Private Quantiles Estimation in the Presence of Atoms

Clément Sébastien Lalanne, Clément Gastaud, Nicolas Grislain, Aurélien Garivier, Rémi Gribonval

Private Quantiles Estimation in the Presence of Atoms

Clément Lalanne[1], Clément Gastaud[2], Nicolas Grislain[2], Aurélien Garivier[1,3], and Rémi Gribonval[1,4]

[1]LIP Laboratory, École Normale Supérieure de Lyon, Lyon, France
[2]Sarus Technologies, Paris, France
[3]UMPA Laboratory, École Normale Supérieure de Lyon, Lyon, France
[4]INRIA, France

February 14, 2022

Abstract

We address the differentially private estimation of multiple quantiles (MQ) of a dataset, a key building block in modern data analysis. We apply the recent non-smoothed Inverse Sensitivity (IS) mechanism to this specific problem and establish that the resulting method is closely related to the current state-of-the-art, the JointExp algorithm, sharing in particular the same computational complexity and a similar efficiency. However, we demonstrate both theoretically and empirically that (non-smoothed) JointExp suffers from an important lack of performance in the case of peaked distributions, with a potentially catastrophic impact in the presence of atoms. While its smoothed version would allow to leverage the performance guarantees of IS, it remains an open challenge to implement. As a proxy to fix the problem we propose a simple and numerically efficient method called Heuristically Smoothed JointExp (HSJointExp), which is endowed with performance guarantees for a broad class of distributions and achieves results that are orders of magnitude better on problematic datasets.

Introduction

As more and more data is collected on individuals and as data science techniques become more powerful, threats to privacy have multiplied and serious concerns have emerged [NS06, BDK07, FJR15, DN03, HSR$^+$08, LDM10, NS08, Swe00, WE18, Swe02]. Against this background, *differential privacy* [DR$^+$14] has become the *gold standard* in privacy protection. By introducing randomness at a level calibrated to the *sensitivity* of a query [DMNS06], it enables the inference of global statistics on a dataset while bounding each sample's influence and ensuring that the presence or absence of an individual in the dataset cannot be deduced from the result. In the last decade, research results have brought nice building blocks and composition theorems [DMNS06, KOV15, DRS19, DDR20, ACG$^+$16]. They paved the way for many applications in data analysis, from basic statistics to advanced *artificial intelligence* algorithms. Notably, differential privacy is now used in production by the US Census Bureau [Abo18], Google [EPK14], Apple [TVV$^+$17] and Microsoft [DKY17] among others.

In this paper, we focus on the problem of estimating one or many *quantiles* with privacy guarantees. Beyond the interest that quantiles have in themselves, they are also important primitives in many advanced applications in machine learning from synthetic data generation to decision tree training. Indeed, since quantiles are a reasonable choice of bins for quantizing a cumulative distribution function, they are commonly used in algorithms based on *decision trees*, such as *Random Forests* and *Boosted Trees* [CG16], where variables are binned using quantiles before they are considered for a split. Besides, in most recent synthetic data models, continuous features are binned to reduce the output space's dimension (see [ZCP$^+$17, MMS21]).

Several techniques have been proposed to solve the problem of private quantiles estimation. A naïve one is to add Laplace noise to non-private quantile estimates [DMNS06]. It is straightforward and easy to compute

but the amount of added noise is based on a pessimistic scenario that could not materialize simultaneously for all quantiles. To reduce the variance of the estimates, a variant that uses smoothed sensitivity instead of the worst case scenario was introduced [NRS07]. It has the drawback, however, of using approximate differential privacy [DKM$^+$06] instead of pure differential privacy which allows some catastrophic failures with small probability. The current state of the art for single quantile estimation uses a fine tuned version of the exponential mechanism that is called *ExponentialQuantile* [Smi11, MT07]. It is cheap to compute and the recent theory of *Inverse Sensitivity* [AD20b] shows that a conceptually simple smoothing operation allows levels of privacy that are quadratically better than the smoothed sensitivity approaches while sticking to pure differential privacy, achieving near-instance optimality. As a result, it has been adopted by the main DP-ready libraries [All, IBM] used in production. All the algorithms previously mentioned are designed to estimate a single quantile from a dataset. Composition theorems make them usable for multiquantile estimation by evaluating each quantile independently but with an increased privacy noise. Recent work by Gillenwater, Joseph and Kulesza [GJK21] has presented a new algorithm *JointExp* that is the first to exploit the non-decreasing constraint of quantiles at the core of its sampling procedure and has since become the new state of the art. Our first main contribution is to formally show that JointExp is indeed deeply related to the general principle of the Inverse Sensivity approach.

The accuracy of JointExp and ExponentialQuantiles has been empirically demonstrated on multiple real-world datasets. Our second main contribution is to show that for highly concentrated datasets, the utility of these mechanisms in fact decreases drastically. To fix this, our third contribution is to propose a new heuristic tractable smoothing technique for those algorithms that vastly improves their utility on problematic datasets without noticeable repercussions on non-degenerate distributions. Our technique differs from the general smoothing trick for *inverse sensitivity* based mechanisms introduced by Asi and Duchi [AD20b, AD20a] in the sense that it smoothes the data distribution rather than taking a maximum convolution of the density over the output space resulting in a much better computational complexity and making it a viable solution in high dimension.

The paper is organized as follows. In Section 1 we recall the needed background on differential privacy, inverse sensitivity, and (private) multiquantile estimation. Section 2 is devoted to characterizing the inverse sensitivity mechanism for private multiquantile estimation and showing its close connections with JointExp. The failure of JointExp on atomic distributions, and its analysis, are discussed in Section 3. The proposed algorithm HSJointExp is introduced in Section 4, where we also establish its privacy and consistency guarantees. Finally, Section 5 gathers the numerical experiments demonstrating the behavior of the proposed algorithm and its comparison with the state of the art.

1 Background

We consider a dataset $\mathbf{X} = (X_1, \dots, X_n) \in \mathfrak{X}^n$, where $\mathfrak{X}$ denotes our feature space and $n \geq 1$ is the sample size. The *Hamming distance* $d(\mathbf{X}, \mathbf{Y})$ between two datasets $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}^n$ is defined as the minimal number of changes (i.e., substitutions of entries) required to transform $\mathbf{X}$ into a permutation of $\mathbf{Y}$. Hence, $d(\mathbf{X}, \mathbf{Y}) = 0$ iff there exists a permutation σ on $\{1, \dots, n\}$ such that $\forall i \in \{1, \dots, n\}, X_i = Y_{\sigma(i)}$. We say that $\mathbf{X}$ and $\mathbf{Y}$ are *neighbors* (noted $\mathbf{X} \sim \mathbf{Y}$) if $d(\mathbf{X}, \mathbf{Y}) = 1$, that is if there exist two permutations σ_1, σ_2 such that $\forall i \in \{1, \dots, n-1\}, X_{\sigma_1(i)} = Y_{\sigma_2(i)}$. Note that $d(\mathbf{X}, \mathbf{Y})$ is the minimum length of a path on consecutive neighbors linking $\mathbf{X}$ to $\mathbf{Y}$.

1.1 Differential Privacy

Given a privacy budget $\epsilon > 0$, a randomized algorithm $\mathcal{A} : \mathfrak{X}^n \to O$ is called ϵ-*differentially private* (ϵ-DP, see [DKM$^+$06]) if for all pairs of datasets $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}^n$ and all measurable sets $S \subseteq O$,

$$\mathbf{X} \sim \mathbf{Y} \Rightarrow \mathbb{P}\big(\mathcal{A}(\mathbf{X}) \in S\big) \leq e^{\epsilon} \times \mathbb{P}\big(\mathcal{A}(\mathbf{Y}) \in S\big) .$$

A classical way to design ϵ-DP algorithms is the Exponential Mechanism [MT07]. For a utility function $u : \mathfrak{X}^n \times O \to \mathbb{R}$ that measures the relevance $u(\mathbf{X}, o)$ of the output o for the dataset $\mathbf{X}$, the Exponential Mechanism $\mathcal{E}_u^{(\alpha)}$ defined by u and with parameter $\alpha > 0$ outputs a random variable on O with a density

proportional to $e^{u(\mathbf{X},o)/\alpha}$ with respect to some reference measure on O. For example, when O is discrete,

$$\mathbb{P}\left(\mathcal{E}_u^{(\alpha)}(\mathbf{X}) = o\right) = \frac{e^{u(\mathbf{X},o)/\alpha}}{\sum_{o' \in O} e^{u(\mathbf{X},o')/\alpha}} \quad \forall o \in O \, .$$

Defining the *sensitivity* of the utility function

$$\Delta u := \sup_{o \in O, \mathbf{X}, \mathbf{Y} \in \mathfrak{X}^n : \mathbf{X} \sim \mathbf{Y}} |u(\mathbf{X}, o) - u(\mathbf{Y}, o)| \, ,$$

$\mathcal{E}_u^{(\alpha)}$ is ϵ-DP as soon as $\alpha \geq \frac{2\Delta u}{\epsilon}$ [MT07].

1.2 The inverse sensitivity mechanism

When considering some deterministic function $\mathcal{Q} : \mathfrak{X}^n \to O$ as the target of the estimation, a specific choice of utility function in the exponential mechanism, long known as folklore, was proved to have remarkable optimality properties under certain assumptions [AD20b]. The *inverse sensitivity function*

$$u_{\mathrm{IS}}(\mathbf{X}, o) := -\inf \left\{ d(\mathbf{X}, \mathbf{Y}) \text{ s.t. } \mathbf{Y} \in \mathcal{Q}^{-1}(o) \right\}$$

is easily seen to have sensitivity $\Delta u_{\mathrm{IS}} = 1$. The resulting ϵ-DP mechanism $\mathcal{E}_{u_{\mathrm{IS}}}^{(2/\epsilon)}(\mathbf{X})$ has then a behavior that is quite intuitive: the likelihood of its output o decreases when $\mathcal{Q}^{-1}(o)$ becomes further apart from $\mathbf{X}$ in the Hamming distance.

1.3 Private Multiquantile Estimation

In order to fix the notations for the multiquantile estimation, given $a < b \in \mathbb{R}$, let $\mathfrak{X} = [a, b]$ be the feature space and let $O = [a, b]^{m\nearrow}$ be the set of vectors of m increasing points in $[a, b]$ representing m quantiles. The hypothesis that the feature space is bounded is necessary to the analysis and is reasonable for many applications. For applications where this is unrealistic, solutions have been proposed at the expense of having an algorithm that has a small chance of halting [DL09, BAM20]. Given a probability vector $\mathbf{p} = (p_1, \ldots, p_m) \in (0, 1)^{m\nearrow}$, the goal is to estimate the empirical quantile function associated to $\mathbf{p}$ and defined as

$$\begin{array}{rccc} \mathcal{Q} & : & \mathfrak{X}^n & \to & O \\ & & \mathbf{X} & \mapsto & (X_{(\lceil np_1 \rceil)}, \ldots, X_{(\lceil np_m \rceil)}) \end{array}$$

where $X_{(i)}$ denotes the i-th order statistics of $X_1, \ldots, X_n$. As a safety check, we assume that $\forall j \in \{1, \ldots, m-1\}, n(p_{j+1} - p_j) \geq 1$ to ensure that no data point will be chosen twice as a quantile representant. Note that given $\mathbf{p}$ this condition can always be satisfied provided that we have enough data, i.e., that n is large enough. For any $\mathbf{X} \in \mathfrak{X}^n$, $\mathbf{q} \in O$ and $\mathbf{p} \in (0, 1)^m$, we use the convention $X_{i \leq 0} = q_{i \leq 0} = a$, $X_{i \geq n+1} = q_{i \geq m+1} = b$, $p_{i \leq 0} = 0$ and $p_{i \geq m+1} = 1$. Finally, for brevity of notations, vectors are interpreted when needed as the set containing their components.

The definition of JointExp, the state-of-the-art mechanism for multiquantile private estimation (also called ExponentialQuantile when $m = 1$) corresponds [GJK21] to the exponential mechanism $\mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\mathbf{X})$ with

$$-u_{\mathrm{JE}}(\mathbf{X}, \mathbf{q}) := \frac{1}{2} \sum_{i=1}^{m+1} \left| \delta^{\mathrm{JE}}(i, \mathbf{X}, \mathbf{q}) \right| \, ,$$

(which is of sensitivity 1) where

$$\delta^{\mathrm{JE}}(i, \mathbf{X}, \mathbf{q}) := n(p_i - p_{i-1}) - \#\left(\mathbf{X} \cap (q_{i-1}, q_i]\right) \, .$$

This mechanism works by penalizing the result whenever the number of data points in each quantile interval ($\#\left(\mathbf{X} \cap (q_{i-1}, q_i]\right)$) deviates from what should be expected ($n(p_i - p_{i-1})$).

2 JointExp meets Inverse Sensitivity

To the best of our knowledge, the inverse sensitivity mechanism for the private estimation of multiple quantiles has never been studied yet. We do so in the rest of this section.

Deriving the expression of the inverse sensitivity for a dataset $\mathbf{X}$ and an output candidate $\mathbf{q}$ boils down to answering the question: What is the minimal number of points from $\mathbf{X}$ that need to be changed in order to obtain a vector that has $\mathbf{q}$ as its empirical quantiles? Theorem 1 solves this question for Lebesgue-almost-any $\mathbf{q}$.

Theorem 1. *For any $\mathbf{X} \in \mathfrak{X}^n$ and $\mathbf{q} \in ([a,b] \setminus \mathbf{X})^{m\nearrow}$ without collision,*

$$-u_{IS}(\mathbf{X},\mathbf{q}) = \frac{1}{2}\sum_{i=2}^{m+1} |\delta(i,\mathbf{X},\mathbf{q})| + \sum_{i=2}^{m} \mathbb{1}_{\mathbb{R}_+}\left(\delta(i,\mathbf{X},\mathbf{q})\right) + \frac{1}{2}\left|\delta_{\texttt{closed}}(1,\mathbf{X},\mathbf{q})\right| + \mathbb{1}_{\mathbb{R}_+}\left(\delta_{\texttt{closed}}(1,\mathbf{X},\mathbf{q})\right)$$

with

$$\delta(i,\mathbf{X},\mathbf{q}) = \#\left(\mathbf{X} \cap (q_{i-1}, q_i]\right) - \left(\lceil np_i \rceil - \lceil np_{i-1} \rceil\right)$$
$$\delta_{\texttt{closed}}(i,\mathbf{X},\mathbf{q}) = \#\left(\mathbf{X} \cap [q_{i-1}, q_i]\right) - \left(\lceil np_i \rceil - \lceil np_{i-1} \rceil\right).$$

We postpone the proof to subsection D.1 for brevity. The cases where $\mathbf{q}$ has collisions or shares some common points with the dataset are more difficult. Luckily, those cases can be neglected when considering the sampling mechanism. Indeed, $\mathcal{E}^{(2/\epsilon)}_{u_{\mathrm{IS}}}(\mathbf{X})$ has a density that is absolutely continuous w.r.t Lebesgue measure, the expression of the resulting mechanism can be further simplified (see Corollary 1) by modifying the density on outcomes of null Lebesgue measure.

Corollary 1. *For any $\mathbf{X} \in \mathfrak{X}^n$, $\mathcal{E}^{(2/\epsilon)}_{u_{IS}}(\mathbf{X})$ has the same output distribution as $\mathcal{E}^{(2/\epsilon)}_{\tilde{u}_{IS}}(\mathbf{X})$ where $\forall \mathbf{X} \in \mathfrak{X}^n$, $\forall \mathbf{q} \in O$,*

$$-\tilde{u}_{IS}(\mathbf{X},\mathbf{q}) = \frac{1}{2}\sum_{i=1}^{m+1} |\delta(i,\mathbf{X},\mathbf{q})| + \sum_{i=1}^{m} \mathbb{1}_{\mathbb{R}_+}(\delta(i,\mathbf{X},\mathbf{q}))\,.$$

Remark 1. *We discuss the sampling from $\mathcal{E}^{(2/\epsilon)}_{\tilde{u}_{IS}}(\mathbf{X})$ in Appendix A but our conclusion is that it can be done by making some minor adjustments to the JointExp sampling algorithm and that in particular, the two algorithms share the same complexity of $O(nm\log n + nm^2)$.*

Remark 2. *One can check that $|\tilde{u}_{IS}(\mathbf{X},\mathbf{q}) - u_{IS}(\mathbf{X},\mathbf{q})| \leq 2(m+1)$ and thus the distributions differ significantly only on outcomes of high utility (when the number of misclassified points is of the order $O(m)$). The bad outcomes are almost equally penalized and for this reason, we can expect the two algorithms to perform almost identically when n is large enough. This is indeed confirmed by numerical examples as illustrated in Section 5. As a consequence, we will mainly focus on JointExp for the rest of this article.*

3 JointExp fails on atomic distributions

To the best of our knowledge, no theoretical utility guarantee for JointExp has been derived yet, and the performance of this algorithm has only been demonstrated experimentally. Even if it outperforms by multiple orders of magnitude previous techniques on many real life datasets [GJK21], we prove in this section that it can also completely fail on some distributions (see Proposition 1). As illustrated in Section 5, JointExp is indeed observed to be suboptimal on several real world datasets associated to peaked distributions, such as the US Census Bureau "Dividends" and "Earnings" data.

In order to understand the origin of this weakness of JointExp, we analyse the density of the distribution of its output. This density is constant on the *"blocks"* $\left([X_{i_1}, X_{i_1+1}) \times \ldots \times \cdots \times [X_{i_m}, X_{i_m+1})\right) \cap [a,b]^{m\nearrow}$ for each $\mathbf{i} = (i_1,\ldots,i_m) \in O'$ where

$$O' = \left\{\mathbf{i} \in \{0,\ldots,n\}^m, i_1 \leq \cdots \leq i_m\right\}\,.$$

The probability of the output of JointExp being in a given block is proportional to the volume of this block. What can happen in practice is that even though a block is interesting in terms of utility level, its volume can in fact be close to zero if the data points are close. The volume can even be zero in case of equality, hence this block is never selected by the exponential mechanism. This phenomenon occurs particularly often for data drawn from distributions with isolated atoms: asymptotically, the dataset will almost surely contain collisions among the data points as n grows and JointExp will fail on the corresponding quantiles.

To formally capture this phenomenon, from now on $\mathbf{X}$ is supposed to be a collection of n i.i.d. samples of a random variable X with distribution $\mathbb{P}_X$ and with cumulative distribution function (CDF) F_X.

Proposition 1. *Suppose that there exist $q \in (a, b)$ and $\eta > 0$ such that $I := (q - \eta, q + \eta) \subset [a, b]$ satisfies $\mathbb{P}_X(\{q\}) > 0$ and $\mathbb{P}_X(I \setminus \{q\}) = 0$. Then there exist some probability vectors $\mathbf{p}$ such that*

$$\mathbb{E}_{\mathbf{X}, \mathcal{E}_{u_{JE}}^{(2/\epsilon)}} \left(||\mathcal{E}_{u_{JE}}^{(2/\epsilon)}(\mathbf{X}) - F_X^{-1}(\mathbf{p})||_2^2 \right) = \Omega_n(1) \ , \tag{1}$$

where we use the vector notation $F_X^{-1}(\mathbf{p}) = (F_X^{-1}(p_1), \dots, F_X^{-1}(p_n))$. Furthermore, the Lebesgue measure of those problematic probability vectors is lower bounded by $\mathbb{P}_X(\{q\})^m$.

$\Omega_n(1)$ refers to a quantity that is lower bounded by a positive constant when n varies. We postpone the proof to subsection D.2 for brevity.

This result shows that for certain data distributions with isolated atoms, JointExp is not consistent, even asymptotically. This behavior is all the more counter intuitive as one would think that on datasets with a lot of collisions, very little noise would be needed to ensure privacy since the points are already indistinguishable.

Example 1. *Consider the private estimation of the median (i.e. $m = 1$ quantile, and $\mathbf{p} = (1/2)$) on $[a, b] = [-1, 1]$. Since $m = 1$, JointExp coincides with ExponentialQuantile, and when all data points are equal to 0 (i.e. $\mathbb{P}_X = \delta_0$) its output is uniformly distributed in $[-1, 1]$* whatever the sample size n *as long as it is even.*

It is worth noting that defavorable cases are not simply irrealistic corner cases. Indeed, many real-life datasets show *accumulation points* and are continuous distributions with some Diracs at specific points. A famous example is the revenue statistics of the US Census Bureau: many participants in surveys are not qualified to have some categorie of revenue (too young or not investing in some assets) hence the presence of accumulations at the zero value for these categories. In fact, any continuous variable that is censored, conditional on some other variable or generated by mimetic agents tending to repeat exactly some values, will show accumulation points where JointExp has great chances to fail.

4 Heuristic smoothing, with guarantees

The type of failure of JointExp highlighted in Section 3 may seem surprising given a) the strong connection between JointExp and the Inverse Sensitivity established in Section 2; and b) existing performance guarantees for *smoothed* Inverse Sensitivity mechanisms [AD20b, AD20a]. Indeed, while JointExp is not smoothed, smoothing convolves the output distribution with a max kernel, increasing the volume of the maximum of the distribution to circumvent the difficulties raised by isolated atoms. We discuss such an approach in Appendix B and conclude that while it would increase the utility of the resulting mechanism, it would also make it computationally intractable.

As a tractable alternative, we present in this section a heuristic algorithm based on noise addition prior to the application of JointExp, and we show that this mechanism is endowed with privacy and consistency guarantees.

4.1 Introducing the HSJointExp algorithm

Since JointExp has a density that is constant on the blocks $\big([X_{i_1}, X_{i_1+1}) \times \ldots \times \cdots \times [X_{i_m}, X_{i_m+1})\big) \cap [a, b]^{m\nearrow}$ for $\mathbf{i} = (i_1, \dots, i_m) \in O'$, it fails when the blocks that have a great utility have a too small volume. By adding noise to the data points, we ensure a minimal volume for the volume of the interesting regions while only shifting the quantiles of the distribution by a small amount.

Let $w_1, \dots, w_n$ be i.i.d variables, and let

$$\tilde{\mathbf{X}} = (X_1 + w_1, \dots, X_n + w_n) \, . \tag{2}$$

The Heuristically Smoothed JointExp (HSJointExp) algorithm returns $\mathcal{E}^{(2/\epsilon)}_{u_{\mathrm{JE}}}(\tilde{\mathbf{X}})$, the output of the JointExp on $\tilde{\mathbf{X}}$.

Let us now discuss the choice of the distribution $\mathbb{P}_w$ of the $(w_i)'s$. Discrete noise distributions (for instance Bernoulli noise scaled by some $\alpha > 0$: $\frac{w}{\alpha} \sim \mathcal{B}(\frac{1}{2})$) may seem interesting because they lead to easily tuneable data gaps. However, it often just creates new instances where JointExp fails. Indeed, adding discrete noise to data distributions with accumulation points creates new accumulation points.

For this reason, we focus in the sequel on continuous noise distributions with a density denoted by π_w. The density $\pi_{\tilde{X}}$ of the noisy data $\tilde{X}$ is hence given by the convolution formula

$$\forall t \in \mathbb{R}, \quad \pi_{\tilde{X}}(t) = \int \pi_w(t - x) \mathbb{P}_X(dx) \, . \tag{3}$$

A typical choice of noise discussed in the sequel is the uniform distribution on the interval $[-\alpha, \alpha]$.

Before discussing the choice of the scale parameter α, we remark that HSJointExp consists of the addition of i.i.d. noise prior to running JointExp. Its privacy guarantees are thus a direct consequence of the following generic composition lemma. Its proof, which we did not find elsewhere, is in the supplementary material (see subsection D.3).

Proposition 2. *Let* $\mathbf{w}$ *be a random variable on* $\mathbb{R}^n$ *with probability distribution* $\mathbb{P}_{\mathbf{w}}$ *that is invariant by permutations of the components of the vector. If* $\mathcal{A}$ *is* ϵ*-DP on* $\mathfrak{X}^n$*, then* $\mathbf{X} \mapsto \mathcal{A}(\texttt{proj}_{\mathfrak{X}^n}(\mathbf{X} + \mathbf{w}))$ *is also* ϵ*-DP.*

The projection step `proj` onto the data space $\mathfrak{X}^n$ is necessary because JointExp needs to know the range of the data. Note that $\mathfrak{X}^n$ could be replaced by any set of the form $[a - \delta_{\alpha,n}, b + \delta_{\alpha,n}]^n$ where $\delta_{\alpha,n}$ is a quantity that depends on α and n. So for instance, if the the noise follows a uniform distribution on the interval $[-\alpha, \alpha]$, projecting on $[a - \alpha, b + \alpha]^n$ (does nothing) and then running JointExp on $[a - \alpha, b + \alpha]$ ensures that no point will overflow.

4.2 Consistency of HSJointExp on constant data

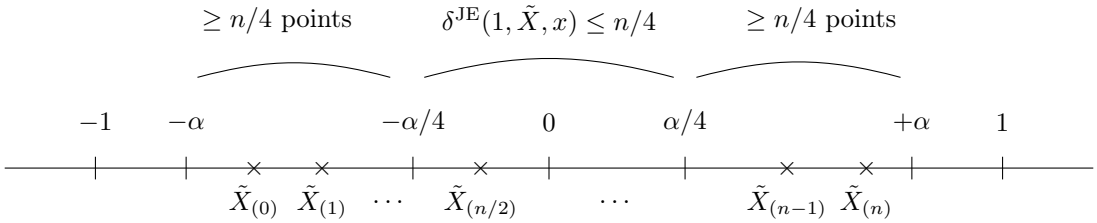


Figure 1: $\delta^{\mathrm{JE}}(1, \tilde{X}, x)$ is bounded by $n/4$ for $-\alpha/4 \leq x \leq \alpha/4$ on the event G.

In order to give some insight on the general analysis of HSJointExp, and to explain the choice that we suggest for the amplitude α of the noise, we start by discussing the simple setting of Example 1 where $X_i \equiv 0$ and JointExp is known to fail. We consider uniform noise with distribution $d\mathbb{P}_w(w) = \frac{\mathbb{1}_{[-\alpha,\alpha]}(w)}{2\alpha} dw$, and HSJointExp returns the output of ExponentialQuantile/JointExp with $m = 1$ on the noisy data $\tilde{X}$:

$$M := \mathcal{E}^{(2/\epsilon)}_{u_{\mathrm{JE}}}(\tilde{\mathbf{X}}).$$

The true median of the dataset is 0, and we study the quadratic risk $\mathbb{E}(M^2)$ of our mechanism. Denoting by $N(x, y) = \sum_{i=1}^n \mathbb{1}_{[x,y)}(0 + w_i)$ the number of noisy points falling in the interval $[x, y)$, we define the event

$$G := \big\{ N(-\alpha, -\alpha/4) \geq n/4 \big\} \cap \big\{ N(\alpha/4, \alpha) \geq n/4 \big\} \, .$$

Since $N(-\alpha,-\alpha/4) \overset{\mathcal{L}}{=} N(\alpha/4,\alpha) \sim \mathcal{B}(n,3/8)$, by Hoeffding's inequality the probability of G is a least $1-2\exp(-n/32)$. Moreover, on the event G, for every $x \in [-\alpha/4,\alpha/4]$ one has $N(-\alpha,x) \geq n/4$ and $N(x,\alpha) \geq n/4$; hence, the minimal number of sample points that need to be changed so as to reach a median equal to x is at most $\delta^{\mathrm{JE}}(1,\tilde{X},x) = \big|n/2 - N(-1,x)\big| \leq n/4$ (see Figure 1), and $u_{\mathrm{JE}}(\tilde{X},x) \leq n/8$. On the other hand, for every $x \notin [-\alpha,\alpha]$, $\delta^{\mathrm{JE}}(1,\tilde{X},x) = n/2$ and $u_{\mathrm{JE}}(\tilde{X},x) = n/4$. Since the density of M at $x \in [-1,1]$ is equal to $\exp\big(-u_{\mathrm{JE}}(\tilde{X},x)\epsilon/2\big)/\int_{-1}^{1}\exp\big(-u_{\mathrm{JE}}(\tilde{X},t)\epsilon/2\big)dt$,

$$\begin{aligned}\mathbb{P}\big(|M|>\alpha\big|G\big) &\leq \frac{\mathbb{P}\big(|M|>\alpha\big|G\big)}{\mathbb{P}\big(|M|\leq\alpha/4\big|G\big)} \\ &\leq \frac{2\times e^{-n\epsilon/8}}{\alpha/2\times e^{-n\epsilon/16}} = \frac{4e^{-n\epsilon/16}}{\alpha}\,.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}\left(M^2\right) &\leq 1^2\left(\mathbb{P}(\bar{G})+\mathbb{P}\big(|M|>\alpha\big|G\big)\right)+\alpha^2\,\mathbb{P}\big(|M|\leq\alpha\big|G\big) \\ &\leq e^{-n/32}+\frac{4e^{-n\epsilon/16}}{\alpha}+\alpha^2\,.\end{aligned}$$

Choosing $\alpha = e^{-n\epsilon/48}$ yields

$$\mathbb{E}\left(M^2\right) \leq 5e^{-n\epsilon/24}+e^{-n/32}\,.$$

We conclude that, contrary to JointExp, HSJointExp is here consistent as soon as $n\epsilon \to \infty$, which is anyway a necessary condition. Besides, the analysis provides a simple and generic way to tune the noise amplitude α as a function of n and ϵ.

4.3 General Consistency of HSJointExp

The consistency of HSJointExp is established by making modifications to the density (via the noise) so that we fall into the favorable cases of JointExp. The ideal cases where JointExp gives an estimator that converges in probability are given by the following theorem.

Theorem 2. *If X is a random variable with density π_X w.r.t. Lebesgue measure that is piecewise continuous and if there exists $\beta > 0$ such that $\pi_X > 0$ and is continuous on $\cup_{i=1}^{n}[F_{\tilde{X}}^{-1}(p_i)-\beta, F_{\tilde{X}}^{-1}(p_i)+\beta]$, then*

$$\mathbb{P}\left(||\mathcal{E}_{u_{JE}}^{(2/\epsilon)}(\mathbf{X}) - F_X^{-1}(\mathbf{p})||_\infty > \beta\right) = o_n(1)\,.$$

The proof that uses similar techniques as in subsection 4.2 is in subsection D.5. The reader can find an expression of the upper bound $o_n(1)$ that does not hide any problem parameter in the proof. This theorem states that for data distributions with continuous densities, JointExp is consistent and this is to the best of our knowledge the first general result stating the consistency of JointExp.

Back to our problem, the data distribution is not so regular. In particular, we are interested in the case where it contains atoms. Following the method that we propose for HSJointExp, we add some independent and identically distributed noise to the data points. We decompose the error on the estimation in two terms: The error measuring the gap between the quantiles of $\mathbb{P}_{\mathbf{X}}$ and the ones of $\mathbb{P}_{\tilde{\mathbf{X}}}$ and the error made by JointExp on the estimation of the quantiles of $\mathbb{P}_{\tilde{\mathbf{X}}}$. The first term can be controlled by the following general purpose proposition which proof is postponed to subsection D.4.

Proposition 3. *For any non-increasing $f : \mathbb{R} \to [0,1]$ such that $\forall t \geq 0, \mathbb{P}(|w| > t) \leq f(t)$, then for every $p \in (0,1)$, for every $t \geq 0$ such that $1-f(t) > 0$,*

$$F_{\tilde{X}}^{-1}(p) \leq F_X^{-1}\left(\frac{p}{1-f(t)}\right)+t\,,$$

$$\sup_{\delta\in(0,p)} -F_{-X}^{-1}\left(\frac{1-p+\delta}{1-f(t)}\right)-t \leq F_{\tilde{X}}^{-1}(p)\,.$$

For instance, when applied to some noise with distribution $d\mathbb{P}_w(w) = \frac{\mathbb{1}_{[-\alpha,\alpha]}(w)}{2\alpha}dw$ with $t = \alpha$ and $f(t) = 0$, if F_X is continuous and strictly increasing on a neighborhood of $F_X^{-1}(p)$, we can say that $|F_X^{-1}(p) - F_{\tilde{X}}^{-1}(p)| \leq \alpha$. The second error term can be controlled with Theorem 2 assuming that we fall into its hypothesis. By adding some uniform noise in $[-\alpha, \alpha]$, we then obtain the following result:

Theorem 3. *If the distribution of X is a mixture of a finite number of Diracs in (a, b) and of a random variable Y with a continuous density π_Y on $[a, b]$ w.r.t. Lebesgue measure such that $\pi_Y > 0$ on $[a, b] \setminus \mathcal{O}$ where $\mathcal{O}$ is a finite union of intervals and $\pi_Y = 0$ on $\mathcal{O}$, then for any precision δ and Lebesgue-almost-any probability vector $\mathbf{p}$, there exist a noise level $\alpha > 0$ such that the ϵ-DP estimator $\mathbf{q}$ based on HSJointExp satisfies*

$$||\mathbf{q} - F_X^{-1}(\mathbf{p})||_\infty \leq \delta$$

with high probability (as n grows).

The proof is in subsection D.6. Theorem 3 states in particular that many distributions that satisfy the hypothesis of Proposition 1 and on which JointExp is not consistent also satisfy the hypothesis of Theorem 3 and HSJointExp can thus achieve arbitrary levels of precision on them (provided n is large enough).

As highlighted by subsection 4.2, working on much stricter distribution classes can lead to numerically tractable optimal levels of noise.

4.4 Privacy Amplification of HSJointExp

One would think that adding noise to the data does not only preserve the privacy guarantees of the original mechanism (as stated by Proposition 2) but actually makes it more private. In order to evaluate the actual privacy of our mechanism, we investigate its privacy loss:

$$\mathcal{L}(\mathbf{X}, \mathbf{Y}, \mathbf{q}) := \frac{d\mathbb{P}/d\mathbf{q}\left(\mathcal{E}_{u_{\text{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) = \mathbf{q}\right)}{d\mathbb{P}/d\mathbf{q}\left(\mathcal{E}_{u_{\text{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{Y}}) = \mathbf{q}\right)}$$

for $\mathbf{Y} \sim \mathbf{X}$ and $\mathbf{q} \in O$ where $d\mathbb{P}/d\mathbf{q}\left(\mathcal{E}_{u_{\text{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) = \mathbf{q}\right)$ refers to the value of the density of HSJointExp applied to $\mathbf{X}$ at $\mathbf{q}$. For a given dataset $\mathbf{X}$, we define $\epsilon_{\text{eff}} := \sup_{\mathbf{X}\sim\mathbf{Y}} \sup_{\mathbf{q}} \log\left(\mathcal{L}(\mathbf{X}, \mathbf{Y}, \mathbf{q})\right)$ the effective difficulty of distinguishing $\mathbf{X}$ from any of its neighbors. We always have that $\epsilon_{\text{eff}} \leq \epsilon$ but we would like to measure the difference between the two and its dependence on the noise level. The theoretical study of such is out of the scope of this article and is left as future work. However, we conduct a numerical analysis of such in Section 5.

5 Numerical results

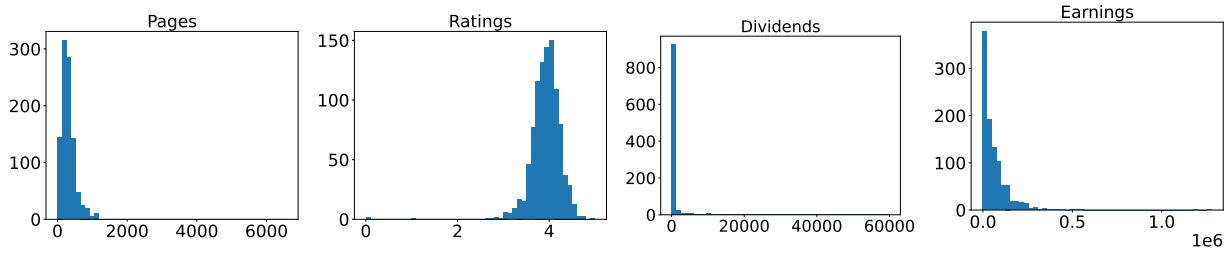
This section presents the behavior of HSJointExp on real world datasets. subsection 5.2 discusses its numerical utility and subsection 5.3 its privacy.

5.1 Datasets

The experiments are conducted on four datasets. *Pages* and *Ratings* (see Figure 2) are subsampled from a collection of ratings and of pages of books from the Goodreads-Books dataset [Sou]. Gillenwater, Joseph and Kulesza [GJK21] used the same ones as numerical evidences of the performances of JointExp. *Earnings* (resp. *Dividends*) are subsampled from the family earnings (resp. from the family earnings from dividends) entries from the 2021 Census [Bur].

Pages and Ratings: Due to the relatively smooth distributions. JointExp performed well on those datasets in its original paper [GJK21] and we expect HSJointExp to perform identically good on them.

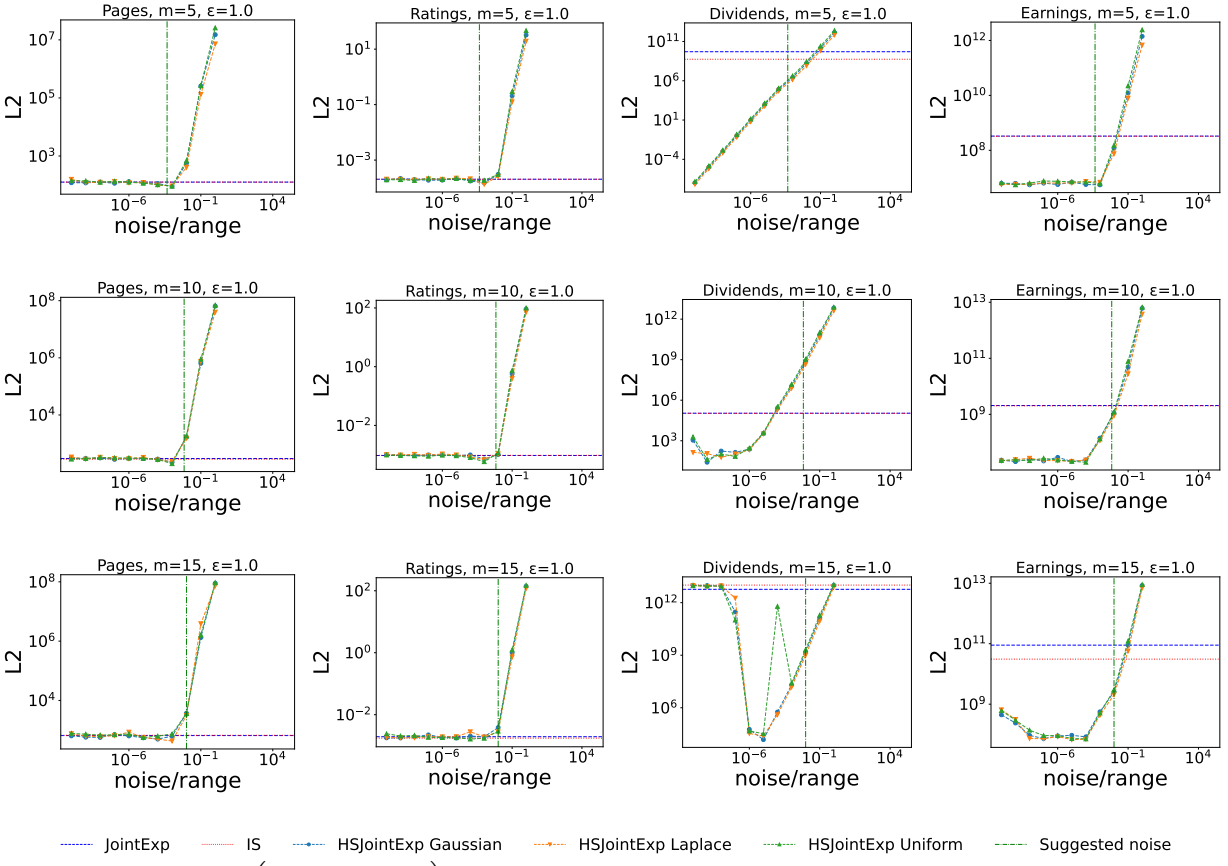
Dividends and Earnings: With accumulation points at 0, we envision JointExp to be the suboptimal. Comparatively, HSJointExp should perform much better on those datasets.



$n = 1000$ data points subsampled from the original datasets and binned in 50 bins.

Figure 2: Datasets used for experiments

5.2 Measuring the utility:



$n = 1000$, $\epsilon = 1$, $\mathbf{p} = \left(\frac{1}{m+1}, \ldots, \frac{m}{m+1}\right)$, $L2 =$ MSE between the private estimator and the quantiles of the dataset averaged over 100 runs. noise/range indicates the ratio between the standard deviation of the noise and the range of the data. Vertical bars indicate the level of noise that we recommend. Horizontal bars indicate the $L2$ scores that achieved respectively JointExp and the Inverse Sensitivity mechanism (indistinguishable on some plots).

Figure 3: Dependence to the smoothing level of the private estimators, $\epsilon = 1$.

Figure 3 represents the utility of HSJointExp for different noise structures, different levels of noise and different m's. The horizontal bars represent the scores of JointExp and of the Inverse Sensitivity mechanism. The first striking fact is how close JointExp and IS perform, confirming our allegations. The scores of

JointExp will serve as reference to compare our heuristically smoothed variants to.

Limit behaviors: When the standard deviation of the noise tends to 0, the performances of HSJointExp and of JointExp match. In this scenario, the noisy points converge to their original place. On the other hand, when the noise level exceeds a certain threshold, the performances degrade quickly. Indeed, the order statistics of the noisy dataset has lost all significance.

About the structure of the noise: In our analysis we focused mainly on uniform noise but as we can see experimentally, the structure of the noise does not make any substantial difference.

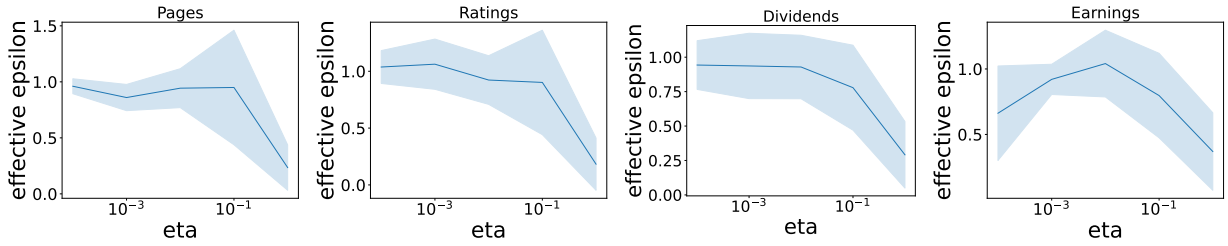
On Pages and Ratings: JointExp already performed well because the datasets did not really have any accumulation point. As long as the noise is not unreasonably high, HSJointExp performs equally good.

On Earnings and Dividends: The concentration of the data forces JointExp to be suboptimal. Indeed, experiments show that HSJointExp gives results that are orders of magnitude better for certain values of the noise.

Tuning the noise: We generalize the optimal value of noise that we found in subsection 4.2 to give an heuristic on the general choice of σ the standard deviation of the noise. In Figure 3, the vertical bars represent $\sigma = (b-a)\min\left\{10^{-2}, e^{-\frac{n\epsilon}{20\sqrt{m}}}\right\}$ which seems to be a decent choice for the noise level. We added the scaling in $(b-a)$ and we replaced ϵ by $\epsilon/\sqrt{m}$ as if each each quantile was to be estimated independently by a strong composition theorem. The term 10^{-2} is just here to make sure that the points are not shifted extensively and serves as a safety check.

Dependency on ϵ: Since the estimation of quantiles is mainly used as a subroutine, we conducted some experiments to see if our observations still hold for smaller values of ϵ (in order to be used with composition theorems). We present the results in Figure 5 for $\epsilon = 0.5$ and Figure 6 for $\epsilon = 0.1$ both in Appendix C. The improvements of HSJointExp over regular JointExp are still present even though they are less noticeable due to higher privacy noise. The heuristic for the noise calibration also seems to work.

5.3 Privacy Amplification



η = noise/range. The blued region represents the mean estimation plus or minus its standard deviation. $\epsilon = 1$.

Figure 4: Evolution of ϵ_{eff} for the median estimation

In Figure 4 we numerically estimate ϵ_{eff} in the following setup: For each or the the datasets (noted $\mathbf{X}$), we estimate the median using HSJointExp with Laplace noise tuned with $\epsilon = 1$. We estimate $\mathcal{L}(\mathbf{X}, \mathbf{Y}, \mathbf{q})$ for any $\mathbf{Y} \sim \mathbf{X}$ by discretizing the search space of $\mathbf{Y}$ and by Monte Carlo averaging to integrate with respect to the noise.

The variance of the resulting ϵ_{eff} is high but we can see two regimes: For low values of the noise, the privacy of the mechanism is unchanged. For high values of noise on the other hand, $\epsilon_{\text{eff}} < \epsilon$ and differentiating the datasets from their neighbors is harder.

By crossing the results with Figure 3 however, it seems that the privacy amplification only occurs for values of the noise for which the utility of HSJointExp is already degraded compared to regular JointExp.

References

[Abo18] John M Abowd. The us census bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2867–2867, 2018.

[ACG+16] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.

[AD20a] Hilal Asi and John C Duchi. Instance-optimality in differential privacy via approximate inverse sensitivity mechanisms. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 14106–14117. Curran Associates, Inc., 2020.

[AD20b] Hilal Asi and John C Duchi. Near instance-optimality in differential privacy. *arXiv preprint arXiv:2005.10630*, 2020.

[All] Joshua et. al Allen. Smartnoise core differential privacy library. `https://github.com/opendp/smartnoise-core`.

[BAM20] Victor-Emmanuel Brunel and Marco Avella-Medina. Propose, test, release: Differentially private estimation with high probability. *arXiv preprint arXiv:2002.08774*, 2020.

[BDK07] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. Wherefore art thou r3579x? anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th international conference on World Wide Web*, pages 181–190, 2007.

[Bur] United States Census Bureau. 2021 annual social and economic supplements. `https://www.census.gov/data/datasets/2021/demo/cps/cps-asec-2021.html`.

[CG16] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[DDR20] Jinshuo Dong, David Durfee, and Ryan Rogers. Optimal differential privacy composition for exponential mechanisms. In *International Conference on Machine Learning*, pages 2597–2606. PMLR, 2020.

[DKM+06] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006.

[DKY17] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. *arXiv preprint arXiv:1712.01524*, 2017.

[DL09] Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 371–380, 2009.

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.

[DN03] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 202–210, 2003.

[DR+14] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

[DRS19] Jinshuo Dong, Aaron Roth, and Weijie J Su. Gaussian differential privacy. *arXiv preprint arXiv:1905.02383*, 2019.

[EPK14] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067, 2014.

[FJR15] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 1322–1333, 2015.

[GJK21] Jennifer Gillenwater, Matthew Joseph, and Alex Kulesza. Differentially private quantiles. *arXiv preprint arXiv:2102.08244*, 2021.

[HSR+08] Nils Homer, Szabolcs Szelinger, Margot Redman, David Duggan, Waibhav Tembe, Jill Muehling, John V Pearson, Dietrich A Stephan, Stanley F Nelson, and David W Craig. Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS Genet*, 4(8):e1000167, 2008.

[IBM] IBM. Smartnoise core differential privacy library. `https://github.com/IBM/differential-privacy-library`.

[KOV15] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. In *International conference on machine learning*, pages 1376–1385. PMLR, 2015.

[LDM10] Grigorios Loukides, Joshua C Denny, and Bradley Malin. The disclosure of diagnosis codes can breach research participants' privacy. *Journal of the American Medical Informatics Association*, 17(3):322–327, 2010.

[MMS21] Ryan McKenna, Gerome Miklau, and Daniel Sheldon. Winning the nist contest: A scalable and general approach to differentially private synthetic data. *arXiv preprint arXiv:2108.04978*, 2021.

[MT07] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.

[NRS07] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 75–84, 2007.

[NS06] Arvind Narayanan and Vitaly Shmatikov. How to break anonymity of the netflix prize dataset. *arXiv preprint cs/0610105*, 2006.

[NS08] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (sp 2008)*, pages 111–125. IEEE, 2008.

[Smi11] Adam Smith. Privacy-preserving statistical estimation with optimal convergence rates. In *Proceedings of the forty-third annual ACM symposium on Theory of computing*, pages 813–822, 2011.

[Sou] Soumik. Goodreads-books dataset. `https://www.kaggle.com/jealousleopard/goodreadsbooks`.

[Swe00] Latanya Sweeney. Simple demographics often identify people uniquely. *Health (San Francisco)*, 671(2000):1–34, 2000.

[Swe02] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[TVV+17] Abhradeep Guha Thakurta, Andrew H Vyrros, Umesh S Vaishampayan, Gaurav Kapoor, Julien Freudiger, Vivek Rangarajan Sridhar, and Doug Davidson. Learning new words. *Granted US Patents*, 9594741, 2017.

[WE18] Isabel Wagner and David Eckhoff. Technical privacy metrics: a systematic survey. *ACM Computing Surveys (CSUR)*, 51(3):1–38, 2018.

[ZCP+17] Jun Zhang, Graham Cormode, Cecilia M Procopiuc, Divesh Srivastava, and Xiaokui Xiao. Privbayes: Private data release via bayesian networks. *ACM Transactions on Database Systems (TODS)*, 42(4):1–41, 2017.

A Sampling from the inverse sensitivity mechanism

In this subsection we explain how to sample exactly from the inverse sensitivity mechanism for multiple quantiles in polynomial time and memory. It is essentially an adaptation from the JointExp algorithm and hence we will use the same notations when possible. For simplicity, $\mathbf{X}$ is assumed to be sorted.

The sampling density of $\mathcal{E}^{(2/\epsilon)}_{\tilde{u}_{\text{IS}}}(\mathbf{X})$ is constant on sets $\big([X_{i_1}, X_{i_1+1}) \times \ldots \times \cdots \times [X_{i_m}, X_{i_m+1})\big) \cap [a,b]^m \nearrow$ for $\mathbf{i} = (i_1, \ldots, i_m) \in O'$ where

$$O' = \{\mathbf{i} \in \{0, \ldots, n\}^m, 0 \leq i_1 \leq \cdots \leq i_m \leq m\} \ .$$

Hence, a finite sampling algorithm for $\mathcal{E}^{(2/\epsilon)}_{u_{\text{IS}}}(\mathbf{X})$ is to:

- sample $\mathbf{i} = (i_1, \ldots, i_m) \in O'$ under $\mathbb{P}_{O'}$;
- sample q'_j uniformly in $[X_{i_j}, X_{i_j+1})$, independently for all j in $\{1 \ldots m\}$;
- output $(q'_j)_{j \in \{1 \ldots m\}}$ sorted by increasing order;

with the probability $\mathbb{P}_{O'}$ defined on O' as

$$\mathbb{P}_{O'}(\mathbf{i}) \propto \frac{1}{\gamma(\mathbf{i})} \Pi_{j=1}^{m+1} \phi(i_{j-1}, i_j, j) \Pi_{j=1}^{m} \tau(i_j) \tag{4}$$

where, if we denote by $\text{count}_{\mathbf{i}}(i)$ the number of occurences of integer i in the ordered tuple $\mathbf{i}$,

$$\forall \mathbf{i} \in O', \gamma(\mathbf{i}) = \Pi_{i=0}^{m} \text{count}_{\mathbf{i}}(i)! \ ,$$

$$\forall i \in \{0, \ldots, m\}, \tau(i) = X_{i+1} - X_i \ ,$$

and for $0 \leq i, i' \leq m$ and $1 \leq j \leq m+1$,

$$\phi(i, i', j) = \begin{cases} 0, & \text{if } i' < i \\ e^{-\frac{\epsilon}{2}\left(\frac{1}{2}|\hat{\delta}(i,i',m+1)|\right)}, & \text{if } j = m+1 \\ e^{-\frac{\epsilon}{2}\left(\frac{1}{2}|\hat{\delta}(i,i',j)|\right) + \mathbb{1}_{\mathbb{R}_+}(\hat{\delta}(i,i',j))}, & \text{otherwise} \end{cases}$$

with $\hat{\delta}(i, i', j) = i' - i - (\lceil np_j \rceil - \lceil np_{j-1} \rceil)$.

Since O' has a finite cardinality bounded by $(n+1)^m$, it is possible to compute the probability of all the elements in that space and to sample this way. However, the fact that this complexity is exponential in m makes it unusable in practice. Gillenwater, Joseph and Kulesza [GJK21] present an algorithm that allows to sample from any distribution that factorizes in an analog form of (4) that has a complexity (both in time and space) of $O(n^2 m + m^2 n)$. Furthermore, if the function $\phi(i, i', j)$ can be rewritten as $\phi'(i' - i, j)$ (which is the case in our problem), the complexity becomes $O(mn \log n + m^2 n)$. Overall, in order to sample efficiently from the inverse sensitivity mechanism, one can use Algorithm 1 proposed by Gillenwater, Joseph and Kulesza [GJK21] by taking great care of using a sensitivity of 1 (instead of 2) and by replacing the function ϕ by the one used in this article.

B Inverse sensitivity smoothing [AD20b]

B.1 General principle

When the output space O is a subset of a Euclidean space, the theory of Inverse Sensitivity comes with a a smoothing operation with some parameter $\rho > 0$. The utility function u_{IS} can be replaced with

$$u^{\rho}_{\text{IS}}(\mathbf{X}, o) = \sup_{o' \in O : ||o - o'||_2 \leq \rho} u_{\text{IS}}(\mathbf{X}, o') \ .$$

It is easy to see that this new utility function has sensitivity $\Delta u^{\rho}_{\text{IS}} = 1$. Contrary to the non smoothed inverse sensitivity mechanism which only comes with guarantees for finite output spaces O, this smoothed

version gives more general results that only rely on a few mathematical tools. Using the notion of *modulus of continuity* of the target function $\mathcal{Q}$, defined as

$$\omega_{\mathcal{Q}}(\mathbf{X}, k) = \sup_{\mathbf{X}' \in \mathfrak{X}^n : d(\mathbf{X}, \mathbf{X}') \leq k} \left\{ ||\mathcal{Q}(\mathbf{X}) - \mathcal{Q}(\mathbf{X}')||_2 \right\},$$

and the corresponding image

$$W_{\mathcal{Q}}(\mathbf{X}, k) = \{\mathcal{Q}(\mathbf{X}') - \mathcal{Q}(\mathbf{X}) : d(\mathbf{X}, \mathbf{X}') \leq k\},$$

one can bound the estimation error [AD20a] assuming that $\mathrm{diam}_2(\mathcal{Q}(\mathfrak{X}^n)) \leq D$: for $1 \leq k \leq n$

$$\begin{aligned} \mathbb{P}\Big(||\mathcal{E}^{(2/\epsilon)}_{u^{\rho}_{\mathrm{IS}}}(\mathbf{X}) - \mathcal{Q}(\mathbf{X})||_2 \geq \omega_{\mathcal{Q}}(\mathbf{X}, k) + \rho \Big) \\ \leq e^{-k\epsilon/2} \left(\tfrac{D}{\rho}\right)^m \end{aligned}$$

with m the ambient space dimension (i.e., $O \subset \mathbb{R}^m$). The original authors consider a smoothing parameter $\rho = 1/n^r$ for some $r > 0$ and k of the order of $(4rm \log n)/\epsilon$, which yields an estimation error bounded with high probability by the modulus of continuity: With high probability

$$||\mathcal{E}^{(2/\epsilon)}_{u^{1/n^r}_{\mathrm{IS}}}(\mathbf{X}) - \mathbf{X}||_2 \leq O\left(\omega_{\mathcal{Q}}(\mathbf{X}, (4rm \log n)/\epsilon) + 1/n^r\right) . \tag{5}$$

This theory also provides an optimality result. Under the rather strong hypothesis that there exists a uniform $c > 0$ such that for all $1 \leq k \leq n$ and $\mathbf{X} \in \mathfrak{X}^n$,

$$W_{\mathcal{Q}}(\mathbf{X}, k) \supseteq c.\omega_{\mathcal{Q}}(\mathbf{X}, k).\mathbb{B}_2^m \tag{6}$$

where $\mathbb{B}_2^m$ is the l_2 ball in $\mathbb{R}^m$, the best ϵ-DP algorithm roughly behaves the same way in a local minimax sense up to a logarithmic factor [AD20a]:

$$\begin{aligned} \inf_{\mathcal{A} \in \mathcal{A}_\epsilon} \sup_{\mathbf{X}' : d(\mathbf{X}, \mathbf{X}') \leq m/\epsilon} & \mathbb{E}(||\mathcal{A}(\mathbf{X}') - \mathcal{Q}(\mathbf{X}')||_2) \\ & \geq \Omega(\omega_{\mathcal{Q}}(\mathbf{X}, m/\epsilon)) \end{aligned}$$

where $\mathcal{A}_\epsilon$ is the class of ϵ-DP algorithms.

B.2 For the multiquantile problem

The problem of sampling: The hypothesis of Theorem 1 restrict our ability to compute the inverse sensitivity of quantile candidates without collisions and that do not overlap with any of the data intervals. Computing the supremum in the definition of the smoothed inverse sensitivity would not only require to handle those cases but also to have an algorithm faster than looking at all the possibilities. We did not manage to overcome that difficulty hence the reason for our heuristic smoothing (see Section 4).

Behavior of the modulus of continuity: The modulus of continuity measures the maximal variation of a function on a ball for Hamming distance k. Here we derive a majoration for the multiquantile problem. Assuming that $\mathbf{X} \in \mathfrak{X}^n$ is sorted, by moving a single point ($k = 1$) of $\mathbf{X}$, two behaviors can happen: Either it exactly matches one of the quantiles, and the corresponding estimate can then vary continuously in the interval between the data point below and the data point above; or it did not match any quantile but it can still shift the entire ordered statistics by one data point. This bounds the values of the function $\mathcal{Q}$ at

Hamming distance 1 of $\mathbf{X}$ (i.e., $W_{\mathcal{Q}}(\mathbf{X}, 1)$):

$$\begin{aligned} W_{\mathcal{Q}}(\mathbf{X}, 1) + \mathcal{Q}(\mathbf{X}) \subseteq \bigcup_{i=1}^{m} \Big\{ & \mathbf{X}_{[\lceil np_1 \rceil - 1 : \lceil np_1 \rceil + 1]} \\ & \times \dots \\ & \times \mathbf{X}_{[\lceil np_{i-1} \rceil - 1 : \lceil np_{i-1} \rceil + 1]} \\ & \times [X_{\lceil np_i \rceil - 1}, X_{\lceil np_i \rceil + 1}] \\ & \times \mathbf{X}_{[\lceil np_{i+1} \rceil - 1 : \lceil np_{i+1} \rceil + 1]} \\ & \times \dots \\ & \times \mathbf{X}_{[\lceil np_m \rceil - 1 : \lceil np_m \rceil + 1]} \Big\} \end{aligned}$$

where we use the notation from computer science $\mathbf{X}_{[i:j]} = (X_i, \dots, X_j)$. When $k \geq 1$ points are moving, the same mechanics arise where the ordered statistics is shifted by at most k points and where we have at most k degrees of freedom. This yields an analog majoration with k intervals in each cartesian product:

$$W_{\mathcal{Q}}(\mathbf{X}, k) + \mathcal{Q}(\mathbf{X}) \subseteq \bigcup_{1 \leq i_1 \leq \dots \leq i_k \leq m} \bigotimes_{j=1}^{m} \mathcal{I}(\mathbf{X}, j, \mathbf{i}, k) \ , \tag{7}$$

where, if we note $\mathbf{i} = (i_1, \dots, i_k)$,

$$\mathcal{I}(\mathbf{X}, j, \mathbf{i}, k) = \begin{cases} [X_{\lceil np_j \rceil - k}, X_{\lceil np_j \rceil + k}], & \text{if } j \in \mathbf{i} \\ \mathbf{X}_{[\lceil np_j \rceil - k : \lceil np_j \rceil + k]}, & \text{if } j \notin \mathbf{i} \end{cases} .$$

Using (7), the modulus of continuity is bounded as

$$\omega_{\mathcal{Q}}(\mathbf{X}, k) \leq \sqrt{\sum_{i=1}^{m} \left(X_{\lceil np_i \rceil + k} - X_{\lceil np_i \rceil - k} \right)^2} \ . \tag{8}$$

Hence, when the dataset $\mathbf{X}$ has many points close to its empirical quantiles, the modulus of continuity $\omega_{\mathcal{Q}}(\mathbf{X}, k)$ should not grow too fast in k.

Convergence with high probability : The concentration bound (i.e., Equation (5)) along with the upper bound on the modulus of continuity (i.e., Equation (8)) gives that, with high probability

$$||\mathcal{E}^{(2/\epsilon)}_{u_{\text{IS}}^{1/n^r}}(\mathbf{X}) - O(\mathbf{X})||_2 \leq O\left(\sqrt{\sum_{i=1}^{m} \left(X_{\lceil np_i \rceil + \delta} - X_{\lceil np_i \rceil - \delta} \right)^2} + \frac{1}{n^r} \right) \ ,$$

where $\delta = \lceil (4rm \log n)/\epsilon \rceil$. Hence, whenever the dataset has many points in the neighborhood of the empirical quantiles, we can expect the smoothed inverse sensitivity mechanism for multiquantile (i.e., $\mathcal{E}^{(2/\epsilon)}_{u_{\text{IS}}^{1/n^r}}(\mathbf{X})$) to perform well. This will typically be the case when the data distribution has a Dirac on a quantile whose mass accounts for more than $\frac{4rm \log n}{n\epsilon}$ or when the data distribution has a density that is strictly positive on a neighborhood of its quantiles and δ is not too large.

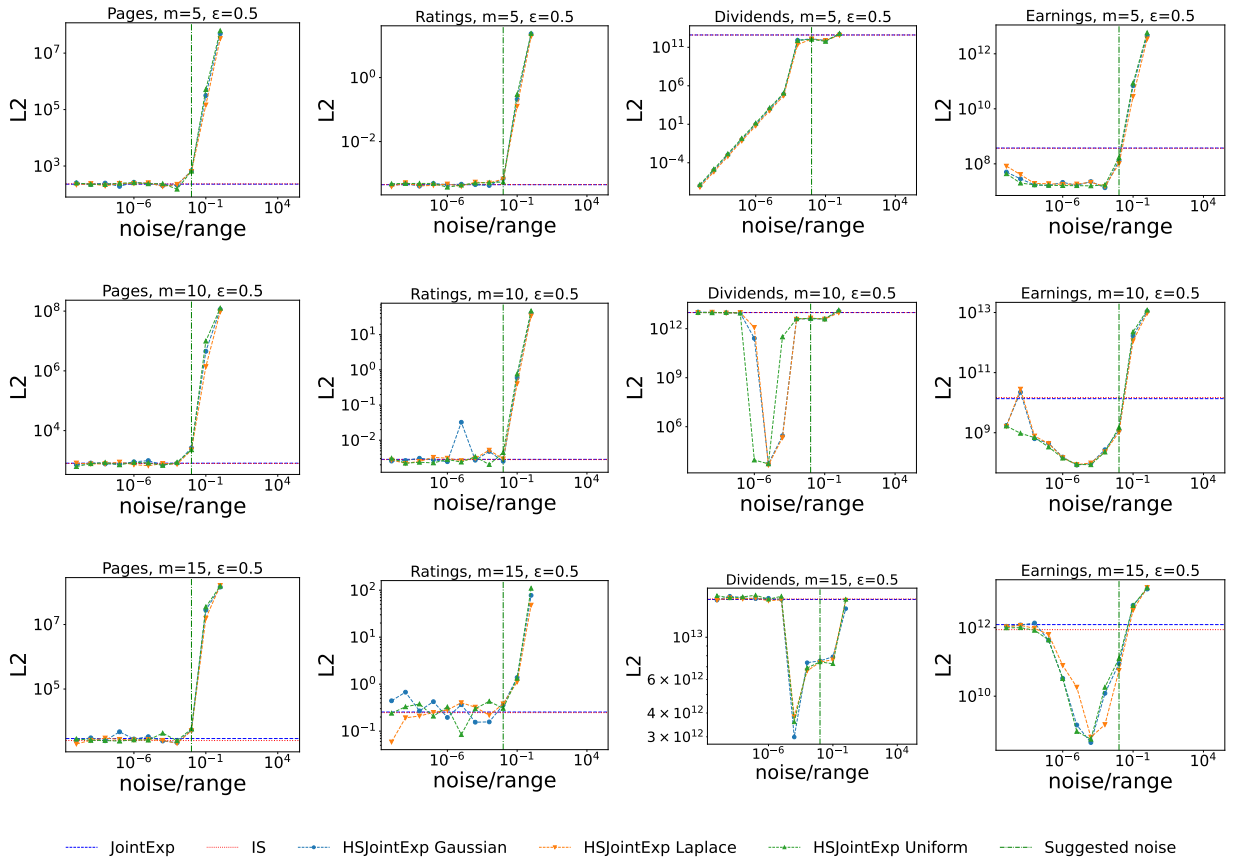
Local minimax bound In (7), the inner term $\bigotimes_{j=1}^{m} \mathcal{I}(\mathbf{X}, j, \mathbf{i}, k)$ has null Lebesgue measure if $k < m$. Indeed, in this case, at least one factor in the product is countable. As a consequence, denoting λ the Lebesgue measure,

$$k < m \Rightarrow \lambda(W_{\mathcal{Q}}(\mathbf{X}, k)) = 0.$$

Hence $W_{\mathcal{Q}}$ does not satisfy the uniform condition expressed in (6) and the local minimax bound falls. The only setup where this bound holds is in the case of single quantile estimation (i.e., $m = 1$). As a consequence,

we may not expect the optimality of the smoothed inverse sensitivity mechanism for multiquantile in the class of ϵ-DP algorithm. We included this negative result for the sake of exploring the theory of inverse sensitivity as a whole. It shows that even if a sampling procedure for the smoothed inverse sensitivity mechanism for multiquantile was to be found, finding an optimal sampling algorithm for private multiple quantiles is still an open question.

C More numerical results



$n = 1000$, $\epsilon = 0.5$, $\mathbf{p} = \left(\frac{1}{m+1}, \ldots, \frac{m}{m+1}\right)$, $L2$ = MSE between the private estimator and the quantiles of the dataset averaged over 100 runs. noise/range indicates the ratio between the standard deviation of the noise and the range of the data. Vertical bars indicate the level of noise that we recommend. Horizontal bars indicate the $L2$ scores that achieved respectively JointExp and the Inverse Sensitivity mechanism (indistinguishable on some plots).

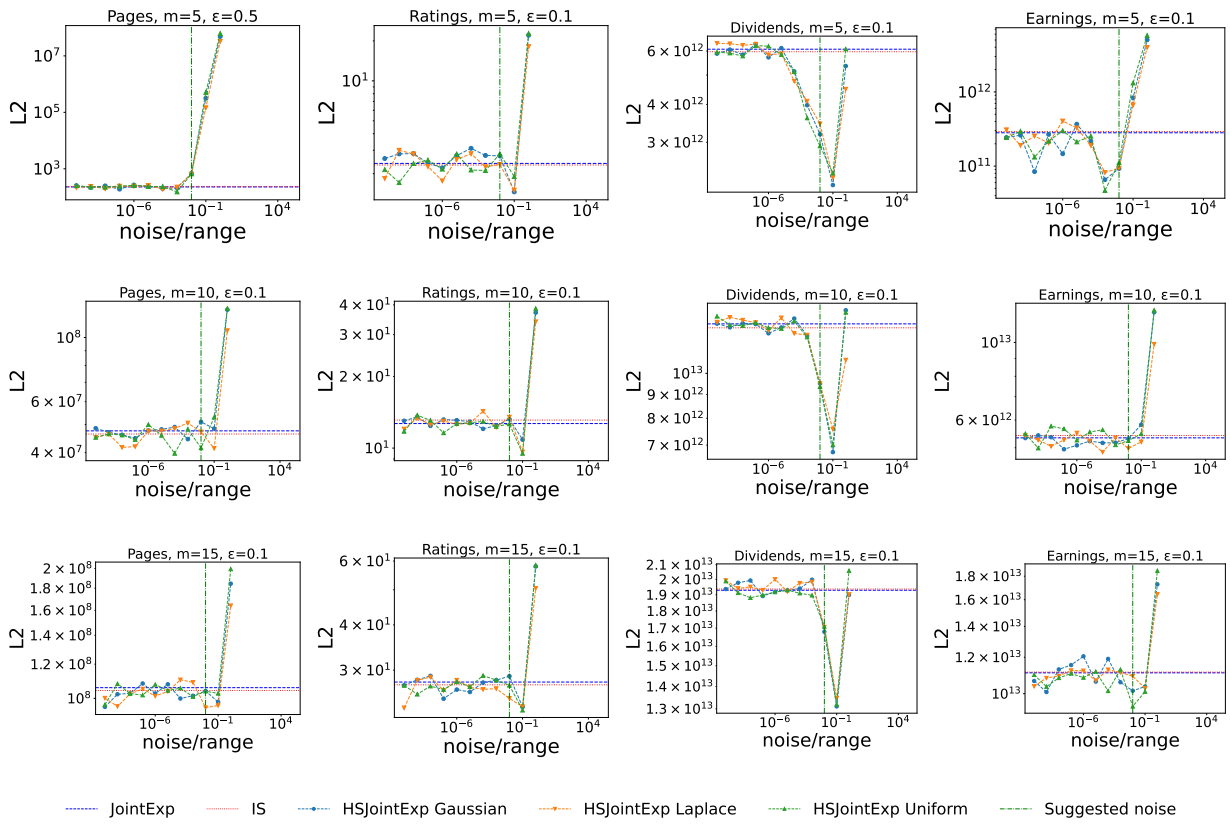
Figure 5: Dependence to the smoothing level of the private estimators, $\epsilon = 0.5$.

D Omitted proofs

D.1 Proof of Theorem 1

If $\mathbf{Y} \in Q^{-1}(\mathbf{q})$ then:

- each "bin" has the right number of points: $\delta(i, \mathbf{Y}, \mathbf{q}) = 0$, $i \in \{2 \ldots m+1\}$, and $\delta_{\texttt{closed}}(1, \mathbf{Y}, \mathbf{q}) = 0$.
- every point of $\mathbf{q}$ appears in $\mathbf{Y}$: $\mathbf{q} \subseteq \mathbf{Y}$.



$n = 1000$, $\epsilon = 0.1$, $\mathbf{p} = \left(\frac{1}{m+1}, \ldots, \frac{m}{m+1}\right)$, $L2$ = MSE between the private estimator and the quantiles of the dataset averaged over 100 runs. noise/range indicates the ratio between the standard deviation of the noise and the range of the data. Vertical bars indicate the level of noise that we recommend. Horizontal bars indicate the $L2$ scores that achieved respectively JointExp and the Inverse Sensitivity mechanism (indistinguishable on some plots).

Figure 6: Dependence to the smoothing level of the private estimators, $\epsilon = 0.1$.

Then we can understand the modifications that have to be made to $\mathbf{X}$ in order to obtain a $\mathbf{Y} \in Q^{-1}(\mathbf{q})$. For the first condition, some points have to be moved from bins in excess to bins in deficit. This procedure accounts for $\sum_{i=2}^{m+1} \delta(i, \mathbf{X}, \mathbf{q})_+ + \delta_{\text{closed}}(1, \mathbf{X}, \mathbf{q})_+$ operations which can be reformulated as $\frac{1}{2}\sum_{i=2}^{m+1} |\delta(i, \mathbf{X}, \mathbf{q})| + \frac{1}{2}|\delta_{\text{closed}}(1, \mathbf{X}, \mathbf{q})|$. For the second condition, we have to make sure that for all i, q_i belongs to the dataset. For a bin in strict deficit, at least a point has to be added to it due to the first condition. Hence, we can make sure to add the associated quantile at no extra cost. For a bin in excess on the other hand, since by hypothesis $\mathbf{q} \cap \mathbf{X} = \emptyset$, a point in the bin will have to be be replaced by the associated quantile at an extra cost of 1. In the end, we find the desired result.

D.2 Proof of Proposition 1

Since $\mathbb{P}_X(\{q\}) > 0$ and $\mathbb{P}_X(I \setminus \{q\}) = 0$, there exists a nonempty interval A of $[0, 1]$ such that $\{q\} = F_X^{-1}(A)$ with $\lambda(A) \geq \mathbb{P}_X(\{q\})$, λ refering to Lebesgue measure. Let us prove that any $\mathbf{p}$ with at least one component in A satisfies (1). For this, assume that $\mathbf{p}$ has its i^{th} entry p_i in A. Due to the structure of $\mathbb{P}_X$, $\mathbb{P}_X(\mathbf{X} \cap (I \setminus \{q\}) \neq \emptyset) = 0$, hence almost surely it holds that for every j we have either $|X_j - q| \geq \eta > 0$ or $|X_j - q| = 0$. Remember that the output density is a mixture of uniforms on the sets $\left([X_{i_1}, X_{i_1+1}) \times \ldots \times \cdots \times [X_{i_m}, X_{i_m+1})\right) \cap [a, b]^{m\nearrow}$ for $\mathbf{i} = (i_1, \ldots, i_m) \in O'$. If the i^{th} component of the output q_i was to be sampled from a data interval that doesn't admit q in its closure, then $||\mathbf{q} - F_X^{-1}(\mathbf{p})||_2^2 \geq \eta^2$. If on the other hand q_i was to be sampled from a data interval that does admit q in its closure, then it belongs to an interval $[X_k, X_{k+1})$ for some k such that

$q \in [X_k, X_{k+1}]$ and $X_{k+1} - X_k \geq \eta$. Conditionally to the fact that there are $m' \leq m$ other quantiles that are sampled from $[X_k, X_{k+1}]$, the conditional expectation of $||\mathbf{q} - F_X^{-1}(\mathbf{p})||_2^2$ can be lower by a (strictly) positive functional ($f(\eta, m')$) of η and m' (because the corresponding slice of the output is uniform on $[X_k, X_{k+1}]^{m'} \nearrow$. This shows that the risk can be lower bounded by a quantity in $\text{Span}\{\eta^2, f(\eta, 1), \ldots, f(\eta, m)\}$ which is then bigger than $\min\{\eta^2, f(\eta, 1), \ldots, f(\eta, m)\}$ which is positive.

D.3 Proof of Proposition 2

Let $\mathcal{A}$ be a ϵ-DP algorithm on $\mathfrak{X}^n$, $\mathbf{X}, \mathbf{X}' \in \mathfrak{X}^n$ such that $\mathbf{X} \sim \mathbf{X}'$. Then, for every $\mathbf{w} \in \mathbb{R}^n, \text{proj}_{\mathfrak{X}^n}(\mathbf{X} + \mathbf{w}) \sim \text{proj}_{\mathfrak{X}^n}(\mathbf{X}' + \sigma(\mathbf{w}))$ for a specific permutation of the components σ. For each measurable set $\mathcal{S} \subseteq O$ we get

$$\begin{aligned}
&\mathbb{P}(\mathcal{A}(\text{proj}_{\mathfrak{X}^n}(\mathbf{X} + \mathbf{w})) \in \mathcal{S}) \\
&= \int_{\mathbb{R}^n} \mathbb{P}_{\mathcal{A}}(\mathcal{A}(\text{proj}_{\mathfrak{X}^n}(\mathbf{X} + \mathbf{w})) \in \mathcal{S}) \mathbb{P}_{\mathbf{w}}(d\mathbf{w}) \\
&\leq e^{\epsilon} \int_{\mathbb{R}^n} \mathbb{P}_{\mathcal{A}}(\mathcal{A}(\text{proj}_{\mathfrak{X}^n}(\mathbf{X}' + \sigma(\mathbf{w}))) \in \mathcal{S}) \mathbb{P}_{\mathbf{w}}(d\sigma(\mathbf{w})) \\
&= e^{\epsilon} \int_{\mathbb{R}^n} \mathbb{P}_{\mathcal{A}}(\mathcal{A}(\text{proj}_{\mathfrak{X}^n}(\mathbf{X}' + \mathbf{w})) \in \mathcal{S}) \mathbb{P}_{\mathbf{w}}(d\mathbf{w}) \\
&= e^{\epsilon} \mathbb{P}(\mathcal{A}(\text{proj}_{\mathfrak{X}^n}(\mathbf{X}' + \mathbf{w})) \in \mathcal{S})
\end{aligned}$$

which completes the proof.

D.4 Proof of Proposition 3

Let $t \geq 0$ such that $1 - f(t) > 0$,

$$\begin{aligned}
&\mathbb{P}\left(X + w \leq F_X^{-1}\left(\frac{p}{1 - f(t)}\right) + t\right) \\
&\geq \mathbb{P}\left(X + w \leq F_X^{-1}\left(\frac{p}{1 - f(t)}\right) + t, |w| \leq t\right) \\
&\geq \mathbb{P}\left(X \leq F_X^{-1}\left(\frac{p}{1 - f(t)}\right), |w| \leq t\right) \\
&\geq \mathbb{P}\left(X \leq F_X^{-1}\left(\frac{p}{1 - f(t)}\right)\right) \mathbb{P}\left(|w| \leq t\right) \\
&\geq \frac{p}{1 - f(t)}\left(1 - f(t)\right) \geq p\,.
\end{aligned}$$

So, $F_{\tilde{X}}^{-1}(p) \leq F_X^{-1}\left(\frac{p}{1-f(t)}\right) + t$. Let $\delta \in (0, p)$, the same arguments give

$$\mathbb{P}\left(X + w \leq -F_{-X}^{-1}\left(\frac{1 - p + \delta}{1 - f(t)}\right) - t\right) \leq p - \delta < p$$

which allows to conclude with the desired result.

D.5 Proof of Theorem 2

Lemma 1. *Let $\tilde{X}$ be a real random variable with density $\pi_{\tilde{X}}$ and $p \in (0, 1)$. We suppose that $\pi_{\tilde{X}} \geq \pi_{min} > 0$ on an open neighborhood $\mathcal{N}$ of $F_{\tilde{X}}^{-1}(p)$. If we have access to $\tilde{\mathbf{X}} = (\tilde{X}_1, \ldots, \tilde{X}_n)$ i.i.d. realisations of $\tilde{X}$ then for every $\gamma > 0$,*

$$[F_{\tilde{X}}^{-1}(p) - \gamma, F_{\tilde{X}}^{-1}(p) + \gamma] \subset \mathcal{N} \implies \mathbb{P}\left(\left|F_{\tilde{X}}^{-1}(p) - \tilde{X}_{(np)}\right| > \gamma\right) \leq e^{-n\left(\frac{\gamma^2 \pi_{min}^2}{2(1-p)}\right)} + e^{-n\left(\frac{\gamma^2 \pi_{min}^2}{2p}\right)}$$

Proof. Let $\gamma > 0$ such that $[F_{\tilde{X}}^{-1}(p) - \gamma, F_{\tilde{X}}^{-1}(p) + \gamma] \subset \mathcal{N}$. Let us define

$$N = \sum_{i=1}^{n} \mathbb{1}_{(F_{\tilde{X}}^{-1}(p)+\gamma, +\infty)}(\tilde{X}_i) \ .$$

N is a sum of n independent Bernoulli random variable with probabilities of success lower than $\eta = 1 - p - \gamma\pi_{\min}$. If $\tilde{X}_{(np)} > F_{\tilde{X}}^{-1}(p) + \gamma$, then $N \geq n(1-p)$. So,

$$\begin{aligned}
&\mathbb{P}\left(\tilde{X}_{(np)} > F_{\tilde{X}}^{-1}(p) + \gamma\right) \leq \mathbb{P}\left(N \geq n(1-p)\right) \\
&= \mathbb{P}\left(N \geq n\eta\left(1 + \frac{\gamma\pi_{\min}}{\eta}\right)\right) \\
&\leq e^{-n\eta\left(\frac{\gamma\pi_{\min}}{\eta}\right)^2/\left(2+\frac{\gamma\pi_{\min}}{\eta}\right)} \leq e^{-n\left(\frac{\gamma^2\pi_{\min}^2}{2(1-p)-\gamma\pi_{\min}}\right)} \leq e^{-n\left(\frac{\gamma^2\pi_{\min}^2}{2(1-p)}\right)}
\end{aligned}$$

where line 3 is deduced from line 2 by a multiplicative Chernoff bounds. Looking at the event $\left(\tilde{X}_{(np)} < F_{\tilde{X}}^{-1}(p) - \gamma\right)$ and a union bound give the expected result. □

Lemma 2. *Let $\tilde{X}$ be a real random variable with density $\pi_{\tilde{X}}$ and $p \in (0,1)$. We suppose that $\pi_{max} \geq \pi_{\tilde{X}} \geq \pi_{min} > 0$ on an interval I of $\mathbb{R}$. If we note $N = \sum_{i=1}^{n} \mathbb{1}_I(\tilde{X}_i)$ the number of points that fall in I, we have*

$$\mathbb{P}\left(N \geq 2n\lambda(I)\pi_{max}\right) \leq e^{-\frac{n\lambda(I)\pi_{max}}{3}} \ ,$$

$$\mathbb{P}\left(N \leq \frac{1}{2}n\lambda(I)\pi_{min}\right) \leq e^{-\frac{n\lambda(I)\pi_{min}}{8}} \ .$$

.

Proof. This is a simple application of multiplicative Chernoff bounds to the sum N of independent Bernoulli random variables. □

Let $0 < \gamma < \beta$ such that $\pi_{\tilde{X}} > 0$ on $\mathcal{O} := \cup_{i=1}^{n}[F_{\tilde{X}}^{-1}(p_i) - \beta, F_{\tilde{X}}^{-1}(p_i) + \beta]$. We note $\pi_{\min} = \inf_{\mathcal{O}} \pi_{\tilde{X}}$ and $\pi_{\max} = \sup_{\mathcal{O}} \pi_{\tilde{X}}$. We also define the following events:

$$A : \forall i, \left|\tilde{X}_{(np_i)} - F_{\tilde{X}}^{-1}(p_i)\right| \leq \gamma \ ,$$

$$B : \forall i, \#(\tilde{\mathbf{X}} \cap [F_{\tilde{X}}^{-1}(p_i)+\gamma, F_{\tilde{X}}^{-1}(p_i)+\beta]) \geq \frac{1}{2}n(\beta-\gamma)\pi_{\min} \text{ and } \#(\tilde{\mathbf{X}} \cap [F_{\tilde{X}}^{-1}(p_i)-\beta, F_{\tilde{X}}^{-1}(p_i)-\gamma]) \geq \frac{1}{2}n(\beta-\gamma)\pi_{\min} \ ,$$

$$C : \forall i, \#(\tilde{\mathbf{X}} \cap [F_{\tilde{X}}^{-1}(p_i) - \gamma, F_{\tilde{X}}^{-1}(p_i) + \gamma]) \leq 2n2\gamma\pi_{\max} \ .$$

Then we can compute,

$$\frac{\mathbb{P}\left(||\mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) - F_{\tilde{X}}^{-1}(\mathbf{p})||_\infty > \beta | A, B, C\right)}{\mathbb{P}\left(||\mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) - F_{\tilde{X}}^{-1}(\mathbf{p})||_\infty \leq \beta | A, B, C\right)} \leq \frac{\mathbb{P}\left(||\mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) - F_{\tilde{X}}^{-1}(\mathbf{p})||_\infty > \beta | A, B, C\right)}{\mathbb{P}\left(||\mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) - F_{\tilde{X}}^{-1}(\mathbf{p})||_\infty \leq \gamma | A, B, C\right)}$$

Conditionally to A and B, $-u_{\mathrm{JE}}(\tilde{\mathbf{X}}, \mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}})) \leq \frac{1}{2}\left(\frac{1}{2}n(\beta-\gamma)\pi_{\min}\right) \implies ||\mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) - F_{\tilde{X}}^{-1}(\mathbf{p})||_\infty \leq \beta$. Furthermore, conditionally to A and C, $||\mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) - F_{\tilde{X}}^{-1}(\mathbf{p})||_\infty \leq \gamma \implies -u_{\mathrm{JE}}(\tilde{\mathbf{X}}, \mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}})) \leq \frac{1}{2}\left(4(m+1)n\gamma\pi_{\max}\right)$. So,

$$\frac{\mathbb{P}\left(||\mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) - F_{\tilde{X}}^{-1}(\mathbf{p})||_\infty > \beta | A, B, C\right)}{\mathbb{P}\left(||\mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) - F_{\tilde{X}}^{-1}(\mathbf{p})||_\infty \leq \gamma | A, B, C\right)} \leq \frac{(b-a)^m}{(2\gamma)^m/m!} e^{-\frac{\epsilon}{4}\left(\frac{(\beta-\alpha)\pi_{\min}}{2} - 4(m+1)\gamma\pi_{\max}\right)n}$$

and by fixing $\gamma = \frac{\beta\pi_{\min}}{16(m+1)\pi_{\max}+2\pi_{\min}}$ we end up with

$$\frac{\mathbb{P}\left(||\mathcal{E}^{(2/\epsilon)}_{u_{\mathrm{JE}}}(\tilde{\mathbf{X}}) - F^{-1}_{\tilde{X}}(\mathbf{p})||_\infty > \beta | A, B, C\right)}{\mathbb{P}\left(||\mathcal{E}^{(2/\epsilon)}_{u_{\mathrm{JE}}}(\tilde{\mathbf{X}}) - F^{-1}_{\tilde{X}}(\mathbf{p})||_\infty \leq \beta | A, B, C\right)} \leq \frac{2^m (b-a)^m m!}{\beta^m} \left(\frac{4(m+1)\pi_{\max} + \pi_{\min}/2}{\pi_{\min}}\right)^m e^{-\frac{\epsilon\beta\pi_{\min}}{16}n} .$$

We can use Lemma 1, Lemma 2 and union bounds to obtain the following:

$$\begin{aligned}
\mathbb{P}\left(||\mathcal{E}^{(2/\epsilon)}_{u_{\mathrm{JE}}}(\tilde{\mathbf{X}}) - F^{-1}_{\tilde{X}}(\mathbf{p})||_\infty > \beta\right) &\leq \mathbb{P}\left(||\mathcal{E}^{(2/\epsilon)}_{u_{\mathrm{JE}}}(\tilde{\mathbf{X}}) - F^{-1}_{\tilde{X}}(\mathbf{p})||_\infty > \beta | A, B, C\right) + \mathbb{P}(A^c) + \mathbb{P}(B^c) + \mathbb{P}(C^c) \\
&\leq \frac{2^m (b-a)^m m!}{\beta^m} \left(\frac{4(m+1)\pi_{\max} + \pi_{\min}/2}{\pi_{\min}}\right)^m e^{-\frac{\epsilon\beta\pi_{\min}}{16}n} \\
&+ \sum_{i=1}^m e^{-n\left(\frac{\beta^2\pi^4_{\min}}{2(1-p_i)(16^2(m+1)^2\pi^2_{\max}+4\pi^2_{\min})}\right)} + \sum_{i=1}^m e^{-n\left(\frac{\beta^2\pi^4_{\min}}{2p_i(16^2(m+1)^2\pi^2_{\max}+4\pi^2_{\min})}\right)} \\
&+ me^{-n\frac{\beta\pi_{\min}\pi_{\max}}{24(m+1)\pi_{\max}+3\pi_{\min}}} + 2me^{-n\frac{\pi_{\min}}{8}\left(\beta - \frac{\beta\pi_{\min}}{16(m+1)\pi_{\max}+2\pi_{\min}}\right)} .
\end{aligned}$$

D.6 Proof of Theorem 3

We tune the noise w to have density $d\mathbb{P}_w(w) = \frac{\mathbb{1}_{[-\delta/2,\delta/2]}(w)}{\delta}dw$. Under the hypothesis, F^{-1}_X has a finite number of discontinuity points. We can apply Proposition 3 with $t = \delta/2$ and $f(t) = 0$ to get that for Lebesgue-almost-any $\mathbf{p}$,

$$\|F^{-1}_{\tilde{X}}(\mathbf{p}) - F^{-1}_X(\mathbf{p})\|_\infty \leq \delta/2 .$$

In order to conclude, we can describe the density $\pi_{\tilde{X}}$ of the noisy random variable. It is piecewise continuous on $[a,b]$, $\pi_{\tilde{X}} > 0$ on $[a,b] \setminus \mathcal{O}'$ where $\mathcal{O}'$ is a finite union of intervals and $\pi_{\tilde{X}} = 0$ on $\mathcal{O}'$. Consequently, there only are a finite number of p's in $(0,1)$ such that it is not possible to find a $\beta > 0$ such that $\pi_{\tilde{X}} > 0$ on $[F^{-1}_{\tilde{X}}(p) - \beta, F^{-1}_{\tilde{X}}(p) + \beta]$ and where $\pi_{\tilde{X}}$ is continuous on that interval. Any $\mathbf{p}$ that has no such p as any of its components qualifies and we can apply Theorem 2 to get that

$$\|\mathbf{q} - F^{-1}_{\tilde{X}}(\mathbf{p})\|_\infty \leq \delta/2$$

with high probability. We get the result by the triangle inequality.