofe, [10]. Similarly, the distance sampling, as discussed by Woodroofe & Zhang, [12] also relates to our problem, where the chances of sampling depends on the length or distance of an object. We consider an observer searching for hidden objects (e.g. bird nests in the bushes) and records the distance X from the object to his line of approach. If $g(x)$ is the probability of finding an object x distance away, then it is logical to assume that g is decreasing and $g(0) = 1$. The conditional density function of X given that it is detected, is $f(x) = g(x)/\mu$, where $\mu = \int g(x)dx$ is the unconditional probability of finding an object. Clearly, f is decreasing, and $\mu = 1/f(0+)$ is our parameter of interest.

As discussed in [11], the Grenander estimator is inconsistent at the end-point. This is referred to as the *Spiking Problem* in the literature. To consistently estimate $f(0+)$, [11] used a penalization of the likelihood. Recently, Kulikov and Lopuhaa, in [5], investigated the behavior of the Grenander estimator near 0, with the rate $n^{-1/3}$, so that it does not suffer from the inflation near zero. It produces another consistent estimator of $f(0+)$, with the rate of convergence being similar to that of Grenander estimator at interior points.

In case of regression, Banerjee and Wellner in [1], developed asymptotic likelihood ratio inference for the functional value at an interior point. However, in case of the end-point estimation, because of the inconsistency mentioned earlier, we focus on the penalized likelihood. We derive the constrained maximizer of the penalized log-likelihood under the null hypothesis. Since we derive the distribution of the likelihood ratio under an optimal penalization, we can numerically estimate the quantiles from this distribution and use them to construct confidence intervals for $c = f(0+)$.

Section 2 characterizes the different penalized estimators, both without and in the presence of a null hypothesis specifying the value of the mode. In Section 3, we derive the large sample distribution of the penalized likelihood ratio. The proof of the main theorem has been presented in this Section and the main concepts explained, with the proofs of auxiliary results being deferred to the Appendix. In Section 4 we compare the performance of different type of confidence intervals using simulations from familiar decreasing densities, namely exponential. Finally, the technical derivations are presented in Section 7, the Appendix.

2. Characterization of the estimators. In this Section, we characterize the different estimators of f in terms of the Grenander estimator $\tilde{f}$, the MLE. Let $X_1, \ldots, X_n$ be a random sample from f, and $0 = x_0 < x_1 \ldots < x_n < \infty$ its order statistics. Then, $\tilde{f}$ is a left continuous step function, expressed as,

$$\tilde{f}(x) = \begin{cases} \min_{0 \leq r < k} \max_{k \leq s \leq n} \frac{s-r}{n(x_s - x_r)}, & x_{k-1} < x \leq x_k,\ 1 \leq k \leq n \\ 0, & x > x_n \end{cases}$$

However, $\tilde{f}(0+) = \tilde{f}(x_1)$ is not consistent for $f(0+)$. It systematically overestimates and in fact, $\tilde{f}(0+)/f(0+) \Rightarrow \sup_{k \geq 1}(k/\Gamma_k)$ where $\Gamma_1, \Gamma_2, \ldots$ are partial sums of i.i.d. exponential(1) random variables, and $\Rightarrow$ denotes convergence in law. Observe that the right side is greater than 1 w.p. 1. Instead of the log-likelihood $\sum_{i=1}^n \log f(x_i)$, we consider the penalized likelihood

$$l_\alpha(f) = \sum_{i=1}^n \log f(x_i) - n\alpha f(0+)$$

on the set $\mathcal{F}$, the collection of all left-continuous decreasing densities with support $(0, \infty)$, for some $\alpha > 0$. We formally characterize the estimator $\hat{f}$

using the following Theorem. This is analogous to theorem 1 in [11].

THEOREM 2.1. *The penalized MLE $\hat{f}$ can be characterized as the unpenalized MLE of a deformed data set $\alpha + \hat{\gamma} x_k, k = 1, \ldots, n$ where $\hat{\gamma}$ is a solution of the equation $\gamma = 1 - \alpha \max_{1 \le s \le n} s/n(\alpha + \gamma x_s)$, and*

$$\hat{f}(x) = \begin{cases} \frac{1-\hat{\gamma}}{\alpha} = \max_{1 \le s \le n} \frac{s}{n(\alpha + \hat{\gamma} x_s)}, & \text{if } x \le x_{m_0} \\ \frac{1}{\hat{\gamma}} \tilde{f}(x), & \text{otherwise.} \end{cases}$$

where $m_0 = argmax_s [s/n(\alpha + \hat{\gamma} x_s)]$.

This estimator is consistent at 0, i.e. $\hat{f}(0+) \xrightarrow{p} f(0+)$, when $0 < f(0+) < \infty$, as $\alpha = \alpha_n \to 0$ but $n\alpha_n \to \infty$ as $n \to \infty$, Under this condition, $h(f, \hat{f}) \xrightarrow{p} 0$ as well, where h denotes the Hellinger metric. We will investigate its properties later.

To construct a likelihood ratio based statistic, our target is to test the hypothesis $H_0^c : f(0+) = c$ for some value of $c > 0$ against the alternative $H_1 : f(0+) \neq c$, under the standard assumption of $f \in \mathcal{F}$. Define, $\mathcal{F}_c = \mathcal{F} \cap H_0^c$. The α-penalized log-likelihood ratio statistic is

$$\begin{aligned} lr_\alpha(c) &= \max_{f \in \mathcal{F}} [\sum_{i=1}^n \log f(x_i) - n\alpha f(0+)] - \max_{f \in \mathcal{F}_c} [\sum_{i=1}^n \log f(x_i) - n\alpha f(0+)] \\ &= \sum_{i=1}^n \log \hat{f}(x_i) - n\alpha(\hat{f}(0+) - c) - \max_{f \in \mathcal{F}_c} [\sum_{i=1}^n \log f(x_i)]. \end{aligned}$$

We need to compute $\max_{f \in \mathcal{F}_c} [\sum_{i=1}^n \log f(x_i)]$. It is again easy to observe that the maximizer is a step function. The following Theorem provides the necessary characterization for the constrained penalized MLE $\hat{f}^c$. The proof of this theorem is fairly involved and we discuss them in the Appendix 7.1, in an order that may enable the reader to understand the technical steps without much difficulty.

THEOREM 2.2. *The constrained MLE can be explicitly written as :*

$$\hat{f}^c(x) = \begin{cases} c & \text{if } x \leq x_{s_0} \\ \frac{1}{\hat{\lambda}}\tilde{f}(x) & \text{otherwise.} \end{cases}$$

where $s_0 = argmin_{cx_s<1}(1-s/n)/(1-cx_s)$ *and* $\hat{\lambda} = (1-s_0/n)/(1-cx_{s_0})$.

Since both $\hat{f}$ and $\hat{f}^c$ are related to $\tilde{f}$, we need to look at the behavior of the Grenander estimator in open intervals near zero. Also, both m_0 (defined in Theorem 2.1) and s_0 (defined in Theorem 2.2) are original breakpoints of $\tilde{f}$. This will easily follow once we prove the theorems, and will play a crucial role in the proofs. Finally, both $\hat{\gamma}$ and $\hat{\lambda}$ converge to 1 in probability, meaning both the constrained and unconstrained estimators are consistent once we stay sufficiently away from 0.

3. The distribution of the likelihood ratio. The penalized log-likelihood ratio, or $lr_\alpha(c)$, as defined in Section 2 measures the effect of constraining the value of $f(0+) = c$. So, a large value of $lr_\alpha(c)$ will indicate a strong evidence in favor of the alternative hypothesis. Our main target centers upon the large sample distribution of $lr_\alpha(c)$ under H_0^c, so that we can use that to build confidence intervals for $f(0+)$, as we derive in this Section. From now on, the true underlying density $f \in \mathcal{F}_c$. We also assume that $f'(0+) < 0$ and f'' is bounded uniformly. Define $a = -f'(0+)/2 > 0$. To characterize the limit distribution (we denote it by $\mathbb{J}$), we need to define the following collection of random processes as functions of the Brownian motion $\{\mathbb{B}(t)\}$ on $t \geq 0$. Define,

$$\mathbb{G}(t) = \mathbb{B}(t) + t^2 \qquad \text{and} \qquad \mathbb{G}_1(t) = (\mathbb{G}(t) + 1)/t,$$

and let $\mathbb{Z}(t)$ be the right derivative process of the Least Concave Majorant (LCM) of the process $-\mathbb{G}(t)$. Further, we define

$$\mathbb{X} = \text{argmin}\mathbb{G}, \qquad \mathbb{U} = \inf \mathbb{G},$$
$$\mathbb{X}_1 = \text{argmin}\mathbb{G}_1, \qquad \mathbb{U}_1 = \inf \mathbb{G}_1.$$

Then, our main theorem that characterizes the limit behavior of $lr_\alpha(c)$ can be stated as follows.

THEOREM 3.1. *Let $\alpha = (ac)^{-1/3}n^{-2/3}$ and $f \in \mathcal{F}_c$. Then, under H_0^c,*

$$2lr_\alpha(c) \Rightarrow \mathbb{X}_1\mathbb{U}_1^2 + \int_{\mathbb{X}_1}^{\mathbb{X}} \mathbb{Z}^2(u)du \equiv \mathbb{J}$$

where the integral is interpreted with a negative sign if $\mathbb{X}_1 > \mathbb{X}$.

Outline of the proof. We start with a proposition that is central in the sense that it characterizes the dominant terms in an expansion of $lr_\alpha(c)$. The proof is deferred to Appendix 7.3.

PROPOSITION 3.1. *Under the assumptions of Theorem 3.1,*

$$\sum_{i=1}^n \log\Big(\frac{\hat{f}(x_i)}{\hat{f}^c(x_i)}\Big) = \begin{cases} n(1-\hat{\gamma})\frac{\hat{f}(0+)-c}{c} + \frac{nx_{m_0}}{2c}(\hat{f}(0+)-c)^2 & \\ \qquad + \frac{n}{2c}\int_{x_{m_0}}^{x_{s_0}}(\tilde{f}(x)-c)^2dx + o_P(1), & x_{m_0} < x_{s_0} \\ n(1-\hat{\lambda})\frac{\hat{f}(0+)-c}{c} - \frac{nx_{s_0}}{2c}(\hat{f}(0+)-c)^2 & \\ \qquad - \frac{n}{2c}\int_{x_{s_0}}^{x_{m_0}}(\tilde{f}(x)-\hat{f}(0+))^2dx + o_P(1), & x_{s_0} < x_{m_0}. \end{cases}$$

Once we have identified the key terms, we only need the joint distribution of them to derive the limit distribution. Define the process,

$$\mathbb{Z}_n(t) = (n/ac)^{1/3}[\tilde{f}((c/na^2)^{1/3}t) - c], t \in [0, \infty)$$

The next Lemma is comparable to Theorem 3.1, part (ii) of [5].

Lemma 3.1. *The process $\mathbb{Z}_n(t)$ converges in distribution, in the uniform topology on compacta, to the process $\mathbb{Z}(t)$.*

Proof : We sketch a streamlined proof using *Hungarian embedding*. This representation is also useful in deriving the limit of other terms. See Page 268 of [8] for a reference. We formally state the result as follows.

Hungarian Representation : There exists a probability space carrying i.i.d. random variables $X_1', \ldots, X_n'$ with law F and a sequence of Wiener Processes $\{\mathbb{B}_n\}$ and corresponding Brownian Bridges $\{\mathbb{G}_n\}$ (associated by $\mathbb{G}_n(t) = \mathbb{B}_n(t) - t\mathbb{B}_n(1)$) such that, for $F_n'(t) = \sum_{i=1}^n 1_{X_i' \le t}/n$, we have,

$$\limsup(\sqrt{n}/(\log n)^2)\|\sqrt{n}(F_n' - F) + \mathbb{G}_n \circ F\|_\infty < \infty \text{ a.s.}$$

To avoid notational complication, since we look at distributional convergence only, we will continue with F_n, as they have the same distribution as F_n'. Moreover, we have, $F(x) = \int_0^x f(t)dt = \int_0^x [f(0) + f'(0)t + f''(t^*)\frac{t^2}{2}]dt$ (for some $0 < t^* < t$) $= cx - ax^2 + o(x^2)$. Now, since $\tilde{f}$ is the right hand slope of the LCM of F_n, and adding a linear function to F_n modifies its LCM by the same linear term, we deduce that $\mathbb{Z}_n(t)$ is the right hand slope of the LCM of the process

$$\mathbb{V}_n(t) = (n^2 a/c^2)^{1/3}[F_n((c/na^2)^{1/3}t) - c(c/na^2)^{1/3}t].$$

Using the Hungarian Embedding and a modulus of continuity by Lévy, we can show that, $\mathbb{V}_n$ is approximately equal to $[-t^2 - \mathbb{B}_n^1(t)]$, where $\mathbb{B}_n^1(t) = (na^2/c^4)^{1/6}\mathbb{B}_n((c^4/na^2)^{1/3}t)$ is also a standard Wiener Process. The result follows by observing the facts that $[-t^2 - \mathbb{B}_n^1(t)]$ has the same distribution as $-\mathbb{G}$, and taking the right hand slope of the LCM is a continuous function on the set of all continuous processes. □

Remark : This approximation will be made more rigorous in the proofs given in the Appendix.

The next proposition and the immediate Corollary characterize the limit distribution of the different terms in the expression of $lr_\alpha(c)$ in terms of the same Brownian motion $\mathbb{B}$, which is necessary for the proof of Theorem 3.1. In the Proposition, the topology for convergence is the product of Euclidean in all the co-ordinates except the last and the uniform topology on compacta for the process $\mathbb{Z}(t)$. This is the main distributional convergence result, and we defer the proof to Appendix 7.3.

PROPOSITION 3.2. *Under the assumptions of Theorem 3.1, and under the topology mentioned above, as $n \to \infty$,*

$$\begin{pmatrix} n^{2/3}(1-\hat{\gamma}) \\ n^{2/3}(1-\hat{\lambda}) \\ n^{1/3}(\hat{f}(0+)-c) \\ n^{1/3}x_{m_0} \\ n^{1/3}x_{s_0} \\ \{\mathbb{Z}_n(t)\}_{t\geq 0} \end{pmatrix} \Rightarrow \begin{pmatrix} (c^2/a)^{1/3} \\ -(c^2/a)^{1/3}\mathbb{U} \\ -(ca)^{1/3}\mathbb{U}_1 \\ (c/a^2)^{1/3}\mathbb{X}_1 \\ (c/a^2)^{1/3}\mathbb{X} \\ \{\mathbb{Z}(t)\}_{t\geq 0} \end{pmatrix}.$$

COROLLARY 3.1. *Under the assumptions of Theorem 3.1, as $n \to \infty$*

$$\begin{pmatrix} 1_{x_{m_0}<x_{s_0}} \\ n\alpha(\hat{f}(0+)-c) \\ n(1-\hat{\gamma})(\hat{f}(0+)-c/c) \\ n(1-\hat{\lambda})(\hat{f}(0+)-c/c) \\ nx_{m_0}(\hat{f}(0+)-c)^2/c \\ nx_{s_0}(\hat{f}(0+)-c)^2/c \\ (n/c)\int_{x_{m_0}}^{x_{s_0}}(\tilde{f}(x)-c)^2dx \\ (n/c)\int_{x_{s_0}}^{x_{m_0}}(\tilde{f}(x)-\hat{f}(0+))^2dx \end{pmatrix} \Rightarrow \begin{pmatrix} 1_{\mathbb{X}_1<\mathbb{X}} \\ -\mathbb{U}_1 \\ -\mathbb{U}_1 \\ \mathbb{U}\mathbb{U}_1 \\ \mathbb{X}_1\mathbb{U}_1^2 \\ \mathbb{X}\mathbb{U}_1^2 \\ \int_{\mathbb{X}_1}^{\mathbb{X}}\mathbb{Z}^2(u)du \\ \int_{\mathbb{X}}^{\mathbb{X}_1}(\mathbb{Z}(u)+\mathbb{U}_1)^2du \end{pmatrix}.$$

Proof : The only non-trivial part is to show the convergence of the integrals. Using a change of variable $u = (na^2/c)^{1/3}x$ we obtain,

$$\frac{n}{c}\int_{x_{m_0}}^{x_{s_0}} (\tilde{f}(x) - c)^2 dx = \int_{(na^2/c)^{1/3}x_{m_0}}^{(na^2/c)^{1/3}x_{s_0}} \mathbb{Z}_n^2(u)du \Rightarrow \int_{\mathbb{X}_1}^{\mathbb{X}} \mathbb{Z}^2(u)du, \text{ and}$$

$$\begin{aligned} & \frac{n}{c}\int_{x_{s_0}}^{x_{m_0}} (\tilde{f}(x) - \hat{f}(0+))^2 dx \\ = & \int_{(na^2/c)^{1/3}x_{s_0}}^{(na^2/c)^{1/3}x_{m_0}} \left(\mathbb{Z}_n(u) - (n/ac)^{1/3}(\hat{f}(0+) - c)\right)^2 du \\ \Rightarrow & \int_{\mathbb{X}}^{\mathbb{X}_1} (\mathbb{Z}(u) + \mathbb{U}_1)^2 du. \end{aligned}$$

□

Proof of Theorem 3.1. If we use the approximation of Proposition 3.1,

$$lr_\alpha(c) = \begin{cases} -n\alpha(\hat{f}(0+) - c) + n(1-\hat{\gamma})\frac{\hat{f}(0+)-c}{c} + \frac{nx_{m_0}}{2c}(\hat{f}(0+) - c)^2 \\ \quad + \frac{n}{2c}\int_{x_{m_0}}^{x_{s_0}} (\tilde{f}(x) - c)^2 dx + o_P(1), \qquad x_{m_0} < x_{s_0} \\ -n\alpha(\hat{f}(0+) - c) + n(1-\hat{\lambda})\frac{\hat{f}(0+)-c}{c} - \frac{nx_{s_0}}{2c}(\hat{f}(0+) - c)^2 \\ \quad - \frac{n}{2c}\int_{x_{s_0}}^{x_{m_0}} (\tilde{f}(x) - \hat{f}(0+))^2 dx + o_P(1), \qquad x_{s_0} < x_{m_0}. \end{cases}$$

Therefore, using Corollary 3.1, we deduce,

$$lr_\alpha(c) \Rightarrow \begin{cases} \mathbb{U}_1 - \mathbb{U}_1 + \frac{1}{2}\mathbb{X}_1\mathbb{U}_1^2 + \frac{1}{2}\int_{\mathbb{X}_1}^{\mathbb{X}} \mathbb{Z}^2(u)du, & \text{if } \mathbb{X}_1 < \mathbb{X} \\ \mathbb{U}_1 + \mathbb{U}\mathbb{U}_1 - \frac{1}{2}\mathbb{X}\mathbb{U}_1^2 - \frac{1}{2}\int_{\mathbb{X}}^{\mathbb{X}_1} (\mathbb{Z}(u) + \mathbb{U}_1)^2 du, & \text{if } \mathbb{X} < \mathbb{X}_1. \end{cases}$$

In the second case, the limit equals

$$\mathbb{U}_1 + \mathbb{U}\mathbb{U}_1 - \frac{1}{2}\mathbb{X}_1\mathbb{U}_1^2 - \frac{1}{2}\int_{\mathbb{X}}^{\mathbb{X}_1} \mathbb{Z}^2(u)du - \mathbb{U}_1\int_{\mathbb{X}}^{\mathbb{X}_1} \mathbb{Z}(u)du$$

Since both x_{s_0} and x_{m_0} are jump-points of $\tilde{f}$, the cumulative sum diagram of the data and its LCM agrees at both $\mathbb{X}$ and $\mathbb{X}_1$. Consequently, using the process convergence of $\tilde{f}$, $-\mathbb{G}$ and its LCM also agrees at both $\mathbb{X}$ and $\mathbb{X}_1$. Therefore,

$$\int_{\mathbb{X}}^{\mathbb{X}_1} \mathbb{Z}(u)du = \mathbb{G}(\mathbb{X}) - \mathbb{G}(\mathbb{X}_1).$$

We also observe that,

$$\mathbb{G}(\mathbb{X}_1) + 1 = \mathbb{X}_1\mathbb{U}_1.$$

Combining this two observations, the limit equals

$$\frac{1}{2}\mathbb{X}_1\mathbb{U}_1^2 - \frac{1}{2}\int_{\mathbb{X}}^{\mathbb{X}_1} \mathbb{Z}^2(u)du.$$

The theorem follows. □

3.1. *An adaptive choice of optimal penalization.* Since the penalization depends on the unknown parameter ac, we need a consistent estimator of ac to compute the likelihood ratio. The following lemma yields such an estimator.

LEMMA 3.2. *Let* $0 < \epsilon < 1/3$. *Define* $\hat{f}^*$ *as the penalized maximum likelihood estimator of* f *with penalization parameter* $\alpha^* = n^{-2/3}$, *with* $\hat{\gamma}^*$ *and* $x^*_{m_0}$ *defined accordingly. Also, define* $u = n^{\epsilon}x^*_{m_0}$. *Then,*

$$\hat{f}^*(0+)(\hat{f}^*(0+) - \hat{\gamma}^*\hat{f}^*(u))/2u \xrightarrow{p} ac. \tag{3.1}$$

Proof: We recall that, $\hat{f}^*(0+) - c = O_P(n^{-1/3})$. Now, $u > x^*_{m_0}$ implies that $\hat{\gamma}^*\hat{f}^*(u) = \tilde{f}(u)$. Since u is of the order $n^{\epsilon-1/3}$,

$$\begin{aligned}
(\hat{f}^*(0+) - c)/2u &= O_P(n^{-\epsilon}),\\
(c - f(u))/2u &= (1/2[-f'(0+) + O_P(u)] = a + o_P(1), \text{ and}\\
(\tilde{f}(u) - f(u))/2u &= (\tilde{f}(n^{\epsilon}x^*_{m_0}) - f(n^{\epsilon}x^*_{m_0}))/(2n^{\epsilon}x^*_{m_0}) = \frac{O_P(n^{-1/3})}{n^{\epsilon}x^*_{m_0}} = O_P(n^{-\epsilon})
\end{aligned}$$

using Theorem 3.1(iii) from [5]. Combining these relations yields,

$$(\hat{f}^*(0+) - \hat{\gamma}^*\hat{f}^*(u))/2u \xrightarrow{p} a.$$

The lemma follows. □

The lemma yields an adaptive choice of the penalization parameter α.

4. Simulations and Data Analysis. In this Section, we start with a brief description of the random variable $\mathbb{J} = \mathbb{X}_1\mathbb{U}_1^2 + \int_{\mathbb{X}_1}^{\mathbb{X}} \mathbb{Z}^2 du$. This random variable is positive with probability 1, as evident from its limit characterization of a sequence of positive random variables. Its quantiles are numerically computed and can be used to construct confidence intervals for $f(0+)$. Table 1 shows a few relevant quantiles of the distribution of $\mathbb{J}$.

TABLE 1
Some of the relevant quantiles of the limit distribution $\mathbb{J}$*.*

p	.5	.75	.90	.95	.99
$\mathbb{J}^{-1}(p)$	0.88	1.94	3.03	3.71	4.93

Figure 1 plots the distribution function of $\mathbb{J}$. Clearly, the limit distribution has support $[0, \infty)$, but most of its mass is concentrated in $[0, 6]$. However, computing the actual quantiles and CDF of it is theoretically very complicated, and numerical estimates are used. We simulated Brownian motions based on a grid of size .00001, and then computed the corresponding functionals to simulate from the proposed distribution.

FIG. 1. *The distribution function of the limit random variable*

4.1. *Simulation from a decreasing density.* Since we have simulated the limit distribution and obtained its quantiles, we can proceed to compare it with the theoretical results. For that, we have simulated from exponential distribution, which is a decreasing density on the positive real line. To compare the performances of the penalized likelihood ratio based estimator with the Woodroofe-Sun penalized MLE($\hat{f}$) and the Kulikov-Lopuhaa estimator (KL), we focus on the asymptotic coverage and the relative length of the three confidence intervals for sample sizes varying from $n = 100, 200, \ldots, 1000$. Since we have already discussed $\hat{f}$ in detail, here we briefly discuss the other estimator before comparison.

The estimator is based on the Grenander estimator $\tilde{f}$. However, because of the inconsistency near zero, [5] consider $\tilde{f}(c_0 n^{-1/3})$, for some optimal value of c_0 also depending on the parameters a and c. They show that $\hat{f}^A(0) = \tilde{f}(c_0 n^{-1/3})$ is a consistent estimator for $f(0)$ and its asymptotic distribution can be derived as the value of the function $\mathbb{Z}$ described in Theorem 3.1, at a point that can be estimated from computer simulations. To compare the relative performances of three methods in the same level, we use the true values of $c = 1$ and $a = .5$ in the simulation.

We simulated 1000 replications each of sample size $n = 100, \ldots, 1000$ from a standard exponential distribution with unit rate. The tables 2 and 3 compare the relative coverage and expected interval length, when the nominal coverage rate should be .95. It seems that the penalized likelihood ratio based estimator has systematically shorter confidence intervals than the estimator $\hat{f}$ and near nominal coverage, whereas the estimator $\hat{f}^A$ yields shorter intervals but its coverage is compromised. However, for small sample the likelihood ratio based interval performs less admirably, since this asymptotics is of the slower type, and takes a substantial sample size to manifest.

Table 2
The relative comparison of the 95% coverage confidence intervals for the three methods.

n	100	200	300	400	500	600	700	800	900	1000
$\hat{f}(0+)$	.926	.931	.940	.937	.939	.941	.951	.948	.942	.943
$\hat{f}^A$	.902	.941	.936	.932	.926	.932	.935	.942	.945	.946
lr_α	.915	.932	.942	.946	.952	.956	.943	.946	.951	.954

Table 3
The relative comparison of the 95% confidence interval lengths for the three methods.

n	100	200	300	400	500	600	700	800	900	1000
$\hat{f}(0+)$	.5197	.4393	.3973	.3469	.2885	.2759	.2479	.2156	.2073	.2017
$\hat{f}^A$	.3774	.2946	.2576	.2182	.1556	.1406	.1230	.0821	.0801	.1057
lr_α	.4300	.3829	.3342	.2911	.2629	.2307	.2179	.1840	.1938	.1984

In all three methods, we use the true values of the parameters, instead of the adaptive estimators. Using the plug-in estimators, the relative behavior remains the same. However, all the three estimators have slightly less coverage for small sample sizes. The interval lengths remain almost the same, since they are given by the same quantiles of the limit distribution, with consistent estimators replacing the true value of the parameter.

5. Conclusion. The problem of estimating a decreasing density occurs in some interesting problems, including but not limited to the examples discussed in Section 1. Since the MLE, the Grenander estimator does not give a nice estimator, different methods have been used to adapt it to yield consistency. Penalization has been used to get rid of the problem. As seen in [5], other methods using the Grenander estimator itself can be used as well. However, the likelihood principle tells us that the difference in the constrained and unconstrained likelihoods can be used as a method for testing the validity of a hypothesis, and therefore can be inverted to construct confidence sets for the parameter of interest.

The limiting distribution, namely $\mathbb{J}$ is similar to the corresponding distribution obtained by Banerjee and Wellner, in [1], in case of the estimation of a regression function at an interior point. However, in their case, no parameters need to be estimated, except the covariate density. In the end-point estimation, however, the estimation of a and c, needs to be done to construct any confidence interval, since the limit distribution always depends on them in a non-linear way.

6. Acknowledgments. These results are derived as a part of my doctoral dissertation work with my advisor Dr. Michael Woodroofe. He helped with substantial feedback, general comments and useful suggestions. Special thanks to Dr. Moulinath Banerjee, for many fruitful discussions.

7. Appendix.

7.1. *Proof of Theorem 2.2.* Depending on different values of c, we have different estimators.

Case 1 : $c \leq 1/x_n$. The maximizer in this case is, $\hat{f}^c(x) = c$ for $0 < x \leq 1/c$.

Case 2 : $c = \tilde{f}(0+)$. The maximizer in this case is, $\hat{f}^c(x) = \tilde{f}(x)$ for all x.

Case 3 : $c > \tilde{f}(0+)$. The maximizer does not exist. In fact, the Grenander estimator $\tilde{f}$ is as close one can get to the MLE in this case.

Case 4 : $1/x_n < c < \tilde{f}(0+)$. This is the most common case, and we deal it separately. The following lemma ensures that, to find the behavior of $lr_\alpha(c)$ under H_0^c, we can ignore the first three scenarios, and assume that $1/x_n < c < \tilde{f}$, to find out the maximizer here.

LEMMA 7.1. *Suppose that $f \in \mathcal{F}_c$ is not the uniform density $1_{[0,1]}$. Then, as $n \to \infty$, then $P(1/x_n < c < \tilde{f}(0+)) \to 1$.*

Proof : As f is not uniform and $f(0+) = c$, $P(X_1 \geq 1/c) > 0$. Therefore, $P(c \leq 1/x_n) = P(x_n \leq 1/c) \to 0$. Again, $P(c \geq \tilde{f}(0+)) = P(\tilde{f}(0+)/f(0+) \leq$

1) $\to P(\sup_{k\geq 1} k/\Gamma_k \leq 1) = 0$. The lemma follows. □

Define the maximizer as $\hat{f}^c(x) = f_k$ (say), for all $x_{k-1} < x \leq x_k$ and all $k = 1, \dots, n$. Then the problem reduces to maximizing $\sum_{i=1}^{n} \log f_i$ subject to

$$f_1 \geq f_2 \dots \geq f_n \geq 0, \tag{7.1}$$

$$f_1 = c, \text{ and} \tag{7.2}$$

$$\sum_{i=1}^{n} f_i(x_i - x_{i-1}) = 1. \tag{7.3}$$

We will incorporate the equality constraints (7.2) and (7.3) using lagrange multipliers. Define the vectors $f = (f_1, \dots, f_n)$, $\theta = \log f = (\theta_1, \dots, \theta_n)$ where the logarithm is defined coordinate-wise and a new optimality function

$$L_{\lambda,\beta}(\theta) = \sum_{i=1}^{n} \theta_i - n\lambda \sum_{i=1}^{n} e^{\theta_i}(x_i - x_{i-1}) - n\beta e^{\theta_1} = \sum_{i=1}^{n} [\theta_i - n u_i(\lambda,\beta) e^{\theta_i}]$$

where $u_1(\lambda, \beta) = \lambda x_1 + \beta$ and $u_i(\lambda, \beta) = \lambda(x_i - x_{i-1})$ for $i = 2, \dots, n$. Let $\Omega = \{\theta \in \mathbb{R}^n : \theta_1 \geq \theta_2 \dots \geq \theta_n\}$. Define for $\lambda, \beta > 0$, $\theta_{\lambda,\beta} = argmax_{\theta\in\Omega} L_{\lambda,\beta}(\theta)$. Let $\hat{\lambda}, \hat{\beta} > 0$ be such that $\hat{\theta} := \theta_{\hat{\lambda},\hat{\beta}}$ satisfies

$$e^{\hat{\theta}_1} = c, \text{ and} \tag{7.2a}$$

$$\sum_{i=1}^{n} e^{\hat{\theta}_i}(x_i - x_{i-1}) = 1 \tag{7.3a}$$

Then, $\hat{f} = e^{\hat{\theta}}$ maximizes $\sum_{i=1}^{n} \log f_i$ subject to (7.1), (7.2) and (7.3) (Using the theory of Lagrange multipliers). The next lemma characterizes the maximizer $\theta_{\lambda,\beta}$ we are looking for.

LEMMA 7.2. *Given $\lambda, \beta > 0$, $L_{\lambda,\beta}$ is maximized by $\theta_{\lambda,\beta} = \log f_{\lambda,\beta}$ in Ω, where $f_{\lambda,\beta}$ is the step function*

$$f_{\lambda,\beta}(x) = \min_{0\leq r<k} \max_{k\leq s\leq n} \frac{s-r}{n(u_{r+1} + \dots u_s)} \qquad x_{k-1} < x \leq x_k \tag{7.4}$$

Proof: Observe,

$$\frac{\partial L_{\lambda,\beta}}{\partial \theta_i} = 1 - nu_i(\lambda,\beta)e_i^{\theta}$$
$$\frac{\partial^2 L_{\lambda,\beta}}{\partial \theta_i \partial \theta_j} = \begin{cases} 0 & i \neq j \\ -nu_i(\lambda,\beta)e^{\theta_i} & i = j \end{cases}$$

Hence, $L_{\lambda,\beta}$ is concave for $\lambda, \beta > 0$. Using Kuhn-Tucker criteria, a necessary and sufficient condition for $\theta_{\lambda,\beta} \in \Omega$ to maximize $L_{\lambda,\beta}$ is, $\nabla L^t_{\lambda,\beta}(\theta_{\lambda,\beta})(\theta - \theta_{\lambda,\beta}) \leq 0$ for all $\theta \in \Omega$, or equivalently, $\sum_{i=1}^n [1 - nu_i(\lambda,\beta) f_{i,\lambda,\beta}](\theta_i - \theta_{i,\lambda,\beta}) \leq 0$ for all $\theta \in \Omega$. The solutions are given by the projection of $[1/nu_1, \ldots, 1/nu_n]^t$ on Ω with respect to the weights $[u_1, \ldots, u_n]^t$, and can be explicitly given as in (7.4). □

The following lemma characterizes the $\hat{\theta}$ we require to find the CPMLE.

LEMMA 7.3. *There exists unique $\hat{\lambda}, \hat{\beta} > 0$ such that $\hat{\theta} = \theta_{\hat{\lambda},\hat{\beta}}$ satisfies (7.2a) and (7.3a).*

Proof: Define,
$$\hat{\lambda} = \min_{cx_s < 1} \frac{1 - s/n}{1 - cx_s}.$$
The definition holds since $cx_1 < \tilde{f}(x_1)x_1 < 1$, and therefore the set $\{cx_s < 1\}$ is non-empty. Clearly, since $cx_n > 1$, $1 - s/n > 0$ for all s such that $cx_s < 1$, and therefore, $\hat{\lambda} > 0$. So, $1 - s/n \geq \hat{\lambda}(1 - cx_s)$ for all s, with equality at least once. Therefore, $\hat{\lambda}cx_s + 1 - \hat{\lambda} \geq s/n$ and equivalently

$$\frac{s}{n(\hat{\lambda}cx_s + 1 - \hat{\lambda})} \leq 1$$

for all s, with equality at least once. Also define $\hat{\beta} = (1 - \hat{\lambda})/c$. Then,

$$e^{\hat{\theta}_1} = \max_{1 \leq s \leq n} \frac{s}{n(\hat{\lambda}x_s + \hat{\beta})} = c \max_{1 \leq s \leq n} \frac{s}{n(\hat{\lambda}cx_s + 1 - \hat{\lambda})} = c,$$

since $\min_s n(1-\hat{\lambda}(1-cx_s))/s = 1$.

Let k denote the last index for which $\tilde{f}(x_k) = \tilde{f}(0+) > c$. Then, $cx_k < \tilde{f}(x_k)x_k = k/n < 1$. It also follows that $1-k/n < 1-cx_k$ and therefore $\hat{\lambda} < 1$. Therefore $\hat{\beta} > 0$. Again, $\sum_{i=1}^n [1 - nu_i(\hat{\lambda}, \hat{\beta})e^{\hat{\theta}_i}](\theta_i - \hat{\theta}_i) \leq 0$ for all $\theta \in \Omega$. In particular, since $\hat{\theta} \pm 1/n \in \Omega$, it follows that $\sum_{i=1}^n [1-nu_i(\hat{\lambda}, \hat{\beta})e^{\hat{\theta}_i}] = 0$, which implies $1 = (\hat{\lambda}x_1 + \hat{\beta})c + \sum_{i=2}^n \hat{\lambda}(x_i - x_{i-1})e^{\hat{\theta}_i} = \hat{\lambda}(\sum_{i=1}^n e^{\hat{\theta}_i}(x_i - x_{i-1})) + 1 - \hat{\lambda}$, which in turn implies that $\sum_{i=1}^n e^{\hat{\theta}_i}(x_i - x_{i-1}) = 1$. Hence, $\hat{\theta}$ satisfies both (7.2a) and (7.3a). The uniqueness follows, since (7.2a) and (7.3a) holds if and only if $\hat{\lambda} = 1 - c\hat{\beta}$ and $\min_s n(1-\hat{\lambda}(1-cx_s))/s = 1$. □

7.2. *Proof of Theorem 2.2 .* Let s_1 denote the last index k for which $\hat{f}^c(x_k) = \hat{f}^c(0+) = c$. Now, from Equation (7.4), it is clear that it is indeed the non-penalized unconstrained MLE for the deformed data set $\{\hat{\beta}+\hat{\lambda}x_i, i = 1, \ldots, n\}$. Since $\hat{\beta}+\hat{\lambda}x_{s_1}$ is a break-point of this new data set, with common value c before that, it follows that, $s_1/n = c(\hat{\beta} + \hat{\lambda}x_{s_1}) = 1 - \hat{\lambda} + \hat{\lambda}cx_{s_1}$, implying $s_0 = s_1$. Moreover, for $k > s_0$,

$$\begin{aligned}
\hat{f}^c(x_k) &= \min_{s_0 \leq r < k} \max_{k \leq s \leq n} \frac{s-r}{n(u_{r+1} + \ldots + u_s)} \\
&= \min_{s_0 \leq r < k} \max_{k \leq s \leq n} \frac{s-r}{n\hat{\lambda}(x_s - x_r)} \\
&= \frac{\tilde{f}(x_k)}{\hat{\lambda}}
\end{aligned}$$

x_{s_0} is a break-off point for $\tilde{f}$ as well, since $\sum_{i=s_0+1}^n \tilde{f}(x_i)(x_i - x_{i-1}) = \hat{\lambda}\sum_{i=s_0+1}^n \hat{f}^c(x_i)(x_i - x_{i-1}) = \hat{\lambda}(1 - cx_{s_0}) = 1 - s_0/n$. The result follows. □

7.3. *Proof of Propositions 3.1 and 3.2.* We start with a Proposition that gives us the limit behavior of $\hat{\lambda}$, $\hat{\gamma}$, x_{s_0} and x_{m_0}.

PROPOSITION 7.1. *Recall the sequence of standard Brownian motions $\mathbb{B}_n^1$ from Lemma 3.1. Then,*

(a)

$$\hat{\lambda} = \inf_{[0,1/c)} \frac{1 - F_n(x)}{1 - cx} \to 1, \qquad \textit{almost surely.}$$

(b)

$$n^{2/3}(\hat{\lambda} - 1) = (c^2/a)^{1/3} \inf_{[0,(na^2/c^4)^{1/3})} [t^2 + \mathbb{B}_n^1(t)] + o_p(1), \ \textit{and}$$

$$n^{1/3} x_{s_0} = (c/a^2)^{1/3} \textit{argmin}_{[0,(na^2/c^4)^{1/3})} [t^2 + \mathbb{B}_n^1(t)] + o_p(1)$$

(c) Let $\alpha = (ac)^{-1/3} n^{-2/3}$. *Then,*

$$n\alpha[\hat{f}(0+) - c] = -\inf_{t>0} \frac{\mathbb{B}_n^1(t) + 1 + t^2}{t + (a/nc^2)^{1/3}} + o_p(1), \ \textit{and} \tag{7.5}$$

$$n^{1/3} x_{m_0} = (c/a^2)^{1/3} \textit{argmin}_{t \geq 0} \frac{\mathbb{B}_n^1(t) + 1 + t^2}{t + (a/nc^2)^{1/3}} + o_p(1).$$

where the approximations hold up to an identical copy of the data.

Proof : (a) For any $x < 1/c$, let $x_i \leq x < x_{i+1}$. Then, $(1 - F_n(x))/(1 - cx) = (1 - i/n)/(1 - cx) \geq (1 - i/n)/(1 - cx_i)$. Hence, $\hat{\lambda} = \min_{cx_s<1}(1 - s/n)/(1 - cx_s) = \inf_{[0,1/c)}(1 - F_n(x))/(1 - cx)$. Also, $\|F_n - F\|_\infty \to 0$ a.s. (using Glivenko - Cantelli theorem), implying $\inf_{[0,1/c)}(1-F_n(x))/(1-cx) \to \inf_{[0,1/c)}(1 - F(x))/(1 - cx) = 1$ a.s. as well. The last equality follows as,

$$\frac{1 - F(x)}{1 - cx} = \frac{1 - \int_0^x f(t)dt}{1 - cx} \geq \frac{1 - cx}{1 - cx} = 1$$

for any $x \geq 0$, with equality at $x = 0$. □

(b) For these, we recall the Hungarian representation in Lemma 3.1. To approximate the term $\mathbb{G}_n(F(x))$, we need to use a *modulus of continuity* given by Lévy as follows. For any n, as $\{\mathbb{B}_n\}$ is a Wiener process, we have

$$P[\limsup_{\delta \downarrow 0} \frac{\max_{t-s<\delta} |\mathbb{B}_n(t) - \mathbb{B}_n(s)|}{\sqrt{2\delta \log(1/\delta)}} = 1] = 1$$

See Theorem 9.25 of [3] for a reference. Taking countable intersection over all n, we get

$$P[\limsup_{\delta \downarrow 0} \frac{\max_{t-s<\delta} |\mathbb{B}_n(t) - \mathbb{B}_n(s)|}{\sqrt{2\delta \log(1/\delta)}} = 1 \text{ for all } n] = 1$$

Consequently, we can write

$$\begin{aligned}|\mathbb{G}_n(F(x)) - \mathbb{G}_n(cx)| &= |\mathbb{B}_n(F(x)) - \mathbb{B}_n(cx) - (F(x) - cx)\mathbb{B}_n(1)| \\ &= \sqrt{cx - F(x)}O_p(1) + (cx - F(x))O_p(1) \\ &= O_p(x) + O_p(x^2)\end{aligned}$$

as $x \downarrow 0$ and $n \to \infty$. Now,

$$\begin{aligned}\frac{1 - F_n(x)}{1 - cx} &= \frac{1 - F(x) + (1/\sqrt{n})\mathbb{G}_n(F(x)) - R_n(x)}{1 - cx} \\ &= \frac{1 - cx + ax^2 + (1/\sqrt{n})\mathbb{G}_n(F(x)) - R_n(x) + o(x^2)}{1 - cx} \\ &= 1 + [ax^2 + (1/\sqrt{n})\mathbb{G}_n(cx)] + O_p(x/\sqrt{n}) + O_p(x^2/\sqrt{n}) + O(\log n/n) \\ &= 1 + ax^2 + (1/\sqrt{n})\mathbb{B}_n(cx) + I_n(x)\end{aligned}$$

where $I_n(x) = O_p(x/\sqrt{n}) + O_p(x^2/\sqrt{n}) + O(\log n/n)$.

Define $b = (na^2/c)^{-1/3}$. Then,

$$\begin{aligned}\inf_{[0,1/c)}[ax^2 + (1/\sqrt{n})\mathbb{B}_n(cx)] &= \inf_{[0,1/bc)}[ab^2t^2 + (1/\sqrt{n})\mathbb{B}_n(cbt)] \\ &= \inf_{[0,1/bc)}[ab^2t^2 + \sqrt{(cb/n)}\mathbb{B}_n^1(t)] \\ &= (c^2/an^2)^{1/3} \inf_{[0,(na^2/c^4)^{1/3})}[t^2 + \mathbb{B}_n^1(t)]\end{aligned}$$

since $\mathbb{B}_n^1(t) = \mathbb{B}_n(cbt)/\sqrt{cb}$. Therefore,

$$\begin{aligned}n^{2/3}(\hat{\lambda} - 1) &= n^{2/3} \inf_{[0,1/c)}[ax^2 + (1/\sqrt{n})\mathbb{B}_n(cx) + I_n(x)] \\ &= (c^2/a)^{1/3} \inf_{[0,(na^2/c^4)^{1/3})}[t^2 + \mathbb{B}_n^1(t)] + o_p(1)\end{aligned}$$

since $n^{2/3}I_n(bt^*) = o_p(1)$ at the minimizer t^*. Moreover, the minimizer can be approximated as

$$\text{argmin}_{[0,1/c)}[ax^2 + \mathbb{B}_n(cx)/\sqrt{n}] = (c/a^2n)^{1/3}\text{argmin}_{[0,(na^2/c^4)^{1/3})}[t^2 + \mathbb{B}_n^1(t)].$$

Using the definition of x_{s_0} in Theorem 2.2, the second part follows. □

(c) We need a comparison result as,

$$0 \leq \hat{f}(0+) - \sup_{x>0} \frac{F_n(x)}{\alpha + x} = O_p(\alpha)$$

from [11]. It follows that the maximizers are asymptotically equivalent, since the maximum value for the latter quantity is of higher order than α.

Now,

$$\begin{aligned}
\sup_{x>0} \frac{F_n(x)}{\alpha + x} - c &= \sup_{x>0} \frac{F_n(x) - F(x) - \alpha c - cx + F(x)}{\alpha + x} \\
&= \sup_{x>0} \frac{-(1/\sqrt{n})\mathbb{G}_n \circ F(x) - \alpha c - ax^2 + o(x^2)}{\alpha + x} \\
&= -\inf_{x>0} \frac{(1/\sqrt{n})\mathbb{B}_n(cx) + \alpha c + ax^2}{\alpha + x} + II_n(x)
\end{aligned}$$

as before. Using the same scaling as before, we get,

$$\begin{aligned}
\frac{(1/\sqrt{n})\mathbb{B}_n(cx) + \alpha c + ax^2}{\alpha + x} &= \frac{(1/\sqrt{n})\mathbb{B}_n(cbt) + (c^2/an^2)^{1/3} + ab^2t^2}{(acn^2)^{-1/3} + bt} \\
&= (ca/n)^{1/3} \frac{\mathbb{B}_n^1(t) + 1 + t^2}{t + (a/nc^2)^{1/3}}
\end{aligned}$$

Clearly, (7.5) follows by taking infimum of both sides and noting that the remainder $II_n(x)$ is of smaller order than the minimum value when x is the minimizer. The second part follows noting that $x_{m_0} = argmax_{x_s \geq 0}(s/n)/(\alpha + \hat{\gamma}x_s) = argmax_{x \geq 0}(F_n(x))/(\alpha + \hat{\gamma}x)$ as in the definition of x_{s_0}, and hence is asymptotically same as $argmax_{x \geq 0}(F_n(x))/(\alpha + x)$ using the comparison result. The rest of the proof is similar to the proof of the last part of (b). □

Proof of Proposition 3.1. Let $\hat{F}_n$ and $\hat{F}_n^c$ be the CDF's corresponding to the penalized density estimators $\hat{f}$ and $\hat{f}^c$ respectively. Since $\hat{F}_n$ can be represented as the LCM of the CSD of the data set $\{\alpha + \hat{\gamma}x_i, i = 1, \ldots, n\}$, we can deduce, $F_n = 1 - \hat{\gamma}(1 - \hat{F}_n)$ at all the jump-points of $\hat{f}$. Similarly, $F_n = 1 - \hat{\lambda}(1 - \hat{F}_n^c)$ at all the jump-points of $\hat{f}_c$. Additionally, we know that,

- $\hat{f}(0+) - c = O_P(n^{-1/3})$.
- $\hat{\lambda} - 1 = O_P(n^{-2/3})$ and $\hat{\gamma} - 1 = -\alpha \hat{f}(0+) = O_P(n^{-2/3})$.
- Both x_{m_0} and x_{s_0} are $O_P(n^{-1/3})$. Therefore, using Theorem 3.1 of [5], we have, $\tilde{f}(x_{m_0}) - c = O_P(n^{-1/3})$ and $\tilde{f}(x_{s_0}) - c = O_P(n^{-1/3})$. The theorem is actually proved for points approaching zero at a fixed rate, but can be readily modified for random shrinking neighborhoods. Therefore, $\tilde{f}(x_{s_0}) - \hat{\lambda}\hat{f}(0+) = O_P(n^{-1/3}) = \tilde{f}(x_{m_0}) - \hat{\lambda}\hat{f}(0+)$.

We will consider two different cases.

Case 1: $x_{m_0} < x_{s_0}$. Then, the breakpoints of $\hat{f}^c$ also become breakpoints of $\hat{f}$. To start with,

$$
\begin{aligned}
\sum_{i=1}^{n} \log\Big(\frac{\hat{f}(x_i)}{\hat{f}^c(x_i)}\Big) &= n \log \frac{\hat{f}(0+)}{c} F_n(x_{m_0}) + n \int_{x_{m_0}}^{\infty} \log(\hat{f}/\hat{f}^c) dF_n \\
&= n \log \frac{\hat{f}(0+)}{c} [1 - \hat{\gamma}(1 - \hat{F}_n(x_{m_0}))] + n\hat{\gamma} \int_{x_{m_0}}^{\infty} \log(\hat{f}/\hat{f}^c) d\hat{F}_n \\
&= n(1-\hat{\gamma}) \log \frac{\hat{f}(0+)}{c} + n\hat{\gamma} \int \log(\hat{f}/\hat{f}^c) d\hat{F}_n \\
&= n(1-\hat{\gamma}) \frac{\hat{f}(0+) - c}{c} + n\hat{\gamma} \int \log(\hat{f}/\hat{f}^c) d\hat{F}_n + o_P(1)
\end{aligned}
$$

The last equality follows from a Taylor series expansion with negligible second derivative. Again,

$$
\begin{aligned}
n \int \log(\hat{f}/\hat{f}^c) d\hat{F}_n &= -n \int \log(\hat{f}^c/\hat{f}) d\hat{F}_n \\
&= -n \int \log(\hat{f}^c/\hat{f}) \hat{f} dx \\
&= -n \int (\hat{f}^c/\hat{f} - 1) \hat{f} dx + \frac{n}{2} \int (\hat{f}^c/\hat{f} - 1)^2 \hat{f} dx + I_n \\
&= \frac{n}{2} \int_0^{x_{s_0}} (\hat{f}^c/\hat{f} - 1)^2 \hat{f} dx + I_n + II_n
\end{aligned}
$$

Now,

$$I_n = \frac{n}{3}\int_0^{x_{m_0}} (\frac{c}{\hat f(0+)} - 1)^3 g\hat f(0+)dx \quad + \quad \frac{n}{3}\int_{x_{m_0}}^{x_{s_0}} (\frac{\tilde f}{\hat\lambda \hat f(0+)} - 1)^3 g\hat f(0+)dx$$
$$+ \quad O_P(n \sup_{(x_{s_0},\infty)} |\hat f^c/\hat f - 1|^3),$$

where g lies between $\hat f^c/\hat f$ and 1, and clearly bounded. The first term is $O_p(n^{-1/3})$, and on the range (x_{s_0},∞), $|\hat f^c/\hat f - 1|^3 = |\hat\gamma/\hat l - 1|^3 = O_P(1/n^2)$, so the last term is $O_P(n^{-1})$. The middle term is also $O_p(n^{-1/3})$, since

$$\begin{aligned}
\frac{n\hat f(0+)}{3}\Big|\int_{x_{m_0}}^{x_{s_0}} (\frac{\tilde f}{\hat\lambda\hat f(0+)} - 1)^3 g dx\Big| &\le \frac{n\|g\|}{3\hat\lambda^3 \hat f(0+)^2}\int_{x_{m_0}}^{x_{s_0}} |\tilde f - \hat\lambda \hat f(0+)|^3 dx \\
&\le \frac{n\|g\|}{3\hat\lambda^3 \hat f(0+)^2}(x_{s_0} - x_{m_0})\Big(|\tilde f(x_{m_0}) - \hat\lambda\hat f(0+)|^3 \\
&\qquad + |\tilde f(x_{s_0}) - \hat\lambda \hat f(0+)|^3\Big) \\
&= O_P(n^{-1/3})
\end{aligned}$$

Combining all of them, $I_n = O_P(n^{-1/3})$. Also, $II_n = \frac{n}{2}\int_{x_{s_0}}^{\infty} (\hat f^c/\hat f - 1)^2 \hat f dx \le \frac{n}{2}(\frac{\hat\gamma}{\hat\lambda} - 1)^2 = O_P(n^{-1/3})$. Therefore,

$$\begin{aligned}
n\int \log{(\hat f/\hat f^c)} d\hat F_n &= \frac{n x_{m_0}}{2\hat f(0+)}(\hat f(0+) - c)^2 + \frac{n}{2}\int_{x_{m_0}}^{x_{s_0}} (\hat f^c/\hat f - 1)^2 \hat f dx + o_P(1) \\
&= \frac{n x_{m_0}}{2c}(\hat f(0+) - c)^2 + \frac{n}{2}\int_{x_{m_0}}^{x_{s_0}} (\frac{\hat\gamma c}{\tilde f} - 1)^2 \frac{\tilde f}{\hat\gamma} dx + o_P(1) \\
&= \frac{n x_{m_0}}{2c}(\hat f(0+) - c)^2 + \frac{n}{2}\int_{x_{m_0}}^{x_{s_0}} \frac{(\tilde f - c)^2}{\tilde f} dx + o_P(1) \\
&= \frac{n x_{m_0}}{2c}(\hat f(0+) - c)^2 + \frac{n}{2c}\int_{x_{m_0}}^{x_{s_0}} (\tilde f - c)^2 dx + o_P(1)
\end{aligned}$$

Since both terms are stochastically bounded, and $\hat\gamma \xrightarrow{P} 1$, the representation follows. □

Case 2: $x_{s_0} \le x_{m_0}$. We reverse the role of $\hat f$ and $\hat f^c$. Here,

$$\sum_{i=1}^n \log\Big(\frac{\hat f(x_i)}{\hat f^c(x_i)}\Big) = n(1-\hat\lambda)\frac{\hat f(0+) - c}{c} + n\hat\lambda\int \log{(\hat f/\hat f^c)} d\hat F_n^c + o_P(1)$$

Again,

$$
\begin{aligned}
& n\int \log(\hat f/\hat f^c)\,d\hat F_n^c \\
= {} & n\int \log(\hat f/\hat f^c)\hat f^c dx \\
= {} & -\frac{n}{2}\int_0^{x_{m_0}}(\hat f/\hat f^c-1)^2\hat f^c dx + o_P(1) \qquad \text{using similar argument,} \\
= {} & -\frac{n x_{s_0}}{2c}(\hat f(0+)-c)^2-\frac{n}{2}\int_{x_{s_0}}^{x_{m_0}}(\hat f/\hat f^c-1)^2\hat f^c dx + o_P(1) \\
= {} & -\frac{n x_{s_0}}{2c}(\hat f(0+)-c)^2-\frac{n}{2}\int_{x_{s_0}}^{x_{m_0}}(\frac{\hat\lambda\hat f(0+)}{\tilde f}-1)^2\frac{\tilde f}{\hat\lambda}dx + o_P(1) \\
= {} & -\frac{n x_{s_0}}{2c}(\hat f(0+)-c)^2-\frac{n}{2}\int_{x_{s_0}}^{x_{m_0}}\frac{(\tilde f-\hat f(0+))^2}{\tilde f}dx + o_P(1) \\
= {} & -\frac{n x_{s_0}}{2c}(\hat f(0+)-c)^2-\frac{n}{2c}\int_{x_{s_0}}^{x_{m_0}}(\tilde f-\hat f(0+))^2 dx + o_P(1),
\end{aligned}
$$

and the proof follows similarly as in Case 1. □

Proof of Proposition 3.2. We use Proposition 7.1 and observe that $\mathbb{B}_n^1$ is a standard Brownian motion having the same law as $\mathbb{B}$. The rest follows. □

REFERENCES

[1] BANERJEE, M. and WELLNER, J. (2001). Likelihood ratio tests for monotone functions. *Ann. Statist.* **29** 1699-1731.

[2] GRENANDER, U. (1956). On the theory of mortality measurement, II *Skand. Aktuarietidskr.* **39** 125-153.

[3] KARATZAS, I. and SHREVE, S. E. (1991). Brownian motion and stochastic calculus. *Springer, New York*

[4] KEIDING, N., KVIST, K., HARTVIG, H. and TVEDE, M. (2002). Estimating time to pregnancy from current duration in a cross-sectional sample, *Biostatistics* **3**, 565-578.

[5] KULIKOV, V. N. and LOPUHAA, H.P. (2006). The behavior of the NPMLE of a decreasing density near the boundaries of the support. *Ann. Statist.* **34**, 2, 742-768

[6] LYNDEN-BELL, D. (1991). Eddington Malmquist Bias, streaming motions, and the distribution of galaxies. *Statistical Challenges in Modern Astronomy*, Edited by J.Babu and E. Feigelson.

[7] PAL, J. K. (2006). Statistical analaysis and inference in shape-restricted problems with applications to astronomy. Doctoral Dissertation. *University of Michigan*

[8] VAN DER VAART, A. W. (1991). Asymptotic Statistics. *Cambridge series in Statistical and Probabilistic Mathematics.*

[9] VARDI, Y. (1989). Multiple censoring, renewal processes, deconvolution and decreasing Density : nonparametric estimation. *Biometrika* **76** 751-761.

[10] WOODROOFE, M. B. (1991). Discussion of "Eddington Malmquist bias, streaming motions, and the distribution of galaxies". *Statistical Challenges in Modern Astronomy.* Edited by J.Babu and E. Feigelson, 217-220.

[11] WOODROOFE, M. B. and SUN, J. (1993). A penalized maximum likelihood estimate of f(0+) when f is non-increasing. *Statist. Sinica* **3** 501-515.

[12] WOODROOFE, M. B. and ZHANG, R.(1999). Isotonic estimation for grouped data. *Statist. Probab. Lett.* **45**, 41-47.

19 T. W. Alexander Drive,
Research Triangle Park,
NC-27606
E-mail: jpal@samsi.info