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

ESTER MARIUCCI¹, KOLYAN RAY² and BOTOND SZABÓ²

¹Institut für Mathematik, Universität Potsdam, Karl-Liebknecht-Str. 24-25, Potsdam 14476, Germany. E-mail: mariucci@uni-potsdam.de
²Mathematical 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 \( \varepsilon \), and positive universal constants \( D, d > 0 \), such that for all \( L \geq L_0 \) and \( n \geq n_0 \),

\[
\sup_{g_0 \in \mathcal{F}_0} P_n\left( h(\hat{g}_n, g_0) \geq Ln^{-2/5} \right) \leq C \exp \left( -cn^1/\left(4+2\varepsilon\right) \right) + D \exp \left( -dL^2n^{1/5} \right),
\]

where \( \hat{g}_n \) denotes the log-concave maximum likelihood estimator based on an i.i.d. sample \( Z_1, \ldots, Z_n \) 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 \( L \geq L_0(\eta) = L_0(\varepsilon) \),

\[
\sup_{g_0 \in \mathcal{F}_0} P_n\left( h(\hat{g}_n, g_0) \geq Ln^{-2/5} \right) \cap \{ \hat{g}_n \in \mathcal{F}_1, \eta \} \leq 2^{15/2} \exp \left( -\frac{L^2n^{1/5} \eta^2}{28} \right),
\]

so that it remains only to control \( \sup_{g_0 \in \mathcal{F}_0} P_n(\hat{g}_n \notin \mathcal{F}_1, \eta) \). 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

\[
\sup_{g_0 \in \mathcal{F}_0} P_n(\hat{g}_n \notin \mathcal{F}_1, \eta) \leq \sup_{g_0 \in \mathcal{F}_0} P_n(\|\hat{\mu}_n\| > 1) + \sup_{g_0 \in \mathcal{F}_0} P_n(\sigma^2_{\hat{g}_n} > 1 + \eta) + \sup_{g_0 \in \mathcal{F}_0} P_n(\sigma^2_{\hat{g}_n} < 1 - \eta).
\]

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

\[
\sup_{g \in \mathcal{F}} g(x) \leq e^{\beta_0 - \alpha_0|x|}.
\]
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 $\sigma^2_{\theta_0} \leq \eta: = n^{-1}\sum_{i=1}^{n}(Z_i - \bar{Z}_n)^2 \leq n^{-1}\sum_{i=1}^{n}Z_i^2$, where $\sigma^2_{\theta}$ 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

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

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

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

$$
\int |x|^{2+\epsilon}dP(x) \leq 4e^{\beta_0^{-1}3^{-3-\epsilon}} \Gamma(3 + \epsilon) =: \tau_c,
$$

where $\Gamma(t) = \int_0^\infty x^{-1} e^{-x}dx$. 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 4th-moment condition with a $(2 + \epsilon)$-moment condition. Following the rest of the proof of Lemma 6 of [3] (noting that $\sup_{g_0 \in \mathcal{F}_{\eta, 1}} \int |x|^{2+\epsilon}g_0(x)dx \leq \tau_c/2$ and that uniform integrability of $\{Y^n_k \in \mathcal{F}_{\eta, 1} \mid k \in \mathbb{N}\}$ in that proof follows from the $(2 + \epsilon)$-moment condition), one can similarly conclude that for some $\eta = \eta(\alpha_0, \beta_0, \epsilon) \in (0, 1)$,

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

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

$$
\sup_{g_0 \in \mathcal{F}_{\eta, 1}} P^n_{g_0}(|\hat{Z}_n| > 1/10) + \sup_{g_0 \in \mathcal{F}_{\eta, 1}} P^n_{g_0}(|\hat{\sigma}^2_{\hat{g}_n} - 1| > 1/2)
$$

$$
+ \sup_{g_0 \in \mathcal{F}_{\eta, 1}} P^n_{g_0} \left( \int |x|^{2+\epsilon}(d\mathbb{P}_n(x) - g_0(x)dx) > \tau_c/2 \right).
$$

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

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

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

$$
\sup_{g_0 \in \mathcal{F}_{\eta, 1}} P^n_{g_0} (\hat{\theta}_n \notin \mathcal{F}^{1, \eta}) \leq C(\epsilon) \exp \left( - c(\epsilon)n^{1+\epsilon} \right)
$$
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 \( \bar{w}_m \) denote the piecewise linear approximation of \( w \) given by

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

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

\[
\bar{w}_m(a) := \frac{1}{x_2^* - x_1^*} \left( \frac{x_2^* - a}{x_2^* - x_1^*} \theta_1 + \frac{x_1^* - a}{x_2^* - x_1} \theta_2 \right),
\]

\[
\bar{w}_m(b) := \frac{1}{x_m^* - x_m} \left( \frac{b - x_m^*}{b - x_m} \theta_m - \frac{x_m^* - b}{x_m - x_m^*} \theta_{m+1} \right).
\]

The function \( \bar{w}_m \) takes value \( \bar{w}_m(x_i^*) = \frac{1}{x_i^* - x_{i-1}} \int_{x_{i-1}}^{x_i} w(s) \, ds \) 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 \( \bar{w}_m \) denote the piecewise linear approximation of \( w \) defined in (3) and (4). Then \( \bar{w}_m \) is a concave function.

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

\[
\bar{w}_m(x_i^*) \geq \frac{x_i^* - x_{i-1}^*}{x_{i+1}^* - x_i^*} \bar{w}_m(x_{i+1}^*) + \frac{x_{i+1}^* - x_i^*}{x_{i+1}^* - x_{i-1}^*} \bar{w}_m(x_{i-1}^*)
\]  

for \( i = 2, \ldots, m - 1 \). Indeed, since \( \bar{w}_m \) 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 \( \bar{w}_m \) is discontinuous (at most) at the points \( x_i^*, i = 2, \ldots, m - 1 \), where the desired inequality is:

\[
\frac{\bar{w}_m(x_{i+1}^*) - \bar{w}_m(x_i^*)}{x_{i+1}^* - x_i^*} \leq \frac{\bar{w}_m(x_i^*) - \bar{w}_m(x_{i-1}^*)}{x_i^* - x_{i-1}^*},
\]

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 \( \bar{w}_m(x_i^*) < \frac{x_i^* - x_{i-1}^*}{x_{i+1}^* - x_i^*} \bar{w}_m(x_{i+1}^*) + \frac{x_{i+1}^* - x_i^*}{x_{i+1}^* - x_{i-1}^*} \bar{w}_m(x_{i-1}^*) \). Consider the linear function \( l \),

\[
l(x) := \frac{x - x_{i-1}^*}{x_{i+1}^* - x_i^*} \bar{w}_m(x_{i+1}^*) + \frac{x_{i+1}^* - x}{x_{i+1}^* - x_{i-1}^*} \bar{w}_m(x_{i-1}^*).
\]
In particular, we have that \( \tilde{w}_m(x^*_i-1) - l(x^*_i-1) = \tilde{w}_m(x^*_i+1) - l(x^*_i+1) = 0 \) and \( \tilde{w}_m(x^*_i) < l(x^*_i) \). We further denote \( g := w - l \) and observe that
\[
\tilde{g}_m(x^*_i) = \frac{1}{x_i - x_{i-1}} \int_{x_{i-1}}^{x_i} (w(s) - l(s))ds = \tilde{w}_m(x^*_i) - \int_{x_{i-1}}^{x_i} l(s)ds = \tilde{w}_m(x^*_i) - l(x^*_i).
\]
It follows that \( \tilde{g}_m(x) = \tilde{w}_m(x) - l(x) \) 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 \( l \).

\[\square\]

**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 \( \pi = x_0 < x_1 < \ldots < x_m = b \) of \([a,b]\) with \( \min_{i=1,\ldots,m} (x_i - x_{i-1}) \geq (b-a)(2m)^{-2} \) and such that
\[
\sup_{x \in [a,b]} |w(x) - \tilde{w}_m(x)| \leq C \frac{M(b-a)}{m^2},
\]
where \( \tilde{w}_m \) 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 \( \mathcal{P}_1 := \{x_{i,1} := \frac{r}{r+i}, i = 0,\ldots,r\} \) be the uniform partition of \([0,b]\) and let \( \tilde{x}_{i,2}, i = 1,\ldots,r_2 \), 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 \( w \),
\[
M \geq w'_-(0) - w'_+(b) \geq \sum_{i=1}^{r_2} \Delta w' (\tilde{x}_{i,2}) \geq \frac{Mr_2}{r},
\]
so that \( r_2 \leq r \).

Consider a new partition \( \mathcal{P}_2 := \{x_{0,2} < \cdots < x_{r_2+1,2}\} \) 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 \( \tilde{x}_{2,i} \) 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 \( x_{i+1,2} \)
\[
y := \sup \left\{ x > x_{i,2} : w'_-(x_{i,2}) - w'_+(x) \leq \frac{2M}{r} \right\}
\]
By construction $m \leq 3$. Since if both new intervals and each interval has size at least $(3M/r)$, we can separate the new intervals to the previous partition, thereby obtaining further add the points if this were the case, $y$ and restricting to the points $x$ where $w$ is differentiable yields $\Delta w'(y) \geq M/r$. This is a contradiction however, because if this were the case, $y$ would already belong to $|P|$. Since $w'_{+}(x_{i,2}) - w'_-(y) \geq M/r$, using a similar argument to (6) gives $r_3 \leq r$ so that $r_3 \leq 7r$.

Finally, if the function $w$ is not differentiable at the point $x_{1,3} = \frac{x_1 + x_2 + 2}{3}$, we split $[x_{1-1,3}, x_{1,3}]$ 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 $(x_{1,3} - x_{1-1,3})/3$. We add the points separating the new intervals to the previous partition, thereby obtaining $P_4 := \{0 = x_{0,4} < x_{1,4} < \cdots < x_{\nu,4} = b\}$. The cardinality of $P_4$ satisfies $\nu + 1 \leq 14r + 1$. We now create a new partition $P$ with polynomially separated points using the following algorithm.

1. Set $P = P_4$, keeping track of all the points coloured red. Set $\tilde{x} = x_{0,4}$.
2. If $b - \tilde{x} \leq br^{-2}$, remove all points in $P$ strictly between $\tilde{x}$ and $b$ skip to Step 4.
3. Set $y = \inf\{t \in P : t > \tilde{x} + br^{-2}\}$. Remove all elements of $P$ between $\tilde{x}$ and $y$. If at least one element was removed, add to $P$ the point $s = \tilde{x} + br^{-2} + \varepsilon$ for some $0 < \varepsilon < br^{-2}$ such that $w$ is differentiable at $(s + \tilde{x})/2$ 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 $\tilde{x}$, set $\tilde{x} := y$. Go to Step 2.
4. If $\tilde{x} = b$ then stop. Otherwise set $y = \max\{t \in P : t < \tilde{x}\}$ and remove $\tilde{x}$ from $P$. If $b - y > 2br^{-2}$, add the point $s := b - br^{-2} - \varepsilon$ to $P$ 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 $b - y \leq 2br^{-2}$, add the point $s := (y + b)/2$ and colour both $y$ and $s$ red.

Relabel the final partition $P := \{0 = x_0 < x_1 < \cdots < x_m = b\}$ and note that $r \leq m \leq \nu \leq 14r + 1$.

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

$$x_{i+1} - x_i \leq \begin{cases} 2C_0^2bm^{-2} & \text{if } x_i \text{ is coloured red (with } C_0 = 15), \\ C_0bm^{-1} & \text{otherwise,} \end{cases}$$

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


We now show that \( \|w - \bar{w}_m\|_{\infty} = O(bm^{-2}) \). 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 \( \bar{w}_m(x_i^*) = w(\xi_i) \). Using the Lipschitz continuity of \( w \) and \( \bar{w}_m \),

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

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 \( P \)) gives

\[
w(x) = w(x_i^*) + w'(x_i^*)(x - x_i^*) + R_i(x), \quad x \in J_i.
\]

Due to the construction of \( P \),

\[
|R_i(x)| = \left| w(x) - w(x_i^*) - w'(x_i^*)(x - x_i^*) \right| = \left| (w'(\xi_i) - w'(x_i^*)) (x - x_i^*) \right|,
\]

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 \( |R_i(x)| \leq 2M \frac{C_0 b}{m} \leq C_0^2 bM/m^2 \).

Since \( \bar{w}_m \) is piecewise linear, we can write

\[
\bar{w}_m(x) = \bar{w}_m(x_i^*) + \bar{w}'_m(x_i^*)(x - x_i^*), \quad x \in J_i,
\]

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

\[
\bar{w}_{m,+}'(x_i^*) = \frac{\bar{w}_m(x_{i+1}) - \bar{w}_m(x_i^*)}{x_{i+1} - x_i^*} = \frac{1}{x_{i+1} - x_i^*} \left( \frac{\theta_{i+1} - \theta_i}{x_{i+1} - x_i} - \frac{\theta_i}{x_i - x_i-1} \right)
\]

and

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

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

\[
w'_{\rightarrow}(x_{i+1}) \leq \bar{w}_{m,+}'(x_i^*) \leq w'_{\rightarrow}(x_i^*).
\]

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

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

We hence deduce that $|w'(x_i^*) - \bar{w}'_m(x_i^*)| \leq 5MC_0 m^{-1}$. Finally, using (7) and the fact that $\int_{J_i} (x - x_i^*) dx = 0$,

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

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

$$|w(x) - \bar{w}_m(x)| \leq |w(x_i^*) - \bar{w}_m(x_i^*)| + |x - x_i^*| |\bar{w}_m(x_i^*) - w'(x_i^*)| + |R_i(x)| \leq \frac{9}{2} C_0^2 M b m^{-2}.$$ 

\[ \square \]

**Lemma 4.** Any piecewise linear concave function $w : [a, b] \rightarrow \mathbb{R}$ 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 $i = 1, ..., N$.

**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 $\gamma_1$. It is easy to see that $\gamma_1 \leq (b - a)(w'_+(a) - w'_-(b))$. The step function can therefore be represented as

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

with $P_N$ an atomic probability measure with $N$ atoms on $[0, b-a]$ and $\gamma_1$ 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 $C > 0$. \[ \square \]
Lemma 5. Let $Z_1, \ldots, Z_n$ 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 $r \geq 1$, $t \geq r^\tau$ and $n \geq e^r$,

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

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

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

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

$$E\Psi_\lambda(|Z_1|^{1/\lambda}/a) \leq \frac{\lambda \tau_\lambda e^\beta}{a} \int_0^{x_\lambda} u^{\lambda} e^{-\alpha u^\lambda} du + \lambda e^\beta \int_{x_\lambda}^{\infty} u^{-(1-\lambda)} e^{-(\alpha - \alpha \lambda)u^\lambda} du$$

$$= K_0(a, \alpha, \beta, \lambda) < \infty.$$ 

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 $K_\lambda$ such that

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

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 $nE(|Z|^{1/\lambda} - E|Z|^{1/\lambda})^2 \leq nE|Z|^{1/\lambda}$. For any $\delta \in (0, 1]$,

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

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 $X$, $\|X - EX\|_{\Psi_\lambda} \leq 2\|X\|_{\Psi_\lambda}$. Indeed, setting $a = \|X\|_{\Psi_\lambda}$, since $\Psi_\lambda$ is convex and increasing,

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

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

$$\max_{1 \leq i \leq n} \left\| |Z_i|^{1/\lambda} - E|Z_i|^{1/\lambda} \right\|_{\Psi_\lambda} \leq K(\Psi_\lambda)\Psi_\lambda^{-1}(n) \max_{1 \leq i \leq n} \left\| |Z_i|^{1/\lambda} \right\|_{\Psi_\lambda} \leq K(\alpha, \beta, \lambda)(\log n)^{1/\lambda},$$
so that we have shown
\[
\sum_{i=1}^{n} \left| \frac{|Z_i|^\lambda - \mathbb{E}|Z_i|^\lambda}{\Psi_{\lambda}} \right| \leq C(\alpha, \beta, \lambda) \sqrt{n}.
\]

By Markov’s inequality, for any random variable \( X \), \( P(|X| \geq x \| X \| \Psi_{\lambda}) = P(\Psi_{\lambda}(\|X\| \Psi_{\lambda}) \geq \Psi_{\lambda}(x)) \leq 1/\Psi_{\lambda}(x) \). 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 \( \text{Beta}(2, 5) \) and \( \text{Gamma}(2, 1) \) 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](image-url)

**Figure 1.** The marginal posterior distribution for the mode based on the empirical Bayes posterior for the \( \text{Beta}(2, 5) \) distribution with increasing sample size 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


Figure 3. Empirical Bayes posterior distribution for the mixture of two log-concave densities for sample size $n = 100$. Top left corner: $0.5 N(1, 1) + 0.5 N(4, 0.5)$, top right corner: $0.5 \times \text{Laplace}(0, 1) + 0.5 \times \text{Laplace}(5, 2.5)$, bottom left corner: $0.5 \times \text{Gamma}(2, 2) + 0.5 \times \text{Laplace}(-3, 2)$, bottom right corner: $0.5 \times \text{Beta}(2, 5) + 0.5 \times \text{Beta}(5, 1)$. 
Figure 4. Empirical Bayes posterior distribution for the mixture of two log-concave densities for sample size $n = 500$. Top left corner: $0.5 N(1,1) + 0.5 N(4,0.5)$, top right corner: $0.5 \ast Laplace(0,1) + 0.5 \ast Laplace(5,2.5)$, bottom left corner: $0.5 \ast Gamma(2,2) + 0.5 \ast Laplace(-3,2)$, bottom right corner: $0.5 \ast Beta(2,5) + 0.5 \ast Beta(5,1)$. 
