Electronic Journal of Statistics
Vol. 15 (2021) 1783–1813
ISSN: 1935-7524
https://doi.org/10.1214/21-EJS1830

On density estimation at a fixed point under local differential privacy*

Martin Kroll†

Ruhr-Universität Bochum
Fakultät für Mathematik
Universitätsstraße 150
D-44801 Bochum
e-mail: martin.kroll-k9x@ruhr-uni-bochum.de

Abstract: We consider non-parametric density estimation in the framework of local, both pure and approximate, differential privacy. In contrast to centralized privacy scenarios with a trusted curator, in the local setup anonymization must be guaranteed already on the individual data owners' side and must therefore precede any data mining tasks. Thus, the published anonymized data should be compatible with as many statistical procedures as possible. We consider different mechanisms to establish pure and approximate differential privacy, respectively. We obtain minimax type results over Sobolev classes indexed by a smoothness parameter $s > 1/2$ for the mean squared error at a fixed point. In particular, we show that appropriately defined kernel density estimators can attain the optimal rate of convergence if the bandwidth parameter is correctly specified. Notably, the optimal convergence rate in terms of the sample size n is $n^{-(2s-1)/(2s+1)}$ under pure differential privacy and thus deteriorated to the rate $n^{-(2s-1)/(2s)}$ which holds both without privacy restrictions and under approximate differential privacy. Since the optimal choice of the bandwidth parameter depends on the smoothness s and is thus not accessible in practise, adaptive methods for bandwidth selection are necessary and must, in the local privacy framework, be performed based on the anonymized data only. We address this problem by means of variants of Lepski's method tailored to the privacy setups at hand and obtain general oracle inequalities for private kernel density estimators. In the Sobolev case, the resulting adaptive estimators attain the optimal rates of convergence at least up to logarithmic factors. On the side, we discuss some critical issues related with the notion of approximate differential privacy.

MSC2020 subject classifications: Primary 62G05; secondary 68P25.

Keywords and phrases: Kernel density estimation, approximate local differential privacy, rates of convergence, adaptive estimation, Lepski's method.

Received May 2020.

*The author gratefully acknowledges financial support from GENES and by the French National Research Agency (ANR) under the grant Labex Ecodec (ANR-11-LABEX-0047). The work has equally been supported in part by the research grant DFG DE 502/27-1 of the German Research Foundation (DFG).

†Equal parts of this research were performed at CREST, ENSAE, Institut Polytechnique de Paris and Ruhr-Universität Bochum.

1. Introduction

In the modern information era data are routinely collected in all areas of private and public life. Although the availability of massive data sets is essential to answer important scientific and societal questions, the individual data owners (who may be individuals, households, research institutions, companies, ...) might refuse to share their, maybe sensitive, raw data with others. Even more, in view of regularly reported data leaks, they may not even want to entrust their data to a central curator who stores the data and publishes anonymized summary statistics. Finding ourselves in such a dilemma, the question whether and, if yes, how data analytics can still be performed is of special importance. For the evaluation of this question, several aspects have to be taken into account.

Firstly, in absence of a trusted curator, privacy of the data has to be achieved already *locally* at the individual data owners' level. The i-th data holder takes its datum, say X_i, as the input of a privacy mechanism and creates an output Z_i that is considered sufficiently anonymized, for instance, in the sense of any of the privacy definitions listed below. For the purpose of the present paper, a privacy mechanism is a Markov kernel Q_i between measurable spaces $(\mathfrak{X}, \mathscr{X})$ and $(\mathfrak{Z}, \mathscr{Z})$ generating Z_i given $X_i = x$ according to the distribution $Q^{Z_i|X_i=x}$. This definition of *local* privacy is in contrast to the framework of *centralized* or *global* privacy where the trusted curator can take the whole data set $\{X_1, \ldots, X_n\}$ to create an output Z. In this sense, the local privacy model can be seen as a proper submodel of the global one because the trusted curator can also mimic any conceivable procedure in the local model.

Secondly, for the quantification of privacy, different solutions have been proposed so far (see [1], Section 2 for a comprehensive overview of existing privacy definitions):

- In this paper, we will exclusively work in the framework of α-differential privacy and its generalization (α, β)-differential privacy as defined in Definition 2.1 below. These two privacy definitions are also referred to as *pure* and *approximate differential privacy*, respectively. Originally, these concepts have been suggested for the anonymization of microdata tables in a global privacy setup, more precisely in a framework where queries are answered by a server that has direct access to the sensitive data [12, 14, 13]. In the statistics community, working under privacy constraints has been popularized in the past decade, amongst others, through the articles [25, 17] (in the global setup) and [11] (in the local privacy setup). Another strict relaxation of pure differential privacy is *random differential privacy* as introduced in [16].
- An alternative quantification of privacy can be given as follows: Let φ be a function from $[0, \infty]$ to $\mathbb{R} \cup \{+\infty\}$ with $\varphi(1) = 0$. Then, the associated *φ-divergence* between two distributions $\mathbf{P}, \mathbf{Q}$ is

$$D_\varphi(\mathbf{P}\|\mathbf{Q}) = \int \varphi\left(\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}\mathbf{Q}}\right) \mathrm{d}\mathbf{Q} = \int \varphi\left(\frac{p}{q}\right) q \mathrm{d}\mu$$

  where μ is a measure such that $\mathbf{P}, \mathbf{Q} \ll \mu$ and p, q denote the corresponding

Radon-Nikodym densities. Then, the mechanism Q is called *β-φ-divergence private* if

$$\sup_{x,x' \in \mathfrak{X}} D_\varphi(Q(\cdot|X=x) \| Q(\cdot|X=x')) \leqslant \beta.$$

The intersection of these two concepts is non-empty: For instance, taking $\varphi(x) = |x-1|/2$, the φ-divergence $D_\varphi(\mathbf{P}\|\mathbf{Q})$ is the total variation distance, and the resulting β-φ-divergence is equivalent to $(0,\beta)$-differential privacy.

Thirdly, the published data $Z_1, \ldots, Z_n$ should ideally be multi-purpose in the sense that they can serve as input data for several types of analyses. Thus, when the unmasked data are for instance a sample from an unknown probability distribution, the anonymized data should contain as much information as possible about the whole distribution and not only about certain of its characteristics. One main motivation for this work is to introduce novel methodology in the framework of density estimation that aims to address also this issue by proposing a local approximate ($\beta > 0$) differential private mechanism whose output can be used for various types of analyses.

Roadmap of the article

Throughout the article, we consider the paradigmatic example of non-parametric density estimation. For the sake of simplicity, we assume that each of n data holders D_i observes a size-one sample X_i from a (in this paper) univariate target density f, but refuses to share this observation. In Section 2, we introduce several mechanisms to estrange the datum X_i. The first approach is based on adding appropriately scaled Laplace noise to a kernel density estimator at a single fixed point $t \in \mathbb{R}$. The idea of the second approach is to publish the original datum X_i with a certain probability p and publishing another random value with probability $1-p$ (thus, the distribution of the Z_i is the discrete mixture of the target density and another distribution).

In Section 3, we consider estimation of the unknown density function under approximate differential privacy from a minimax point of view. As the performance measure to evaluate arbitrary estimators, we consider the mean squared error at the fixed point $t \in \mathbb{R}$. Via the Laplace perturbation approach, we attain the convergence rate $n^{-(2s-1)/(2s+1)}$ in terms of n over Sobolev ellipsoids with smoothness index s under $(\alpha, 0)$-differential privacy which is slower than the optimal rate $n^{-(2s-1)/(2s)}$ in the setup without privacy constraints. However, this slower rate can be shown to be optimal for the case of pure differential privacy when $\beta = 0$. In turn, the standard rate $n^{-(2s-1)/(2s)}$ from the setup without privacy can be attained under (α,β)-differential for $\beta > 0$ by means of the second, mixture approach to obtain privacy combined with suitable kernel density estimators. As a consequence, the second approach attains the optimal rate of convergence and furthermore does not hinge on *a priori* knowledge of the point t that has to be chosen prior to the anonymization procedure. Hence, this approach enables the statistician to apply a wider spectrum of inference procedures. Investigating theoretical guarantees of such general procedures, however, is outside the scope of this work and deferred to future research.

As usual for kernel density estimators, the choice of the bandwidth parameter is crucial. In the considered minimax framework over Sobolev classes, the optimal order of the bandwidth that leads to a rate optimal estimator depends on the smoothness index s which is typically unknown. In Section 4, we apply a Lepski scheme tailored to the privacy framework to overcome this problem and obtain an adaptive choice of the bandwidth. In the case $\beta > 0$, there is no additional problem since a standard approach via Lepski's method can be applied with the data $Z_1, \dots, Z_n$ and one can conclude as in the case for known smoothness. However, in the case $\beta = 0$, the considered privacy mechanism depends already on the choice of the bandwidth that one actually wants to choose in an adaptive way. In order to perform the Lepski scheme here, any data owner has to publish the kernel density estimator not only for one single bandwidth but for a finite set of potential bandwidths. Such a multiple output still guarantees the desired privacy condition provided that the additive noise is multiplied with a factor proportional to the number of potential bandwidths which is logarithmic in the number of data sources in our case. Note that this issue specifically arises in the local privacy setup since in the global framework the trusted curator can apply the existing plethora of methods for bandwidth selection on the unmasked data, and then only publish the resulting estimator with the adaptively determined bandwidth in its anonymized form. We derive general oracle type inequalities for the estimator resulting from the Lepski procedure adapted to the privacy framework. For the specific example of Sobolev ellipsoids, the rates of convergence are merely deteriorated by logarithmic factors with respect to the case of *a priori* known smoothness.

2. Privacy mechanisms

2.1. Definition of approximate differential privacy

Let $(\mathfrak{X}, \mathscr{X})$ and $(\mathfrak{Z}, \mathscr{Z})$ be measurable spaces. A privacy mechanism is a Markov kernel $Q : \mathfrak{X} \times \mathscr{Z} \to [0,1]$ with the interpretation that, given original data $X = x$, an anonymized output is randomly drawn from the probability measure $Q(\cdot|X = x)$. In the non-interactive setup that we are going to consider, we work under the following definition of *approximate* or (α, β)*-differential privacy*.

Definition 2.1. Let $\alpha \geqslant 0, \beta \in [0,1]$. We say that $Z \sim Q(\cdot \mid X)$ is a local (α, β)-differentially private view of X if for all $x, x' \in \mathfrak{X}$, $A \in \mathscr{Z}$ the estimate

$$Q(A|X = x) \leqslant \exp(\alpha) Q(A|X = x') + \beta \tag{1}$$

holds true.

Let us emphasize that in Definition 2.1 the spaces $(\mathfrak{X}, \mathscr{X})$ and $(\mathfrak{Z}, \mathscr{Z})$ do not necessarily need to coincide. In the literature, the case $\beta = 0$ is also referred to as α-differential privacy or *pure* differential privacy. Evidently, the privacy condition (1) becomes more restrictive for smaller values of the two parameters α and β. Although Definition 2.1 smoothly bridges the cases $\beta = 0$ and $\beta > 0$, the classical anonymization techniques used for $\beta = 0$ and $\beta > 0$ are essentially

different: In the case $\beta = 0$, Laplace perturbation as well as randomization techniques as considered in [11, 21] can be used. In the case $\beta > 0$, adding appropriately scaled Gaussian noise has been suggested in [17]. However, as proved in [18], appropriately scaled Laplace noise can also lead to approximately differential private outputs (see Propositions 2.2 and 3.1 as well as Remark 3.2). In the sequel, we discuss how to achieve approximate differential privacy in the scenario of non-parametric density estimation at a fixed point.

2.2. Pure differential privacy by adding Laplace noise

Throughout, we exclusively consider the case that both the input and the output of the privacy mechanism are univariate and real-valued, that is $(\mathfrak{X}, \mathscr{X}) = (\mathfrak{Z}, \mathscr{Z}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. First, we consider so-called Laplace perturbation which is also used to derive an upper bound in Section 3. To introduce this mechanism, let $Y_i = g(X_i) \in \mathbb{R}$ a quantity derived from the X_i that should be masked. Define the sensitivity of g as

$$\Delta(g) = \sup_{x,x' \in \mathfrak{X}} |g(x) - g(x')|.$$

Recall that the univariate Laplace distribution, denoted by $\mathcal{L}(b)$, is given by the probability density function $p_b(x) = \frac{1}{2b} \exp(-|x|/b)$ (we include also the case $b = 0$; then the Laplace distribution is, by convention, the Dirac measure concentrated at 0). In particular, the variance of an $\mathcal{L}(b)$ distributed random variable is $2b^2$. The following proposition establishes approximate differential privacy by Laplace perturbation.

Proposition 2.2 (See [18], Example 5)**.** *Let $\alpha > 0$, $\beta \in [0, 1]$. Then*

$$Z = g(X) + b\xi$$

with $\xi \sim \mathcal{L}(1)$ for $b \geqslant \Delta(g)/(\alpha - \log(1 - \beta))$ provides an (α, β)-differential private view of $g(X)$ (and of X as well).

A benefit of Proposition 2.2 in contrast to the often proposed perturbation by Gaussian noise to establish approximate differential privacy is that it allows to deal with the cases $\beta = 0$ and $\beta > 0$ by the same approach. Moreover, letting the parameter β vary permits natural interpretations: If $\beta = 0$, the variance of $\sqrt{2}b\xi$ corresponds to the one that is usually encountered in the case of pure differential privacy. When β tends to one, the privacy constraint gets weaker and the variance of the centred noise $\sqrt{2}b\xi$ tends to 0. In the extreme case $\beta = 1$ it is even allowed to publish $g(X)$ directly. Let us however note that, with the exception of the extreme cases when $\beta = 1$, we are not able to obtain optimal convergence rates for $\beta \in (0, 1)$ via this approach following the calculations in the proof of Proposition 3.1 below. This is why we consider different strategies for the anonymization in the case $\beta \in (0, 1)$ in Subsection 2.3.

We now introduce kernel density estimators giving the main example that we have in mind for the function g in Proposition 2.2.

Example 2.3. Let $X_1, \dots, X_n$ i.i.d. according to an unknown probability density function $f\colon \mathbb{R} \to \mathbb{R}$. Let $t \in \mathbb{R}$ be fixed. Then the i-th dataholder, who observes $X_i \in \mathbb{R}$, can compute

$$K_h(X_i - t) := \frac{1}{h} K\left(\frac{X_i - t}{h}\right)$$

for a bounded kernel function K, that is, $K\colon \mathbb{R} \to \mathbb{R}$ is integrable and $\int K(u)\mathrm{d}u = 1$. The quantity $K_h(X_i - t)$ will play the role of $g(X)$ in Proposition 2.2. By the triangle inequality $\Delta(K_h(\cdot - t)) \leqslant 2\|K\|_\infty/h$, and one can take any $b \geqslant 2\|K\|_\infty/(h(\alpha - \log(1-\beta)))$ to obtain an approximate differential private view $Z_{i,h}$ of $K_h(X_i - t)$. Note that $t \in \mathbb{R}$ has to be fixed in advance before the anonymization procedure.

2.3. Approximate differential privacy by random replacement

Although Proposition 2.2 provides us with a mechanism in order to achieve (α, β)-differential privacy, we will see later on in Proposition 3.1 that this mechanism leads to a convergence rate proportional to $n^{-(2s-1)/(2s+1)}$ which is not optimal for $\beta \in (0,1)$. In order to resolve this defect, we suggest another (quite simple) mechanism by which even the standard optimal rate $n^{-(2s-1)/(2s)}$ is attainable. For this mechanism, the data holders have to agree on an arbitrary but fixed $(\alpha, 0)$-differentially private mechanism which we denote with $\widetilde{Q}$. Then, any of the n data holders D_i draws (independently of the others) a random number $U_i \in [0,1]$ according to the uniform distribution on the unit interval, and a random outcome $Y_i \sim \widetilde{Q}(\cdot|X = X_i)$. We have the following result.

Proposition 2.4. *The mechanism where the i-th data holder publishes Z_i given through*

$$Z_i = \begin{cases} X_i, & \text{if } U_i \leqslant \beta, \\ Y_i, & \text{if } U_i > \beta, \end{cases} \tag{2}$$

guarantees (α, β)-differential privacy.

For $\beta \in [0,1]$, the mechanism in Proposition 2.4 publishes the original datum with probability β, and the $(\alpha, 0)$-private datum Y_i with probability $1 - \beta$. In the special case of $(0, \beta)$-differential privacy, the mechanism $\widetilde{Q}(\cdot|X = x)$ must indeed be independent of x (this choice of $\widetilde{Q}$ is then evidently also admissible for (α, β)-differential privacy with $\alpha > 0$). Of course, for $\alpha = 0$ and $\beta = 0$, the mechanism guarantees even perfect privacy but is completely useless for further analyses. For $\beta > 0$, the mixture structure of the density of Z_i allows to estimate the density f with the usual rate of convergence that holds without privacy restrictions (see Proposition 3.3). A delicate aspect of the conception of (α, β)-differential privacy for $\beta > 0$ becomes evident via the admissible privacy procedure (2): it does not exclude algorithms that publish the original datum X_i with a strictly positive probability which will surely not be acceptable in certain applications. Even worse, even procedures where an observer does not

only observe the original datum but also knows that this is the case are not excluded by the very concept of approximate differential privacy. This important issue will be further discussed below.

2.4. A composition lemma for approximate differential privacy

For kernel density estimation, bandwidth selection is usually a delicate issue and so it is in our local privacy setup. Whereas in the centralized setup existing methods can be applied by the trusted curator on the unmasked data, this is not possible in our local setup when working with the Laplace mechanism from Example 2.3. Thus the data holders have to publish versions of the kernel density estimator for different bandwidths, and one has to adapt general strategies from the non-private framework to the one with approximate local differential private data. In order to do this under our privacy constraint it is necessary to understand how multiple outputs influence the defining condition of approximate differential privacy. The following lemma provides a result of this flavour and is known in the research literature on privacy for statistical databases. The setup is the following: Given the unmasked datum X, the data owner does not only want to publish $Z_1 = Z_1(X)$ but also $Z_2 = Z_2(X)$, i.e., the vector (Z_1, Z_2). The following result tells us how α and β for the single components have to be scaled in order to obtain (α, β)-differential privacy for multiple outputs.

Lemma 2.5 (Composition lemma for (α, β)-differential privacy)**.** *Let Z_i, $i = 1, 2$ be (α_i, β_i)-differential private and conditionally (on X) independent views of X, respectively. Then $Z = (Z_1, Z_2)$ is an $(\alpha_1+\alpha_2, \beta_1+\beta_2)$-differential private view of X.*

Of course, Lemma 2.5 can be successively applied. For instance, if we want to publish $Z_{i,h}$ from Example 2.3 for different h in a finite set $\mathcal{H}$, then α and β should be replaced with $\alpha' = \alpha/\#\mathcal{H}$ and $\beta' = \beta/\#\mathcal{H}$, respectively, in order to get differential privacy for $Z = (Z_{i,h})_{h\in\mathcal{H}}$.

3. Private minimax estimation

Minimax theory provides a standard framework to study convergence properties of estimators in non-parametric statistics [24]. In this section, we apply this general toolbox to the specific case of density estimation under privacy constraints. For fixed $t \in \mathbb{R}$ and any estimator $\widehat{\ell}$ of the linear functional $f(t)$ based on the private views $Z = \{Z_1, \ldots, Z_n\}$, we study its mean squared error

$$\mathbf{E}[(\widehat{\ell} - f(t))^2].$$

The guiding principle of minimax theory is to look for estimators that perform best in a worst-case scenario. However, due to the privacy framework, we have not only the freedom of choosing the estimator $\widehat{\ell}$ but also the privacy mechanism Q that generates the private outputs. Hence, following [11], classical minimax

theory has to be adapted and a natural quantity to consider is the private minimax risk

$$\inf_{\substack{\widehat{\ell}\in\sigma(Z)\\ Q\in\mathcal{Q}_{\alpha,\beta}}} \sup_{f\in\mathcal{P}} \mathbf{E}[(\widehat{\ell}-f(t))^2]$$

where $\mathcal{P}$ is some function class containing probability densities and the infimum is taken over all local (α,β)-differential private Markov kernels $Q\in\mathcal{Q}_{\alpha,\beta}$ and all estimators based on the local approximate differential private views Z of the corresponding original sample $X_1,\dots,X_n$. We specify the function class $\mathcal{P}$ by so called Sobolev ellipsoids $\mathcal{S}(s,L)$ that we define for $s>1/2$ and $L>0$ by means of

$$\mathcal{S}(s,L)=\{f\colon\mathbb{R}\to[0,\infty):\int f(x)\mathrm{d}x=1,\int|\mathcal{F}[f](\omega)|^2|\omega|^{2s}\mathrm{d}\omega\leqslant 2\pi L^2\},$$

which, for $s\in\mathbb{N}^*$, is equivalent to the definition

$$\mathcal{S}(s,L)=\{f\colon\mathbb{R}\to[0,\infty):\int f(x)\mathrm{d}x=1,\int(f^{(s)}(x))^2\mathrm{d}x\leqslant L^2\}.$$

In the first definition, $\mathcal{F}[f]$ denotes the Fourier transform of the density f, in the second one $f^{(s)}$ denotes the weak s-th derivative of f.

3.1. Upper bound

Upper bound for Laplace perturbation

We first derive an upper bound on the minimax risk by specializing both the privacy mechanism $Q\in\mathcal{Q}_{\alpha,\beta}$ and the estimator of $f(t)$. Concerning the privacy mechanism, we first consider the mechanism mapping X_i to private views $Z_{i,h}$ of $K_h(X_i-t)$ from Example 2.3 for one single $h>0$. More precisely, we consider the Laplace mechanism given through

$$Z_{i,h}(t)=K_h(X_i-t)+\underbrace{\frac{2\|K\|_\infty}{h(\alpha-\log(1-\beta))}}_{=:C_{\alpha\beta}/(\sqrt{2}h)}\xi_{i,h},\qquad \xi_{i,h}\text{ i.i.d.}\sim\mathcal{L}(1). \tag{3}$$

Given $Z_{1,h},\dots,Z_{n,h}$ as in (3), a natural estimator of $f(t)$ is given by

$$\widehat{f}_h(t)=\frac{1}{n}\sum_{i=1}^n Z_{i,h}(t). \tag{4}$$

The following proposition provides a uniform upper risk bound for this estimators over the Sobolev ellipsoids $\mathcal{S}(s,L)$ introduced above.

The proof of the results exploits the special choice of the kernel function as the so-called *sinc-kernel* defined via

$$K_{\mathrm{sinc}}(x)=\mathrm{sinc}(x)=\frac{\sin(\pi x)}{\pi x}. \tag{5}$$

Proposition 3.1. *Consider the kernel density estimator* $\widehat{f}_h(t)$ *for some fixed* $t \in \mathbb{R}$ *where the kernel used in the anonymization procedure* (3) *is the* sinc*-kernel given in* (5)*. Then, for any* $s > 1/2$*,*

$$\sup_{f \in \mathcal{S}(s,L)} \mathbf{E}[(\widehat{f}_h(t) - f(t))^2] \leqslant C \left[h^{2s-1} + \frac{1}{nh} + \frac{1}{nh^2} \right]$$

for some $C = C(\alpha, \beta, L, s, \|f\|_\infty, K_{\mathrm{sinc}})$*. In particular, setting* $h = h^\star$ *with* $h^\star \asymp n^{-1/(2s+1)}$*, we obtain*

$$\sup_{f \in \mathcal{S}(s,L)} \mathbf{E}[(\widehat{f}_{h^\star}(t) - f(t))^2] \lesssim n^{-\frac{2s-1}{2s+1}}.$$

Since the noise added by the privacy mechanisms is centred, the bias term in the proof of Proposition 3.1 remains unchanged in comparison to the standard setup without privacy constraints. However, the variance term changes due to the additional Laplace noise, and the classical variance term $1/(nh)$ is augmented by the additional term $1/(nh^2)$ which is of higher order for $h \to 0$. Consequently, the optimal bandwidth is no longer of order $n^{-1/(2s)}$ as in the standard setup but of the larger order $n^{-1/(2s+1)}$. However, consistency of $\widehat{f}_h$ is already guaranteed if $h \to 0$ and $nh^2 \to \infty$ simultaneously (in the standard density estimation setup one only needs $nh \to \infty$ in addition to $h \to 0$).

Remark 3.2. Proposition 3.5 below shows that the rate obtained in Proposition 3.1 is optimal for $\beta = 0$. However, in the following we will show that the upper bound on the rate of convergence in Proposition 3.1 is distinct from the optimal rate of convergence for $\beta > 0$. In this latter regime, one can even achieve the optimal rate of convergence from the non-privacy setup by appropriately chosen privacy mechanisms.

Upper bound for random replacement mechanisms

For our first specialization of the general approach in (2) we assume that the support of the X_i is bounded, say $X_i \in [0, 1]$ without loss of generality. Then, Proposition 2.2 shows that

$$Y_i = X_i + 2\xi_i \tag{6}$$

with $\xi_i \sim \mathcal{L}(1)$ is a α-differential private view of X_i, and can be used as a building block in the definition (2) of an (α, β)-differential private algorithm.

Let us denote with g the density of the random variable 2ξ. Then, the density φ of Z_i is (recall that $X_i \sim f$)

$$\varphi = \beta f + (1 - \beta)(f \star g).$$

In terms of the Fourier transform this yields

$$\mathcal{F}[f] = \frac{\mathcal{F}[\varphi]}{\beta + (1 - \beta)\mathcal{F}[g]}$$

which motivates to consider the kernel K_n defined via the Fourier transform

$$\mathcal{F}[K_n](\omega) = \mathcal{F}[K_{\mathrm{sinc}}](\omega) \cdot (\beta + (1-\beta)\mathcal{F}[g](\omega/h))^{-1} \tag{7}$$

with a bandwidth parameter $h = h_n \to 0$. Define the estimator $\widehat{f}_h(x)$ of f as

$$\widehat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K_n\left(\frac{x - Z_i}{h}\right). \tag{8}$$

We have the following result.

Proposition 3.3. *Assume $X_i \in [0,1]$. Consider the kernel density estimator $\widehat{f}_h(t)$ from Equation (8) for some fixed $t \in [0,1]$ with the kernel K_n defined through its Fourier transform (7). Then, for any $s > 1/2$,*

$$\sup_{f \in \mathcal{S}(s,L)} \mathbf{E}[(\widehat{f}_h(t) - f(t))^2] \leqslant C \left[h^{2s-1} + \frac{1}{n\beta^2 h}\right]$$

for some numerical constant C. In particular, setting $h = h^\star$ with $h^\star \asymp (n\beta^2)^{-1/(2s)}$, we obtain

$$\sup_{f \in \mathcal{S}(s,L)} \mathbf{E}[(\widehat{f}_{h^\star}(t) - f(t))^2] \lesssim (n\beta^2)^{-\frac{2s-1}{2s}}.$$

If X_i is not assumed to be bounded, the above Laplace mechanism is not applicable to obtain differential privacy. In this case, the only obvious mechanism to obtain $(0,\beta)$-differential privacy is to draw Y_i according to some arbitrary density g (on which the individual data holders have to agree) and then to publish

$$Z_i = \begin{cases} X_i, & \text{if } U_i \leqslant \beta, \\ Y_i, & \text{if } U_i > \beta. \end{cases} \tag{9}$$

In this case, the density φ of Z is a mixture with components f and g, namely

$$\varphi = \beta \cdot f + (1-\beta) \cdot g. \tag{10}$$

Then, the statistician can estimate the density φ via a usual kernel density estimator, say $\widehat{\varphi}_h$, and then estimate f via

$$\widehat{f}_h(t) := \frac{1}{\beta}\widehat{\varphi}_h(t) - \frac{1-\beta}{\beta} g(t). \tag{11}$$

This leads to the risk decomposition

$$\begin{aligned} \mathbf{E}[(\widehat{f}_h(t) - f(t))^2] &= \mathbf{E}\left[(\widehat{f}_h(t) - \frac{1}{\beta}\varphi(t) - \frac{1-\beta}{\beta} g(t))^2\right] \\ &= \beta^{-2}\mathbf{E}[(\widehat{\varphi}_h(t) - \varphi(t))^2]. \end{aligned} \tag{12}$$

If the density g is at least as smooth as the unknown density f, then φ inherits the smoothness s from f, and we obtain the following result (the easy proof of which is left to the reader).

Proposition 3.4. *Consider the estimator $\widehat{f}_h(t)$ from Equation (11) where we use the* sinc*-kernel in the definition of $\widehat{\varphi}_h$ and assume that the density g in (10) and (11) belongs to $\mathcal{S}(s,R)$ for some large enough R. Then, for all $s > \frac{1}{2}$,*

$$\sup_{f\in\mathcal{S}(s,L)} \mathbf{E}[(\widehat{f}_h(t) - f(t))^2] \leqslant C\beta^{-2}\left[h^{2s-1} + \frac{1}{nh}\right]$$

for some $C = C(\beta, s, L, R)$. In particular, setting $h = h^\star$ with $h^\star \asymp n^{-1/(2s)}$, we obtain

$$\sup_{f\in\mathcal{S}(s,L)} \mathbf{E}[(\widehat{f}_{h^*}(t) - f(t))^2] \lesssim \beta^{-2} n^{-\frac{2s-1}{2s}}.$$

3.2. Lower bound

The following result states a lower bound over Sobolev ellipsoids in the case of pure differential privacy ($\beta = 0$).

Proposition 3.5. *Let $\alpha > 0$ arbitrary. Then,*

$$\inf_{\substack{\widehat{\ell}\in\sigma(Z)\\ Q\in\mathcal{Q}_{\alpha,0}}} \sup_{f\in\mathcal{S}(s,L)} \mathbf{E}[(\widehat{\ell} - f(t))^2] \geqslant C(\alpha) n^{-\frac{2s-1}{2s+1}}$$

where $C(\alpha) > 0$ depends on the privacy parameter, and the infimum is taken over all estimators $\widehat{\ell}$ based on private views $Z_1, \ldots, Z_n$ and privacy mechanisms providing $(\alpha, 0)$-differential privacy.

Remark 3.6. The lower bound of Proposition 3.5 still holds true when one allows a slight amount of interaction between the data holders, namely when the distribution of every Z_i is determined by X_i and the previously masked values $Z_1, \ldots, Z_{i-1}$. The proof remains the same because the data processing inequality (14) from [11] still holds true in this more general setup.

Proposition 3.5 shows that, regarding the privacy parameter α as an *a priori* fixed constant, the estimators $\widehat{f}_h(t)$ from Proposition 3.1 attain the optimal rate $n^{-(2s-1)/(2s+1)}$ in terms of n under pure local differential privacy.

Recall that without privacy restrictions the optimal rate over Sobolev ellipsoids is $n^{-(2s-1)/(2s)}$ (as mentioned in [3], this rate can, other than by a reduction scheme as used in our proof, be easily obtained via the theory developed in [10], see also [22]).

Of course, lower bounds on the rate of convergence in the scenario without privacy still hold true in the setup of differential privacy (since the privacy restriction can be interpreted as restricting the set of admissible estimators). Thus, for approximate differential privacy the rates of convergence derived in Propositions 3.3 and 3.4 are optimal.

In this work, we consider the parameter α (and also β in the case of approximate differential privacy) as fixed and are interested in the behaviour of the rate as a function of n only but remarks concerning α analogous to the ones made

in [4] could be made (as in that paper, α and β could also be allowed to vary with n). The optimal behaviour, however, of the rates in terms of the privacy parameters α and β, especially if $\beta > 0$, remains an open issue which is outside the scope of this paper.

Instead, we give in the following a heuristic which links (α, β)-differential privacy to *missing data problems*. We have designed the privacy procedures considered in Propositions 3.3 and 3.4 such that either the original value is published with zero probability (Proposition 3.3) or that it is at least not evident whether the original value has been published (provided that g in Proposition 3.4 is appropriately chosen, for instance assuming that g has the same support as the original data). However, this (from a privacy point of view reasonable assumption) is not enforced by the notion of (α, β)-differential privacy alone: For example, the local mechanism that publishes

$$Z_i = \begin{cases} X_i, & \text{if } U_i \leqslant \beta, \\ \emptyset, & \text{if } U_i > \beta, \end{cases} \tag{13}$$

where $\emptyset$ denotes the empty set ensures (α, β)-differential privacy, the distribution of the random variable Y_i in (9) corresponding here to the trivial distribution on the one point set $\{\emptyset\}$. This mechanism, though, will certainly not be regarded as a legitimate privacy mechanism in most applications since it does not only reveal the true value with a positive probability but also tells the observer of the 'privatized' data if this is the case or not. This problem might become even more severe in multivariate setups where exact knowledge of one (maybe not per se sensitive) value associated with a certain individual might help to identify this individual and in addition reveals values associated with this person that are considered as sensitive.

Concerning the (optimal) rate of convergence, the missing data problem (13) is asymptotically equivalent to a standard nonparametric experiment where the number of observations is now βn instead of n which would lead to a rate of order $(n\beta)^{-\frac{2s-1}{2s}}$ over Sobolev spaces. Of course this rate is better than the one we have obtained above in Proposition 3.3 which is in turn better than the one from Proposition 3.4 (the latter fact being intuitive since in the setup of Proposition 3.4 the random variable Y_i does not contain any information on f any more which still holds in Proposition 3.3).

Finally, we would like to mention that the mixture structure in Equation (10) is well-known under the name of Huber's contamination model in the theory on robust statistics. In this contamination model, however, the nuissance component g is itself unknown which leads to slower rates when the supremum in the minimax formulation is also taken over a set of potential contamination distributions g [8]. In the framework of differential privacy, the data holders can agree on a suitable choice of g which even allows to maintain the standard rate of convergence that holds for unmasked observations.

4. Adaptation to unknown smoothness

The estimators of the previous section are not completely satisfying since the optimal choice $h_n^\star$ of the bandwidth, as usually in non-parametric statistics, depends on *a priori* knowledge of the smoothness of the unknown function f. Such knowledge is usually not available in practise. At least, using the approach suggested in Subsection 2.3 (with its specializations considered in Propositions 3.3 and 3.4) we relieved ourselves from the drawback of the Laplace method that one can privatize only one functional of the form $f(t)$ for one single t that has to be fixed even before the anonymization. Note that this drawback is, for instance, also present in the mechanisms suggested in [21]. From this point of view, (α, β)-differential privacy with strictly positive β via one of these approaches should be preferred.

The purpose of this section is to address the remaining issue of adapting to the unknown smoothness of f. Note that the problem of adaptation has, to the best of the author's knowledge, only been addressed in the recent work [4] so far, where the authors use wavelet estimators for density estimation on a compact interval. The approach in that paper is thus conceptionally different from the one presented in the sequel. In order to tackle this problem, we use a variant of Lepski's method (see [20] for a general account in the Gaussian white noise model, and [7] for an application to a tomography problem whose concise presentation has inspired our one).

Recall that in the case of global privacy (which is not considered here) the trusted data curator can choose the bandwidth in an adaptive way using all the data $X_1, \dots, X_n$ and, as a consequence, can build on the existing plethora of methods and theoretical results for this standard case; hence bandwidth selection does not provide any additional difficulty for centralized privacy since only the final output is anonymized. Nearly the same holds true for our approaches to achieve (α, β)-differential privacy presented in Subsection 2.3. Here, slight modifications of standard Lepski's method yield completely data-driven estimators that attain the optimal rates of convergence up to logarithmic factors. For these cases, we will state results but omit the proofs since these are obtained by adapting the one for the Laplace perturbation case (or standard results from the literature). However, we study in detail the case of pure differential privacy which is the one that differs most from the standard framework.

4.1. Adaptive estimation for Laplace perturbation

In order to apply Lepski's method for the case of Laplace perturbation, the observations (3) must be available for different values of the bandwidth parameter h, say $h \in \mathcal{H}_n$. This can be realized using Lemma 2.5 provided that the privacy parameters α and β are appropriately scaled. Thus, we can assume that $Z_{i,h}(t)$ in (3) are accessible for any $i \in [\![1, n]\!]$ and $h \in \mathcal{H}_n$ if we replace α and β by $\alpha' = \alpha/\#\mathcal{H}_n$ and $\beta' = \beta/\#\mathcal{H}_n$, respectively. For any $h \in \mathcal{H}_n$ and $t \in \mathbb{R}$, we can then consider the estimator defined in (4). In our case, we define the set of

potential bandwidths by a geometric grid,

$$\mathcal{H}_n = \{h \in [\underline{h}_n, \overline{h}_n] : h = a^{-j}\overline{h}_n, j \in \mathbb{N}\}, \tag{14}$$

where $a > 1$ is a fixed constant, $\overline{h}_n$ is such that $a\log(\overline{h}_n\sqrt{n})/\sqrt{n} \leqslant \overline{h}_n \leqslant 1$, and $\underline{h}_n$ satisfies $\underline{h}_n = (\log(\overline{h}_n\sqrt{n}) \vee 1)/\sqrt{n}$. For $h \in \mathcal{H}_n$ and some $M > 0$, define

$$v^2(h) = \frac{M\int K^2(u)\mathrm{d}u}{nh} + \frac{C^2_{\alpha'\beta'}}{nh^2}$$

where $C_{\alpha'\beta'}$ is defined as in Section 3. The proof of Proposition 3.1 shows that

$$\mathrm{Var}(\widehat{f}_h) \leqslant v^2(h)$$

if $\|f\|_\infty \leqslant M$. Put $\lambda(h) = \max(1, (\kappa\log(\overline{h}_n/h))^{1/2})$ with κ being a sufficiently large constant (an explicit value can be determined from the proof of Theorem 4.3) and define

$$h_n^* = h_n^*(t,f) = \max\{h \in \mathcal{H}_n : |f_\eta(t) - f(t)| \leqslant \frac{v(h)\lambda(h)}{2} \text{ for all } \eta \in \mathcal{H}_n, \eta \leqslant h\} \tag{15}$$

where $f_\eta := \mathbf{E}\widehat{f}_\eta$. If the set in the definition of h_n^* is empty, we set $h_n^* = \underline{h}_n$ by convention. However, in the proof of Proposition 4.1 we will show that this set is non-empty for n large enough. The bandwidth h_n^* is an oracle in the sense that it is not accessible by the statistician since it depends on the unknown parameter f. The definition of h_n^* provides some kind of ideal criterion: The bandwidth h is increased along the grid $\mathcal{H}_n$ as long as the bias term $|f_\eta(t) - f(t)|$ is bounded by the 'rate' $v(h)\lambda(h)$, a procedure that aims at mimicking the classical bias-variance tradeoff. In order to state a risk bound for the pseudo estimator $\widehat{f}_{h_n^*}$, we further define

$$r_n(t,f) = \inf_{\underline{h}_n \leqslant h \leqslant 1}\left[\sup_{0\leqslant\eta\leqslant h}(f_\eta(t) - f(t))^2 + \frac{M\int K^2(u)\mathrm{d}u\log(n)}{nh} + \frac{C^2_{\alpha'\beta'}\log(n)}{nh^2}\right].$$

Proposition 4.1. *Consider the pseudo-estimator $\widehat{f}_{h_n^*}$ defined via (4) and (15) where α and β are replaced with α' and β', respectively. Assume that*

$$\lim_{h\to 0}\frac{1}{h}\int K\left(\frac{x-t}{h}\right)f(x)\mathrm{d}x = f(t). \tag{16}$$

Choose $\overline{h}_n = 1$. Then, for n sufficiently large,

$$\mathbf{E}[(\widehat{f}_{h_n^*}(t) - f(t))^2] \leqslant \frac{5}{4}v^2(h_n^*)\lambda^2(h_n^*) \leqslant C(a)r_n(t,f)$$

uniformly for all f with $\|f\|_\infty \leqslant M$.

Remark 4.2. Assumption (16) is satisfied in many cases. For instance, if $\int|K(u)|\mathrm{d}u < \infty$, then (16) is a special case of Bochner's lemma (see [23], Lemma 1.1). However, the sinc-kernel is not absolutely integrable and thus Bochner's lemma cannot be applied. In this case, one can alternatively assume that f belongs at least to some Sobolev space $\mathcal{S}(s,L)$ for some $s > 1/2$. Then, the analysis of the bias term as in the proof of Proposition 3.1 guarantees the validity of (16).

The pseudo estimator $\widehat{f}_{h_n^*}$ is a stopover on our road to an adaptive estimator. We now construct a genuine estimator of f that aims at mimicking this oracle. For this, we first define

$$v^2(h,\eta) = \frac{M}{n}\int (K_h(u) - K_\eta(u))^2 \mathrm{d}u + \frac{C^2_{\alpha'\beta'}}{nh^2} + \frac{C^2_{\alpha'\beta'}}{n\eta^2}.$$

Then, calculations similar to those in the proof of Proposition 3.1 show that

$$\mathrm{Var}(\widehat{f}_h - \widehat{f}_\eta) \leqslant v^2(h,\eta)$$

if $\|f\|_\infty \leqslant M$. For $h, \eta \in \mathcal{H}_n$, put

$$\psi(h,\eta) = v(h)\lambda(h) + v(h,\eta)\lambda(\eta).$$

Then, we define an adaptive choice of the bandwidth parameter by

$$\widehat{h}_n = \max\{h \in \mathcal{H}_n : |\widehat{f}_h(t) - \widehat{f}_\eta(t)| \leqslant \psi(h,\eta) \text{ for all } \eta \leqslant h, \eta \in \mathcal{H}_n\}. \tag{17}$$

This choice of the bandwidth is well-defined since the maximum is taken over a non-empty set. The definition of $\widehat{h}_n$ is characteristic for Lepski's method [19], and the motivation of this procedure is neatly described in [7], p. 67: One chooses the largest bandwidth h such that the difference between the two estimators $\widehat{f}_h$ and $\widehat{f}_\eta$ is not too large (in the sense of (17)) for all $\eta \leqslant h$. Evidently, the motivation of this procedure is to mimick the trade-off between squared bias and variance in a purely data-driven manner. Note also that (17) provides, as well as the oracle version (15), a local choice of the bandwidth in the sense that $\widehat{h}_n$ depends on t. Such a local criterion might result in a better adaptation to spatial inhomogeneity of the target density than global selection rules.

Theorem 4.3. *Consider the estimator $\widehat{f}_{\widehat{h}_n}$ defined via (4) and (17) where $Z_{i,h}(t)$ for $h \in \mathcal{H}_n$ are defined via (3) with α and β replaced with α' and β', respectively. Then, uniformly for all f with $\|f\|_\infty \leqslant M$,*

$$\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - f(t))^2] \leqslant C(a)v^2(h_n^*)\lambda^2(h_n^*).$$

As a consequence, taking $\overline{h}_n = 1$, we obtain

$$\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - f(t))^2] \leqslant C(a)r_n(t,f).$$

The following corollary is obtained by specializing Theorem 4.3 with the sinc-kernel and $\overline{h}_n = 1$. Note that a logarithmic loss for adaptation is commonly accepted and even known to be indispensable for pointwise estimation in the non-private framework [2].

Corollary 4.4. *Consider the estimator $\widehat{f}_{\widehat{h}_n}$ defined via (4) and (17) where $Z_{i,h}(t)$ for $h \in \mathcal{H}_n$ are defined via (3) for the* sinc*-kernel with α and β replaced with α' and β', respectively. Then,*

$$\sup_{\substack{f \in \mathcal{S}(s,L) \\ \|f\|_\infty \leqslant M}} \mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - f(t))^2] \leqslant C(\alpha', \beta', M, L)\left(\frac{n}{\log n}\right)^{-\frac{2s-1}{2s+1}}.$$

4.2. Adaptive estimation for approximate differential privacy

We now state the adaptation results for approximate differential privacy that are obtained via Lepski's method in analogy to the ones for pure differential privacy. For this, we have to redefine some of the quantities from the previous subsection. Taking $\mathcal{H}_n$ as defined in (14) with the exception that we now take $\overline{h}_n$ and $\underline{h}_n$ satisfying $\log(\overline{h}_n n)/n \leqslant \overline{h}_n \leqslant 1$ and $\underline{h}_n = (\log(\overline{h}_n n) \vee 1)/n$. We put (defining $\lambda(h) = \max(1, (\kappa \log(\overline{h}_n/h))^{1/2})$ as above)

$$
\begin{aligned}
&v^2(h) = \frac{M \int K_n^2(u)\mathrm{d}u}{nh},\\
&h_n^* = \max\{h \in \mathcal{H}_n : |f_\eta(t) - f(t)| \leqslant \frac{v(h)\lambda(h)}{2} \text{ for all } \eta \in \mathcal{H}_n, \eta \leqslant h\}\\
&r_n(t,f) = \inf_{\underline{h}_n \leqslant h \leqslant 1} \left[\sup_{0 \leqslant \eta \leqslant h} (f_\eta(t) - f(t))^2 + \frac{M \int K_n^2(u)\mathrm{d}u \log(n)}{nh} \right],\\
&v^2(h,\eta) = \frac{M}{n} \int (K_{n,h}(u) - K_{n,\eta}(u))^2 \mathrm{d}u\\
&\psi(h,\eta) = v(h)\lambda(h) + v(h,\eta)\lambda(\eta), \quad \text{and}\\
&\widehat{h}_n = \max\{h \in \mathcal{H}_n : |\widehat{f}_h(t) - \widehat{f}_\eta(t)| \leqslant \psi(h,\eta) \text{ for all } \eta \leqslant h, \eta \in \mathcal{H}_n\}. \qquad (18)
\end{aligned}
$$

With this redefinition, we obtain the following result.

Theorem 4.5. *Assume $X_i \in [0,1]$. Consider the estimator $\widehat{f}_{\widehat{h}_n}$ defined via (8) and (18) with Z_i defined in (2) and Y_i as in (6). Then, uniformly for all f with $\|f\|_\infty \leqslant M$,*

$$\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - f(t))^2] \lesssim v^2(h_n^*)\lambda^2(h_n^*).$$

As a consequence, taking $\overline{h}_n = 1$, we obtain

$$\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - f(t))^2] \lesssim r_n(t,f).$$

Corollary 4.6. *Consider the estimator $\widehat{f}_{\widehat{h}_n}$ defined via (8) and (18) where $Z_i(t)$ is defined via (2) with Y_i as in (6). Then,*

$$\sup_{\substack{f \in \mathcal{S}(s,L) \\ \|f\|_\infty \leqslant M}} \mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - f(t))^2] \leqslant C(\alpha,\beta,M,L) \left(\frac{n}{\log n}\right)^{-\frac{2s-1}{2s}}.$$

For *a priori* not bounded X_i we can consider the estimation procedure considered in Propostion 3.4. In the definitions preceeding Theorem 4.5, we replace the deconvolution-type kernel K_n by a general kernel in all of the quantities.

Theorem 4.7. *Consider the estimator $\widehat{f}_{\widehat{h}_n}$ defined via (11) and (18) with Z_i defined in (2) and $Y_i \sim g$ for some distribution g. Then, uniformly for all f with $\|f\|_\infty \leqslant M$,*

$$\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - f(t))^2] \lesssim \lambda^2(h_n^*).$$

As a consequence, taking $\overline{h}_n = 1$*, we obtain*

$$\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - f(t))^2] \lesssim r_n(t, f).$$

Corollary 4.8. *Consider the estimator* $\widehat{f}_{\widehat{h}_n}$ *defined via* (11) *and* (18) *where* $Z_i(t)$ *is defined via* (2) *with* $Y_i \sim g$ *for some infinitely smooth distribution (that is,* g *belongs to* $\mathcal{S}(s, L)$ *for some* L *and all* $s > 0$*; for instance, one can take* g *as the density of a standard normal). Then,*

$$\sup_{\substack{f \in \mathcal{S}(s,L) \\ \|f\|_\infty \leqslant M}} \mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - f(t))^2] \leqslant C(\alpha, \beta, M, L) \left(\frac{n}{\log n}\right)^{-\frac{2s-1}{2s}}.$$

5. Discussion

We have investigated the optimal rates of convergence for pointwise estimation of a probability density over Sobolev classes for both pure and approximate local differential privacy. We have found two regimes of convergence rates: for approximate differential privacy the rate of convergence is of the same order $n^{-\frac{2s-1}{2s}}$ as in the case of non-private observations, whereas the rate is deteriorated to $n^{-\frac{2s-1}{2s+1}}$ for pure differential privacy. We have suggested approaches to adaptive kernel density estimation via Lepski's method in the framework of local differential privacy for both setups. Although we have studied its theoretical properties in the prototypical example of univariate density estimation only, our methodology should be transferable to the multivariate case. We also conjecture that it might be possible to extend our results to the case of general linear functionals (different from pointwise evaluation of the density function at a fixed point) as investigated in [15] via Lepski's method in a inverse problem setup. Moreover, the optimal power of the logarithmic factor in the adaptive rate of convergence deserves further investigation. We have also pointed out the potentially misleading conception of (α, β)-differential privacy for the case of strictly positive β since it does not rule out mechanisms where the original data are published and the observer is even aware of this fact. This point certainly deserves conceptional work in the future since approximate differential privacy is receiving rather large interest in the computer science and official statistics literature. Note that this point of criticism is not only valid for local privacy but also in the (even more popular) setup of global privacy (see, for instance, [6] for a rigorous defintion of privacy in this case): the mechanism that publishes a whole original database with positive probability β and the symbol $\emptyset$ with probablity $1 - \beta$ guarantees (α, β)-differential privacy but is certainly not admissible in practise when dealing with really sensitive data. Hence, a notion that on the hand lessens the restrictive property of strict $(\alpha, 0)$-differential privacy but on the other hand rules out non-acceptable procedures is desirable. Finally, note that pointwise rates of convergence have in the non-privacy setup also been studied for wavelet estimators [5]. Transferring such results to privacy setups (using, for instance, the anonymization techniques suggested in [4]) and investigating their properties provides another direction for future research.

Appendix A: Proofs of Section 2

A.1. Proof of Proposition 2.2

Let $A \in \mathscr{B}(\mathbb{R})$ be arbitrary. It has to be shown that

$$\int_A \frac{1}{2b} \exp\left(-\frac{|z-g(x)|}{b}\right) \mathrm{d}z \leqslant e^{\alpha} \int_A \frac{1}{2b} \exp\left(-\frac{|z-g(x')|}{b}\right) \mathrm{d}z + \beta$$

for any $x, x' \in \mathfrak{X}$. By the triangle inequality this holds true if

$$\int_A \frac{1}{2b} \exp\left(-\frac{|z-g(x)|}{b}\right) \mathrm{d}z \leqslant e^{\alpha - \frac{|g(x)-g(x')|}{b}} \int_A \frac{1}{2b} \exp\left(-\frac{|z-g(x)|}{b}\right) \mathrm{d}z + \beta,$$

and the latter holds true as soon as $1 \leqslant \exp(\alpha - \Delta(g)/b) + \beta$ which is equivalent to $b \geqslant \Delta(g)/(\alpha - \log(1-\beta))$.

A.2. Proof of Proposition 2.4

Let $A \in \mathscr{B}(\mathbb{R})$ be arbitrary. Then, the condition on approximate differential privacy reads

$$\beta \mathbf{1}_{\{x\}}(A) + (1-\beta)\widetilde{Q}(A \mid X = x) \leqslant e^{\alpha}(\beta \mathbf{1}_{\{x'\}}(A) + (1-\beta)\widetilde{Q}(A \mid X = x')) + \beta$$

for all $x, x' \in \mathfrak{X}$. This is certainly satisfied if

$$\beta + (1-\beta)\widetilde{Q}(A \mid X = x) \leqslant e^{\alpha}(1-\beta)\widetilde{Q}(A \mid X = x')) + \beta$$

which in turn holds true for any $(\alpha, 0)$-differentially private mechanism $\widetilde{Q}$.

A.3. Proof of Lemma 2.5

Let $A \in \mathscr{Z}_1 \otimes \mathscr{Z}_2$ be a measurable set. Denote $A_{z_1} = \{z_2 \in \mathfrak{Z}_2 : (z_1, z_2) \in A\}$ which is measurable. By Cavalieri's principle and the independence assumption

$$\begin{aligned}
\mathbf{P}^{Z|X=x}(A) &= \int_{\mathfrak{Z}_1} \mathbf{P}^{Z_2|X=x}(A_{z_1}) \mathbf{P}^{Z_1|X=x}(\mathrm{d}z_1) \\
&\leqslant \int_{\mathfrak{Z}_1} (e^{\alpha_2} \mathbf{P}^{Z_2|X=x'}(A_{z_1}) \wedge 1 + \beta_2) \mathbf{P}^{Z_1|X=x}(\mathrm{d}z_1) \\
&= \int_{\mathfrak{Z}_1} (e^{\alpha_2} \mathbf{P}^{Z_2|X=x'}(A_{z_1}) \wedge 1) \mathbf{P}^{Z_1|X=x}(\mathrm{d}z_1) + \int_{\mathfrak{Z}_1} \beta_2 \mathbf{P}^{Z_1|X=x}(\mathrm{d}z_1).
\end{aligned}$$

Now put $\Omega = \{\mathrm{d}\mathbf{P}^{Z_1|X=x}/\mathrm{d}\mathbf{P}^{Z_1|X=x'} \leqslant e^{\alpha_1}\} \subseteq \mathfrak{Z}_1$. Then $\mathbf{P}^{Z_1|X=x}(\Omega^{\mathrm{c}}) \leqslant \beta_1$ since otherwise there would be a contradiction to approximate differential pri-

vacy. Hence,

$$\begin{aligned}\mathbf{P}^{Z|X=x}(A) &\leqslant \int_{\mathfrak{Z}_1\cap\Omega} e^{\alpha_1+\alpha_2}\mathbf{P}^{Z_2|X=x'}(A_{z_1})\mathbf{P}^{Z_1|X=x'}(\mathrm{d}z_1)\\ &\quad+\int_{\mathfrak{Z}_1\cap\Omega^{\mathsf{c}}}\mathbf{P}^{Z_1|X=x}(\mathrm{d}z_1)+\beta_2\\ &\leqslant e^{\alpha_1+\alpha_2}\mathbf{P}^{Z|X=x'}(A)+\mathbf{P}^{Z_1|X=x}(\Omega^{\mathsf{c}})+\beta_2\\ &\leqslant e^{\alpha_1+\alpha_2}\mathbf{P}^{Z|X=x'}(A)+\beta_1+\beta_2\end{aligned}$$

which shows the claim assertion.

Appendix B: Proofs of Section 3

B.1. Proof of Proposition 3.1

The bias-variance decomposition for the estimator $\widehat{f}_h(t)$ is

$$\mathbf{E}[(\widehat{f}_h(t)-f(t))^2]=(f_h(t)-f(t))^2+\mathbf{E}[(\widehat{f}_h(t)-f_h(t))^2]$$

where $f_h(t)=\mathbf{E}[\widehat{f}_h(t)]$. We begin with the analysis of the bias. First recall that

$$f(t)=\frac{1}{2\pi}\int e^{-\mathrm{i}t\omega}\mathcal{F}[f](\omega)\mathrm{d}\omega,$$

and due to centredness of the error added by the privacy mechanism

$$\begin{aligned}f_h(t)&=\int\frac{1}{h}K_{\mathrm{sinc}}\left(\frac{u-t}{h}\right)f(u)\mathrm{d}u\\ &=\frac{1}{2\pi h}\int\mathcal{F}\left[K_{\mathrm{sinc}}\left(\frac{\cdot-t}{h}\right)\right](\omega)\mathcal{F}[f](\omega)\mathrm{d}\omega\\ &=\frac{1}{2\pi}\int e^{-\mathrm{i}t\omega}\mathcal{F}\left[K_{\mathrm{sinc}}\right](h\omega)\mathcal{F}[f](\omega)\mathrm{d}\omega.\end{aligned}$$

Thus, using that $\mathcal{F}\left[K_{\mathrm{sinc}}\right](\cdot)=\mathbf{1}_{[-\pi,\pi]}(\cdot)$, we obtain

$$\begin{aligned}(f_h(t)-f(t))^2&=\frac{1}{4\pi^2}\left(\int_{\mathbb{R}}e^{-\mathrm{i}t\omega}\left[1-\mathcal{F}\left[K_{\mathrm{sinc}}\right](h\omega)\right]\mathcal{F}[f](\omega)\mathrm{d}\omega\right)^2\\ &=\frac{1}{4\pi^2}\left(\int_{\mathbb{R}}e^{-\mathrm{i}t\omega}\mathbf{1}_{\{|\omega|>1/h\}}\mathcal{F}[f](\omega)\mathrm{d}\omega\right)^2\\ &\leqslant\frac{1}{4\pi^2}\int|\mathcal{F}[f](\omega)|^2|\omega|^{2s}\mathrm{d}\omega\cdot\int\mathbf{1}_{\{|\omega|>1/h\}}|\omega|^{-2s}\mathrm{d}\omega\\ &\leqslant\frac{2\pi L^2}{4\pi^2}\cdot\frac{2}{2s-1}h^{2s-1}=C(L,s)h^{2s-1}.\end{aligned}$$

Let us now consider the variance. We denote

$$\widetilde{f}_h(t) = \frac{1}{nh}\sum_{i=1}^{n} K_{\mathrm{sinc}}\left(\frac{X_i - t}{h}\right).$$

By denoting $\xi \sim \mathcal{L}(1)$, we have

$$\begin{aligned}\mathbf{E}[(\widehat{f}_h(t) - f_h(t))^2] &= \mathrm{Var}(\widetilde{f}_h) + \frac{1}{n}\,\mathrm{Var}\left(\frac{\Delta(K_{\mathrm{sinc}}((t-\cdot)/h)/h)}{\alpha - \log(1-\beta)}\,\xi\right)\\ &\leqslant \frac{\|f\|_\infty \int K_{\mathrm{sinc}}^2(u)\mathrm{d}u}{nh} + \frac{8\|K_{\mathrm{sinc}}\|_\infty^2}{nh^2(\alpha - \log(1-\beta))^2}\\ &\leqslant C(\|f\|_\infty, K_{\mathrm{sinc}}, \alpha, \beta)\left[\frac{1}{nh} + \frac{1}{nh^2}\right].\end{aligned}$$

The statement of the proposition follows now by combining the obtained bounds for squared bias and variance.

B.2. Proof of Proposition 3.3

By the very definition of the kernel K_n, we have with $f_h(t) := \mathbf{E}[\widehat{f}_h(t)]$

$$\begin{aligned}f_h(t) &= \int \frac{1}{h} K_n\left(\frac{u-t}{h}\right)\varphi(u)\mathrm{d}u\\ &= \frac{1}{2\pi h}\int \mathcal{F}\left[K_n\left(\frac{\cdot - t}{h}\right)\right](\omega)\mathcal{F}[\varphi](\omega)\mathrm{d}\omega\\ &= \frac{1}{2\pi h}\int \mathcal{F}\left[K_n\left(\frac{\cdot - t}{h}\right)\right](\omega)\mathcal{F}[f](\omega)(\beta + (1-\beta)\mathcal{F}[g](\omega))\mathrm{d}\omega\\ &= \frac{1}{2\pi}\int \frac{\mathcal{F}\left[K_{\mathrm{sinc}}(\cdot)\right](h\omega)}{\beta + (1-\beta)\mathcal{F}[g](\omega)}\mathcal{F}[f](\omega)(\beta + (1-\beta)\mathcal{F}[g](\omega))\mathrm{d}\omega\\ &= \frac{1}{2\pi}\int e^{-\mathrm{i}t\omega}\mathcal{F}\left[K_{\mathrm{sinc}}\right](h\omega)\mathcal{F}[f](\omega)\mathrm{d}\omega.\end{aligned}$$

This is the same expression as in the proof of Proposition 3.1 and we obtain the same bound for the squared bias as in this proposition (note that this bound does not depend on β).

In order to study the variance, note that

$$\begin{aligned}K_n(t) &= \frac{1}{2\pi}\int e^{-\mathrm{i}t\omega}\mathcal{F}[K_n](\omega)\mathrm{d}\omega\\ &= \frac{1}{2\pi}\int e^{-\mathrm{i}t\omega}\mathcal{F}[K_{\mathrm{sinc}}](\omega)\cdot(\beta + (1-\beta)\mathcal{F}[g](\omega/h))^{-1}\,\mathrm{d}\omega.\end{aligned}$$

Hence, since $\mathcal{F}[g] \geqslant 0$,

$$\int |K_n(t)|^2\mathrm{d}t = \frac{1}{2\pi}\int |\mathcal{F}[K_n](\omega)|^2\mathrm{d}\omega \lesssim \beta^{-2}.$$

Using this estimate, we obtain the following bound for the variance term:

$$\mathrm{Var}(\widehat{f}_h) \leqslant \frac{C(\|f\|_\infty)}{nh} \cdot \beta^{-2}.$$

The proof of the proposition follows now by combining the bounds for bias and variance.

B.3. Proof of Proposition 3.5

Let $\widehat{\ell}, Q \in \mathcal{Q}_\alpha$ be arbitrary as in the statement of the proposition. Define $\psi_n > 0$ via $\psi_n^2 = n^{-\frac{2s-1}{2s+1}}$. Let $f_{0,n}$, $f_{1,n}$ be two functions in $\mathcal{S}(s, L)$ (to be specified later on) such that $(f_{0,n}(t) - f_{1,n}(t))^2 \gtrsim \psi_n^2$. Using a general reduction argument (see [24], Section 2.2) it can be shown that

$$\begin{aligned}\sup_{f \in \mathcal{S}(s,L)} \psi_n^{-2} \mathbf{E}[(\widehat{\ell} - f(t))^2] &\geqslant \psi_n^{-2} \sup_{\theta \in \{0,1\}} \mathbf{E}[(\widehat{\ell} - f_{\theta,n}(t))^2] \\ &\gtrsim \psi_n^{-2} \inf_\tau \max_{\theta \in \{0,1\}} \mathbf{P}_\theta(\tau = 1 - \theta)\end{aligned}$$

where the infimum is taken over all $\{0,1\}$-valued test functions τ based on the observations $Z_1, \ldots, Z_n$ and $\mathbf{P}_\theta$ denotes the distribution of $Z_1, \ldots, Z_n$ if the true density of $X_1, \ldots, X_n$ is $f_{\theta,n}$. In view of [24], Theorem 2.2, Statement (iii), the claim assertion follows if we can choose the functions $f_{0,n}$ and $f_{1,n}$ such that

(1) $f_{0,n}, f_{1,n} \in \mathcal{S}(s, L)$,

(2) $(f_{0,n}(t) - f_{1,n}(t))^2 \gtrsim \psi_n^2$, and

(3) $\mathrm{KL}(\mathbf{P}_0, \mathbf{P}_1) \leqslant C < \infty$ for some C independent of n.

To construct such $f_{0,n}, f_{1,n}$ we use ideas from Section 6 of [3] and refer to this paper also for some of the computations. First, take a strictly positive probability density f on $\mathbb{R}$ that is infinitely often continously differentiable. Setting $\|f^{(s)}\|_2^2 = \frac{1}{2\pi} \int_{\mathbb{R}} |\mathcal{F}[f](\omega)|^2 |\omega|^{2s} \mathrm{d}\omega$, we can further assume that $\|f^{(s)}\|_2 \leqslant L$. Then, for $\delta \in (0, 1/2)$, define the function $f_{0,n}$ by

$$f_{0,n}(x) = f_0(x) = \left(\frac{\delta}{2}\right)^{\frac{1}{s+1/2}} f\left(x \left(\frac{\delta}{2}\right)^{\frac{1}{s+1/2}}\right).$$

In order to define the second hypothesis $f_{1,n}$ we consider the auxiliary function $\widetilde{K}_s$ as introduced on p. 26 of [3] (its construction in that paper is borrowed from [22]). In particular, note that $\widetilde{K}_s$ is compactly supported and satisfies $\|\widetilde{K}_s^{(s)}\|_2 \leqslant 1 - \delta/2$ (thus $\widetilde{K}_s \in \mathcal{S}(s, 1)$) and $\widetilde{K}_s(0) \geqslant (1 - \delta) C(s) > 0$. Set $h_n = (n(\exp(\alpha) - 1)^2)^{-1/(2s+1)}$, and put

$$g_{n,s}(x) = cLh_n^{s-\frac{1}{2}} \widetilde{K}\left(\frac{x - t}{h_n}\right)$$

for some constant $c > 0$. Defining $\gamma_{n,s} = \int g_{n,s}(x)\mathrm{d}x < \infty$, set

$$f_{n,1}(x) = f_{n,0}(x)(1 - \gamma_{n,s}) + g_{n,s}(x).$$

We now check conditions (1)–(3) from above.

Verification of (1): The proof follows step by step along the lines of the one in [3] and we omit the details. We only record the fact that

$$\gamma_{n,s} = cLh_n^{s+\frac{1}{2}} \int \widetilde{K}_s(u)\mathrm{d}u = O(h_n^{s+\frac{1}{2}})$$

which will be used below.

Verification of (2): We have

$$\begin{aligned}(f_{n,0}(t) - f_{n,1}(t))^2 &= (f_{n,0}(t) - (1-\gamma_n)f_{0,n}(t) - g(t))^2 \\ &= |\gamma_{n,s}f_{0,n}(t) - g(t)|^2 \\ &\geqslant ||g(t)| - |\gamma_{n,s}f_{0,n}(t)||^2.\end{aligned}$$

Now, since $g_n(t) = Ch_n^{s-\frac{1}{2}}$ and $\gamma_n = O(h_n^{s+\frac{1}{2}})$, the last expression inside the outer absolute values is greater than $Ch_n^{s-\frac{1}{2}}$ for sufficiently large n, say $n \geqslant n_0$. Hence for $n \geqslant n_0$,

$$(f_{n,0}(t) - f_{n,1}(t))^2 \geqslant Ch_n^{2s-1} = C(\alpha)n^{-\frac{2s-1}{2s+1}}$$

which is the desired bound.

Verification of (3): By Equation (14) in [11] we have

$$\mathrm{KL}(\mathbf{P}_0, \mathbf{P}_1) \leqslant 4n(\exp(\alpha) - 1)^2\mathrm{TV}^2(\mathbf{P}_0^{X_1}, \mathbf{P}_1^{X_1}). \tag{19}$$

Now

$$\begin{aligned}\mathrm{TV}(\mathbf{P}_0^{X_1}, \mathbf{P}_1^{X_1}) &= \int |f_{n,0}(x) - f_{n,1}(x)|\mathrm{d}x \\ &= \int_{\mathbb{R}} |-\gamma_{n,s}f_{n,0}(x) + g_n(x)|\mathrm{d}x \\ &\leqslant \gamma_{n,s}\int |f_{n,0}(x)|\mathrm{d}x + \int_{\mathbb{R}} |g_n(x)|\mathrm{d}x \\ &\leqslant O(h_n^{s+\frac{1}{2}}) + Ch_n^{s-\frac{1}{2}} \int \widetilde{K}\left(\frac{x-t}{h_n}\right)\mathrm{d}x \\ &\leqslant O(h_n^{s+\frac{1}{2}}) + C(n(\exp(\alpha)-1)^2)^{-\frac{1}{2}} \int \widetilde{K}(u)\mathrm{d}u \\ &\leqslant C(n(\exp(\alpha)-1)^2)^{-\frac{1}{2}}.\end{aligned}$$

for n sufficiently large. Thus, by (19), for n sufficiently large

$$\mathrm{KL}(\mathbf{P}_0, \mathbf{P}_1) \leqslant Cn(\exp(\alpha) - 1)^2\mathrm{TV}^2(\mathbf{P}_0^{X_1}, \mathbf{P}_1^{X_1}) \leqslant C.$$

Appendix C: Proofs of Section 4

C.1. Proof of Proposition 4.1

Under Assumption (16), we have that $\sup_{0<\eta\leqslant h}|f_\eta(t)-f(t)|^2$ converges to zero as $h\to 0$. Let $n\geqslant 3$. By definition of $v^2(\cdot)$, $\lambda(\cdot)$ and $\underline{h}_n=\log(\sqrt{n})/\sqrt{n}$ (since $\overline{h}_n=1$),

$$v^2(\underline{h}_n)\lambda^2(\underline{h}_n)\geqslant\frac{M\int K^2(u)\mathrm{d}u}{\sqrt{n}\log(\sqrt{n})}+\frac{C^2_{\alpha'\beta'}}{\log(\sqrt{n})}\cdot\kappa\log(\sqrt{n}/\log(\sqrt{n})),$$

hence $\liminf_{n\to\infty}v(\underline{h}_n)\lambda(\underline{h}_n)>0$, and the set in the definition of h_n^* is non-empty provided that n is sufficiently large. Now, the bias-variance decomposition of the pseudo estimator is

$$\begin{aligned}\mathbf{E}[(\widehat{f}_{h_n^*}(t)-f(t))^2]&=(f_{h_n^*}(t)-f(t))^2+\mathrm{Var}_f(\widehat{f}_{h_n^*})\\&\leqslant(f_{h_n^*}(t)-f(t))^2+v^2(h_n^*)\\&\leqslant\frac{v^2(h_n^*)\lambda^2(h_n^*)}{4}+v^2(h_n^*)\\&\leqslant\frac{5}{4}v^2(h_n^*)\lambda^2(h_n^*).\end{aligned}$$

Let now h_0 be the minimizer in the definition of $r_n(t,f)$. We distinguish the cases $h_0<ah_n^*$ and $h_0\geqslant ah_n^*$. First, if $h_0<ah_n^*$, then

$$\begin{aligned}r_n(t,f)&=\sup_{0\leqslant\eta\leqslant h_0}(f_\eta(x)-f(x))^2+\frac{M\int K^2(u)\mathrm{d}u\log(n)}{nh_0}+\frac{C^2_{\alpha'\beta'}\log(n)}{nh_0^2}\\&\geqslant\frac{M\int K^2(u)\mathrm{d}u\log(n)}{nh_0}+\frac{C^2_{\alpha'\beta'}\log(n)}{nh_0^2}\\&\geqslant\frac{M\int K^2(u)\mathrm{d}u\log(n)}{anh_n^*}+\frac{C^2_{\alpha'\beta'}\log(n)}{na^2(h_n^*)^2}\\&\geqslant C(a,\kappa)v^2(h_n^*)\lambda^2(h_n^*).\end{aligned}$$

If $h_0\geqslant ah_n^*$, then by the very definition of h_n^* we obtain

$$r_n(t,f)\geqslant\sup_{0\leqslant\eta\leqslant h_0}(f_\eta(t)-f(t))^2\geqslant\sup_{0\leqslant\eta\leqslant ah_n^*}(f_\eta(t)-f(t))^2>\frac{v^2(ah_n^*)\lambda^2(ah_n^*)}{4},$$

and thus $r_n(t,f)\gtrsim v^2(h_n^*)\lambda^2(h_n^*)$ also in this case.

C.2. Proof of Theorem 4.3

We consider the risk decomposition

$$\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t)-f(t))^2]=\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t)-f(t))^2\mathbf{1}_{\{\widehat{h}_n\geqslant h_n^*\}}]+\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t)-f(t))^2\mathbf{1}_{\{\widehat{h}_n<h_n^*\}}],$$

and study the two terms on the right-hand side separately.

Analysis of the first term (Case $\widehat{h}_n \geqslant h_n^*$). Note that the quantities $v(\cdot), \lambda(\cdot)$ satisfy $v(h) \geqslant v(h')$ and $\lambda(h) \geqslant \lambda(h')$ for $h' \geqslant h$. Thus, using the inequality $(a+b)^2 \leqslant 2a^2 + 2b^2$, we have for $h \leqslant h'$ that

$$\begin{aligned}\psi(h',h) &= v(h')\lambda(h') + v(h',h)\lambda(h) \\ &\leqslant v(h)\lambda(h) + 2\sqrt{2}v(h)\lambda(h) \\ &= (1+2\sqrt{2})v(h)\lambda(h).\end{aligned}$$

By the definition of ψ and $\widehat{h}_n$, we obtain

$$\begin{aligned}|\widehat{f}_{\widehat{h}_n}(t) - \widehat{f}_{h_n^*}(t)|\mathbf{1}_{\{\widehat{h}_n \geqslant h_n^*\}} &\leqslant \psi(\widehat{h}_n, h_n^*) \\ &\leqslant \sup\{\psi(\eta, h_n^*) : \eta \in \mathcal{H}_n, \eta \geqslant h_n^*\} \\ &\leqslant (1+2\sqrt{2})v(h_n^*)\lambda(h_n^*).\end{aligned}$$

Hence (recall that we denote $f_h(t) = \mathbf{E}[\widehat{f}_h(t)]$),

$$\begin{aligned}&\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - f(t))^2 \mathbf{1}_{\{\widehat{h}_n \geqslant h_n^*\}}] \\ &\quad\leqslant 2\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - \widehat{f}_{h_n^*}(t))^2 \mathbf{1}_{\{\widehat{h}_n \geqslant h_n^*\}}] + 2\mathbf{E}[(\widehat{f}_{h_n^*}(t) - f(t))^2] \\ &\quad= 2\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t) - \widehat{f}_{h_n^*}(t))^2 \mathbf{1}_{\{\widehat{h}_n \geqslant h_n^*\}}] + 2\mathbf{E}[(\widehat{f}_{h_n^*}(t) - f_{h_n^*}(t))^2] \\ &\qquad + 2(f_{h_n^*}(t) - f(t))^2 \\ &\quad\lesssim v^2(h_n^*)\lambda^2(h_n^*)\end{aligned}$$

where we used the bound $\mathrm{Var}(\widehat{f}_{h_n^*}) \leqslant v^2(h_n^*)$ for the term $2\mathbf{E}[(\widehat{f}_{h_n^*}(t) - f_{h_n^*}(t))^2]$ and the definition of h_n^* to bound the term $(f_{h_n^*}(t) - f(t))^2$.

Analysis of the second term (Case $\widehat{h}_n < h_n^*$). For $h, \eta \in \mathcal{H}_n$ with $\eta < h$, set

$$B_n(t,h,\eta) = \{|\widehat{f}_h(t) - \widehat{f}_\eta(t)| > \psi(h,\eta)\}.$$

Let h in $\mathcal{H}_n$. Then, by definition of $\widehat{h}_n$,

$$\{\widehat{h}_n = a^{-1}h\} \subseteq \bigcup_{\substack{\eta \in \mathcal{H}_n \\ \eta < h}} B_n(t,h,\eta),$$

and thus

$$\begin{aligned}\{\widehat{h}_n < h_n^*\} &= \bigcup_{\substack{h \in \mathcal{H}_n \\ h < h_n^*}} \{\widehat{h}_n = h\} \\ &\subseteq \bigcup_{\substack{h \in \mathcal{H}_n \\ h < a h_n^*}} \{\widehat{h}_n = a^{-1}h\}\end{aligned}$$

$$= \bigcup_{\substack{h\in\mathcal{H}_n\\ h<ah_n^*}} \bigcup_{\substack{\eta\in\mathcal{H}_n\\ \eta<h}} B_n(t,h,\eta).$$

We obtain

$$\begin{aligned}\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t)-f(t))^2\mathbf{1}_{\{\widehat{h}_n<h_n^*\}}] &\leqslant \sum_{\substack{h\in\mathcal{H}_n\\ h<ah_n^*}} \mathbf{E}[(\widehat{f}_{a^{-1}h}(t)-f(t))^2\mathbf{1}_{\{\widehat{h}_n=a^{-1}h\}}]\\ &\leqslant \sum_{\substack{h\in\mathcal{H}_n\\ h<ah_n^*}} \sum_{\substack{\eta\in\mathcal{H}_n\\ \eta<h}} \mathbf{E}[(\widehat{f}_{a^{-1}h}(t)-f(t))^2\mathbf{1}_{B_n(t,h,\eta)}].\end{aligned}$$

By definition of h_n^*, for all $\eta, h \in \mathcal{H}_n$ with $\eta < h \leqslant h_n^*$, it holds

$$|f_\eta(t)-f(t)| \leqslant \frac{v(h_n^*)\lambda(h_n^*)}{2} \leqslant \frac{v(h)\lambda(h)}{2}.$$

Now, for $\eta < h \leqslant h_n^*$,

$$\begin{aligned}B_n(t,h,\eta) &= \{|\widehat{f}_h(t)-\widehat{f}_\eta(t)|>\psi(h,\eta)\}\\ &= \{|\widehat{f}_h(t)-\widehat{f}_\eta(t)-(f_h(x)-f_\eta(t))+f_h(t)\\ &\qquad -f_\eta(t)-f(t)+f(t)|>\psi(h,\eta)\}\\ &\subseteq \left\{v(h)\lambda(h)+\left|\frac{1}{n}\sum_{i=1}^n\zeta_i\right|>\psi(h,\eta)\right\}\\ &\subseteq \left\{\left|\frac{1}{n}\sum_{i=1}^n\zeta_i\right|>v(h,\eta)\lambda(\eta)\right\}\end{aligned}$$

where $\zeta_i = \zeta_{i,h,\eta} = Z_{i,h}(t) - Z_{i,\eta}(t) - (f_h(t)-f_\eta(t))$. Note that $\mathbf{E}\zeta_i = 0$ and $\mathrm{Var}(\zeta_i) \leqslant nv^2(h,\eta)$. Now, by the Cauchy-Schwarz inequality,

$$\begin{aligned}&\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t)-f(t))^2\mathbf{1}_{\{\widehat{h}_n<h_n^*\}}]\\ &\quad\leqslant \sum_{\substack{h\in\mathcal{H}_n\\ h<ah_n^*}} \sum_{\substack{\eta\in\mathcal{H}_n\\ \eta<h}} \mathbf{E}[(\widehat{f}_{a^{-1}h}(t)-f(t))^2\mathbf{1}_{\{|\frac{1}{n}\sum_{i=1}^n\zeta_i|>v(h,\eta)\lambda(\eta)\}}]\\ &\quad\leqslant \sum_{\substack{h\in\mathcal{H}_n\\ h<ah_n^*}} \sum_{\substack{\eta\in\mathcal{H}_n\\ \eta<h}} \left(\mathbf{E}[(\widehat{f}_{a^{-1}h}(t)-f(t))^4]\right)^{1/2}\left(\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n\zeta_i\right|>v(h,\eta)\lambda(\eta)\right)\right)^{1/2}.\end{aligned}$$

For the first term in the sum, we have

$$\begin{aligned}\mathbf{E}[(\widehat{f}_{a^{-1}h}(t)-f(t))^4] &= \mathbf{E}[(\widehat{f}_{a^{-1}h}(t)-f_{a^{-1}h}(t)+f_{a^{-1}h}(t)-f(t))^4]\\ &\leqslant 8\mathbf{E}[(\widehat{f}_{a^{-1}h}(t)-f_{a^{-1}h}(t))^4]+8(f_{a^{-1}h}(t)-f(t))^4.\end{aligned}$$

Putting $\zeta_i' = Z_{i,a^{-1}h}(t) - f_{a^{-1}h}(t)$, we have

$$\mathbf{E}\left[(\widehat{f}_{a^{-1}h}(t) - f_{a^{-1}h}(t))^4\right] = \mathbf{E}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\zeta_i'\right)^4\right] \leqslant \frac{\mathbf{E}[(\zeta_i')^4]}{n^3} + \frac{3(\mathbf{E}[(\zeta_i')^2])^2}{n^2}.$$

On the one hand,

$$\begin{aligned}\mathbf{E}[(\zeta_i')^4] &\lesssim \frac{C_{\alpha'\beta'}^4}{a^{-4}h^4} + \frac{1}{a^{-4}h^4}\mathbf{E}\left[\left(K\left(\frac{X-t}{a^{-1}h}\right) - \mathbf{E}\left[K\left(\frac{X-t}{a^{-1}h}\right)\right]\right)^4\right] \\ &\lesssim \frac{C_{\alpha'\beta'}^4}{a^{-4}h^4} + \frac{8}{a^{-4}h^4}\mathbf{E}\left[\left(K\left(\frac{X-t}{a^{-1}h}\right)\right)^4\right] + \frac{8}{a^{-4}h^4}\left(\mathbf{E}\left[K\left(\frac{X-t}{a^{-1}h}\right)\right]\right)^4 \\ &\lesssim \frac{1}{a^{-4}h^4} + \frac{1}{a^{-3}h^3} + 1,\end{aligned}$$

on the other hand

$$\mathbf{E}[(\zeta_i')^2] \lesssim \frac{C_{\alpha'\beta'}^2}{a^{-2}h^2} + \frac{1}{a^{-1}h}.$$

Hence,

$$\mathbf{E}[(\widehat{f}_{a^{-1}h}(t) - f_{a^{-1}h}(t))^4] \leqslant Cv^4(a^{-1}h).$$

Moreover, for $a^{-1}h < h_n^*$,

$$(f_{a^{-1}h}(t) - f(t))^4 \leqslant \frac{v^4(h_n^*)\lambda^4(h_n^*)}{16}$$

by the very definition of h_n^*. Thus, altogether,

$$\mathbf{E}[(\widehat{f}_{a^{-1}h}(t) - f(t))^4] \leqslant C(v^4(a^{-1}h) + v^4(h_n^*)\lambda^4(h_n^*)),$$

and by the monotonicity of $v(\cdot)$ and $\lambda(\cdot)$, for $\eta < h \leqslant h_n^*$

$$\mathbf{E}[(\widehat{f}_{a^{-1}h}(t) - f(t))^4] \leqslant C\lambda^4(\eta)v^4(a^{-1}h).$$

Write $\zeta_i = \zeta_i^{(1)} + \zeta_i^{(2)}$ where $\zeta_i^{(1)} = K_h(X_i - t) - K_\eta(X_i - t) - (f_h(t) - f_\eta(t))$ and $\zeta_i^{(2)} = \frac{C_{\alpha'\beta'}}{\sqrt{2}h}\xi_{i,h} + \frac{C_{\alpha'\beta'}}{\sqrt{2}\eta}\xi_{i,\eta}$ with $\xi_{i,h}, \xi_{i,\eta}$ i.i.d.$\sim \mathcal{L}(1)$ for $i = 1, \ldots, n$. We have

$$\begin{aligned}\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i\right| > v(h,\eta)\lambda(\eta)\right) &\leqslant \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i^{(1)}\right| > \frac{v(h,\eta)\lambda(\eta)}{2}\right) \\ &\quad + \mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i^{(2)}\right| > \frac{v(h,\eta)\lambda(\eta)}{2}\right).\end{aligned}$$

Consider $\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i^{(1)}\right| > \frac{v(h,\eta)\lambda(\eta)}{2}\right)$ first. By Bernstein's inequality (see Lemma D.1) with $b = 4\|K\|_\infty/\eta$,

$$\begin{aligned}
&\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i^{(1)}\right| > \frac{v(h,\eta)\lambda(\eta)}{2}\right)\\
&\quad\leqslant 2\max\left\{\exp\left(-\frac{nv^2(h,\eta)\lambda^2(\eta)}{4nv^2(h,\eta)}\right), \exp\left(-\frac{nv(h,\eta)\lambda(\eta)\eta}{32\|K\|_\infty}\right)\right\}\\
&\quad= 2\max\left\{\exp\left(-\frac{\lambda^2(\eta)}{4}\right), \exp\left(-\frac{nv(h,\eta)\lambda(\eta)\eta}{32\|K\|_\infty}\right)\right\}.
\end{aligned}$$

Note that

$$v(h,\eta) \geqslant \frac{C_{\alpha'\beta'}}{\sqrt{n}\eta}.$$

For any $h \in \mathcal{H}_n$ and n large enough, it holds

$$\sqrt{n} \geqslant \sqrt{\kappa}\log(\sqrt{n}) \geqslant \sqrt{\kappa}\log(\overline{h}_n\sqrt{n}) = \sqrt{\kappa}\log(\overline{h}_n/(1/\sqrt{n}))) \geqslant \sqrt{\kappa}\log(\overline{h}_n/h).$$

Thus

$$\begin{aligned}
&\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i^{(1)}\right| > \frac{v(h,\eta)\lambda(\eta)}{2}\right)\\
&\quad\leqslant 2\max\left\{\exp\left(-\frac{\lambda^2(\eta)}{4}\right), \exp\left(-\frac{C_{\alpha'\beta'}\sqrt{n}\lambda(\eta)}{32\|K\|_\infty}\right)\right\}\\
&\quad\leqslant 2\exp\left(-\kappa\left(\frac{1}{4}\wedge\frac{C_{\alpha'\beta'}}{32\|K\|_\infty}\right)\log\left(\frac{\overline{h}_n}{\eta}\right)\right). \qquad (20)
\end{aligned}$$

Let us now consider the probability in terms of $\zeta_i^{(2)}$. We decompose

$$\begin{aligned}
\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\zeta_i^{(2)}\right| > \frac{v(h,\eta)\lambda(\eta)}{2}\right) &\leqslant \mathbf{P}\left(\left|\frac{C_{\alpha'\beta'}}{\sqrt{2}nh}\sum_{i=1}^{n}\xi_{i,h}\right| > \frac{v(h,\eta)\lambda(\eta)}{4}\right)\\
&\quad+\mathbf{P}\left(\left|\frac{C_{\alpha'\beta'}}{\sqrt{2}n\eta}\sum_{i=1}^{n}\xi_{i,\eta}\right| > \frac{v(h,\eta)\lambda(\eta)}{4}\right).
\end{aligned}$$

Consider only the first probability on the right-hand side, the bound for the second one following analogously. By Bernstein's inequality (see Lemma D.1, take the version with control on the moments applied with $t = v(h,\eta)\lambda(\eta)/4$, $v^2 = C_{\alpha'\beta'}^2/h^2$ and $b = C_{\alpha'\beta'}/h$)

$$\begin{aligned}
&\mathbf{P}\left(\left|\frac{C_{\alpha'\beta'}}{\sqrt{2}nh}\sum_{i=1}^{n}\xi_{i,h}\right| > \frac{v(h,\eta)\lambda(\eta)}{4}\right)\\
&\quad\leqslant 2\max\left\{\exp\left(-\frac{nt^2}{4v^2}\right), \exp\left(-\frac{nt}{4b}\right)\right\}
\end{aligned}$$

$$\leqslant 2\max\left\{\exp\left(-\frac{\lambda^2(\eta)}{64}\right),\exp\left(-\frac{\sqrt{n}\lambda(\eta)}{16}\right)\right\},$$

and hence by using $\sqrt{n}\geqslant\sqrt{\kappa}\log(\overline{h}_n/\eta)$,

$$\mathbf{P}\left(\left|\frac{C_{\alpha'\beta'}}{\sqrt{2}nh}\sum_{i=1}^n\xi_{i,h}\right|>\frac{v(h,\eta)\lambda(\eta)}{4}\right)\leqslant 2\exp\left(-\frac{\kappa}{64}\log\left(\frac{\overline{h}_n}{\eta}\right)\right).$$

Finally, we obtain with $\kappa'=\frac{\kappa}{64}\wedge\frac{\kappa C_{\alpha'\beta'}}{32\|K\|_\infty}$ that

$$\mathbf{P}\left(\left|\frac{1}{n}\sum_{i=1}^n\zeta_i^{(2)}\right|>\frac{v(h,\eta)\lambda(\eta)}{2}\right)\leqslant 4\exp\left(-\kappa'\log\left(\frac{\overline{h}_n}{\eta}\right)\right).$$

Now,

$$\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t)-f(t))^2\mathbf{1}_{\{\widehat{h}_n<h_n^*\}}]\lesssim\sum_{\substack{h\in\mathcal{H}\\ h<ah_n^*}}\sum_{\substack{\eta\in\mathcal{H}\\ \eta<h}}\lambda^2(\eta)v^2(a^{-1}h)\exp\left(-\frac{\kappa'}{2}\log\left(\frac{\overline{h}_n}{\eta}\right)\right).$$

For sufficiently small $\gamma>0$[1], we have

$$\begin{aligned}\sum_{\substack{\eta\in\mathcal{H}_n\\ \eta<h}}\lambda^2(\eta)\exp\left(-\frac{\kappa'}{2}\log\left(\frac{\overline{h}_n}{\eta}\right)\right)&\lesssim\left(\frac{h}{\overline{h}_n}\right)^{\kappa'/2-\gamma}\sum_{\substack{\eta\in\mathcal{H}_n\\ \eta<h}}\log\left(\frac{\overline{h}_n}{\eta}\right)\left(\frac{\eta}{\overline{h}_n}\right)^{\gamma}\\&\lesssim\left(\frac{h}{\overline{h}_n}\right)^{\kappa'/2-\gamma}\sum_{j=0}^{\infty}ja^{-\gamma j}\log(a)\\&\lesssim\left(\frac{h}{\overline{h}_n}\right)^{\kappa'/2-\gamma}.\end{aligned}$$

Recall that $v^2(h)\asymp\frac{1}{nh}+\frac{1}{nh^2}$. Thus,

$$\begin{aligned}&\mathbf{E}[(\widehat{f}_{\widehat{h}_n}(t)-f(t))^2\mathbf{1}_{\{\widehat{h}_n<h_n^*\}}]\\&\lesssim\sum_{\substack{h\in\mathcal{H}_n\\ h<ah_n^*}}\left(\frac{h}{\overline{h}_n}\right)^{\kappa'/2-\gamma}v^2(a^{-1}h)\\&\lesssim\frac{\overline{h}_n}{n}\sum_{\substack{h\in\mathcal{H}\\ h<ah_n^*}}\left(\frac{h}{\overline{h}_n}\right)^{\kappa'/2-\gamma-1}+\frac{\overline{h}_n^2}{n\alpha^2}\sum_{\substack{h\in\mathcal{H}\\ h<ah_n^*}}\left(\frac{h}{\overline{h}_n}\right)^{\kappa'/2-\gamma-2}.\end{aligned}$$

The sums on the right-hand side converge and the bound for the case $\widehat{h}_n<h_n^*$ is negligible with respect to the upper bound $v^2(h_n^*)\lambda^2(h_n^*)$.

[1]Our calculations show that $\gamma>0$ has to satisfy also that $\kappa'/2-\gamma-2>0$. Such a choice is possible whenever $\kappa'/2-2>0$ which holds for κ large enough.

Appendix D: Bernstein inequality

The following version of the Bernstein inequality is taken from [9].

Lemma D.1. *Let $X_1, \dots, X_n$ be i.i.d. random variables and put $S_n = \sum_{i=1}^n (X_i - \mathbf{E}[X_i])$. Then, for any $t > 0$,*

$$\begin{aligned}\mathbf{P}(|S_n - \mathbf{E}[S_n]| \geqslant nt) &\leqslant 2\exp\left(-\frac{nt^2}{2v^2 + 2b\eta}\right)\\ &\leqslant 2\max\left\{\exp\left(-\frac{nt^2}{4v^2}\right), \exp\left(-\frac{nt}{4b}\right)\right\}\end{aligned}$$

where $\operatorname{Var}(X_1) \leqslant v^2$ and $|X_1| \leqslant b$ (or $\mathbf{E}[|X_i|^m] \leqslant \frac{m!}{2} v^2 b^{m-2}$ for $m \geqslant 2$).

Acknowledgments

I am thankful to an anonymous referee as well as an AE for their detailed comments that considerably improved the article. Especially the critical discussion of the concept of (α, β)-differential privacy which now appears at several passages of the paper has been motivated by their reviews. I also thank Sandra Schluttenhofer for sending me comments concerning a previous version of this work collected in the framework of a reading group on differential privacy at the Ruprecht-Karls-Universität Heidelberg.

References

[1] R. F. Barber and J. C. Duchi. Privacy and Statistical Risk: Formalisms and Minimax Bounds. *arXiv-preprint, available at* https://arxiv.org/abs/1412.4451v1, 2014.

[2] L. D. Brown and M. G. Low. A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.*, 24(6):2524–2535, 1996. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1032181166. MR1425965

[3] C. Butucea. Exact adaptive pointwise estimation on Sobolev classes of densities. *ESAIM Probab. Statist.*, 5:1–31, 2001. ISSN 1292-8100. URL https://doi.org/10.1051/ps:2001100. MR1845320

[4] C. Butucea, A. Dubois, M. Kroll, and A. Saumard. Local differential privacy: elbow effect in optimal density estimation and adaptation over Besov ellipsoids. *Bernoulli*, 26(3):1727–1764, 2020. ISSN 1350-7265. URL https://doi.org/10.3150/19-BEJ1165. MR4091090

[5] T. T. Cai. Rates of convergence and adaptation over Besov spaces under pointwise risk. *Statist. Sinica*, 13(3):881–902, 2003. ISSN 1017-0405. MR1997178

[6] T. T. Cai, Y. Wang, and L. Zhang. The cost of privacy: Optimal rates of convergence for parameter estimation with differential privacy. *arXiv-preprint, available at* https://arxiv.org/abs/1902.04495, 2020.

[7] L. Cavalier. On the problem of local adaptive estimation in tomography. *Bernoulli*, 7(1):63–78, 2001. ISSN 1350-7265. URL https://doi.org/10.2307/3318602. MR1811744

[8] M. Chen, C. Gao, and Z. Ren. A general decision theory for Huber's ε-contamination model. *Electron. J. Stat.*, 10(2):3752–3774, 2016. URL https://doi.org/10.1214/16-EJS1216. MR3579675

[9] F. Comte. *Estimation non-paramétrique*. Spartacus, Paris, 2015.

[10] D. L. Donoho and M. G. Low. Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.*, 20(2):944–970, 1992. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1176348665. MR1165601

[11] J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Minimax optimal procedures for locally private estimation. *J. Amer. Statist. Assoc.*, 113(521):182–201, 2018. ISSN 0162-1459. URL https://doi.org/10.1080/01621459.2017.1389735. MR3803452

[12] C. Dwork. Differential privacy. In *Automata, languages and programming. Part II*, volume 4052 of *Lecture Notes in Comput. Sci.*, pages 1–12. Springer, Berlin, 2006. URL https://doi.org/10.1007/11787006_1. MR2307219

[13] C. Dwork. Differential privacy: a survey of results. In *Theory and applications of models of computation*, volume 4978 of *Lecture Notes in Comput. Sci.*, pages 1–19. Springer, Berlin, 2008. URL https://doi.org/10.1007/978-3-540-79228-4_1. MR2472670

[14] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography*, volume 3876 of *Lecture Notes in Comput. Sci.*, pages 265–284. Springer, Berlin, 2006. URL https://doi.org/10.1007/11681878_14. MR2241676

[15] A. Goldenshluger and S. V. Pereverzev. Adaptive estimation of linear functionals in Hilbert scales from indirect white noise observations. *Probab. Theory Related Fields*, 118(2):169–186, 2000. ISSN 0178-8051. URL https://doi.org/10.1007/s440-000-8013-3. MR1790080

[16] R. Hall, A. Rinaldo, and L. Wasserman. Random Differential Privacy. *arXiv-preprint, available at* https://arxiv.org/abs/1112.2680, 2011.

[17] R. Hall, A. Rinaldo, and L. Wasserman. Differential privacy for functions and functional data. *Journal of Machine Learning Research*, 14:703–727, 2013. ISSN 1532-4435. MR3033345

[18] N. Holohan, D. J. Leith, and O. Mason. Differential privacy in metric spaces: numerical, categorical and functional data under the one roof. *Inform. Sci.*, 305:256–268, 2015. ISSN 0020-0255. URL https://doi.org/10.1016/j.ins.2015.01.021. MR3317616

[19] O. V. Lepski. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3):459–470, 1990. ISSN 0040-361X. URL https://doi.org/10.1137/1135065. MR1091202

[20] O. V. Lepski and V. G. Spokoiny. Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, 25(6):2512–2546, 1997. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1030741083.

MR1604408

[21] A. Rohde and L. Steinberger. Geometrizing rates of convergence under local differential privacy constraints. *Ann. Statist.*, 48(5):2646–2670, 2020. ISSN 0090-5364. URL https://doi.org/10.1214/19-AOS1901. MR4152116

[22] A. B. Tsybakov. Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.*, 26(6):2420–2469, 1998. ISSN 0090-5364. URL https://doi.org/10.1214/aos/1024691478. MR1700239

[23] A. B. Tsybakov. *Introduction à l'estimation non-paramétrique*. Springer-Verlag, Berlin, 2004. ISBN 3-540-40592-5. MR2013911

[24] A. B. Tsybakov. *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-79051-0. URL https://doi.org/10.1007/b13794. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats. MR2724359

[25] L. Wasserman and S. Zhou. A statistical framework for differential privacy. *J. Amer. Statist. Assoc.*, 105(489):375–389, 2010. ISSN 0162-1459. URL https://doi.org/10.1198/jasa.2009.tm08651. MR2656057