Health Serv Outcomes Res Method (2010) 10:1–15
DOI 10.1007/s10742-010-0061-3

Assessing database privacy using the area under the receiver-operator characteristic curve

Gregory J. Matthews · Ofer Harel · Robert H. Aseltine Jr.

Received: 12 March 2010 / Revised: 15 July 2010 / Accepted: 19 July 2010 /
Published online: 31 July 2010

Abstract One of the most pressing issues in the confidentiality literature is the quantification of privacy. One proposal, ϵ-differential privacy, moves away from absolute guarantees of privacy to relative guarantees. However, the selection of an appropriate ϵ is difficult because its interpretation is unclear. Further, when comparing different privacy preserving techniques to one another, a direct comparison cannot be made by simply comparing the respective values of ϵ. The aim of this work is to provide a measure that allows for direct comparison across different privacy schemes and is more easily interpreted. In turn, this will aid in policy debate pertaining to how much privacy is acceptable. Our proposal sets the problem in a hypothesis testing framework and uses the area under the receiver-operator characteristic (ROC) curve as a measure of privacy.

Keywords Privacy · Confidentiality · ROC curve · Likelihood ratio testing · Hypothesis testing · Health policy

1 Motivation

According to the National Institute of Health (NIH) Data Sharing Policy and Implementation Guidance, "Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data." While the NIH is just one example, this statement summarizes the often conflicting goals of many organizations that release data to the public. Scientific research is fueled by data, without which, progress in research may be brought to a halt. However, these same organizations must also acknowledge the ethical and legal consequences for failing to maintain the privacy of an individual within any publicly released data set.

G. J. Matthews · O. Harel (✉)
Department of Statistics, University of Connecticut, Storrs, CT, USA
e-mail: oharel@stat.uconn.edu

R. H. Aseltine Jr.
Division of Behavioral Sciences and Community Health, Institute for Public Health Research
University of Connecticut Health Center, East Hartford, CT, USA

Many large data collecting institutions, such as the U.S. Bureau of the Census, strive to release their data to the public while maintaining privacy. Also, many other agencies that provide grants, such as NIH, may require grantees to release their data to the public. Regardless of the agency, however, releasing useful microdata and maintaining privacy are often in direct conflict with each other. From a scientific perspective, it is ideal to release a complete data set in its unaltered state. This way researchers can use the same data to replicate past experiments, which is central to the foundations of good science. At the other extreme, total privacy protection, no data would be released making it impossible to perform research. As such, some compromise must be reached between utility of the released data and individual privacy.

When releasing microdata to the public, many different privacy preserving techniques have been proposed, and many of these methods allow a consumer of the perturbed data to make valid scientific inferences about a population of interest. Examples will be reviewed in the next section. Using measures such as bias, mean squared error, and confidence interval coverage, these methods have been studied for their utility properties, and they often work quite well. However, while each of these methods offer some degree of privacy, quantifying the level of privacy is often difficult. Even after measuring privacy protection, the question of how much privacy is enough still needs to be addressed.

The proposal in this paper builds off the framework of ϵ-differential privacy proposed in Dwork (2006). Here privacy is assessed in a hypothesis testing framework. Using this framework, the end result gives a metric which allows for direct comparisons to be made across different privacy preserving schemes. Further, the scale of the metric ranges from 0.5 to 1 with values close to 0.5 associated with high levels of privacy and the opposite for values near 1. This scale is more easily interpretable and offers some intuition as to what is an acceptable level of privacy. The metric and its interpretation are described in detail in Sect. 3.

The paper is organized as follows. Section 2 contains a review of methods used for statistical disclosure limitation, as well as past proposed measures of privacy. Section 3 discusses differential privacy, its definitions, and extensions. Section 4 proposes using the receiver-operator characteristic (ROC) curve and associated area under curve metric to assess privacy in a more intuitive way, followed by several examples of this privacy metric in Sect. 5. The paper concludes with a discussion and some suggestions for possible directions of future work.

2 Statistical disclosure limitation and quantifying privacy

When an agency wishes to privately release microdata to the public in conformance with applicable laws (for example, the Health Insurance Portability and Accountability Act (HIPAA) for medical data) a primary defense is the removal of obvious identifiers, such as name, social security number, or address. However, Sweeney (2002) showed that by using auxiliary information, privacy breaches can occur even after removing all of these identifiers. Some methods that go beyond removing obvious identifiers for preserving privacy include sampling, top coding, suppression, and limitation of detail. Further privacy preserving methods include matrix masking (Cox 1980, 1994), data swapping (Dalenius and Reiss 1982), and synthetic data (Rubin 1993). These methods all, to varying degrees, achieve greater levels of privacy than simply removing obvious identifiers, while still maintaining many useful properties for making inference about the population.

An important consideration in the evaluation of these methods is the quantification of utility and privacy. Assessing the utility of privacy preserved data has been studied often

and is fairly easy to assess. Assessing utility for perturbed data are discussed in Keller and Bethlehem (1992) for matrix masking, Little (1993) for the addition of white noise, and Raghunathan et al. (2003), Reiter (2002), and Matthews et al. (2010) for synthetic data.

Quantifying the privacy of the data, on the other hand, is more difficult and a less studied topic especially in the statistics and biostatistics literature. Several proposed methods involving record linkage are discussed in Bethlehem et al. (1990), Dale and Elliot (2001), Duncan and Lambert (1989), Paass (1988), Skinner and Elliot (2002), Yancey et al. (2002), and Domingo-Ferrer and Torra (2004). Another proposal suggests using canonical correlation analysis to assess privacy as in Sarathy and Muralidhar (2002), and Reiter (2005) discusses predictive disclosure risk in the context of synthetic data.

Sweeney (2002) proposes a measure called k-anonymity which seeks to limit the number of unique combinations of identifying features in the data. However, Machanavajjhala et al. (2007) demonstrated how privacy can be breached even when the data achieves k-anonymity and, in response, proposes l-diversity. Li et al. (2007) notes that k-anonymity can prevent identity disclosures, but attribute disclosures are still possible. Further, the authors note that l-diversity fails to account for the knowledge of the adversary and offer a measure of privacy called t-closeness. All three of these methods, k-anonymity, l-diversity, and t-closeness, assess privacy based on some version of the actual released data, which may be appropriately coarsened to achieve desired properties.

Alternatively, instead of releasing a private version of the original data, information could be released via a query system. Rather than release a full private data set, an interested party could query a database consisting of the private data. Based on some measure of privacy, if the query is deemed to achieve an acceptable level of privacy the query could be released. For example, a user may pose a query to the database requesting the means and variances of several variables or, perhaps, the regression coefficients of one variable regressed against several others. As an added layer of protection, rather than releasing the actual response to a query, noise could be added to perturb the actual response and a randomized version of the query is returned.

2.1 Differential privacy

ϵ-differential privacy (Dwork 2006) assesses privacy based on a randomized version of the data for release. ϵ-differential privacy moves away from offering absolute guarantees involving privacy to relative guarantees. Achieving ϵ-differential privacy ensures that if a disclosure does take place, that the likelihood of such a disclosure is nearly the same, within some bound determined by the choice of ϵ, whether the individual's record is present in or absent from the database. One major advantage of assessing privacy in this manner is that it makes no assumptions about the auxiliary information known to the intruder. In practice this means that the risk of an individual choosing to participate in a database or not is nearly the same regardless of the auxiliary information possessed by the intruder. This is a very strong form of privacy which, if enforced, may prevent many of the datasets currently disseminated by organizations for research purposes from being released.

While offering strong privacy guarantees, this may come at a cost to utility. For the release of microdata, it may be the case that to achieve these privacy guarantees, the releasing organization will not be able to release any actual data and will have to settle for a fully simulated release. Even then, however, the fully simulated data are not guaranteed to achieve differential privacy. There are similar problems for query systems. It may be the case that the amount of noise that must be added to a query response to achieve differential

privacy may be so great as to render the private, released response nearly useless. However, recent work, including Smith (2008) and Dwork and Lei (2009), address the utility aspects of differentially private releases. Smith (2008) shows that, for certain parametric models, differentially private point estimators can be constructed that converge to the maximum likelihood estimator.

Dwork and Lei (2009) explore robust statistics as a means of constructing differentially private point estimators. Both of these papers are examples that privacy can be maintained while still allowing for useful statistical analysis of the data; Or as Smith (2008, p. 1) says about the result of that paper, "This provides (further) strong evidence that rigorous notions of database privacy can be consistent with statistically valid inference."

Other works leading up to differential privacy include Dinur and Nissam (2003), Dwork and Nissam (2004), and Blum et al. (2005), while Dwork and Smith (2009) and Wasserman and Zhou (2010) both present overviews of differential privacy form a statistical point of view.

A relaxed form ϵ-differential privacy, (ϵ,δ)-indistinguishability, fully described in Sect. 2.1, is proposed in Nissim et al. (2007), which addresses the problem that many release mechanisms cannot achieve ϵ-differential privacy. Machanavajjhala et al. (2008) then builds off this privacy mechanism to propose probabilistic differential privacy, which is similar to ϵ-differential privacy but does not take into account extremely rare events. The authors argue that there are certain events that are so rare that they do not wish to consider them in the assessment of privacy.

While all of these methods offer certain privacy guarantees, the interpretation of the metric is unclear. For example, how should one who releases data choose suitable values of ϵ (and/or δ) such that enough privacy is guaranteed? There will likely come a day when legal requirements are established prescribing how much privacy must be guaranteed to individuals. If this discussion centers around differential privacy, a suitable value of ϵ needs to be quantified. However, the meaning of this ϵ does not allow for easy comparisons between different privacy methods, thus, choosing a suitably protective value becomes challenging.

Below, ϵ-differential privacy is defined, and an example of a mechanism that achieves ϵ-differential privacy is given.

Let D be a database with n observations and p attributes, and let D' be a neighboring database if it differs from D by at most one observation. Consider query f to the database where $f : D \rightarrow \mathbb{R}^d$ for some d. However, rather than releasing the results of the query, $f(D)$, which could cause a privacy breach, a randomized version, $\kappa_f(D)$, of the query is released. Using this notation, Dwork (2006) defines differential privacy as follows:

Definition 1 A randomized function κ has ϵ-differential privacy if $Pr[\kappa(D) \in S] \leq e^{\epsilon} Pr[\kappa(D') \in S]$ for any two databases D and D' that differ by at most one element and for any subset $S \subseteq range(\kappa)$.

By this definition, ϵ-differential privacy is bounding the ratio of the distributions of $\kappa_f(D)$ and $\kappa_f(D')$ by e^{ϵ} for all possible D and D'. When the value of ϵ is "small", the distributions of $\kappa_f(D)$ and $\kappa_f(D')$ are close to one another. This means that the presence of absence of any observation from D will not dramatically change the distribution of $\kappa_f(D)$, ensuring a certain level of protection against an inferential attack.

For any query f, the L_1 sensitivity is defined as

Definition 2 $\Delta f = \mathrm{Max}_{D,D'} ||f(D) - f(D')||_1$.

Dwork (2006) offers the following example of a randomized release function that achieves ϵ-differential privacy: The mean of the database D, $\bar{X}$, is computed, and random

Laplace noise with mean zero and scale parameter σ (thus a corresponding variance of $2\sigma^2$) is added to the mean. This results in a randomized release function which has a Laplace distribution centered at $\bar{X}$ with scale parameter σ. The distribution of this release function is $Pr[K_f(D) = r] \propto exp\left(\frac{-||f(D)-r||_1}{\sigma}\right)$, and Dwork (2006) shows that $\frac{Pr[K_f(D)=r]}{Pr[K_f(D')=r]} \leq e^{\epsilon}$ for all D and D' when $\epsilon = \frac{\Delta f}{\sigma}$. As a result, ϵ-differential privacy can be achieved, in this case, by adding noise from a Laplace distribution with mean 0 and scale parameter $\frac{\Delta f}{\epsilon}$.

In this case, adding Laplace noise to this query, ϵ-differential privacy can be achieved. However, adding other types of noise (i.e. normal noise), ϵ-differential privacy cannot be achieved since the ratio of two normal distributions with the same variance but different means is not bounded. As such, ϵ-differential privacy should be generalized to (ϵ,δ)-indistinguishability, as in Nissim et al. (2007). There they define a negligible function $\delta()$ to be a positive function that is asymptotically smaller than any inverse polynomial: $\delta(n) = \frac{1}{n^{\omega(1)}}$. This leads to the following definition of (ϵ,δ)-indistinguishability.

Definition 3 Let $\delta = \delta(n)$ be a negligible function. A randomized function κ achieves (ϵ,δ)-indistinguishability if $Pr[\kappa(D) \in S] \leq e^{\epsilon} Pr[\kappa(D') \in S] + \delta$ for all neighboring databases D and D' that differ by at most one element and for any subset $S \subseteq range(\kappa)$.

Consider the previous example of releasing the mean of a database $\bar{X}$. If, instead of Laplace noise, one chose to release the mean with the addition of normal noise ϵ-differential privacy cannot be achieved. However, by adding noise from a normal distribution with mean 0 and variance $\sigma^2 \geq 2ln(\frac{2}{\delta})(\frac{\Delta f}{\epsilon})^2$, (ϵ,δ)-indistinguishability can be achieved.

In the example where Laplace noise is added to the results of a query, $f(D)$, it is easy to compare the relative amounts of privacy between mechanisms that both add Laplace noise to $f(D)$. As the amount of noise added to $f(D)$ increases, ϵ gets smaller, and more privacy is guaranteed.

However, two issues arise regarding the interpretation of ϵ. The first is that the choice of an appropriate ϵ is difficult. For example, if a mechanism achieves 0.1-differential privacy, is that enough privacy? Dwork (2008, p. 6) states, "The choice of ϵ is essentially a social question and is beyond the scope of this paper. That said, we tend to think of ϵ as, say, 0.01, 0.1, or in some cases, ln 2 or ln 3." The discussion of an appropriate ϵ is further complicated as the interpretation of what $\epsilon = 0.1$ means is not easy to grasp. The methods proposed in this paper do not directly solve this problem, as that choice is still a social question. However, the proposed measure has a clearer meaning than ϵ which, it is hoped, will aid in answering the question of what amount of privacy should be required.

A second issue with this method is that it may be difficult to compare the privacy of randomized release mechanisms when differing types of noise are added. Continuing with the example of releasing a query, $f(D)$, with some noise addition, consider that we wish to compare the privacy of adding Laplace noise to the privacy achieved by adding normal noise. When the randomized release function adds Laplace noise, the privacy can be measured by ϵ, as this mechanism achieves ϵ-differential privacy (with an implicit value of $\delta = 0$). However, with normal noise the privacy will be measured with ϵ and a necessary δ. With these metrics it is not clear which randomized release mechanism is better, as they cannot be directly compared. Further, one must again decide on a suitable value of δ, which is difficult as there seems to be little intuitive meaning behind this value other than, as with ϵ, smaller is more private, and thus better. The method proposed here overcomes this and allows for direct comparisons to be made between different privacy preserving methods.

3 Using the receiver-operator characteristic (ROC) curve and area under the curve (AUC) to assess privacy

Both ϵ-differential privacy and its relaxed form, (ϵ,δ)-indistinguishability guarantee that if an individual decides to participate in a database, their information will have little effect on the released values of the queries. Essentially, this means that the distribution of the randomized function based on the entire data base, $Pr[\kappa_f(D) = r]$, should not change dramatically when this distribution is based on any neighboring database, $Pr[\kappa_f(D') = r]$. If there were a dramatic change in these two distributions, it would be possible to infer that the actual values of the queries $f(D)$ and $f(D')$ were different based only the realizations from the random release functions $\kappa_f(D)$ and $\kappa_f(D')$.

With this observation, we can view database D as a population about which an intruder wishes to make inference, and, further, we can consider the realizations from $Pr[\kappa_f(D) = r]$ and $Pr[\kappa_f(D') = r]$ as data that can be used to make inference about D. As such, one can frame this problem as a hypothesis test of H_0: $f(D) = f(D')$ versus H_1: $f(D) \neq f(D')$ using the released value from the randomized functions $\kappa_f(D)$ and $\kappa_f(D')$ as data. One way in which this test could be performed is by using a likelihood ratio test (LRT). In that case, the test statistic would be based on the ratio of $\Lambda = \frac{Pr[\kappa_f(D)=r]}{Pr[\kappa_f(D')=r]}$, and the null hypothesis would not be rejected for values of Λ near 1. This form of the likelihood ratio test is exactly the ratio that must be bounded in order to achieve ϵ-differential privacy, and so in this sense, ϵ can be viewed as the largest value of a test statistic over all possible neighboring databases. Viewing ϵ as a test statistic underscores the difficulty in comparing the privacy across different type of randomized release functions, as test statistics cannot be directly compared to one another without considering their underlying distribution.

If this test has high power, that indicates that an intruder can easily distinguish between $f(D)$ and $f(D')$ based solely on the randomized released versions of the query, $\kappa_f(D)$ and $\kappa_f(D')$, creating a situation with low privacy. Therefore, good privacy will result when enough noise is added to the released query to create a test with low power. There are many ways of assessing how well a hypothesis test performs, and one such method is the receiver-operator characteristic (ROC) curve (Pepe 2003). The ROC curve is calculated by computing the power ($1 - \beta$, where β is the type II error rate) of the test for each value of α, the type I error rate, between 0 and 1. Power is defined as Pr[choosing $H_1|H_1$ is true]. In a classic hypothesis testing setting, this probability can be difficult to calculate, as we usually do not know the exact value of the population parameters about which we are making inference. Here, however, we know exactly the true values about which inference is to be made, $f(D)$ and $f(D')$, so for a given level of α, the power can be calculated exactly. Once these quantities are calculated, plotting α versus $1 - \beta$ gives us a theoretical ROC curve. A useful measure associated with the ROC curve is the area under the ROC curve (AUC).

Definition 4 Let power(α) be the power function at α. AUC $= \int_0^1$ power(α) dα.

The AUC takes on values in the interval from 0 to 1. An AUC of 1 indicates a perfect tests which is correct all of the time, and an AUC of 0 means that the test is incorrect all of the time. However, a test that is always correct can easily be constructed from a test which is always incorrect by choosing the opposite of the poor test's results. Therefore, practically, AUC takes on values in the interval 0.5 to 1 where 0.5 would indicate that the test is no better than random guessing and 1 indicates a perfect test.

The AUC has a few interesting interpretations. In our context, it can be interpreted as follows: Consider an experiment in which two values are drawn at random, one from the

distribution of the test statistic under the null and one from the distribution of the test statistic under the alternative hypothesis. We are then forced to classify these two random draws as coming from the distribution of the null hypothesis or the distribution of the alternative hypothesis. The probability of correct classification is the AUC.

Therefore, an AUC near 0.5 indicates that correctly classifying these random draws using this test is nearly the same as simply flipping a fair coin. Alternatively, an AUC near 1 indicates that this classification could be done almost perfectly indicating that it is not difficult to distinguish between the two distributions based on respective neighboring databases, D and D'. Pepe (2003, p. 78) discusses the ROC curve in terms of diagnostic testing and offers several other interpretations of the AUC from this perspective.

In most situations, we desire a test whose AUC is large. Here, however, a test with high power, and respective large AUC, indicates that when the two randomized release functions generated by D and D', respectively, are actually different, a statistical test will be able to detect this difference with high probability, which is exactly what we are trying to prevent. We desire that when the two distributions are actually different, it falsely chooses the null hypothesis, creating a type II error. Therefore, the AUC, either interpreted as the average power over all values of α or the probability of correct classification, is an intuitive, easily interpreted assessment of risk. This leads to the following definition of risk.

Definition 5 $Risk = \max_{D,D'} AUC$.

Similarly, we can define $Privacy = 1 - Risk$.

The main idea of the differential privacy framework is the comparison of two distributions. This seems to make the ROC curve a natural choice in the assessment of privacy as the ROC curve a well established statistical method for measuring the difference between two distributions (Pepe 2003, p. 81).

Addressing privacy in this manner is essentially the same as ϵ-differential privacy, as both methods use that ratio of the distributions of the randomized release functions of neighboring databases as a starting point for assessing privacy. However, in the hypothesis testing framework, the bound defined in differential privacy can be thought of as a test statistic which cannot always be compared directly to another test statistic. For example, if in one test the distribution of the test statistic is normal and the distribution of the second test statistic is χ^2 the test statistics can take on the same value but are not directly comparable. So, two different randomized release mechanisms can achieve the same level of ϵ, but, again, may not be directly comparable as they are on different scales. This problem is overcome by using the ROC curve which allows the comparison of diagnostic tests to be on the same scale (Pepe 2003, p. 72). Similarly, in our setting, the ROC curve lets the assessment of privacy occur on the same scale, which allows for direct comparisons to be made between different randomized release functions.

Using $Privacy = 1 - Risk$ as a measure of privacy is one way to compare different randomized release function to one another. We illustrate this in the next section.

4 Examples

4.1 Adding Laplace noise

Consider again a data set D and a neighboring data set D'. As an example, one may wish to release the mean of the data D. Thusly, our query is $f(D) = \bar{X}$ and $f(D') = \bar{X}'$. Due to privacy constraints, rather than release $\bar{X}$, a random draw from $Pr[\kappa_f(D) = r]$ will be

released, which we will call $\bar{X}_R$. In this case, our randomized function, κ, adds Laplace noise with scale parameter σ to the query. Therefore, here the distribution of $\kappa_f(D)$ is Laplace centered at $\bar{X}$ with scale parameter σ. The randomized released version of the mean based on the database D is then $\bar{X}_R = \bar{X} + \eta$ and the released mean based on D' is $\bar{X}'_R = \bar{X}' + \eta'$ where η, η' are distributed i.i.d. $Laplace(\sigma)$. As mentioned in Sect. 3, we wish to test H_0: $f(D) = f(D')$ versus H_1: $f(D) \neq f(D')$. Specifically, in this case we are testing $H_0 : \bar{X} = \bar{X}'$ versus $H_1 : \bar{X} \neq \bar{X}'$.

The LRT statistic for this hypothesis is

$$\Lambda = \frac{|\bar{X}_R - \bar{X}'_R|}{\sigma} \tag{1}$$

and we reject H_0 for large values or Λ. Notice that (1) is of the same form as ϵ when Laplace noise is added in Dwork (2006).

Ideally, we would want an intruder to be unable to distinguish between $\bar{X}$ and $\bar{X}'$. Therefore, if they are unable to reject this null hypothesis at some level α then no difference can be distinguished. In the ideal situation, when H_1 is correct we want to add enough noise so that the null hypothesis cannot be rejected.

In the Laplace case, the power of the test, for a given level of α, is $Pr(\Lambda_{H_1} > \Lambda_{\frac{\alpha}{2}}) + Pr(\Lambda_{H_1} < -\Lambda_{\frac{\alpha}{2}})$ where Λ_{H_1} is the test statistic under the alternative hypothesis and $\Lambda_{\frac{\alpha}{2}}$ is the critical value of the test under H_0. Often this quantity is difficult to compute as we may not know the true value of the parameters that we are making inferences about. However, in this context, we know the true values of the parameters are, $\bar{X}$ and $\bar{X}'$.

In the privacy context, we are interested in the most extreme case, when $|\bar{X} - \bar{X}'|$ is the largest over all possible neighboring data sets, which is defined previously as Δf.

Dwork (2006) shows that by adding Laplace noise you can achieve ϵ-differential privacy as long as $\sigma \geq \frac{\Delta f}{\epsilon}$ where Δf is the largest difference between $\bar{X}$ and $\bar{X}'$ over all neighboring data sets.

Figure 1 shows the associated ROC curves for $\epsilon = 0.1$, $\epsilon = 1.0$, and $\epsilon = 4.0$. Notice, as ϵ gets larger, in this case $\epsilon = 4.0$, one can see that the ROC curve stretches toward the upper left corner of the graph indicating that, when D and D' are the most different over all neighboring databases, the random released distributions based on D and D' are much different. ROC curves that are close to the 45° line, such as the ROC curve associated with $\epsilon = 0.1$, achieve a high level of privacy. Using these ROC curves, one can compute their associated AUC, which was defined as risk in Definition 5. The risk associated with the three ROC curves in Fig. 1, as well as the risks associated with several other ROC curves of different values of ϵ, are displayed in Table 1. Notice that as ϵ gets close to 0, the corresponding value of the risk gets very close to 0.5, achieving maximum privacy. This reinforces the intuition that 0.01 and 0.1 are good choices for ϵ (Dwork 2008, p. 6) in terms of confidentiality protection.

Rather than focusing on specific values of ϵ, Fig. 2a shows a graph of the risk corresponding to each ϵ between 0 and 10. The risk based on AUC increases dramatically in the range of ϵ from 0 to 5. Clearly, any reasonable value of ϵ has to come from this range, but any actual decision on how to choose an "acceptable" ϵ or risk must be decided in a social debate on privacy. Figure 2b focuses on the interval where ϵ is between 0 and 1.

Table 2 displays the risk values based on different values of Δf and σ^2. As σ^2 gets larger, the risk value tends to the optimal privacy value of 0.5. Likewise, as Δf increases the risk tends towards 1, for example, when $\Delta f = 10.0$ and $\sigma^2 = 0.1$ the risk is very nearly 1 indicating almost no privacy. At the other extreme, when $\Delta f = 0.01$ and $\sigma^2 = 10.0$, we

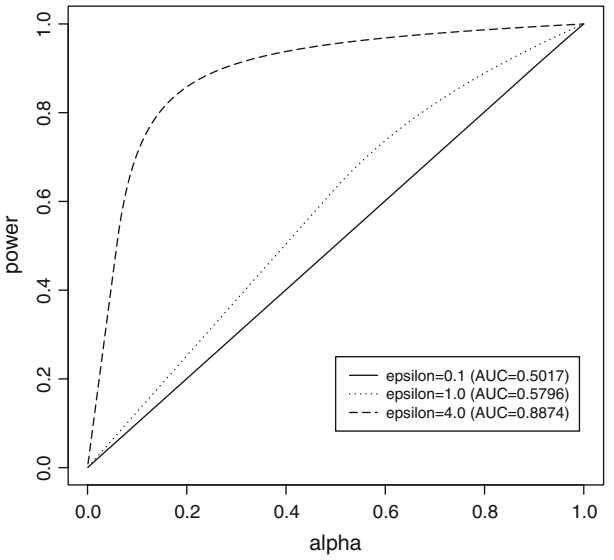

Fig. 1 ROC curves associated with selected values of ϵ

Table 1 Comparison of ϵ from differential privacy with the risk values based on Definition 5 when Laplace noise is added to the mean of a database

ϵ	0.01	0.1	0.25	0.5	1.0	2.0	4.0	8.0
Risk	0.5005	0.5017	0.5075	0.5253	0.5796	0.7070	0.8874	0.9889

High privacy is associated with small values of ϵ (close to 0) and risk (close to 0.5). Alternatively, low privacy is associated with large values of ϵ and risk (close to 1.0)

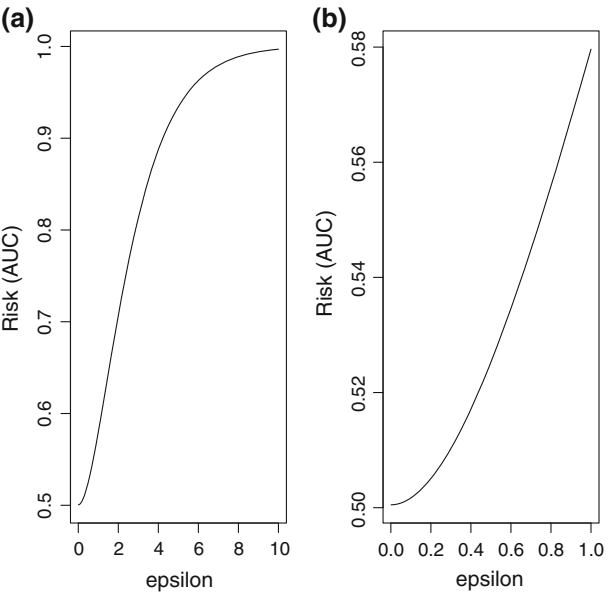

Fig. 2 A graph of the relationship between ϵ from differential privacy and risk (AUC) based on Definition 5 when Laplace noise is used to perturb the released value of the mean of a database

Table 2 Risk values associated with Δf and σ^2 values when adding Laplace noise to the mean of a database D

Δf	σ^2				
	0.1	0.5	1.0	5.0	10.0
0.01	0.5010	0.5006	0.5005	0.5005	0.5005
0.5	0.5378	0.5093	0.5051	0.5015	0.5010
1.0	0.8276	0.6326	0.5796	0.5208	0.5113
5.0	0.9692	0.7975	0.7070	0.5668	0.5378
10.0	1.000	0.9998	0.9970	0.9124	0.8276

Δf is the largest different of the means, $\bar{X}$ and $\bar{X}'$, based, respectively, on neighboring databases D and D'. σ^2 is the variance of the Laplace noise added to the released value of the mean $\bar{X}$

observe that the risk is very close to 0.5 indicating very strong privacy. However, when $\Delta f = 0.01$ adding Laplace noise with variance $\sigma^2 = 10$ is not necessary. One can see from the first row of Table 2 that there is only a very small gain in privacy by increasing σ^2 from 0.1 to 10, so very little is gained by perturbing the data with this extra noise.

4.2 Adding normal noise

Again, consider a data set D and a neighboring data set D'. We wish to release the mean of the data D, but due to privacy constraints we will release a draw from $Pr[\kappa_f(D) = r]$, where, here, $f(D)$ is the sample mean based on D, $\bar{X}$. As before, the released version of the mean is denoted $\bar{X}_R$. In the previous example, we released the mean with the addition of random Laplace noise. Here, the randomized release function will add normal noise to $\bar{X}$.

The randomized released version of the mean based on D is $\bar{X}_R = \bar{X} + \xi$ and the released mean based on D' is $\bar{X}'_R = \bar{X}' + \xi'$ where ξ, ξ' are distributed i.i.d. $normal(0, \tau^2)$. Again, we wish to test $H_0 : \bar{X} = \bar{X}'$ versus $H_1 : \bar{X} \neq \bar{X}'$.

The LRT for this hypothesis is

$$\Lambda = \frac{\bar{X}_R - \bar{X}'_R}{\tau} \tag{2}$$

and we reject H_0 for large values of $|\Lambda|$.

Let $\Phi(\cdot)$ be the standard normal cumulative distribution and $Z_{\frac{\alpha}{2}}$ be the critical value of the test under H_0 at level α. The corresponding power of this test is $1 - \Phi\left(Z_{\frac{\alpha}{2}} - \frac{\Delta f}{\sqrt{2}\tau}\right) + \Phi\left(-Z_{\frac{\alpha}{2}} - \frac{\Delta f}{\sqrt{2}\tau}\right)$ which yields $\beta = 1 - power = \Phi\left(Z_{\frac{\alpha}{2}} - \frac{\Delta f}{\sqrt{2}\tau}\right) - \Phi\left(-Z_{\frac{\alpha}{2}} - \frac{\Delta f}{\sqrt{2}\tau}\right)$.

Figure 3 displays a contour plot of the risk when normal noise is added to the mean of the data as a function of δ and ϵ. Lower risk values are located in the lower left corner of the graph, while higher risk values, near 1.0, are located as one moves in the direction of the upper right hand corner towards the bound where the risk is nearly 1.0. One can see that there are many combinations of ϵ and δ that yield the same risk, and thus the same level of privacy. Table 3 displays some exact risk values for selected values of ϵ and δ.

Table 4 displays the risk for different values of Δf and variance τ^2 when normal noise is used to perturb the output. Again, similar to Table 2, as the variance τ^2 gets larger risk tends to 0.5 indicating high amounts of privacy and as Δf gets large the risk tends to 1 indicating low levels of privacy. Comparing Tables 2 and 4, one can observe that for the

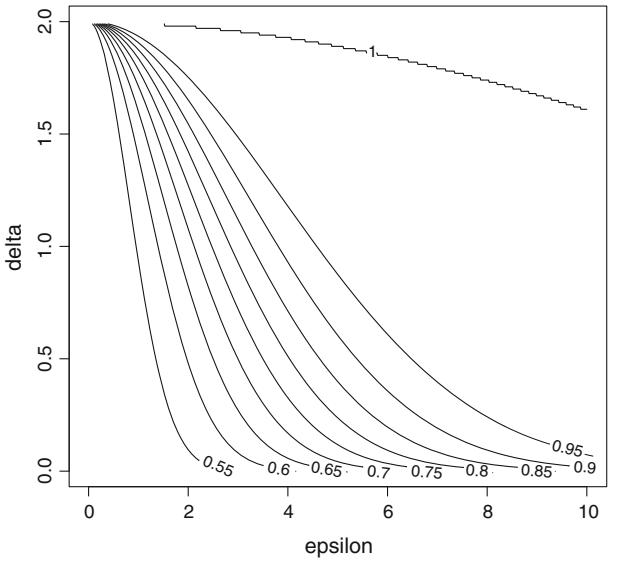

Fig. 3 Contour plot of risk as a function of ϵ and δ. The y-axis is values of δ, whereas, the x-axis is different values of ϵ. Risk values range from 0.5 to 1

Table 3 The risk values based on Definition 5 associated with different values of ϵ and δ used in (ϵ,δ)-indistinguishability

δ	ϵ							
	0.01	0.1	0.25	0.5	1.0	2.0	4.0	8.0
0.75	0.5005	0.5009	0.5030	0.5105	0.5394	0.6382	0.8589	0.9959
1	0.5005	0.5012	0.5041	0.5146	0.5546	0.6831	0.9151	0.9994
1.5	0.5005	0.5019	0.5091	0.5339	0.6207	0.8307	0.9919	0.9999

Table 4 Risk values associated with Δf and τ^2 values when adding normal noise to the mean of a database D

Δf	τ^2				
	0.1	0.5	1.0	5.0	10.0
0.01	0.5006	0.5005	0.5005	0.5005	0.5005
0.5	0.5084	0.5021	0.5013	0.5007	0.5006
1.0	0.6634	0.5387	0.5200	0.5045	0.5025
5.0	0.8931	0.6360	0.5738	0.5162	0.5084
10.0	1.000	0.9997	0.9880	0.7717	0.6634

Δf is the largest different of the means, $\bar{X}$ and $\bar{X}'$, based, respectively, on neighboring databases D and D'. τ^2 is the variance of the normal noise added to the released value of the mean $\bar{X}$

same values of Δf and when $\sigma^2 = \tau^2$ when the noise from a normal distribution is added the value of the risk is less than the value of risk when Laplace noise is added for every value in the table. Therefore, from a privacy perspective, it may be more desirable to use normal noise to perturb the query result rather than Laplace noise.

As an example, consider a situation where one wishes to release the mean of some data with noise added. Further, say that this data achieves $\Delta f = 1$, and one wants to add noise with variance equal to 1. Using the risk defined using AUC, one can see that by comparing Tables 2 and 4, adding normal noise rather than Laplace ensures a greater level of privacy, respectively, 0.5200 for normal noise versus 0.5796 for the addition of Laplace noise. Figure 4, which shows the ROC curves for the addition of normal and Laplace noise when $\Delta f = 1$ and variance equals 1, clearly indicates that the AUC, and thus risk, is smaller for the addition of normal noise. However, when α is small, one should note that the power for the addition of Laplace noise is smaller than for that of the normal distribution indicating that, for certain levels of α, the Laplace distribution ensures more privacy. Figure 4b displays this in more detail over the interval when α ranges from 0 to 0.10. Since, these two ROC curves cross one another, assessing privacy using the AUC may be problematic when comparing different types of random noise.

When comparing release mechanisms, using the AUC offers a method for making a direct comparison. Consider trying to compare two release schemes, one which releases the mean plus normal noise and the other which releases the mean plus Laplace noise, using differential privacy. Say, for example, that the Laplace mechanism achieves 0.5-differential privacy (or (0.5,0)-indistinguishability) and the normal mechanism achieves (0.25,1.5)-indistinguishability. It seems difficult to directly compare these two situations to each other using the provided metrics and make a claim about the relative levels of privacy. However, using the AUC metric we can see that with the normal noise added the AUC is 0.5091 (from Fig. 3) and with the Laplace noise added the AUC is 0.5253 (from Fig. 1).

While using the AUC allows for direct comparison between mechanisms, Fig. 4 is used to illustrate the fact that a smaller AUC does not necessarily indicate that one ROC curve is uniformly less than the other curve. In Fig. 4a there are two ROC curves, one for releasing the mean of the data plus Laplace noise and the other for the mean plus normal noise. Both these release mechanisms add noise with the same variance ($\sigma^2 = 1$), however, they have different ROC curves. The ROC curve when normal noise is added is much closer to the

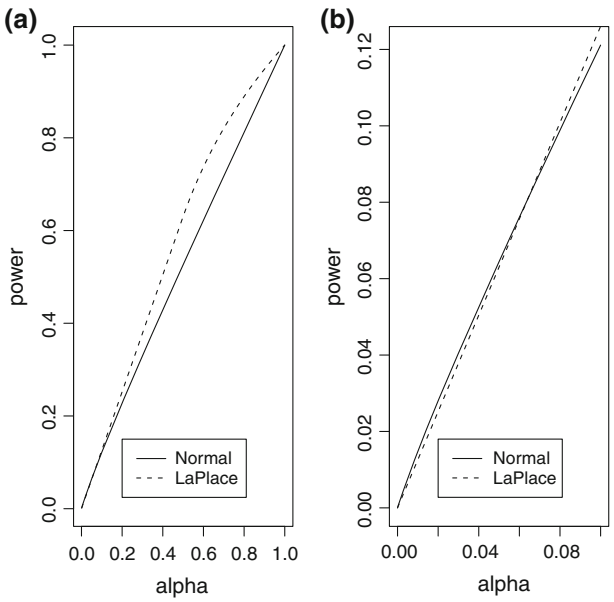

Fig. 4 ROC curves comparing normal and Laplace noise addition to database D mean when $\Delta f = 1$ and variance of both distributions is 1. Both figures are the same, but the right figure is focused in on the interval where α extends from 0 to 0.1 to illustrate that the ROC curves cross one another

45° line indicating a greater level of security, whereas, the ROC curve for the instance where Laplace noise is added extends much closer to the upper left corner of the graph indicating a better test for distinguishing between D and D' and, thusly, a lower level of privacy. However, Fig. 4b shows that for certain small values of α, the Laplace noise achieves less power, implying more privacy.

5 Discussion

Many data collecting organizations wish to release useful data to the public for research purposes. However, these organizations are bound by ethical and sometimes legal obligations to maintain the privacy of the individuals whose data they are releasing. Many methods for maintaining the privacy of the released data have been proposed and shown to provide varying degrees of utility. However, the assessment of how much privacy is ensured for these different types of secure releases has been less often discussed. Recently, the notion of ϵ-differential privacy, and its relaxed version (ϵ,δ)-indistinguishability, have been suggested.

The theoretical improvement in ϵ-differential privacy proposed in Dwork (2006), which provides a relative, as opposed to an absolute, guarantee of privacy, was a major advance in the area of statistical disclosure limitation. As such, the proposal to use the ROC curve and associated AUC metric closely follows the framework of differential privacy. However, using the ROC curve allows us to compare release mechanisms on the same scale regardless of the type of randomized release function. In the hypothesis testing framework, the ϵ in ϵ-differential privacy can be thought of as a test statistic. However, a test statistic must be evaluated along with its distribution, which means meaningful comparisons on the relative amounts of privacy across different randomized release functions cannot be made by directly comparing the two respective values of ϵ. By using the ROC curve and the AUC metric, all types of randomized release functions can be compared on the same scale, making claims about the relative amount of privacy ensured by two different types of noise addition easier. This makes the policy debate regarding the appropriate level of privacy easier. An AUC $= 0.5$ indicates that enough noise has been added to the query result that any attempt by an intruder to distinguish between a randomized release based on the entire data set and a randomized release based on a neighboring data set will be no better than random guessing. At the other extreme, an AUC $= 1.0$ indicates that an intruder will always be able to distinguish the difference between two randomized release functions, one based on all the data and the other based on a neighboring set of data.

Many federal institutions require grantees to provide a plan in their grant proposal for sharing the collected data with other researchers while still maintaining the privacy of the individuals' data. One possibility for achieving this requirement is to store the data in a database and allow queries to this database. Rather than release actual query responses, noise could be added in such a way so as to provide privacy guarantees set forth by differential privacy. However, this may compromise the utility of the query responses.

Another possible method for releasing data privately is the use of synthetic data. The utility properties of synthetic data have been discussed often, however, quantification of privacy concerning synthetic data is less often discussed. ϵ-differential privacy and its relaxation (ϵ,δ)-indistinguishability are methods for assessing the privacy of randomized versions of data, including synthetic data. However, granting institutions and researchers may have a difficult time determining what is an appropriate level of ϵ and/or δ. Further, it may be difficult for researchers to choose between privacy preserving mechanisms based

on values of ϵ and/or δ. Specifically, there are many possible methods for creating synthetic data, but their relative levels of privacy are difficult to interpret using ϵ and/or δ.

The definition of risk proposed in this paper provides an easily interpretable scale of privacy, with AUC ranging from 0.5 to 1.0, which will aid both researchers and granting institutions in deciding on appropriate levels of privacy. Also, levels of privacy of randomized release functions can now be compared to one another on the same scale. This allows researchers to choose between competing randomized release functions in a more methodological way.

One drawback to this method is that it may be difficult to calculate AUC for complicated release scenarios. All of the examples in this paper are fairly simple: the addition of some noise (i.e. normal or Laplace) to the mean of a database. As more complicated release scenarios are discussed the calculation of the AUC metric gets more complicated. However, this problem would exist when trying to compute ϵ (and δ) in the differential privacy framework, so this should not prevent us from using the AUC metric. Also, as was seen in Sect. 4.2, when comparing different types of noise with the same amount of variance, the ROC curves may cross each other. This makes using the AUC metric for comparing the two curves less desirable. As such, it may be of future interest to consider other summaries of the ROC curve, including, possibly, a weighted version of the AUC metric.

Future work could include an examination into the meaning and practical implications of achieving a certain level of AUC privacy. For instance, what are the practical implications of moving from AUC privacy of 0.51 to 0.505? What does this really mean from a practical privacy point of view? Further future work could also include applying this metric to more complicated randomized release functions, including synthetic data. In this way, different synthetic data generation techniques could be assessed based on the privacy protections they guarantee.

References

Bethlehem, J.G., Keller, W., Pannekoek, J.: Disclosure control of microdata. J. Am. Stat. Assoc. **85**, 38–45 (1990)

Blum, A., Dwork, C., McSherry, F., Nissam, K.: Practical privacy: the sulq framework. In: Proceedings of the 24th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 128–138 (2005)

Cox, L.H.: Suppression methodology and statistical disclosure control. J. Am. Stat. Assoc. **75**, 377–385 (1980)

Cox, L.H.: Matrix masking methods for disclosure limitation in microdata. Surv. Methodol. **6**, 165–169 (1994)

Dale, A., Elliot, M.: Proposals for 2001 samples of anonymized records: an assessment of disclosure risk. J. R. Stat. Soc. Ser. A **164**(3), 427–447 (2001)

Dalenius, T., Reiss, S.P.: Data-swapping: a technique for disclosure control. J. Stat. Plan. Inference **6**, 73–85 (1982)

Dinur, I., Nissam, K.: Revealing information while preserving privacy. In: Proceedings of the 22nd ACM SIGMOD-SIGACT-SIGART Symposium on Principlesof Database Systems, pp. 202–210 (2003)

Domingo-Ferrer, J., Torra, V.: Disclosure risk assessment in statistical data protection. J. Comput. Appl. Math. **164–165**(1), 285–293 (2004)

Duncan, G., Lambert, D.: The risk of disclosure for microdata. J. Bus. Econ. Stat. **7**, 207–217 (1989)

Dwork, C.: Differential privacy. In: ICALP, pp.1–12. Springer, New York (2006)

Dwork, C.: An ad omnia approach to defining and achieving private data analysis. In: Lecture Notes in Computer Science, 10 pp. Springer, New York (2008)

Dwork, C., Lei, J.: Differential privacy and robust statistics. In: Proceedings of the 41th Annual ACM Symposium on Theory of Computing (STOC), pp. 371–380 (2009)

Dwork, C., Nissam, K.: Privacy-preserving datamining on vertically partitioneddatabases. In: Advances in Cryptology: Proceedings of Crypto, pp. 528–544 (2004)
Dwork, C., Smith, A.: Differential privacy for statistics: what we know and what we want to learn. J. Priv. Confid. **1**(2), 135–154 (2009)
Keller, W.J., Bethlehem, J.G.: Disclosure protection of microdata: problems and solutions. Stat. Neerl. **46**, 5–19 (1992)
Li, N., Li, T., Venkatasubramanian, S.: t-Closeness: privacy beyond k-anonymity and l-diversity. In: IEEE 23rd International Conference on Data Engineering, 2007. ICDE 2007, pp. 106–115 (2007)
Little, R.J.A.: Statistical analysis of masked data (Disc: P455-474) (Corr: 94V10 p469). J. Off. Stat. **9**, 407–426 (1993)
Machanavajjhala A., Kifer D., Gehrke J., Venkitasubramaniam M.: L-diversity: Privacy beyond k-anonymity. ACM Trans. Knowl. Discov. Data **1**(1), 3 (2007)
Machanavajjhala, A., Kifer, D., Abowd, J., Gehrke, J., Vilhuber, L.: Privacy: theory meets practice on the map. In: International Conference on Data Engineering, April, 10 pp. Cornell University Computer Science Department, Cornell (2008)
Matthews, G.J., Harel, O., Aseltine, R.H.: Examining the robustness of fully synthetic data techniques for data with binary variables. J. Stat. Comput. Simul. **80**(6), 609–624 (2010)
Nissim, K., Raskhodnikova, S., Smith, A.: Smooth sensitivity and sampling in private data analysis. In: STOC '07: Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing, New York, NY, USA, pp. 75–84. ACM Press, New York (2007)
Paass, G.: Disclosure risk and disclosure avoidance for microdata. J. Bus. Econ. Stat. **6**(4), 487–500 (1988)
Pepe, M.S.: The Statistical Evaluation of Medical Tests for Classification and Prediction. Oxford University Press, Oxford (2003)
Raghunathan, T.E., Reiter, J.P., Rubin, D.B.: Multiple imputation for statistical disclosure limitation. J. Off. Stat. **19**(1), 1–16 (2003)
Reiter, J.P.: Satisfying disclosure restriction with synthetic data sets. J. Off. Stat. **18**(4), 531–543 (2002)
Reiter, J.P.: Releasing multiply imputed, synthetic public use microdata: an illustration and empirical study. J. R. Stat. Soc. Ser. A **168**(1), 185–205 (2005)
Rubin, D.B.: Comment on "statistical disclosure limitation". J. Off. Stat. **9**, 461–468 (1993)
Sarathy, R., Muralidhar, K.: The security of confidential numerical data in databases. Inf. Syst. Res. **13**(4), 389–403 (2002)
Skinner, C.J., Elliot, M.J.: A measure of disclosure risk for microdata. J. R. Stat. Soc. Ser. B **64**(4), 855–867 (2002)
Smith, A.: Efficient, differentially private point estimators. arXiv:0809.4794v1 [cs.CR] (2008)
Sweeney, L.: k-Anonymity: A model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl. Based Syst. **10**(5), 557–570 (2002)
Wasserman, L., Zhou S.: A statistical framework for differential privacy. J. Am. Stat. Assoc. **105**(489), 375–389 (2010)
Yancey, W.E., Winkler, W.E., Creecy, R.H.: Disclosure risk assessment in perturbative microdata protection. In: Inference Control in Statistical Databases, From Theory to Practice, pp. 135–152. Springer, London (2002)