

UNIQUENESS OF THE MAXIMUM LIKELIHOOD ESTIMATOR FOR k -MONOTONE DENSITIES

ARSENI SEREGIN

(Communicated by Edward C. Waymire)

ABSTRACT. We prove uniqueness of the maximum likelihood estimator for the class of k -monotone densities.

1. INTRODUCTION

A density f on $[0, +\infty)$ is k -monotone if and only if it can be represented in the form

$$f(x) = \int_0^{+\infty} \frac{1}{y} g_0\left(\frac{x}{y}\right) dG(y),$$

where G is some probability distribution on $(0, +\infty)$ and $g_0(x) \equiv k(1-x)_+^{k-1}$ is the $Beta(1, k)$ density. Thus the family \mathcal{M}_k of k -monotone densities is a mixture model defined by the basic density g_0 .

This class generalizes classes of monotone (1-monotone) and convex decreasing (2-monotone) densities on \mathbb{R}_+ . As k increases, the classes \mathcal{M}_k decrease and the class of completely monotone densities lies in the intersection of all \mathcal{M}_k . The paper [2] proves consistency and [3] gives the asymptotic limit theory for the maximum likelihood estimator (MLE) of a k -monotone density. For $k = 1$ the uniqueness of the MLE follows from a well known characterization of the Grenander estimator; for $k = 2$ the uniqueness was proved in [4] and for $k = 3$ in [1].

The approach suggested in [6] allows us to prove the uniqueness of the MLE values at data points. For k -monotone densities this property is mentioned in [2]. However, the uniqueness of the mixing measure which defines the MLE depends on the properties of a particular mixture model. In [5] the uniqueness and weak consistency were proved for the MLE of a completely monotone density. In [9] a general method was developed which allows us to prove uniqueness of the MLE in mixture models defined by strictly totally positive kernels such as the exponential kernel for completely monotone densities. For k -monotone densities the kernel g_0 is not strictly totally positive as follows from [11]. The uniqueness of the MLE for $k > 3$ has not been investigated before. Below, we answer this question by proving the following theorem:

Received by the editors December 20, 2009 and, in revised form, March 2, 2010.

2000 *Mathematics Subject Classification*. Primary 62G07.

Key words and phrases. Uniqueness, k -monotone density, mixture models, density estimation, maximum likelihood, nonparametric estimation, shape constraints.

This research was supported in part by NSF grant DMS-0804587.

Theorem 1.1. *Let $k \geq 2$ be a positive integer and let X_1, \dots, X_n be i.i.d. random variables with k -monotone density $f_0 \in \mathcal{M}_k$. With probability 1 the MLE \hat{f}_n is uniquely defined by some discrete mixing probability measure \hat{G}_n . The cardinality of the support set $\text{supp}(\hat{G}_n)$ is less than or equal to n . If $\text{supp}(\hat{G}_n)$ consists of points $Y_1 < \dots < Y_m$, $m < n$, then there exists a subset of order statistics $X_{(i_1)} < \dots < X_{(i_m)}$ such that*

$$X_{(i_j)} < Y_j < X_{(i_{j+k})},$$

where we assume $X_{(i_l)} = +\infty$ for $l > m$. Also we have $Y_m > X_{(n)}$.

2. PROOF

With probability 1 we have that the order statistics are distinct:

$$0 < X_{(1)} < \dots < X_{(n)}.$$

We assume that $X_{(l)} = +\infty$ for $l > n$.

The proof consists of several lemmas. The first lemma provides a tool for counting zeroes of a function.

Lemma 2.1. *Let $A \subseteq \mathbb{R}$ be an interval. Consider a function $f \in C^k(A)$. Then the following inequality is true for the number of zeroes $N(f, A)$ of f counted with multiplicities:*

$$N(f^{(k)}, A) \geq N(f, A) - k.$$

Here the multiplicity of zero x of the function $f \in C^k(A)$, $k \geq 0$, $x \in A$, is an integer $n \leq k+1$ such that

$$f(y) = \dots = f^{(n-1)}(y) = 0$$

and either $n = k+1$ or $f^{(n)}(y) \neq 0$.

Proof. Let x_i , $1 \leq i \leq m$, be zeroes of f on A and let n_i be its multiplicities. Then by Rolle's theorem the derivative $f' \in C^{k-1}(A)$ has $m-1$ zeroes between points x_i with multiplicities at least 1 and also zeroes x_i with multiplicities n_i-1 . Thus,

$$N(f', A) \geq m-1 + \sum_{i=1}^m (n_i - 1) = -1 + \sum_{i=1}^m n_i = N(f, A) - 1.$$

Applying this inequality k times, we obtain the result. \square

The proof of the second lemma builds the necessary tools for the main result. Note that discreteness of the mixing measure \hat{G}_n actually follows from general results about the MLE for mixture models proved in [6]. We repeat a part of that argument to make our proof self-contained and also to provide references to the classic convex analysis textbook [10] for the interested reader.

Lemma 2.2. *There exists a set Y of points $Y_1 < \dots < Y_m$, $m \leq n$, such that k -monotone density \hat{f}_n defined by a mixing probability measure \hat{G}_n is the MLE only if $\text{supp}(\hat{G}_n) \subseteq Y$. In other words, any MLE has the form*

$$(2.1) \quad \hat{f}_n(x) = \sum_{j=1}^m a_j \frac{k(Y_j - x)_+^{k-1}}{Y_j^k},$$

where $a_j \geq 0$, $\sum a_j = 1$ and $Y_j > X_{(j)}$. Moreover, the vector $(\hat{f}_n(X_{(i)}))_{i=1}^n$ is unique.

Proof. Following [6], [8] we define a curve $\Gamma \subset \mathbb{R}_+^n$ parameterized by $y \in \mathbb{R}_+$ as

$$\Gamma(y) = \left(\frac{1}{y} g_0 \left(\frac{X_{(i)}}{y} \right) \right)_{i=1}^n = \left(\frac{k(y - X_{(i)})_+^{k-1}}{y^k} \right)_{i=1}^n.$$

Since $(1/y)g_0(x/y)$ as a function of y is continuous and equal to zero at 0 and $+\infty$, we have that Γ is compact. Suppose that $f \in \mathcal{M}_k$ with corresponding mixing probability measure G . Then the vector $(f(X_{(i)}))_{i=1}^n$ belongs to the set $\bar{\Gamma} = \text{cl}(\text{conv}(\Gamma))$:

$$(f(X_{(i)}))_{i=1}^n = \int_0^{+\infty} \Gamma(y) dG(y).$$

Since Γ is compact it follows that $\bar{\Gamma}$ is also compact and $\bar{\Gamma} = \text{conv}(\Gamma)$ by ([10], Theorem 17.2). Therefore, the continuous function $L(z) \equiv \sum_{i=1}^n \log z_i$ attains its maximum on $\bar{\Gamma}$, where the z_i are the coordinates in the space \mathbb{R}_+^n . Let us denote $S \equiv \text{argmax}_{\bar{\Gamma}} L(z)$.

The function $L(z)$ is $-\infty$ on $\partial\mathbb{R}_+^n$ and $L(z) > -\infty$ on the interior $\text{ri}(\mathbb{R}_+^n)$ of \mathbb{R}_+^n . Since the intersection of Γ and $\text{ri}(\mathbb{R}_+^n)$ is not empty, we have $S \subset \text{ri}(\mathbb{R}_+^n)$. The function $L(z)$ is strictly concave; therefore S consists of a single point $b = (b_i)_{i=1}^n > 0$. Therefore for any MLE \hat{f}_n it follows that the vector $(\hat{f}_n(X_{(i)}))_{i=1}^n$ is equal to b and is unique. The gradient of $L(z)$ at b is proportional to $1/b \equiv (1/b_i)_{i=1}^n$.

We have $\dim(\bar{\Gamma}) = n$. Indeed if we consider n points $t_i \in (X_{(i)}, X_{(i+1)})$, then the vectors $\Gamma(t_i)$ are linearly independent ([11] or direct observation). By ([10], Theorem 27.4) the vector $1/b$ belongs to the normal cone of $\bar{\Gamma}$ at b . Since $1/b_i > 0$ for all i we have $b \in \partial\bar{\Gamma}$ and the plane α defined by the equation $\sum_{i=1}^n z_i/b_i = n$ is a support plane of $\bar{\Gamma}$ at b . Thus for $v_i = k/nb_i$ we have

$$p(y) \equiv y^k - \sum_{i=1}^n v_i (y - X_{(i)})_+^{k-1} \geq 0,$$

for all $y \geq 0$ and $p(y) = 0$ if and only if $y = 0$ or $\Gamma(y) \in \alpha$. Let us denote by Y the set of y such that $\Gamma(y) \in \alpha$; i.e. $\Gamma(Y) = \alpha \cap \Gamma$.

The intersection $\alpha \cap \bar{\Gamma}$ is an *exposed face* ([10], p. 162) of $\bar{\Gamma}$. By Theorem 18.3 from [10] we have $\alpha \cap \bar{\Gamma} = \text{conv}(\Gamma(Y))$ and by Theorem 18.1 we have $\text{supp}(\hat{G}_n) \subseteq Y$.

The function $p(y)$ belongs to $C^{(k-2)}(\mathbb{R}_+)$; at $y = 0$ it has a zero of multiplicity $k-1$ and at each $y \in Y$ it has a zero of multiplicity greater than or equal to 2. By Lemma 2.1 we have

$$N(p^{(k-2)}, [0, X_{(j)}]) \geq N(p, [0, X_{(j)}]) - (k-2) \geq 2\#(Y \cap [0, X_{(j)}]) + 1.$$

The function $p^{(k-2)}(y)$ is piecewise quadratic and therefore on each interval $[X_{(i)}, X_{(i+1)}]$ it may have at most two zeroes and on $[0, X_{(1)}]$ it has only one zero at zero. Thus

$$1 + 2(j-1) \geq N(p^{(k-2)}, [0, X_{(j)}]) \geq 2\#(Y \cap [0, X_{(j)}]) + 1,$$

or equivalently the set Y has no more than $j-1$ points on $[0, X_{(j)}]$. In particular, the set Y contains no more than n points. Let $Y_1 < \dots < Y_m$ be the points of Y . Then we have $Y_j > X_{(j)}$.

This implies that for any MLE \hat{f}_n the support of the corresponding mixing measure \hat{G}_n is a subset of Y and thus any MLE \hat{f}_n has the form

$$\hat{f}_n(x) = \sum_{j=1}^m a_j \frac{k(Y_j - x)_+^{k-1}}{Y_j^k},$$

where $a_j \geq 0$, $\sum a_j = 1$. \square

The last lemma proves uniqueness of the MLE.

Lemma 2.3. *The discrete mixing probability measure \hat{G}_n which defines an MLE is unique.*

Proof. Suppose there exist two different MLEs \hat{f}_n^1 and \hat{f}_n^2 . Then by Lemma 2.2 we have

$$\hat{f}_n^l(x) = \sum_{j=1}^m a_j^l \frac{k(Y_j - x)_+^{k-1}}{Y_j^k},$$

and therefore the function $q(x)$ defined by

$$q(x) = \hat{f}_n^1(x) - \hat{f}_n^2(x) = \sum_{j=1}^m s_j \frac{k(Y_j - x)_+^{k-1}}{Y_j^k},$$

where $s_j = a_j^1 - a_j^2$, admits at least n zeroes $X_{(i)}$ and not all s_j are equal to zero. Therefore, by Lemma 2.1,

$$N(q^{(k-2)}, [0, X_i)) \geq N(q, [0, X_i)) - (k-2) \geq i-1-(k-2).$$

The function $q^{(k-2)}$ is piecewise linear with knots at Y_j . Therefore $N(q^{(k-2)}, [0, X_i)) \leq \#\{Y \cap [0, X_{(i)})\} + 1$ and

$$\#\{Y \cap [0, X_{(i)})\} \geq i-k.$$

This implies

$$Y_{i-k} < X_{(i)},$$

and by Lemma 2.2

$$Y_{i-k} < X_{(i)} < Y_i.$$

Now, by Corollary 1 of [11] we have $\det((Y_i - X_{(j)})_+^k)_{i,j=1}^m > 0$. Thus the vectors $\Gamma(Y_i)$, $1 \leq i \leq m$, are independent and all $s_i = 0$. This contradiction completes the proof. \square

Note that from Lemma 2.2 it follows that any MLE is defined by some mixing probability measure on a finite set $Y = \{Y_j\}$. Lemma 2.3 shows that the MLE is unique and hence the vectors $\Gamma(Y_j)$ are independent. In particular, there exists a subset of order statistics $X_{(i_1)} < \dots < X_{(i_m)}$ such that

$$\det((Y_j - X_{(i_l)})_+^{k-1})_{j,l=1}^m > 0.$$

Therefore, by Corollary 1 of [11] we obtain $Y_{j-k} < X_{(i_j)} < Y_j$.

The equality (2.1) also implies that for any MLE, $Y_m > X_{(n)}$, since otherwise the last coordinate of all vectors $\Gamma(Y_j)$ is equal to zero and $\hat{f}_n(X_{(n)}) = 0$. This completes the proof of Theorem 1.1.

ACKNOWLEDGMENT

The author would like to thank his advisor, Professor Jon A. Wellner, for introducing this problem to him and for helpful comments and suggestions.

REFERENCES

- [1] Balabdaoui, F. (2004). *Nonparametric estimation of k -monotone density: A new asymptotic distribution theory*. Ph.D. thesis, University of Washington, Department of Statistics.
- [2] Balabdaoui, F. and Wellner, J. (2010). Estimation of a k -monotone density: characterizations, consistency and minimax lower bounds. *Statist. Neerl.* **64**, 45–70.
- [3] Balabdaoui, F. and Wellner, J. A. (2007). Estimation of a k -monotone density: limit distribution theory and the spline connection. *Ann. Statist.* **35**, 2536–2564. <http://dx.doi.org.offcampus.lib.washington.edu/10.1214/09053607000000262>. MR2382657 (2009b:62077)
- [4] Groeneboom, P., Jongbloed, G. and Wellner, J. A. (2001). Estimation of a convex function: characterizations and asymptotic theory. *Ann. Statist.* **29**, 1653–1698. MR1891742 (2003a:62047)
- [5] Jewell, N. P. (1982). Mixtures of exponential distributions. *Ann. Statist.* **10**, 479–484. MR653523 (83f:62057)
- [6] Lindsay, B. G. (1983). The geometry of mixture likelihoods: a general theory. *Ann. Statist.* **11**, 86–94. MR684866 (85m:62008a)
- [7] Lindsay, B. G. (1983). The geometry of mixture likelihoods. II. The exponential family. *Ann. Statist.* **11**, 783–792. MR707929 (85m:62008b)
- [8] Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications*, vol. 5. NSF-CBMS Regional Conference Series in Probability and Statistics, IMS, Hayward, CA.
- [9] Lindsay, B. G. and Roeder, K. (1993). Uniqueness of estimation and identifiability in mixture models. *Canad. J. Statist.* **21**, 139–147. <http://dx.doi.org.offcampus.lib.washington.edu/10.2307/3315807>. MR1234757 (94k:62062)
- [10] Rockafellar, R. T. (1970). *Convex analysis*. Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J. MR0274683 (43:445)
- [11] Schoenberg, I. J. and Whitney, A. (1953). On Pólya frequency functions. III. The positivity of translation determinants with an application to the interpolation problem by spline curves. *Trans. Amer. Math. Soc.* **74**, 246–259. MR0053177 (14:732g)

DEPARTMENT OF STATISTICS, UNIVERSITY OF WASHINGTON, BOX 354322, SEATTLE, WASHINGTON 98195-4322

E-mail address: arseni@stat.washington.edu