A Proofs of the lemmas

A.1 Proof of Lemma 1

Because q dominates f, we have $\mathbb{E}[W(X)] = 1$. The first inequality is due to Jensen's inequality: $1 = \mathbb{E}[W(X)]^\eta \geq \mathbb{E}[W(X)^\eta]$. When $W(X)$ is not a constant, equality holds if and only if $\eta = 1$.

For the second inequality, write

$$\mathbb{V}\mathrm{ar}[W(X)^\eta] \leq \mathbb{V}\mathrm{ar}[W(X)^\eta] + (\mathbb{E}[W(X)^\eta] - 1)^2 = \mathbb{E}[(W(X)^\eta - 1)^2] \leq \mathbb{E}[(W(X) - 1)^2].$$

The first inequality has already been justified. The second inequality holds because $|w^\eta - 1| \leq |w - 1|$ for all $w \geq 0$. □

A.2 Proof of Lemma 2

Recall that the general definition of the Kullback-Leibler divergence is given by

$$\mathrm{KL}\left(f \| q\right) = \int f \log\left(\frac{f}{q}\right) + \int q - \int f \ .$$

Note that it extends the definition given in Lemma 2 to unnormalized densities. Let q be a probability density function and set $\tilde{q} = f^\eta q^{1-\eta}$. Then, using that $\log(u) \leq u - 1$ for all $u > 0$, we have that

$$\begin{aligned} KL\left(f \middle\| \frac{\tilde{q}}{\int \tilde{q}}\right) &= \int f \log\left(\frac{f}{\tilde{q}} \cdot \int \tilde{q}\right) \\ &= \int f \log\left(\frac{f}{\tilde{q}}\right) + \log\left(\int \tilde{q}\right) \\ &\leq \mathrm{KL}\left(f \| \tilde{q}\right) \ . \end{aligned} \tag{9}$$

Furthermore, by definition of $\tilde{q}$, it holds that

$$\begin{aligned} \mathrm{KL}\left(f \| \tilde{q}\right) &= \int \log\left(\frac{f}{f^\eta q^{1-\eta}}\right) f + \int f^\eta q^{1-\eta} - 1 \\ &= (1-\eta) \int \log(f/q) f + \int f^\eta q^{1-\eta} - 1 \\ &= (1-\eta) KL(f \| q) + \int f^\eta q^{1-\eta} - 1 \\ &\leq (1-\eta)\, \mathrm{KL}\left(f \| q\right) \ , \end{aligned}$$

where the last inequality results from Jensen's inequality applied to the convex function $u \mapsto u^\eta$:

$$\int f^\eta q^{1-\eta} = \int q \left(\frac{f}{q}\right)^\eta \leq \left(\int f\right)^\eta = 1 \ .$$

Combining with (9) and letting $(q_k^*)_{k \geq 1}$ be defined by (2) starting from an initial probability density function q_1, by recursion we have for all $n \in \mathbb{N}^\star$,

$$\mathrm{KL}\left(f \| q_{n+1}^*\right) \leq \mathrm{KL}\left(f \| q_1\right) \prod_{k=1}^{n} (1 - \eta_k).$$

By applying Pinsker's inequality, we finally obtain

$$\int |f - q_{n+1}^*| \leq \sqrt{2\, \mathrm{KL}\left(f \| q_1\right)} \prod_{k=1}^{n} (1 - \eta_k)^{1/2}.$$

□

Convergence rates obtained from Lemma 2. Assuming that $\mathrm{KL}\left(f\|q_1\right) < +\infty$ and noticing that

$$\log\left(\prod_{k=1}^{n}(1-\eta_k)^{1/2}\right) \leq -\frac{1}{2}\sum_{k=1}^{n}\eta_k,$$

we get the following convergence rates:

- taking $\eta_k = c/k$ with $0 < c < 1$ yields $\int |f - q^*_{n+1}| = O(n^{-c/2})$,
- taking $\eta_k = c/k^\beta$ with $0 < c < 1$ and $\beta \in [0, 1)$ yields $\int |f - q^*_{n+1}| = O(\exp(-Cn^{(1-\beta)})$, with $C = c/(2(1-\beta))$.

B Deriving (2) from an optimisation perspective

One way to approximate an unknown probability density is to formulate an optimisation problem over a certain space of distributions, as it is typically done in Variational Inference. The common choice in Variational Inference then often corresponds selecting the Kullback-Leibler divergence and to try to find

$$q^\star = \mathrm{arginf}_{q\in\mathcal{Q}} \ \mathrm{KL}\left(q\|f\right) \ ,$$

where $\mathcal{Q}$ is a valid set of probability densities on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $\mathcal{B}(\mathbb{R}^d)$ denoting the Borel σ-field of $\mathbb{R}^d$, and where $\mathrm{KL}\left(f\|q\right)$ stands for the Kullback-Leibler divergence between f and q, i.e $\mathrm{KL}\left(f\|q\right) = \int \log\left(f/q\right) f$.

Following the approach of [26], one way to solve this optimisation problem is to resort to the *Entropic Mirror Descent* algorithm applied to the objective function $q \mapsto \mathrm{KL}\left(q\|f\right)$. When $\mathcal{Q}$ corresponds to the set of probability density functions on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, this algorithm admits a closed-form solution and generates a sequence $(q_k)_{k\geq 1}$ in $\mathcal{Q}$ satisfying (2).

To see this, let us start with a preliminary result.

B.1 A preliminary result

Let ν be a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ and let h be a real-valued measurable function on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. For any probability measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, we define

$$\Psi(\mu) = \int h \mathrm{d}\mu + \int \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) \mathrm{d}\mu.$$

The minimum of the function Ψ over the set $\left\{\mu : \mu \preceq \nu, \int |h| \mathrm{d}\mu < \infty\right\}$ is attained for μ such that

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu} \propto \exp(-h) \ .$$

Proof. By applying Jensen's inequality to the convex function $u \mapsto \exp(-u)$, we obtain

$$\exp(-\Psi(\mu)) \leq \int \exp\left(-\left[h + \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\right]\right) \mathrm{d}\mu \ .$$

Thus, we have that

$$\Psi(\mu) \geq -\log\left(\int \exp\left(-h\right) \mathrm{d}\nu\right) \ ,$$

where the r.h.s does not depend on μ and equality is attained if and only if $h + \log\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right) \propto 1$ μ-almost surely, that is if we take

$$\frac{\mathrm{d}\mu}{\mathrm{d}\nu} \propto \exp(-h) \ , \quad \mu\text{-almost surely.}$$

□

Next, we rewrite (2) as an Entropic Mirror Descent step.

B.2 Seeing (2) as an Entropic Mirror Descent

For any $x \in \mathbb{R}^d$ probability density $q \in \mathcal{Q}$, we set $h_q(x) = \log(q(x)/f(x)) + 1$. Given a probability density q_k^* and $\eta_k > 0$, one iteration of the (Infinite-Dimensional) Entropic Mirror Descent algorithm applied to the objective function $q \mapsto \mathrm{KL}\left(q \| f\right)$ with a learning rate η_k corresponds to finding

$$q_{k+1}^* = \mathrm{argmin}_{q \in \mathcal{Q}} \ \eta_k \int h_{q_k^*}(x) q(x) \mathrm{d}x + \mathrm{KL}\left(q \| q_k^*\right) \ .$$

In this expression, which is called the proximal form of the Entropic Mirror Descent, the function $h_{q_k^*}$ plays the role of the gradient of $\mathrm{KL}\left(q \| f\right)$ w.r.t the probability density q_k^* (here, it corresponds to its Fréchet differential). Based on the previous paragraph, we deduce that if $\mathcal{Q}$ corresponds to the set of probability density functions on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, then

$$q_{k+1}^* = \frac{q_k^*(x) e^{-\eta_k h_{q_k^*}(x)}}{\int q_k^*(x') e^{-\eta_k h_{q_k^*}(x')} \mathrm{d}x'} \propto f^{\eta_k}(x) q_k^*(x)^{1-\eta_k} \ ,$$

that is, we recover (2).

B.3 Convergence of the algorithm

Under minimal assumptions, the convergence towards f can be established with a known convergence rate for an appropriate choice of learning policy $(\eta_k)_{k \geq 1}$.

Lemma 6. Let $(\eta_k)_{k \geq 1}$ be a sequence of positive learning rates and let $\mathcal{Q}$ be set of probability density functions on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. Let $q_1 \in \mathcal{Q}$ and let the sequence $(q_k^*)_{k \geq 1}$ be defined by (2). Assume that $x \mapsto h_q(x)$ is bounded by a positive constant L for all $x \in \mathbb{R}^d$ and $q \in \mathcal{Q}$. Then, for all $n \in \mathbb{N}^\star$, we have

$$\mathrm{KL}\left(\sum_{k=1}^n \frac{\eta_k q_k^*}{\sum_{k'=1}^n \eta_{k'}} \middle\| f \right) \leq \frac{\sum_{k=1}^n \eta_k^2 L^2/2}{\sum_{k=1}^n \eta_k} + \frac{\mathrm{KL}\left(f \| q_1\right)}{\sum_{k=1}^n \eta_k} \ .$$

In particular, taking $\eta_k = c_0/\sqrt{k}$ with $c_0 > 0$ yields an $O(\log(n)/\sqrt{n})$ convergence rate. If the total number of iterations n is known in advance, setting $\eta_k = c_0/\sqrt{n}$ for all $k = 1 \dots n$ with $c_0 > 0$ yields an $O(1/\sqrt{n})$ convergence rate.

The proof of this result can be adapted from [50, Theorem 4.2]. It is provided here for the sake of completeness.

Proof. For all $k \geq 1$, set $\Delta_k = \mathrm{KL}\left(q_k^* \| f\right)$. By convexity of the function $u \mapsto u \log u$, we have

$$\begin{aligned}
\Delta_k &= \int \log\left(\frac{q_k^*(x)}{f(x)}\right) q_k^*(x) \mathrm{d}x \\
&\leq \int \left[\log\left(\frac{q_k^*(x)}{f(x)}\right) + 1\right] (q_k^*(x) - f(x)) \mathrm{d}x \ , \\
&= \int h_{q_k^*}(x) (q_k^*(x) - f(x)) \mathrm{d}x \ .
\end{aligned}$$

Since the integral of any constant w.r.t $q_k^* - f$ is null, we deduce

$$\begin{aligned}
\eta_k \Delta_k &\leq \int \log\left(\frac{q_k^*(x)}{q_{k+1}^*(x)}\right) (q_k^*(x) - f(x)) \mathrm{d}x \\
&= \int \log\left(\frac{q_k^*(x)}{q_{k+1}^*(x)}\right) q_k^*(x) \mathrm{d}x - \int \log\left(\frac{q_k^*(x)}{q_{k+1}^*(x)}\right) f(x) \mathrm{d}x \\
&= \int \log\left(\frac{q_k^*(x)}{q_{k+1}^*(x)}\right) (q_k^*(x) - q_{k+1}^*(x)) \mathrm{d}x - \mathrm{KL}\left(q_{k+1}^* \| q_k^*\right) \\
&\quad + \left[\mathrm{KL}\left(f \| q_k^*\right) - \mathrm{KL}\left(f \| q_{k+1}^*\right)\right]
\end{aligned}$$

Let us consider the first term of the r.h.s. of the latter inequality. We have that

$$\int \log\left(\frac{q_k^*(x)}{q_{k+1}^*(x)}\right)(q_k^*(x)-q_{k+1}^*(x))\mathrm{d}x = \eta_k \int h_{q_k^*}(x)(q_k^*(x)-q_{k+1}^*(x))\mathrm{d}x$$
$$\leq \eta_k L \int |q_k^* - q_{k+1}^*| \,.$$

since by assumption $h_{q_k^*}$ is bounded by L. Additionally, we have by Pinsker's inequality that

$$-\operatorname{KL}\left(q_{k+1}^* \| q_k^*\right) \leq -\frac{1}{2}\left(\int |q_k^* - q_{k+1}^*|\right)^2 \,.$$

Now combining with the fact that $\eta_k L a - a^2/2 \leq (\eta_k L)^2/2$ for all $a \geq 0$, we get:

$$\int \log\left(\frac{q_k^*(x)}{q_{k+1}^*(x)}\right)(q_k^*(x)-q_{k+1}^*(x))\mathrm{d}x - \operatorname{KL}\left(q_{k+1}^* \| q_k^*\right) \leq \frac{(\eta_k L)^2}{2}$$

and as a consequence we deduce

$$\eta_k \Delta_k \leq \frac{(\eta_k L)^2}{2} + \left[\operatorname{KL}\left(f \| q_k^*\right) - \operatorname{KL}\left(f \| q_{k+1}^*\right)\right] \,.$$

Finally, as we recognize a telescoping sum in the right-hand side, we have

$$\sum_{k=1}^n \eta_k \Delta_k \leq \sum_{k=1}^n \eta_k^2 L^2/2 + \operatorname{KL}\left(f \| q_1\right)$$

that is, by convexity of the mapping $q \mapsto \operatorname{KL}(q\|f)$,

$$\operatorname{KL}\left(\sum_{k=1}^n \frac{\eta_k q_k^*}{\sum_{k'=1}^n \eta_{k'}} \middle\| f\right) - \operatorname{KL}\left(f\|f\right) \leq \frac{\sum_{k=1}^n \eta_k^2 L^2/2}{\sum_{k=1}^n \eta_k} + \frac{\operatorname{KL}\left(f\|q_1\right)}{\sum_{k=1}^n \eta_k}$$

Then, notice that taking $\eta_k = \eta_0/\sqrt{k}$ with $\eta_0 > 0$ yields an $O(\log(n)/\sqrt{n})$ convergence rate and that setting $\eta_k = \eta_0/\sqrt{n}$ for all $k = 1 \dots n$ yields an $O(1/\sqrt{n})$ convergence rate. □

C Proof of Proposition 3

The proof is organized in three parts. First we provide high-level results related to Freedman's inequality for martingales. Then we provide some intermediary technical results, and finally we conclude with the proof of Proposition 3.

C.1 Bernstein inequalities for martingale processes

The two following propositions can be found in [27].

Proposition 7. Let $(\Omega, \mathcal{F}, (\mathcal{F}_k)_{k\geq 1}, \mathbb{P})$ be a filtered space. Let $(Y_k)_{1\leq k\leq n}$ be real valued random variables such that

$$\mathbb{E}[Y_k|\mathcal{F}_{k-1}] = 0, \quad \text{for all } k = 1, \dots, n \,.$$

Then, for all $t \geq 0$ and all $v, m > 0$,

$$\mathbb{P}\left(\left|\sum_{k=1}^n Y_k\right| \geq t, \max_{k=1,\dots,n} |Y_k| \leq m, \sum_{k=1}^n \mathbb{E}[Y_k^2|\mathcal{F}_{k-1}] \leq v\right) \leq 2\exp\left(-\frac{t^2}{2(v+tm/3)}\right) \,.$$

Proposition 8. Let $(\Omega, \mathcal{F}, (\mathcal{F}_k)_{k\geq 1}, \mathbb{P})$ be a filtered space. Let $(Y_k)_{k\geq 1}$ be a sequence of real valued stochastic processes defined on $\mathbb{R}^d$, adapted to $(\mathcal{F}_k)_{k\geq 1}$, such that for any $x \in \mathbb{R}^d$,

$$\mathbb{E}[Y_k(x)|\mathcal{F}_{k-1}] = 0, \quad \text{for all } k \geq 1 \,.$$

Consider $\epsilon > 0$ and let $(\tilde{Y}_k)_{k\geq 1}$ be another $(\mathcal{F}_k)_{k\geq 1}$-adapted sequence of nonnegative stochastic processes defined on $\mathbb{R}^d$ such that for all $k \geq 1$ and $x \in \mathbb{R}^d$

$$\sup_{\|y\|\leq\epsilon} |Y_k(x+y) - Y_k(x)| \leq \tilde{Y}_k(x) \; .$$

Let $n \geq 1$ and assume that for some $A \geq 0$ and some set $\Omega_1 \subset \Omega$, there exist $m, v, \tau \in \mathbb{R}^+$ such that for all $\omega \in \Omega_1$ and $\|x\| \leq A$,

$$\max_{k=1,\dots,n} |Y_k(x)| \leq m \tag{10}$$

$$\sum_{k=1}^{n} \mathbb{E}\big[Y_k(x)^2 | \mathcal{F}_{k-1}\big] \leq v \tag{11}$$

$$\sum_{k=1}^{n} \mathbb{E}\big[\tilde{Y}_k(x) | \mathcal{F}_{k-1}\big] \leq \tau \; . \tag{12}$$

Then, for all $t \geq 0$,

$$\mathbb{P}\Big(\sup_{\|x\|\leq A} \big| \sum_{k=1}^{n} Y_k(x) \big| > t + \tau, \Omega_1 \Big) \leq 4(1 + 2A/\varepsilon)^d \exp\left(-\frac{t^2}{8(\tilde{v} + 2mt/3)} \right) \; ,$$

with $\tilde{v} = \max(v, 2m\tau)$.

Recall also that the *predictable quadratic variation* [37] of a martingale $\sum_{k=1}^{n} \beta_k$ is given by

$$\sum_{k=1}^{n} \mathbb{E}[\beta_k^2 | \mathcal{F}_{k-1}].$$

This quantity is important as it appears as an essential factor in the Bernstein inequalities above. In particular, the predictable quadratic variation of the function M_n defined in (7) can be found in Proposition 9 below.

Proposition 9. Suppose that the kernel function $K : \mathbb{R}^d \to \mathbb{R}^+$ is bounded. For $n \geq 1$, let $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ be the σ-algebra generated by the random variables $X_1, \dots, X_n$, and $\mathcal{F}_0 = \emptyset$. Then, for each $x \in \mathbb{R}^d$, $(nM_n(x))_{n\geq 1}$ is a $(\mathcal{F}_n)_{n\geq 1}$-martingale with predictable quadratic variation $\sum_{k=1}^{n} V_{\eta_k, h_n}(x)$ where $V_{\eta,h}(x) = \int K_h(x-y)^2 f^{2\eta}(y) q^{1-2\eta}(y) dy - f \star K_h(x)^2$.

Proof. To lighten notation, we set $g(y) = K_{h_n}(x - y)$ for the next few lines. Because q_k is positive on $\mathbb{R}^d$ and $\mathcal{F}_k$-measurable, $\mathbb{E}[W_k^{\eta_k} g(X_k) \mid \mathcal{F}_{k-1}] = \int f^{\eta_k} q_{k-1}^{1-\eta_k} g$. The formula for the predictable quadratic variation follows in the same way. □

C.2 Technical results

We start with some notation.

Notation. Recall from Section 3.2 that for all $n \geq 1$ and all $x \in \mathbb{R}^d$,

$$f_n(x) = \frac{N_n(x)}{D_n}$$

$$N_n(x) = n^{-1} \sum_{k=1}^{n} W_k^{\eta_k} K_{h_n}(x - X_k)$$

$$D_n = n^{-1} \sum_{k=1}^{n} W_k^{\eta_k} \; .$$

In addition, for all $k \geq 1$ and all $x \in \mathbb{R}^d$, we introduce the helpful notation

$$\tilde{f}_k(x) = f^{\eta_k}(x) q_{k-1}^{1-\eta_k}(x),$$

$$Z_k^{(1)}(x) = W_k^{\eta_k} K_{h_k}(x - X_k) - (\tilde{f}_k \star K_{h_k})(x),$$

$$Z_k^{(2)} = W_k^{\eta_k} - \int \tilde{f}_k.$$

The three following propositions focus on convergence results for $Z_k^{(1)}$ and $Z_k^{(2)}$.

Proposition 10. Under $(\mathbf{A}_1)$ and $(\mathbf{A}_2)$ we have that almost-surely,

$$\forall k = 1, \dots, n, \qquad 0 \leq W_k^{\eta_k} \leq (c\lambda_n)^{-1} \, .$$

Proof. Note that by definition of q_k in (3) combined with $(\mathbf{A}_2)$, we have that for all $k = 1, \dots, n$ and all $x \in \mathbb{R}^d$,

$$q_k(x) \geq \lambda_k q_0(x) \geq \lambda_k c f(x) \, . \tag{13}$$

As a consequence, $0 \leq W_k \leq 1/(c\lambda_{k-1}) \leq 1/(c\lambda_n)$ using that $(\lambda_k)_{k\geq 1}$ is nonincreasing under $(\mathbf{A}_1)$. Note also that (13), we have that $\lambda_n \in (0, 1/c]$. Thus, we can write, for all $k = 1, \dots, n$, $(c\lambda_n)^{-\eta_k} \leq (c\lambda_n)^{-1}$, which implies the stated result. □

Proposition 11. Under $(\mathbf{A}_1)$ and $(\mathbf{A}_2)$ we have that almost-surely

$$n^{-1} \sum_{k=1}^{n} Z_k^{(2)} = O\left(\sqrt{\frac{\log(n)}{n\lambda_n}} \right).$$

Proof. We first show that without loss of generality, we can derive the proof assuming that the sequence $(\eta_k)_{k\geq 1}$ is valued in $(1/2, 1]$. From Proposition 10, it easy to see that for all $k = 1, \dots, n$,

$$|Z_k^{(2)}| = |W_k^{\eta_k} - \mathbb{E}[W_k^{\eta_k} | \mathcal{F}_{k-1}]| \leq (c\lambda_n)^{-1} \, . \tag{14}$$

Since the sequence $(\eta_k)_{k\geq 1}$ goes to 1 under Assumption $(\mathbf{A}_1)$(iii), there is an integer $k_0 \geq 1$ such that $\eta_k \in (1/2, 1]$ for all $k > k_0$. Hence, whenever $n > k_0$, we have

$$\sum_{k=1}^{n} Z_k^{(2)} = \sum_{k=1}^{k_0} Z_k^{(2)} + \sum_{k=k_0+1}^{n} Z_k^{(2)}.$$

By (14), the first term is bounded by $k_0/(c\lambda_{k_0})$ and hence has a negligible contribution in the bound we need to establish. The only term that matters is then the second one. Hence, from now on we assume that $(\eta_k)_{k\geq 1}$ is valued in $(1/2, 1]$. The goal will be to apply Proposition 7 to $Y_k(x) = Z_k^{(2)}(x)$. Using that $0 < 2\eta_k - 1 \leq 1$ and Proposition 10, we can write

$$\begin{aligned}
\mathbb{E}[Z_k^{(2)2} | \mathcal{F}_{k-1}] &\leq \mathbb{E}[f(X_k)/q_{k-1}(X_k))^{2\eta_k} | \mathcal{F}_{k-1}] \\
&= \int f(x)^{2\eta_k} q_{k-1}(x)^{1-2\eta_k} \, \mathrm{d}x \\
&= \int (f(x)/q_{k-1}(x))^{2\eta_k - 1} f(x) \, \mathrm{d}x \\
&\leq (1/(c\lambda_{k-1}))^{2\eta_k - 1} \\
&\leq (1/(c\lambda_n))^{2\eta_k - 1} \\
&\leq (c\lambda_n)^{-1} \, .
\end{aligned}$$

It follows that

$$\sum_{k=1}^{n} \mathbb{E}[Z_k^{(2)2} | \mathcal{F}_{k-1}] \leq n(c\lambda_n)^{-1}.$$

Consequently, and using (14), we can apply Proposition 7 with $m = (c\lambda_n)^{-1}$, $v = n(c\lambda_n)^{-1}$ and we get that for

all n large enough such that $\log(n)/(n\lambda_n) \leq c$; this is made possible by $(\mathbf{A}_1)$; and all $\gamma \geq 9$,

$$\begin{aligned}
\mathbb{P}\left(\Big|\sum_{k=1}^{n} Z_k^{(2)}\Big| \geq \sqrt{\gamma(c\lambda_n)^{-1} n\log(n)}\right) &\leq 2\exp\left(-\frac{\gamma n\log(n)}{2(n+\sqrt{\gamma(c\lambda_n)^{-1}n\log(n)}/3)}\right) \\
&\leq 2\exp\left(-\frac{\gamma}{2(1+\sqrt{\gamma}/3)}\log(n)\right) \\
&= 2\exp\left(-\frac{\sqrt{\gamma}}{2}\frac{\sqrt{\gamma}}{1+\sqrt{\gamma}/3}\log(n)\right) \\
&\leq 2\exp\left(-\frac{9}{4}\log(n)\right) \\
&\leq 2\exp\left(-2\log(n)\right) = 2n^{-2},
\end{aligned}$$

which series is convergent. We obtain the desired result by invoking the Borel-Cantelli lemma.

□

Proposition 12. Under $(\mathbf{A}_1)$, $(\mathbf{A}_2)$, $(\mathbf{A}_3)$ and $(\mathbf{A}_4)$, we have that for any $r > 0$,

$$\sup_{\|x\|\leq n^r}\left|n^{-1}\sum_{k=1}^{n} Z_k^{(1)}(x)\right| = O\left(\sqrt{\frac{\log(n)}{nh_n^d\lambda_n}}\right) .$$

Proof. Using similar arguments as in the beginning of the proof of Proposition 11, we can assume that $(\eta_k)_{k\geq 1}$ is valued in $(1/2, 1]$. The proof consists in applying Proposition 8 to $Y_k(x) = Z_k^{(1)}(x)$ that is

$$Y_k(x) = W_k^{\eta_k} K_{h_n}(x - X_k) - \mathbb{E}[W_k^{\eta_k} K_{h_n}(x - X_k)|\mathcal{F}_{k-1}] .$$

In the next few lines, we derive the quantities m, v, τ that appear in Proposition 8. Under Assumption $(\mathbf{A}_4)$ combined with Proposition 10, we have that

$$|Y_k(x)| \leq \frac{K_\infty}{c\lambda_n h_n^d} .$$

The previous bound corresponds to m in Proposition 8. Moreover, using Proposition 10,

$$\begin{aligned}
\mathbb{E}[Y_k(x)^2|\mathcal{F}_{k-1}] &\leq \mathbb{E}[W_k^{2\eta_k} K_{h_n}(x - X_k)^2|\mathcal{F}_{k-1}] \\
&= h_n^{-2d}\int\left(\frac{f(y)}{q_{k-1}(y)}\right)^{2\eta_k-1} f(y)K((x-y)/h_n)^2\mathrm{d}y \\
&\leq h_n^{-2d}(1/(c\lambda_{k-1}))^{2\eta_k-1}\int f(y)K((x-y)/h_n)^2\mathrm{d}y \\
&= h_n^{-d}(1/(c\lambda_n))^{2\eta_k-1}\int f(x-h_nu)K(u)^2\mathrm{d}u \\
&\leq h_n^{-d}(c\lambda_n)^{-1}UK_\infty
\end{aligned}$$

where we used a variable change $u = (x - y)/h_n$ in the penultimate inequality, and the last inequality results from Assumption $(\mathbf{A}_4)$, $0 < 2\eta_k - 1 \leq 1$, and Assumption $(\mathbf{A}_3)$.

Hence, we get

$$\sum_{k=1}^{n}\mathbb{E}[Y_k(x)^2|\mathcal{F}_{k-1}] \leq nh_n^{-d}(c\lambda_n)^{-1}UK_\infty .$$

The previous bound corresponds to v in Proposition 8. Under $(\mathbf{A}_4)$, $|K_h(x+y-X_k)-K_h(x-X_k)| \leq L_K\|y/h\|h^{-d}$ and it holds that for all $\|y\| \leq \epsilon$,

$$|Y_k(x+y) - Y_k(x)| \leq (W_k^{\eta_k} + \mathbb{E}[W_k^{\eta_k}|\mathcal{F}_{k-1}])L_K\epsilon h_n^{-d-1}.$$

The l.h.s. of the previous inequality corresponds to $\tilde{Y}_k(x)$ in Proposition 8. We have

$$\sum_{k=1}^{n} \mathbb{E}[\tilde{Y}_k(x)|\mathcal{F}_{k-1}] \leq 2nL_K \epsilon h_n^{-d-1}.$$

where we have used that $\mathbb{E}[W_k^{\eta_k}|\mathcal{F}_{k-1}] \leq 1$.

Taking $\epsilon = h_n^{d+1}/n$, the value for τ in Proposition 8 is $2L_K$. Let us now summarize the different factors taken to apply Proposition 8:

$$\begin{aligned}
m &= \frac{K_\infty}{c\lambda_n h_n^d} \\
v &= nh_n^{-d}(c\lambda_n)^{-1} U K_\infty \\
\tau &= 2L_K \\
\tilde{v} &= \max(v, 2m\tau) \leq C \max(n/(\lambda_n h_n^d), \lambda_n^{-1} h_n^{-d}) = Cn/(\lambda_n h_n^d)
\end{aligned}$$

where C is a positive constant. Let $\gamma > 1$. We have, taking $t = \sqrt{\gamma n \log(n)/(h_n^d \lambda_n)}$, $A = n^r$ and $\Omega_1 = \Omega$, for n large enough ($t \geq \tau$ and $\tilde{v}\sqrt{\gamma} \geq 2mt/3$; this is made possible by $(\mathbf{A}_1)$),

$$\begin{aligned}
\mathbb{P}\Big(\sup_{\|x\| \leq n^r} |\sum_{k=1}^{n} Y_k(x)| > 2t \Big) &\leq \mathbb{P}\Big(\sup_{\|x\| \leq n^r} |\sum_{k=1}^{n} Y_k(x)| > t + \tau \Big) \\
&\leq 4(1 + 2n^{r+1}/h_n^{d+1})^d \exp\left(-\frac{t^2}{8(1+\sqrt{\gamma})\tilde{v}} \right) \\
&\leq 4(1 + 2n^{r+1}/h_n^{d+1})^d \exp\left(-\frac{\gamma \log(n)}{16C} \right)
\end{aligned}$$

The last inequality holds because $\gamma > 1$. It remains to choose γ large enough in order to ensure the summability condition in the Borel Cantelli lemma. □

C.3 End of the proof of Proposition 3

Since $f_n(x) = N_n(x)/D_n$ for all $n \geq 1$ and all $x \in \mathbb{R}^d$, it is enough to show that $|D_n - 1| = o(1)$, and $\sup_{\|x\| \leq n^r} |N_n(x) - f(x)| = o(1)$. Both results are obtained independently, starting with D_n.

Proof for D_n. First note that for all $n \geq 1$, we can write the following decomposition for D_n

$$D_n = n^{-1} \sum_{k=1}^{n} \int \tilde{f}_k + n^{-1} \sum_{k=1}^{n} Z_k^{(2)} \ .$$

Furthermore, under $(\mathbf{A}_1)$ and $(\mathbf{A}_2)$, Proposition 11 implies that the second term of the r.h.s. is $O(\sqrt{\log(n)/(n\lambda_n)})$ almost surely. Hence, by $(\mathbf{A}_1)$ the sequence $(D_n)_{n\geq 1}$ converges to 1 as soon as $(n^{-1}\sum_{k=1}^{n} \int \tilde{f}_k)_{n\geq 1}$ does. By the Cesaro lemma, this will be a consequence of having proven that $(\int \tilde{f}_k)_{k\geq 1}$ goes to 1, which is what we set out to do next. For all $k \geq 1$, Jensen's inequality yields $\int \tilde{f}_k \leq 1$ and setting $\lambda_0 = 1$, we deduce using $(\mathbf{A}_2)$ that

$$\int \tilde{f}_k = \int f^{\eta_k} q_{k-1}^{1-\eta_k} \geq \lambda_{k-1}^{1-\eta_k} \int f^{\eta_k} q_0^{1-\eta_k} \geq (\lambda_{k-1} c)^{1-\eta_k} \ .$$

Thus, $(\int \tilde{f}_k)_{k\geq 1}$ goes to 1 under $(\mathbf{A}_1)$ and we can conclude that $(D_n)_{n\geq 1}$ converges to 1.

Proof for N_n. For the numerator N_n, we follow a similar approach except that we need to deal with some convolution operator. For all $n \geq 1$ and all $x \in \mathbb{R}^d$, we can write

$$N_n(x) = n^{-1} \sum_{k=1}^{n} (\tilde{f}_k \star K_{h_n})(x) + n^{-1} \sum_{k=1}^{n} Z_k^{(1)},$$

where given $r > 0$, the second term of the r.h.s is $O(\sqrt{\log(n)/(nh_n^d \lambda_n)})$ as a consequence of Proposition 12.

To treat the first term, use $(\mathbf{A}_2)$ and $(\mathbf{A}_4)$, to obtain that for all $x \in \mathbb{R}^d$,

$$\lambda_k c f(x) \leq q_k(x) \leq h_k^{-d} K_\infty + U_{q_0} := h_k^{-d} C,$$

for some $C > 0$. It follows that

$$\left(n^{-1} \sum_{k=1}^{n} f_k^- \right) \star K_{h_n}(x) \leq n^{-1} \sum_{k=1}^{n} (\tilde{f}_k \star K_{h_n})(x) \leq \left(n^{-1} \sum_{k=1}^{n} f_k^+ \right) \star K_{h_n}(x)$$

with $f_k^-(x) = f(x)(c\lambda_k)^{1-\eta_k}$ and $f_k^+(x) = f(x)^{\eta_k}(h_k^{-d}C)^{1-\eta_k}$.

It remains to show that the previous lower and upper bounds converge to f uniformly. For $1 \leq p \leq +\infty$, let $\|\cdot\|_p$ denote the $L_p(\lambda)$-norm. Using that $\|g \star \tilde{g}\|_\infty \leq \|g\|_\infty \|\tilde{g}\|_1$, we find, for all collection of bounded functions $(g_1, \ldots, g_n)$,

$$\begin{aligned}
|\left(n^{-1} \sum_{k=1}^{n} g_k \right) \star K_{h_n}(x) - f(x)| &\leq |\left(n^{-1} \sum_{k=1}^{n} (g_k - f) \right) \star K_{h_n}(x)| + |f \star K_{h_n}(x) - f(x)| \\
&\leq \|n^{-1} \sum_{k=1}^{n} (g_k - f)\|_\infty \|K_{h_n}\|_1 + \|f \star K_{h_n} - f\|_\infty \\
&\leq n^{-1} \sum_{k=1}^{n} \|g_k - f\|_\infty + \|f \star K_{h_n} - f\|_\infty
\end{aligned}$$

Hence, in virtue of the Cesaro lemma, the fact that f_k^- and f_k^+ both converge uniformly to f enables to conclude that the first term in the latter upper bound goes to 0. The fact that $\|f \star K_{h_n} - f\|_\infty$ goes to 0 is an easy consequence of $(\mathbf{A}_4)$.

□

Remark 5. Notice that in the latter proof, a different bandwith h_k for each point X_k, $k = 1, \ldots, n$ could have been set (instead of $h_k = h_n$ for all k). Indeed, in the latter inequalities, the term $\|f \star K_{h_n} - f\|_\infty$ would be replaced by $1/n \sum_{k=1}^{n} \|f \star K_{h_k} - f\|_\infty$, which also goes to 0 by the Cesaro Lemma as soon as $\|f \star K_{h_n} - f\|_\infty$ goes to 0 .

D Proof of Proposition 4

For the sake of completeness, we first recall the Inequality Reversal lemma as written in [51, Lemma 1].

Lemma 13 (Inequality Reversal lemma)**.** Let X be a random variable and let $a, b > 0$, $c, d \geq 0$ be such that

$$\forall t > 0 \ , \quad \mathbb{P}(X \geq t) \leq a \exp\left(-\frac{bt^2}{c + dt} \right) \ .$$

Then, with probability at least $1 - \delta$,

$$|X| \leq \sqrt{\frac{c}{b} \ln \frac{a}{\delta}} + \frac{d}{b} \ln \frac{a}{\delta} \ .$$

The proof of Proposition 4 is an easy consequence of the following Lemma.

Lemma 14. Under $(\mathbf{A}_1)$, $(\mathbf{A}_2)$, $(\mathbf{A}_5)$ and $(\mathbf{A}_6)$, there exists $s_0 \in \mathbb{N}$ large enough such that

$$\sup_{\|x\| > n^{s_0}} f_n(x) = o(1) \ , \ a.s \ , \tag{15}$$

$$\sup_{\|x\| > n^{s_0}} f(x) = o(1) \ . \tag{16}$$

Proof. We start with (15). Let $n \geq 1$ and set $A = n^{s_0}/2$ with $s_0 \in \mathbb{N}$. For all $x \in \mathbb{R}^d$, we have the following decomposition

$$f_n(x) = \sum_{k=1}^{n} W_{n,k}^{(\eta_k)} K_{h_n}(x - X_k) \mathbb{I}_{\{\|X_k\| \leq A\}} + \sum_{k=1}^{n} W_{n,k}^{(\eta_k)} K_{h_n}(x - X_k) \mathbb{I}_{\{\|X_k\| > A\}} \ . \tag{17}$$

Our goal is to prove that for s_0 large enough, both terms on the r.h.s of (17) go to 0. We start by studying the first term of the r.h.s.

(i) Proof for the first term of the r.h.s of (17). For any $\|x\| > n^{s_0}$, we can write

$$\begin{aligned}\sum_{k=1}^{n} W_{n,k}^{(\eta_k)} K_{h_n}(x - X_k)\mathbb{I}_{\{\|X_k\|\leq A\}} &\leq \sup_{1\leq k\leq n}\sup_{\|y\|\leq A} K_{h_n}(x-y) \\ &\leq C_K h_n^{-d} \sup_{\|y\|\leq A,\, \|x\|>n^{s_0}} (1 + \|x-y\|/h_n)^{-r_K} \\ &\leq C_K h_n^{-d} \sup_{\|y\|\leq A,\, \|x\|>n^{s_0}} (1 + \|x-y\|/h_1)^{-r_K} \\ &\leq C_K h_n^{-d} (1 + n^{s_0}/(2h_1))^{-r_K},\end{aligned}$$

where the last inequality follows from the fact that for all $x, y \in \mathbb{R}^d$, $|\|x\| - \|y\|| \leq \|x - y\|$. We can then ensure that the previous term goes to 0 by letting s_0 be large enough.

(ii) Proof for the second term of the r.h.s of (17). For the second term of the r.h.s, we have

$$\begin{aligned}\sum_{k=1}^{n} W_{n,k}^{(\eta_k)} K_{h_n}(x - X_k)\mathbb{I}_{\{\|X_k\|>A\}} &= \left(\sum_{k=1}^{n} W_k^{\eta_k}\right)^{-1} \sum_{k=1}^{n} W_k^{\eta_k} K_{h_n}(x - X_k)\mathbb{I}_{\{\|X_k\|>A\}} \\ &\leq \left(\sum_{k=1}^{n} W_k^{\eta_k}\right)^{-1} K_\infty h_n^{-d} \sum_{k=1}^{n} W_k^{\eta_k}\mathbb{I}_{\{\|X_k\|>A\}}\,, \end{aligned} \tag{18}$$

and we are thus interested in studying the r.h.s of (18). A first remark is that using Proposition 11, we obtain that almost surely

$$\left(\sum_{k=1}^{n} W_k^{\eta_k}\right)^{-1} = n^{-1}(1 + o(1))^{-1}\,. \tag{19}$$

We now move on to the study of $\sum_{k=1}^{n} W_k^{\eta_k}\mathbb{I}_{\{\|X_k\|>A\}}$ in (18). To do so, for all $k \geq 1$, let us define $p_k(A) = \mathbb{E}[W_k^{\eta_k}\mathbb{I}_{\{\|X_k\|>A\}}|\mathcal{F}_{k-1}]$ and $Z_k^{(3)}(A) = W_k^{\eta_k}\mathbb{I}_{\{\|X_k\|>A\}} - p_k(A)$ so that

$$\sum_{k=1}^{n} W_k^{\eta_k}\mathbb{I}_{\{\|X_k\|>A\}} = \sum_{k=1}^{n} Z_k^{(3)}(A) + \sum_{k=1}^{n} p_k(A)\,.$$

Then, for all $k \geq 1$, it holds that

$$\begin{aligned}p_k(A) &= \int_{\|x\|>A} f(x)^{\eta_k} q_{k-1}(x)^{1-\eta_k}\,\mathrm{d}x \\ &\leq (c\lambda_{k-1})^{1-\eta_k} \int_{\|x\|>A} f(x)\,\mathrm{d}x \\ &= (c\lambda_{k-1})^{1-\eta_k} p(A)\,,\end{aligned}$$

with $p(A) := \int_{\|x\|>A} f(x)\,\mathrm{d}x$. Using $(\mathbf{A}_1)$, we deduce that there exists a constant $C > 0$ such that

$$\sum_{k=1}^{n} p_k(A) \leq Cnp(A)\,.$$

Furthermore, under $(\mathbf{A}_5)$, Markov's inequality yields

$$p(A) \leq A^{-\delta} \int \|x\|^{\delta} f(x)\,\mathrm{d}x \tag{20}$$

and as a consequence, we obtain

$$\sum_{k=1}^{n} p_k(A) \leq CnA^{-\delta} \int \|x\|^{\delta} f(x)\,\mathrm{d}x\ . \tag{21}$$

Additionally, observe that $\sum_{k=1}^{n} Z_k^{(3)}(A)$ is a sum of martingale increments so our next step will be to apply Proposition 7. For this purpose, note that under $(\mathbf{A}_2)$ and $(\mathbf{A}_5)$, we can write for all $k = 1, \ldots, n$,

$$W_k \leq \lambda_{k-1}^{-1} C_0 (1 + \|X_k\|^{\delta})^{-1} \leq \lambda_n^{-1} C_0 (1 + \|X_k\|^{\delta})^{-1}$$

so that

$$|Z_k^{(3)}(A)| \leq \lambda_n^{-\eta_k} C_0^{\eta_k} \sup_{\|x\|>A} (1 + \|x\|^{\delta})^{-\eta_k} \leq \lambda_n^{-1} C_0 (1 + A)^{-\delta}\ . \tag{22}$$

We now treat the two case $k \geq k_0$ and $k < k_0$ separately.

- When $k \geq k_0$ (such that $0 < 2\eta_k - 1 \leq 1$), we can write

$$\begin{aligned}
\mathbb{E}[Z_k^{(3)}(A)^2 | \mathcal{F}_{k-1}] &\leq \mathbb{E}[(f(X_k)/q_{k-1}(X_k))^{2\eta_k} \mathbb{I}_{\{\|X_k\|>A\}} | \mathcal{F}_{k-1}] \\
&= \int_{\|x\|>A} f(x)^{2\eta_k} q_{k-1}(x)^{1-2\eta_k}\,\mathrm{d}x \\
&= \int_{\|x\|>A} (f(x)/q_{k-1}(x))^{2\eta_k - 1} f(x)\,\mathrm{d}x \\
&\leq (1/(c\lambda_{k-1}))^{2\eta_k - 1} p(A) \\
&\leq (c\lambda_n)^{-1} p(A) \\
&\leq (c\lambda_n)^{-1} A^{-\delta} \int \|x\|^{\delta} f(x)\,\mathrm{d}x\ .
\end{aligned}$$

  where we have used (20) in the last inequality.

- When $k < k_0$, we have $\mathbb{E}[Z_k^{(3)}(A)^2 | \mathcal{F}_{k-1}] \leq MA^{-\delta}$ for some constant $M > 0$ that can be deduced from the almost sure bound (22) given just before.

It follows that, when n is large enough, for all $k = 1, \ldots, n$,

$$\mathbb{E}[Z_k^{(3)}(A)^2 | \mathcal{F}_{k-1}] \leq (c\lambda_n)^{-1} A^{-\delta} \int \|x\|^{\delta} f(x)\,\mathrm{d}x\ .$$

and therefore,

$$\sum_{k=1}^{n} \mathbb{E}[Z_k^{(3)}(A)^2 | \mathcal{F}_{k-1}] \leq n(c\lambda_n)^{-1} A^{-\delta} \int \|x\|^{\delta} f(x)\,\mathrm{d}x\ .$$

Consequently, we can apply Proposition 7 with $m = C_0 \lambda_n^{-1} (1 + A)^{-\delta}$, $v = n(c\lambda_n)^{-1} A^{-\delta} \int \|x\|^{\delta} f(x)\,\mathrm{d}x$ and we obtain that for all $t > 0$

$$\mathbb{P}\left(\Big| \sum_{k=1}^{n} Z_k^{(3)}(A) \Big| \geq t \right) \leq 2 \exp\left(-\frac{t^2}{2(v + tm/3)} \right)\ .$$

Inverting this inequality using Lemma 13, we get that, with probability $1 - 1/n^2$,

$$\begin{aligned}
\left| \sum_{k=1}^{n} Z_k^{(3)}(A) \right| &\leq \sqrt{4v \log(2n)} + (2m/3)\log(2n) \\
&= \sqrt{4n(c\lambda_n)^{-1} A^{-\delta} \left(\int \|x\|^{\delta} f(x)\,\mathrm{d}x \right) \log(2n)} + (2/3) C_0 \lambda_n^{-1} (1 + A)^{-\delta} \log(2n).
\end{aligned} \tag{23}$$

Invoking the Borel-Cantelli lemma we obtain that the previous bound is an almost sure rate.

Putting together (19), (21) and (23) in (18), we obtain the almost-sure bound

$$\begin{aligned}&\left|\sum_{k=1}^{n} W_{n,k}^{(\eta_k)} K_{h_n}(x - X_k)\mathbb{I}_{\{\|X_k\|>A\}}\right| \\ &= O\left(n^{-1}h_n^{-d}\left(\sqrt{nA^{-\delta}\lambda_n^{-1}\log(2n)} + (A^{-\delta}\lambda_n)^{-1}\log(2n) + nA^{-\delta}\right)\right) .\end{aligned}$$

We easily obtain that the previous bound goes to 0 provided that s_0 is large enough, which concludes the proof of (15).

As for (16), notice that by $(\mathbf{A}_5)$, for all $\|x\| > n^{s_0}$

$$f(x) \leq \frac{C_0 q_0(x)}{1 + \|x\|^{\delta}} \leq \frac{C_0 q_0(x)}{1 + n^{s_0\delta}},$$

we obtain (16) when s_0 is large enough.

□

E Proof of Proposition 5

Proof. We start by proving (i) and (ii).

Proof of (i) and (ii). First note that for our choice of $\mathbb{P}$ and $\mathbb{Q}$, we can write

$$\begin{aligned}D_1(\mathbb{P}||\mathbb{Q}) &= \sum_{\ell=1}^{m_k} W_{k,\ell}\log\left(\frac{W_{k,\ell}}{1/m_k}\right) = \sum_{\ell=1}^{m_k} W_{k,\ell}\log\left(W_{k,\ell}\right) + \log(m_k) \\ &\leq \sum_{\ell=1}^{m_k} W_{k,\ell}\left(W_{k,\ell} - 1\right) + \log(m_k)\end{aligned}$$

where we have used that $\log(x) \leq x - 1$ for $x > 0$. Thus, we have that

$$D_1(\mathbb{P}||\mathbb{Q}) \leq \log(m_k) . \tag{24}$$

In addition, [39, Theorem 3] implies that for all $\alpha \in [0, 1]$

$$0 \leq D_\alpha(\mathbb{P}||\mathbb{Q}) \leq D_1(\mathbb{P}||\mathbb{Q})$$

where equality is reached if and only if $\mathbb{P} = \mathbb{Q}$. Now combining with (24) and by definition of $\eta_{k,\alpha}$ in (8), we deduce that for all $\alpha \in [0, 1]$,

$$0 \leq \eta_{k,1} \leq \eta_{k,\alpha} \leq 1 ,$$

and that $\eta_{k,\alpha} = 1$ if and only if $\mathbb{P} = \mathbb{Q}$.

Proof of (iii). We assume that $\lim_{k\to\infty} m_k = m$. Hence, for k big enough, $m_k = m$. Hence we will derive the proof as if $m_k = m$ for all k. A first remark is that thanks to (ii), it is enough to prove that $\lim_{k\to\infty} \eta_{k,1} = 1$ in L_1 to obtain that for all $\alpha \in [0, 1]$, $\lim_{k\to\infty} \eta_{k,\alpha} = 1$ in L_1, that is $\lim_{k\to\infty} \mathbb{E}[|\eta_{k,\alpha} - 1|] = 0$.

Since that for all $k \geq 1$,

$$\eta_{k,1} = 1 - \frac{D_1(\mathbb{P}||\mathbb{Q})}{\log(m)} = -\frac{\sum_{\ell=1}^{m} W_{k,\ell}\log\left(W_{k,\ell}\right)}{\log(m)}$$

the proof is concluded if we can prove that, in L_1

$$\lim_{k\to\infty} \sum_{\ell=1}^{m} W_{k,\ell}\log\left(W_{k,\ell}\right) = -\log(m) . \tag{25}$$

To see this, let us define the two maps $g_1 : (w_1, \dots, w_m) \mapsto \sum_{\ell=1}^m w_\ell \log(w_\ell)$ and $g_2 : (w_1, \dots, w_m) \mapsto (\sum_{\ell'=1}^m w'_\ell)^{-1} (w_1, \dots, w_m)$. Observe then that the map g_1 is a continuous transformation (defined on the simplex) and g_2 is a continuous transformation on the space of nonnegative weights.

If we further denote by $(\tilde{W}_{k,\ell})_{\ell=1}^m$ the unnormalised weights (i.e $\tilde{W}_{k,\ell} = f(X_{k,\ell})/q_{k-1}(X_{k,\ell})$ for all $\ell = 1 \dots m$), then it is enough to show that $(\tilde{W}_{k,1}, \dots, \tilde{W}_{k,m})$ converges to $(1, \dots, 1)$ in L_1 to prove (25). Since for all $\ell = 1, \dots, m$,

$$\mathbb{E}[|\tilde{W}_{k,\ell} - 1|] = \int |f - q_{k-1}| \,,$$

this follows from Scheffé's lemma and the proof is concluded.

□

F Additional Experiments

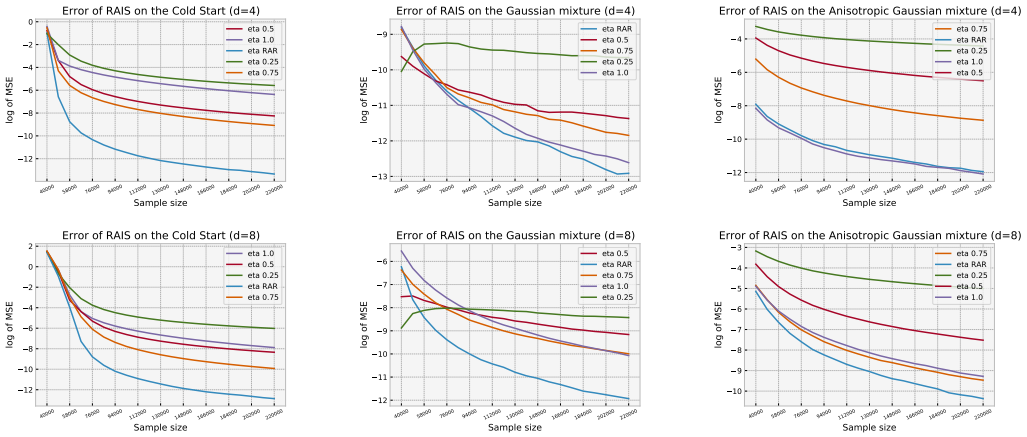

Figure 4: Logarithm of the average squared error, computed over 50 replicates.

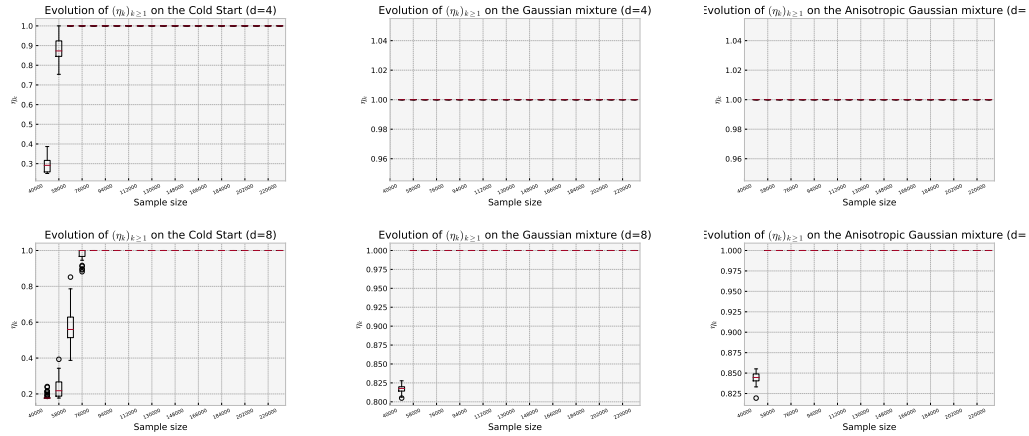

Figure 5: Boxplot of the values of η obtained from the ADA strategy.

We report in Figure 4 the results of the proposed method, and in Figure 5 the evolution of η in smaller dimensions. We can see in Figure 4 that Algorithm 1 along with the subroutine Algorithm 2 always outperform the competitive schedules for the regularization, and in Figure 5 that Algorithm 2 converges to 1.