A. Proofs of the main results

A.1. Proof of Theorem 1

Using Assumptions 1 and 2 and the fact that $\|\cdot\|_2 \le \|\cdot\|_s$, since $s \le 2$, we have that for any $x \in [0, A]^d$,

$$\begin{aligned}
\left|\mathbb{E}\left[\widehat{f}(x)\right] - f(x)\right| &\le \frac{c'' A^d}{A^d h^d} \int_{\mathbb{R}^d} \left| K\left(\frac{y-x}{h}\right)\right| \|y-x\|_s^s dy \\
&= \frac{c''}{h^d} \sum_{i=1}^d \int_{\mathbb{R}^d} |y_i - x_i|^s \prod_{j=1}^d K_0\left(\frac{y_j - x_j}{h}\right) dy_j \\
&= \frac{c''}{h^d} \sum_{i=1}^d \int_{\mathbb{R}} K_0\left(\frac{y_i - x_i}{h}\right) |y_i - x_i|^s dy_i \prod_{j\neq i} \underbrace{\int_{\mathbb{R}} K_0\left(\frac{y_j - x_j}{h}\right) dy_j}_{=h} \\
&= \frac{c''}{h} \sum_{i=1}^d \int_{\mathbb{R}} K_0\left(\frac{y_i - x_i}{h}\right) |y_i - x_i|^s dy_i \\
&= c'' d h^s \int_{\mathbb{R}} K_0(u) |u|^s du.
\end{aligned}$$

Since $0 < s \le 2$, we have that $s \log|u| \le \max(0, 2\log|u|)$, which means that $|u|^s \le \max(1, |u|^2) \le 1 + |u|^2$. Therefore, we can write

$$\int_{\mathbb{R}} K_0(u)|u|^s du \le \int_{\mathbb{R}} K_0(u)\left(1 + |u|^2\right) du \le 1 + C'.$$

Thus,

$$\left|\mathbb{E}\left[\widehat{f}(x)\right] - f(x)\right| \le c''(1+C') d h^s. \tag{4}$$

Let $0 < \delta < \frac{1}{2}$. Note that since K_0 is non-negative and bounded by C, we have that $K \le C^d$. For $y \in \mathbb{R}^d$, let us write $Y_i = A^d f(X_i) K\left(\frac{y - X_i}{h}\right)$, where $X_i \sim \mathcal{U}([0, A^d])$. This implies that $|Y_i| \le c(CA)^d$. Moreover, since $f \le c$, we have

$$\begin{aligned}
\mathbb{V}(Y_i) \le \mathbb{E}\left[Y_i^2\right] &= \frac{A^{2d}}{A^d} \int_{[0,A]^d} K^2\left(\frac{y-x}{h}\right) f^2(x) dx \\
&\le c^2 A^d \int_{\mathbb{R}^d} K^2\left(\frac{y-x}{h}\right) dx \\
&\le c^2 A^d h^d \int_{\mathbb{R}^d} K^2(u) du \\
&\le c^2 C^d A^d h^d \int_{\mathbb{R}^d} K(u) du \\
&= c^2 A^d C^d h^d.
\end{aligned}$$

Therefore, by Bernstein's inequality, for any $x \in [0, A]^d$, we know that with probability larger than $1 - \delta$

$$\begin{aligned}
\left|\mathbb{E}\left[\widehat{f}(x)\right] - \widehat{f}(x)\right| &= \left| \frac{A^d}{N h^d} \sum_{k=0}^n \left(f(X_i) K\left(\frac{X_i - x}{h}\right) - \mathbb{E}\left[f(X_i) K\left(\frac{X_i - x}{h}\right)\right]\right)\right| \\
&\le \frac{1}{N h^d}\left(2c\sqrt{C^d A^d h^d N \log(1/\delta)} + 2c A^d C^d \log(1/\delta)\right) \\
&\le 2c\sqrt{C^d A^d \frac{\log(1/\delta)}{N h^d}} + 2c A^d C^d \frac{\log(1/\delta)}{N h^d},
\end{aligned} \tag{5}$$

for n large enough with respect to δ.

By Equations 4 and 5, we have for any $x \in [0, A]^d$, with probability larger than $1-\delta$, that

$$\left|\widehat{f}(x) - f(x)\right| \le 2c\sqrt{C^d A^d \frac{\log(1/\delta)}{Nh^d}} + 2cC^d A^d \frac{\log(1/\delta)}{Nh^d} + c''(1+C')dh^s. \tag{6}$$

Therefore, for $h \overset{\text{def}}{=} h_s(\delta) = \left(\frac{\log(NA/\delta)}{N}\right)^{\frac{1}{2s+d}}$, we get that with probability larger than $1-\delta$,

$$\left|\widehat{f}(x) - f(x)\right| \le \left(2c\sqrt{C^d A^d} + c''(1+C')d\right)\left(\frac{\log(NA/\delta)}{N}\right)^{\frac{s}{2s+d}} + 2cC^d A^d \left(\frac{\log(NA/\delta)}{N}\right)^{\frac{2s}{2s+d}}. \tag{7}$$

Let $\mathcal{X}$ be a $1/N^v$ covering set in $\|.\|_2$ norm and of minimal cardinality of the hypercube $[0, A]^d$. Its cardinality is at most $(AN^v\sqrt{d})^d \le (ANd)^{d(v+1)}$, since the covering number of $[0, A]^d$ with small hypercubes of side $N^{-v}/\sqrt{d}$ is smaller than $\left(A/\frac{N^{-v}}{\sqrt{d}}\right)^d = (AN^v\sqrt{d})^d$, and hypercubes of side $N^{-v}/\sqrt{d}$ are contained in ℓ_2 balls of diameter N^{-v}. By a union bound and Equation 7, it holds that with probability larger than $1-\delta$, for any $x \in \mathcal{X}$, we have

$$\left|\widehat{f}(x) - f(x)\right| \le 4d(v+1)\left(2c\sqrt{A^d C^d} + c''(1+C')d\right)\left(\frac{\log(NAd/\delta)}{N}\right)^{\frac{s}{2s+d}} + 4cd(v+1)A^d C^d \left(\frac{\log(NAd/\delta)}{N}\right)^{\frac{2s}{2s+d}}.$$

Let ξ be the event of probability larger than $1-\delta$ where this is satisfied. Let $y \in [0, A]^d$. Then, there exists $x \in \mathcal{X}$ such that $\|x-y\|_2 \le 1/N^v$. Since K_0 is ε-Hölder and since f is bounded by c, we have that

$$\left|\widehat{f}(x) - \widehat{f}(y)\right| \le cN(C'')^d \frac{c^d}{h_s(\delta)^d N^{d\varepsilon v}} \le (C'')^d c^{d+1}\frac{N^{1+\frac{d}{2s+1}}}{N^{d\varepsilon v}},$$

and for $v > 3/\varepsilon \log(1 + 1/(C''c))$, we get

$$\left|\widehat{f}(x) - \widehat{f}(y)\right| \le N^{-1}.$$

In the same way, by Assumption 3, we get also that for $v \ge 2\log(1 + 1/(c''+c))/\min(1, s)$,

$$|f(x) - f(y)| \le c''N^{-v} + cN^{-vs} \le N^{-1}.$$

This implies that

$$\left|\widehat{f}(y) - f(y)\right| \le \left|\widehat{f}(x) - f(x)\right| + 2/N,$$

which means that on ξ, for any $x \in [0, A]^d$ we have

$$\left|\widehat{f}(x) - f(x)\right| \le 4d(v+1)\left(2\sqrt{cA^d C^d} + cdC'\right)\left(\frac{\log(NAd/\delta)}{N}\right)^{\frac{s}{2s+d}} + 4cd(v+1)A^d C^d\left(\frac{\log(NAd/\delta)}{N}\right)^{\frac{2s}{2s+d}} + 2/N,$$

where $v = \log\left(1 + \frac{1}{c''+c}\right)\frac{2}{\min(1,s)} + \frac{3}{\varepsilon}\log\left(1 + \frac{1}{C''c}\right)$. Therefore, we get that for any $x \in [0, A]^d$,

$$\begin{aligned}
&\left|\widehat{f}(x) - f(x)\right| \\
&\le 4d(v+1)\left(2\sqrt{cA^d C^d} + c''(1+C')d\right)\left(\frac{\log(NAd/\delta)}{N}\right)^{\frac{s}{2s+d}} + 4cd(v+1)A^d C^d\left(\frac{\log(NAd/\delta)}{N}\right)^{\frac{2s}{2s+d}} + 2/N \\
&\le \left(2\sqrt{cA^d C^d} + c''(1+C')d + c(AC)^d\left(\frac{\log(NAd/\delta)}{N}\right)^{\frac{s}{2s+d}}\right) 8d(v+1)\left(\frac{\log(NAd/\delta)}{N}\right)^{\frac{s}{2s+d}}, \\
&\le H_0\left(\frac{\log(NAd/\delta)}{N}\right)^{\frac{s}{2s+d}},
\end{aligned}$$

where $v = \log\left(1 + \frac{1}{c''+c}\right)\frac{2}{\min(1,s)} + \frac{3}{\varepsilon}\log\left(1 + \frac{1}{C''c}\right)$ and H_0 a constant that depends on d, v, c, c'', C, C', and A.

B. Extension to densities with unbounded support

In this part of the appendix, we extend our method to densities that do not have a compact support. For the sake of clarity, we assume that f is normalized here. In fact, we have already shown how to deal with the unnormalized case in the Section 3, where the normalization constant is estimated by a Monte-Carlo sum over the same samples used to estimate f.

Assumption 3 (Assumption on the density)**.** *The density f, defined on $\mathbb{R}^d$, is sub-Gaussian, i.e., there exist constants $c, c' > 0$ such that the density f satisfies for any $x \in \mathbb{R}^d$*

$$f(x) \leq c \exp\left(-c'\|x\|_2^2\right).$$

Moreover, f can be uniformly expanded by a Taylor expansion in any point up to degree s for some $0 < s \leq 2$, i.e., there exists $c'' > 0$ such that, for any $x \in \mathbb{R}^d$, and for any $u \in \mathbb{R}^d$, we have

$$|f(x+u) - f(x) - \langle \bigtriangledown f(x), u \rangle \mathbf{1}\{s > 1\}| \leq c''\|u\|_2^s.$$

The above assumption means that the tails of f are sub-Gaussian, and also that f is in a Hölder ball of smoothness s. Note that a bounded function f with a compact support in $\mathbb{R}^d$ is sub-Gaussian. The fact that f is in a Hölder ball of smoothnes s is also not very restrictive, in particular for a small s.

B.1. Uniform bounds for kernel density estimation

Let $X_1, \ldots, X_N$ be N points generated by f. Let us define for $h \stackrel{\text{def}}{=} h_s(\delta) = \left(\frac{\log(N/\delta)}{N}\right)^{\frac{1}{2s+1}}$,

$$\widehat{f}(x) = \frac{1}{Nh^d} \sum_{k=1}^{N} K\left(\frac{X_i - x}{h}\right),$$

and $\tilde{f}$ be such that

$$\tilde{f}(x) = \widehat{f}(x) \mathbf{1}\left\{\|x\|_2 \leq \log(N)\right\}. \tag{8}$$

Theorem 3. *Assume that Assumptions 2 and 3 hold with $0 < s \leq 2$, $C, C', C'', c, c', c'' > 0$, and $\varepsilon > 0$. The estimate $\tilde{f}$ is such that with probability larger than $1 - \delta$, for any point $x \in \mathbb{R}^d$,*

$$\begin{aligned}
\left|\tilde{f}(x) - f(x)\right| &\leq 8d(v+1)(2\sqrt{cC^d} + c''(1+C')d + C^d)\left(\frac{\log(N/\delta)}{N}\right)^{\frac{s}{2s+d}} + c\exp\left(-c'\|x\|_2^2\right)\mathbf{1}_{\|x\|_2 \geq \log(N)} \\
&\leq H_1 \left(\frac{\log(N/\delta)}{N}\right)^{\frac{s}{2s+d}} + c\exp\left(-c'\|x\|_2^2\right)\mathbf{1}_{\|x\|_2 \geq \log(N)}
\end{aligned}$$

where $v = \log\left(1 + \frac{1}{c''+c}\right)\frac{2}{\min(1,s)} + \frac{3}{\varepsilon}\log\left(1 + \frac{1}{C''c}\right)$, and H_1 is a constant that depends on d, v, c, c'', C, C'.

Theorem 3 provides a uniform bound on the error of $\tilde{f}$ on a large centered ball of radius $\log n$ denoted by $B_d(\log n)$), and bounds the far fluctuations by an upper bound on f itself. Note that the previous bound implies in particular that

$$\left|\tilde{f}(x) - f(x)\right| \leq \mathcal{O}\left(\left(\frac{\log(N/\delta)}{N}\right)^{\frac{s}{2s+d}}\right).$$

B.2. Proof of Theorem 3

By Assumptions 2 and 3, similarly to the starting point of proof of Theorem 1 (see Appendix A), we have that for any $x \in \mathbb{R}^d$

$$\left|\mathbb{E}\left[\widehat{f}(x)\right] - f(x)\right| \leq c''(1+C')dh^s. \tag{9}$$

Let $0 < \delta < \frac{1}{2}$. Note that since K_0 is non-negative and bounded by C, we have that $K \leq C^d$. For $y \in \mathbb{R}^d$, let us write $Y_i = K\left(\frac{y-X_i}{h}\right)$, where $X_i \sim f$. This implies that $|Y_i| \leq C^d$. Moreover, since $f(x) \leq c\exp\left(-c'\|x\|_2^2\right) \leq c$, we have

$$
\begin{aligned}
\mathbb{V}(Y_i) &\leq \int_{\mathbb{R}^d} K^2\left(\frac{y-x}{h}\right) f(x)dx \\
&\leq c\int_{\mathbb{R}^d} K^2\left(\frac{y-x}{h}\right) dx \\
&\leq ch^d \int_{\mathbb{R}^d} K^2(u)du \\
&\leq cC^d h^d \int_{\mathbb{R}^d} K(u)du \\
&= cC^d h^d.
\end{aligned}
$$

Therefore, by Bernstein's inequality, for any $x \in \mathbb{R}^d$, we know that with probability larger than $1-\delta$

$$
\begin{aligned}
\left|\mathbb{E}\left[\widehat{f}(x)\right] - \widehat{f}(x)\right| &= \left|\frac{1}{Nh^d}\sum_{k=0}^{n}\left(K\left(\frac{X_i - x}{h}\right) - \mathbb{E}\left[K\left(\frac{X_i-x}{h}\right)\right]\right)\right| \\
&\leq \frac{1}{Nh^d}\left(2\sqrt{cC^dh^dN\log(1/\delta)} + 2C^d\log(1/\delta)\right) \\
&\leq 2\sqrt{cC^d\frac{\log(1/\delta)}{Nh^d}} + 2C^d\frac{\log(1/\delta)}{Nh^d},
\end{aligned} \tag{10}
$$

for n large enough with respect to δ.

By Equations 9 and 10, we thus know that for any $x \in \mathbb{R}^d$, with probability larger than $1-\delta$,

$$
\left|\widehat{f}(x) - f(x)\right| \leq 2\sqrt{cC^d\frac{\log(1/\delta)}{Nh^d}} + 2C^d\frac{\log(1/\delta)}{Nh^d} + c''(1+C')dh^s. \tag{11}
$$

Therefore, for $h = h_s(\delta) = \left(\frac{\log(N/\delta)}{N}\right)^{\frac{1}{2s+d}}$, we get that with probability larger than $1-\delta$,

$$
\left|\widehat{f}(x) - f(x)\right| \leq \left(2\sqrt{cC^d} + c''(1+C')d\right)\left(\frac{\log(N/\delta)}{N}\right)^{\frac{s}{2s+d}} + 2C^d\left(\frac{\log(N/\delta)}{N}\right)^{\frac{2s}{2s+d}}. \tag{12}
$$

Now, since the Ψ_2 norm of f is bounded by c', we know that for any $x \in \mathbb{R}^d$,

$$
f(x) \leq c\exp\left(-c'\|x\|_2^2\right).
$$

This implies in particular that for any x such that $\|x\|_2 \geq \log(N)$, we have

$$
f(x) \leq c\exp\left(-c'\|x\|_2^2\right)\mathbf{1}_{\|x\|_2 \geq \log N}. \tag{13}
$$

Let $\mathcal{X}$ be a $1/N^v$ covering set in $\|.\|_2$ norm and of minimal cardinality of the ball of $\mathbb{R}^d$ of center 0 and radius $\log N$ that we will denote by $B_d(\log N)$. Its cardinality is at most $(2N^v\log N)^d \leq N^{d(v+1)}$ (by a similar reasoning as in the proof of Thm 1). By a union bound and Equation 12, it holds that with probability larger than $1-\delta$, for any $x \in \mathcal{X}$, we have

$$
\left|\widehat{f}(x) - f(x)\right| \leq 4d(v+1)\left(2\sqrt{cC^d} + c''(1+C')d\right)\left(\frac{\log(N/\delta)}{N}\right)^{\frac{s}{2s+d}} + 4d(v+1)C^d\left(\frac{\log(N/\delta)}{N}\right)^{\frac{2s}{2s+d}}. \tag{14}
$$

Let ξ be the event of probability larger than $1-\delta$ where this is satisfied. Let $y \in B_d(\log N)$. Then, there exists $x \in \mathcal{X}$ such that $\|x-y\|_2 \leq 1/N^v$. Since K_0 is ε-Hölder, we have that

$$
\left|\widehat{f}(x) - \widehat{f}(y)\right| \leq N(C'')^d\frac{c^d}{h_s(\delta)^dN^{d\varepsilon v}} \leq (C'')^dc^d\frac{N^{1+\frac{d}{2s+1}}}{N^{d\varepsilon v}},
$$

and for $v > 3/\varepsilon \log(1 + 1/(C''c))$, we get

$$\left|\widehat{f}(x) - \widehat{f}(y)\right| \leq N^{-1}.$$

In the same way, by Assumption 3, we also get that for $v \geq 2\log(1 + 1/(c'' + c))/(\min(1, s))$,

$$|f(x) - f(y)| \leq c''N^{-v} + cN^{-vs} \leq N^{-1}.$$

This implies that

$$\left|\widehat{f}(y) - f(y)\right| \leq \left|\widehat{f}(x) - f(x)\right| + 2/N,$$

which means that on ξ, for any $x \in B_d(\log N)$ we have

$$\left|\widehat{f}(x) - f(x)\right| \leq 4d(v+1)\left(2\sqrt{cC^d} + c''(1+C')d\right)\left(\frac{\log(N/\delta)}{N}\right)^{\frac{s}{2s+d}} + 4d(v+1)C^d\left(\frac{\log(N/\delta)}{N}\right)^{\frac{2s}{2s+d}} + 2/N,$$

where $v = \log\left(1 + \frac{1}{c''+c}\right)\frac{2}{\min(1,s)} + \frac{3}{\varepsilon}\log\left(1 + \frac{1}{C''c}\right)$. Combining this with the definition of $\tilde{f}$, and Equation 13, we get that for any $x \in \mathbb{R}^d$,

$$\begin{aligned}
&\left|\tilde{f}(x) - f(x)\right| \\
&\leq 4d(v+1)\left(2\sqrt{cC^d} + c''(1+C')d\right)\left(\frac{\log(N/\delta)}{N}\right)^{\frac{s}{2s+d}} + 4d(v+1)C^d\left(\frac{\log(N/\delta)}{N}\right)^{\frac{2s}{2s+d}} + 2/N + f(x)\mathbf{1}_{\|x\|_2 \geq \log N} \\
&\leq 8d(v+1)\left(2\sqrt{cC^d} + c''(1+C')d\right)\left(\frac{\log(N/\delta)}{N}\right)^{\frac{s}{2s+d}} + 4d(v+1)C^d\left(\frac{\log(N/\delta)}{N}\right)^{\frac{2s}{2s+d}} + f(x)\mathbf{1}_{\|x\|_2 \geq \log N} \\
&\leq 8d(v+1)\left(2\sqrt{cC^d} + c''(1+C')d\right)\left(\frac{\log(N/\delta)}{N}\right)^{\frac{s}{2s+d}} + 4d(v+1)C^d\left(\frac{\log(N/\delta)}{N}\right)^{\frac{2s}{2s+d}} + c\exp\left(-c'\|x\|_2^2\right)\mathbf{1}_{\|x\|_2 \geq \log N},
\end{aligned}$$

where $v = \log\left(1 + \frac{1}{c''+c}\right)\frac{2}{\min(1,s)} + \frac{3}{\varepsilon}\log\left(1 + \frac{1}{C''c}\right)$

B.3. Extended pliable rejection sampling

Our modified algorithm, ***e**xtended **p**liable **r**ejection **s**ampling (EPRS)*, aims at sampling as many i.i.d. points distributed according to f as possible with a fixed budget of evaluations of f. It consists of three main steps: (i) a first rejection sampling step where it generates $\widehat{N}$ initial samples from f by rejection sampling using an *initial proposal*. Then, (ii) EPRS uses these samples to estimate f by a kernel density estimation method. Finally, (iii) EPRS uses the newly obtained estimate, plus a uniform bound on it, as a new *extended pliable proposal* for rejection sampling. Since this pliable proposal is closer to the target density than the initial proposal, the rejection sampling will reject significantly fewer points by using it. Our EPRS method is described as Algorithm 2.

As mentioned, our method makes use of an *initial proposal* density g that must satisfy the following properties with respect to the target density.

Assumption 4 (Assumption on the initial proposal)**.** *Let $M > 0$. We have*

$$f \leq Mg.$$

Furthermore, the density g is sub-Gaussian, i.e., there exist constants $a, a' > 0$ such that the density g defined on $\mathbb{R}^d$ satisfies

$$g(x) \leq a\exp\left(-a'\|x\|_2^2\right).$$

We set the following constants:

$$T_s \stackrel{\text{def}}{=} n^{\frac{2s+d}{3s+d}}, \quad \text{and} \quad \bar{N} \stackrel{\text{def}}{=} T_s/M - 2\sqrt{T_s\log(1/\delta)}.$$

$\bar{N}$ is actually a high probability lower bound of $\widehat{N}$ given by our algorithm (it is the number of samples obtained by the initial rejection sampling step). T_s is the number of calls needed for the first estimation step that will optimize the number of accepted samples in the second step.

We also define

$$r_{\bar{N}} \stackrel{\text{def}}{=} \mathcal{V}_n H_E \left(\frac{\log(\bar{N}/\delta)}{\bar{N}}\right)^{\frac{s}{2s+d}}, \quad \text{and}$$

$$\bar{g}_{\bar{N}} \stackrel{\text{def}}{=} \int_{B_d(\log(\bar{N}))^C} g(x)dx,$$

where $\mathcal{V}_n = \mathcal{V}\left(B_d\left(\log\left(n\right)\right)\right)$ is the volume of $B_d\left(\log n\right)$ the centered ball in $\mathbb{R}^d$ of radius $\log n$ and where H is a parameter of the algorithm.

Our method samples most of the samples by rejection sampling according to a *pliable proposal* that is defined as

$$\hat{g}^\star \stackrel{\text{def}}{=} \frac{1}{1 + r_{\bar{N}} + M\bar{g}_{\bar{N}}} \left(\tilde{f} + r_{\bar{N}}\mathcal{U}_{B_d(\log(n))} + Mg\mathbf{1}_{\|x\|_2 \geq \log(\bar{N})}\right),$$

where $\mathcal{U}_{B_d(\log(n))}$ is the uniform distribution on $B_d(\log(n))$, and $\tilde{f}$ the estimate of f defined in (8).

Algorithm 2 Extended pliable rejection sampling (EPRS)

Parameters: s, n, δ, H g, and M.
Initial sampling
  Draw T_s samples at random according to g, and evaluate f on them
Estimation of f
  Perform rejection sampling on the samples using M as the constant
  Obtain this way $\widehat{N}$ samples from f
  Estimate f by $\tilde{f}$ on these $\widehat{N}$ samples (Section B.1)
Generating the sample
  Sample $n - T_s$ samples from the *pliable proposal* $\hat{g}^\star$
  Perform rejection sampling on these samples using $1 + r_{\bar{N}}/\mathcal{V}_n + M\bar{g}_{\bar{N}}$ as a constant
  Obtain this way $\widehat{n}$ samples from f
Output: Return the $\widehat{n}$ samples.

Theorem 4. *Assume that Assumptions 2, 3, and 4 hold with $0 < s \leq 2$, g, $M > 0$, and that $H_E > 0$ is an upper bound on the constant H_1 defined in Theorem 3 (applied to f and $\tilde{f}$). The number $\widehat{n}$ of samples generated in an i.i.d. way according to f in this way is such that for n large enough, with probability larger than $1 - \delta$,*

$$\widehat{n} \geq n\left[1 - \mathcal{O}\left(\log\left(n/\delta\right)^{d+1} n^{-\frac{s}{3s+d}}\right)\right].$$

B.4. Proof of Theorem 4

By definition of $\bar{g}_{\bar{N}}$ and since the g is sub-Gaussian by Assumption 3, we have that

$$\bar{g}_{\bar{N}} = \int_{B_d(\log \bar{N})^C} g(x)dx \leq \int_{B_d(\log \bar{N})^C} a\exp(-a'\|x\|_2^2)dx \leq a(\bar{N})^{-da'(\log \bar{N})/4} \leq \bar{N}^{-1},$$

for n (and thus $\bar{N}$) large enough.

By definition of rejection sampling, the probability of accepting a sample is exactly $1/M$, where M is the upper bound used in the rejection sampling (provided that $f \leq Mg$). Therefore, and since the first rejection sampling step uses T_s samples, $\widehat{N}$ is a sum of T_s independent Bernoulli random variables of parameter $1/M$. Thus, we have with probability larger than $1 - \delta$ that

$$\widehat{N} \geq T_s/M - 2\sqrt{T_s\log(1/\delta)} \stackrel{\text{def}}{=} \bar{N}. \tag{15}$$

Let us write ξ' for the event where this happens. On ξ', we have by Theorem 3 (end of the proof) that with probability larger than $1-\delta$, for any $x \in \mathbb{R}^d$

$$\left|\tilde{f}(x) - f(x)\right| \le \frac{r_{\widehat{N}}}{\mathcal{V}_n} + f\mathbf{1}_{\|x\|_2 \ge \log \widehat{N}} \le \frac{r_{\bar{N}}}{\mathcal{V}_n} + Mg\mathbf{1}_{\|x\|_2 \ge \log \bar{N}}.$$

Let ξ be the intersection of ξ' and the event where Equation 15 holds. It has probability larger than $1-2\delta$. On ξ, we thus have that

$$\tilde{f} + r_{\bar{N}}\mathcal{U}_{B_d(\log n)} + Mg\mathbf{1}_{\|x\|_2 \ge \log \bar{N}} \ge f.$$

Therefore, the rejection sampling is going to provide samples that are i.i.d. according to f, and $\widehat{n}$ will be a sum of Bernoulli random variables of parameter $\frac{1}{1+r_{\bar{N}}+M\bar{g}_{\bar{N}}}$. We thus have that on ξ, with probability larger than $1-\delta$,

$$\widehat{n} \ge (n - T_s)\frac{1}{1 + r_{\bar{N}} + M\bar{g}_{\bar{N}}} - 2\sqrt{n\log(1/\delta)}.$$

This implies, together with the definition of $r_{\bar{N}}$ and the upper bound on $\bar{g}_{\bar{N}}$, that $\widehat{n}$ is with probability larger than $1-3\delta$ lower bounded as

$$\begin{aligned}
\widehat{n} &\ge (n - T_s)\left(1 - \pi^d \log\left(\bar{N}\right)^d H\left(\frac{\log(\bar{N}/\delta)}{\bar{N}}\right)^{\frac{s}{2s+d}} - M\bar{N}^{-1}\right) - 2\sqrt{n\log(1/\delta)}\\
&\ge n\left[1 - \frac{T_s}{n} - \pi^d \log(n)^d H\left(\frac{\log(n/\delta)}{T_s/M - 2\sqrt{T_s\log(1/\delta)}}\right)^{\frac{s}{2s+d}} - \frac{M}{T_s/M - 2\sqrt{T_s\log(1/\delta)}} - 4\sqrt{\frac{\log(1/\delta)}{n}}\right].
\end{aligned}$$

Since

$$T_s = n^{\frac{2s+d}{3s+d}},$$

we have that with probability larger than $1-3\delta$, for n large enough, there exists a constant K such that

$$\widehat{n} \ge n\left[1 - K\log(n/\delta)^{d+1} n^{-\frac{s}{3s+d}}.\right] \tag{16}$$