Submitted to Bernoulli

Supplement to: A Bayesian nonparametric approach to log-concave density estimation

ESTER MARIUCCI[1], KOLYAN RAY[2] and BOTOND SZABÓ[2]

[1] *Institut für Mathematik, Universität Potsdam, Karl-Liebknecht-Str. 24-25, Potsdam 14476, Germany. E-mail:* mariucci@uni-potsdam.de

[2] *Mathematical Institute, Leiden University, Niels Bohrweg 1, Leiden 2333 CA, Netherlands. E-mail:* k.m.ray@math.leidenuniv.nl*;* b.t.szabo@math.leidenuniv.nl*,*

We collect here the proofs of Lemmas 1-4 and some additional simulations.

1. Technical lemmas

Lemma 1. *For every $\varepsilon > 0$, there exist positive constants L_0, C, c, n_0, depending only on ε, and positive universal constants $D, d > 0$, such that for all $L \geq L_0$ and $n \geq n_0$,*

$$\sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P_{g_0}^n \left(h(\hat{g}_n, g_0) \geq L n^{-2/5} \right) \leq C \exp\left(-c n^{1/(4+2\varepsilon)} \right) + D \exp\left(-d L^2 n^{1/5} \right),$$

where $\hat{g}_n$ denotes the log-concave maximum likelihood estimator based on an i.i.d. sample $Z_1, ..., Z_n$ from g_0.

Proof. It is shown in the proof of Theorem 5 of Kim and Samworth [3], p. 2772, that for $\eta = \eta(\varepsilon) \in (0,1)$ to be defined below and $L \geq L_0(\eta) = L_0(\varepsilon)$,

$$\sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P_{g_0}^n \left(\left\{ h(\hat{g}_n, g_0) \geq L n^{-2/5} \right\} \cap \left\{ \hat{g}_n \in \widetilde{\mathcal{F}}^{1,\eta} \right\} \right) \leq 2^{15/2} \exp\left(-\frac{L^2 n^{1/5}}{2^{28}} \right), \quad (1)$$

so that it remains only to control $\sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P_{g_0}^n (\hat{g}_n \notin \widetilde{\mathcal{F}}^{1,\eta})$. Lemma 6 of Kim and Samworth [3] shows that this quantity is $O(n^{-1})$ as $n \to \infty$; we essentially follow their proof, suitably sharpening the probability bounds in the case $d = 1$. Then

$$\begin{aligned}\sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P_{g_0}^n (\hat{g}_n \notin \widetilde{\mathcal{F}}^{1,\eta}) \leq & \sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P_{g_0}^n (|\mu_{\hat{g}_n}| > 1) + \sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P_{g_0}^n (\sigma_{\hat{g}_n}^2 > 1 + \eta) \\ & + \sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P_{g_0}^n (\sigma_{\hat{g}_n}^2 < 1 - \eta).\end{aligned} \quad (2)$$

By Lemma 13 of [2], there exist universal constants $\alpha_0, \beta_0 > 0$ such that for all $x \in \mathbb{R}$,

$$\sup_{g \in \overline{\mathcal{F}}^{0,1}} g(x) \leq e^{\beta_0 - \alpha_0 |x|}.$$

imsart-bj ver. 2014/10/16 file: LCBayes_supp.tex date: November 29, 2018

It is shown in [3], p. 2774, that the first term in (2) is bounded by $2\alpha_0^{-1}e^{\beta_0-\alpha_0\sqrt{n}}$. For the second term, by Remark 2.3 of Dümbgen et al. [1], one has $\sigma^2_{\hat{g}_n} \leq \widetilde{\sigma}^2_n := n^{-1}\sum_{i=1}^n (Z_i - \bar{Z}_n)^2 \leq n^{-1}\sum_{i=1}^n Z_i^2$, where $\widetilde{\sigma}^2_n$ denotes the sample variance and $\bar{Z}_n = n^{-1}\sum_{i=1}^n Z_i$ the sample mean. Letting $C_0 = C_0(\alpha_0, \beta_0, 2)$ denote the constant in Lemma 5 below, we can apply that lemma to bound the second term in (2) by

$$\sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P^n_{g_0}\left(\frac{1}{n}\sum_{i=1}^n Z_i^2 > 1+\eta\right) \leq \exp(-\sqrt{\eta/C_0}\,n^{1/4})$$

for $n \geq \max(16C_0^2/\eta, e^2)$.

Consider now the third term in (2). For $\varepsilon > 0$, let $\overline{\mathcal{P}}^{1/10,1/2}_\varepsilon$ denote the class of probability distributions on $\mathbb{R}$ such that $\mu_P = \int x dP(x)$ and $\sigma_P^2 = \int (x-\mu_P)^2 dP(x)$ satisfy $|\mu_P| \leq 1/10$ and $1/2 \leq \sigma_P^2 \leq 3/2$ and

$$\int |x|^{2+\varepsilon} dP(x) \leq 4e^{\beta_0}\alpha_0^{-3-\varepsilon}\Gamma(3+\varepsilon) =: \tau_\varepsilon,$$

where $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$. This is exactly the same as the class $\mathcal{P}^{1/10,1/2}$ considered in Lemma 6 of [3], except that we have replaced the 4^{th}-moment condition with a $(2+\varepsilon)$-moment condition. Following the rest of the proof of Lemma 6 of [3] (noting that $\sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} \int |x|^{2+\varepsilon} g_0(x)dx \leq \tau_\varepsilon/2$ and that uniform integrability of $\{Y^2_{n_k} : k \in \mathbb{N}\}$ in that proof follows from the $(2+\varepsilon)$-moment condition), one can similarly conclude that for some $\eta = \eta\big(\alpha_0, \beta_0, \varepsilon\big) \in (0,1)$,

$$\sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P^n_{g_0}(\sigma^2_{\hat{g}_n} < 1-\eta) \leq \sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P^n_{g_0}\left(\mathbb{P}_n \notin \overline{\mathcal{P}}^{1/10,1/2}_\varepsilon\right),$$

where $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{Z_i}$ is the empirical measure and δ_x is the Dirac measure at x. This last probability can be bounded by

$$\sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P^n_{g_0}(|\bar{Z}_n| > 1/10) + \sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P^n_{g_0}(|\widetilde{\sigma}^2_n - 1| > 1/2)$$
$$+ \sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P^n_{g_0}\left(\int |x|^{2+\varepsilon}(d\mathbb{P}_n(x) - g_0(x)dx) > \tau_\varepsilon/2\right).$$

Using similar arguments to those used previously, the first two terms can be bounded by $Ce^{-c\sqrt{n}}$ and $Ce^{-cn^{1/4}}$, respectively, where the constants depend only on α_0 and β_0. For the last term, using Lemma 5 with $C_0 = C_0(\alpha_0, \beta_0, 2+\varepsilon)$ the constant in that lemma,

$$\sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P^n_{g_0}\left(\frac{1}{n}\sum_{i=1}^n \left(|Z_i|^{2+\varepsilon} - \mathbb{E}_{g_0}|Z_i|^{2+\varepsilon}\right) > \frac{\tau_\varepsilon}{2}\right) \leq \exp\left(-\left(\frac{\tau_\varepsilon}{2C_0}\right)^{1/(2+\varepsilon)} n^{\frac{1}{4+2\varepsilon}}\right)$$

for $n \geq \max(4(2+\varepsilon)^{4+2\varepsilon}C_0^2/\tau_\varepsilon^2, e^{2+\varepsilon})$. In conclusion, we have shown that for any $\varepsilon > 0$ there exists $\eta = \eta(\varepsilon) \in (0,1)$ such that

$$\sup_{g_0 \in \overline{\mathcal{F}}^{0,1}} P^n_{g_0}\big(\hat{g}_n \notin \widetilde{\mathcal{F}}^{1,\eta}\big) \leq C(\varepsilon)\exp\Big(-c(\varepsilon)n^{\frac{1}{4+2\varepsilon}}\Big)$$

for all $n \geq n_0(\varepsilon)$, where $C(\varepsilon)$, $c(\varepsilon)$ and $n_0(\varepsilon)$ are positive constants. Together with (1), this establishes the result. □

We recall the piecewise linear approximation of a continuous concave function w on a compact interval $[a,b]$ considered in the main article. For any partition $a = x_0 < x_1 < \cdots < x_m = b$ of $[a,b]$, let $\widetilde{w}_m$ denote the piecewise linear approximation of w given by

$$\widetilde{w}_m(x) := \sum_{i=2}^{m} \left(\frac{x - x_{i-1}^*}{x_i^* - x_{i-1}^*} \frac{1}{x_i - x_{i-1}} \theta_i + \frac{x_i^* - x}{x_i^* - x_{i-1}^*} \frac{1}{x_{i-1} - x_{i-2}} \theta_{i-1} \right) \mathbb{1}_{(x_{i-1}^*, x_i^*]}(x), \quad (3)$$

where $\theta_i := \int_{x_{i-1}}^{x_i} w(s)ds$ and $x_i^* := \frac{x_i + x_{i-1}}{2}$. On $[a, x_1^*]$ and $(x_m^*, b]$, the function is defined by linearly extending the piecewise linear function defined above, that is

$$\begin{aligned} \widetilde{w}_m(a) &:= \frac{1}{x_2^* - x_1^*} \Big(\frac{x_2^* - a}{x_1 - a} \theta_1 - \frac{x_1^* - a}{x_2 - x_1} \theta_2 \Big), \\ \widetilde{w}_m(b) &:= \frac{1}{x_m^* - x_{m-1}^*} \Big(\frac{b - x_{m-1}^*}{b - x_{m-1}} \theta_m - \frac{b - x_m^*}{x_{m-1} - x_{m-2}} \theta_{m-1} \Big). \end{aligned} \quad (4)$$

The function $\widetilde{w}_m$ takes value $\widetilde{w}_m(x_i^*) = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} w(s)ds$ at the midpoint $x_i^* = \frac{x_i + x_{i-1}}{2}$ of the interval $[x_{i-1}, x_i]$ and interpolates linearly in between.

Lemma 2. *Let $w : [a,b] \to \mathbb{R}$ be a continuous concave function, where $-\infty < a < b < \infty$. For any partition $a = x_0 < x_1 < \cdots < x_m = b$ of $[a,b]$, let $\widetilde{w}_m$ denote the piecewise linear approximation of w defined in (3) and (4). Then $\widetilde{w}_m$ is a concave function.*

Proof. By rescaling, we may without loss of generality assume that $[a,b] = [0,1]$. Note that $\widetilde{w}_m$ is concave if and only if

$$\widetilde{w}_m(x_i^*) \geq \frac{x_i^* - x_{i-1}^*}{x_{i+1}^* - x_{i-1}^*} \widetilde{w}_m(x_{i+1}^*) + \frac{x_{i+1}^* - x_i^*}{x_{i+1}^* - x_{i-1}^*} \widetilde{w}_m(x_{i-1}^*) \quad (5)$$

for $i = 2, \ldots, m-1$. Indeed, since $\widetilde{w}_m$ is piecewise linear, it is concave if and only if at every point where the derivative is discontinuous (i.e. a knot), the left derivative is greater than or equal to the right derivative. The above statement follows since the derivative of $\widetilde{w}_m$ is discontinuous (at most) at the points x_i^*, $i = 2, \ldots, m-1$, where the desired inequality is:

$$\frac{\widetilde{w}_m(x_{i+1}^*) - \widetilde{w}_m(x_i^*)}{x_{i+1}^* - x_i^*} \leq \frac{\widetilde{w}_m(x_i^*) - \widetilde{w}_m(x_{i-1}^*)}{x_i^* - x_{i-1}^*},$$

which is equivalent to (5).

To see that (5) holds if w is concave, we argue by contradiction and suppose that there exists i such that $\widetilde{w}_m(x_i^*) < \frac{x_i^* - x_{i-1}^*}{x_{i+1}^* - x_{i-1}^*} \widetilde{w}_m(x_{i+1}^*) + \frac{x_{i+1}^* - x_i^*}{x_{i+1}^* - x_{i-1}^*} \widetilde{w}_m(x_{i-1}^*)$. Consider the linear function l,

$$l(x) := \frac{x - x_{i-1}^*}{x_{i+1}^* - x_{i-1}^*} \widetilde{w}_m(x_{i+1}^*) + \frac{x_{i+1}^* - x}{x_{i+1}^* - x_{i-1}^*} \widetilde{w}_m(x_{i-1}^*).$$

In particular, we have that $\widetilde{w}_m(x^*_{i-1}) - l(x^*_{i-1}) = \widetilde{w}_m(x^*_{i+1}) - l(x^*_{i+1}) = 0$ and $\widetilde{w}_m(x^*_i) < l(x^*_i)$. We further denote $g := w - l$ and observe that

$$\tilde{g}_m(x^*_i) = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} \big(w(s) - l(s)\big) ds = \widetilde{w}_m(x^*_i) - \int_{x_{i-1}}^{x_i} l(s) ds = \widetilde{w}_m(x^*_i) - l(x^*_i).$$

It follows that $\tilde{g}_m(x) = \widetilde{w}_m(x) - l(x)$ for all $x \in [0,1]$ and hence by the mean value theorem,

$$\tilde{g}_m(x^*_i) = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} g(s) ds = g(\xi_i)$$

for some $\xi_i \in [x_{i-1}, x_i]$. One can similarly prove the existence of two points, $\xi_{i-1} \in [x_{i-2}, x_{i-1}]$ and $\xi_{i+1} \in [x_i, x_{i+1}]$, such that $\tilde{g}_m(x^*_{i-1}) = g(\xi_{i-1})$ and $\tilde{g}_m(x^*_{i+1}) = g(\xi_{i+1})$. Using the above results, we deduce the existence of three points $\xi_{i-1} < \xi_i < \xi_{i+1}$ such that $g(\xi_{i-1}) = 0 = g(\xi_{i+1})$ and $g(\xi_i) < 0$, which is a contradiction since g is concave by the concavity of w and l. □

Lemma 3. *Let $w : [a,b] \to \mathbb{R}$ be a continuous concave function with $w'_+(a) - w'_-(b) \leq M$ and where $-\infty < a < b < \infty$. Then there exists a partition $a = x_0 < x_1 < \cdots < x_m = b$ of $[a,b]$ with $\min_{i=1,\dots,m}(x_i - x_{i-1}) \geq (b-a)(2m)^{-2}$ and such that*

$$\sup_{x \in [a,b]} |w(x) - \widetilde{w}_m(x)| \leq C \frac{M(b-a)}{m^2},$$

where $\widetilde{w}_m$ is the piecewise linear approximation of w defined in (3) and (4) and $C > 0$ is a universal constant (i.e. not depending on a, b, m).

Proof. By translation we may without loss of generality take $a = 0$. Recall that since w is a continuous concave function, it has left and right derivatives at every point $x \in [0,b]$. Define $\Delta w'(x) = w'_-(x) - w'_+(x)$. For every $r \geq 1$, let $\mathcal{P}_1 := \{x_{i,1} := \frac{ib}{r}, i = 0, \dots, r\}$ be the uniform partition of $[0,b]$ and let $\tilde{x}_{i,2}$, $i = 1, \dots, r_2$, be the points such that $\Delta w'(\tilde{x}_{i,2}) \geq M/r$, setting $r_2 = 0$ if no such point exists. By concavity of w,

$$M \geq w'_+(0) - w'_-(b) \geq \sum_{i=1}^{r_2} \Delta w'(\tilde{x}_{i,2}) \geq \frac{M r_2}{r}, \tag{6}$$

so that $r_2 \leq r$.

Consider a new partition $\mathcal{P}_2 := \{x_{0,2} < \cdots < x_{r'_2,2}\}$ of $[0,b]$, consisting of the points $\{x_{i,1}\} \cup \{\tilde{x}_{i,2}\} \cup \{\tilde{x}_{i,2} - br^{-2}\} \cup \{\tilde{x}_{i,2} + br^{-2}\}$ written in increasing order. Note that $r'_2 \leq r + 3r_2$. Colour in red all the points of the form $\tilde{x}_{i,2}$ and $\tilde{x}_{i,2} - br^{-2}$, so that each red point is the left endpoint of an interval of length at most br^{-2}. This colouring will be used to keep track of points that have a close neighbour on the right.

We next refine the partition $\mathcal{P}_2$ by adding the point y between $x_{i,2}$ and $x_{i+1,2}$

$$y := \sup\left\{ x > x_{i,2} : w'_+(x_{i,2}) - w'_-(x) \leq \frac{2M}{r} \right\}$$

if

$$w'_+(x_{i,2}) - w'_-(x_{i+1,2}) > \frac{2M}{r}.$$

Denote by r_3 the total number of points y added in this manner to the sequence. We further add the points $y - br^{-2}$, $y + br^{-2}$ and colour in red all points of the form y and $y - br^{-2}$, similarly to the previous case. Repeating this procedure results in a new partition that separates intervals where the derivative decreases by at most $2M/r$. Denote by $\mathcal{P}_3 := \{0 = x_{0,3} < x_{1,3} < \cdots < x_{r'_3,3} = b\}$ this new partition. We now show that $r'_3 \le 7r$.

Let y by any point added in the way just described. Suppose by contradiction that $w'_+(x_{i,2}) - w'_-(y) < M/r$. By definition, we know that for all $x > y$, $w'_+(x_{i,2}) - w'_-(x) > 2M/r$. Subtracting the two inequalities gives $w'_-(y) - w'_-(x) > M/r$. However, since the right derivative of a concave function is right continuous, taking the limit $x \to y^+$ (and restricting to the points x where w is differentiable) yields $\Delta w'(y) \ge M/r$. This is a contradiction however, because if this were the case, y would already belong to $\mathcal{P}_2$. Since $w'_+(x_{i,2}) - w'_-(y) \ge M/r$, using a similar argument to (6) gives $r_3 \le r$ so that $r'_3 \le 7r$.

Finally, if the function w is not differentiable at the point $x^*_{i,3} = \frac{x_{i,3}+x_{i-1,3}}{2}$, we split $[x_{i-1,3}, x_{i,3}]$ into two parts in such a way that w is differentiable at the midpoints of both new intervals and each interval has size at least $(x_{i,3} - x_{i-1,3})/3$. We add the points separating the new intervals to the previous partition, thereby obtaining $\mathcal{P}_4 := \{0 = x_{0,4} < x_{1,4} < \cdots < x_{\nu,4} = b\}$. The cardinality of $\mathcal{P}_4$, satisfies $\nu + 1 \le 14r + 1$. We now create a new partition $\mathcal{P}$ with polynomially separated points using the following algorithm.

1. Set $\mathcal{P} = \mathcal{P}_4$, keeping track of all the points coloured red. Set $\tilde{x} = x_{0,4}$.
2. If $b - \tilde{x} \le br^{-2}$, remove all points in $\mathcal{P}$ strictly between $\tilde{x}$ and b skip to Step 4.
3. Set $y = \inf\{t \in \mathcal{P} : t > \tilde{x} + br^{-2}\}$. Remove all elements of $\mathcal{P}$ between $\tilde{x}$ and y. If at least one element was removed, add to $\mathcal{P}$ the point $s = \tilde{x} + br^{-2} + \varepsilon$ for some $0 < \varepsilon < br^{-2} \wedge (y - \tilde{x} - br^{-2})$ such that w is differentiable at $(s+y)/2$ and $(s+\tilde{x})/2$. Colour $\tilde{x}$ red to mark that $s - \tilde{x} < 2br^{-2}$. Set $\tilde{x} := s$. If no point was removed from $\mathcal{P}$, set $\tilde{x} := y$. Go to Step 2.
4. If $\tilde{x} = b$ then stop. Otherwise set $y = \max\{t \in \mathcal{P} : t < \tilde{x}\}$ and remove $\tilde{x}$ from $\mathcal{P}$. If $b - y > 2br^{-2}$, add the point $s := b - br^{-2} - \varepsilon$ to $\mathcal{P}$ and colour it red, where $0 < \varepsilon < (b - y - 2br^{-2}) \wedge br^{-2}$ is such that w is differentiable at $(y + s)/2$. If $b - y \le 2br^{-2}$, add the point $s := (y + b)/2$ and colour both y and s red.

Relabel the final partition $\mathcal{P} := \{0 = x_0 < x_1 < \cdots < x_m = b\}$ and note that

$$r \le m \le \nu \le 14r + 1.$$

By construction $\min_{i=0,\dots,m-1}(x_{i+1} - x_i) \ge \frac{1}{2}br^{-2} \ge \frac{1}{2}bm^{-2}$ and

$$x_{i+1} - x_i \le \begin{cases} 2C_0^2 bm^{-2} & \text{if } x_i \text{ is coloured red (with } C_0 = 15), \\ C_0 bm^{-1} & \text{otherwise,} \end{cases}$$

since if x_i is coloured red, $x_{i+1} - x_i \le 2br^{-2} \le 2C_0^2 bm^{-2}$.

We now show that $\|w-\widetilde{w}_m\|_\infty = O(bm^{-2})$. If x_{i-1} is red, then by the mean value theorem, there exists $\xi_i \in J_i := [x_{i-1}, x_i]$ such that $\widetilde{w}_m(x_i^*) = w(\xi_i)$. Using the Lipschitz continuity of w and $\widetilde{w}_m$,

$$|w(x) - \widetilde{w}_m(x)| \leq 2C_0^2 \frac{bM}{m^2}, \quad \forall x \in J_i := [x_{i-1}, x_i].$$

If x_{i-1} is not red, Taylor expanding w at the points $x_i^* := \frac{x_i + x_{i-1}}{2}$ (at which w is differentiable by the construction of $\mathcal{P}$) gives

$$w(x) = w(x_i^*) + w'(x_i^*)(x - x_i^*) + R_i(x), \quad x \in J_i. \tag{7}$$

Due to the construction of $\mathcal{P}$,

$$\begin{aligned} |R_i(x)| &= \Big| w(x) - w(x_i^*) - w'(x_i^*)(x - x_i^*) \Big| \\ &= \big| (w'(\xi_i) - w'(x_i^*))(x - x_i^*) \big|, \end{aligned}$$

where $w'(\xi_i)$ stands here for some value in the interval $[w'_+(\xi_i), w'_-(\xi_i)]$ for some point $\xi_i \in J_i$. We then deduce that $|R_i(x)| \leq \frac{2M}{r}\frac{C_0 b}{2m} \leq C_0^2 bM/m^2$.

Since $\widetilde{w}_m$ is piecewise linear, we can write

$$\widetilde{w}_m(x) = \widetilde{w}_m(x_i^*) + \widetilde{w}'_m(x_i^*)(x - x_i^*), \qquad x \in J_i,$$

where $\widetilde{w}'_m$ denotes the left or right derivative of $\widetilde{w}_m$ at x_i^*, depending on whether $x < x_i^*$ or $x > x_i^*$. We now show that $|\widetilde{w}'_m(x_i^*) - w'(x_i^*)| \leq \max\{w'_+(x_{i-1}) - w'(x_i^*), w'(x_i^*) - w'_-(x_{i+1})\}$ for $i = 1, \ldots, m-1$. Consider the case of right derivatives (the same argument also works for left derivatives). Using the definition of $\widetilde{w}_m$ and that $\theta_i = \int_{J_i} w(s)ds$,

$$\begin{aligned} \widetilde{w}'_{m,+}(x_i^*) &= \frac{\widetilde{w}_m(x_{i+1}^*) - \widetilde{w}_m(x_i^*)}{x_{i+1}^* - x_i^*} = \frac{1}{x_{i+1}^* - x_i^*}\left(\frac{\theta_{i+1}}{x_{i+1} - x_i} - \frac{\theta_i}{x_i - x_{i-1}}\right) \\ &= \frac{1}{(x_i - x_{i-1})(x_{i+1}^* - x_i^*)} \int_{x_{i-1}}^{x_i} \left[w\Big(\frac{x_{i+1} - x_i}{x_i - x_{i-1}} t + \frac{x_i^2 - x_{i+1}x_{i-1}}{x_i - x_{i-1}} \Big) - w(t) \right] dt \end{aligned}$$

and

$$\int_{x_{i-1}}^{x_i} \left[\frac{x_{i+1} - x_i}{x_i - x_{i-1}} t + \frac{x_i^2 - x_{i+1}x_{i-1}}{x_i - x_{i-1}} - t \right] dt = (x_{i+1}^* - x_i^*)(x_i - x_{i-1}).$$

By the continuity and concavity of w, $(v-u)w'_-(x_{i+1}) \leq w(v) - w(u) \leq (v-u)w'_+(x_{i-1})$ for any $x_{i-1} \leq u \leq v \leq x_{i+1}$. Combining all of the above yields

$$w'_-(x_{i+1}) \leq \widetilde{w}'_{m,+}(x_i^*) \leq w'_+(x_{i-1}). \tag{8}$$

We remark that $\max\{w'_+(x_{i-1}) - w'(x_i^*), w'(x_i^*) - w'_-(x_{i+1})\} \leq \frac{5M}{r}$. Indeed, since $x_i - x_{i-1} > C_0^2 b/m^2$, the point x_i is not equal to $x'_{j,4}$ for any j and hence both $\Delta w'(x_i) < M/r$

and $w'_+(x_i) - w'_-(x_{i+1}) \leq 2M/r$ hold. Together with $w'_+(x_{i-1}) - w'(x_i^*) \leq 2M/r$ and $w'(x_i^*) - w'_-(x_i) \leq 2M/r$, this verifies the preceding statement. Then

$$w'(x_i^*) - w'_-(x_{i+1}) \leq w'(x_i^*) - w'_-(x_i) + w'_-(x_i) - w'_+(x_i) + w'_+(x_i) - w'_-(x_{i+1}) \leq \frac{5M}{r}.$$

We hence deduce that $|w'(x_i^*) - \widetilde{w}'_m(x_i^*)| \leq 5MC_0m^{-1}$. Finally, using (7) and the fact that $\int_{J_i}(x - x_i^*)dx = 0$,

$$\big|w(x_i^*) - \widetilde{w}_m(x_i^*)\big| = \frac{1}{x_i - x_{i-1}}\left|\int_{J_i}\big(w(x_i^*) - w(x)\big)dx\right| \leq \sup_{x\in J_i}|R_i(x)| \leq \frac{C_0^2bM}{m^2}.$$

Collecting together all the pieces, we have that for any $x \in [x_{i-1}, x_i]$,

$$|w(x) - \widetilde{w}_m(x)| \leq |w(x_i^*) - \widetilde{w}_m(x_i^*)| + |x - x_i^*||\widetilde{w}'_m(x_i^*) - w'(x_i^*)| + |R_i(x)| \leq \frac{9}{2}C_0^2Mbm^{-2}.$$

□

Lemma 4. *Any piecewise linear concave function $w : [a, b] \to \mathbb{R}$ with N knots $\{z_1, ..., z_N\}$ can be written in the form*

$$w(x) = \gamma_1\sum_{i=1}^{N}\frac{z_i \wedge (x - a)}{z_i}p_i - \gamma_2(x - a) + \gamma_3,$$

with parameters $0 \leq \gamma_1 \leq (w'_+(a) - w'_-(b))(b - a)$, $|\gamma_2| \leq |w'_-(b)|$, $\gamma_3 \in \mathbb{R}$, $\sum_{i=1}^N p_i = 1$ and $p_i \geq 0$ for $i = 1, \dots, N$.

Proof. The left derivative of w is a step function $g : (a, b] \mapsto \mathbb{R}$ with $g(a+\varepsilon) = w'_+(a)$ for sufficiently small $\varepsilon > 0$ and $g(b) = w'_-(b)$. By shifting this function vertically by $-w'_-(b)$, we arrive at a non-negative, bounded, monotone decreasing step function, which can therefore be written as a monotone decreasing probability density times a normalizing constant γ_1. It is easy to see that $\gamma_1 \leq (b - a)(w'_+(a) - w'_-(b))$. The step function can therefore be represented as

$$g(x) = \gamma_1\int_{x-a}^{b-a}\frac{1}{z}dP_N(z) + w'_-(b), \qquad x \in [a, b],$$

with P_N an atomic probability measure with N atoms on $[0, b-a]$ and γ_1 the normalizing constant. Integrating the step function g yields

$$\bar{w}_N(x) = \gamma_1\int_0^{b-a}\frac{z \wedge (x - a)}{z}dP_N(z) + w'_-(b)(x - a) + C,$$

which is equal to w for an appropriately chosen constant $C > 0$. □

Lemma 5. *Let $Z_1,...,Z_n$ be i.i.d. random variables from a density $f \in \mathcal{F}_{\alpha,\beta}$, where $\alpha > 0$ and $\beta \in \mathbb{R}$. Then for any $r \geq 1$, $t \geq r^r$ and $n \geq e^r$,*

$$P_f^n\left(\frac{1}{\sqrt{n}}\left|\sum_{i=1}^n (|Z_i|^r - \mathbb{E}|Z_i|^r)\right| \geq C_0(\alpha,\beta,r)t\right) \leq \exp(-t^{1/r}).$$

Proof. For notational convenience, write $\lambda = 1/r \in (0,1]$. Let $x_\lambda = (1/\lambda)^{1/\lambda}$ and define

$$\Psi_\lambda(x) = \begin{cases} e^{x^\lambda} & x \geq x_\lambda, \\ \tau_\lambda x & x < x_\lambda, \end{cases}$$

where $\tau_\lambda = \Psi_\lambda(x_\lambda)/x_\lambda = (\lambda e)^{1/\lambda}$. This defines a Young function, that is a convex, increasing function $\Psi_\lambda : \mathbb{R}^+ \to \mathbb{R}^+$ with $\Psi_\lambda(0) = 0$. Denote the corresponding Orlicz norm $\|X\|_{\Psi_\lambda} := \inf\{a > 0 : \mathbb{E}[\Psi_\lambda(|X|/a)] \leq 1\}$. Note that the density function g_λ of $|Z_1|^{1/\lambda}$ satisfies $g_\lambda(x) = \lambda x^{-(1-\lambda)}\big(f(x^\lambda) + f(-x^\lambda)\big)\mathbb{1}_{\{x\geq 0\}} \leq \lambda x^{-(1-\lambda)} e^{\beta - \alpha x^\lambda}\mathbb{1}_{\{x \geq 0\}}$. Then for fixed $a > \alpha^{-1/\lambda}$,

$$\begin{aligned}\mathbb{E}\Psi_\lambda(|Z_1|^{1/\lambda}/a) &\leq \frac{\lambda\tau_\lambda e^\beta}{a}\int_0^{x_\lambda} u^\lambda e^{-\alpha u^\lambda} du + \lambda e^\beta \int_{x_\lambda}^\infty u^{-(1-\lambda)} e^{-(\alpha - a^{-\lambda})u^\lambda} du \\ &= K_0(a,\alpha,\beta,\lambda) < \infty.\end{aligned}$$

If $K_0 = K_0(a,\alpha,\beta,\lambda) > 1$, it follows by convexity that $\||Z_1|^{1/\lambda}\|_{\Psi_\lambda} \leq aK_0$.

By Theorem 6.21 of Ledoux and Talagrand [4], there exists a constant K_λ such that

$$\left\|\sum_{i=1}^n (|Z_i|^{\frac{1}{\lambda}} - \mathbb{E}|Z_i|^{\frac{1}{\lambda}})\right\|_{\Psi_\lambda} \leq K_\lambda\left(\left\|\sum_{i=1}^n (|Z_i|^{\frac{1}{\lambda}} - \mathbb{E}|Z_i|^{\frac{1}{\lambda}})\right\|_1 + \left\|\max_{1\leq i\leq n}\big||Z_i|^{\frac{1}{\lambda}} - \mathbb{E}|Z_i|^{\frac{1}{\lambda}}\big|\right\|_{\Psi_\lambda}\right).$$

The first-term on the right-hand side can be bounded by the $\|\cdot\|_2$-norm of the same quantity, which equals the square root of $n\mathbb{E}(|Z_i|^{\frac{1}{\lambda}} - \mathbb{E}|Z_i|^{\frac{1}{\lambda}})^2 \leq n\mathbb{E}|Z_i|^{\frac{2}{\lambda}}$. For any $\delta \in (0,1]$,

$$\mathbb{E}|Z_1|^{\frac{1}{\delta}} = \int_0^\infty x g_\delta(x)dx \leq 2\delta e^\beta \int_0^\infty x^\delta e^{-\alpha x^\delta} dx = \frac{2e^\beta}{\alpha^{1+\frac{1}{\delta}}}\int_0^\infty y^{\frac{1}{\delta}} e^{-y} dy = \frac{2e^\beta}{\alpha^{1+\frac{1}{\delta}}}\Gamma(1+1/\delta).$$

Since this is finite for any $\delta \in (0,1]$, we can bound the $\|\cdot\|_1$-norm above by $C(\alpha,\beta,\lambda)\sqrt{n}$.

Note that for any random variable X, $\|X - \mathbb{E}X\|_{\Psi_\lambda} \leq 2\|X\|_{\Psi_\lambda}$. Indeed, setting $a = \|X\|_{\Psi_\lambda}$, since Ψ_λ is convex and increasing,

$$\mathbb{E}\Psi_\lambda\left(\frac{|X - \mathbb{E}X|}{2a}\right) \leq \frac{1}{2}\mathbb{E}\Psi_\lambda\left(\frac{|X|}{a}\right) + \frac{1}{2}\Psi_\lambda\left(\frac{\mathbb{E}|X|}{a}\right) \leq \mathbb{E}\Psi_\lambda\left(\frac{|X|}{a}\right) \leq 1.$$

Using this and Lemma 2.2.2 of van der Vaart and Wellner [5],

$$\big\|\max_{1\leq i\leq n}\big||Z_i|^{\frac{1}{\lambda}} - \mathbb{E}|Z_i|^{\frac{1}{\lambda}}\big|\big\|_{\Psi_\lambda} \leq K(\Psi_\lambda)\Psi_\lambda^{-1}(n)\max_{1\leq i\leq n}\big\||Z_i|^{\frac{1}{\lambda}}\big\|_{\Psi_\lambda} \leq K(\alpha,\beta,\lambda)(\log n)^{\frac{1}{\lambda}},$$

so that we have shown

$$\left\| \sum_{i=1}^{n} (|Z_i|^{\frac{1}{\lambda}} - \mathbb{E}|Z_i|^{\frac{1}{\lambda}}) \right\|_{\Psi_\lambda} \leq C(\alpha, \beta, \lambda)\sqrt{n}.$$

By Markov's inequality, for any random variable X, $\mathbb{P}(|X| \geq x\|X\|_{\Psi_\lambda}) = \mathbb{P}(\Psi_\lambda(|X|/\|X\|_{\Psi_\lambda}) \geq \Psi_\lambda(x)) \leq 1/\Psi_\lambda(x)$. Applying this to the above sum completes the proof. □

2. Additional simulations

We firstly provide some additional figures for the empirical Bayes posterior of the mode. We consider the $Beta(2, 5)$ and $Gamma(2, 1)$ distributions and take i.i.d. samples of size ranging from $n = 50$ to $n = 20000$ in both cases. As in the Gaussian example, we run the Gibbs sampler for 20000 iterations of which half were considered as burn in and discarded. The resulting marginal posteriors for the mode based on the empirical Bayes posterior are displayed in Figures 1 and 2.

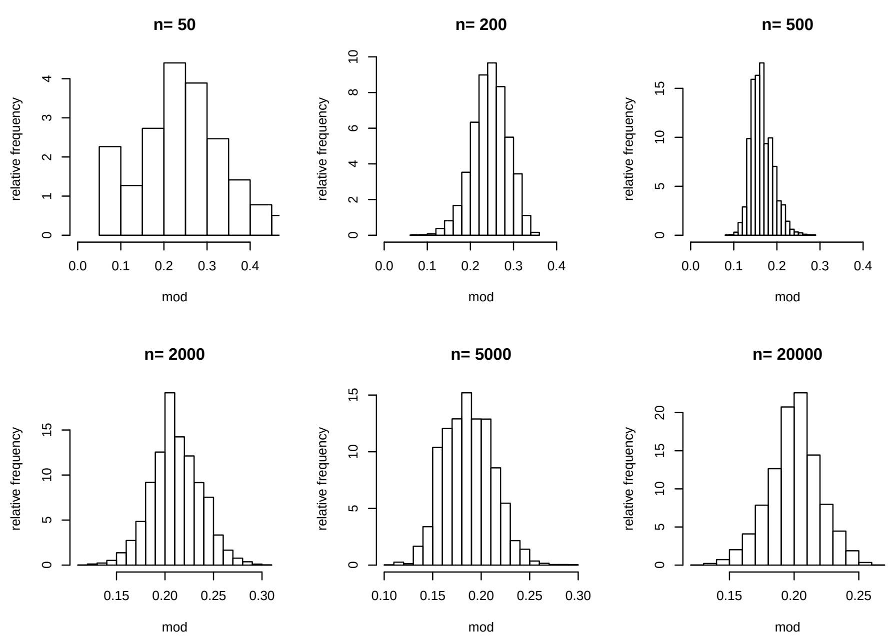

Figure 1. The marginal posterior distribution for the mode based on the empirical Bayes posterior for the $Beta(2, 5)$ distribution with increasing sample size from left to right and top to bottom, ranging from $n = 50$ to $n = 20000$.

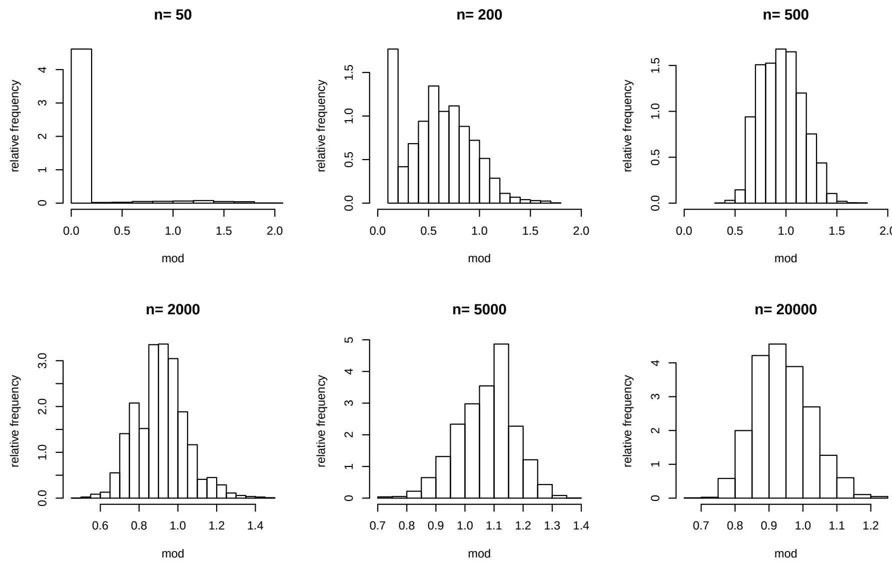

Figure 2. The marginal posterior distribution for the mode based on the empirical Bayes posterior for the $Gamma(2, 1)$ distribution with increasing sample sizes from left to right and top to bottom, ranging from $n = 50$ to $n = 20000$.

Next we provide a brief simulation study to demonstrate the (possible) applicability of our procedure for cluster analysis. Here we consider only the simple case where the underlying density is a mixture of two log-concave densities. We have modified our Bayesian procedure to accommodate mixtures of log-concave densities in a straightforward way, again using a Gibbs sampler. We consider various combinations of log-concave densities including two Gaussians, two Laplace distributions, a Laplace and a Gamma distribution and two Beta distributions. We consider two sample sizes, $n = 100$ and $n = 500$, see Figures 3 and 4, respectively. In all pictures we have plotted the posterior mean (solid blue), 95% pointwise credible sets (dashed blue) and the underlying density (solid red). One can see that in all cases our procedure provides reasonable estimators and seemingly reliable uncertainty quantification.

References

[1] Dümbgen, L., Samworth, R., and Schuhmacher, D. Approximation by log-concave distributions, with applications to regression. *Ann. Statist. 39*, 2 (2011), 702–730.

[2] Fresen, D. A multivariate Gnedenko law of large numbers. *Ann. Probab. 41*, 5 (2013), 3051–3080.

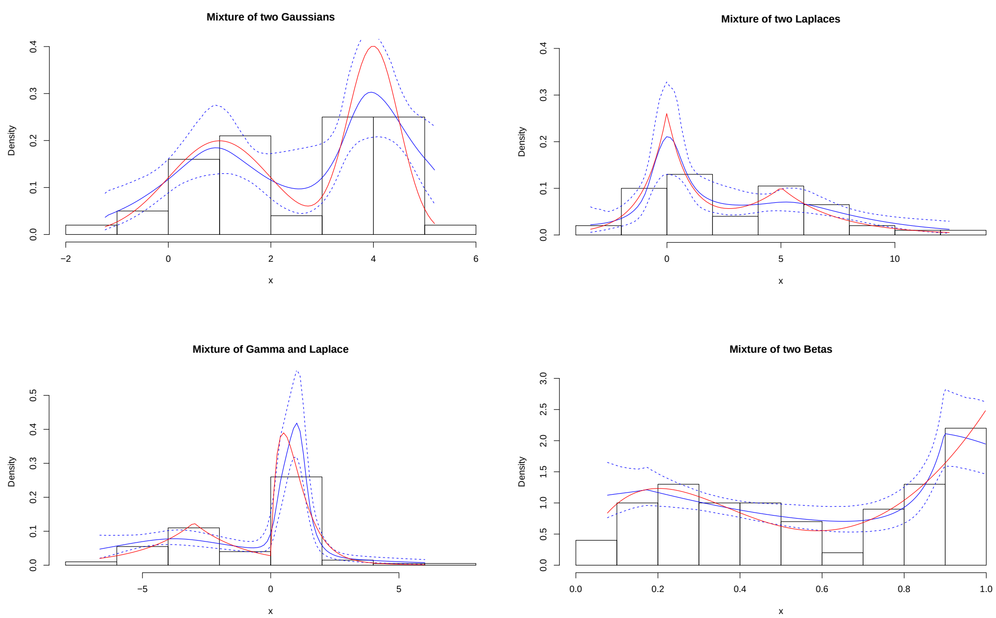

Figure 3. Empirical Bayes posterior distribution for the mixture of two log-concave densities for sample size $n = 100$. Top left corner: $0.5N(1,1) + 0.5N(4,0.5)$, top right corner: $0.5 * Laplace(0,1) + 0.5 * Laplace(5,2.5)$, bottom left corner: $0.5 * Gamma(2,2) + 0.5 * Laplace(-3,2)$, bottom right corner: $0.5 * Beta(2,5) + 0.5 * Beta(5,1)$.

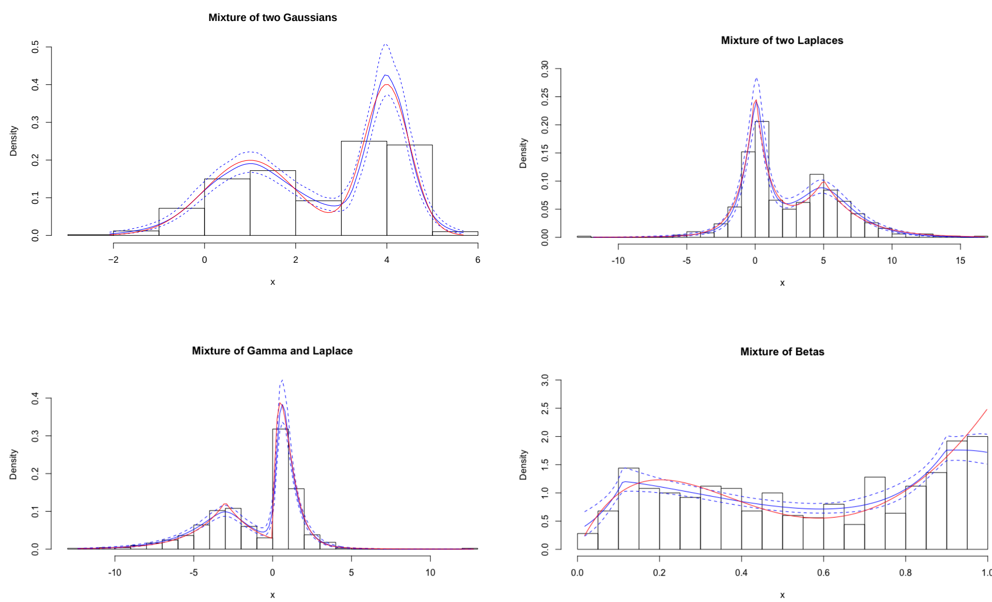

Figure 4. Empirical Bayes posterior distribution for the mixture of two log-concave densities for sample size $n = 500$. Top left corner: $0.5N(1,1) + 0.5N(4,0.5)$, top right corner: $0.5 * Laplace(0,1) + 0.5 * Laplace(5,2.5)$, bottom left corner: $0.5 * Gamma(2,2) + 0.5 * Laplace(-3,2)$, bottom right corner: $0.5 * Beta(2,5) + 0.5 * Beta(5,1)$.

[3] Kim, A. K. H., and Samworth, R. J. Global rates of convergence in log-concave density estimation. *Ann. Statist. 44*, 6 (2016), 2756–2779.

[4] Ledoux, M., and Talagrand, M. *Probability in Banach spaces.* Classics in Mathematics. Springer-Verlag, Berlin, 2011. Isoperimetry and processes, Reprint of the 1991 edition.

[5] van der Vaart, A. W., and Wellner, J. A. *Weak convergence and empirical processes.* Springer Series in Statistics. Springer-Verlag, New York, 1996. With applications to statistics.