

Private Quantiles Estimation in the Presence of Atoms

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

▶ **To cite this version:**

Clément Lalanne, Clément Gastaud, Nicolas Grislain, Aurélien Garivier, Rémi Gribonval. Private Quantiles Estimation in the Presence of Atoms. Information and Inference, 2023, 10.1093/imaiai/iaad030 . hal-03572701v2

HAL Id: hal-03572701
https://hal.science/hal-03572701v2

Submitted on 8 Feb 2023

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.



Distributed under a Creative Commons Attribution 4.0 International License

Private Quantiles Estimation in the Presence of Atoms

Clément Lalanne
LIP, Univ Lyon, EnsL, UCBL, CNRS, Inria, LYON Cedex 07 F-69342,France
clement.lalanne@ens-lyon.fr

Clément Gastaud
Sarus Technologies SAS, 128 rue la Boétie, 75008 Paris, France

Nicolas Grislain
Sarus Technologies SAS, 128 rue la Boétie, 75008 Paris, France

Aurélien Garivier
UMPA UMR 5669, Univ. Lyon, ENS de Lyon, 46 allée d'Italie,
Lyon cedex 07 F-69364, France

Rémi Gribonval
LIP, Univ Lyon, EnsL, UCBL, CNRS, Inria, LYON Cedex 07 F-69342,France

February 8, 2023

Abstract

We consider the differentially private estimation of multiple quantiles (MQ) of a distribution from a dataset, a key building block in modern data analysis. We apply the recent non-smoothed Inverse Sensitivity (IS) mechanism to this specific problem. We establish that the resulting method is closely related to the recently published ad hoc algorithm *JointExp*. In particular, they share the same computational complexity and a similar efficiency. We prove the statistical consistency of these two algorithms for continuous distributions. Furthermore, we demonstrate both theoretically and empirically that this method suffers from an important

lack of performance in the case of peaked distributions, which can degrade up to a potentially catastrophic impact in the presence of atoms. Its smoothed version (i.e. by applying a max kernel to its output density) would solve this problem, but remains an open challenge to implement. As a proxy, 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.

1 Introduction

As more and more data is collected on individuals and data science techniques become more powerful, threats to privacy have multiplied and serious concerns have emerged [31, 6, 21, 13, 23, 27, 32, 37, 41, 38]. Against this background, *differential privacy* (DP) [19] has become the gold standard in privacy protection. By introducing randomness calibrated to the *sensitivity* of a query [18], 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 [18, 25, 15, 14, 1]. 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 [2], Google [20], Apple [39] and Microsoft [12] 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 [11], where variables are binned using quantiles before they are considered for a split. Besides, in many recent synthetic data models, continuous features are binned to reduce the output space's dimension (see [42, 29]).

Given a real random variable X of probability distribution $\mathbb{P}_X$, the cumulative distribution function (CDF) of X (or $\mathbb{P}_X$), noted F_X (or $F_{\mathbb{P}_X}$) is classically defined as $F_X(t) = \mathbb{P}_X(X \leq t), \forall t \in \mathbb{R}$. Its pseudo-inverse, the quantile function, F_X^{-1} (or $F_{\mathbb{P}_X}^{-1}$) is defined as

$$F_X^{-1}(p) = \inf \{t \in \mathbb{R} | F_X(t) \geq p\}, \quad \forall p \in [0, 1],$$

with the convention $\inf \emptyset = +\infty$. The quantity $F_X^{-1}(p)$ is the *quantile* of order p of the distribution of X. Furthermore, when given a dataset (a collection of real numbers), the

empirical quantile of order p of this dataset is the quantile of order p of its empirical distribution.

When privacy is not an issue, it is well known that the empirical quantiles of a dataset of i.i.d. random variables are good estimators of the quantiles of the underlying distribution [40]. There has been a lot of recent research on how to privately estimate multiple empirical quantiles from a dataset with a privacy overhead as small as possible. This article builds on this framework by considering private estimates of the empirical quantiles as estimators of the quantiles of the true distribution.

Among several ways to privately estimate empirical quantiles, a naive one is to add Laplace noise to non-private quantile estimates [18]. It is straightforward and easy to compute, but the amount of added noise is based on a pessimistic scenario that cannot materialize simultaneously for all quantiles. To reduce the variance of the estimates, a variant that uses so-called *smoothed sensitivity* instead of the worst case scenario was introduced [33]. It has the drawback, however, of using approximate differential privacy [16] 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 standard exponential mechanism for differential privacy [30] that is called *ExponentialQuantile* [35]. It is cheap to compute and the recent theory of *Inverse Sensitivity* [5] shows that a conceptually simple smoothing operation allows levels of privacy that are quadratically better than those guaranteed for the smoothed sensitivity approaches while sticking to pure differential privacy, achieving near-instance optimality (i.e. with a risk comparable to the local minimax risk when the set of hypotheses is the set of neighboring datasets). As a result, it has been adopted by the main DP-ready software libraries [3, 24] 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 [22] has presented a new algorithm, *JointExp*, that was the first to exploit the non-decreasing constraint of quantiles at the core of its sampling procedure and became the empirical state of the art for a small period. Even more recent work [26] cleverly exploits the structure of quantiles by computing them recursively / hierarchically on disjoint subsets of the dataset, which reduces the privacy overhead in composition. It is the current empirical state of the art for estimating many quantiles.

1.1 Contributions and organization of the paper

In this paper, we study the use of the Inverse Sensitivity mechanism [5, 4] in order to estimate multiple empirical quantiles from a dataset, and their statistical properties as estimators of the quantiles of the underlying distribution.

In Section 2 we recall the needed background on differential privacy, Inverse Sensitivity, and JointExp. Section 3 is devoted to characterizing the Inverse Sensitivity mechanism for private multiquantile estimation. In particular, our first contribution is to obtain the precise expression of the utility function of the Inverse Sensitivity mechanism for an arbitrary number of empirical quantiles, which was previously known for a single quantile only. In particular, this expression is quite similar to the utility function of JointExp [22], which was proposed as an ad-hoc algorithm based on a heuristic. We draw the explicit link between the two algorithms. Due to their similarities, the two algorithms are used interchangeably in most of the article.

It was noted by the authors of JointExp [22] that their algorithm struggles when the gaps between the data points are too small, but without more details. Our next contribution is to statistically quantify this empirical phenomenon by proving that JointExp/IS is in fact inconsistent on *atomic distributions*. This result is presented in Section 4.

Section 5 serves two purposes : we prove the consistency of JointExp/IS on *continuous distributions* (i.e. with a continuous density w.r.t. Lebesgue measure). This is the first consistency result of JointExp and is thus a big step towards the understanding of this algorithm. Furthermore, we propose a new heuristic, tractable smoothing technique based on *jittering* for JointExp/IS that vastly improves their utility on peaked distributions without noticeable repercussions on non-degenerate distributions. In particular, this estimator is consistent on atomic distributions. Our technique differs from the general smoothing trick for *inverse sensitivity* based mechanisms introduced by [5, 4], a conceptual trick consisting in taking a maximum convolution of the density over the output space, in the sense that our technique is a concrete mechanism to smooth the data distribution. As such, our technique results in a much better computational complexity making it a viable solution in high dimension. While similar techniques have already been used in order to fix ill-posed problems (see [34, 28, 10]) when dealing with peaked distributions, the motivations behind the addition of jitter here are fairly different : it fixes an over-penalization of the exponential mechanism when dealing with concentrated data.

Finally, Section 6 gathers the numerical experiments demonstrating the behavior of the proposed algorithms.

2 Background

Considering datasets of the form $\mathbf{X} = (X_1, \dots, X_n) \in \mathfrak{X}^n$, 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, $d(\mathbf{X}, \mathbf{Y}) = 0$ i.f.f. 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 $\mathbf{Y}$.

2.1 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 [16]) if for all pairs of datasets $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}^n$ and all measurable sets $S \subseteq O$,

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

Differential privacy offers strong privacy protections by bounding the efficiency of any test trying to distinguish two neighboring databases. A classical way to design ϵ-DP algorithms is the Exponential Mechanism [30]. 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

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

a classical result is that $\mathcal{E}_u^{(\alpha)}$ is ϵ-DP as soon as $\alpha \geq \frac{2\Delta u}{\epsilon}$ [30]. All the algorithms presented in this paper build on this mechanism.

2.2 The inverse sensitivity mechanism

When considering some deterministic function $\mathcal{Q} : \mathfrak{X}^n \to O$ as the target of privatization, a specific choice of utility function in the exponential mechanism, long known as folklore, was proved to have remarkable optimality properties under certain assumptions [5]. 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, which means that the probability of an output decreases the more points have to be modified in the dataset for it to be a viable deterministic output.

2.3 JointExp

Specifying 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 [17, 8]. 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 the brevity of notation, vectors are interpreted when needed as the set containing their components.

For reasons that will become clear later, we take the time to redefine the JointExp [22] (also called ExponentialQuantile when $m = 1$) mechanism. It corresponds to a specific instantiation of the exponential mechanism $\mathcal{E}^{(2/\epsilon)}_{u_{\mathrm{JE}}}(\mathbf{X})$ 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 ($n(p_i - p_{i-1})$).

3 JointExp meets Inverse Sensitivity

At first glance, there is no connection between the theory of the Inverse Sensitivity and JointExp. The first one is born from the need to build a general mechanism that is endowed with optimality properties [5, 4] for a broad class of problems, while the second comes from the idea that good empirical quantiles should separate the data points proportionally. In the case of the estimation of a single quantile (i.e. $m = 1$), it was observed [5] that the two algorithms are similar. Our contribution in this section is to prove that, up to minor differences, this remains true with an arbitrary number of quantiles. For this, we provide the precise expression of the inverse sensitivity function for the multiquantile problem.

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,*

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

with

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

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

Corollary 2. *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* $|\tilde{u}_{IS}(\mathbf{X},\mathbf{q}) - u_{JE}(\mathbf{X},\mathbf{q})| \leq 2(m+1)$ *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 6. As a consequence, we will mainly focus on JointExp for the rest of this article, all the results being applicable to IS as well (with some minor tweaks).*

4 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 [22], we prove in this section that it can also completely fail on some distributions (see Proposition 3). As illustrated in Section 6, 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"* $\big([X_{i_1}, X_{i_1+1}) \times \ldots \times \cdots \times [X_{i_m}, X_{i_m+1})\big) \cap [a,b]^m \nearrow$ for each $\mathbf{i} = (i_1, \ldots, i_m) \in O'$ where $O' = \{\mathbf{i} \in \{0,\ldots,n\}^m, i_1 \leq \cdots \leq i_m\}$.

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 3. *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

$$\mathbb{E}_{\mathbf{X},\mathcal{E}_{u_{JE}}^{(2/\epsilon)}}\left(\|\mathcal{E}_{u_{JE}}^{(2/\epsilon)}(\mathbf{X}) - F_X^{-1}(\mathbf{p})\|_\infty\right) = \Omega_n(1)\ , \tag{4.1}$$

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

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

This result shows that for certain data distributions with isolated atoms, JointExp is not consistent, even asymptotically, on many instances of the estimation problem (i.e. not on unrealistic corner cases). This behavior is all the more counterintuitive 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.*

When considering estimation on real-world distributions, many real-life datasets show *accumulation points* and can be modeled as 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 category 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.

5 Heuristic smoothing, with guarantees

The type of failure of JointExp highlighted in Section 4 may seem surprising given a) the strong connection between JointExp and the Inverse Sensitivity established in Section 3; and b) existing performance guarantees for *smoothed* Inverse Sensitivity mechanisms [5, 4]. 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. Note that the exposed problems with atomic distribution also occur for highly concentrated continuous distributions. Hence simpler and more naive solutions such as multi-indices do not fix them.

Another possible solution would be to discretize the output space. However, the resulting algorithm would have a complexity of $O(f(m,n,\delta) + 1/\delta^m)$ where δ is the precision of the discretization and f is some function. Since this is exponential in the number of quantiles, it suffers from the curse of dimensionality, and we argue that jittering is a better alternative.

5.1 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 $\mathbf{i} = (i_1, \ldots, i_m) \in O'$, it fails when the blocks that have a great utility (i.e. the ones leading to interesting quantile candidates) have a volume that is too small. By adding noise to the data points, we ensure a minimal volume for the blocks, and in particular for the interesting regions, while only shifting the empirical quantiles of the dataset by a small amount.

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

$$\tilde{\mathbf{X}} = (X_1 + w_1, \ldots, X_n + w_n) \; . \tag{5.1}$$

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

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, this 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,

$$\forall t \in \mathbb{R}, \quad \pi_{\tilde{X}}(t) = \int \pi_w(t-x)\mathbb{P}_X(dx) \; . \tag{5.2}$$

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 Appendix C.3).

Proposition 4. *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 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.

5.2 Consistency of HSJointExp on constant data

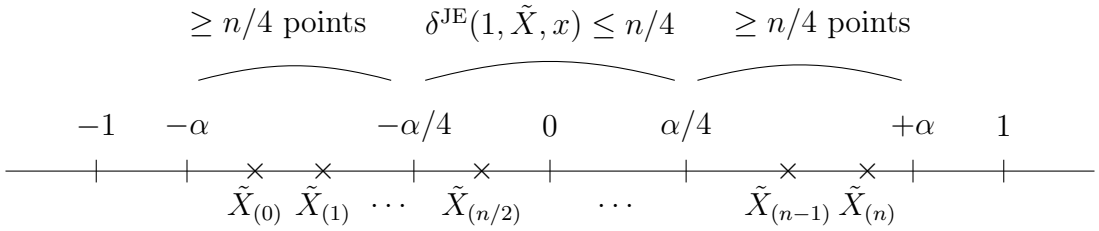


Figure 1: $\delta^{\mathrm{JE}}(1, \tilde{X}, x)$ is bounded by $n/4$ for $-\alpha/4 \le x \le \alpha/4$ on the event G.

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. Note that the classical way of analyzing exponential mechanisms is to use the utility bounds found in [30]. However, here we do not have the required level of control on the normalization factor. We hence go for a more direct way of controlling the output distribution. 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.

5.3 General Consistency of HSJointExp

For multiquantile estimation, JoitExp/IS is not endowed with any satisfying statistical utility bounds. We start by proving their consistency in the favorable case of continuous distributions (see Theorem 5). We then leverage this result in order to prove the consistency of HSJointExp on a larger class of distributions. Indeed, 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.

Theorem 5. *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 Section 5.2 is in Appendix C.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 Appendix C.4.

Proposition 6. *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 5 assuming that we fall into its hypothesis. By adding some uniform noise in $[-\alpha, \alpha]$, we then obtain the following result:

Theorem 7. *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's 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 Appendix C.6. Theorem 7 states in particular that many distributions that satisfy the hypothesis of Proposition 3 and on which JointExp is not consistent also satisfy the hypothesis of Theorem 7 and HSJointExp can thus achieve arbitrary levels of precision on them (provided n is large enough).

As highlighted by Section 5.2, working on much stricter distribution classes can lead to numerically tractable optimal levels of noise.

5.4 Privacy Amplification of HSJointExp

A final property that we would like to explore is the possible amplification of privacy of HSJointExp. Indeed, adding Laplace or Gaussian noise to bounded quantities is a common way to make them private [18]. Furthermore, it is well known that some preprocessing steps (prior to the application of an already private mechanism) increase the provable privacy of the overall mechanism. This is for instance the case with subsampling [7]. Consequently, one would think that adding noise to the data does not only preserve the privacy guarantees of the original mechanism (as stated by Proposition 4), but has reasonable chances to make it more private. In order to evaluate the actual privacy of our mechanism, we investigate its privacy loss:

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

for $\mathbf{Y} \sim \mathbf{X}$ and $\mathbf{q} \in O$ where $d\mathbb{P}/d\mathbf{q}\left(\mathcal{E}_{u_{\mathrm{JE}}}^{(2/\epsilon)}(\tilde{\mathbf{X}}) = \mathbf{q}\right)$ 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 for future work, but we conduct a numerical analysis in Section 6.

6 Numerical Results

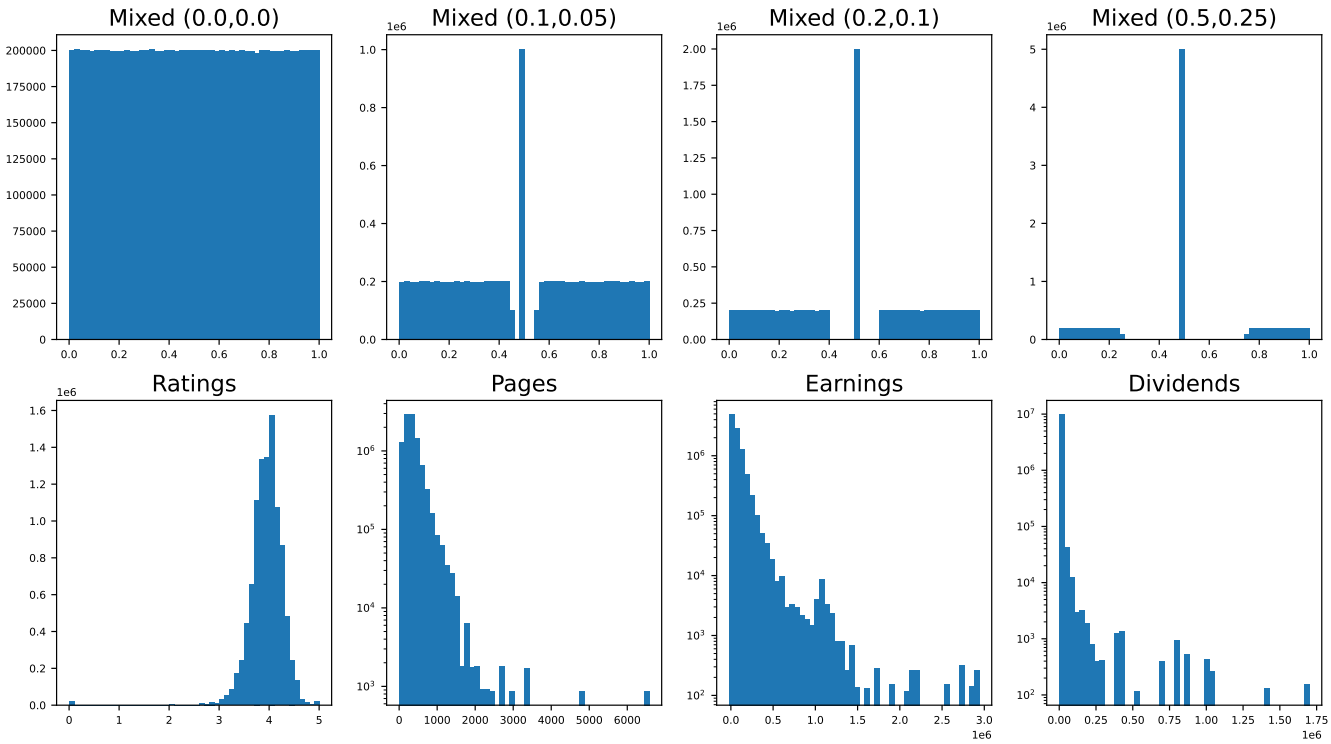
This section presents the behaviors of JointExp, the Inverse Sensitivity mechanism and HSJointExp on synthetic and on on real-world distributions. In particular, 6.1 is devoted to the presentation of the distributions of interest. Section 6.2 numerically studies the performance of the algorithms on the above-mentioned distributions. And finally, Section 6.3 looks at the possible numerical gain of privacy resulting of the noise addition.

6.1 Distributions

We claimed that HSJointExp has a huge advantage over regular JointExp in the case of distributions with isolated atoms. In order to test it numerically, we propose to do so with synthetic data in the first place. Indeed, in allows us to tune various interesting quantities. For real world distributions, it is harder to identify which ones satisfy the condition of having isolated atoms. We propose to evaluate the performance of the algorithms by identifying a real-wold distribution with the empirical distribution of a real-world dataset. The concentration of this dataset (i.e. how peaked its histogram is) is then the decisive criterion: The more concentrated it is, the more suboptimal JointExp/IS is expected to be compared to the smoothed variants.

Mixed distributions (synthetic). For $p \in [0,1]$ and $\delta \in [0,1/2]$, we define the *Mixed* distribution of parameters (p,δ) as the distribution with support in $[0,1/2-\delta] \cup \{1/2\} \cup [1/2+\delta,1]$ such that if a random variable X follows this distribution, we have $\mathbb{P}(X = 1/2) = p$, $\mathbb{P}(X \in [0,1/2-\delta]) = \mathbb{P}(X \in [1/2+\delta,1])$, and conditionally to the event $(X \in [0,1/2-\delta])$ or to the event $(X \in [1/2+\delta,1])$, X is uniform. In particular, the mixed distribution of parameters $(0,0)$ is the uniform distribution on $[0,1]$. In order to better visualize such distributions, sampled histograms are represented in Figure 2. The parameters ϵ and δ allow tuning, respectively, the probability of the atom and its isolation. The bigger they are, the more HSJointExp is expected to outperform the non-smoothed variants.

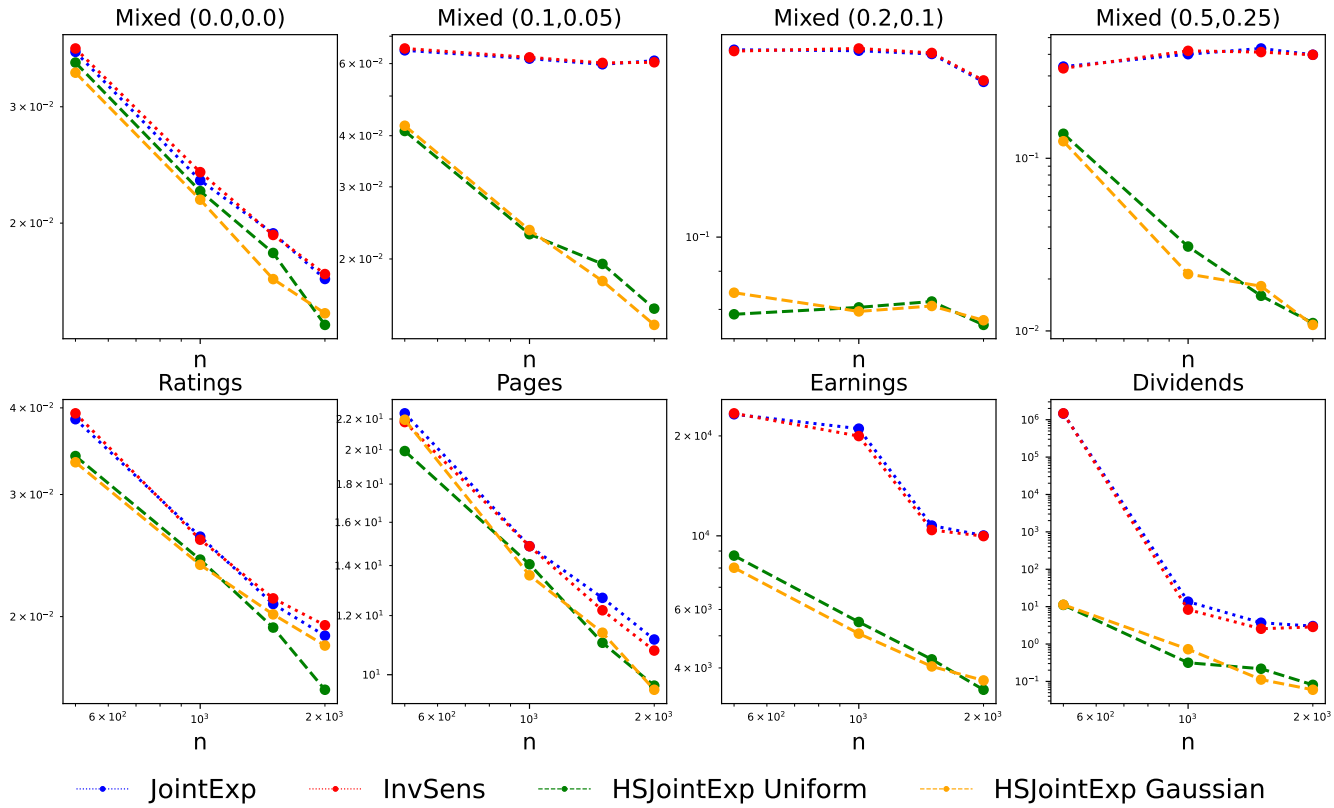
Pages and Ratings (real-world). The distributions that we call *Pages* and *Ratings* correspond to the empirical distributions of a collection of ratings and of number of pages of books from the Goodreads-Books dataset [36]. Gillenwater et al. [22] used the same datasets as numerical evidences of the performance of JointExp for estimating empirical quantiles. Again, sampled histograms are represented in Figure 2. The distributions look relatively smooth (i.e. not too peaked and with a relatively small support), and as a result, we can expect the gap between JointExp/IS and HSJointExp to be negligible.



Histograms representing $n = 10^7$ data points sampled from the original distributions and binned in 50 bins. Note that for Pages, Earnings and Dividends, the vertical axis is in $\log_{10}$-scale.

Figure 2: Distributions used for experiments

Earnings and Dividends (real-world). The distributions that we call *Earnings* and *Dividends* correspond respectively to the personal incomes and personal incomes from dividends categories of the US 2021 Census [9]. Again, sampled histograms are represented in Figure 2. We can notice that contrary to the previous two real-world distributions, these two are much more concentrated. For Earnings, the concentration is due to the existence of categories of extremely high revenues. As a consequence, the support of the distribution is necessary big, and the algorithms that seek for privately estimating the quantiles have little information about the localization of the data points. On the other hand, the vast majority of people declare revenues inferior to 500000 dollars, resulting in the high concentration of the distribution close to 0. For Dividends, the support is smaller, but since a big part of the population simply does



The vertical axis reads the error $\mathbb{E}\left(\|\hat{\mathbf{q}} - F^{-1}(\mathbf{p})\|_\infty\right)$ where $\mathbf{p} = \left(\frac{1}{m+1}, \ldots, \frac{m}{m+1}\right)$ for $m = 8$, $\epsilon = 1$, $\hat{\mathbf{q}}$ is the private estimator, and $\mathbb{E}$ is estimated by Monte-Carlo averaging over 50 runs. For HSJointExp Uniform and Gaussian, the optimal noise level on a discretization of $\log_{10}$-resolution 2 of $[10^{-10}, 10^4]$ is selected. Note that both axis are in $\log_{10}$-scale.

Figure 3: Error of the estimators as a function of n

not have any revenues from dividends, the distribution shows an accumulation point at 0. With both distributions, we expect the smoothing operation to vastly improve the performance of JointExp/IS.

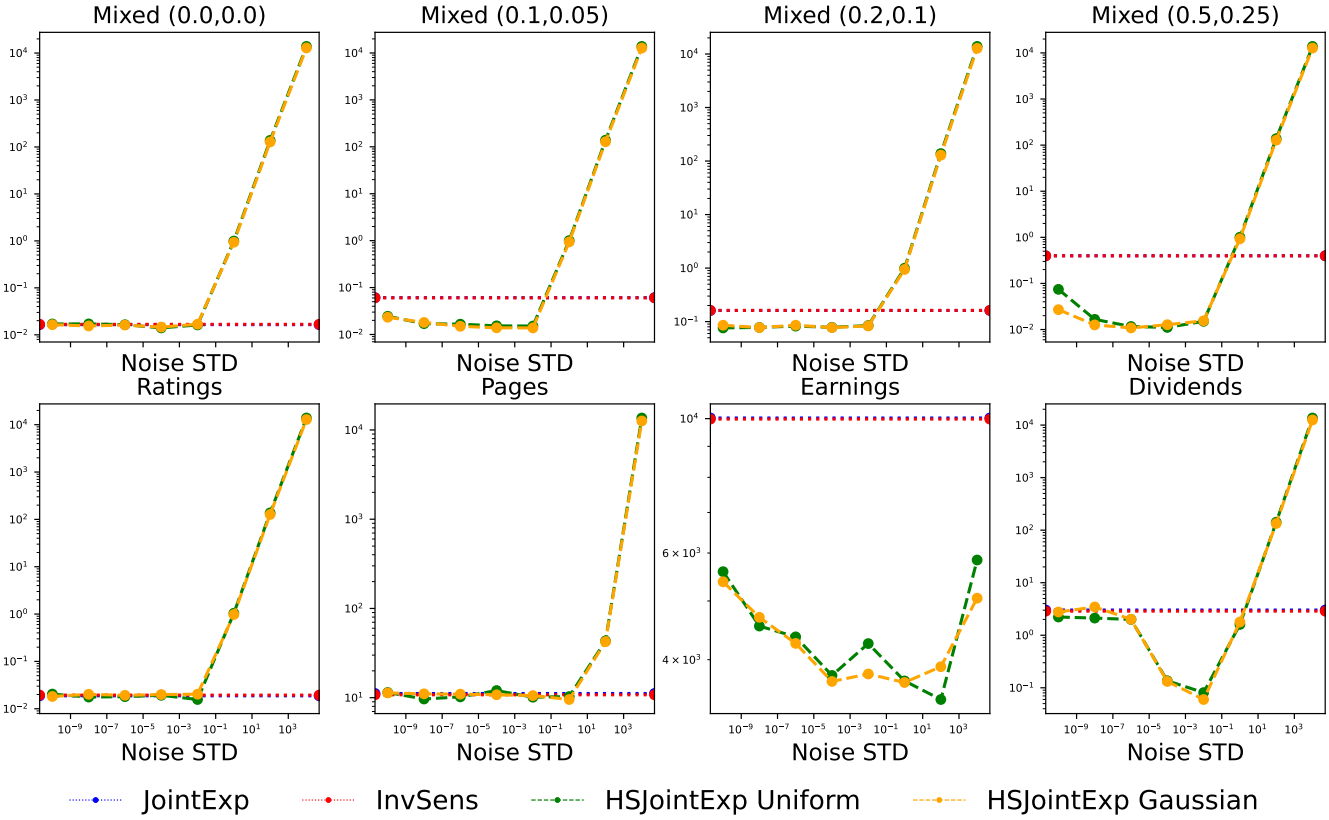
6.2 Numerical Performance

Figure 3 and Figure 4 Compare the performance of JointExp, the Inverse Sensitivity mechanism and two variants of HSJointExp with uniform and Gaussian noise structure respectively on the distributions presented in Figure 2.

Complements on HSJointExp Uniform and Gaussian. The mechanism that we call *HSJointExp Uniform* is the application of JointExp post addition of centered uniform noise. If $[a, b]$ was our estimate of the support of the distribution, we apply JointExp on $[a - \sigma\sqrt{3}, b + \sigma\sqrt{3}]$ where σ is the standard deviation of the noise. In *HSJointExp Gaussian*, the centered uniform noise is replaced by centered Gaussian noise. The support of the resulting distribution is now infinite, and the projection step is therefore mandatory. We chose to project the data points in $[a - 5\sigma, b + 5\sigma]$ where σ is the standard deviation of the noise in order to make sure that most of the points will remain untouched by the projection step.

Analyzing the results of Figure 3. The first important fact to notice is the similar performance of JointExp and the Inverse Sensitivity mechanism, confirming the theoretical results of Section 3. The second is the similar performance of HSJointExp Uniform and Gaussian, showing that the structure of the noise, given that it is regular enough, is not of critical importance. Finally, and probably the most important, we can compare the performance of JointExp/IS and of HSJointExp. On Mixed$(0, 0)$ (i.e. the uniform distribution on $[0, 1]$), Ratings and Pages, the two algorithms perform identically. This is what we expected given the smoothness of the distributions. On more concentrated distributions like Earnings and Dividends on the other hand, we see that HSJointExp vastly improves the performance of JointExp, sometimes by multiple orders of magnitude. Finally, Mixed$(0.1, 0.05)$, Mixed$(0.2, 0.1)$ and Mixed$(0.5, 0.25)$ demonstrate that the more isolated and probable the atoms of the distribution are, the more suboptimal JointExp is compared to the smoothed variants.

Analyzing the results of Figure 4. Figure 4 shows the same results as Figure 3 but with an emphasis on the dependence on the noise level. For instance, we can see that when the smoothing operation allows for better performance, it is often the case for a large range of smoothing levels. Finally, we can numerically observe two limit behaviors that are quite intuitive : When the noise level tends to 0, HSJointExp performs as JointExp. Indeed, in this case, the smoothing trick has almost no effect on the distribution. When the noise level tends to $+\infty$ on the other hand, the performance of HSJointExp is terrible. This is also quite intuitive, since the smoothed distribution has lost almost all correlation with the original distribution. For all these reasons, we recommend tuning the noise as in the extreme case of the Dirac (see



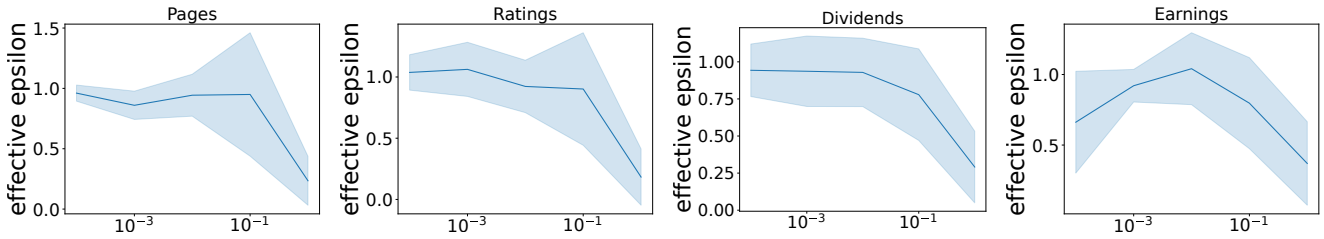
The vertical axis reads the error $\mathbb{E}\left(\|\hat{\mathbf{q}} - F^{-1}(\mathbf{p})\|_\infty\right)$ where $n = 2000$, $\mathbf{p} = \left(\frac{1}{m+1}, \ldots, \frac{m}{m+1}\right)$ for $m = 8$, $\epsilon = 1$, $\hat{\mathbf{q}}$ is the private estimator, and $\mathbb{E}$ is estimated by Monte-Carlo averaging over 50 runs. For HSJointExp Uniform and Gaussian, the optimal noise level on a discretization of $\log_{10}$-resolution 2 of $[10^{-10}, 10^4]$ is selected. Note that both axis are in $\log_{10}$-scale. The horizontal axis reads the standard deviation of the smoothing noise. JointExp and the Inverse Sensitivity mechanism are represented by horizontal bars, since they do net depend on the noise level.

Figure 4: Dependence on the smoothing level

Section 5.2) since this value is small enough to not fall in the regime where the performance are degraded by the smoothing, but it still greatly improves the performance on degenerated distributions.

6.3 Privacy Amplification

In Figure 5 we numerically estimate ϵ_{eff} in the following setup: For each of 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 horizontal axis represents the standard deviation of the noise divided by the length of the support of the distribution. $\epsilon = 1$.

Figure 5: Evolution of ϵ_{eff} for the median estimation

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 4 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.

7 Conclusion

We highlight the connections between the general inverse sensitivity mechanism applied to the private estimation of multiple empirical quantiles and the recently published ad-hoc algorithm JointExp [22]. We prove the consistency of this algorithm when used as a statistical estimator of the statistical quantiles of the underlying distribution for smooth enough distributions. These results are key to the understanding of JointExp, which wasn't endowed with any theoretical utility results before despite a excellent numerical behavior. Furthermore, we demonstrate that isolated atoms in the distribution cause JointExp to be inconsistent, and propose a numerically tractable fix that solves this issue. The numerical experiments on both real-world and synthetic

distributions backup the theoretical claims of this article, and support our suggestion to use our variant instead of JointExp in practice.

Very recently, a new approach was proposed by Kaplan et al. in [26]: they present an algorithm that experimentally beats JointExp when the number of empirical quantiles of a dataset is high. Determining under what conditions and in what circumstances this empirical gap reflects in terms of statistical utility is left as future work.

Acknowledgments

Aurélien Garivier acknowledges the support of the Project IDEXLYON of the University of Lyon, in the framework of the Programme Investissements d'Avenir (ANR-16-IDEX-0005), and Chaire SeqALO (ANR-20-CHIA-0020-01). This project was supported in part by the AllegroAssai ANR project ANR-19-CHIA-0009.

References

[1] 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. 2

[2] 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. 2

[3] Joshua et. al Allen. Smartnoise core differential privacy library. `https://github.com/opendp/smartnoise-core`. 3

[4] 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. 3, 4, 7, 9, 27

[5] Hilal Asi and John C Duchi. Near instance-optimality in differential privacy. *arXiv preprint arXiv:2005.10630*, 2020. 3, 4, 5, 7, 9, 27

[6] 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. 2

[7] Borja Balle, Gilles Barthe, and Marco Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. *CoRR*, abs/1807.01647, 2018. 14

[8] Victor-Emmanuel Brunel and Marco Avella-Medina. Propose, test, release: Differentially private estimation with high probability. *arXiv preprint arXiv:2002.08774*, 2020. 6

[9] United States Census Bureau. 2021 annual social and economic supplements. `https://www.census.gov/data/datasets/2021/demo/cps/cps-asec-2021.html`. 16

[10] Jien Chen and Nicole A Lazar. Quantile estimation for discrete data via empirical likelihood. *Journal of Nonparametric Statistics*, 22(2):237–255, 2010. 4

[11] 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. 2

[12] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. *arXiv preprint arXiv:1712.01524*, 2017. 2

[13] 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. 2

[14] 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. 2

[15] Jinshuo Dong, Aaron Roth, and Weijie J Su. Gaussian differential privacy. *arXiv preprint arXiv:1905.02383*, 2019. 2

[16] 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. 3, 5

[17] 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. 6

[18] 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. 2, 3, 14

[19] 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. 2

[20] 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. 2

[21] 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. 2

[22] Jennifer Gillenwater, Matthew Joseph, and Alex Kulesza. Differentially private quantiles. *In International conference on machine learning. PMLR*, 2021. 3, 4, 6, 8, 15, 20, 26

[23] 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. 2

[24] IBM. Smartnoise core differential privacy library. `https://github.com/IBM/differential-privacy-library`. 3

[25] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. In *International conference on machine learning*, pages 1376–1385. PMLR, 2015. 2

[26] Haim Kaplan, Shachar Schnapp, and Uri Stemmer. Differentially private approximate quantiles. In *International Conference on Machine Learning*, pages 10751–10761. PMLR, 2022. 3, 21

[27] 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. 2

[28] José A F Machado and JMC Santos Silva. Quantiles for counts. *Journal of the American Statistical Association*, 100(472):1226–1237, 2005. 4

[29] 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. 2

[30] 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. 3, 5, 11

[31] Arvind Narayanan and Vitaly Shmatikov. How to break anonymity of the netflix prize dataset. *arXiv preprint cs/0610105*, 2006. 2

[32] 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. 2

[33] 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. 3

[34] Art B Owen. *Empirical likelihood*. Chapman and Hall/CRC, 2001. 4

[35] 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. 3

[36] Soumik. Goodreads-books dataset. `https://www.kaggle.com/jealousleopard/goodreadsbooks`. 15

[37] Latanya Sweeney. Simple demographics often identify people uniquely. *Health (San Francisco)*, 671(2000):1–34, 2000. 2

[38] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002. 2

[39] 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. 2

[40] Aad W Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000. 3

[41] Isabel Wagner and David Eckhoff. Technical privacy metrics: a systematic survey. *ACM Computing Surveys (CSUR)*, 51(3):1–38, 2018. 2

[42] 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. 2

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 $\mathcal{E}^{(2/\epsilon)}_{\tilde{u}_{\mathrm{IS}}}(\mathbf{X})$ 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

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

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

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

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

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

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

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. [22] present an algorithm that allows to sample from any distribution that factorizes in an analog form of (A.1) 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 [22] 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 [5]

B.1 General principle

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

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

It is easy to see that this new utility function has sensitivity $\Delta u_{\text{IS}}^{\rho} = 1$. 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 [4] assuming that $\text{diam}_2(\mathcal{Q}(\mathfrak{X}^n)) \leq D$: for $1 \leq k \leq n$

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

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

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

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 $\mathbf{X} \in \mathfrak{X}^n$,

$$W_{\mathcal{Q}}(\mathbf{X}, k) \supseteq c.\omega_{\mathcal{Q}}(\mathbf{X}, k).\mathbb{B}_2^m \tag{B.2}$$

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 [4]:

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

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 5).

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 $\mathbf{X} \in \mathcal{X}^n$ 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:

$$W_{\mathcal{Q}}(\mathbf{X}, k) + \mathcal{Q}(\mathbf{X}) \subseteq \bigcup_{1 \leq i_1 \leq \cdots \leq i_k \leq m} \bigotimes_{j=1}^{m} \mathcal{I}(\mathbf{X}, j, \mathbf{i}, k) \ , \tag{B.3}$$

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

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

Using (B.3), the modulus of continuity is bounded as

$$\omega_{\mathcal{Q}}(\mathbf{X}, k) \leq \sqrt{\sum_{i=1}^{m} \left(X_{\lceil np_i \rceil + k} - X_{\lceil np_i \rceil - k} \right)^2} \ . \tag{B.4}$$

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 (B.1)) along with the upper bound on the modulus of continuity (i.e., Equation (B.4)) gives that, with high probability

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

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., $\mathcal{E}^{(2/\epsilon)}_{u_{\mathrm{IS}}^{1/n^r}}(\mathbf{X})$) 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 (B.3), 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 (B.2) 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 Omitted proofs

C.1 Proof of Theorem 1

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

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

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 replaced by the associated quantile at an extra cost of 1. In the end, we find the desired result.

C.2 Proof of Proposition 3

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 referring to Lebesgue measure. Let us prove that any $\mathbf{p}$ with at least one component in A satisfies (4.1). 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 $\big([X_{i_1}, X_{i_1+1}) \times \ldots \times \cdots \times [X_{i_m}, X_{i_m+1})\big) \cap [a, b]^{m\nearrow}$ for $\mathbf{i} = (i_1, \dots, i_m) \in O'$. 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 $\|\mathbf{q} - F_X^{-1}(\mathbf{p})\|_\infty \geq \eta$. 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 $\|\mathbf{q} - F_X^{-1}(\mathbf{p})\|_\infty$ 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 $\mathrm{Conv}\{\eta, f(\eta, 1), \ldots, f(\eta, m)\}$ which is then bigger than $\min\{\eta, f(\eta, 1), \ldots, f(\eta, m)\}$ which is positive.

C.3 Proof of Proposition 4

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 $\mathbf{w} \in \mathbb{R}^n, \mathrm{proj}_{\mathfrak{X}^n}(\mathbf{X} + \mathbf{w}) \sim \mathrm{proj}_{\mathfrak{X}^n}(\mathbf{X}' + \sigma(\mathbf{w}))$ for a specific permutation of the components σ. For each measurable set $\mathcal{S} \subseteq O$ we get

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

which completes the proof.

C.4 Proof of Proposition 6

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 concluding with the desired result.

C.5 Proof of Theorem 5

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 $\gamma > 0$, if $n \geq \frac{2}{\gamma \pi_{min}}$,*

$$[F_{\tilde{X}}^{-1}(p) - \gamma, F_{\tilde{X}}^{-1}(p) + \gamma] \subset \mathcal{N} \implies \mathbb{P}\left(\left|F_{\tilde{X}}^{-1}(p) - \tilde{X}_{(\lceil np \rceil)}\right| > \gamma\right) \leq e^{-n\left(\frac{\gamma^2 \pi_{min}^2}{8(1-p)}\right)} + e^{-n\left(\frac{\gamma^2 \pi_{min}^2}{8p}\right)}$$

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 $\tilde{X}_{(\lceil np \rceil)} > F_{\tilde{X}}^{-1}(p) + \gamma$, then $N \geq n(1-p)$. So,

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

where line 3 is deduced from line 2 by a multiplicative Chernoff bounds. If we further impose that $n \geq \frac{2}{\gamma\pi_{\min}}$,

$$\begin{aligned}
&\mathbb{P}\left(\tilde{X}_{(\lceil np \rceil)} > F_{\tilde{X}}^{-1}(p) + \gamma\right) \leq e^{-\frac{n\eta}{4}\left(\frac{\gamma\pi_{\min}}{\eta}\right)^2 / \left(2 + \frac{\gamma\pi_{\min}}{\eta}\right)} \\
&\leq e^{-\frac{n}{4}\left(\frac{\gamma^2\pi_{\min}^2}{2(1-p) - \gamma\pi_{\min}}\right)} \leq e^{-n\left(\frac{\gamma^2\pi_{\min}^2}{8(1-p)}\right)}
\end{aligned}$$

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:

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

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

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

Then we can compute,

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

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

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

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

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

We can use Lemma 1, Lemma 2 and union bounds to obtain the following for n big enough:

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

C.6 Proof of Theorem 7

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_X^{-1} has a finite number of discontinuity points. We can apply Proposition 6 with $t = \delta/2$ and $f(t) = 0$ to get that for Lebesgue-almost-any $\mathbf{p}$,

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

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 $[F_{\tilde{X}}^{-1}(p) - \beta, F_{\tilde{X}}^{-1}(p) + \beta]$ 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 5 to get that

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

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