Optimal Data-Independent Noise for Differential Privacy

Jordi Soria-Comas and Josep Domingo-Ferrer

Universitat Rovira i Virgili
Department of Computer Engineering and Mathematics
UNESCO Chair in Data Privacy
Av. Països Catalans 26, E-43007 Tarragona, Catalonia
Tel.: +34 977558270
Fax: +34 977559710
E-mail {jordi.soria,josep.domingo}@urv.cat

Abstract

ε-Differential privacy is a property that seeks to characterize privacy in data sets. It is formulated as a query-response method, and computationally achieved by output perturbation. Several noise-addition methods to implement such output perturbation have been proposed in the literature. We focus on data-independent noise, that is, noise whose distribution is constant across data sets. Our goal is to find the optimal data-independent noise distribution to achieve ε-differential privacy. We propose a general optimality criterion based on the concentration of the probability mass of the noise distribution around zero, and we show that any noise optimal under this criterion must be optimal under any other sensible criterion. We also show that the Laplace distribution, commonly used for noise in ε-differential privacy, is not optimal, and we build the optimal data-independent noise distribution. We compare the Laplace and the optimal data-independent noise distributions. For univariate query functions, both introduce a similar level of distortion; for multivariate query functions, optimal data-independent noise offers responses with substantially better data quality.

Key words: Data privacy, Differential privacy, Noise addition, Privacy-preserving data mining, Statistical disclosure control

1 Introduction

ε-Differential privacy [6,5] is a statistical disclosure control methodology for queryable databases. A remarkable fact about ε-differential privacy is that,

unlike other methods, it is not based on the understanding that some specific output may be disclosive. Instead it seeks to limit the knowledge gain that any database user may obtain from a response.

The initial formulation of differential privacy only in a query-response setting was justified by previous results [1,3,4] showing the imposibility of answering a large number of queries with a bounded error while preserving the utility of the data. This seemed to preclude using differential privacy for data set releases. In [2,9] it was shown that those previous results were overpessimistic, which opened the door to the generation of ε-differentially private data sets [13]. Nonetheless, the initial query-response formulation remains the basic use case for differential privacy, and the methods developed for such use case can also be leveraged to generate ε-differentially private data sets.

Computationally, ε-differential privacy is usually achieved by output perturbation; responses are computed on the real data and masked by adding a random noise. Other methods for attaining ε-differential privacy not based on directly adding noise to the real query response are, for instance, the exponential mechanism [14], and the sample and aggregate framework [16]. For a more complete overview of differential privacy and, in particular, of a variety of methods used to attain it, see [7,8,12].

Several methods to generate the required random noise have been proposed in the differential privacy literature. We classify them in two categories, according to whether the noise distribution takes the original data into account: data-independent noise and data-dependent noise. Methods based on adding data-independent noise conform the most basic approach. Laplace noise addition [6] belongs to this category. Methods based on adding data-dependent noise are more complex, but usually they lead to less distortion being introduced. Calibration to smooth sensitivity [16] belongs to the data-dependent noise category. In this paper we focus on the data-independent noise approach, which is the most frequently used one (and the one that was first proposed).

To maximize the utility of the results provided by ε-differential privacy, the magnitude of the random noise should be as small as possible. Some criticisms have appeared to the data utility that results from using Laplace noise addition as the mechanism to obtain ε-differential privacy [15,17]. The question of the optimality of Laplace noise addition arises: is it possible to achieve ε-differential privacy with substantially more data utility using other noise distributions?

Our goal is to determine the optimal distribution to achieve ε-differential privacy with data-independent random noise. We will limit our discussion to absolutely continuous random noise distributions, as they provide the greatest level of generality. Similar results can also be obtained for discrete random

noise; however, this type of noise is only applicable in very specific circumstances.

By using an optimal noise, the distortion required to achieve a certain level ε of differential privacy is minimized. This may lead to under-protection if the disclosure limitation offered by ε-differential privacy is measured by how much noise is added to the data (as in traditional noise addition for disclosure control, see [11]), rather than by the theoretical guarantee offered by differential privacy in terms of ε (see Definition 1 below). In what follows, we assume that a protection level ε is chosen such that the theoretical guarantee provides sufficient protection.

Before going into the details of the construction of the optimal data-independent random noise we briefly introduce some basic concepts about ε-differential privacy. The following formal definition of ε-differential privacy can be found in [5].

Definition 1 *A randomized function κ gives ε-differential privacy if, for all data sets D and D' differing in at most one row (that is, one record), and all $S \subset Range(\kappa)$ measurable, it holds that*

$$P(\kappa(D) \in S) \leq e^{\varepsilon} \times P(\kappa(D') \in S) \tag{1}$$

The interpretation of the above definition is as follows. Assume that we want to query the database with a function $f : \mathcal{D} \longrightarrow \mathbb{R}^d$ that maps each of the data sets to a value in $\mathbb{R}^d$. ε-Differential privacy returns a randomization κ_f of f such that the probability of obtaining a given response changes at most by a factor $\exp(\varepsilon)$ when adding or removing a record from the database.

The privacy guarantee provided by ε-differential privacy to an individual is that, no matter whether the record containing the individual's data is included in the data set, the responses returned for any query will be similar. Hence, the presence or absence of the individual's data are not easily noticed, which means privacy for the individual.

Definition 1 is stated in terms of data sets D and D' differing in at most one row. Data sets differing in one row, called neighbor data sets, can be obtained from one another in two ways: either by adding/removing one record (as assumed in [5]) or by modifying a single record (as assumed in [6]). Depending on the definition used, the magnitude of the required random noise may slightly change, but the methods used for noise calibration remain the same. For the sake of concreteness, in the sequel we will focus on addition and removal of records.

The randomization κ in Definition 1 can be seen as the addition of a random

noise, whose distribution may depend on the data set D, to the real value of the query function $f(D)$:

$$\kappa(D) = f(D) + (\kappa(D) - f(D)) = f(D) + Y(D)$$

If the distribution of the random noise depends on the actual data set D, we say that noise is data-dependent. If the random noise distribution is constant across data sets, we say that noise is data-independent. As mentioned above, we focus on data-independent noise.

Data-independent noise for ε-differential privacy is usually implemented as proposed by Dwork *et al.* in [6]. These authors proposed to generate noise using a Laplace distribution whose scale parameter depends on the maximum variation of the query function between neighbor data sets. This variation is known as the L_1-sensitivity of the function, and it is formally introduced next.

Definition 2 (L_1-sensitivity) *The L_1-sensitivity of a function $f : \mathcal{D} \to \mathbb{R}^d$ is defined as*

$$\Delta f = \sup_{D,D'} \|f(D) - f(D')\|_1 = \sup_{D,D'} \sum_{i=1}^{d} |f_i(D) - f_i(D')|$$

where f_i is the i-th component of f, for all D, D' such that one can be obtained from the other by adding or removing one record.

In order to reach ε-differential privacy, Laplace-distributed random noise with zero mean and $\Delta f/\varepsilon$ scale parameter is added to each component of f .

1.1 Contribution and plan of this paper

The randomized function κ that provides ε-differential privacy can be viewed as the addition of a random noise to the real value of the query function f. Hence, the quality of the resulting differentially-private data critically depends on the noise distribution. Taking this into account, the aim of this paper is to build the optimal data-independent noise distribution for ε-differential privacy.

Section 2 states the criteria that will be used in later sections to determine the optimal noise distribution. Section 3 elaborates further on the definition of ε-differential privacy using absolutely continuous (a.c.) noise distributions with the goal of characterizing the noise distribution in terms of its density function. Section 4 shows that the Laplace distribution is not the optimal a.c. noise distribution to achieve ε-differential privacy. Other distributions with the probability mass more concentrated towards zero exist. Section 5 is devoted to the construction of an optimal a.c. noise distribution to achieve ε-differential

privacy for the case of a query whose function has fixed L_1-sensitivity. To construct this optimal distribution we need to characterize the properties of the density functions that satisfy the ε-differential privacy definition. While Section 5 shows that the Laplace distribution is actually near-optimal for a single query, Section 6 illustrates that, for multiple queries or for a query with a multivariate response, it can be substantially far from optimality. Conclusions are summarized in Section 7.

2 Optimal random noise

To improve the utility of the outputs provided by an ε-differentially private access mechanism, the random noise must be adjusted to minimize the distortion to the real query result. When using Laplace noise, the scale parameter is set to $\Delta f/\varepsilon$ (see Section 1); this yields a noise distribution optimal within the class of Laplacian noises, because a smaller scale parameter would no longer satisfy ε-differential privacy. In Section 4 below, we will study whether the Laplace distribution itself is optimal within all possible noise distributions, an issue that has not been addressed in the literature. We devote the present section to a previous and more fundamental topic: the concept of optimality of a random noise distribution.

Deciding which among a pair of random noises, Y_1 and Y_2, leads to greater utility is a question that may depend on the users' preferences. The goal of this section is to come up with an optimality notion that is independent from the users' preferences: if Y_1 is better than Y_2 according to our criterion, any rational user must prefer Y_1 to Y_2. Later, in Section 5, we will determine the form of all optimal random noises that provide ε-differential privacy to a given query function.

Let Y_1 and Y_2 be two random noise distributions. If Y_1 can be constructed from Y_2 by moving some of the probability mass towards zero (but without going beyond zero), then Y_1 must always be preferred to Y_2. The reason is that the probability mass of Y_1 is more concentrated around zero, and thus the distortion introduced by Y_1 is smaller. A rational user always prefers less distortion and, therefore, prefers Y_1 to Y_2.

We use the notation $\langle 0, \alpha\rangle$, where $\alpha \in \mathbb{R}$, to denote the interval $[0, \alpha]$ when $\alpha \geq 0$, and the interval $[\alpha, 0]$ when $\alpha \leq 0$. If Y_1 can be constructed from Y_2 by moving some of the probability mass towards zero, it must be $P(Y_1 \in \langle 0, \alpha\rangle) \geq P(Y_2 \in \langle 0, \alpha\rangle)$ for any $\alpha \in \mathbb{R}$: otherwise, some of the probability mass that Y_2 had in $\langle 0, \alpha\rangle$ would have been moved outside $\langle 0, \alpha\rangle$, which is not possible (by assumption mass is moved towards zero without crossing zero). This leads to the following definition.

Definition 3 *Let Y_1 and Y_2 be two random noise distributions on $\mathbb{R}$. We say that Y_1 is smaller (or better) than Y_2, denoted by $Y_1 \leq Y_2$, if $P(Y_1 \in \langle 0, \alpha \rangle) \geq P(Y_2 \in \langle 0, \alpha \rangle)$ for any $\alpha \in \mathbb{R}$. We say that Y_1 is strictly smaller than Y_2, denoted by $Y_1 < Y_2$, if some of the previous inequalities are strict.*

For $\alpha = (\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$, we use $\langle 0, \alpha \rangle$ to denote the set $\langle 0, \alpha_1 \rangle \times \ldots \times \langle 0, \alpha_d \rangle$. Consider a set $S \subset \mathbb{R}^d$ such that for every point $x \in S$ we have $\langle 0, x \rangle \subset S$, and a pair of random noises $Y_1 = (Y_1^1, \ldots, Y_d^1)$ and $Y_2 = (Y_1^2, \ldots, Y_d^2)$ such that Y_1 can be constructed from Y_2 by moving some probability mass towards zero. It is obvious that we must have $P(Y_1 \in S) \geq P(Y_2 \in S)$: if that was not the case, it would mean that some of the probability mass that Y_2 had in S has been moved outside S, which is not possible because of the form of S. This leads to the definition for the multivariate case.

Definition 4 *Let Y_1 and Y_2 be two random noise distributions on $\mathbb{R}^d$. We say that Y_1 is smaller (or better) than Y_2, denoted by $Y_1 \leq Y_2$, if $P(Y_1 \in S) \geq P(Y_2 \in S)$ for every set $S \subset \mathbb{R}^d$ such that for any $x \in S$ we have $\langle 0, x \rangle \subset S$. We say that Y_1 is strictly smaller than Y_2, denoted by $Y_1 < Y_2$, if some of the previous inequalities are strict.*

Definitions 3 and 4 induce an order relationship between random noises. We use that order relationship to define the concept of optimal random noise.

Definition 5 *A random noise distribution Y_1 is optimal within a class $\mathcal{C}$ of random noise distributions if Y_1 is minimal within $\mathcal{C}$; in other words, there is no other random $Y_2 \in \mathcal{C}$ such that $Y_2 < Y_1$.*

As stated in the previous definition, the concept of optimality is relative to a specific class $\mathcal{C}$ of random noise distributions. In Section 5 we will determine the form of all optimal random noise distributions that provide ε-differential privacy to a specific query function f; to do so, we will take $\mathcal{C}$ to be the class of all random noise distributions that provide ε-differential privacy for f.

3 Characterization of differential privacy in terms of the noise

To build the optimal data-independent random noise distribution satisfying ε-differential privacy, we will have to analyze the properties that such a distribution must satisfy. The first step to perform this analysis is to express the condition in the definition of ε-differential privacy in terms of the random noise. Assuming a data-independent random noise Y, if we let $\kappa = f + Y$ then Inequality (1) becomes

$$P\left(Y \in S - f(D)\right) \leq e^{\varepsilon} P\left(Y \in S - f(D')\right)$$

As this inequality holds for all S, we can think of S as being of the form $S + f(D)$.

$$P\left(Y \in S\right) \leq e^{\varepsilon} P\left(Y \in S + \left(f(D) - f(D')\right)\right) \tag{2}$$

For the case of absolutely continuous random noise, the characterization in Inequality (2) can be expressed in terms of the density function f_Y of Y. To simplify the notation, we will assume that Y takes values in $\mathbb{R}$. Consider that f_Y is continuous except for a finite or countable set of removable discontinuities and a finite or countable set of jump discontinuities. If the set of jump discontinuities is countable, we will assume that it has no accumulation points; that is, around any jump discontinuity point in $\mathbb{R}$ we assume we can find an interval with no other jump discontinuity points. If f_Y has removable discontinuities we will modify f_Y to remove them. As we are modifying f_Y in at most a countable set, the modification will not affect the distribution of Y.

Let x be a continuity point of f_Y such that $x + d$ is also a continuity point, where $d = f(D) - f(D')$ for some data sets D and D' that differ in one row. Let I be an interval of size m centered at x such that f_Y is continuous in I and $I + d$. We know that such I exists because there are no accumulation points in the set of jump discontinuities. We can upper- and lower-bound the integrals by multiplying the maximum and minimum by the size of the interval:

$$m \times \inf_I(f_Y) \leq \textstyle\int_I f_Y \leq m \times \sup_I(f_Y)$$
$$m \times \inf_{I+d}(f_Y) \leq \textstyle\int_{I+d} f_Y \leq m \times \sup_{I+d}(f_Y)$$

As f_Y is continuous in I, the limit of $\inf_I(f_Y)$ and $\sup_I(f_Y)$ as the size m of I goes to zero is $f_Y(x)$. In the same way, as f_Y is continuous in $I + d$, the limit of $\inf_{I+d}(f_Y)$ and $\sup_{I+d}(f_Y)$ as m tends to 0 is $f_Y(x+d)$. Dividing both expressions by m and taking limits as m goes to zero, we have

$$f_Y(x) \leq lim_{m\to 0} \frac{\int_I f_Y}{m} \leq f_Y(x)$$
$$f_Y(x+d) \leq lim_{m\to 0} \frac{\int_{I+d} f_Y}{m} \leq f_Y(x+d)$$

Hence, combining the above limits and Expression (2) we have

$$\begin{array}{ccc} \frac{\int_I f_Y}{m} & \leq & e^{\varepsilon} \times \frac{\int_{I+d} f_Y}{m} \\ \downarrow & & \downarrow \\ f_Y(x) & & e^{\varepsilon} \times f_Y(x+d) \end{array}$$

Thus for all $x \in \mathbb{R}$ continuity point of f_Y, if $x + d$ is a continuity point we have

$$f_Y(x) \leq e^{\varepsilon} \times f_Y(x+d), \quad d = f(D) - f(D') \tag{3}$$

It is immediate to see that, if Inequality (3) holds, by integrating it over a set we recover Inequality (2). Hence, Inequality (3) is in fact an equivalent definition of ε-differential privacy for the case of a.c. random noise.

4 Non-optimality of the Laplace noise

Since the inception of differential privacy up to now [6,10], Laplace noise addition has been proposed as a method to achieve ε-differential privacy for an arbitrary function f in terms of its L_1-sensitivity. Also, as we said in the introduction, this practice has raised some criticisms.

In this section we show, for a univariate function f with values in $\mathbb{R}$, that the Laplace distribution is not optimal in the sense of Definition 5. To that end, we build another distribution, based on the Laplace distribution, that still fulfills the conditions of differential privacy and has its probability mass more concentrated towards zero, that is, it is strictly smaller than Laplace according to Definition 3. Although the distribution we build is optimal, we leave the formal proof of this assertion for Section 5.

The basic idea is to concentrate the probability mass around 0 as much as possible. This can only be done to a certain extent, because Inequality (3) limits our capability to do so. For example, increasing the value of the density at a point x may increase the minimum value that f_Y may take in the interval $[x - \Delta f,\, x + \Delta f]$.

In the construction of the distribution we will split the domain of f_Y into intervals of the form $[i\Delta f,\, (i+1)\Delta f]$ where $i \in \mathbb{Z}$. For each interval we will redistribute the probability mass that f_X assigns to that interval. The new density function $\tilde{f_Y}$ will take only two values (see Fig.1): $\max_{[i\Delta f,\,(i+1)\Delta f]} f_X$ at the portion of the interval closer to zero and $\min_{[i\Delta f,\,(i+1)\Delta f]} f_X$ at the portion of the interval farther from zero. The result is an absolutely continuous distribution where the probability mass has clearly been moved towards zero. We still have to check that it fulfills the conditions of ε-differential privacy.

To simplify, we will detail the argument only for intervals at the right of zero (positive reals); the argument for intervals at the left of zero is symmetrical. The probability mass at $[i\Delta f,\, (i+1)\Delta f]$ is $e^{-i\varepsilon}\frac{1-e^{-\varepsilon}}{2}$. The maximum value of the density of the Laplace distribution, $\frac{\varepsilon e^{-i\varepsilon}}{2\Delta f}$, occurs at the beginning of the interval and the minimum, $\frac{\varepsilon e^{-(i+1)\varepsilon}}{2\Delta f}$, occurs at the end. Let us determine the size m_i of the interval portion where the new density will be set to the maximum.

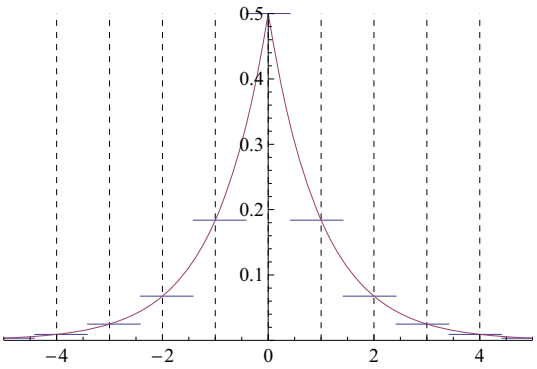

Fig. 1. Construction of the new distribution based on the Laplace(0,1) distribution

Since the probability mass of the interval must be preserved, we have

$$\frac{\varepsilon e^{-i\varepsilon}}{2\Delta f} m_i + \frac{\varepsilon e^{-(i+1)\varepsilon}}{2\Delta f}(\Delta f - m_i) = e^{-i\varepsilon}\frac{1-e^{-\varepsilon}}{2}$$

By solving for m_i in the above equality, we obtain:

$$m_i = \frac{\Delta f}{\varepsilon(1-e^{-\varepsilon})}(1 - e^{-\varepsilon} - \varepsilon e^{-\varepsilon})$$

The important fact about m_i is that it does not depend on i. Also, note that the maximum density of the current interval is equal to the minimum density of the previous interval. Hence, by joining the portion of the previous interval which evaluates to the minimum with the portion of the current interval which evaluates to the maximum, we obtain an interval of size $(\Delta f - m_{i-1}) + m_i = (\Delta f - m_i) + m_i = \Delta f$ which evaluates to a constant density value (such joined intervals are depicted as horizontal segments in Fig. 1). This way, except for the maximum of the first interval, we have split the domain of the density function into intervals of size Δf such that the density function evaluates to $\frac{\varepsilon e^{-i\varepsilon}}{2\Delta f}$. This clearly satisfies the density-based characterization of differential privacy specified by Inequality (3).

5 Optimal data-independent absolutely continuous noise for univariate queries

Section 4 has shown that the Laplace noise distribution is not optimal to achieve differential privacy. A new distribution has been built that satisfies differential privacy and has the probability mass more concentrated towards zero. This section will determine the optimal data-independent absolutely continuous random noise distribution to achieve ε-differential privacy for any univariate function with finite L_1-sensitivity. Optimal noise distributions need not

be symmetric; however, we focus on the symmetric case, because it is the most usual one.

Showing that optimal absolutely continuous noise distributions are of a certain form requires using some properties that will be stated as lemmata. Some of the proofs place additional regularity requirements on the noise distribution, beyond being absolutely continuous. These additional requirements are hardly a limitation as they are satisfied by any practical distribution, and can be overlooked if the reader is not interested in the proofs. In particular, we restrict the discussion to absolutely continuous random noises, Y, whose density function, f_Y, is continuous except for a finite or countable set of jump or removable discontinuities, with the set of jump discontinuities having no accumulation points. To avoid being unnecessarily cumbersome, we will not mention this again in the sequel.

It was shown in Section 3 that for a.c. noise distributions the definition of ε-differential privacy can be stated in terms of the density function. Now we show that if the inequality in terms of the probability function is satisfied at the extreme (*i.e.* it is satisfied as an equality), it also must be the case for the inequality in terms of density functions.

Lemma 6 *Let Y be an a.c. noise random variable that provides ε-differential privacy to a function f with a given L_1-sensitivity. Consider an interval $I = [i_0, i_1] \subset \mathbb{R}$. Then $P(Y \in I + \Delta f) = e^{-\varepsilon} P(Y \in I)$ if and only if $f_Y(x + \Delta f) = e^{-\varepsilon} f_Y(x)$, $\forall x \in I$, except at those points $x \in I$ such that f_Y is not continuous at x or at $x + \Delta f$. Similarly, $P(Y \in I - \Delta f) = e^{-\varepsilon} P(Y \in I)$ if and only if $f_Y(x - \Delta f) = e^{-\varepsilon} f_Y(x)$, $\forall x \in I$, except at those points $x \in I$ such that f_Y is not continuous at x or at $x - \Delta f$.*

Proof. See Appendix. □

We are trying to find the optimal a.c. noise distribution that provides ε-differential privacy. The goal is to concentrate as much probability mass around the mean as possible; ε-differential privacy limits our capability to do so. We will see how the probability mass must be distributed to achieve optimality.

Lemma 7 *Let Y be a symmetric a.c. noise random variable with zero mean that satisfies ε-differential privacy for a function f. If Y is optimal at providing ε-differential privacy, then for all $i \in \mathbb{Y}$*

$$P(Y \in [(i+1)\Delta f,\, (i+2)\Delta f]) = e^{-\varepsilon} P(Y \in [i\Delta f,\, (i+1)\Delta f])$$
$$P(Y \in [-(i+2)\Delta f,\, -(i+1)\Delta f]) = e^{-\varepsilon} P(Y \in [-(i+1)\Delta f,\, -i\Delta f])$$

Proof. See Appendix. □

Corollary 8 *Let Y be a symmetric a.c. noise random variable with zero mean that provides ε-differential privacy to a function f. If Y is optimal at providing ε-differential privacy then*

$$f_Y(x + \Delta f) = e^{-\varepsilon} f_Y(x) \quad \forall x \geq 0$$
$$f_Y(x - \Delta f) = e^{-\varepsilon} f_Y(x) \quad \forall x \leq 0$$

when the points x and $x + \Delta f$ in the first equality above and x and $x - \Delta f$ in the second equality are continuity points of f_Y.

Proof. The proof follows from Lemmata 6 and 7. □

Now we will show that for any symmetric a.c. noise distribution that provides ε-differential privacy for a function f we can find another noise distribution, similar to the one used in the proof that the Laplace distribution is not optimal, that performs at least as well according to Definition 3.

Theorem 9 *Let Y be an a.c. noise random variable with zero mean that provides ε-differential privacy to a query function f. Then there exists a noise random variable $\widetilde{Y}$ with density function $f_{\widetilde{Y}}$ of the form*

$$f_{\widetilde{Y}}(x) = \begin{cases} M_0 e^{-i\varepsilon} & x \in [-d - (i+1)\Delta f, -d - i\Delta f], \, i \in \mathbb{N} \\ M_0 & x \in [-d, 0] \\ M_0 & x \in [0, d] \\ M_0 e^{-i\varepsilon} & x \in [d + i\Delta f, d + (i+1)\Delta f], \, i \in \mathbb{N} \end{cases}$$

that provides ε-differential privacy to f and satisfies $\widetilde{Y} \leq Y$ as per Definition 3.

Proof. We will assume that Y is optimal and that its density function is not of the form of $f_{\widetilde{Y}}$ for any M_0 and d. The goal is to build another distribution $\widetilde{Y}$ from Y such that the density $f_{\widetilde{Y}}(x)$ is as stated above and satisfies $\widetilde{Y} \leq Y$. Note that, from the definition of $f_{\widetilde{Y}}(x)$, the condition of ε-differential privacy immediately holds for f.

Since Y fulfills the conditions of Corollary 8, we have

$$f_Y(x + \Delta f) = e^{-\varepsilon} f_Y(x) \quad \forall x \geq 0$$
$$f_Y(x - \Delta f) = e^{-\varepsilon} f_Y(x) \quad \forall x \leq 0$$

Now we apply the same procedure we used in Section 4 for the Laplace noise. First we split the domain of f_Y into intervals of the form $[i\Delta f, (i+1)\Delta f]$ where $i \in \mathbb{Z}$. At a given interval, we redistribute the probability mass that f_Y assigns to that interval. The new density function $f_{\widetilde{Y}}(x)$ takes only two

values: $\max_{[i\Delta f,(i+1)\Delta f]} f_Y$ at the portion of the interval closer to zero and $\min_{[i\Delta f,(i+1)\Delta f]} f_Y$ at the portion of the interval farther from zero. The result is an absolutely continuous distribution $\widetilde{Y}$ with $\widetilde{Y} \leq Y$.

To make sure that the distribution $\widetilde{Y}$ has the specified form, and thus satisfies ε-differential privacy, it remains to check that the length of the interval where we assign maximum value is constant across intervals.

The probability mass at $[i\Delta f, (i+1)\Delta f]$ is $e^{-i\varepsilon}\frac{1-e^{-\varepsilon}}{2}$. It is clear from $f_Y(x + \Delta f) = e^{-\varepsilon} f_Y(x)$, $\forall x \geq 0$, that the maximum and the minimum of each interval, M_i and m_i respectively, satisfy $M_i = e^{-i\varepsilon} M_0$ and $m_i = e^{-i\varepsilon} m_o$. Let d_i be the size of the interval where the new density evaluates to the maximum. We have

$$e^{-i\varepsilon} M_0 \times d_i + e^{-i\varepsilon} m_o \times (\Delta f - d_i) = e^{-i\varepsilon} \frac{1 - e^{-\varepsilon}}{2}$$

This formula leads to $d_i = \frac{1-e^{-\varepsilon}-2m_0\Delta f}{2(M_0-m_0)}$ which does not depend on i, as we wanted to see. □

Theorem 9 states that, for any random noise that provides ε-differential privacy to f, we can find another random noise distribution, of the specified form, that is smaller. However, we still have to prove that such a distribution is optimal.

Theorem 10 *Let Y be a random noise distribution with a density function f_Y of the form specified in Theorem 9. Then Y is optimal at providing ε-differential privacy.*

Proof. To prove that Y is optimal, we have to show that if we move some probability mass of Y towards zero then ε-differential privacy no longer holds. We only show it for the probability mass to the right of zero; a symmetric argument can be used for the probability mass to the left of zero.

First of all, we must show that it is not possible to move any probability mass from an interval $I_i = [i\Delta f, (i+1)\Delta f]$ to an interval $I_j = [j\Delta f, (j+1)\Delta f]$ with $0 \leq j < i$. This is straightforward: as the density f_Y specified in Theorem 9 has the maximum decrease rate between consecutive intervals compatible with the constraints of ε-differential privacy, moving probability mass from I_i to I_j would break ε-differential privacy.

To conclude the proof, we need to check that it is not possible to redistribute the probability mass within an interval I_i so that it gets closer to zero. Within the interval I_i, the density function f_Y takes values $M_0 \exp(-i\varepsilon)$ at I_i^l (the left portion of the interval) and $M_0 \exp(-(i+1)\varepsilon)$ at I_i^r (the right portion of the interval). We cannot move any probability mass from I_i^r towards zero, because the density would go below $M_0 \exp(-(i+1)\varepsilon)$ and, thus, ε-differential privacy would not hold. We cannot move any probability mass from I_i^l towards

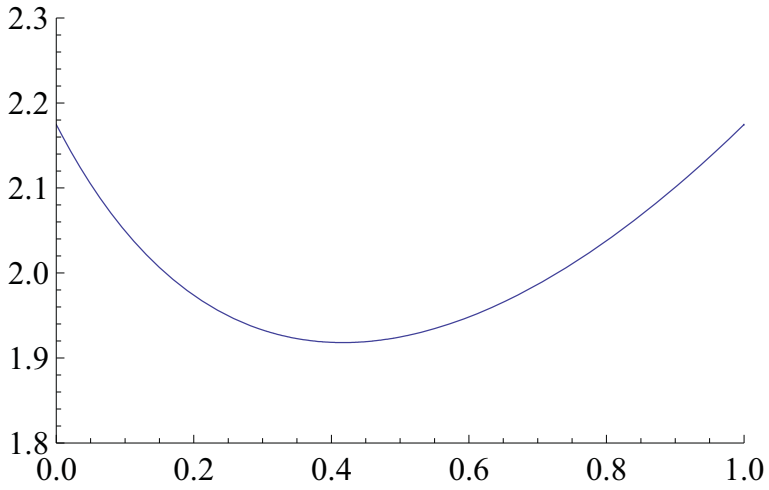

Fig. 2. Variance for $\varepsilon = 1$ and $\Delta f = 1$

zero, because the density would go above $M_0 \exp(-i\varepsilon)$ and, thus, ε-differential privacy would not hold. □

Although the theorems above are stated in terms of a fixed query function f, the optimal distribution depends only on Δf; hence, all query functions with the same L_1-sensitivity share the same optimal noise distribution.

The values of M_0 and d can be freely chosen according to the user's preferences. In fact the two parameters M_0 and d of the optimal family of distributions can be reduced to one because, as shown in the proof of Theorem 9,

$$d = \frac{1 - e^{-\varepsilon} - 2M_0 e^{-\varepsilon} \Delta f}{2(1 - e^{-\varepsilon}) M_0}$$

For instance, let us assume that the user prefers to minimize the noise variance. We compute the variance of candidate optimal distributions in terms of the parameters d and M_0, and find the values that yield the minimum:

$$V(Z) = 2M_0 \int_0^d x^2 dx + 2M_0 e^{-\varepsilon} \sum_{i=0\ldots\infty} e^{-i\varepsilon} \int_{d+i\Delta f}^{d+(i+1)\Delta f} x^2 dx$$

The variance can be computed by performing the integrals and calculating the sum of the power series. Fig. 2 shows the variance obtained in terms of the parameter d for the case of $\varepsilon = 1$ and $\Delta f = 1$. In this case, the minimum is reached at $d = 0.416737$ and the variance is 1.9181. This is below 2, the variance of the Laplace noise with scale parameter 1.

Table 1 shows a comparison of the variance achieved by the Laplace distribution and the optimal a.c. random noise with minimum variance, for different values of ε when $\Delta f = 1$. The table shows that the Laplace variance is only slightly greater than the minimum variance; we can say that, for a single

Table 1
Variance comparison between Laplace random noise and a.c. optimal random noise with minimum variance, for $\Delta f = 1$

	$\varepsilon = 0.1$	$\varepsilon = 0,5$	$\varepsilon = 1$
Laplace distribution	200.00	8.00	2.00
Optimal a.c. noise with min. var.	199.92	7.92	1.92

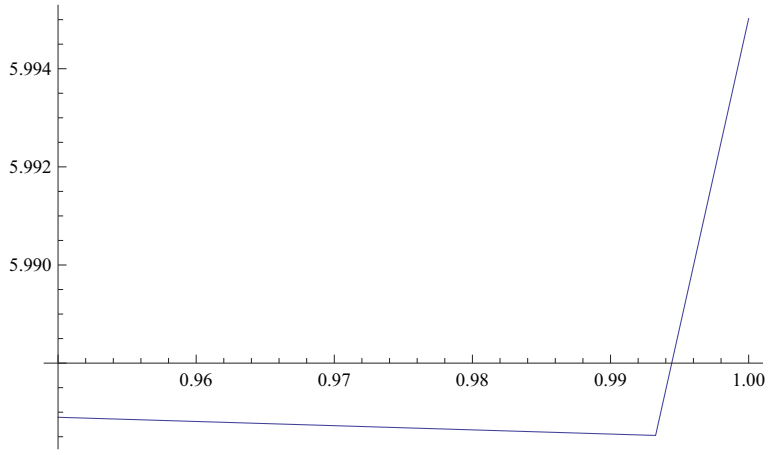

Fig. 3. Size of the 95% symmetric confidence interval centered at zero

univariate query, although the Laplace distribution is not optimal, it is near-optimal. Therefore, *if the utility of the differentially private answer to a single univariate query obtained using Laplace noise is poor, not much improvement can be expected from using a data-independent variance-optimal random noise distribution.*

Assume now that the user wants the noise distribution that minimizes the size of the symmetric confidence interval around the differentially private query answer that contains the real query value at 95% confidence level. In this case, we must solve a minimization problem, as before, but now the objective function is the size of the confidence interval in terms of the parameters d and M_0. Fig. 3 shows the size of the confidence interval, when $\Delta f = 1$ and $\varepsilon = 1$, in terms of parameter d. The minimal length for this case is achieved for $d = 0.993$, approximately; in general, however, the actual value of d where the minimum is reached depends on Δf and ε. Table 2 shows a comparison between the optimal lengths of the confidence intervals at 95% confidence level for several values of ε when $\Delta f = 1$. As expected, the results obtained from the Laplace distribution are worse but close to those obtained using the optimal distribution.

Table 2
Comparison of the size of the symmetric 95% confidence interval between Laplace random noise and a.c. optimal random noise with minimum confidence interval, for $\Delta f = 1$

	$\varepsilon = 0.1$	$\varepsilon = 0,5$	$\varepsilon = 1$
Laplace distribution	59.91	11.98	5.99
Optimal a.c. noise with min. conf. int.	59.91	11.97	5.98

6 Optimal data-independent absolutely continuous noise for multivariate queries

In Section 5 we worked out the optimal a.c. random noise for a query with values in $\mathbb{R}$. We deal here with multiple queries or with a single query whose response is a value in $\mathbb{R}^d$: both cases are equivalent, because d queries with answers in $\mathbb{R}$ can be viewed as a single query with answer in $\mathbb{R}^d$. Determining the form of all optimal multivariate a.c. random noises is out of scope; we restrict to a class of noise distributions whose density consists of several steps (as was the case for optimal univariate distributions) and show that they are optimal. The optimal distributions constructed will be shown to be substantially better than Laplace. Hence, *while Laplace is near-optimal in the univariate case, in general it is far from optimal for multivariate or multiple queries.*

We will be less formal here and, to simplify even more, examples will be presented for the case of two queries/two dimensions, that is, $d = 2$; generalization to arbitrary d is easy.

For the case of a.c. random noise for a single query, it was shown in Section 3 that the ε-differential privacy condition can be expressed in terms of the density function. The result is easily generalizable to greater dimensions, and therefore here we can also express the condition in terms of the density function.

Proposition 11 *Let $Y = (Y_1, \ldots, Y_d)$ be an absolutely continuous random noise that provides ε-differential privacy to a query $f : \mathcal{D} \to \mathbb{R}^d$. Then ε-differential privacy can be characterized in terms of the density function as:*

$$f_Y(x) \leq e^{\varepsilon} \times f_Y(x + d), \quad d = f(D) - f(D')$$

for all x and $x + d$ continuity points of f_Y, where D and D' are data sets that differ in one row.

Similarly to the case of a single univariate query, we will construct a noise density with several steps, which reaches its maximum all over a set that contains zero and decreases by a factor $e^{-\varepsilon}$ as we move away from it.

The main difference with other, non-optimal distributions, such as multivariate Laplace noise, is that the various components (dimensions) of the random noise do not need to be independent. This allows more freedom in the definition of the distribution, which we will employ to achieve a finer calibration to the query function. This is illustrated below in an example, but prior to it we define a set that will be repeatedly used in the remainder of this section.

Definition 12 *Let $f : \mathcal{D} \to \mathbb{R}^d$ be a query function. The set of differences between neighbor data sets is defined as*

$$S_f = \cup_{D,D'} \langle 0, f(D) - f(D') \rangle$$

where D and D' are data sets that differ in at most one row.

The set S_f contains all possible variations in f when one record changes. The boundary of S_f can be seen as a generalization of the L_1-sensitivity used in the univariate case. Instead of summarizing the variability of f with a single figure, as L_1-sensitivity does, S_f keeps track of the maximum variability in each direction.

Example 1 *Consider a query function $f = (f_1, f_2)$ such that $S_f = [-1, 1] \times [-1, 1]$. From Definition 2, the L_1-sensitivity of f is*

$$\Delta f = \sup_{D,D'} \|f(D) - f(D')\|_1 = \sup_{D,D'} (|f_1(D) - f_1(D')| + |f_2(D) - f_2(D')|) = 1 + 1 = 2$$

As stated in Proposition 11, the density of the random noise, f_Y, in each of the points of the set $[-1, 1] \times [-1, 1]$ must be in the range $[e^{-\varepsilon} f_Y(0), e^{\varepsilon} f_Y(0)]$. When using independent Laplace-distributed components with zero mean and $\Delta f/\varepsilon$ scale parameter, the top value for the density is reached at zero, and it decreases exponentially as we move away from it. Points with density $e^{-\varepsilon} f_Y(0)$ are those that have L_1-norm equal to Δf. Fig. 4 depicts S_f as a gray shaded box. If all points in S_f are protected with independent Laplace-distributed random noise components, all points within $[-1, 1] \times [-1, 1]$ must have density within the range $[e^{-\varepsilon} f_Y(0), f_Y(0)]$.

As it can be appreciated in Fig. 4, to satisfy ε-differential privacy at points $(1, 1)$, $(1, -1)$, $(-1, -1)$ and $(-1, 1)$ with independent Laplace noise addition for each dimension, we are overprotecting those points with L_1-norm less than or equal to $\Delta f = 2$ that do not belong to $[-1, 1] \times [-1, 1]$; the density at these points is greater or equal to $e^{-\varepsilon} f_Y(0)$, while this is not a requirement of ε-differential privacy (which only requires a density greater or equal to $e^{-\varepsilon} f_Y(0)$ for the points in S_f).

The ratio between the size of the overprotected region and the size of S_f may become still larger if the variability of one of the components is greater than the variability of the other. Fig. 5 illustrates the case of S_f being the set $[-1, 1] \times$

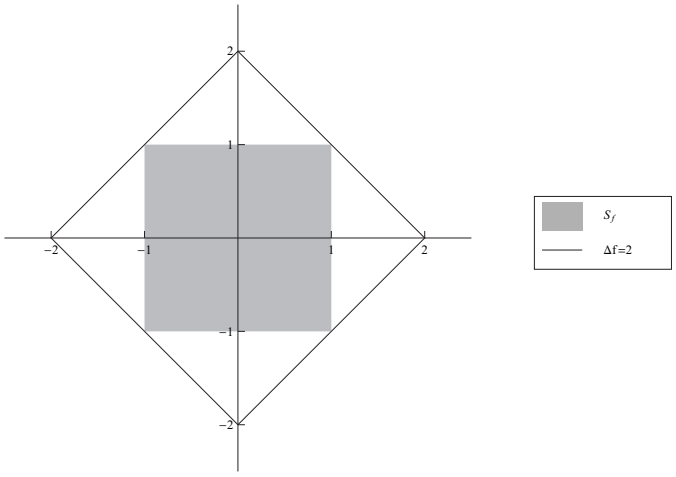

Fig. 4. Achieving ε-differential privacy by Laplace noise addition for $S_f = [-1, 1] \times [-1, 1]$. The shaded box represents the possible differences in the query result between data sets that differ in one record. Differential privacy requires the density of the noise in the shaded box to be within a factor in $[\exp(-\varepsilon), \exp(\varepsilon)]$ of the density at zero. The square that encloses the shaded box represents the points that satisfy the previous condition when using Laplace noise.

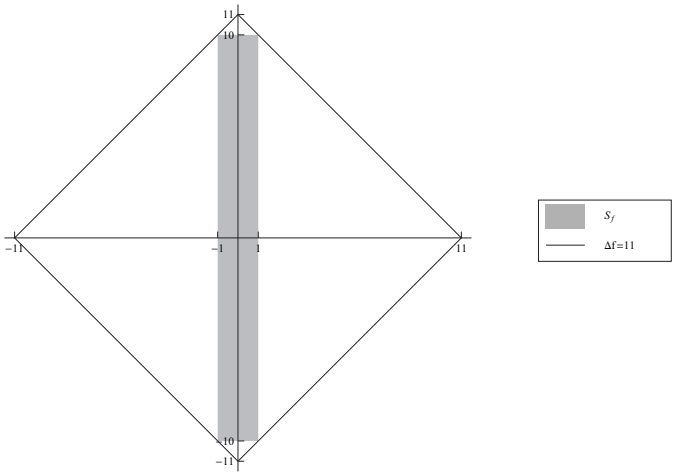

Fig. 5. Achieving ε-differential privacy by Laplace noise addition for $S_f = [-1, 1] \times [-10, 10]$ The shaded box represents the possible differences in the query result between data sets that differ in one record. Differential privacy requires the density of the noise in the shaded box to be within a factor in $[\exp(-\varepsilon), \exp(\varepsilon)]$ of the density at zero. The square that encloses the shaded box represents the points that satisfy the previous condition when using Laplace noise.

$[-10, 10]$.

In the construction of the piecewise constant noise density, we will fix a set $S_0 \subset S_f$ with $\langle 0, x\rangle \subset S_0$ for all $x \in S_0$, where the maximum density will be reached. From this S_0, we will define S_i as the set that contains the points

that are reachable from S_{i-1} in one step, that is, by adding a value from S_f:

$$S_i = \{x \in \mathbb{R}^d | x = z + \delta,\, z \in S_{i-1},\, \delta \in S_f\} \setminus \cup_{j=0}^{i-1} S_j$$

The density value over the points in S_i will be $e^{-\varepsilon}$ times the density value over the points in S_{i-1}. Therefore, for x in S_i it will be

$$f_Y(x) = Me^{-i\varepsilon}$$

The value M must be calibrated so that the total probability equals 1. Such calibration is possible because the density function decreases exponentially as i grows.

The following theorem shows that the constructed distribution is optimal at providing ε-differential privacy to the function f.

Theorem 13 *Let $f = (f_1, \ldots, f_d)$ be a query function with values in $\mathbb{R}^d$. Let $Y = (Y_1, \ldots, Y_d)$ be an a.c. random noise with density*

$$f_Y(x) = \sum_{i \geq 0} M \exp(-i\varepsilon) \mathbb{I}_{S_i}(x)$$

where $\mathbb{I}_{S_i}(x)$ is the indicator function for set S_i and M has been calibrated to adjust the total probability mass to one. If the following conditions hold, then Y is optimal at providing ε-differential privacy to f:

- $S_0 \subset S_f$
- $\langle 0, x \rangle \subset S_0$ *for all* $x \in S_0$
- $S_{i+1} = (S_i + S_f) \setminus \cup_{j=0}^{i-1} S_j$ *for all* $i \geq 0$

Proof. See Appendix. □

Example 2 *Let f be a function with $S_f = [-1, 1] \times [-10, 10]$, and take $\varepsilon = 1$. Hence, the sensitivity of f is $\Delta f = 1 + 10 = 11$ and ε-differential privacy with two independent Laplace-distributed random noise components requires these components to have zero mean and $\frac{11}{\varepsilon}$ scale parameter. Our proposal to achieve ε-differential privacy is to use the piecewise constant density construction by setting $S_0 = [-0.1, 0.1] \times [-1, 1]$. Fig. 6 shows the density function of both distributions. Note that with the Laplace distribution the noise densities for both components of f decrease at the same rate, even if the second component of f has ten times the sensitivity of the first one.*

It is easily appreciated in the figure that the piecewise constant distribution has much more probability concentrated around zero, which agrees with our optimality definition in Section 2. To compare both distributions, we compute the variance of the components, and the minimal size of a confidence region at some confidence levels.

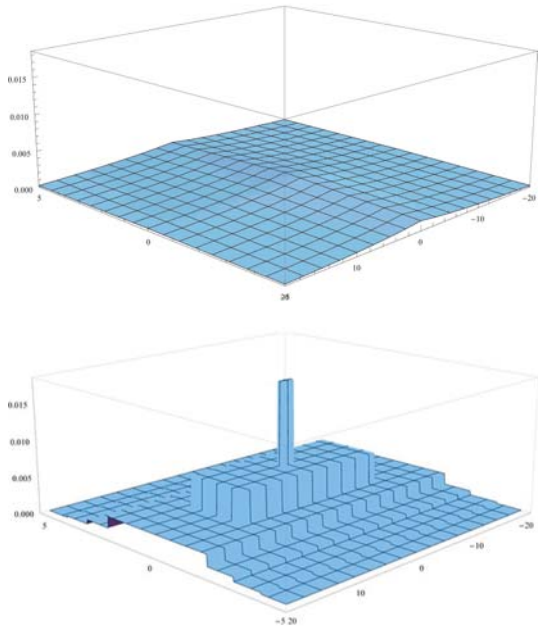

Fig. 6. Density functions of the Laplace and piecewise constant noise distributions required to achieve 1-differential privacy for a bivariate function $f = (f_1, f_2)$ with $\Delta f_1 = 1$ and $\Delta f_2 = 10$

Table 3
Minimal size of the confidence region for two-dimensional Laplace-distributed random noise with scale parameter 11

Confidence level	α	Size
0.99	73.02	10663
0.95	52.18	5445
0.90	42.79	3662

For Laplace-distributed random noise (Y_1, Y_2), the computations are easy. Since we know that Y_1 and Y_2 follow a Laplace distribution, their variance is twice the square of the scale factor

$$Var(Y_1) = 242$$
$$Var(Y_2) = 242$$

With the Laplace-distributed random noise (Y_1, Y_2), points with equal L_1-norm are assigned the same noise density. Therefore the confidence region of minimal size, for a given confidence level, is of the form $\{x| \, \|x\| \leq \alpha\}$. Table 3 shows the size of the confidence region for several confidence levels.

Computing the variance of the components of the piecewise constant distribution will be done in terms of the sets S_f and S_0. If we let $S_f = [-s_1, s_1] \times$

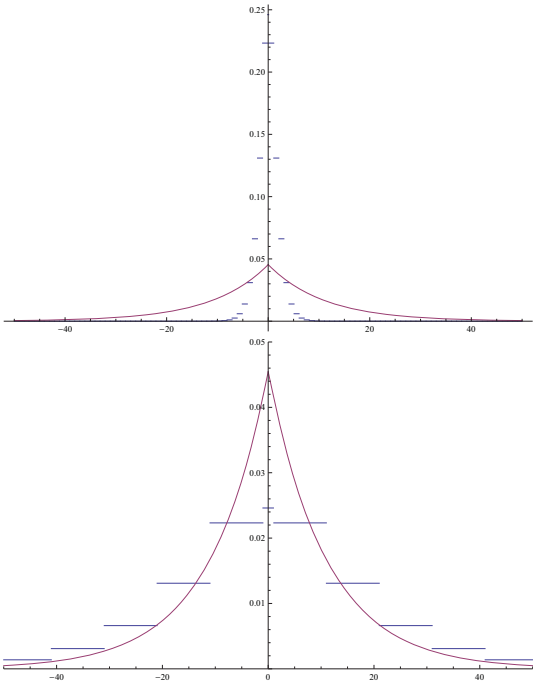

Fig. 7. Comparison of the components of the Laplace and the piecewise constant random noise distributions required to achieve 1-differential privacy for a bivariate function $f = (f_1, f_2)$ with $\Delta f_1 = 1$ and $\Delta f_2 = 10$. Top, comparison of the first component; bottom, comparison of the second component.

$[-s_2, s_2]$ *and* $S_0 = [-z_1, z_1] \times [-z_2, z_2]$ *then the density of the components* Y_1 *and* Y_2 *is*

$$f_{Y_1}(x) = 2Me^{-i_1\varepsilon} \times (z_2 + s_2 i_1 + s_2/(e^{\varepsilon} - 1))$$
$$f_{Y_2}(x) = 2Me^{-i_2\varepsilon} \times (z_1 + s_1 i_2 + s_1/(e^{\varepsilon} - 1))$$

where $i_1 = \lfloor (|x| - z_1)/s_1 + 1 \rfloor$ *is the index of the first set* S_i *such that* $(x, 0)$ *belongs to it,* $i_2 = \lfloor (|x| - z_2)/s_2 + 1 \rfloor$ *is the index of the first set* S_i *such that* $(0, x)$ *belongs to it, and* M *is a constant adjusted so that the random distribution* (Y_1, Y_2) *has probability mass one. Fig. 7 compares the first and second components of the Laplace and the piecewise constant random noise. Note that the piecewise constant distribution seems to slightly underperform Laplace for the second component, but it clearly outperforms Laplace for the first component.*

Since the mean of the components is zero, their variance can be computed by integrating $\int_{\mathbb{R}} x^2 f_{Y_i}(x)dx$*, which results in:*

$$Var(Y_1) = 4.0338$$
$$Var(Y_2) = 403.38$$

Table 4
Minimal size of the confidence region for the piecewise constant noise distribution needed for a bivariate function $f = (f_1, f_2)$ with $\Delta f_1 = 1$ and $\Delta f_2 = 10$

Confidence level	β	Size
0.99	6.99	1790.2
0.95	4.79	916.6
0.90	3.90	611.2

Compared to the variances obtained for the Laplace-distributed random noise, we observe that the variance for Y_2 when using the piecewise constant distribution is about twice as big as when using Laplace distribution. On the other side, the variance of Y_1 is much smaller when using the piecewise constant distribution. These results are consistent with the previous observation about Fig. 7.

We compute now confidence regions for the piecewise constant distribution. To obtain a confidence region with minimal size, we make sure to include all the points in S_i before including any point in S_{i+1}. We will consider confidence regions of the form $[-z_1 - \beta s_1, z_1 + \beta s_1] \times [-z_2 - \beta s_2, z_2 + \beta s_2]$. Table 4 shows the confidence regions obtained. By comparing with Table 3, it can be observed in the table that the minimal size for a confidence level is much smaller when using the piecewise constant distribution.

Note that in Example 1 we considered S_f to be the product of two intervals. This case models the situation where the query function components are independent, in the sense that any combination of values for the difference of the query function is possible. That is, $S_f = [-1, 1] \times [-1, 1]$ means that, for any $[\delta_1, \delta_2] \in [-1, 1] \times [-1, 1]$, we can find two data sets D and D' differing in one row such that $f_1(D) - f_1(D') = \delta_1$ and $f_2(D) - f_1(D') = \delta_2$. Taking S_f to be the product of intervals is the natural option in the case of an interactive mechanism [6], where we get to know each of the components of the query function (*i.e.* each successive query if we view the multivariate query as a group of queries) at different times. In an interactive mechanism it is not possible to construct the distribution that best matches the multiquery function f, because at the time of the first query we only know f_1. Clearly, it is possible to achieve a better noise calibration for a non-interactive query than for an interactive one, but using independent Laplace noise addition for each component fails to exploit non-interactivity.

7 Conclusions

The goal of this paper was to analyze the optimality of data-independent random noise distributions to achieve ε-differential privacy. The first step was to define the concept of optimal distribution as a distribution that concentrates the probability around zero as much as possible while ensuring differential privacy. This criterion led to a family of optimal distributions, which can be refined by using additional criteria. In the examples, we have computed optimal distributions using as additional criteria the minimization of the response variance or the minimization of the size of the confidence interval around the response.

For a single univariate query, the optimal absolutely continuous noise distributions to achieve ε-differential privacy were built; as a result, we obtained a family of piecewise constant density functions. The comparison with the Laplace noise distribution showed that Laplace performs only slightly worse than the optimal absolutely continuous distributions. Comparison figures were provided for the variance and the size of the confidence interval.

For a multivariate query or multiple queries, a piecewise constant construction similar to that of a single query was presented. Comparisons in terms of variance and of size of the minimal confidence interval showed that, *for multivariate and/or multiple queries, the Laplace distribution is far from being optimal*. Given the popularity of the Laplace distribution, this is a very relevant result. We also observed that the proposed mechanism provides better responses for non-interactive queries, as it is able to exploit the global knowledge on the query function. This is not possible for mechanisms that assume the components of the query function to be independent, as it is the case for Laplace noise addition.

Appendix: Proofs

Proof (Lemma 6). We will prove the first claim; the second one is completely symmetric. The proof of ($\Longleftarrow$) is straightforward by computing the probability as the integral of the density function. We will focus on the ($\Longrightarrow$) implication. By the ε-differential privacy condition we know that $f_Y(x + \Delta f) \geq e^{-\varepsilon} f_Y(x)$. Assuming that the implication does not hold, a continuity point $a \in I$ exists such that $f_Y(a + \Delta f) > e^{-\varepsilon} f_Y(a)$. Because of the constraints on the set of discontinuity points, an interval $[a_0, a_1] \subseteq I$ exists such that $f_Y(x + \Delta f) > e^{-\varepsilon} f_Y(x)$, $\forall x \in [a_0, a_1]$. Now we can decompose the probabilities in the state-

ment of the Lemma as follows:

$$P(Y \in I) = \int_{i_0}^{a_0} f_Y(x)dx + \int_{a_0}^{a_1} f_Y(x)dx + \int_{a_1}^{i_1} f_Y(x)dx$$

$$P(Y \in I + \Delta f) = \int_{i_0}^{a_0} f_Y(x + \Delta f)dx + \int_{a_0}^{a_1} f_Y(x + \Delta f)dx + \int_{a_1}^{i_1} f_Y(x + \Delta f)dx$$

Since $f_Y(a + \Delta f) \geq e^{-\varepsilon} f_Y(a)$ and, for $x \in [a_0, a_1]$, $f_Y(a + \Delta f) > e^{-\varepsilon} f_Y(a)$, we have $P(Y \in I + \Delta f) > e^{-\varepsilon} P(Y \in I)$, which is a contradiction that comes from the assumption that a continuity point $a \in I$ exists such that $f_Y(a + \Delta f) > e^{-\varepsilon} f_Y(a)$. □

Proof (Lemma 7). The second claim is completely symmetric to the first one; a symmetric distribution that satisfies the first claim will also satisfy the second one. We will show that, if the claims do not hold, we can build another distribution that fulfills ε-differential privacy and has the probability mass more concentrated towards zero.

We will assume that the claim for Y does not hold and we will build another distribution $\widetilde{Y}$ that provides ε-differential privacy and has $\widetilde{Y} \leq Y$. If the claim held, by Lemma 6, it would be $f_Y(x + \Delta f) = e^{-\varepsilon} f_Y(x) \, \forall x \in \mathbb{R}$ where x and $x + \Delta f$ are continuity points. Let $i_0 \geq 0$ be the index of the first interval $[i\Delta f, (i + 1)\, \Delta f]$ such that $f_Y(x + \Delta f) = e^{-\varepsilon} f_Y(x)$ does not hold for all x in the interval. Let $\tilde{f}_{i_0}$ be the function defined as follows

$$\tilde{f}_{i_0}(x) = \begin{cases} e^{-\varepsilon} f_Y(x + \Delta f) & x \in [-(i_0 + 1)\, \Delta f, -\Delta f] \\ f_Y(x) & x \in [-\Delta f, +\Delta f] \\ e^{-\varepsilon} f_Y(x - \Delta f) & x \in [\Delta f, (i_0 + 1)\, \Delta f] \end{cases}$$

Since $\tilde{f}_{i_0}$ has been defined in such a way that the decrease of the density between points at distance Δf is maximum as we move away from zero, it is clear that we will have $f_Y > \tilde{f}_{i_0}$. As both f_Y and $\tilde{f}_{i_0}$ are symmetric, we will only consider the points on the right of zero; the same transformations must be applied to the points on the left. For each $x \in [\Delta f, (i_0 + 1)\, \Delta f]$ we will consider $e_x = f_Y(x) - \tilde{f}_{i_0}(x)$, the excess density of f_Y over $\tilde{f}_{i_0}$. We will build another function f_{i_0} by distributing e_x among the points $\{x + i\Delta f : 0 \leq i \leq i_0\}$ in such a way that the new function concentrates as much as possible around the mean, and ε-differential privacy is satisfied. The density added to $\tilde{f}_{i_0}$ at $x + i\Delta f$ will be $\alpha_x e^{-i\varepsilon}$ where α_x is determined by imposing $\sum_{i=0,\ldots,i_0} \alpha_x e^{-i\varepsilon} = e_x$. Note that f_{i_0} still satisfies that images of points at distance Δf exponentially decrease as we move away from zero, that is $f_{i_0}(x + \Delta f) = e^{-\varepsilon} f_{i_0}(x)$.

It is important to note that the new function f_{i_0} satisfies ε-differential privacy in the range $[-i_0\Delta f, i_0\Delta f]$. We will show that ε-differential privacy is satisfied in the interval $[-\Delta f, \Delta f]$; then by using that the images by f_{i_0} of points at distance Δf exponentially decrease as we move away from zero, ε-differential

privacy will be satisfied in $[-i_0\Delta f, i_0\Delta f]$. In fact we will only check that ε-differential privacy is satisfied in $[0, \Delta f]$; if it is so, by the symmetry of f_{i_0}, differential privacy will be satisfied in the whole interval $[-\Delta f, \Delta f]$.

We must check that $f_{i_0}(x+\delta) \leq e^{\varepsilon} \times f_{i_0}(x)$ for all $x \in [0, \Delta f]$ and all $\delta \in [-\Delta f, \Delta f]$. Let us assume that there exist $x \in [0, \Delta f]$ and $\delta \in [-\Delta f, \Delta f]$ such that the condition is not satisfied, that is, $f_{i_0}(x+\delta) > e^{\varepsilon} f_{i_0}(x)$. If $x + \delta \in [\Delta f, 2\Delta f]$, by multiplying by $e^{-(i_0-1)\varepsilon}$ we have that $x + (i_0 - 1)\Delta f$, the corresponding point in the interval $[(i_0-1)\Delta f, i_0\Delta f]$, does not fulfill the ε-differential privacy condition, but this is not possible as we had $f_Y(x+i_0\Delta f) \leq e^{\varepsilon} f_Y(x + (i_0 - 1)\Delta f)$ and when building f_0 we have increased the value at $x + (i_0 - 1)\Delta f$ and decreased the value at $x + i_0\Delta f$. If $x + \delta \in [0, \Delta f]$, by multiplying by $e^{-i_0\varepsilon}$ we have that the corresponding point in the interval $[i_0\Delta f, (i_0+1)\Delta f]$ does not satisfy the differential privacy condition. This is impossible as we know that $\tilde{f}_{i_0}$ and f_Y do satisfy it and that f_{i_0} lies between them; therefore f_{i_0} must also satisfy the differential privacy condition. In the case $x + \delta \in [-\Delta f, 0]$, the justification is different. The point $-x - \delta$ belongs to the interval $[0, \Delta f]$ and, by the symmetry of f_{i_0}, we have $f_{i_0}(-x-\delta) = f_{i_0}(x+\delta)$; therefore, as we have already checked that the condition is satisfied when $x + d \in [0,\ \Delta f]$, it must also be satisfied when $x + d \in [-\Delta f, 0]$.

Now we iterate this process and define functions f_i, $i \in \mathbb{N}$. To be able to do this, it is important to note that, when defining f_i, we are reducing the density amount in the interval $[i\Delta f, (i+1)\Delta f]$ and that $\tilde{f}_{i+1}$ is defined in $[(i+1)\Delta f, (i+2)\Delta f]$ by reducing the value in the previous interval as much as possible while still satisfying ε-differential privacy. This means that $f_Y > \tilde{f}_{i+1}$ at $[(i+1)\Delta f, (i+2)\Delta f]$ and thus we can compute the excess and distribute it among the corresponding points in the previous intervals.

The resulting $\tilde{f}_\infty$ satisfies the ε-differential privacy condition. By construction it also satisfies $f_Y(x + \Delta f) = e^{-\varepsilon} f_Y(x)\, \forall x \in \mathbb{R}$ which by integration over the desired intervals leads to the claim of the lemma. Moreover, as all the probability mass translation has been done towards zero, we have $\widetilde{Y} \leq Y$. $\square$

Proof (Theorem 13). First of all we check that Y satisfies the ε-differential privacy condition as stated in Proposition 11. Consider $x \in \mathbb{R}^d$ and $\delta \in S_f$. The sets S_i form a cover of $\mathbb{R}^d$; therefore we have $x \in S_i$ for some $i \in \mathbb{N}$. For $x + \delta$ we have one of the following possibilities: $x + \delta \in S_{i-1}$, $x + \delta \in S_i$, or $x + \delta \in S_{i+1}$. The value of the density function will, respectively, be $Me^{-(i-1)\varepsilon}$, $Me^{-i\varepsilon}$, or $Me^{-(i+1)\varepsilon}$; in all three cases, the ε-differential privacy condition is satisfied.

To show that Y is optimal at providing ε-differential privacy to f we have to check that if we move some probability mass towards zero, the resulting random noise does not provide ε-differential privacy to f. We partition $\mathbb{R}^d$, and

check, for each set in the partition, that it is not possible to move any probability mass towards zero and still satisfy ε-differential privacy. The partition is $\{S_f^i,\, i \geq 1\}$ where $S_f^1 = S_f$ and $S_f^{i+1} = (S_f^i + S_f) \setminus \cup_{j=1}^{i} S_f^j$.

We start by checking that it is not possible to move any probability mass contained in S_f^1 towards zero and still satisfy ε-differential privacy. The density f_Y in S_f^1 can be expressed as

$$f_Y(x) = M \times \mathbb{I}_{S_0}(x) + M \exp(-\varepsilon) \times \mathbb{I}_{S_f^1 \setminus S_0}(x)$$

Note that f_Y already has the maximum change in the density that ε-differential privacy allows: $\exp(\varepsilon)$. In other words, if we increase the density above M or decrease it below $M \times \exp(-\varepsilon)$, ε-differential privacy will not hold. Let $U \subset S_f^1$ be the set that will have its probability mass reduced. It must be $U \subset S_0$; otherwise some points would have its density reduced below $M \times \exp(-\varepsilon)$, which is not possible. Now, as we have $\langle 0, x \rangle \subset S_0$ for all $x \in S_0$ (*i.e* for any point in S_0 the points closer to zero are already in S_0), if we move probability mass from U towards zero, this probability mass must go to a set of points U' contained in S_0. This way the density of points in U' would be greater than M, which would also break ε-differential privacy.

To conclude the proof we have to check that it is not possible to move any probability mass belonging to a set S_f^{i+1} with $i \geq 1$ towards zero and still satisfy ε-differential privacy. Note that the density function f_Y decreases as fast as possible as we move away from S_0: according to Proposition 11 the density at a point y reachable from a point x by adding a value from S_f must satisfy $f_Y(y) \geq \exp(-\varepsilon) f_Y(x)$. We have set the density f_Y at S_{i+1} to be $\exp(-\varepsilon)$ times the density at S_i; that is, the minimum value that satisfies ε-differential privacy.

To move some probability mass belonging to S_f^{i+1} towards zero we must select a set $U \subset S_f^{i+1}$ and reduce its probability mass. In other words, the density function at the points in U is to be reduced. But this is not possible, if we want to preserve ε-differential privacy (as pointed out in the previous paragraph, when we move away from S_0, the density f_Y already decreases as fast as differential privacy permits). □

Acknowledgments

The authors are with the UNESCO Chair in Data Privacy, but the views expressed in this paper are their own and do not commit UNESCO. The second author is partly supported as an ICREA-Acadèmia researcher by the Government of Catalonia. This work was partly funded by the European Commis-

sion under FP7 project "DwB", by the Spanish Government through projects TSI2007-65406-C03-01 "E-AEGIS", TIN2011-27076-C03-01 "CO-PRIVACY" and CONSOLIDER INGENIO 2010 CSD2007-0004 "ARES", and by the Government of Catalonia under grant 2009 SGR 1135.

References

[1] A. Blum, C. Dwork, F. McSherry, and K. Nissim. Practical privacy: the SuLQ framework. In *Proceedings of the 24th ACM Symposium on Principles of Database Systems-PODS 2005*, pages 128–138, 2005.

[2] A. Blum, K. Ligett, and A. Roth. A learning theory approach to non-interactive database privacy. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing-STOC 2008*, pages 609–618, 2008.

[3] I. Dinur, and K. Nissim. Revealing information while preserving privacy. In *Proceedings of the 32nd ACM Symposium on Principles of Database Systems*, pages 202–210, 2003.

[4] C. Dwork, and K. Nissim. Privacy-preserving datamining on vertically partitioned databases. In *Proceedings of the 24th Annual International Cryptology Conference-CRYPTO 2004*, pages 528–544, 2004.

[5] C. Dwork. Differential privacy. In *33rd International Colloquium on Automata, Languages and Programming-ICALP 2006, Part II*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12, 2006.

[6] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *TCC 2006-Theory of Cryptography Conference*, pages 265–284, 2006.

[7] C. Dwork. Differential privacy: a survey of results. In M. Agrawal, D. Du, Z. Duan, A. Li editors, *Theory and Applications of Models of Computation*, *Lecture Notes in Computer Science*, pages 1–19, 2008.

[8] C. Dwork and A. Smith. Differential privacy for statistics: what we know and what we want to learn. *Journal of Privacy and Confidentiality*, 1(2): 135–154, 2009.

[9] C. Dwork, M. Naor, O. Reingold, G. N. Rothblum, and S. Vadhan. On the complexity of differentially private data release: efficient algorithms and hardness results. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing-STOC 2009*, pages 381–390, 2009.

[10] C. Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54:86–95, 2011.

[11] A. Hundepool, J. Domingo-Ferrer, L. Franconi, S. Giessing, E. Schulte Nordholt, K. Spicer, and P.-P. de Wolf. *Statistical Disclosure Control*. Wiley, 2012.

[12] D. Leoni. Non-interactive differential privacy: a survey. In *Proceedings of the First International Workshop on Open Data-WOD 2012*, pages 40–52, 2012.

[13] A. Machanavajjhala, D. Kifer, J. Abowd, J. Gehrke, and L. Vilhuber. Privacy: theory meets practice on the map. In *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering-ICDE 2008*, pages 277–286, 2008.

[14] F. Mcsherry, and K. Talwar. Mechanism design via differential privacy. In *Proceedings of the 48th Annual Symposium on Foundations of Computer Science-FOCS 2007*, pages 94–103, 2007.

[15] K. Muralidhar and R. Sarathy. Does differential privacy protect Terry Gross' privacy? In *Privacy in Statistical Databases-PSD 2010*, volume 6344 of *Lecture Notes in Computer Science*, pages 200–209, 2010.

[16] K. Nissim, S. Raskhodnikova, and A. Smith. Smooth sensitivity and sampling in private data analysis. In D. S. Johnson and U. Feige, editors, *39th ACM Symposium on Theory of Computing-STOC 2007*, pages 75–84. ACM, 2007.

[17] R. Sarathy and K. Muralidhar. Some additional insights on applying differential privacy for numeric data. In *Privacy in Statistical Databases*, volume 6344 of *Lecture Notes in Computer Science*, pages 210–219, 2010.