Log-concave density estimation for independent, identically distributed observations

Kaspar Rufibach

September 28, 2006

1 Introduction

This note shortly describes how a log-concave density can be estimated and what algorithms are used in the package `logcondens`. It is by far not intended to give full reference about the subject, more details can be found in Rufibach (2006a,b), Dümbgen and Rufibach (2006), and Dümbgen, Hüsler, and Rufibach (2006).

2 Log-concave density estimation

A probability density f on the real line is called log–concave if it may be written as

$$f(x) \ = \ \exp \varphi(x)$$

for some concave function $\varphi : \mathbb{R} \to [-\infty, \infty)$. Let $X_1, X_2, \ldots, X_n$ be independent random variables with such a log–concave probability density. If $x_1 < x_2 < \cdots < x_n$ are the corresponding order statistics, the normalized log–likelihood function is given by

$$\ell(\varphi) \ := \ n^{-1} \sum_{i=1}^{n} \varphi(x_i).$$

It may happen that due to rounding errors one observes $\widetilde{X}_i$ in place of X_i. In that case, let $x_1 < x_2 < \cdots < x_m$ be the different elements of $\{\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_n\}$ and define $w_j := n^{-1}\#\{i : \widetilde{X}_i = x_j\}$. Then an appropriate surrogate for the normalized log–likelihood is

$$\ell(\varphi) \ := \ \sum_{i=1}^{m} w_i \varphi(x_i). \tag{1}$$

In what follows we consider the functional (1) for arbitrary given points $x_1 < x_2 < \cdots < x_m$ and probability weights $w_1, w_2, \ldots, w_m > 0$, i.e. $\sum_{i=1}^{m} w_i = 1$. Suppose that we want to maximize $\ell(\varphi)$ over all functions

φ that are concave and induce a probability density. This is equivalent to maximizing

$$L(\varphi) \;:=\; \sum_{i=1}^{m} w_i \varphi(x_i) - \int \exp \varphi(x)\, dx$$

over all concave functions φ. From Theorem 3.2.1 in Rufibach (2006a) we know that

$$\widehat{\varphi}_m \quad := \quad \operatorname*{arg\,min}_{\varphi \text{ concave}} L(\varphi)$$

is piecewise linear on $[X_1, X_m]$ with knots only in $\widehat{\mathcal{S}}_m := \{x_1, \ldots, x_m\}$ and $\widehat{\varphi}_m = -\infty$ on $\mathbb{R} \setminus [x_1, x_m]$. Therefore, we can restrict our attention to functions of this type and rewrite the log-likelihood function as

$$L(\varphi) \;=\; L(\boldsymbol{\varphi}) := \sum_{j=1}^{m} w_j \varphi_j - \sum_{k=2}^{m} \Delta x_{k+1} J(\varphi_k, \varphi_{k+1})$$

with

$$J(r,s) \;:=\; \int_0^1 \exp\bigl((1-t)r + ts\bigr)\, dt$$

for arbitrary $r, s, \in \mathbb{R}$ where we tacitly introduced the following notation: Any continuous concave function that is piecewise linear with knots only in $\widehat{\mathcal{S}}_m$ can be identified with the vector $\boldsymbol{\varphi} := (\varphi(x_j))_{j=1}^m = (\varphi_j)_{j=1}^m \in \mathbb{R}^m$. Likewise, any vector $\boldsymbol{\varphi} \in \mathbb{R}^m$ defines a function φ via

$$\varphi(x) \;:=\; \Bigl(1 - \frac{x - x_k}{\Delta x_{k+1}}\Bigr) \varphi_k + \frac{x - x_k}{\Delta x_{k+1}} \varphi_{k+1} \quad \text{for } x \in [x_k, x_{k+1}], 1 \le k < m,$$

where $\Delta x_k := x_k - x_{k-1}$. The maximization problem can now be reformulated to

$$\max_{\boldsymbol{\varphi} \in \mathbb{R}^m} L(\boldsymbol{\varphi})$$

under the constraints

$$\frac{\Delta \varphi_j}{\Delta x_j} - \frac{\Delta \varphi_{j-1}}{\Delta x_{j-1}} \;\le\; 0 \quad \text{for } j = 3, \ldots, m\,.$$

Standard optimization techniques are now suitable to find this maximum.

3 An active set algorithm

This algorithm maximizes L by alternately going into the ordinary Newton direction (only as far as the constraints allow) and altering the set of constraints. To find the Newton direction, the gradient and the Hesse matrix of L are needed and are given in Dümbgen, Hüsler, and Rufibach (2006). In the latter paper, the general framework for active set algorithms is accounted for in Section 3.

This algorithm is implemented in the function `activeSetLogCon`.

4 An iterative convex minorant algorithm

To be able to apply such an algorithm, the function L needs to be reparametrized, see Rufibach (2006a,b). Define

$$\boldsymbol{\eta} \;=\; \Big(\varphi_1, \Big(\frac{\Delta\varphi_i}{\Delta x_i}\Big)_{i=2}^{m}\Big),$$

the vector of successive slopes of the piecewise linear concave function φ. The back-parametrization is then

$$\boldsymbol{\varphi} \;=\; \Big(\eta_1, \eta_1 + \Big(\sum_{j=2}^{i} \Delta x_i \eta_i\Big)_{i=2}^{m}\Big).$$

Inserting this new parametrization $\boldsymbol{\eta}$ into L, the reparametrized log-likelihood function L becomes

$$\begin{aligned} L(\boldsymbol{\eta}) \;&:=\; L(\boldsymbol{\eta}(\boldsymbol{\varphi})) \\ &=\; \eta_1 \sum_{i=1}^{m} w_i + \sum_{i=2}^{m} \eta_i \sum_{i=k}^{m} w_i \Delta x_i - e^{\eta_1} \sum_{i=2}^{m} \exp\Big(\sum_{k=2}^{i-1} \Delta x_k \eta_k\Big) \frac{\exp(\Delta x_i \eta_i) - 1}{\eta_i}\,. \end{aligned} \tag{2}$$

The point of the reparametrization is, that the optimization problem now writes

$$\max_{\boldsymbol{\eta} \in \mathbb{R}^m} L(\boldsymbol{\eta})$$

under the constraints

$$\eta_{i-1} \;\geq\; \eta_i \quad \text{for } i = 3, \ldots, m\,.$$

Now approximate (2) quadratically around a given $\boldsymbol{\eta}_o \in \mathbb{R}^m$ by the quadratic function $\tilde{L}$:

$$\begin{aligned} \tilde{L}(\boldsymbol{\eta}) \;&=\; \tilde{L}(\boldsymbol{\eta}|\boldsymbol{\eta}_o) \\ &=\; L(\boldsymbol{\eta}_o) + \nabla_{\boldsymbol{\eta}} L(\boldsymbol{\eta}_o)'(\boldsymbol{\eta} - \boldsymbol{\eta}_o) + 2^{-1}(\boldsymbol{\eta} - \boldsymbol{\eta}_o)'\boldsymbol{W}(\boldsymbol{\eta}_o)(\boldsymbol{\eta} - \boldsymbol{\eta}_o) \end{aligned}$$

where $\nabla_{\boldsymbol{\eta}} L$ is the gradient of L and $\boldsymbol{D}$ some positive definite matrix, which we choose to be the diagonal matrix that equals the Hesse on the diagonal. For ease of notation, introduce $\boldsymbol{g} := \nabla_{\boldsymbol{\eta}} L(\boldsymbol{\eta}_o)$ and $\boldsymbol{d} = \operatorname{diag}(\boldsymbol{D}(\boldsymbol{\eta}_o))$. Then rewrite $\tilde{L}(\boldsymbol{\eta})$ as

$$\begin{aligned} \tilde{L}(\boldsymbol{\eta}) \;&=\; \tilde{L}(\boldsymbol{\eta}_o) + \sum_{i=1}^{m} g_i(\eta_i - \eta_{o,i}) + 2^{-1} \sum_{i=1}^{m} d_i(\eta_i - \eta_{o,i})^2 \\ &=\; \tilde{L}(\boldsymbol{\eta}_o) - 2^{-1} \sum_{i=1}^{m} (g_i/w_i)^2 + 2^{-1} \sum_{i=1}^{m} d_i\Big(\eta_i - (\eta_{o,i} - g_i/d_i)\Big)^2 \end{aligned}$$

and the maximization problem to solve becomes

$$\max_{\eta_2 \geq \cdots \geq \eta_m} \sum_{i=1}^{m} d_i \Big(\eta_i - (\eta_{o,i} - g_i / d_i) \Big)^2 .$$

But this is exactly what a (weighted) pool-adjacent-violaters algorithm delivers.

Using this, one can set up an iterative algorithm by computing the Newton step as usually, but then applying the above procedure to make this step fulfilling the constraints, and this does only increase the likelihood. Supplemented by a robustification procedure and an Hermite interpolation as first described in Dümbgen, Freitag and Jongbloed (2006), the algorithm implemented as the function `icmaLogCon` comes out.

References

DÜMBGEN, L., FREITAG S., JONGBLOED, G. (2006). Estimating a Unimodal Distribution from Interval-Censored Data. *J. Amer. Statist. Assoc., to appear.*

L. DÜMBGEN, A. HÜSLER AND K. RUFIBACH (2006). Active Set and EM Algorithms for Log–Concave Densities Based on Complete and Censored Data. Preprint

L. DÜMBGEN AND K. RUFIBACH (2006). Maximum likelihood estimation of a log–concave density and its distribution function: basic properties and uniform consistency. Preprint

K. RUFIBACH (2006a). *Log–Concave Density Estimation and Bump Hunting for I.I.D. Observations.* Dissertation, Universities of Bern and Göttingen

K. RUFIBACH (2006b). Computing maximum likelihood estimators of a log–concave density function. *J. Statist. Comp. Sim.*, to appear