A curious fact about k-monotone functions ($k > 2$)

Fadoua Balabdaoui [1]

Department of Statistics, University of Washington

November 17, 2003

Abstract

It is proved that for any given integer $k > 2$ and a k-monotone function g, there exists a k-monotone function $\tilde{g} \leq g$ of the form $c(a-x)_+^{k-1}$ and passing through a fixed point in the support of g. The result is motivated by the problem of the existence of the LSE of a k-monotone density.

Keywords: k-monotone functions; k-monotone densities; Nonincreasing densities; Convex densities; Nonparametric estimation; Least Squares Estimator; Maximum Likelihood Estimator, One-sided approximation, Splines.

Background

The class of k-monotone functions has been of interest over the past fifty years (see e.g. [1], [2], [3], [4]). For $k \geq 2$, a function g defined on $(0, \infty)$ is said to be k-monotone if and only if for all $m = 0, \cdots, k-2$, $(-1)^m g^{(m)}$ is nonnegative, nonincreasing and convex (see e.g. [1], [2], [3]). For $k = 1$, g is simply nonnegative and nonincreasing on $(0, \infty)$. Williamson [2] studied the properties of k-monotone functions and gave a very useful characterization in the form of an integral representation. He established that g is k-monotone on $(0, \infty)$ if and only if there exists a nondecreasing function γ such that

$$g(x) = \int_0^\infty (1-ux)_+^{k-1} d\gamma(u), \quad \text{for all} \quad x > 0, \tag{1}$$

where

$$t_+ = \begin{cases} t, & \text{if } t \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

[1]Research supported in part by National Science Foundation grant DMS-0203320

In statistics, problems related to the estimation of a k-monotone density when $k = 1$ and $k = 2$; i.e., nonincreasing, and nonincreasing and convex density received special attention (see e.g. [5], [6], [7], [8], [9], [10], [11], [12], [13]). Based on n independent observations from the true density, two nonparametric estimators were mainly considered: the Maximum Likelihood Estimator (MLE) and the Least Squares Estimator (LSE). The LSE of a nonincreasing and convex density is defined in Groeneboom et al. [11] as the minimizer of the criterion function

$$\Phi_n(g) = \frac{1}{2}\int_0^\infty g^2(t)dt - \int_0^\infty g(t)d\mathbb{G}_n(t)$$

over the class of nonincreasing and convex functions that are square integrable, where $\mathbb{G}_n$ is the empirical distribution function based on n i.i.d. observations from the true density. As it was pointed out by Groeneboom et al. [11], minimizing Φ_n has a very intuitive appeal. In fact, if we assume that $\mathbb{G}_n$ has a density g_n with respect to Lebesgue measure, then we can rewrite Φ_n as

$$\Phi_n(g) = \frac{1}{2}\int_0^\infty (g(x) - g_n(x))^2 dx - \frac{1}{2}\int_0^\infty g_n^2(x)dx,$$

where the second term does not depend on g. The LSE can be viewed as the closest 2-monotone function to the "empirical density" g_n in the L_2 sense.

It was shown by Groeneboom et al. [11] that both the MLE and LSE of a nonincreasing and convex density are asymptotically equivalent. However, the LSE appears to be intimately connected with the stochastic process involved in the asymptotic distribution. Beside the fact that they have very similar characterizations, Groeneboom et al. [12] showed that this limiting stochastic process is equal to $\lim_{c\to\infty} g_c$ (in an appropriate sense), where g_c is nothing but the minimizer of the functional

$$\phi_c(g) = \frac{1}{2}\int_{-c}^{c} g^2(t)dt - \int_{-c}^{c} g(t)dX(t)$$

over the class of convex functions g such that $g(-c) = g(c) = 12c^2$, where $dX(t) = dW(t) + 12t^2$ and W is a two-sided Brownian motion starting at 0. Through the LSE, a nice parallelism is established between the "finite sample world" and the "infinite sample" one. In the latter, $\lim_{c\to\infty} g_c$ would be the analogue of the LSE and the parabola $t \mapsto 12t^2, t \in \mathbb{R}$, would be the analogue of the convex density that is being estimated.

The appeal of the LSE was the first driving motivation for the author to investigate its existence for k-monotone densities in the case of $k > 2$. In addition to a compactness argument, the proof of its existence in the particular case of $k = 2$ is partially based on the following fact: for any given nonincreasing and convex function g and n different points in its support, we can find a nonincreasing and convex piecewise linear function that stays underneath g and passes through the n different points (see Groeneboom et al. [11]). It is natural to wonder whether such a fact continues to hold for k-monotone functions when $k > 2$. In other words, given an integer $k > 2$, a k-monotone function g and n different points in its support, can we find a function $\tilde{g}$ of the form

$$\tilde{g}(x) = \tilde{c}_1(\tilde{a}_1 - x)_+^{k-1} + \cdots + \tilde{c}_m(\tilde{a}_m - x)_+^{k-1}, \quad x > 0$$

such that $\tilde{g} \leq g$ and $\tilde{g}$ passes through the n fixed points?

In the proof of Theorem 1, the author gives an answer to the question when $n = 1$. The difficulty can be realized already for $n = 2$. Whereas for $k = 2$, it is very easy to find $\tilde{g}$ for an arbitrary n by taking the tangents of g at the considered points, such a procedure cannot be generalized easily for functions that are as smooth and shape-constrained as the k-monotone functions. Existence of the LSE when $n = 1$ follows easily from the following theorem.

Theorem 1. Let $k > 2$ be an integer. For any given k-monotone function g and $\tau > 0$ in its support, there exists a function $\tilde{g}$ of the form

$$\tilde{g}(x) = \tilde{c}(\tilde{a} - x)_+^{k-1},$$

such that

$$\begin{cases} \tilde{g}(x) \leq g(x), & \text{for} \quad x \in (0, \infty), \\ \tilde{g}(\tau) = g(\tau). \end{cases}$$

Proof. We preface the proof of Theorem 1 by the following lemma:

Lemma. g is k-monotone on $(0, \infty)$ if and only if there exists a positive measure μ such that for all $x > 0$

$$g(x) = \int_0^\infty (t - x)_+^{k-1} d\mu(t).$$

Proof. See Williamson [2].

Proof of Theorem 1. We would like to find $\tilde{c}$ and $\tilde{a}$ such that the function $\tilde{g}$ passes through $(\tau, g(\tau))$ and satisfies $\tilde{g} \leq g$. It is natural to impose that $\tilde{g}'(\tau) = g'(\tau)$. Then, $\tilde{c}$ and $\tilde{a}$ are uniquely determined and given by

$$\tilde{c} = \frac{g(\tau)}{\left[\frac{(k-1)g(\tau)}{-g'(\tau)}\right]^{k-1}} = \frac{\left(\int_0^\infty (t-\tau)_+^{k-2} d\mu(t)\right)^{k-1}}{\left(\int_0^\infty (t-\tau)_+^{k-1} d\mu(t)\right)^{k-2}},$$

and

$$\tilde{a} = \tau + \frac{(k-1)g(\tau)}{-g'(\tau)} = \tau + \frac{\left(\int_0^\infty (t-\tau)_+^{k-1} d\mu(t)\right)}{\left(\int_0^\infty (t-\tau)_+^{k-2} d\mu(t)\right)}.$$

Define $h = g - \tilde{g}$. Let $0 < x \le \tau$. For ease of notation, $h^{(k-1)}(x)$ can be interpreted either as the left or right $(k-1)$-th derivative of h at the point x. We have

$$\begin{aligned}
&(-1)^{k-1} h^{(k-1)}(x) \\
&= (-1)^{k-1} g^{(k-1)}(x) - (k-1)!\tilde{c} \\
&= (k-1)! \left(\int_0^\infty 1_{[x,\infty)}(t) d\mu(t) - \frac{\left(\int_0^\infty (t-\tau)_+^{k-2} d\mu(t)\right)^{k-1}}{\left(\int_0^\infty (t-\tau)_+^{k-1} d\mu(t)\right)^{k-2}} \right) \\
&= (k-1)! \frac{\left(\int_0^\infty 1_{[x,\infty)}(t) d\mu(t)\right) \left(\int_0^\infty (t-\tau)_+^{k-1} d\mu(t)\right)^{k-2} - \left(\int_0^\infty (t-\tau)_+^{k-2} d\mu(t)\right)^{k-1}}{\left(\int_0^\infty (t-\tau)_+^{k-1} d\mu(t)\right)^{k-2}}.
\end{aligned}$$

Fix $0 < x \le \tau$. Let $r = k-1$ and $s = (k-1)/(k-2)$. $r > 1$ and $1/r + 1/s = 1$. By Hölder's inequality, we have

$$\begin{aligned}
&\left(\int_0^\infty 1_{[x,\infty)}(t)d\mu(t)\right)^{\frac{1}{k-1}}\left(\int_0^\infty (t-\tau)_+^{k-1}d\mu(t)\right)^{\frac{k-2}{k-1}}\\
&\geq \int_0^\infty (t-\tau)_+^{k-2}1_{[x,\infty)}(t)d\mu(t)\\
&\geq \int_0^\infty (t-\tau)_+^{k-2}d\mu(t)
\end{aligned}$$

since $x \leq \tau$. This implies that $(-1)^{k-1}h^{(k-1)}(x) \geq 0$ for all $0 < x \leq \tau$. It follows that $(-1)^{k-2}h^{(k-2)}$ is nonincreasing on $(0,\tau]$. Furthermore, we have

$$(-1)^{k-2}h^{(k-2)}(\tau) = (k-1)!\left(\int_0^\infty (t-\tau)_+ d\mu(t) - \frac{\left(\int_0^\infty (t-\tau)_+^{k-2}d\mu(t)\right)^{k-2}}{\left(\int_0^\infty (t-\tau)_+^{k-1}d\mu(t)\right)^{k-3}}\right) \geq 0$$

by applying again Hölder's inequality where $r = k-2$ and $s = (k-2)/(k-3)$. Hence, $(-1)^{k-2}h^{(k-2)} \geq 0$ on $(0,\tau]$. Similarly, we show that $(-1)^j h^{(j)} \geq 0$ on $(0,\tau]$ for $0 \leq j \leq k-3$. In particular, $h \geq 0$ or equivalently $\tilde{g} \leq g$ on $(0,\tau]$.

To finish off, we only need to show that the same inequality holds on (τ,∞) and for that it suffices to establish that

$$\left(\frac{\tilde{a}-x}{\tilde{a}-\tau}\right)^{k-1} \leq \frac{\int_0^\infty (t-x)_+^{k-1}d\mu(t)}{\int_0^\infty (t-\tau)_+^{k-1}d\mu(t)}, \tag{2}$$

for all $x \in (\tau,\tilde{a}]$ since

$$\tilde{g}(x) = g(\tau)\left(\frac{\tilde{a}-x}{\tilde{a}-\tau}\right)_+^{k-1}, \quad \text{for} \quad x \in (0,\infty).$$

Now, the function

$$x \mapsto \int_0^\infty \left(\frac{t-x}{\tilde{a}-x}\right)_+^{k-1} d\mu(t)$$

is nondecreasing in $x \in (\tau,\tilde{a})$ since its first derivative

$$x \mapsto \int_0^\infty (k-1)\frac{(t-\tilde{a})}{(\tilde{a}-x)^2}\left(\frac{t-x}{\tilde{a}-x}\right)_+^{k-2} d\mu(t), \qquad x \in (\tau,\tilde{a})$$

is nonnegative. In fact, the sign of this derivative is the same as the sign of the function defined by

$$x \mapsto \int_0^\infty (t-\tilde{a})(t-x)_+^{k-2}d\mu(t), \qquad x \in (\tau,\tilde{a}).$$

But for all $x \in (\tau, \tilde{a})$, we have

$$\begin{aligned}
&\int_0^\infty (t-\tilde{a})(t-x)_+^{k-2} d\mu(t) \\
&= \int_0^\infty (t-\tau)(t-x)_+^{k-2} d\mu(t) - (\tilde{a}-\tau)\int_0^\infty (t-x)_+^{k-2} d\mu(t) \\
&= \int_0^\infty (t-\tau)(t-x)_+^{k-2} d\mu(t) - \frac{\left(\int_0^\infty (t-\tau)_+^{k-1} d\mu(t)\right)}{\left(\int_0^\infty (t-\tau)_+^{k-2} d\mu(t)\right)} \int_0^\infty (t-x)_+^{k-2} d\mu(t) \quad (3)
\end{aligned}$$

where in (3), $\tilde{a}$ is replaced by its expression. To conclude, we only need to show that for all $x \in (\tau, \tilde{a})$

$$\frac{\int_0^\infty (t-\tau)(t-x)_+^{k-2} d\mu(t)}{\int_0^\infty (t-x)_+^{k-2} d\mu(t)} \geq \frac{\int_0^\infty (t-\tau)_+^{k-1} d\mu(t)}{\int_0^\infty (t-\tau)_+^{k-2} d\mu(t)}.$$

By writing $(t-\tau)(t-x)_+^{k-2} = (t-x)_+^{k-1} + (x-\tau)(t-x)_+^{k-2}$, the latter inequality is equivalent to

$$\frac{\int_0^\infty (t-x)_+^{k-1} d\mu(t)}{\int_0^\infty (t-x)_+^{k-2} d\mu(t)} + x - \tau \geq \frac{\int_0^\infty (t-\tau)_+^{k-1} d\mu(t)}{\int_0^\infty (t-\tau)_+^{k-2} d\mu(t)}. \quad (4)$$

Now, consider the function f defined by

$$f(x) = \frac{\int_0^\infty (t-x)_+^{k-1} d\mu(t)}{\int_0^\infty (t-x)_+^{k-2} d\mu(t)} + x - \tau, \qquad x \in (\tau, \tilde{a}).$$

We have for all $x \in (\tau, \tilde{a})$

$$
\begin{aligned}
&f'(x) \\
&= \frac{-(k-1)\left(\int_0^\infty (t-x)_+^{k-2}d\mu(t)\right)^2 + (k-2)\left(\int_0^\infty (t-x)_+^{k-3}d\mu(t)\right)\left(\int_0^\infty (t-x)_+^{k-1}d\mu(t)\right)}{\left(\int_0^\infty (t-x)_+^{k-2}d\mu(t)\right)^2} + 1 \\
&= -(k-2) + (k-2)\frac{\left(\int_0^\infty (t-x)_+^{k-3}d\mu(t)\right)\left(\int_0^\infty (t-x)_+^{k-1}d\mu(t)\right)}{\left(\int_0^\infty (t-x)_+^{k-2}d\mu(t)\right)^2} \\
&= (k-2)\frac{\left(\int_0^\infty (t-x)_+^{k-1}d\mu(t)\right)\left(\int_0^\infty (t-x)_+^{k-3}d\mu(t)\right) - \left(\int_0^\infty (t-x)_+^{k-2}d\mu(t)\right)^2}{\left(\int_0^\infty (t-x)_+^{k-2}d\mu(t)\right)^2}.
\end{aligned}
$$

Let $r = s = 1/2$ and consider $f_1(t) = (t-x)_+^{(k-3)/2}$ and $f_2(t) = (t-x)_+^{(k-1)/2}$ defined on $(0,\infty)$. $1/r + 1/s = 1$ and $(f_1 f_2)(t) = (t-x)_+^{k-2}$ for $t \in (0,\infty)$. By Hölder's inequality, we have

$$
\left(\int_0^\infty f_1^2(t)d\mu(t)\right)^{\frac{1}{2}}\left(\int_0^\infty f_2^2(t)d\mu(t)\right)^{\frac{1}{2}} \geq \int_0^\infty (f_1 f_2)(t)d\mu(t).
$$

Therefore, $f' \geq 0$ on $(\tau, \tilde{a})$. Hence, f is nondecreasing and $f(x) \geq f(\tau)$ for all $x \in (\tau, \tilde{a}]$. This implies the inequality in (4) and (2) follows. ■

We apply the previous result to show that the LSE exists when the sample size n is equal to 1.

Theorem 2. Let X be an observation from a k-monotone density g_0 such that $g_0(X) > 0$. Φ_1 admits a unique minimizer $\tilde{g}_1(x) = \tilde{c}_1(\tilde{a}_1 - x)_+^{k-1}$ where

$$
\tilde{a}_1 = (2k-1)X, \quad \text{and} \quad \tilde{c}_1 = \frac{(2(k-1))^{k-1}}{(2k-1)^{2k-2}}\frac{1}{X^k}. \tag{5}
$$

Proof. From Theorem 1, we see that it is beneficial to "move" from any candidate g for the minimization problem to the function $\tilde{g}(x) = \tilde{c}(\tilde{a} - x)^{k-1}$ passing through the point X and staying underneath g. In fact,

$$
\tilde{g}(X) = g(X)
$$

and

$$
\int_0^\infty \tilde{g}^2(x)dx \leq \int_0^\infty g^2(x)dx.
$$

Hence, $\Phi_1(\tilde{g}) \leq \Phi_1(g)$ implying that if the LSE $\tilde{g}_1$ exists, it has to be of the form $\tilde{g}_1(x) = \tilde{c}_1(\tilde{a}_1 - x)_+^{k-1}$, for some $\tilde{c}_1 > 0$ and $\tilde{a}_1 > 0$.

Minimizing Φ_1 on the space of k-monotone functions g of the form $g(x) = c(a - x)_+^{k-1}$, $x > 0$ is equivalent to minimizing the function Q defined on $(0, \infty) \times (X, \infty)$ by

$$Q(a, c) = \frac{1}{2(2k-1)} c^2 a^{2k-1} - c(a - X)^{k-1}.$$

Easy calculations yield the unique solution given in (5). ■

Remark. Whereas for $k = 2$, Groeneboom et al. [11] showed that the LSE $\tilde{g}_n$ is a genuine density, the author conjectures that this fact does not hold for $k > 2$. Its failure can be already seen when $n = 1$. In fact, one can easily verify that

$$\int_0^\infty \tilde{g}_1(x)dx = \frac{2k-2}{k}\left(1 - \frac{1}{2k-1}\right)^{k-2} \rightarrow 2e^{-1/2} \approx 1.21$$

as $k \rightarrow \infty$.

Discussion

As it was mentioned earlier, finding a k-monotone spline that stays below a given k-monotone function and passes through $n > 1$ fixed points is a much harder problem. If this spline exists for $n > 1$, it would be a one-sided approximation of the given k-monotone function. In the literature on approximation via splines, the problem of finding the best one-sided approximation of a given function in $C^1[a, b]$, where a and b are finite was already considered. When the measure of optimality is the L_1-distance, it was shown that the problem admits a unique solution in the space of splines with degree $m \geq 2$ and fixed $p \geq 1$ knots (see e.g. [14, Theorem 5.6]). Also, there has been some very recent work done by Kopotun and Shadrin [15] on shape-preserving approximation by free knot splines. However, the functions approximated are k-convex; i.e., only their $(k-2)$-th derivative has to be convex and in this sense, the functions considered in this paper are more constrained. Moreover, it seems that most of the literature is concerned with functions that are defined on a compact set, an undesirable restriction in the current problem. This more general question is then still open.

Acknowledgment: The author would like to thank Professor Tilmann Gneiting for valuable discussions on the proof of Theorem 1. Thanks are also due to Professor Jon Wellner for his helpful comments on this work.

References

[1]. D. V. Widder, "*The Laplace Transform,*" *Princeton University Press, Princeton, 1947.*

[2]. *R. E. Williamson, Multiply monotone functions and their Laplace transforms. Duke Math. J.* **23** *(1955), 189–207.*

[3]. *W. Feller, "An Introduction to Probability Theory and its Applications," Wiley, New York, 1971.*

[4]. *T. Gneiting, Radial positive definite functions generated by Euclid's hat, J. Multivariate Anal.* **69** *(1999), 88–119.*

[5]. *U. Grenander, On the theory of mortality measurement, Part II, Skand. Actuar.* **39** *(1956), 125 –153.*

[6]. *B. L. S. Prakasa Rao, Estimation of a unimodal density, Sankhya. Ser. A* **31** *(1969), 23–36.*

[7]. *P. Groeneboom, Estimating a monotone density, Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II, Lucien M. LeCam and Richard A. Olshen, editors, Wadsworth, New York, 1985, 529–555.*

[8]. *T. Robertson, F. T. Wright, R. L. Dykstra, "Order Restricted Statistical Inference," Wiley, New York, 1988.*

[9]. *F. R. Hampel, Design, modelling and analysis of some biological datasets, Design, data and analysis, by some friends of Cuthbert Daniel, C. L. Mallows, editor, Wiley, New York, 1987, 111–115.*

[10]. *Y. Vardi, Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation, Biometrika* **76** *(1989), 751–761.*

[11]. *P. Groeneboom, G. Jongbloed and J. A. Wellner, Estimation of convex functions: characterizations and asymptotic theory, Ann. Statist.* **29** *(2001b), 1653–1698.*

[12]. *P. Groeneboom, G. Jongbloed and J. A. Wellner, A canonical process for estimation of convex functions: The "invelope" of integrated Brownian motion* $+t^4$*, Ann. Statist.* **29** *(2001a), 1620–1652.*

[13]. *J. A. Wellner, Gaussian white noise models: some results for monotone functions, in Crossing Boundaries: Statistical Essays in Honor of Jack Hall, IMS Lecture Notes - Monograph Series, Vol.43, J. E. Kolassa and D. Oakes, editors, 2003, 87–104.*

[14]. *G. Nürnberger, "Approximation by Spline Functions," Springer-Verlag, 1989.*

[15]. *K. Kopotun and A. Shadrin, On* k*-monotone approximation by free knot splines, Siam J. Math. Anal.* **34** *(2003), 901–924.*