Theoretical properties of the log-concave maximum likelihood estiamtor of a multidimensional density

Richard Samworth

(joint work with Madeleine Cule)

Shape-constrained density estimation in general, and log-concave density estimation in particular, have received a great deal of attention in the statistical literature recently – see, for example, Walther (2002), Cule, Samworth and Stewart (2007), Dümbgen, Hüsler and Rufibach (2007), Pal, Woodroofe and Meyer (2007), Dümbgen and Rufibach (2009), Balabdaoui, Rufibach and Wellner (2009). The following theorem helps to explain this interest:

Theorem 1 (Cule, Samworth and Stewart (2007)). *Let $X_1, \ldots, X_n$ be independent with density f_0 in $\mathbb{R}^d$, and suppose that $n \geq d+1$. Then, with probability one, there exists a unique log-concave maximum likelihood estimator $\hat{f}_n$ of f_0.*

Thus, even though the class of log-concave densities is infinite-dimensional (and contains many well-known and commonly-used families of densities), there exists a fully automatic density estimator within this class, with no smoothing parameters to choose. To understand the theoretical properties of this estimator, we begin by noting that when it is known that a sequence of densities is log-concave, convergence in weak sesnses in fact implies convergence in much stronger senses:

Proposition 2 (Cule and Samworth (2009)). *Let (f_n) be a sequence of log-concave densities on $\mathbb{R}^d$ with $f_n \xrightarrow{d} f$ for some density f. Then:*

(a) *f is log-concave*
(b) *$f_n \to f$, almost everywhere*
(c) *Let $a_0 > 0$ and $b_0 \in \mathbb{R}$ be such that $f(x) \leq e^{-a_0\|x\|+b_0}$. Then for every $a < a_0$, we have $\int_{\mathbb{R}^d} e^{a\|x\|}|f_n(x) - f(x)|\,dx \to 0$ and, if f is continuous, $\sup_{x\in\mathbb{R}^d} e^{a\|x\|}|f_n(x) - f(x)| \to 0$.*

In order to state our main result, we first require appropriate bounds on the behaviour of the log-concave maximum likelihood estimator, as illustrated in the following result. Write E for the support of f_0.

Lemma 3 (Cule and Samworth (2009)). *Suppose that $\int_{\mathbb{R}^d} \|x\| f_0(x)\,dx < \infty$.*

(a) *There exists a constant $C > 0$ such that, with probability one,*

$$\limsup_{n\to\infty} \sup_{x\in\mathbb{R}^d} \hat{f}_n(x) \leq C.$$

(b) *Let S be a compact subset of* $\mathrm{int}(E)$. *There exists a constant $c > 0$ such that, with probability one,*

$$\liminf_{n\to\infty} \inf_{x\in\mathrm{conv}\, S} \hat{f}_n(x) \geq c.$$

Our main result establishes establishes desirable performance properties of $\hat{f}_n$. Recall that the Kullback–Leibler divergence of a density f from f_0 is given by

$d_{KL}(f_0, f) = \int_{\mathbb{R}^d} f_0 \log(f_0/f)$. Jensen's inequality shows that the Kullback–Leibler divergence is non-negative, and equal to zero if and only if $f = f_0$ (almost everywhere). Thus when f_0 is log-concave, Theorem 4 below shows that the log-concave maximum likelihood estimator $\hat{f}_n$ is strongly consistent in certain exponentially weighted total variation metrics. Convergence in exponentially weighted supremum norms also follows if f_0 is continuous.

In the case where the model is misspecified, i.e. f_0 is not log-concave, we prove that the existence and uniqueness of a log-concave density f^* that minimises the Kullback–Leibler divergence from f_0. Moreover, we show that the log-concave maximum likelihood estimator $\hat{f}_n$ converges in the same senses as in the previous paragraph to f^*. We write $\log_+ x = \max(\log x, 0)$.

Theorem 4 (Cule and Samworth (2009))**.** *Let f_0 be any density on $\mathbb{R}^d$ with $\int_{\mathbb{R}^d} \|x\| f_0(x)\, dx < \infty$, $\int_{\mathbb{R}^d} f_0 \log_+ f_0 < \infty$ and $\mathrm{int}(E) \neq \emptyset$. There exists a log-concave density f^*, unique almost everywhere, that minimises the Kullback–Leibler divergence of f from f_0 over all log-concave densities f. Taking $a_0 > 0$ and $b_0 \in \mathbb{R}$ such that $f^*(x) \leq e^{-a_0\|x\|+b_0}$, we have for any $a < a_0$ that*

$$\int_{\mathbb{R}^d} e^{a\|x\|} |\hat{f}_n(x) - f^*(x)|\, dx \stackrel{a.s.}{\to} 0,$$

and, if f^* is continuous, $\sup_{x \in \mathbb{R}^d} e^{a\|x\|} |\hat{f}_n(x) - f^*(x)| \stackrel{a.s.}{\to} 0$.

References

Balabdaoui, F., Rufibach, K. and Wellner, J. A. (2009), Limit distribution theory for maximum likelihood estimation of a log-concave density, *Ann. Statist.*, **37**, 1299–1331.

Cule, M. L. and Samworth, R. J. (2009), *Theoretical properties of the log-concave maximum likelihood estimator of a multidimensional density.* Preprint. Available at `http://arxiv.org/pdf/0908.4400`.

Cule, M. L., Samworth, R. J. and Stewart, M. I. (2007), *Maximum likelihood estimation of a multidimensional log-concave density*, Submitted. Available at `http://www.statslab.cam.ac.uk/~rjs57/Research.html`.

Dümbgen, L., Hüsler, A. and Rufibach, K. (2007), Active Set and EM Algorithms for Log-Concave Densities Based on Complete and Censored Data. Preprint. `http://arxiv.org/abs/0707.4643`.

Dümbgen, L. and Rufibach, K. (2009), Maximum likelihood estimation of a log-concave density and its distribution function: basic properties and uniform consistency, *Bernoulli*, **15**, 40–68.

Pal, J., Woodroofe, M. and Meyer, M. (2007), Estimating a Polya frequency function. In *Complex datasets and Inverse problems: Tomography, Networks and Beyond* (eds. R. Liu, W. Strawderman, C.-H. Zhang), pp. 239–249. IMS Lecture Notes - Monograph Series **54**.

Walther, G. (2002), Detecting the presence of mixing with multiscale maximum likelihood, *J. Amer. Statist. Assoc.*, **97**, 508–513.