



HAL
open science

Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint

Joseph Lam-Weil, Béatrice Laurent, Jean-Michel Loubes

► To cite this version:

Joseph Lam-Weil, Béatrice Laurent, Jean-Michel Loubes. Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint. 2021. hal-02993462

HAL Id: hal-02993462

<https://hal.science/hal-02993462>

Preprint submitted on 15 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Minimax optimal goodness-of-fit testing for densities and multinomials under a local differential privacy constraint

JOSEPH LAM-WEIL¹, BÉATRICE LAURENT² and JEAN-MICHEL LOUBES³

¹Magdeburg University, Germany, E-mail: joseph.lam@ovgu.de

²INSA de Toulouse, France, E-mail: laurentb@insa-toulouse.fr

³Université de Toulouse, France, E-mail: loubes@math.univ-toulouse.fr

Finding anonymization mechanisms to protect personal data is at the heart of recent machine learning research. Here, we consider the consequences of local differential privacy constraints on goodness-of-fit testing, i.e. the statistical problem assessing whether sample points are generated from a fixed density f_0 , or not. The observations are kept hidden and replaced by a stochastic transformation satisfying the local differential privacy constraint. In this setting, we propose a testing procedure which is based on an estimation of the quadratic distance between the density f of the unobserved samples and f_0 . We establish an upper bound on the separation distance associated with this test, and a matching lower bound on the minimax separation rates of testing under non-interactive privacy in the case that f_0 is uniform, in discrete and continuous settings. To the best of our knowledge, we provide the first minimax optimal test and associated private transformation under a local differential privacy constraint over Besov balls in the continuous setting, quantifying the price to pay for data privacy. We also present a test that is adaptive to the smoothness parameter of the unknown density and remains minimax optimal up to a logarithmic factor. Finally, we note that our results can be translated to the discrete case, where the treatment of probability vectors is shown to be equivalent to that of piecewise constant densities in our setting. That is why we work with a unified setting for both the continuous and the discrete cases.

Keywords: local differential privacy; non-interactive privacy; goodness-of-fit testing; minimax separation rates; continuous and discrete distributions

1. Introduction

Ensuring user privacy is at the core of the development of Artificial Intelligence. Indeed datasets can contain extremely sensitive information, and someone with access to a privatized training set or the outcome of an algorithm should not be able to retrieve the original dataset. However, classical anonymization and cryptographic approaches fail to prevent the disclosure of sensitive information in the context of learning. Indeed, with the example of a hospital's database, removing names and social security numbers from databases does not prevent the identification of patients using a combination of other attributes like gender, age or illnesses. [16] cites Cystic Fibrosis as an example which exist with a frequency of around 1/3000. Hence differential privacy mechanisms were developed to cope with such issues. Such considerations can be traced back to [52, 22, 23, 27]. As early as in 1965, [52] presented the first privacy mechanism which is now a baseline method for binary data: Randomized response. Another important result is presented in the works of [22, 23, 27], where they expose a trade-off between statistical utility, or in other terms performance, and privacy in a limited-disclosure setting.

Differential privacy as expressed in [25, 24] is the most common formalization of the problem of privacy. It can be summed up as the following condition: altering a single data point of the training set only affects the probability of an outcome to a limited degree. One main advantage of such a definition

of privacy is that it can be parametrized by some positive parameter α , where α close to 0 corresponds to a more restrictive privacy condition. This definition treats privacy in a global way with respect to the original dataset, in contrast with the privacy constraint that follows.

We now consider a stronger privacy condition called local differential privacy which also depends on a positive parameter α and where the analyst himself is not trusted with the data. Consider unobserved random variables X_1, \dots, X_n taking values in $[0, 1]$, which are independent and identically distributed (i.i.d.) with density f with respect to the Lebesgue measure. We observe Z_1, \dots, Z_n which are α -local differentially private views of X_1, \dots, X_n . That is, there exist probability measures Q_1, \dots, Q_n such that for all $0 \leq i \leq n$, Z_i is a stochastic transformation of X_i by the channel Q_i and

$$\sup_{S \in \mathcal{Z}_i, z_j \in \tilde{\Omega}_j, (x, x') \in [0, 1]^2} \frac{Q_i(Z_i \in S | X_i = x, Z_j = z_j, j < i)}{Q_i(Z_i \in S | X_i = x', Z_j = z_j, j < i)} \leq e^\alpha, \quad (1)$$

where $Q_i(Z_i \in \tilde{\Omega}_i) = 1$ and \mathcal{Z}_i is a σ -algebra such that $\tilde{\Omega}_i$ is its associated sample space. This notion has been extensively studied through the concept of local algorithms, especially in the context of privacy-preserving data mining [52, 4, 3, 51, 26, 5, 44, 36, 39]. Now note that Equation (1) accounts for possible dependencies between Z_i 's, corresponding to the interactive case. The role of interactivity has been further studied in [37, 13, 10], and it can be complete or sequential. Recent results detailed in [18, 21, 19] give information processing inequalities depending on the local privacy constraint via the parameter α . Those can be used to obtain Fano or Le Cam-type inequalities in order to obtain a minimax lower bound for estimation or testing problems. Our proof also relies on Le Cam's inequality, albeit in a more refined way in order to obtain minimax optimal results.

Testing problems have appeared as crucial tools in machine learning in order to assess whether a model fits the observations or to detect anomalies and novelties. In particular, goodness-of-fit testing is a classical hypothesis testing problem in statistics. It consists in testing whether the density f of n independent and identically distributed (i.i.d.) observations equals a specified density f_0 or not. This motivates our study of goodness-of-fit testing under a local differential privacy constraint.

We want to design our tests so that they reject the null hypothesis $H_0 : f = f_0$ if the data is not actually generated from the given model with a given confidence level. Assuming that f and f_0 belong to $\mathbb{L}_2([0, 1]) = \left\{ f : [0, 1] \rightarrow \mathbb{R}, \|f\|_2^2 = \int_0^1 f^2(x) dx < \infty \right\}$, it is natural to propose a test based on an estimation of the squared \mathbb{L}_2 -distance $\|f - f_0\|_2^2$ between f and f_0 . In order to test whether $f = f_0$ from the observation of an i.i.d sample set (X_1, \dots, X_n) with common density f , [45] introduces an orthonormal basis $\{f_0, \phi_l, l \geq 0\}$ of $\mathbb{L}_2([0, 1])$. The goodness-of-fit hypothesis is rejected if the estimator $\sum_{l=1}^D (\sum_{i=1}^n \phi_l(X_i)/n)^2$ exceeds some threshold, where D is a given integer depending on n . Data-driven versions of this test, where the parameter D is chosen to minimize some penalized criterion have been introduced by [11, 42, 38, 32].

Additionally, we want to find the limitations of a test by determining how close the two hypotheses can get while remaining separated by the testing procedure. This classical problem has been studied under the lens of minimax optimality in the seminal work by [33, 34]. Non-asymptotic performances and an extension to composite null hypotheses are provided in [28]. In order to introduce the notion of minimax optimality for a testing procedure, let us recall some definitions. We consider the uniform separation rate as defined in [9]. Let Δ_γ be a γ -level test with values in $\{0, 1\}$, where $\Delta_\gamma = 1$ corresponds to the decision of rejecting the null hypothesis $f = f_0$ and $\mathbb{P}_{f_0}(\Delta_\gamma = 1) \leq \gamma$. The uniform separation rate $\tilde{\rho}_n(\Delta_\gamma, \mathcal{C}, \beta, f_0)$ of the test Δ_γ with respect to the \mathbb{L}_2 -norm, over a class \mathcal{C} of alternatives f such that $f - f_0$ satisfies smoothness assumptions, is defined for all β in $(0, 1)$ as

$$\tilde{\rho}_n(\Delta_\gamma, \mathcal{C}, \beta, f_0) = \inf \left\{ \rho > 0; \sup_{f \in \mathcal{C}, \|f - f_0\|_2 > \rho} \mathbb{P}_f(\Delta_\gamma(X_1, \dots, X_n) = 0) \leq \beta \right\}, \quad (2)$$

where \mathbb{P}_f denotes the distribution of the i.i.d. samples (X_1, \dots, X_n) with common density f .

The uniform separation rate is then the smallest value in the sense of the \mathbb{L}_2 -norm of $(f - f_0)$ for which the second kind error of the test is uniformly controlled by β over \mathcal{C} . This definition extends the notion of critical radius introduced in [34] to the non-asymptotic framework. Note that minimax separation rates are at least as fast as minimax estimation rates and the interest lies in determining problems where testing can be done faster than estimating.

A test with level γ having optimal performances should then have the smallest possible uniform separation rate (up to a multiplicative constant) over \mathcal{C} . To quantify this, [9] introduces the non-asymptotic minimax rate of testing defined by

$$\tilde{\rho}_n^*(\mathcal{C}, \gamma, \beta, f_0) = \inf_{\Delta_\gamma} \tilde{\rho}_n(\Delta_\gamma, \mathcal{C}, \beta, f_0), \quad (3)$$

where the infimum is taken over all tests of level γ . A test is optimal in the minimax sense over the class \mathcal{C} if its uniform separation rate is upper-bounded, up to some constant, by the non-asymptotic minimax rate of testing. Taking \mathcal{C} too general leads to trivial rates. That is the reason why we restrict our study to two cases. On the one hand, we consider multinomial distributions which cover the discrete case. On the other hand, we work in the continuous case with Besov balls, which have been widely used in statistics since the seminal paper by [17]. Non-private results already exist for such sets, which make them meaningful for comparisons. Another motivation is that Besov sets are function classes parametrized by smoothness parameters and the minimax rates depend exclusively on those parameters in a lot of problems. Finally, thanks to their interesting properties from approximation theory, a large variety of signals can be dealt with, especially those built using wavelet bases.

We present a few non-private results from the literature. For Hölder classes with smoothness parameter $s > 0$, [34] establishes the asymptotic minimax rate of testing $n^{-2s/(4s+1)}$. The test proposed in their paper is not adaptive since it makes use of a known smoothness parameter s . Minimax optimal adaptive goodness-of-fit tests over Hölder or Besov classes of alternatives are provided in [35] and [28]. These tests achieve the separation rate $(n/\sqrt{\log \log(n)})^{-2s/(4s+1)}$ over a wide range of regularity classes (Hölder or Besov balls) with smoothness parameter $s > 0$. The $\log \log(n)$ term is the optimal price to pay for adaptation to the unknown parameter $s > 0$.

In the discrete case, the goal is to distinguish between d -dimensional probability vectors p and p^0 using samples from the multinomial distribution with parameters p and n . [46] obtain that the minimax optimal rate with respect to the l_1 -distance, $\sum_{i=1}^d |p_i - p_i^0|$, is $d^{1/4}/\sqrt{n}$. An extension is the study of local minimax rates as in [50], where the rate is made minimax optimal for any p^0 instead of just in the worst choice of p^0 . Finally, [8] presents local minimax rates of testing both in the discrete and continuous cases.

A few problems have already been tackled in order to obtain minimax rates under local privacy constraint. The main question is whether the minimax rates are affected by the local privacy constraint and to quantify the degradation of the rate in that case. We define a sample degradation of $C(\alpha)$ in the following way. If n is the necessary and sufficient sample size in order to solve the classical non-private version of a problem, the α -local differential private problem is solved with $nC(\alpha)$ samples. For a few problems, a degradation of the effective sample size by a multiplicative constant is found. In [21], they obtain minimax estimation rates for multinomial distributions in dimension d and find a sample degradation of α^2/d . In [20], they also find a multiplicative sample degradation of α^2/d for generalized linear models, and α^2 for median estimation. However, in other problems, a polynomial degradation is noted. For one-dimensional mean estimation, the usual minimax rate is $n^{-(1 \wedge (2-2/k))}$, whereas the private rate from [20] is $(n\alpha^2)^{-(0 \wedge (1-1/k))}$ for original observations X satisfying $\mathbb{E}(X) \in [-1, 1]$ and $\mathbb{E}(|X|^k) < \infty$. As for the problem of nonparametric density estimation presented in [20], the rate goes from $n^{-2s/(2s+1)}$ to $(n\alpha^2)^{-2s/(2s+2)}$ over an elliptical Sobolev space with smoothness s . This result

was extended in [12] over Besov ellipsoids. The classical minimax mean squared errors were presented in [54, 53, 49].

Goodness-of-fit testing has been studied extensively under a global differential privacy constraint in [29], [14], [7], [2] and [15]. Further steps into covering other testing problems under global differential privacy have been taken already with works like [6].

Our contributions can be summarized in the following way. Under non-interactive local differential privacy, we provide optimal separation rates for goodness-of-fit testing over Besov balls in the continuous case. To the best of our knowledge, we are the first to provide quantitative guarantees in such a continuous setting. We also provide minimax separation rates for multinomial distributions. In particular, we establish a lower bound when f_0 is uniform, that is completely novel in the definition of the prior distributions leading to optimal rates, and in the way we tackle non-interactive privacy. Indeed, naive applications of previous information processing inequalities under local privacy lead to suboptimal lower bounds. Finally, we provide an adaptive version of our test, which does not rely on the knowledge of the smoothness parameter s and is rate-optimal up to a logarithmic factor. So in shorter terms:

- We provide the first minimax lower bound for the problem of goodness-of-fit testing under local privacy constraint over Besov balls. We focus on non-interactive privacy, and f_0 being uniform, which directly translates to a lower bound when f_0 is nearly uniform.
- We present the first minimax optimal test with the associated non-interactive local differentially private channel in this continuous setting. The upper bound obtained for this test will hold for any density $f_0 \in \mathbb{L}_2([0, 1])$.
- The test is made adaptive to the smoothness parameter of the unknown density up to a logarithmic term.
- A minimax optimal test under non-interactive privacy can be derived for multinomial distributions as well.

We start with citing results pertaining to the study of goodness-of-fit testing in the discrete case under local differential privacy. [30] takes another point of view from ours and provide asymptotic distributions for a chi-squared statistic applied to noisy observations satisfying the local differential privacy condition. [48] takes a closer approach to ours and determines a sufficient number of samples for testing between $p = p^0$ and fixed $\sum |p_i - p_i^0|$, which has been improved upon by [1]. Finally, in parallel with the writing of the present paper, [10] have provided minimax optimal rates of testing for discrete distributions under local privacy, in both l_1 and l_2 norms. In particular, they tackle both interactive and non-interactive privacy channels and point out a discrepancy in the rates between both cases.

Now, the following papers tackle the continuous case. [12] provides minimax optimal rates for density estimation over Besov ellipsoids under local differential privacy. Following this paper, we apply Laplace noise to the projection of the observations onto a wavelet basis, although we tackle the different problem of density testing. The difference between density estimation and testing is fundamental and leads in our case to faster rates. A problem closer to density testing is the estimation of the quadratic functional presented in [13], where the authors find minimax rates over Besov ellipsoids under local differential privacy. They rely on the proof of the lower bound in the non-interactive case given in a preliminary version of our paper – see [40]. It was refined in order to improve on the rate in α , reaching an optimal rate for low values of α . So combining the lower bound we obtain under non-interactive privacy with their study of interactive local privacy, it is possible to note that there is an intrinsic gap in effectiveness that non-interactive privacy cannot hope to close for estimation of the squared functional.

Finally, the present paper is an iteration over [40], which only focused on the continuous case. We extend its scope and construct a unified setting to tackle both Besov classes and multinomial distributions, leading to minimax optimal results in both settings.

The rest of the paper is articulated as follows. In Section 2, we detail our setting and sum up our results. A lower bound on the minimax separation distance for goodness-of-fit testing is introduced in Section 3. Then we introduce a test and a privacy mechanism in Section 4. This leads to an upper bound which matches the lower bound. However, in the continuous case, the proposed test depends on a smoothness parameter which is unknown in general. That is the reason why we present a version of the test in Section 5 that is adaptive to s . Afterwards, we conclude the paper with a final discussion in Section 6. Finally, in the Appendix, the proofs of all the results presented in this paper are contained in Section A.

All along the paper, C will denote some absolute constant, $c(a, b, \dots), C(a, b, \dots)$ will be constants depending only on their arguments. The constants may vary from line to line.

2. Setting

2.1. Local differential privacy

Let n be some positive integer and $\alpha > 0$. Let f, f_0 be densities in $\mathbb{L}_2([0, 1])$ with respect to the Lebesgue measure. Let X_1, \dots, X_n be i.i.d. random variables with density f . Equation (1) defines local differential privacy. However, we define Z_1, \dots, Z_n satisfying a stronger assumption corresponding to the non-interactive case (see [52] and [26]). It is expressed for all $1 \leq i \leq n$ as

$$\sup_{S \in \mathcal{Z}_i, (x, x') \in [0, 1]^2} \frac{Q_i(Z_i \in S | X_i = x)}{Q_i(Z_i \in S | X_i = x')} \leq e^\alpha. \quad (4)$$

Let \mathcal{Q}_α be the set of joint distributions whose marginals satisfy the condition in Equation (4), that is, the set of α non-interactive privacy channels associated with X_1, \dots, X_n .

2.2. A unified setting for discrete and continuous distributions

We present a unified setting and end up dealing with densities in $\mathbb{L}_2([0, 1])$ in both the continuous and discrete cases. In the discrete case, $\widetilde{X}_1, \dots, \widetilde{X}_n$ are i.i.d. random variables taking their values in d classes denoted by $\{0, 1, \dots, d-1\}$ according to the probability vector $p = (p_0, p_1, \dots, p_{d-1})$. For a given probability vector $p^0 = (p_0^0, p_1^0, \dots, p_{d-1}^0)$, we want to test the null hypothesis $H_0 : p = p^0$ against the alternative $H_1 : p \neq p^0$. In order to have a unified setting, we transform these discrete observations into continuous observations X_1, \dots, X_n with values in $[0, 1]$ by the following process. For all $k \in \{0, \dots, d-1\}$, if we observe $\widetilde{X}_i = k$, we generate X_i by a uniform distribution on the interval $[k/d, (k+1)/d)$. Note that the variables X_1, \dots, X_n are i.i.d. with common density f defined for all $x \in [0, 1]$ by

$$f(x) = \sum_{k=0}^{d-1} dp_k \mathbf{1}_{[\frac{k}{d}, \frac{k+1}{d})}(x).$$

Similarly, for the probability vector p^0 , we define the corresponding density f_0 for $x \in [0, 1]$ by

$$f_0(x) = \sum_{k=0}^{d-1} dp_k^0 \mathbf{1}_{[\frac{k}{d}, \frac{k+1}{d})}(x).$$

So, given the definitions of f and f_0 , we have the equivalence $p = p^0 \iff f = f_0$. The following equation highlights the connection between the separation rates for densities and for probability vectors. We have

$$\|f - f_0\|_2^2 = d \sum_{k=0}^{d-1} (p_k - p_k^0)^2. \quad (5)$$

2.3. Separation rates

We now define a privacy mechanism and a testing procedure based on the private views Z_1, \dots, Z_n . We want to test

$$H_0 : f = f_0, \quad \text{versus} \quad H_1 : f \neq f_0, \quad (6)$$

from α -local differentially private views of X_1, \dots, X_n .

The twist on classical goodness-of-fit testing is in the fact that the samples (X_1, \dots, X_n) from f are unobserved, we only observe their private views. For $\alpha > 0$ and $\gamma \in (0, 1)$, we construct an α -local differentially private channel $Q \in \mathcal{Q}_\alpha$ and a γ -level test $\Delta_{\gamma, Q}$ such that

$$\mathbb{P}_{Q_{f_0}^n} (\Delta_{\gamma, Q}(Z_1, \dots, Z_n) = 1) \leq \gamma,$$

where

$$\mathbb{P}_{Q_{f_0}^n} ((Z_1, \dots, Z_n) \in \prod_{i=1}^n S_i) = \int \prod_i Q_i(Z_i \in S_i | X_i = x_i) f_0(x_i) dx_i,$$

and Q_i is the i -th marginal channel of Q .

We then define the uniform separation rate of the test $\Delta_{\gamma, Q}$ over the class \mathcal{C} as

$$\rho_n(\Delta_{\gamma, Q}, \mathcal{C}, \beta, f_0) = \inf \left\{ \rho > 0; \sup_{f \in \mathcal{C}, \|f - f_0\|_2 > \rho} \mathbb{P}_{Q_f^n} (\Delta_{\gamma, Q}(Z_1, \dots, Z_n) = 0) \leq \beta \right\}. \quad (7)$$

A good channel Q and a good test $\Delta_{\gamma, Q}$ are characterized by a small uniform separation rate. This leads us to the definition of the α -private minimax separation rate over the class \mathcal{C}

$$\rho_n^*(\mathcal{C}, \alpha, \gamma, \beta, f_0) = \inf_{Q \in \mathcal{Q}_\alpha} \inf_{\Delta_{\gamma, Q}} \rho_n(\Delta_{\gamma, Q}, \mathcal{C}, \beta, f_0), \quad (8)$$

where the infimum is taken over all possible α -private channels Q and all γ -level test $\Delta_{\gamma, Q}$ based on the private observations Z_1, \dots, Z_n .

Let us now introduce the classes of alternatives \mathcal{C} over which we will establish α -private minimax separation rates.

1. In the discrete case, we define

$$\mathcal{D} = \left\{ f \in \mathbb{L}_2([0, 1]); \exists p = (p_0, \dots, p_{d-1}) \in \mathbb{R}^d, \sum_{j=0}^{d-1} p_j = 1, f = \sum_{j=0}^{d-1} p_j \mathbf{1}_{[j/d, (j+1)/d)} \right\}, \quad (9)$$

which is associated with the class of densities for multinomial distributions over d classes. Then the minimax separation rate of interest will be denoted $\rho_n^*(\mathcal{D}, \alpha, \gamma, \beta, f_0)$.

2. In the continuous case, we consider Besov balls. To define these classes, we consider a pair of compactly supported and bounded wavelets (φ, ψ) such that for all J in \mathbb{N} ,

$$\left\{2^{J/2}\varphi(2^J(\cdot) - k), k \in \Lambda(J)\right\} \cup \left\{2^{j/2}\psi(2^j(\cdot) - k), j \geq J, k \in \Lambda(j)\right\}$$

is an orthonormal basis of $\mathbb{L}_2([0, 1])$. For the sake of simplicity, we consider the Haar basis where $\varphi = \mathbf{1}_{[0,1]}$ and $\psi = \mathbf{1}_{[0,1/2)} - \mathbf{1}_{[1/2,1]}$. In this case, for all $j \in \mathbb{N}$, $\Lambda(j) = \{0, 1, \dots, 2^j - 1\}$. We denote for all $j \geq 0, k \in \Lambda(j)$,

$$\alpha_{j,k}(f) = \int 2^{j/2} f \varphi(2^j(\cdot) - k), \text{ and } \beta_{j,k}(f) = \int 2^{j/2} f \psi(2^j(\cdot) - k).$$

For $R > 0$ and $s > 0$, the Besov ball $\tilde{B}_{s,2,\infty}(R)$ with radius R associated with the Haar basis is defined as

$$\tilde{B}_{s,2,\infty}(R) = \left\{ f \in \mathbb{L}_2([0, 1]), \forall j \geq 0, \sum_{k \in \Lambda(j)} \beta_{j,k}^2(f) \leq R^2 2^{-2js} \right\}.$$

Now note that, if $s < 1$, then there is an equivalence between the definition of $\tilde{B}_{s,2,\infty}(R)$ and the definition of the corresponding Besov space using moduli of smoothness – see e.g. Theorem 4.3.2 in [31]. And for larger s , Besov spaces defined with Daubechies wavelets satisfy this equivalence property, as explained in Section 4.3.5 in [31].

We introduce the following class of alternatives: for any $s > 0$ and $R > 0$, we define the set $\mathcal{B}_{s,2,\infty}(R)$ as follows

$$\mathcal{B}_{s,2,\infty}(R) = \left\{ f \in \mathbb{L}_2([0, 1]), f - f_0 \in \tilde{B}_{s,2,\infty}(R) \right\}. \quad (10)$$

Note that the class $\mathcal{B}_{s,2,\infty}(R)$ depends on f_0 since only the regularity for the difference $f - f_0$ is required to establish the separation rates. Nevertheless, for the sake of simplicity we omit f_0 in the notation of this set. The minimax separation rate of interest will be denoted $\rho_n^*(\mathcal{B}_{s,2,\infty}(R), \alpha, \gamma, \beta, f_0)$.

2.4. Overview of the results

For any $\alpha > 0$, we define $z_\alpha = e^{2\alpha} - e^{-2\alpha} = 2 \sinh(2\alpha)$.

Continuous case. The results presented in Theorems 3.4 and 4.9 can be condensed into the following conclusion that holds if $n z_\alpha^2 \geq (\log n)^{1+3/(4s)}$, $s > 0$, $R > 0$, $\alpha \geq 1/\sqrt{n}$, $(\gamma, \beta) \in (0, 1)^2$ such that $2\gamma + \beta < 1$,

$$\begin{aligned} c(s, R, \gamma, \beta) [(n z_\alpha^2)^{-2s/(4s+3)} \vee n^{-2s/(4s+1)}] \\ \leq \rho_n^* \left(\mathcal{B}_{s,2,\infty}(R), \alpha, \gamma, \beta, \mathbf{1}_{[0,1]} \right) \\ \leq C(s, R, \gamma, \beta) \left[(n \alpha^2)^{-2s/(4s+3)} \vee n^{-2s/(4s+1)} \right]. \end{aligned} \quad (11)$$

Comments.

1. Having $nz_\alpha^2 \geq (\log n)^{1+3/(4s)}$ and $\alpha \geq 1/\sqrt{n}$ reduces to wanting a sample set large enough, which is a classical non-restrictive assumption.
2. The upper bound holds for any density $f_0 \in \mathbb{L}_2([0, 1])$ up to $\|f_0\|_2$ and matches the lower bound when $f_0 = \mathbf{1}_{[0,1]}$, as shown in Equation (11). So we can deduce the minimax separation rate for goodness-of-fit testing of Besov densities under a non-interactive privacy constraint, when f_0 is nearly uniform. It can be decomposed into two different regimes, where the rates of our upper and lower bounds match in n as well as in α , when α tends to 0. When α is larger than $n^{1/(4s+1)}$, then the minimax rate is of order $n^{-2s/(4s+1)}$, which coincides with the rate obtained in the non-private case in [33]. The other regime corresponds to α being smaller than $n^{1/(4s+1)}$. The minimax rate is then of order $(n\alpha^2)^{-2s/(4s+3)}$ and so we show a polynomial degradation in the rate due to the privacy constraints. Very similar results have been found for the estimation of the quadratic functional in [13]. Such a degradation has also been discovered in the problem of second moment estimation and mean estimation, as well as for the density estimation in [12].
3. Due to having z_α instead of α , our upper and lower bounds do not match in α when α is larger than a constant but smaller than $n^{1/(4s+1)}$. This is not an issue in practice, since α will be taken small in order to guarantee privacy.

Discrete case. The results presented in Theorems 3.2 and 4.5 can be condensed into the following conclusion that holds if $n \geq (z_\alpha^{-2} d^{3/2} \log d) \vee ((\alpha^2 d^{-1/2}) \wedge d^{1/2})$, $\alpha > 0$, $(\gamma, \beta) \in (0, 1)^2$ such that $2\gamma + \beta < 1$,

$$\begin{aligned}
c(\gamma, \beta) & \left[(nz_\alpha^2)^{-1/2} d^{1/4} \vee (n^{-1/2} d^{-1/4}) \right] \\
& \leq \rho_n^* \left(\mathcal{D}, \alpha, \gamma, \beta, \mathbf{1}_{[0,1]} \right) / d^{1/2} \\
& \leq C(\gamma, \beta) \left[((n\alpha^2)^{-1/2} d^{1/4}) \vee (n^{-1/2} d^{-1/4}) \right].
\end{aligned} \tag{12}$$

Comments.

1. Assuming that $nz_\alpha^2 \geq d^{3/2} \log d$ means that the problem gets harder with the dimension, which aligns with the interpretation of the private rate.
2. We present matching bounds on $\rho_n^* \left(\mathcal{D}, \alpha, \gamma, \beta, \mathbf{1}_{[0,1]} \right) / d^{1/2}$ since it is the usual rate of interest as justified by the combination of Definition 2 and Equation (5). This helps us conclude on the minimax separation rate for goodness-of-fit testing of discrete distributions under a non-interactive privacy constraint, when p_0 is nearly uniform. We present in Theorem 4.5 an upper bound for other definitions of p_0 as well. Here again, we find two regimes corresponding to the classical rate taking over if α is larger than \sqrt{d} . So we can see that the local privacy condition leaves the rate in n unchanged, but the rate in d changes drastically for the testing problem with respect to the \mathbb{L}_2 -norm. Indeed, the classical testing problem with \mathbb{L}_2 -separation becomes easier as the number of dimension grows, whereas the private rate exhibits the opposite behaviour.
3. Simultaneously and independently of our work, [10] find similar results in the non-interactive case.

3. Lower bound

This section will focus on the presentation of a lower bound on the minimax separation rate defined in Equation (8) for the problem of goodness-of-fit testing under a non-interactive differential privacy

constraint. The result is presented both in the discrete and the continuous cases, when f_0 is the uniform density over $[0, 1]$. As seen in [12], an important contribution of our lower bound focusing on the non-interactive case is that, when combined with an interactive privacy channel and a test reaching a smaller separation distance, one can conclude that interactive privacy can lead to better results than what can be achieved under non-interactive privacy in the problem under scrutiny.

The outline of the lower bound proof relies on a classical scheme, which is recalled below. Nevertheless, the implementation of this scheme in the context of local differential privacy is far from being classical, and we do it in a novel way which leads to a tight lower bound. At the end of the section, a more naive approach will be presented and shown to lead to suboptimal results.

We apply a Bayesian approach, where we will define a prior distribution which corresponds to a mixture of densities such that $\|f - f_0\|_2$ is large enough. Such a starting point has been largely employed for lower bounds in minimax testing, as described in [9]. Its application is mainly due to [34] and inequalities on the total variation distance from [41]. The result of this approach is summarized in the following lemma.

Lemma 3.1. *Let $\mathcal{C} \subset \mathbb{L}_2([0, 1])$. Let $(\gamma, \beta) \in (0, 1)^2$ and $\delta \in [0, 1)$ such that $\gamma + \beta + \delta < 1$. Let $\rho > 0$. We define*

$$\mathcal{F}_\rho(\mathcal{C}) = \{f \in \mathcal{C}, \|f - f_0\|_2 \geq \rho\}.$$

Let $\alpha > 0$ and let $Q \in \mathcal{Q}_\alpha$ be some non-interactive α -private channel. Let ν_ρ be some probability measure such that $\nu_\rho(\mathcal{F}_\rho(\mathcal{C})) \geq 1 - \delta$ and let $Q_{\nu_\rho}^n$ be defined, for all measurable set A by

$$\mathbb{P}_{Q_{\nu_\rho}^n}((Z_1, \dots, Z_n) \in A) = \int \mathbb{P}_{Q_g^n}((Z_1, \dots, Z_n) \in A) d\nu_\rho(g).$$

We note the total variation distance between two probability measures \mathbb{P}_1 and \mathbb{P}_2 as $\|\mathbb{P}_1 - \mathbb{P}_2\|_{TV} = \sup_A |\mathbb{P}_1(A) - \mathbb{P}_2(A)|$.

Then if

$$\|\mathbb{P}_{Q_{\nu_\rho}^n} - \mathbb{P}_{Q_{f_0}^n}\|_{TV} < 1 - \gamma - \beta - \delta,$$

we have

$$\inf_{\Delta_{\gamma, Q}} \rho_n(\Delta_{\gamma, Q}, \mathcal{C}, \beta, f_0) \geq \rho,$$

where the infimum is taken over all possible γ -level test, hence satisfying

$$\mathbb{P}_{Q_{f_0}^n}(\Delta_{\gamma, Q}(Z_1, \dots, Z_n) = 1) \leq \gamma.$$

The idea is to establish the connection between the second kind error and the total variation distance between arbitrary distributions with respective supports in H_0 and $\mathcal{F}_\rho(\mathcal{C})$. It turns out that the closer the distributions from H_0 and $\mathcal{F}_\rho(\mathcal{C})$ are allowed to be, the higher the potential second kind error. So if we are able to provide distributions from H_0 and $\mathcal{F}_\rho(\mathcal{C})$ which are close from one another, we can guarantee that the second kind error of any test will be high. The main difficulty lies in finding the right prior distribution ν_ρ appearing in Lemma 3.1.

In the discrete case, we obtain the following lower bound.

Theorem 3.2. *Let $(\gamma, \beta) \in (0, 1)^2$ such that $2\gamma + \beta < 1$. Let $\alpha > 0$.*

We obtain the following lower bound for the α -private minimax separation rate defined by Equation (8) for non-interactive channels in \mathcal{Q}_α over the class of alternatives \mathcal{D} in Equation (9)

$$\rho_n^* \left(\mathcal{D}, \alpha, \gamma, \beta, \mathbf{I}_{[0,1]} \right) / d^{1/2} \geq c(\gamma, \beta) \left([(nz_\alpha^2)^{-1/2} d^{1/4} \wedge d^{-1/2} (\log d)^{-1/2}] \vee (n^{-1/2} d^{-1/4}) \right).$$

Remark 3.3. In parallel to our work, [10] focus on the case when $\alpha \leq 1$ and find similar results displayed in their Theorem 6.

In the continuous case, we obtain the following theorem for Besov balls.

Theorem 3.4. Let $(\gamma, \beta) \in (0, 1)^2$ such that $2\gamma + \beta < 1$. Let $\alpha > 0, R > 0, s > 0$.

We obtain the following lower bound for the α -private minimax separation rate defined by Equation (8) for non-interactive channels in \mathcal{Q}_α over the class of alternatives $\mathcal{B}_{s,2,\infty}(R)$ defined in Equation (10)

$$\rho_n^* \left(\mathcal{B}_{s,2,\infty}(R), \alpha, \gamma, \beta, \mathbf{I}_{[0,1]} \right) \geq c(\gamma, \beta, R) \left[[(nz_\alpha^2)^{-2s/(4s+3)} \wedge (\log n)^{-1/2}] \vee n^{-2s/(4s+1)} \right].$$

Remark 3.5. These theorems represent a major part of our contributions and lead to the construction of the inequalities presented in Section 2.4. Note that $(nz_\alpha^2)^{-2s/(4s+3)} \wedge (\log n)^{-1/2}$ reduces to $(nz_\alpha^2)^{-2s/(4s+3)}$ for n large enough and we can reduce the formulation of Theorem 3.2 in the same way with a condition on n being large enough.

Sketch of proof. We want to find the largest \mathbb{L}_2 -distance between the initial density f_0 under the null hypothesis and the density in the alternative hypothesis such that their transformed counterparts by an α -private channel Q cannot be discriminated by a test. We will rely on the singular vectors of Q in order to define densities and their private counterparts with ease. Employing bounds on the singular values of Q , we define a mixture of densities such that they have a bounded \mathbb{L}_2 -distance to $f_0 = \mathbf{I}_{[0,1]}$. We obtain a sufficient condition for the total variation distance between the densities in the private space to be small enough for both hypotheses to be indistinguishable. Then we ensure that the functions that we have defined are indeed densities, and in the continuous case belong to the regularity class $\mathcal{B}_{s,2,\infty}(R)$. Collecting all these elements, the conclusion relies on Lemma 3.1.

Remark 3.6. The total variation distance is a good criterion in order to determine whether two distributions are distinguishable. Another natural idea to prove Theorem 3.4 is to bound the total variation distance between two private densities by the total variation distance between the densities of the original samples, up to some constants depending on the privacy constraints. Following this intuitive approach, we can provide a lower bound using Theorem 1 in [19] combined with Pinsker's inequality. This approach has been used with success in density estimation in [12]. However, the resulting lower bound does not match the upper bound for the separation rates of goodness-of-fit testing presented in our Section 4.

4. Definition of a test and privacy mechanism

We will firstly define a testing procedure coupled with a privacy mechanism. Their application provides an upper bound on the minimax separation rate for any density f_0 . The bounds obtained are presented in the right-hand side of Equations (11) and (12) for the continuous and the discrete cases respectively. The test and privacy mechanism will turn out to be minimax optimal since the upper bounds will match the lower bounds obtained in Section 3.

Let us first propose a transformation of the data, satisfying the differential privacy constraints.

4.1. Privacy mechanism

We consider the privacy mechanism introduced in [12]. It relies on Laplace noise, which is classical as a privacy mechanism. However, applying it to the correct basis with the corresponding scaling is critical in finding optimal results. We denote by φ the indicator function on $[0, 1)$

$$\forall x, \varphi(x) = \mathbf{1}_{[0,1)}(x),$$

and for all integer $L \geq 1$, we set, for all $k \in \{0, \dots, L-1\}$, for all $x \in [0, 1)$,

$$\varphi_{L,k}(x) = \sqrt{L}\varphi(Lx - k).$$

The integer L will be taken as $L = 2^J$ for some $J \geq 0$ in the continuous case, and we choose $L = d$ in the discrete case. We define, for all $i \in \{1, \dots, n\}$, the vector $Z_{i,L} = (Z_{i,L,k})_{k \in \{0, \dots, L-1\}}$, by

$$\forall k \in \{0, \dots, L-1\}, Z_{i,L,k} = \varphi_{L,k}(X_i) + \sigma_L W_{i,L,k}, \quad (13)$$

where $(W_{i,L,k})_{1 \leq i \leq n, k \in \{0, \dots, L-1\}}$ are i.i.d. Laplace distributed random variables with variance 1 and

$$\sigma_L = 2\sqrt{2} \frac{\sqrt{L}}{\alpha}.$$

Lemma 4.1. *For any i , denote $q_{i,L}(\cdot|x)$ the density of the random vector $Z_{i,L}$ with respect to the probability measure μ_i conditionally to $X_i = x$. Then*

$$\sup_{S \in \mathcal{Z}_{i,L}, (x, x') \in [0,1]^2} \frac{Q_i(Z_{i,L} \in S | X_i = x)}{Q_i(Z_{i,L} \in S | X_i = x')} \leq e^\alpha$$

if and only if there exists $\Omega \in \mathcal{Z}_{i,L}$ with $\mu_i(Z_{i,L} \in \Omega) = 1$ such that

$$\frac{q_{i,L}(z|x)}{q_{i,L}(z|x')} \leq e^\alpha$$

for any $z \in \Omega$ and any $(x, x') \in [0, 1]^2$.

Proof. Assume there exists Ω with $\mu_i(Z_{i,L} \in \Omega) = 1$ such that $\frac{q_{i,L}(z|x)}{q_{i,L}(z|x')} \leq e^\alpha$ for any $z \in \Omega$. Let $\tilde{S} \in \mathcal{Z}_{i,L}$ and $S = \tilde{S} \cap \Omega$.

$$\frac{Q_i(Z_{i,L} \in \tilde{S} | X_i = x)}{Q_i(Z_{i,L} \in \tilde{S} | X_i = x')} = \frac{Q_i(Z_{i,L} \in S | X_i = x)}{Q_i(Z_{i,L} \in S | X_i = x')}.$$

Then

$$\begin{aligned} \frac{Q_i(Z_{i,L} \in S | X_i = x)}{Q_i(Z_{i,L} \in S | X_i = x')} &= \frac{\int_S q_{i,L}(z|x) d\mu_i(z)}{\int_S q_{i,L}(z|x') d\mu_i(z)} \\ &\leq \frac{\int_S q_{i,L}(z|x') e^\alpha d\mu_i(z)}{\int_S q_{i,L}(z|x) e^{-\alpha} d\mu_i(z)} = \frac{Q_i(Z_{i,L} \in S | X_i = x')}{Q_i(Z_{i,L} \in S | X_i = x)} e^{2\alpha}. \end{aligned}$$

So

$$\frac{Q_i(Z_{i,L} \in \tilde{S} | X_i = x)}{Q_i(Z_{i,L} \in \tilde{S} | X_i = x')} \leq e^\alpha.$$

Assume that $Q \in \mathcal{Q}_\alpha$. Then for any $S \in \mathcal{Z}_{i,L}$, we have $Q_i(Z_{i,L} \in S | X_i = x) \leq e^\alpha Q_i(Z_{i,L} \in S | X_i = x')$. That is, for any $S \in \mathcal{Z}_{i,L}$,

$$\int_S (e^\alpha q_{i,L}(z|x') - q_{i,L}(z|x)) d\mu_i(z) \geq 0.$$

So there exists Ω with $\mu_i(Z_{i,L} \in \Omega) = 1$ such that $\frac{q_{i,L}(z|x)}{q_{i,L}(z|x')} \leq e^\alpha$ for any $z \in \Omega$. \square

Lemma 4.2. *To each random variable X_i of the sample set (X_1, \dots, X_n) , we associate the vector $Z_{i,L} = (Z_{i,L,k})_{k \in \{0, \dots, L-1\}}$. The random vectors $(Z_{1,L}, \dots, Z_{n,L})$ are non-interactive α -local differentially private views of the samples (X_1, \dots, X_n) . Namely, they satisfy the condition in Equation (4).*

The proof in the continuous case can also be found in [12] (see Proposition 3.1). We recall here the main arguments for the sake of completeness.

Proof. The random vectors $(Z_{i,L})_{1 \leq i \leq n}$ are i.i.d. by definition. For any x_i, x'_i in $[0, 1]$, for any $z_i \in \mathbb{R}^L$,

$$\begin{aligned} \frac{q_{i,L}(z_i|x_i)}{q_{i,L}(z_i|x'_i)} &= \prod_{k=0}^{L-1} \exp \left[\sqrt{2} \frac{|z_{i,k} - \varphi_{L,k}(x'_i)| - |z_{i,k} - \varphi_{L,k}(x_i)|}{\sigma_L} \right] \\ &\leq \exp \left[\sum_{k=0}^{L-1} \frac{\sqrt{2}}{\sigma_L} (|\varphi_{L,k}(x'_i)| + |\varphi_{L,k}(x_i)|) \right]. \end{aligned}$$

Since $\varphi_{L,k}(x_i) \neq 0$ for a single value of $k \in \{0, \dots, L-1\}$, we get

$$\frac{q_{i,L}(z_i|x_i)}{q_{i,L}(z_i|x'_i)} \leq \exp \left[\frac{2\sqrt{2} \|\varphi_{L,k}\|_\infty}{\sigma_L} \right] \leq e^\alpha,$$

by definition of σ_L , which concludes the proof by application of Lemma 4.1. \square

4.2. Definition of the test

Let f_0 be some fixed density in $\mathbb{L}_2([0, 1])$. Our aim is now to define a testing procedure for the testing problem defined in Equation (6) from the observation of the vectors (Z_1, \dots, Z_n) . Our test statistic \hat{T}_L is defined as

$$\hat{T}_L = \frac{1}{n(n-1)} \sum_{i \neq l=1}^n \sum_{k=0}^{L-1} \left(Z_{i,L,k} - \alpha_{L,k}^0 \right) \left(Z_{l,L,k} - \alpha_{L,k}^0 \right), \quad (14)$$

where $\alpha_{L,k}^0 = \int_0^1 \varphi_{L,k}(x) f_0(x) dx$.

We consider the test function

$$\Delta_{L,\gamma,Q}(Z_1, \dots, Z_n) = \mathbf{1}_{\hat{T}_L > t_L^0(1-\gamma)}, \quad (15)$$

where $t_L^0(1-\gamma)$ denotes the $(1-\gamma)$ -quantile of \hat{T}_L under H_0 . Note that this quantile can be estimated by simulations, under the hypothesis $f = f_0$. We can indeed simulate the vector (Z_1, \dots, Z_n) if the density of (X_1, \dots, X_n) is assumed to be f_0 . Hence the test rejects the null hypothesis H_0 if

$$\hat{T}_L > t_L^0(1-\gamma).$$

The test is of level γ by definition of the threshold.

Comments.

1. In a similar way as in [28], the test is based on an estimation of the quantity $\|f - f_0\|_2^2$. Note indeed that \hat{T}_L is an unbiased estimator of $\|\Pi_{S_L}(f - f_0)\|_2^2$, where Π_{S_L} denotes the orthogonal projection in $\mathbb{L}_2([0, 1])$ onto the space generated by the functions $(\varphi_{L,k}, k \in \{0, \dots, L-1\})$. In the discrete case, f and f_0 belong to S_L and $\Pi_{S_L}(f - f_0) = f - f_0$. In this case,

$$\|\Pi_{S_L}(f - f_0)\|_2^2 = \|f - f_0\|_2^2 = d \sum_{k=0}^{d-1} (p_k - p_k^0)^2.$$

2. Note that, in the discrete case, we obtain the following expression for the test statistic

$$\hat{T}_d = \frac{d}{n(n-1)} \sum_{k=0}^{d-1} \sum_{i \neq l=1}^n \left(\mathbf{1}_{\tilde{X}_i=k} - p_k^0 \right) \left(\mathbf{1}_{\tilde{X}_l=k} - p_k^0 \right). \quad (16)$$

It is interesting to compare this expression with the χ^2 statistics, which can be written as

$$\sum_{k=0}^{d-1} \sum_{i,l=1}^n \frac{\left(\mathbf{1}_{\tilde{X}_i=k} - p_k^0 \right) \left(\mathbf{1}_{\tilde{X}_l=k} - p_k^0 \right)}{np_k^0}.$$

Hence, besides the normalization of each term in the sum by p_k^0 in the χ^2 test, the main difference lies in the fact that we remove the diagonal terms (corresponding to $i = l$) in our test statistics.

In the next section, we provide non-asymptotic theoretical results for the power of this test.

4.3. Upper bound for the second kind error of the test

We first provide an upper bound for the second kind error of our test and privacy channel in a general setting.

Theorem 4.3. *Let (X_1, \dots, X_n) be i.i.d. with common density f on $[0, 1]$. Let f_0 be some given density on $[0, 1]$. We assume that f and f_0 belong to $\mathbb{L}_2([0, 1])$. From the observation of the random vectors (Z_1, \dots, Z_n) defined by Equation (13), for a given $\alpha > 0$, we test the hypotheses*

$$H_0 : f = f_0, \quad \text{versus} \quad H_1 : f \neq f_0.$$

We consider the test $\Delta_{L,\gamma,Q}$ defined by Equation (15) with \hat{T}_L defined in Equation (14). The test is obviously of level γ by definition of the threshold $t_L^0(1-\gamma)$, namely we have

$$\mathbb{P}_{Q_{f_0}^n} \left(\hat{T}_L \geq t_L^0(1-\gamma) \right) \leq \gamma.$$

Under the assumption that

$$\|\Pi_{S_L}(f - f_0)\|_2^2 \geq \sqrt{\text{Var}_{Q_{f_0}^n}(\hat{T}_L)/\gamma} + \sqrt{\text{Var}_{Q_f^n}(\hat{T}_L)}/\beta, \quad (17)$$

the second kind error of the test is controlled by β , namely we have

$$\mathbb{P}_{Q_f^n} \left(\hat{T}_L \leq t_L^0(1-\gamma) \right) \leq \beta. \quad (18)$$

Moreover, we have

$$\text{Var}_{Q_f^n}(\hat{T}_L) \leq C \left[\frac{(\sqrt{L}\|f\|_2 + \sigma_L^2)}{n} \|\Pi_{S_L}(f - f_0)\|_2^2 + \frac{(\|f\|_2^2 + \sigma_L^4)L}{n^2} \right]. \quad (19)$$

We give here a sketch of proof of Theorem 4.3. The complete proof of this result is given in Section A.3 of the appendix. Note that it is not fundamentally different from non-private proofs given in [28].

Sketch of proof. We want to establish a condition on $f - f_0$, under which the second kind error of the test is controlled by β . Denoting by $t_L(\beta)$ the β -quantile of \hat{T}_L under $\mathbb{P}_{Q_f^n}$, the condition in Equation (18) holds as soon as $t_L^0(1-\gamma) \leq t_L(\beta)$. Hence, we provide an upper bound for $t_L^0(1-\gamma)$ and a lower bound for $t_L(\beta)$. By Chebyshev's inequality, we obtain that on the one hand,

$$t_L^0(1-\gamma) \leq \sqrt{\text{Var}_{Q_{f_0}^n}(\hat{T}_L)/\gamma}, \quad (20)$$

and on the other hand,

$$\|\Pi_{S_L}(f - f_0)\|_2^2 - \sqrt{\text{Var}_{Q_f^n}(\hat{T}_L)}/\beta \leq t_L(\beta). \quad (21)$$

We deduce from the inequalities in Equations (20) and (21) that Equation (18) holds as soon as

$$\|\Pi_{S_L}(f - f_0)\|_2^2 \geq \sqrt{\text{Var}_{Q_{f_0}^n}(\hat{T}_L)/\gamma} + \sqrt{\text{Var}_{Q_f^n}(\hat{T}_L)}/\beta.$$

The main ingredient to control the variance terms is a control of the variance for U-statistics of order two which relies on Hoeffding's decomposition – see e.g. [47] Lemma A p. 183. The proof is given in Section A.3.

We obtain the following corollary of Theorem 4.3. It states a result that will be used in order to obtain an upper bound on the minimax rate both in the discrete and the continuous cases.

Corollary 4.4. *Under the same assumptions as in Theorem 4.3, we obtain that Equation (18) holds, that is, the second kind error of the test is controlled by β provided that*

$$\|\Pi_{S_L}(f - f_0)\|_2^2 \geq C(\gamma, \beta) \frac{(\|f\|_2 + \|f_0\|_2 + \sigma_L^2)\sqrt{L}}{n}. \quad (22)$$

In the next sections, we derive from this result upper bounds for the minimax separation rate over Besov balls in the continuous case, and conditions on the l_2 -distance between p and p^0 to obtain a prescribed power for the test in the discrete case.

4.4. Upper bound for the separation distance in the discrete case

The following theorem provides a sufficient condition on the separation distance between the probability vectors p and p^0 for both error kinds of the test to be controlled by γ and β , respectively. This sufficient condition corresponds to an upper bound on the minimax rate $\rho_n^*(\mathcal{D}, \alpha, \gamma, \beta, f_0) / d^{1/2}$ in the discrete case.

Theorem 4.5. *Let $p^0 = (p_0^0, p_1^0, \dots, p_{d-1}^0)$ be some given probability vector. Let (X_1, \dots, X_n) be i.i.d. with values in the finite set $\{0, 1, \dots, d-1\}$ and with common distribution defined by the probability vector $p = (p_0, p_1, \dots, p_{d-1})$.*

From the observation of the random vectors (Z_1, \dots, Z_n) defined by Equation (13) for a given $\alpha > 0$ with $L = d$, we want to test the hypotheses

$$H_0 : p = p^0, \quad \text{versus} \quad H_1 : p \neq p^0.$$

We consider the test $\Delta_{d, \gamma, Q}$ defined by Equation (15), which has a first kind error of γ . The second kind error of the test is controlled by β , provided that

$$\sqrt{\sum_{i=0}^{d-1} (p_i - p_i^0)^2} \geq C(\gamma, \beta) \frac{d^{-1/4}}{n^{1/2}} \left(d^{1/4} \left[\left(\sum_{k=0}^{d-1} (p_k^0)^2 \right)^{1/4} + n^{-1/2} \right] + d^{1/2} \alpha^{-1} \right).$$

Since $\sum_{k=0}^{d-1} (p_k^0)^2 \leq 1$, the second kind error of the test is controlled by β , provided that

$$\sqrt{\sum_{i=0}^{d-1} (p_i - p_i^0)^2} \geq C(\gamma, \beta) n^{-1/2} (1 \vee [d^{1/4} \alpha^{-1}]). \quad (23)$$

Remark 4.6. *Equation (23) displays a rate that is optimal in d, n , when α is smaller than $d^{1/4}$. Besides, the rate in α matches the lower bound asymptotically when α converges to 0. The upper bound presented in Theorem 1 from [10] tackles the case when α is smaller than 1 and they find the same rate as ours in their Corollary 2. They present an additional test statistic in order to refine their rates when p^0 is not a uniform vector.*

Corollary 4.7. *We assume that there exists an absolute constant κ such that*

$$d \sum_{k=0}^{d-1} (p_k^0)^2 \leq \kappa. \quad (24)$$

Then the second kind error of the test is controlled by β , provided that

$$\sqrt{\sum_i (p_i - p_i^0)^2} \geq C(\gamma, \beta, \kappa) \left[\left(d^{-1/4} n^{-1/2} \right) \vee \left(d^{1/4} n^{-1/2} \alpha^{-1} \right) \vee n^{-1} \right]. \quad (25)$$

Remark 4.8. *If we also assume the bound on $d \sum_{k=0}^{d-1} (p_k^0)^2$ as expressed in Equation (24), we find optimal rates in d, n if $n \geq (\alpha^2 d^{-1/2}) \wedge d^{1/2}$. The assumption in Equation (24) in Lemma 4.7 is equivalent to assuming that the function f_0 defined in Section 2.2 belongs to $\mathbb{L}^2([0, 1])$. It restricts p^0 to vectors that are close to being uniform. This coincides with the lower bound on the rate found when f_0 is a uniform density.*

4.5. Upper bound for the minimax separation rate over Besov balls

We provide an upper bound on the uniform separation rate for our test and privacy channel over Besov balls in Theorem 4.9.

Theorem 4.9. *Let (X_1, \dots, X_n) be i.i.d. with common density f on $[0, 1]$. Let f_0 be some given density on $[0, 1]$. We assume that f and f_0 belong to $\mathbb{L}_2([0, 1])$.*

We observe the random vectors (Z_1, \dots, Z_n) defined by Equation (13) for a given $\alpha > 0$ with the following value for L : we assume that $L = L^$, where $L^* = 2^{J^*}$, and J^* is the smallest integer J such that $2^J \geq (n\alpha^2)^{2/(4s+3)} \wedge n^{2/(4s+1)}$.*

We want to test the hypotheses

$$H_0 : f = f_0, \quad \text{versus} \quad H_1 : f \neq f_0.$$

We consider the test $\Delta_{L^, \gamma, Q}$ defined by Equation (15). The uniform separation rate, defined by Equation (7), of the test $\Delta_{L^*, \gamma, Q}$ over $\mathcal{B}_{s, 2, \infty}(R)$ defined by Equation (10) satisfies for all $n \in \mathbb{N}^*$, $R > 0$, $\alpha \geq 1/\sqrt{n}$, $(\gamma, \beta) \in (0, 1)^2$ such that $\gamma + \beta < 1$*

$$\rho_n(\Delta_{L^*, \gamma, Q}, \mathcal{B}_{s, 2, \infty}(R), \beta, f_0) \leq C(s, R, \|f_0\|_2, \gamma, \beta) \left[(n\alpha^2)^{-2s/(4s+3)} \vee n^{-2s/(4s+1)} \right].$$

The proof of this result is in Section A.3 of the appendix.

Comments.

1. When the sample set (X_1, \dots, X_n) is directly observed, [28] propose a testing procedure with uniform separation rate over the set $\mathcal{B}_{s, 2, \infty}(R)$ controlled by

$$C(s, R, \gamma, \beta) n^{-2s/(4s+1)},$$

which is an optimal result, as proved in [34]. Hence we obtain here a loss in the uniform separation rate, due to the fact that we only observe α -differentially private views of the original sample. This loss occurs when $\alpha \leq n^{1/(4s+1)}$. Otherwise, we get the same rate as when the original sample is observed. Comparing this result with the lower bound from Section 3, we conclude that the rate is optimal.

2. Finally, having $\alpha < 1/\sqrt{n}$ represents an extreme case, where the sample size is really low in conjunction with a very strict privacy condition. In such a range of α , J^* is taken equal to 0, but this does not lead to optimal rates.

The test proposed in Theorem 4.9 depends on the smoothness parameter s of the Besov ball $\mathcal{B}_{s, 2, \infty}(R)$ via the parameter J^* . In a second step, we will propose a test which is adaptive to the smoothness parameter s . Namely, in Section 5, we construct an aggregated testing procedure, which is independent of the smoothness parameter and achieves the minimax separation rates established in Equation (11) over a wide range of Besov balls simultaneously, up to a logarithmic term.

5. Adaptive tests

In Section 4, we have defined in the continuous case a testing procedure which depends on the parameter J . The performances of the test depend on this parameter. We have optimized the choice of J to obtain the smallest possible upper bound for the separation rate over the set $\mathcal{B}_{s,2,\infty}(R)$. Nevertheless, the test is not adaptive since this optimal choice of J depends on the smoothness parameter s .

In order to obtain adaptive procedure, we propose, as in [28] to aggregate a collection of tests. For this, we introduce the set

$$\mathcal{J} = \left\{ J \in \mathbb{N}, 2^J \leq n^2 \right\}$$

and the aggregated procedure will be based on the collection of test statistics $(\hat{T}_{2^J}, J \in \mathcal{J})$ defined by (14).

In Theorem 4.9, the testing procedure is based on the observation of the random vectors (Z_1, \dots, Z_n) defined by Equation (13) with $L = 2^{J^*}$ for the optimized value of J^* . Hence, the private views of the original sample depend on the unknown parameter s . In order to build the aggregated procedure, we can no more use the optimized value J^* of J and we need to observe the random vectors (Z_1, \dots, Z_n) for all $J \in \mathcal{J}$. In order to guaranty the α -local differential privacy, we have to increase slightly the variance of the Laplace perturbation. The privacy mechanism is specified in the following lemma.

Lemma 5.1. *We consider the set $\mathcal{J} = \{J \in \mathbb{N}, 2^J \leq n^2\}$. We define, for all $i \in \{1, \dots, n\}$, for all $J \in \mathcal{J}$, the vector $\tilde{Z}_{i,2^J} = (\tilde{Z}_{i,2^J,k})_{k \in \{0, \dots, 2^J-1\}}$, by*

$$\forall k \in \{0, \dots, 2^J - 1\}, \tilde{Z}_{i,2^J,k} = \varphi_{2^J,k}(X_i) + \tilde{\sigma}_{2^J} W_{i,2^J,k}, \quad (26)$$

where $(W_{i,2^J,k})_{1 \leq i \leq n, k \in \{0, \dots, 2^J-1\}}$ are i.i.d. Laplace distributed random variables with variance 1 and

$$\tilde{\sigma}_{2^J} = 2\sqrt{2}|\mathcal{J}| \frac{2^{J/2}}{\alpha}.$$

For all $1 \leq i \leq n$, we define the random vector $\tilde{Z}_i = (\tilde{Z}_{i,2^J}, J \in \mathcal{J})$. The random vectors $(\tilde{Z}_i, 1 \leq i \leq n)$ are non-interactive α -local differentially private views of the samples (X_1, \dots, X_n) . Namely, they satisfy the condition in Equation (4).

Proof. The random vectors $(\tilde{Z}_i)_{1 \leq i \leq n}$ are i.i.d. by definition. Let us denote by $\tilde{q}_i(\cdot|x_i)$ the density of the vector \tilde{Z}_i , conditionally to $X_i = x_i$. For any x_i, x'_i in $[0, 1]$, for any $z_i \in \mathbb{R}^{\sum_{J \in \mathcal{J}} 2^J}$,

$$\begin{aligned} \frac{\tilde{q}_i(z_i|x_i)}{\tilde{q}_i(z_i|x'_i)} &= \prod_{J \in \mathcal{J}} \prod_{k=0}^{2^J-1} \exp \left[\sqrt{2} \frac{|z_{i,k} - \varphi_{2^J,k}(x'_i)| - |z_{i,k} - \varphi_{2^J,k}(x_i)|}{\tilde{\sigma}_{2^J}} \right] \\ &\leq \exp \left[\sum_{J \in \mathcal{J}} \sum_{k=0}^{2^J-1} \frac{\sqrt{2}}{\tilde{\sigma}_{2^J}} \left(|\varphi_{2^J,k}(x'_i)| + |\varphi_{2^J,k}(x_i)| \right) \right]. \end{aligned}$$

Since $\varphi_{2^J,k}(x_i) \neq 0$ for a single value of $k \in \{0, \dots, 2^J - 1\}$, we get

$$\frac{\tilde{q}_i(z_i|x_i)}{\tilde{q}_i(z_i|x'_i)} \leq \exp \left[2\sqrt{2} \sum_{J \in \mathcal{J}} \frac{\|\varphi_{2^J,k}\|_\infty}{\tilde{\sigma}_{2^J}} \right] \leq e^\alpha,$$

by definition of $\tilde{\sigma}_2^J$, which concludes the proof by application of Lemma 4.1. \square

Note that $|\mathcal{J}| \leq 1 + 2 \log_2(n)$, hence we will have a logarithmic loss for the separation rates due to the privacy condition for the aggregated procedure.

Let us now define the adaptive test. We set, for all $J \in \mathcal{J}$,

$$\tilde{T}_J = \frac{1}{n(n-1)} \sum_{i \neq l=1}^n \sum_{k=0}^{2^J-1} \left(\tilde{Z}_{i,2^J,k} - \alpha_{2^J,k}^0 \right) \left(\tilde{Z}_{l,2^J,k} - \alpha_{2^J,k}^0 \right). \quad (27)$$

For a given level $\gamma \in (0, 1)$, the aggregated testing procedure rejects the hypothesis $H_0 : f = f_0$ if

$$\exists J \in \mathcal{J}, \tilde{T}_J > \tilde{t}_J^0(1 - u_\gamma),$$

where u_γ is defined by

$$u_\gamma = \sup \left\{ u \in (0, 1), \mathbb{P}_{Q_{f_0}^n} \left(\sup_{J \in \mathcal{J}} \left(\tilde{T}_J - \tilde{t}_J^0(1 - u_\gamma) \right) > 0 \right) \leq \gamma \right\} \quad (28)$$

and $\tilde{t}_J^0(1 - u_\gamma)$ denotes the $1 - u_\gamma$ quantile of \tilde{T}_J under H_0 . Hence u_γ is the least conservative choice leading to a γ -level test. We easily notice that $u_\gamma \geq \gamma/|\mathcal{J}|$. Indeed,

$$\begin{aligned} \mathbb{P}_{Q_{f_0}^n} \left(\sup_{J \in \mathcal{J}} \left(\tilde{T}_J - \tilde{t}_J^0(1 - \gamma/|\mathcal{J}|) \right) > 0 \right) &\leq \sum_{J \in \mathcal{J}} \mathbb{P}_{Q_{f_0}^n} \left(\tilde{T}_J > \tilde{t}_J^0(1 - \gamma/|\mathcal{J}|) \right) \\ &\leq \sum_{J \in \mathcal{J}} \gamma/|\mathcal{J}| \leq \gamma. \end{aligned}$$

Let us now consider the second kind error for the aggregated test, which is the probability to accept the null hypothesis H_0 , although the alternative hypothesis H_1 holds. This quantity is upper bounded by the smallest second kind error of the tests of the collection, at the price that γ has been replaced by u_γ . Indeed,

$$\begin{aligned} \mathbb{P}_{Q_f^n} \left(\sup_{J \in \mathcal{J}} \left(\tilde{T}_J - \tilde{t}_J^0(1 - u_\gamma) \right) \leq 0 \right) &= \mathbb{P}_{Q_f^n} \left(\bigcap_{J \in \mathcal{J}} \left(\tilde{T}_J \leq \tilde{t}_J^0(1 - u_\gamma) \right) \right) \\ &\leq \inf_{J \in \mathcal{J}} \mathbb{P}_{Q_f^n} \left(\tilde{T}_J \leq \tilde{t}_J^0(1 - u_\gamma) \right). \end{aligned} \quad (29)$$

We obtain the following theorem for the aggregated procedure.

Theorem 5.2. *Let (X_1, \dots, X_n) be i.i.d. with common density f in $\mathbb{L}_2([0, 1])$. Let f_0 be some given density in $\mathbb{L}_2([0, 1])$. From the observation of the random vectors $(\tilde{Z}_i, 1 \leq i \leq n)$ defined in Lemma 5.1 for a given $\alpha > 0$, we want to test the hypotheses*

$$H_0 : f = f_0, \quad \text{versus} \quad H_1 : f \neq f_0.$$

We assume that $n\alpha^2 / \log^{5/2}(n) \geq 1$.

We consider the set $\mathcal{J} = \{J \in \mathbb{N}, 2^J \leq n^2\}$ and the aggregated test

$$\Delta_{\gamma, Q}^{\mathcal{J}} = \mathbf{1}_{\{\sup_{J \in \mathcal{J}} (\tilde{T}_J - \tilde{t}_J^0(1-u_\gamma)) > 0\}}$$

where \tilde{T}_J is defined by Equation (27) and u_γ by Equation (28).

The uniform separation rate, defined by Equation (7), of the test $\Delta_{\gamma, Q}^{\mathcal{J}}$ over the set $\mathcal{B}_{s, 2, \infty}(R)$ defined by Equation (10) satisfies for all $n \in \mathbb{N}^*$, $s > 0$, $R > 0$, $\alpha > 0$, $(\gamma, \beta) \in (0, 1)^2$ such that $\gamma + \beta < 1$,

$$\begin{aligned} \rho_n \left(\Delta_{\gamma, Q}^{\mathcal{J}}, \mathcal{B}_{s, 2, \infty}(R), \beta, f_0 \right) \\ \leq C(\|f_0\|_2, R, \gamma, \beta) \left[(n\alpha^2 / \log^{5/2}(n))^{-2s/(4s+3)} \vee (n/\sqrt{\log(n)})^{-2s/(4s+1)} \right], \end{aligned}$$

The proof of this result is in Section A.4 of the appendix. We compare this result with the rates obtained in Theorem 4.9, which has been proved to be optimal. Here, we incur a logarithmic loss due to the adaptation. We recall that in the usual setting, the separation rates obtained by [35] and [28] for adaptive procedures over Besov balls was $(n/\sqrt{\log \log(n)})^{-2s/(4s+1)}$. This result was proved to be optimal for adaptive tests in [35]. In their paper, the log-log term is obtained from exponential inequalities for U-statistics involved in the testing procedure under the null hypothesis. In our setting, obtaining exponential inequalities is not trivial due to the Laplace noise. That is why our logarithmic loss originates from a simple upper bound on the variance of our test statistic under the null. The optimality of the adaptive rates presented in Theorem 5.2 remains an open question.

6. Discussion

Our study of minimax testing rates is in line with Ingster's work and we focus on separation rates in \mathbb{L}_2 -norm for goodness-of-fit testing under local differential privacy. We construct a unified setting in order to tackle both discrete and continuous distributions. In the continuous case, we provide the first minimax optimal test and local differentially private channel for the problem of goodness-of-fit testing over Besov balls. This result also holds for multinomial distributions. Besides, in the continuous case, the test and channel remain optimal up to a log factor even if the smoothness parameter is unknown. Among our technical contributions, it is to note that we use a proof technique in the lower bound that could not involve Theorem 1 from [19]. The minimax separation rate turns out to suffer from a polynomial degradation in the private case. However, we point out an elbow effect, where the rate is the same as in the non-private case up to some constant factor if α is large enough. Simultaneously and independently, [10] present minimax testing rates for the l_1 and l_2 norms in the discrete case. We define Besov balls using Haar wavelets, which are equivalent to Besov balls defined using moduli of smoothness when $s < 1$. In order for the equivalence to hold for any s , it is possible to define Besov balls using Daubechies wavelets instead. In the proof of our lower bound, we use the disjoint support property of the Haar wavelets, but this can be circumvented taking fewer wavelets in the definition of the prior distributions. A more critical assumption is that $\varphi_{L,k}^2 = \sqrt{L}\varphi_{L,k}$. Future possible works could extend our results to larger Besov classes and study the optimality of the adaptive procedure. Our bounds match when f_0 is nearly uniform, and matching bounds for any f_0 remain to be proved under non-interactive differential privacy. Finally, finding an interactive privacy channel leading to faster rates than our lower bound would prove that interactive privacy can lead to better results than it is possible under non-interactive privacy for goodness-of-fit testing.

Appendix A: Proof of the results

A.1. Lower bound: proof of Lemma 3.1

Since $\nu_\rho(\mathcal{F}_\rho(\mathcal{C})) \geq 1 - \delta$, we first note that

$$\begin{aligned}
\inf_{\Delta_{\gamma,Q}} \sup_{f \in \mathcal{F}_\rho(\mathcal{C})} \mathbb{P}_{Q_f^n} (\Delta_{\gamma,Q}(Z_1, \dots, Z_n) = 0) &\geq \inf_{\Delta_{\gamma,Q}} \mathbb{P}_{Q_{\nu_\rho}^n} (\Delta_{\gamma,Q}(Z_1, \dots, Z_n) = 0) - \delta \\
&= \inf_{\Delta_{\gamma,Q}} (\mathbb{P}_{Q_{f_0}^n} (\Delta_{\gamma,Q}(Z_1, \dots, Z_n) = 0) \\
&\quad + \mathbb{P}_{Q_{\nu_\rho}^n} (\Delta_{\gamma,Q}(Z_1, \dots, Z_n) = 0) \\
&\quad - \mathbb{P}_{Q_{f_0}^n} (\Delta_{\gamma,Q}(Z_1, \dots, Z_n) = 0)) - \delta \\
&\geq 1 - \gamma - \sup_{\Delta_{\gamma,Q}} \left| \mathbb{P}_{Q_{\nu_\rho}^n} (\Delta_{\gamma,Q}(Z_1, \dots, Z_n) = 0) \right. \\
&\quad \left. - \mathbb{P}_{Q_{f_0}^n} (\Delta_{\gamma,Q}(Z_1, \dots, Z_n) = 0) \right| - \delta
\end{aligned}$$

by definition of $\Delta_{\gamma,Q}(Z_1, \dots, Z_n)$. Finally, by definition of the total variation distance,

$$\inf_{\Delta_{\gamma,Q}} \sup_{f \in \mathcal{F}_\rho(\mathcal{C})} \mathbb{P}_{Q_f^n} (\Delta_{\gamma,Q}(Z_1, \dots, Z_n) = 0) \geq 1 - \gamma - \delta - \|\mathbb{P}_{Q_{\nu_\rho}^n} - \mathbb{P}_{Q_{f_0}^n}\|_{TV}.$$

So we have

$$\inf_{\Delta_{\gamma,Q}} \sup_{f \in \mathcal{F}_\rho(\mathcal{C})} \mathbb{P}_{Q_f^n} (\Delta_{\gamma,Q}(Z_1, \dots, Z_n) = 0) > \beta,$$

provided that

$$\|\mathbb{P}_{Q_{\nu_\rho}^n} - \mathbb{P}_{Q_{f_0}^n}\|_{TV} < 1 - \gamma - \beta - \delta.$$

A.2. Lower bound: proof of Theorem 3.4

An initial version of this proof has been presented in a preprint of [40], which was then improved upon by [13] in order to find the matching rate in α and to account for different channels Q_i for each initial observation X_i . The proof remains fundamentally the same, however. In line with the rest of the paper, both the discrete and the continuous cases are treated in one unified setting.

In this section, $f_0 = \mathbf{1}_{[0,1]}$.

A.2.1. Preliminary results

The following lemma sheds light on the equivalence between the local differential privacy condition and a similar condition on the density of the channel.

Lemma A.1. *Let $Q \in \mathcal{Q}_\alpha$ be an α -private channel and $i \leq n$. Let X_i be a random variable with density $f \in \mathbb{L}_2([0, 1])$ with respect to the Lebesgue measure. Then there exists a probability measure with respect to which $Q_i(\cdot|x)$ is absolutely continuous for any $x \in [0, 1]$.*

Proof. Let $\mu_i = \int_{[0,1]} Q_i(\cdot|x)f(x)dx$. Let $S \in \mathcal{Z}_i$ such that $\mu_i(S) = 0$. Then since $Q_i(S|x) \geq 0$ for any x , there exists x such that $Q_i(S|x) = 0$. Now by α -local differential privacy, $Q_i(S|x) = 0$ for any x . \square

For the sake of completeness, we prove the following classical inequality between the total variation distance and the chi-squared distance. It will be used in order to reduce the study of the distance between the distributions to that of an expected squared likelihood ratio.

Lemma A.2.

$$\|\mathbb{P}_{Q_{\nu_\rho}^n} - \mathbb{P}_{Q_{f_0}^n}\|_{TV} \leq \frac{1}{2} \left(\mathbb{E}_{Q_{f_0}^n} \left[L_{Q_{\nu_\rho}^n}^2(Z_1, \dots, Z_n) - 1 \right] \right)^{1/2},$$

where $L_{Q_{\nu_\rho}^n}(Z_1, \dots, Z_n)$ is the likelihood ratio between $Q_{\nu_\rho}^n$ and $Q_{f_0}^n$.

Proof. We have

$$\begin{aligned} \|\mathbb{P}_{Q_{\nu_\rho}^n} - \mathbb{P}_{Q_{f_0}^n}\|_{TV} &= \frac{1}{2} \int |L_{Q_{\nu_\rho}^n} - 1| d\mathbb{P}_{Q_{f_0}^n} = \frac{1}{2} \mathbb{E}_{Q_{f_0}^n} \left[|L_{Q_{\nu_\rho}^n}(Z_1, \dots, Z_n) - 1| \right] \\ &\leq \frac{1}{2} \left(\mathbb{E}_{Q_{f_0}^n} \left[L_{Q_{\nu_\rho}^n}^2(Z_1, \dots, Z_n) - 1 \right] \right)^{1/2}, \end{aligned}$$

by Cauchy-Schwarz inequality and since $\mathbb{E}_{Q_{f_0}^n} \left(L_{Q_{\nu_\rho}^n}(Z_1, \dots, Z_n) \right) = 1$. \square

The following two lemmas can be interpreted as data processing inequalities. Lemma A.3 describes the contraction of the total variation distance by a stochastic channel.

Lemma A.3. Let $\mathbb{P}_f, \mathbb{P}_g$ be probability measures over the sample space $[0, 1]$ with respective densities f and g with respect to the Lebesgue measure. Let Q be a stochastic channel. Then

$$\|\mathbb{P}_f - \mathbb{P}_g\|_{TV} \geq \|\mathbb{P}_{Q_f} - \mathbb{P}_{Q_g}\|_{TV}.$$

Proof. For any measurable set S ,

$$\begin{aligned} \int_{[0,1]} Q(S|x)(f(x) - g(x))dx &= \int_{[0,1]} Q(S|x)(f(x) - g(x)) \mathbf{1}_{\{f-g \geq 0\}}(x)dx \\ &\quad + \int_{[0,1]} Q(S|x)(f(x) - g(x)) \mathbf{1}_{\{f-g < 0\}}(x)dx. \end{aligned}$$

Now, since $0 \leq Q(S|x) \leq 1$ for any measurable set S and $x \in [0, 1]$,

$$0 \leq \int_{[0,1]} Q(S|x)(f(x) - g(x)) \mathbf{1}_{\{f-g \geq 0\}}(x)dx \leq \int_{[0,1]} (f(x) - g(x)) \mathbf{1}_{\{f-g > 0\}}(x)dx.$$

and

$$0 \geq \int_{[0,1]} Q(S|x)(f(x) - g(x)) \mathbf{1}_{\{f-g < 0\}}(x)dx \geq \int_{[0,1]} (f(x) - g(x)) \mathbf{1}_{\{f-g < 0\}}(x)dx.$$

So for any measurable set S ,

$$\begin{aligned} \int_{[0,1]} (f(x) - g(x)) \mathbf{1}_{\{f-g < 0\}}(x) dx \\ \leq \int_{[0,1]} Q(S|x)(f(x) - g(x)) dx \\ \leq \int_{[0,1]} (f(x) - g(x)) \mathbf{1}_{\{f-g > 0\}}(x) dx. \end{aligned}$$

That is, for any measurable set S

$$\begin{aligned} \left| \int_{[0,1]} Q(S|x)(f(x) - g(x)) dx \right| &\leq \left| \int_{[0,1]} (f(x) - g(x)) \mathbf{1}_{\{f-g > 0\}}(x) dx \right| \\ &\vee \left| \int_{[0,1]} (f(x) - g(x)) \mathbf{1}_{\{f-g < 0\}}(x) dx \right| \\ &= \sup_A \left| \int_A (f(x) - g(x)) dx \right| = \|\mathbb{P}_f - \mathbb{P}_g\|_{TV}. \end{aligned}$$

□

A.2.2. Definition of prior distributions

By Lemma A.1, let $Q \in \mathcal{Q}_\alpha$ be a non-interactive α -private channel with marginal conditional densities $q_i(z_i|x_i)$ with respect to probability measure μ_i over the respective sample space $\tilde{\Omega}_i$ for any $1 \leq i \leq n$. In the discrete case, we assume that p^0 is a uniform probability vector. By Equation (5), we can consider the associated uniform density on $[0, 1]$. So in both the continuous and the discrete cases, we end up considering a uniform density f_0 over $[0, 1]$. Let $\tilde{f}_{0,i}(z_i) = \int_0^1 q_i(z_i|x) f_0(x) dx = \int_0^1 q_i(z_i|x) dx$ with the convention $0/0 = 0$. Let $\mu_i = \int_{[0,1]} Q_i(\cdot|x) f_0(x) dx$ and $K_i : \mathbb{L}_2([0, 1]) \rightarrow \mathbb{L}_2(\tilde{\Omega}_i, d\mu_i)$ such that

$$K_i f = \int_0^1 q_i(\cdot|x) f(x) \frac{dx}{\sqrt{\tilde{f}_{0,i}(\cdot)}}.$$

Let K_i^* denote the adjoint of K_i . Then $K_i^* K_i$ is a symmetric integral operator with kernel

$$F_i(x, y) = \int \frac{q_i(z_i|x) q_i(z_i|y)}{\tilde{f}_{0,i}(z_i)} d\mu_i(z_i). \quad (30)$$

And by Fubini's theorem, for any $f \in \mathbb{L}_2([0, 1])$:

$$K_i^* K_i f(\cdot) = \int_0^1 F_i(\cdot, y) f(y) dy.$$

Note that f_0 is an eigenfunction of $K_i^* K_i$ associated to the eigenvalue $\lambda_{0,i} = 1$ for all $1 \leq i \leq n$. Let

$$K = \sum_{i=1}^n K_i^* K_i / n,$$

which is symmetric and positive semidefinite, and $\lambda_0 = 1$ is an eigenvalue associated with f_0 . It is an integral operator with kernel

$$F(x, y) = \sum_i F_i(x, y)/n.$$

We denote by ψ the difference of indicator functions: $\psi = \mathbb{1}_{[0,1/2)} - \mathbb{1}_{[1/2,1)}$ and for all integer $L \geq 1$, we set, for all $k \in \{0, \dots, L-1\}$, for all $x \in [0, 1)$,

$$\psi_k(x) = \sqrt{L}\psi(Lx - k).$$

The integer L will be taken as $L = 2^J$ for some $J \geq 0$ in the continuous case, and we choose $L = d/2$ in the discrete case (we assume that d is even). We denote by V the linear subspace of $\mathbb{L}^2([0, 1])$ generated by the functions $(f_0, \psi_k, k \in \{0, 1, \dots, L-1\})$. Then we complete (f_0) into an orthogonal basis $(f_0, u_i)_{1 \leq i \leq L}$ of V with eigenfunctions of K such that $\int u_i(x)dx = 0$ by orthogonality with f_0 and $\|u_i\|_2 = 1$. We write the corresponding eigenvalues λ_i .

Let $z_\alpha = e^{2\alpha} - e^{-2\alpha} \leq 2$ for any $\alpha \in (0, 1]$. Let $\tilde{\lambda}_k = (\lambda_k/z_\alpha^2) \vee L^{-1} \geq L^{-1}$. Let

$$f_\eta(x) = f_0(x) + \varepsilon \sum_{j=1}^L \eta_j \tilde{\lambda}_j^{-1/2} u_j(x),$$

where $\eta \in \{-1, 1\}^L$. For all $i \in \{1, \dots, L\}$, $u_i \in \text{Span}(\psi_k, k \in \{0, 1, \dots, L-1\})$, hence we write

$$u_i = \sum_{k=0}^{L-1} a_{i,k} \psi_k.$$

Then

$$f_\eta(x) = f_0(x) + \varepsilon \sum_{j=1}^L \sum_{k=0}^{L-1} \eta_j a_{j,k} \tilde{\lambda}_j^{-1/2} \psi_k(x).$$

We define ν_ρ as the uniform probability measure over $\{f_\eta : \eta \in \{-1, 1\}^L\}$. Now, we can identify the distance between f_η and f_0 . Let $l = \sum_{i=1}^L \mathbb{1}_{\{z_\alpha^{-2} \lambda_i > L^{-1}\}}$. By definition and orthonormality of $(u_i)_{1 \leq i \leq L}$, for any $\eta \in \{-1, 1\}^L$

$$\begin{aligned} \|f_\eta - f_0\|_2 &= \varepsilon \sqrt{\sum_{i=1}^L \tilde{\lambda}_i^{-1} \|u_i\|_2^2} = \varepsilon \sqrt{\sum_{i=1}^L \tilde{\lambda}_i^{-1}} \\ &= \varepsilon \sqrt{\sum z_\alpha^2 \lambda_i^{-1} \mathbb{1}_{\{z_\alpha^{-2} \lambda_i > L^{-1}\}} + L \sum \mathbb{1}_{\{z_\alpha^{-2} \lambda_i \leq L^{-1}\}}} \\ &\geq \varepsilon \sqrt{z_\alpha^2 l^2 \left(\sum_i \lambda_i \mathbb{1}_{\{z_\alpha^{-2} \lambda_i > L^{-1}\}}\right)^{-1} + L(L-l)}, \end{aligned} \quad (31)$$

by Cauchy-Schwarz inequality.

So let us provide guarantees on the singular values in order to determine sufficient conditions for ε to lead to a lower bound on $\rho_n^*(\Delta_{\gamma, Q}, \mathcal{C}, \beta, f_0)$, depending on \mathcal{C} .

A.2.3. Obtaining the inequalities on the eigenvalues

Lemma A.4. *Let K be defined as in Section A.2.2 and $(\lambda_i^2)_{0 \leq i \leq L}$ its eigenvalues associated with the orthonormal basis $(f_0, u_i)_{1 \leq i \leq L}$. Then the following inequality holds.*

$$\sum_{k=1}^L \lambda_k \leq z_\alpha^2.$$

Proof. We have

$$\begin{aligned} \sum_{k=1}^L \lambda_k &= \sum_{k=1}^L \int_0^1 \int_0^1 \frac{u_k(x)u_k(y)}{n} \sum_{i=1}^n F_i(x, y) dx dy \\ &= \frac{1}{n} \sum_{i=1}^n \int_{\tilde{\Omega}_i} \sum_{k=1}^L \left(\int_0^1 \frac{q_i(z_i|x)}{\tilde{f}_{0,i}(z_i)} u_k(x) dx \right)^2 \tilde{f}_{0,i}(z_i) d\mu_i(z_i) \\ &= \frac{1}{n} \sum_{i=1}^n \int_{\tilde{\Omega}_i} \sum_{k=1}^L \left(\int_0^1 \left(\frac{q_i(z_i|x)}{\tilde{f}_{0,i}(z_i)} - e^{-2\alpha} \right) u_k(x) dx \right)^2 \tilde{f}_{0,i}(z_i) d\mu_i(z_i), \end{aligned}$$

since $\int u_k(x) dx = 0$. Now we define $f_{z,i}(x) = \frac{q_i(z_i|x)}{\tilde{f}_{0,i}(z_i)} - e^{-2\alpha}$ and by Lemma 4.1,

$$0 \leq e^{-\alpha} - e^{-2\alpha} \leq f_{z,i}(x) = \left(\int_0^1 \frac{q_i(z_i|s)}{q_i(z_i|x)} ds \right)^{-1} - e^{-2\alpha} \leq e^\alpha - e^{-2\alpha} \leq e^{2\alpha} - e^{-2\alpha}.$$

So $\|f_{z,i}\|_2 \leq e^{2\alpha} - e^{-2\alpha}$. Then, by orthonormality of the u_k 's, we apply Parseval's inequality:

$$\sum_{k=1}^L \left(\int_0^1 \left(\frac{q_i(z_i|x)}{\tilde{f}_{0,i}(z_i)} - e^{-2\alpha} \right) u_k(x) dx \right)^2 = \sum_{k=1}^L \langle f_{z,i}, u_k \rangle^2 = \left\| \sum_{k=1}^L \langle f_{z,i}, u_k \rangle u_k \right\|_2^2 \leq \|f_{z,i}\|_2^2 \leq z_\alpha^2.$$

Finally, $\int_{\tilde{\Omega}_i} \tilde{f}_{0,i}(z_i) d\mu_i(z_i) = 1$ leads to $\sum \lambda_k \leq z_\alpha^2$. □

Then from Equation (31) and by application of Lemma A.4,

$$\|f_\eta - f_0\|_2 \geq L\varepsilon \sqrt{(L^{-1}l)^2 + 1 - (L^{-1}l)} \geq L\varepsilon \sqrt{3/4}. \quad (32)$$

So for the discrete case, by Equation (5),

$$\sqrt{\sum_{i=0}^{d-1} (p_i - p_i^0)^2} \geq \sqrt{d\varepsilon} \sqrt{3/4}. \quad (33)$$

A.2.4. Information bound

Let $\varepsilon > 0$, for all $\eta \in \{-1, 1\}^L$, we define

$$\tilde{f}_{\eta,i}(z) = \tilde{f}_{0,i}(z) + \varepsilon \sum_{j=1}^L \eta_j \tilde{\lambda}_j^{-1/2} \int_0^1 q_i(z_i|x) u_j(x) dx.$$

We consider the expected squared likelihood ratio:

$$\begin{aligned}
\mathbb{E}_{Q_{f_0}^n} \left[L_{Q_{\nu\rho}^n}^2(Z_1, \dots, Z_n) \right] &= \mathbb{E}_{Q_{f_0}^n} \mathbb{E}_{\eta, \eta'} \prod_{i=1}^n \left(1 + \varepsilon \frac{\sum_{j=1}^L \tilde{\lambda}_j^{-1/2} \eta_j \int_0^1 q_i(Z_i|x) u_j(x) dx}{\tilde{f}_0(Z_i)} \right) \\
&\quad \left(1 + \varepsilon \frac{\sum_{j=1}^L \tilde{\lambda}_j^{-1/2} \eta'_j \int_0^1 q_i(Z_i|x) u_j(x) dx}{\tilde{f}_0(Z_i)} \right) \\
&= \mathbb{E}_{Q_{f_0}^n} \mathbb{E}_{\eta, \eta'} \prod_{i=1}^n \left(1 + \frac{\varepsilon \sum_{j=1}^L \tilde{\lambda}_j^{-1/2} \eta_j \int_0^1 q_i(Z_i|x) u_j(x) dx}{\tilde{f}_0(Z_i)} \right. \\
&\quad \left. + \frac{\varepsilon \sum_{j=1}^L \tilde{\lambda}_j^{-1/2} \eta'_j \int_0^1 q_i(Z_i|x) u_j(x) dx}{\tilde{f}_0(Z_i)} \right. \\
&\quad \left. + \frac{\varepsilon^2 \sum_{j,l=1}^L \tilde{\lambda}_j^{-1/2} \tilde{\lambda}_l^{-1/2} \eta_j \eta'_l \int_0^1 q_i(Z_i|x) u_j(x) dx \int_0^1 q_i(Z_i|y) u_l(y) dy}{\tilde{f}_0(Z_i)^2} \right).
\end{aligned}$$

Now, for any j ,

$$\mathbb{E}_{Q_{f_0}} \left[\frac{\int_0^1 q_i(Z_i|x) u_j(x) dx}{\tilde{f}_0(Z_i)} \right] = \int_0^1 \int_{\tilde{\Omega}_i} q_i(z|x) d\mu_i(z) u_j(x) dx = \int_0^1 u_j(x) dx = 0,$$

by orthogonality with uniform vector f_0 .

And, by Equation (30),

$$\mathbb{E}_{Q_{f_0}} \left[\frac{\int_0^1 q_i(Z_i|x) u_j(x) dx \int_0^1 q_i(Z_i|y) u_l(y) dy}{\tilde{f}_0(Z_i)^2} \right] = \int_0^1 \int_0^1 F_i(x, y) u_j(x) u_l(y) dx dy.$$

So since $1 + u \leq \exp u$ for any u ,

$$\begin{aligned}
\mathbb{E}_{Q_{f_0}^n} \left[L_{Q_{\nu\rho}^n}^2(Z_1, \dots, Z_n) \right] \\
\leq \mathbb{E}_{\eta, \eta'} \exp \left(\varepsilon^2 \sum_{j,l=1}^L \tilde{\lambda}_j^{-1/2} \tilde{\lambda}_l^{-1/2} \eta_j \eta'_l n \int_0^1 \int_0^1 F(x, y) u_j(x) u_l(y) dx dy \right).
\end{aligned}$$

Now

$$\int_0^1 \int_0^1 F(x, y) u_j(x) u_l(y) dx dy = \lambda_j \int_0^1 u_j(x) u_l(x) dx = \lambda_j \mathbf{1}_{\{j=l\}},$$

since u_j is an eigenfunction of K and by orthonormality. So

$$\mathbb{E}_{Q_{f_0}^n} \left[L_{Q_{\nu\rho}^n}^2(Z_1, \dots, Z_n) \right] \leq \mathbb{E}_{\eta, \eta'} \exp \left(n \varepsilon^2 \sum_{j=1}^L \tilde{\lambda}_j^{-1} \eta_j \eta'_j \lambda_j \right) \leq \mathbb{E}_{\eta, \eta'} \exp \left(n \varepsilon^2 \sum_{j=1}^L \eta_j \eta'_j z_\alpha^2 \right).$$

Then

$$\mathbb{E}_{Q_{f_0}^n} \left[L_{Q_{\nu_\rho}^n}^2(Z_1, \dots, Z_n) \right] \leq \prod_{j=1}^L \cosh(n\varepsilon^2 z_\alpha^2) \leq \prod_{j=1}^L \exp(n^2 \varepsilon^4 z_\alpha^4) \leq \exp(n^2 \varepsilon^4 z_\alpha^4 L).$$

Then, in order to apply Lemma 3.1 combined with A.2, let us find a sufficient condition for

$$\mathbb{E}_{Q_{f_0}^n} \left[L_{Q_{\nu_\rho}^n}^2(Z_1, \dots, Z_n) \right] < 1 + 4(1 - \gamma - \beta - \gamma)^2.$$

So let us choose ε and J in order to ensure that

$$\exp(n^2 \varepsilon^4 z_\alpha^4 L) < 1 + 4(1 - 2\gamma - \beta)^2,$$

i.e.

$$L\varepsilon^4 \leq (nz_\alpha^2)^{-2} \log \left[1 + 4(1 - 2\gamma - \beta)^2 \right],$$

i.e.

$$\varepsilon \leq (nz_\alpha^2)^{-1/2} \left(\frac{\log [1 + 4(1 - 2\gamma - \beta)^2]}{L} \right)^{1/4}. \quad (34)$$

A.2.5. Sufficient condition for f_η to be non-negative

Lemma A.5. *If*

$$\varepsilon \leq \frac{L^{-1}}{\sqrt{2 \log(2L/\gamma)}},$$

then there exists $A_\gamma \subset \{-1, 1\}^L$ such that $\mathbb{P}_{\nu_\rho}(\eta \in A_\gamma) \geq 1 - \gamma$ and for any $\eta \in A_\gamma$, f_η is a density.

Proof. Let

$$A_\gamma = \left\{ \eta : \left| \sum_j \eta_j a_{j,k} \tilde{\lambda}_j^{-1/2} \right| \leq \sqrt{2L \log(2L/\gamma)} \right\}.$$

Since for all i , u_i is orthogonal to f_0 , uniform density on $[0, 1]$, we have that $\int_0^1 f_\eta(x) dx = 1$ and we just have to prove that f_η is nonnegative. We remind the reader that

$$u_i = \sum_{k=0}^{L-1} a_{i,k} \psi_k.$$

The bases (u_1, \dots, u_L) and $(\psi_k, k \in \{0, 1, \dots, L-1\})$ are orthonormal. This implies that the matrix $A = (a_{i,k})_{1 \leq i \leq L, k \in \{0, 1, \dots, L-1\}}$ is orthogonal. So

$$\forall i, \sum_{k=0}^{L-1} a_{i,k}^2 = 1, \quad \forall k, \sum_{i=1}^L a_{i,k}^2 = 1.$$

Hence we have for all $x \in [0, 1]$,

$$(f_\eta - f_0)(x) = \sum_{k=0}^{L-1} \sum_{i=1}^L \eta_i \varepsilon \tilde{\lambda}_i^{-1/2} a_{i,k} \psi_k(x). \quad (35)$$

The functions $(\psi_k, k \in \{0, 1, \dots, L-1\})$ have disjoint supports and $\sup_{x \in [0,1]} |\psi_k(x)| = L^{1/2}$. Hence f_η is nonnegative if and only if for any $k \in \Lambda(J)$

$$L^{1/2} \left| \sum_{i=1}^L \eta_i \varepsilon \tilde{\lambda}_i^{-1/2} a_{i,k} \right| \leq 1. \quad (36)$$

By definition of ν_ρ , we have that f_η is a density with probability larger than $1 - \gamma$ under the prior ν_ρ as soon as Equation (36) holds with probability larger than $1 - \gamma$. That is,

$$\mathbb{P}_{\nu_\rho} \left(\forall k \in \{0, 1, \dots, L-1\}, L^{1/2} \left| \sum_{i=1}^L \eta_i \varepsilon \tilde{\lambda}_i^{-1/2} a_{i,k} \right| \leq 1 \right) \geq 1 - \gamma,$$

where (η_1, \dots, η_L) are i.i.d. Rademacher random variables. Using Hoeffding's inequality, we get for all $x > 0$, for all $k \in \{0, 1, \dots, L-1\}$,

$$\mathbb{P}_{\nu_\rho} \left(\left| \sum_{i=1}^L \eta_i \varepsilon \tilde{\lambda}_i^{-1/2} a_{i,k} \right| > x \right) \leq 2 \exp \left(\frac{-2x^2}{\sum_{i=1}^L (2\varepsilon \tilde{\lambda}_i^{-1/2} a_{i,k})^2} \right).$$

Hence

$$\mathbb{P}_{\nu_\rho} \left(\exists k \in \{0, 1, \dots, L-1\}, \left| \sum_{i=1}^L \eta_i \varepsilon \tilde{\lambda}_i^{-1/2} a_{i,k} \right| > x \right) \leq 2L \exp \left(\frac{-x^2}{2 \sum_{i=1}^L (\varepsilon \tilde{\lambda}_i^{-1/2} a_{i,k})^2} \right).$$

So the probability of having the existence of some $k \in \{0, 1, \dots, L-1\}$ such that

$$\left| \sum_{i=1}^L \eta_i \varepsilon \tilde{\lambda}_i^{-1/2} a_{i,k} \right| > \sqrt{2 \sum_{i=1}^L (\varepsilon \tilde{\lambda}_i^{-1/2} a_{i,k})^2 \log(2L/\gamma)}$$

is smaller than γ . Hence, f_η is a density with probability larger than $1 - \gamma$ under the prior ν_ρ as soon as for any $k \in \Lambda(J)$,

$$2L \sum_{i=1}^L (\varepsilon \tilde{\lambda}_i^{-1/2} a_{i,k})^2 \log(2L/\gamma) \leq 1.$$

Now by definition, $\tilde{\lambda}_i^{-1/2} \leq L^{1/2}$. So we have the sufficient condition

$$2L\varepsilon^2 L \log(2L/\gamma) \sum_{i=1}^L a_{i,k}^2 \leq 1.$$

And $\sum_{i=1}^L a_{ik}^2 = 1$ leads to the following sufficient condition,

$$\varepsilon \leq \frac{L^{-1}}{\sqrt{2 \log(2L/\gamma)}}.$$

□

A.2.6. Sufficient conditions for $f_\eta \in \mathcal{F}_\rho(\mathcal{B}_{s,2,\infty}(R))$, only in the continuous case

We first prove the following points.

Lemma A.6. *If*

$$\varepsilon \leq \frac{L^{-1}(1 \wedge RL^{-s})}{\sqrt{2 \log(2L/\gamma)}},$$

then there exists $A_\gamma \subset \{-1, 1\}^d$ such that $\mathbb{P}_{\nu_\rho}(\eta \in A_\gamma) \geq 1 - \gamma$ and for any $\eta \in A_\gamma$,

- a) f_η is a density.
- b) $f_\eta \in \mathcal{B}_{s,2,\infty}(R)$.

Proof. We consider the same event as in the proof of Lemma A.5:

$$A_\gamma = \left\{ \eta : \left| \sum_j \eta_j a_{j,k} \tilde{\lambda}_j^{-1/2} \right| \leq \sqrt{2L \log(2L/\gamma)} \right\}.$$

- a) In the same way as in Lemma A.5, f_η is a density.
- b) For all $k \in \{0, 1, \dots, L-1\}$,

$$\langle f_\eta - f_0, \psi_k \rangle = \varepsilon \sum_{i=1}^L \eta_i \tilde{\lambda}_i^{-1/2} a_{i,k}.$$

Hence $f_\eta \in \mathcal{B}_{s,2,\infty}(R)$ if and only if

$$\sum_{k=0}^{L-1} \varepsilon^2 \left(\sum_{i=1}^L \eta_i \tilde{\lambda}_i^{-1/2} a_{i,k} \right)^2 \leq R^2 L^{-2s}.$$

But we also have that for any $\eta \in A_\gamma$,

$$\sum_{k=0}^{L-1} \varepsilon^2 \left(\sum_{i=1}^L \eta_i \tilde{\lambda}_i^{-1/2} a_{i,k} \right)^2 \leq \varepsilon^2 L^2 2 \log(2L/\gamma).$$

So $f_\eta \in \mathcal{B}_{s,2,\infty}(R)$ if

$$\varepsilon \leq RL^{-(s+1)} / \sqrt{2 \log(2L/\gamma)}.$$

A.2.7. Conclusion

1. Discrete case.

So combining Equation (34) and Lemma A.5, we obtain the following sufficient condition in order to apply Lemma 3.1:

$$\varepsilon \leq \left[(nz_\alpha^2)^{-1/2} \left(\frac{\log [1 + 4(1 - 2\gamma - \beta)^2]}{L} \right)^{1/4} \right] \wedge \frac{L^{-1}}{\sqrt{2 \log(2L/\gamma)}}.$$

So, by Equation (33), if

$$\sqrt{\sum_{k=0}^{d-1} (p_k - p_k^0)^2} \leq \sqrt{3/4} \left(\left[(nz_\alpha^2)^{-1/2} d^{1/4} \left(\log [1 + 4(1 - 2\gamma - \beta)^2] \right)^{1/4} \right] \wedge \frac{d^{-1/2}}{\sqrt{2 \log(2d/\gamma)}} \right),$$

then we can define densities f_η such that the errors are larger than γ and β . So

$$\begin{aligned} & \inf_{\Delta_{\gamma, Q}} \rho_n(\Delta_{\gamma, Q}, \mathcal{D}, \beta, f_0) / d^{1/2} \\ & \geq c(\gamma, \beta) \left[((nz_\alpha^2)^{-1/2} d^{1/4}) \wedge (d \log d)^{-1/2} \right]. \end{aligned}$$

Now, we also have

$$\rho_n^*(\mathcal{D}, \alpha, \gamma, \beta, f_0) \geq \rho_n^*(\mathcal{D}, +\infty, \gamma, \beta, f_0),$$

where $\rho_n^*(\mathcal{D}, +\infty, \gamma, \beta, f_0)$ corresponds to the case where there is no local differential privacy condition on Q . In particular, taking Q such that $Z = X$ with probability 1 reduces the private problem to the classical testing problem. Now, the data processing inequality in Lemma A.3 justifies that such a Q is optimal by contraction of the total variation distance. And the classical result leads to having $\rho_n^*(\mathcal{C}, +\infty, \gamma, \beta, f_0) = c(\gamma, \beta) n^{-1/2} d^{-1/4}$.

So, we have

$$\rho_n^*(\mathcal{D}, \alpha, \gamma, \beta, f_0) / d^{1/2} \geq c(\gamma, \beta) \left[((nz_\alpha^2)^{-1/2} d^{1/4}) \wedge (d \log d)^{-1/2} \right] \vee (n^{-1/2} d^{-1/4}).$$

2. Continuous case.

So combining Equation (34) and Lemma A.6, we obtain the following sufficient condition in order to apply Lemma 3.1:

$$\varepsilon \leq \left[(nz_\alpha^2)^{-1/2} \left(\frac{\log [1 + 4(1 - 2\gamma - \beta)^2]}{L} \right)^{1/4} \right] \wedge \frac{L^{-1}(1 \wedge RL^{-s})}{\sqrt{2 \log(2L/\gamma)}}.$$

So, by Equation (32), if

$$\|f - f_0\|_2 \leq \sqrt{3/4} \left(\left[(nz_\alpha^2)^{-1/2} L^{3/4} \log^{1/4} (1 + 4(1 - 2\gamma - \beta)^2) \right] \wedge \frac{(1 \wedge RL^{-s})}{\sqrt{2 \log(2L/\gamma)}} \right),$$

then, taking J as the largest integer such that $2^J \leq c(\gamma, \beta, R) (nz_\alpha^2)^{2/(4s+3)}$, we obtain:

$$\inf_{\Delta_{\gamma, Q}} \rho_n(\Delta_{\gamma, Q}, \mathcal{B}_{s, 2, \infty}(R), \beta, f_0) \geq c(\gamma, \beta, R) (nz_\alpha^2)^{-2s/(4s+3)} (\log n)^{-1/2}.$$

Now, we also have

$$\rho_n^*(\mathcal{B}_{s, 2, \infty}(R), \alpha, \gamma, \beta, f_0) \geq \rho_n^*(\mathcal{B}_{s, 2, \infty}(R), +\infty, \gamma, \beta, f_0),$$

where $\rho_n^*(\mathcal{B}_{s, 2, \infty}(R), +\infty, \gamma, \beta, f_0)$ corresponds to the case where there is no local differential privacy condition on Q . In particular, taking Q such that $Z = X$ with probability 1 reduces the private problem to the classical testing problem. Now, the data processing inequality in Lemma A.3 justifies that such a Q is optimal by contraction of the total variation distance. And the classical result leads to having $\rho_n^*(\mathcal{B}_{s, 2, \infty}(R), +\infty, \gamma, \beta, f_0) = c(\gamma, \beta, R) n^{-2s/(4s+1)}$. So, we have

$$\rho_n^*(\mathcal{B}_{s, 2, \infty}(R), \alpha, \gamma, \beta, f_0) \geq c(\gamma, \beta, R) [(nz_\alpha^2)^{-2s/(4s+3)} (\log n)^{-1/2} \vee n^{-2s/(4s+1)}].$$

□

A.3. Proof of the upper bound

In this section, f_0 is some fixed density in $\mathbb{L}_2([0, 1])$.

A.3.1. Proof of Theorem 4.3

We prove the bound on the variance term $\text{Var}_{Q_f^n}(\hat{T}_L)$ given in Equation (19). Let us define

$$\hat{U}_L = \frac{1}{n(n-1)} \sum_{i \neq l=1}^n \sum_{k=0}^{L-1} (Z_{i,L,k} - \alpha_{L,k}) (Z_{l,L,k} - \alpha_{L,k}),$$

$$\hat{V}_L = 2 \sum_{k=0}^{L-1} (\alpha_{L,k} - \alpha_{L,k}^0) \frac{1}{n} \sum_{i=1}^n (Z_{i,L,k} - \alpha_{L,k}),$$

where $\alpha_{L,k} = \int_0^1 \varphi_{L,k}(x) f(x) dx$ and $\alpha_{L,k}^0 = \int_0^1 \varphi_{L,k}(x) f_0(x) dx$. Then we obtain the Hoeffding's decomposition of the U-statistic \hat{T}_L , namely

$$\hat{T}_L = \hat{U}_L + \hat{V}_L + \|\Pi_{S_L}(f - f_0)\|_2^2.$$

We first control the variance of the degenerate U-statistic \hat{U}_L which can be written as

$$\hat{U}_L = \frac{1}{n(n-1)} \sum_{i \neq l=1}^n h_L(Z_{i,L}, Z_{l,L}),$$

where

$$h_L(Z_{i,L}, Z_{l,L}) = \sum_{k=0}^{L-1} (Z_{i,L,k} - \alpha_{L,k}) (Z_{l,L,k} - \alpha_{L,k}).$$

In order to provide an upper bound for the variance $\text{Var}_{Q_f^n}(\hat{U}_L)$, let us first state a lemma controlling the variance of a U -statistic of order 2. This result is a particular case of Lemma 8 in [43].

Lemma A.7. *Let h be a symmetric function with 2 inputs, Z_1, \dots, Z_n be independent and identically distributed random vectors and U_n be the U -statistic of order 2 defined by*

$$U_n = \frac{1}{n(n-1)} \sum_{i \neq l=1}^n h(Z_i, Z_l).$$

The following inequality gives an upper bound on the variance of U_n ,

$$\text{Var}(U_n) \leq C \left(\frac{\sigma^2}{n} + \frac{s^2}{n^2} \right),$$

where $\sigma^2 = \text{Var}(\mathbb{E}[h(Z_1, Z_2) | Z_1])$ and $s^2 = \text{Var}(h(Z_1, Z_2))$.

We have that $\mathbb{E}_{Q_f^n}[h_L(Z_1, Z_2) | Z_1] = 0$, hence the first term in the upper bound of the variance vanishes. In order to bound the term s^2 , we write

$$\begin{aligned} h_L(Z_1, Z_2) &= \sum_{k=0}^{L-1} (\varphi_{L,k}(X_1) - \alpha_{L,k}) (\varphi_{L,k}(X_2) - \alpha_{L,k}) + \sigma_L^2 \sum_{k=0}^{L-1} W_{1,L,k} W_{2,L,k} \\ &\quad + \sigma_L \sum_{k=0}^{L-1} W_{1,L,k} (\varphi_{L,k}(X_2) - \alpha_{L,k}) + \sigma_L \sum_{k=0}^{L-1} W_{2,L,k} (\varphi_{L,k}(X_1) - \alpha_{L,k}). \end{aligned}$$

So, since $\mathbb{E}_{Q_f^n}(\varphi_{L,k}(X_i) - \alpha_{L,k}) = 0$ and $\mathbb{E}(W_{i,L,k}) = 0$ for any i . Using independence properties, we therefore have

$$\begin{aligned} &\text{Var}_{Q_f^n}(h_L(Z_1, Z_2)) \\ &= \text{Var}_{Q_f^n} \left[\sum_{k=0}^{L-1} (\varphi_{L,k}(X_1) - \alpha_{L,k}) (\varphi_{L,k}(X_2) - \alpha_{L,k}) \right] + \text{Var}_{Q_f^n} \left[\sigma_L^2 \sum_{k=0}^{L-1} W_{1,L,k} W_{2,L,k} \right] \\ &\quad + 2\text{Var}_{Q_f^n} \left[\sigma_L \sum_{k=0}^{L-1} W_{1,L,k} (\varphi_{L,k}(X_2) - \alpha_{L,k}) \right]. \end{aligned}$$

Now, by independence of X_1 and X_2 ,

$$\begin{aligned} &\text{Var}_{Q_f^n} \left[\sum_{k=0}^{L-1} (\varphi_{L,k}(X_1) - \alpha_{L,k}) (\varphi_{L,k}(X_2) - \alpha_{L,k}) \right] \\ &= \sum_{k,k'=0}^{L-1} \mathbb{E} [(\varphi_{L,k}(X_1) - \alpha_{L,k}) (\varphi_{L,k'}(X_1) - \alpha_{L,k'})] \\ &\quad \mathbb{E} [(\varphi_{L,k}(X_2) - \alpha_{L,k}) (\varphi_{L,k'}(X_2) - \alpha_{L,k'})] \end{aligned}$$

$$= \sum_{k,k'=0}^{L-1} \left[\int \varphi_{L,k} \varphi_{L,k'} f - \alpha_{L,k} \alpha_{L,k'} \right]^2.$$

So

$$\begin{aligned} & \text{Var}_{Q_f^n} \left[\sum_{k=0}^{L-1} (\varphi_{L,k}(X_1) - \alpha_{L,k}) (\varphi_{L,k}(X_2) - \alpha_{L,k}) \right] \\ &= \int \int \left(\sum_{k=0}^{L-1} \varphi_{L,k}(x) \varphi_{L,k}(y) \right)^2 f(x) f(y) dx dy - 2 \int \left(\sum_{k=0}^{L-1} \alpha_{L,k} \varphi_{L,k}(x) \right)^2 f(x) dx \\ & \quad + \left(\sum_{k=0}^{L-1} \alpha_{L,k}^2 \right)^2. \end{aligned}$$

In order to control the first term of the variance, note that by definition of the functions $\varphi_{L,k}$, we have that, for all $x \in [0, 1]$, $\varphi_{L,k}(x) \varphi_{L,k'}(x) = 0$ if $k \neq k'$, and that $\varphi_{L,k}^2 = \sqrt{L} \varphi_{L,k}$. Hence,

$$\begin{aligned} \int \int \left(\sum_{k=0}^{L-1} \varphi_{L,k}(x) \varphi_{L,k}(y) \right)^2 f(x) f(y) dx dy &= L \sum_{k=0}^{L-1} \alpha_{L,k}^2 \\ &\leq L \|f\|_2^2. \end{aligned}$$

Since the second term of the variance is nonpositive, and the third term is controlled by $\|f\|_2^4$, we obtain

$$\text{Var}_{Q_f^n} \left[\sum_{k=0}^{L-1} (\varphi_{L,k}(X_1) - \alpha_{L,k}) (\varphi_{L,k}(X_2) - \alpha_{L,k}) \right] \leq L \|f\|_2^2 + \|f\|_2^4 \leq 2L \|f\|_2^2.$$

By independence of the variables $(W_{i,L,k})$,

$$\text{Var} \left(\sigma_L^2 \sum_{k=0}^{L-1} W_{1,L,k} W_{2,L,k} \right) = \sigma_L^4 \sum_{k=0}^{L-1} \text{Var}(W_{1,L,k} W_{2,L,k}) = L \sigma_L^4.$$

Finally, using again the independence of the variables $(W_{1,L,k})_{k \in \{0, \dots, L-1\}}$, and their independence with X_2 ,

$$\begin{aligned} & \text{Var}_{Q_f^n} \left[\sigma_L \sum_{k=0}^{L-1} W_{1,L,k} (\varphi_{L,k}(X_2) - \alpha_{L,k}) \right] \\ &= \sigma_L^2 \mathbb{E}_{Q_f^n} \left[\sum_{k,k'=0}^{L-1} W_{1,L,k} W_{1,L,k'} (\varphi_{L,k}(X_2) - \alpha_{L,k}) (\varphi_{L,k'}(X_2) - \alpha_{L,k'}) \right] \\ &= \sigma_L^2 \sum_{k=0}^{L-1} \mathbb{E}(W_{1,L,k}^2) \mathbb{E}_{Q_f^n} \left[(\varphi_{L,k}(X_2) - \alpha_{L,k})^2 \right] \end{aligned}$$

$$\leq \sigma_L^2 \sum_{k=0}^{L-1} \int \varphi_{L,k}^2 f \leq \sigma_L^2 L \sum_{k=0}^{L-1} \int_{k/L}^{(k+1)/L} f \leq \sigma_L^2 L$$

since $\int_0^1 f = 1$. This leads to the following upper bound for $\text{Var}_{Q_f^n}(h_L(Z_1, Z_2))$,

$$\text{Var}_{Q_f^n}(h_L(Z_1, Z_2)) \leq (2\|f\|_2^2 + \sigma_L^2 + \sigma_L^4)L,$$

from which, by application of Lemma A.7, we deduce that

$$\text{Var}_{Q_f^n}(\hat{U}_L) \leq 2 \frac{(\|f\|_2^2 + \sigma_L^4)L}{n^2}.$$

Let us now compute $\text{Var}_{Q_f^n}(\hat{V}_L)$. Since \hat{V}_L is centered,

$$\begin{aligned} \text{Var}_{Q_f^n}(\hat{V}_L) &= \mathbb{E}_{Q_f^n}(\hat{V}_L^2) \\ &= \frac{4}{n^2} \sum_{k,k'=0}^{L-1} (\alpha_{L,k} - \alpha_{L,k}^0) (\alpha_{L,k'} - \alpha_{L,k'}^0) \sum_{i,l=1}^n \mathbb{E}_{Q_f^n} [(Z_{i,L,k} - \alpha_{L,k})(Z_{l,L,k'} - \alpha_{L,k'})]. \end{aligned}$$

Note that, if $i \neq l$,

$$\mathbb{E}_{Q_f^n} [(Z_{i,L,k} - \alpha_{L,k})(Z_{l,L,k'} - \alpha_{L,k'})] = 0.$$

Moreover,

$$\begin{aligned} \mathbb{E}_{Q_f^n} [(Z_{i,L,k} - \alpha_{L,k})(Z_{i,L,k'} - \alpha_{L,k'})] &= \mathbb{E}[(\varphi_{L,k}(X_i) - \alpha_{L,k})(\varphi_{L,k'}(X_i) - \alpha_{L,k'}) + \sigma_L^2 \mathbb{E}_{Q_f^n}(W_{i,L,k}W_{i,L,k'})] \\ &= \int \varphi_{L,k}\varphi_{L,k'} f - \alpha_{L,k}\alpha_{L,k'} + 2\sigma_L^2 \mathbf{1}_{k=k'}. \end{aligned}$$

Hence,

$$\begin{aligned} \text{Var}_{Q_f^n}(\hat{V}_L) &= \frac{4}{n} \sum_{k,k'=0}^{L-1} (\alpha_{L,k} - \alpha_{L,k}^0) (\alpha_{L,k'} - \alpha_{L,k'}^0) \left(\int \varphi_{L,k}\varphi_{L,k'} f - \alpha_{L,k}\alpha_{L,k'} + 2\sigma_L^2 \mathbf{1}_{k=k'} \right) \\ &= \frac{4}{n} \int \left(\sum_{k=0}^{L-1} (\alpha_{L,k} - \alpha_{L,k}^0) \varphi_{L,k} \right)^2 f - \frac{4}{n} \left(\sum_{k=0}^{L-1} \alpha_{L,k} (\alpha_{L,k} - \alpha_{L,k}^0) \right)^2 \\ &\quad + \frac{8}{n} \sigma_L^2 \sum_{k=0}^{L-1} (\alpha_{L,k} - \alpha_{L,k}^0)^2 \\ &\leq \frac{4}{n} \sum_{k=0}^{L-1} (\alpha_{L,k} - \alpha_{L,k}^0)^2 \int \varphi_{L,k}^2 f + \frac{8}{n} \sigma_L^2 \sum_{k=0}^{L-1} (\alpha_{L,k} - \alpha_{L,k}^0)^2 \end{aligned}$$

$$\leq \frac{1}{n} \left(4\sqrt{L}\|f\|_2 + 8\sigma_L^2 \right) \sum_{k=0}^{L-1} (\alpha_{L,k} - \alpha_{L,k}^0)^2$$

since by Cauchy Schwarz inequality,

$$0 \leq \int \varphi_{L,k}^2 f = \sqrt{L} \int \varphi_{L,k} f \leq \sqrt{L} \|\varphi_{L,k}\|_2 \|f\|_2 = \sqrt{L} \|f\|_2.$$

We finally obtain,

$$\text{Var}_{Q_f^n}(\hat{V}_L) \leq C \frac{(\sqrt{L}\|f\|_2 + \sigma_L^2)}{n} \|\Pi_{S_L}(f - f_0)\|_2^2.$$

Collecting the upper bounds for $\text{Var}_{Q_f^n}(\hat{U}_L)$ and for $\text{Var}_{Q_f^n}(\hat{V}_L)$, we obtain the inequality from Equation (19), that we remind here:

$$\text{Var}_{Q_f^n}(\hat{T}_L) \leq C \left[\frac{(\sqrt{L}\|f\|_2 + \sigma_L^2)}{n} \|\Pi_{S_L}(f - f_0)\|_2^2 + \frac{(\|f\|_2^2 + \sigma_L^4)L}{n^2} \right].$$

A.3.2. Proof of Corollary 4.4

From Equation (19) and taking $f = f_0$, we obtain

$$\sqrt{\text{Var}_{Q_{f_0}^n}(\hat{T}_L)} / \gamma \leq C(\gamma) \frac{(\|f_0\|_2 + \sigma_L^2)\sqrt{L}}{n}.$$

Moreover, we deduce from (19) that

$$\sqrt{\text{Var}_{Q_f^n}(\hat{T}_L)} / \beta \leq C(\beta) \left[\frac{(L^{1/4}\|f\|_2^{1/2} + \sigma_L)}{\sqrt{n}} \|\Pi_{S_L}(f - f_0)\|_2 + \frac{(\|f\|_2 + \sigma_L^2)\sqrt{L}}{n} \right].$$

Using the inequality between geometric and harmonic means, we get

$$\sqrt{\text{Var}_{Q_f^n}(\hat{T}_L)} / \beta \leq \frac{1}{2} \|\Pi_{S_L}(f - f_0)\|_2^2 + C(\beta) \frac{(\|f\|_2 + \sigma_L^2)\sqrt{L}}{n}.$$

We conclude the proof by using the condition in Equation (17).

A.3.3. Proof of Theorem 4.5

We recall that we have defined

$$f = d \sum_{k=0}^{d-1} p_k \mathbf{1}_{[k/d, (k+1)/d)}, \quad f_0 = d \sum_{k=0}^{d-1} p_k^0 \mathbf{1}_{[k/d, (k+1)/d)}.$$

We obtain from Corollary 4.4 that the second kind error of the test is controlled by β if

$$\frac{3}{2} \|\Pi_{S_L}(f - f_0)\|_2^2 \geq C(\gamma, \beta) \frac{(\|f\|_2 + \|f_0\|_2 + \sigma_L^2)\sqrt{L}}{n}.$$

In the discrete case, by definition, f and f_0 belong to S_L , hence $\|\Pi_{S_L}(f - f_0)\|_2 = \|f - f_0\|_2$ and $\|f\|_2 \leq \|f_0\|_2 + \|f - f_0\|_2$. So we have the following sufficient condition.

$$\|f - f_0\|_2^2 \geq C(\gamma, \beta) \frac{(\|f_0\|_2 + \sigma_L^2 + \sqrt{L}/n)\sqrt{L}}{n}$$

Moreover, we have

$$\|f - f_0\|_2^2 = d \sum_{k=0}^{d-1} (p_k - p_k^0)^2.$$

We recall that $L = d$ and $\sigma_L = 2\sqrt{2d}/\alpha$. That is, the sufficient condition turns out to be

$$d \sum_{k=0}^{d-1} (p_k - p_k^0)^2 \geq C(\gamma, \beta) \frac{d^{1/2}}{n} \left(\|f_0\|_2 + d\alpha^{-2} + d^{1/2}n^{-1} \right).$$

By definition of f_0 , we have that

$$\|f\|_2^2 = d \sum_{k=0}^{d-1} (p_{0,k})^2.$$

Finally, we obtain the following condition

$$\sqrt{\sum_{i=0}^{d-1} (p_i - p_i^0)^2} \geq C(\gamma, \beta) \frac{d^{-1/4}}{n^{1/2}} \left(\left[d \sum_{k=0}^{d-1} (p_k^0)^2 \right]^{1/4} + d^{1/2}\alpha^{-1} + d^{1/4}n^{-1/2} \right).$$

A.3.4. Proof of Theorem 4.9

We obtain from Corollary 4.4 that the second kind error of the test is controlled by β if

$$\|f - f_0\|_2^2 \geq \|f - f_0 - \Pi_{S_L}(f - f_0)\|_2^2 + C(\|f_0\|_2, \|f\|_2, \gamma, \beta) \frac{(\sigma_L^2 + 1)\sqrt{L}}{n}.$$

Since $f - f_0 \in \mathcal{B}_{s,2,\infty}(R)$, setting $L = 2^J$, we have, on one hand

$$\|f - f_0 - \Pi_{S_L}(f - f_0)\|_2^2 \leq R^2 2^{-2Js},$$

and on the other hand, $\|f\|_2 \leq C(s, R, \|f_0\|_2)$. This leads to the sufficient condition

$$\|f - f_0\|_2^2 \geq R^2 2^{-2Js} + C(s, R, \|f_0\|_2, \gamma, \beta) \frac{(\sigma_L^2 + 1)2^{J/2}}{n}.$$

We recall that $\sigma_L = 2\sqrt{2L}/\alpha$. That is, the sufficient condition turns out to be:

$$\|f - f_0\|_2^2 \geq C(s, R, \|f_0\|_2, \gamma, \beta) \left(2^{-2Js} + \frac{2^{3J/2}}{\alpha^2 n} + \frac{2^{J/2}}{n} \right). \quad (37)$$

J^* being set as the smallest integer J such that $2^J \geq (n\alpha^2)^{2/(4s+3)} \wedge n^{2/(4s+1)}$, we consider two cases.

- If $1/\sqrt{n} \leq \alpha \leq n^{1/(4s+1)}$, then $(n\alpha^2)^{2/(4s+3)} \leq n^{2/(4s+1)}$ and the right-hand side of the inequality in Equation (37) for $J = J^*$ is upper bounded by

$$C(s, R, \|f_0\|_2, \gamma, \beta)(n\alpha^2)^{-4s/(4s+3)}.$$

- If $\alpha > n^{1/(4s+1)}$, then $(n\alpha^2)^{2/(4s+3)} > n^{2/(4s+1)}$ and the right-hand side of the inequality in Equation (37) for $J = J^*$ is upper bounded by

$$C(s, R, \|f_0\|_2, \gamma, \beta)n^{-4s/(4s+1)}.$$

Hence, the separation rate of our test over the set $\mathcal{B}_{s,2,\infty}(R)$ is controlled by

$$C(s, R, \|f_0\|_2, \gamma, \beta) \left[(n\alpha^2)^{-2s/(4s+3)} \vee n^{-2s/(4s+1)} \right],$$

which concludes the proof of Theorem 4.9.

A.4. Adaptivity: proof of Theorem 5.2

In this section, f_0 is some fixed density in $\mathbb{L}_2([0, 1])$.

Using the inequality from Equation (29), and the fact that $u_\gamma \geq \gamma/|\mathcal{J}|$, we obtain that

$$\mathbb{P}_{Q_f^n} \left(\Delta_{\gamma, Q}^{\mathcal{J}} = 0 \right) \leq \beta \tag{38}$$

as soon as

$$\exists J \in \mathcal{J}, \mathbb{P}_{Q_f^n} \left(\tilde{T}_J \leq \tilde{t}_J^0(1 - u_\gamma) \right) \leq \beta.$$

We use the result of Corollary 4.4, for $L = 2^J$ for some $J \in \mathcal{J}$, where σ_L is replaced by $\tilde{\sigma}_{2^J}$ and γ is replaced by $\gamma/|\mathcal{J}|$.

Using the fact that $|\mathcal{J}| \leq C \log(n)$, we get that Equation (38) holds as soon as there exists $J \in \mathcal{J}$ such that

$$\|\Pi_{S_{2^J}}(f - f_0)\|^2 \geq C(\|f_0\|_2, \|f\|_2, \beta) \left(\frac{(\tilde{\sigma}_{2^J}^2 + 1)2^{J/2}}{n\sqrt{\gamma/|\mathcal{J}|}} \right),$$

or equivalently

$$\|f - f_0\|^2 \geq \inf_{J \in \mathcal{J}} \left[\|f - f_0 - \Pi_{S_{2^J}}(f - f_0)\|^2 + C(\|f_0\|_2, \|f\|_2, \gamma, \beta) \frac{(\tilde{\sigma}_J^2 + 1)2^{J/2} \sqrt{\log(n)}}{n} \right].$$

Assuming that $f \in \mathcal{B}_{s,2,\infty}(R)$, for some $s > 0$ and $R > 0$, we get that Equation (38) holds if

$$\|f - f_0\|^2 \geq \inf_{J \in \mathcal{J}} \left[R^2 2^{-2Js} + C(\|f_0\|_2, R, \gamma, \beta) \left(2^{J/2} + \frac{2^{3J/2} \log^2(n)}{\alpha^2} \right) \frac{\sqrt{\log(n)}}{n} \right].$$

Choosing $J \in \mathcal{J}$ as the smallest integer in \mathcal{J} such that

$$2^J \geq (n^2 \alpha^4 / \log^5(n))^{1/(4s+3)} \wedge (n^2 / \log(n))^{1/(4s+1)},$$

we obtain the sufficient condition

$$\|f - f_0\|^2 \geq C(\|f_0\|_2, R, \gamma, \beta) \left[(n\alpha^2 / \log^{5/2}(n))^{-4s/(4s+3)} \vee (n / \sqrt{\log(n)})^{-4s/(4s+1)} \right].$$

Hence, for all $s > 0$, $R > 0$, the separation rate of the aggregated test over the set $\mathcal{B}_{s,2,\infty}(R)$ is controlled by

$$C(\|f_0\|_2, R, \gamma, \beta) \left[(n\alpha^2 / \log^{5/2}(n))^{-2s/(4s+3)} \vee (n / \sqrt{\log(n)})^{-2s/(4s+1)} \right],$$

which concludes the proof of Theorem 5.2.

Acknowledgements

B. Laurent and J-M. Loubes recognize the funding by ANITI ANR-19-PI3A-0004.

References

- [1] ACHARYA, J., CANONNE, C. L., FREITAG, C. and TYAGI, H. (2018). Test without trust: Optimal locally private distribution testing. *arXiv preprint arXiv:1808.02174*.
- [2] ACHARYA, J., SUN, Z. and ZHANG, H. (2018). Differentially private testing of identity and closeness of discrete distributions. In *Advances in Neural Information Processing Systems* 6878–6891.
- [3] AGRAWAL, D. and AGGARWAL, C. C. (2001). On the design and quantification of privacy preserving data mining algorithms. In *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* 247–255. ACM.
- [4] AGRAWAL, R. and SRIKANT, R. (2000). Privacy-preserving data mining. In *ACM Sigmod Record* **29** 439–450. ACM.
- [5] AGRAWAL, S. and HARITSA, J. R. (2005). A framework for high-accuracy privacy-preserving mining. In *21st International Conference on Data Engineering (ICDE'05)* 193–204. IEEE.
- [6] ALIAKBARPOUR, M., DIAKONIKOLAS, I., KANE, D. and RUBINFELD, R. (2019). Private testing of distributions via sample permutations. In *Advances in Neural Information Processing Systems* 10878–10889.
- [7] ALIAKBARPOUR, M., DIAKONIKOLAS, I. and RUBINFELD, R. (2018). Differentially private identity and equivalence testing of discrete distributions. In *International Conference on Machine Learning* 169–178.
- [8] BALAKRISHNAN, S. and WASSERMAN, L. (2017). Hypothesis testing for densities and high-dimensional multinomials: Sharp local minimax rates. *arXiv preprint arXiv:1706.10003*.
- [9] BARAUD, Y. (2002). Non-asymptotic minimax rates of testing in signal detection. *Bernoulli* **8** 577–606.
- [10] BERRETT, T. B. and BUTUCEA, C. (2020). Locally private non-asymptotic testing of discrete distributions is faster using interactive mechanisms. *arXiv preprint arXiv:2005.12601*.
- [11] BICKEL, P. J. and RITOV, Y. (1992). Testing for goodness-of-fit : A new approach. In *Nonparametric statistics and related topics- North-Holland, Amsterdam* **23** 51–57.
- [12] BUTUCEA, C., DUBOIS, A., KROLL, M. and SAUMARD, A. (2019). Local differential privacy: Elbow effect in optimal density estimation and adaptation over Besov ellipsoids. *arXiv preprint arXiv:1903.01927*.

- [13] BUTUCEA, C., ROHDE, A. and STEINBERGER, L. (2020). Interactive versus non-interactive locally, differentially private estimation: Two elbows for the quadratic functional. *arXiv preprint arXiv:2003.04773*.
- [14] CAI, B., DASKALAKIS, C. and KAMATH, G. (2017). Priv’it: Private and sample efficient identity testing. *arXiv preprint arXiv:1703.10127*.
- [15] CANONNE, C. L., KAMATH, G., MCMILLAN, A., ULLMAN, J. and ZAKYNTHINO, L. (2019). Private identity testing for high-dimensional distributions. *arXiv preprint arXiv:1905.11947*.
- [16] DINUR, I. and NISSIM, K. (2003). Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* 202–210.
- [17] DONOHO, D. L., JOHNSTONE, I. M., KERKYACHARIAN, G. and PICARD, D. (1996). Density estimation by wavelet thresholding. *The Annals of Statistics* 508–539.
- [18] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Local privacy and statistical minimax rates. In *Foundations of Computer Science (FOCS), 2013 IEEE 54th Annual Symposium on* 429–438. IEEE.
- [19] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2013). Local privacy, data processing inequalities, and statistical minimax rates. *arXiv preprint arXiv:1302.3203*.
- [20] DUCHI, J. C., JORDAN, M. I. and WAINWRIGHT, M. J. (2018). Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association* **113** 182–201.
- [21] DUCHI, J. C., WAINWRIGHT, M. J. and JORDAN, M. I. (2013). Local privacy and minimax bounds: Sharp rates for probability estimation. In *Advances in Neural Information Processing Systems* 1529–1537.
- [22] DUNCAN, G. T. and LAMBERT, D. (1986). Disclosure-limited data dissemination. *Journal of the American statistical association* **81** 10–18.
- [23] DUNCAN, G. T. and LAMBERT, D. (1989). The risk of disclosure for microdata. *Journal of Business & Economic Statistics* **7** 207–217.
- [24] DWORK, C., KENTHAPADI, K., MCSHERRY, F., MIRONOV, I. and NAOR, M. (2006). Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques* 486–503. Springer.
- [25] DWORK, C., MCSHERRY, F., NISSIM, K. and SMITH, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference* 265–284. Springer.
- [26] EVFIMIEVSKI, A., GEHRKE, J. and SRIKANT, R. (2003). Limiting privacy breaches in privacy preserving data mining. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* 211–222. ACM.
- [27] FIENBERG, S. E. and STEELE, R. J. (1998). Disclosure limitation using perturbation and related methods for categorical data. *Journal of Official Statistics* **14** 485.
- [28] FROMONT, M. and LAURENT, B. (2006). Adaptive goodness-of-fit tests in a density model. *The Annals of Statistics* **34** 680–720.
- [29] GABOARDI, M., LIM, H.-W., ROGERS, R. M. and VADHAN, S. P. (2016). Differentially private chi-squared hypothesis testing: Goodness of fit and independence testing. In *ICML’16 Proceedings of the 33rd International Conference on International Conference on Machine Learning-Volume 48*. JMLR.
- [30] GABOARDI, M. and ROGERS, R. (2017). Local private hypothesis testing: Chi-square tests. *arXiv preprint arXiv:1709.07155*.
- [31] GINÉ, E. and NICKL, R. (2016). *Mathematical foundations of infinite-dimensional statistical models* **40**. Cambridge University Press.
- [32] INGLOT T., L. T. (1996). Asymptotic optimality of data-driven Neyman’s tests for uniformity. *The Annals of Statistics* **24** 1982–2019.

- [33] INGSTER, Y. I. (1987). Minimax testing of nonparametric hypotheses on a distribution density in the L_p metrics. *Theory of Probability & Its Applications* **31** 333–337.
- [34] INGSTER, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Math. Methods Statist* **2** 85–114.
- [35] INGSTER, Y. I. (2000). Adaptive chi-square tests. *Journal of Mathematical Sciences* **99** 1110–1119.
- [36] JANK, W., SHMUELI, G. and WANG, S. (2008). *Statistical methods in e-commerce research*. Wiley Online Library.
- [37] JOSEPH, M., MAO, J., NEEL, S. and ROTH, A. (2019). The role of interactivity in local differential privacy. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)* 94–105. IEEE.
- [38] KALLENBERG W., L. T. (1995). Consistency and Monte Carlo simulation of a data driven version of smooth goodness-of-fit tests. *The Annals of Statistics* **23** 1594–1608.
- [39] KASIVISWANATHAN, S. P., LEE, H. K., NISSIM, K., RASKHODNIKOVA, S. and SMITH, A. (2011). What can we learn privately? *SIAM Journal on Computing* **40** 793–826.
- [40] LAM-WEIL, J., LAURENT, B. and LOUBES, J.-M. (2020). Minimax optimal goodness-of-fit testing for densities under a local differential privacy constraint. *arXiv preprint arXiv:2002.04254*.
- [41] LE CAM, L. M. (1986). Asymptotic methods in statistical theory.
- [42] LEDWINA, T. (1994). Data-driven version of Neyman’s smooth test of fit. *Journal of the American Statistical Association* **89** 1000–1005.
- [43] MEYNAOUI, A., ALBERT, M., LAURENT, B. and MARREL, A. (2019). Adaptive test of independence based on HSIC measures. *hal-02020084*.
- [44] MISHRA, N. and SANDLER, M. (2006). Privacy via pseudorandom sketches. In *Proceedings of the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* 143–152. ACM.
- [45] NEYMAN, J. (1937). Smooth test for goodness of fit. *Scandinavian Actuarial Journal* **1937** 149–199.
- [46] PANINSKI, L. (2008). A coincidence-based test for uniformity given very sparsely sampled discrete data. *IEEE Transactions on Information Theory* **54** 4750–4755.
- [47] SERFLING, R. J. (2009). *Approximation theorems of mathematical statistics* **162**. John Wiley & Sons.
- [48] SHEFFET, O. (2018). Locally private hypothesis testing. *arXiv preprint arXiv:1802.03441*.
- [49] TSYBAKOV, A. B. (2004). Introduction to nonparametric estimation, 2009. URL <https://doi.org/10.1007/b13794>. Revised and extended from the.
- [50] VALIANT, G. and VALIANT, P. (2017). An automatic inequality prover and instance optimal identity testing. *SIAM Journal on Computing* **46** 429–455.
- [51] VAN DEN HOUT, A. and VAN DER HEIJDEN, P. G. (2002). Randomized response, statistical disclosure control and misclassification: a review. *International Statistical Review* **70** 269–288.
- [52] WARNER, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* **60** 63–69.
- [53] YANG, Y. and BARRON, A. (1999). Information-theoretic determination of minimax rates of convergence. *Annals of Statistics* 1564–1599.
- [54] YU, B. (1997). Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam* 423–435. Springer.