http://dx.doi.org/10.14236/ewic/ODAK22.3

Towards a Standard Testing Data Set in Privacy

Amedeo Roberto Esposito
School of Computer and Communication Sciences
EPFL, Lausanne
Switzerland
amedeo.esposito@epfl.ch

More and more information is being rendered publicly available through open data. Consequently, the need for private mechanisms is growing. The issue of the privacy-accuracy trade-off is more prominent than ever: keeping the information private and secure can seriously hamper the performance of queries of interest. Having perfectly secure open data that no one can interrogate is a paradox against the principles upon which open data themselves were founded. But how can one test said accuracy and performance? Much like in Machine Learning, data-sets for benchmarking are becoming necessary. The best one can do without them is theoretically compare private mechanisms among themselves, while the implications of these theoretical guarantees in daily practice remain unclear. A preliminary analysis that takes ideas from theory and tries to identify the characteristics of a potential benchmark is presented in this work.

Information Measures, Differential Privacy, Privacy, Rényi Divergences, Estimation, Open Data

1. INTRODUCTION

The revolution in the field of Privacy and Security started with the groundbreaking work of Shannon [1]. Shannon laid the very first theoretical foundations of cryptography, introducing the notion of information-theoretic secrecy. He envisioned a Message Source and a Key Source. The message was enciphered using said Key, and then the resulting cryptogram was sent over a potentially interceptable channel to the destination. Denoting with M the message, K the key and E the enciphered message, then $E = f(K, M)$ for some suitably chosen function f. He then defined what is known as Perfect Secrecy. Assuming that all of these objects are random variables, the objective would be to achieve the following equality: for a given message m, $P_E(m) = P_M(m)$ where $P_M(m)$ is the "a priori" probability of message m while $P_E(m)$ represents the "a posteriori" probability of message m if the cryptogram E is intercepted. The idea is simple: if we manage to intercept the cryptogram E, our probability distribution over the messages should be identical to the probability distribution we had before, the prior P_M. *I.e.*, the cryptogram is not giving us any **new information** on the message. To achieve this type of secrecy, one needs the (entropy of the) key to be as large as the (entropy of the) message. Thus, perfect secrecy comes with a cost that can rarely be afforded. Relaxations of this notion of secrecy have been introduced over the years based on Information Theory [2; 3; 4] and not [5; 6; 7]. The desiderata have also meaningfully changed and grown. The need for privacy (rather than simple secrecy) was born to protect part of the information (personal or sensitive information, medical data, etc.) while allowing for the possibility of answering statistical questions on the data. A general framework in this setting is represented by Differential Privacy, which enables to answer aggregate queries about a database while keeping individual records private. Dozens of mechanisms inspired by Differential Privacy have appeared throughout the years, along with several implementations. Some of these contributions are purely theoretical, and some are extremely applied. The two worlds, however, communicate very little, if at all. The issue has become even more prominent with the rise of open data: a paradigm that has at its very core the principle of rendering information and data public so that anyone can access it through a variety of queries. This is a firmly applied field working with a lot of potentially sensitive information. The default in these settings is to privatise as much as possible, with the downside of rendering data effectively useless to any user who might want to interrogate it. New algorithms that offer less restrictive guarantees than the privacy/security approaches mentioned just above are thus being

built. Some favour the queries, some favour the privacy, but little to no comparison is being done for the lack of a framework that allows doing so. This paper aims at underlining the limitations and dangers of pursuing the two directions separately. The questions we will explore are the following: how can one test the properties of a new private theoretical/applied framework? How many samples are needed to guarantee statistical significance? Are these samples accessible?

2. BACKGROUND

A probabilistic framework will be considered with a specific focus on Differential Privacy and Information-theoretic measures of dependence. Random variables are denoted with capital letters X, Y, M, realisations of said random variables will be represented with lower-case letters x, y, m. The corresponding probability measures will be denoted with P_X, P_Y, P_M.

2.1. Rényi's Divergences

Introduced by Rényi to generalise the concept of Entropy and KL-Divergence, the Rényi's α-Divergence has found many applications over the years in hypothesis testing, guessing, and several others statistical inference problems [8; 9]. It has several helpful operational interpretations (e.g., the number of bits by which a mixture of two codes can be compressed, the cut-off rate in block coding and hypothesis testing [9; 10]). It can be defined as follows [9]:

Definition 1. Let $(\Omega, \mathcal{F}, \mathcal{P}), (\Omega, \mathcal{F}, \mathcal{Q})$ be two probability spaces. Let $\alpha > 0$ be a positive real different from 1. Consider a measure μ such that $\mathcal{P} \ll \mu$ and $\mathcal{Q} \ll \mu$ (such a measure always exists, e.g. $\mu = (\mathcal{P} + \mathcal{Q})/2$)) and denote with p, q the densities of $\mathcal{P}, \mathcal{Q}$ with respect to μ. The α-Divergence of $\mathcal{P}$ from $\mathcal{Q}$ is defined as follows:

$$D_\alpha(\mathcal{P}\|\mathcal{Q}) = \frac{1}{\alpha - 1} \log \int p^\alpha q^{1-\alpha} d\mu. \qquad (1)$$

Remark 1. The definition is independent of the chosen measure μ whenever $\infty > \alpha > 0$ and $\alpha \neq 1$. It is indeed possible to show that $\int p^\alpha q^{1-\alpha} d\mu = \int \left(\frac{q}{p}\right)^{1-\alpha} d\mathcal{P}$, and that whenever $\mathcal{P} \ll \mathcal{Q}$ or $0 < \alpha < 1$ one has that $\int p^\alpha q^{1-\alpha} d\mu = \int \left(\frac{p}{q}\right)^{\alpha} d\mathcal{Q}$, see [9].

It can be shown that if $\alpha > 1$ and $\mathcal{P} \not\ll \mathcal{Q}$ then $D_\alpha(\mathcal{P}\|\mathcal{Q}) = \infty$. The behaviour of the measure for $\alpha \in \{0, 1, \infty\}$ can be defined by continuity. In particular, we have that $\lim_{\alpha \to 1} D_\alpha(\mathcal{P}\|\mathcal{Q}) = D(\mathcal{P}\|\mathcal{Q})$, *i.e.*, the classical Kullback-Leibler divergence. We refer the reader to [9] for an extensive treatment of α-Divergences and their properties. One could also consider other information measures, *e.g.*, f–Divergences or even measures of dependence like Sibson's α-Mutual Information. However, for this treatise, we will only consider D_α.

2.2. Differential Privacy

The idea behind Differential Privacy is to provide mechanisms that "obfuscate individual identities"

Definition 2. Let $\mathcal{M} : \mathcal{X}^n \to \mathcal{Y}$ be a randomised algorithm. $\mathcal{M}$ is said to be (ϵ, δ)-Differentially Private if for every $S \in \mathcal{Y}$ and every pair of vectors $x^n, \hat{x}^n$ that differ only in one position:

$$\mathbb{P}(\mathcal{M}(x^n) \in S) \leq e^\epsilon \mathbb{P}(\mathcal{M}(\hat{x}^n) \in S) + \delta. \qquad (2)$$

We can now clarify what we meant with "obfuscate individual identities". Let $x^n = (x_1, \ldots, x_i, \ldots, x_n)$ and $\hat{x}^n = (x_1, \ldots, \hat{x}_i, \ldots, x_n)$, the two vectors differ only in the i-th position (all the other x_j's are equal). If $\mathcal{M}$ is Differentially-Private, we have that the probability of the output of $\mathcal{M}$ when x^n is given as an input *is not too far* from the probability of the output of $\mathcal{M}$ when $\hat{x}^n$ is given as an input. The consequence of this is: if an adversarial party looks at the outcome of the algorithm $\mathcal{M}$ it cannot decide with too much certainty whether x_i was included in the input or not. The two outputs are *statistically indistinguishable*. When $\delta = 0$ then Definition 2 boils down to ϵ-DP. Widespread ways of achieving differential privacy typically require considering functions $f : \mathcal{X}^n \to \mathbb{R}$ with "bounded sensitivity" (*i.e.*, functions whose outcomes do not get too far if the inputs are close) and the addition of Laplacian/Gaussian/Exponential noise whose variance depends on the sensitivity of the function f, ϵ and δ. *E.g.*, denoting with Δf the sensitivity of f (cf. [5]) then the algorithm $\mathcal{M}$ defined as follows

$$\mathcal{M}(x^n) = f(x^n) + L, \text{ where } L \sim \text{Lap}(\Delta f/\epsilon) \qquad (3)$$

is ϵ-DP [5, Theorem 3.6].

3. FROM THEORY TO PRACTICE

3.1. A fair confrontation

Given any mechanism, as easy as an ϵ-DP algorithm implemented through the addition of (Laplace/Gaussian) noise, can its performances be tested in practice? How can one empirically analyse the privacy-utility trade-off? To do this, test databases need to be rendered available. Having access to an ensemble of standardised data sets would allow for direct comparison of performances on well-known queries, much like in the Machine Learning community with MNIST and other basic data-sets often used as a benchmark

for new algorithms. Without such a standardised setting for testing performances, any new algorithmic breakthrough cannot be easily compared to the previously existing ones. This discourages new contributions, especially in bridging the gap between theory and practice. Most of the literature revolves around the notion of Differential Privacy [5] , Attribute Privacy [6] or the recently defined Pufferfish Privacy [7]. Whenever a new type of privacy/secrecy appears (regardless of the literature it is originated from) if it cannot be directly compared with any of these DP-based approaches, it is automatically discredited and set aside. With a standardised testing and performance evaluation framework, these new contributions could be directly compared to others.

3.2. The size of the data-set

The characteristics of an ideal test-set are not evident *a priori*. They depend on the type of query one is trying to render available. The first question that needs to be asked is: how large should such a data-set be? The natural response would be: "as large as possible". Can one do better?

Denote with $\mathcal{X}^n$ the space of data and with $Q : \mathcal{X}^n \to \mathbb{R}^m$ (with usually $m = 1$) the query to be asked on the data-set. Our purpose is to provide a privacy-enhancing mechanism $\mathcal{M} : \mathcal{X}^n \to \hat{\mathcal{X}}^n$ such that if X^n is a data-set that we are willing to protect, then $Q(X^n)$ is as close as possible to $Q(\mathcal{M}(X^n))$. The qualifier "as close as possible" can be defined in a number of ways. *E.g.*, one could ask for every $x \in \mathcal{X}^n$ and $\hat{x} \in \mathcal{X}^n$ to be such that $N(Q(x) - Q(\hat{x})) < \eta$ for some norm $N : \mathbb{R}^m \to \mathbb{R}^+$ and some $\eta > 0$. This would be too restrictive. Moreover, since we are dealing with random objects, it is perhaps (more) reasonable to ask for $Q(X^n)$ and $Q(\mathcal{M}(X^n))$ to be close "in probability" or in expected value. One would thus focus only on objects with "large enough probability of occurring" rather than on every instance $x^n \in \mathcal{X}^n$. Let us specify this a bit more.

Denote with $Y = \mathcal{M}(X^n)$ then $\mathcal{M}$ induces a Markov Kernel $\{K(\hat{x}|x) : x \in \mathcal{X}^n, \hat{x} \in \hat{\mathcal{X}}^n\}$ and a probability measure ν over $\hat{\mathcal{X}}^n$. This is typically denoted as $\nu(\hat{x}) = \mu K(\hat{x}) = \sum_x \mu(x)K(\hat{x}|x)$.

Assume, as a first step, that our purpose is to compute some value $\bar{X}$ that we know we can get arbitrarily close to in probability, using $Q(X^n)$. This represents a reasonable assumption, as if approximating $\bar{X}$ through X^n is not possible then applying a privacy-enhancing mechanism would only make things worse. By enforcing privacy we are adding noise and obfuscating data, our performances can only get worse. Thus, assume that

$$\mu\left(|Q(X^n) - \bar{X}| \geq \eta\right) \leq f_\eta(n)$$

with $f_\eta(n)$ decreasing in n for a given $\eta > 0$. This formalises the idea that the more data samples we see, the better we can estimate $\bar{X}$. Typical accuracy requirements ask for $f_\eta(n) \leq \delta$ with δ fixed beforehand. If f_η is invertible, this would immediately imply that one needs $n \geq f_\eta^{-1}(\delta)$, thus providing a lower-bound on the number of samples **necessary** to achieve the accuracy η with confidence δ. An example of this would be the following. Suppose that X^n is a sequence of iid random variables distributed according to ξ (thus, $\mu = \xi^{\otimes n}$) and that $\bar{X} = \mathbb{E}_\xi[X]$. Assume also that $0 \leq X_i \leq 1$ almost surely, then one could pick $Q(X^n) = \frac{1}{n}\sum_{i=1}^n X_i$ and have that $f_\eta(n) = 2\exp(-2n\eta^2)$, *i.e.*,

$$\mu\left(|Q(X^n) - \bar{X}| \geq \eta\right) \leq 2\exp(-2n\eta^2)$$

[11]. If, for a given $\eta > 0$ and $\delta > 0$, the requirement is that

$$\mu\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \mathbb{E}[X]\right| \geq \eta\right) \leq \delta$$

then, to ensure the confidence δ and accuracy η it is necessary to have a number of samples n such that

$$n \geq \frac{1}{2\eta^2}\log\left(\frac{1}{\delta}\right). \tag{4}$$

If for a given query Q and accuracy η, the required confidence δ is known, this approach provides the minimum number of samples needed to estimate $\bar{X}$ in the absence of privacy. Adding privacy to the mix will make the task of approximating $\bar{X}$ harder. Regardless of the specific method used to enforce privacy, the increased number of required samples can be somewhat quantified.

Theorem 1. *Let $X \sim \xi$ and consider $X^n \sim \mu$. Let $Q : \mathcal{X}^n \to \mathbb{R}$ be a query and assume that $\mathcal{M} : \mathcal{X}^n \to \mathcal{X}^n$ is a privacy-enforcing mechanism that induces a Markov Kernel denoted by K. Assume that $\mu(|\frac{1}{n}\sum_{i=1}^n X_i - \xi(X)| \geq \eta) \leq \exp(-2n\eta^2)$. Then one has that, in order to obtain the same accuracy η with confidence δ after applying $\mathcal{M}$ to X^n, the number of samples necessary n is such that for every $\alpha \in [1, +\infty]$*

$$n \geq \frac{D_\alpha(\mu K\|\mu) - \log(\delta^{\frac{\alpha}{\alpha-1}})}{2\eta^2}. \tag{5}$$

Taking the limit of $\alpha \to \infty$ leads to

$$n \geq \frac{D_\infty(\mu K\|\mu) + \log\left(\frac{1}{\delta}\right)}{2\eta^2}. \tag{6}$$

Proof. Our purpose is to verify how far is $Q(\mathcal{M}(X^n))$ from $\bar{X} = \xi(X)$ in probability. Hence, given that

$X^n \sim \mu$ and that $\mathcal{M}(X^n) \sim \mu K$ we wish to bound $\mu K(|Q(\mathcal{M}(X^n)) - \bar{X}| \geq \eta)$. Denoting with $E = \{|Q(Y^n) - \bar{X}| \geq \eta\}$ for a given $\eta > 0$, the following is true whenever $\mu K \ll \mu$:

$$\mu K(|Q(Y^n) - \bar{X}| \geq \eta) = \mu K(\mathbb{1}_E) \tag{7}$$

$$= \mu\left(\mathbb{1}_E \frac{d\mu K}{d\mu}\right) \tag{8}$$

$$\leq \mu(|Q(X^n) - \bar{X}| \geq \eta)^{\frac{\alpha-1}{\alpha}} \cdot \exp\left(\frac{(\alpha-1)}{\alpha} D_\alpha(\mu K \| \mu)\right) \tag{9}$$

$$\leq \exp(-2n\eta^2)^{\frac{\alpha-1}{\alpha}} \cdot \exp\left(\frac{(\alpha-1)}{\alpha} D_\alpha(\mu K \| \mu)\right) \tag{10}$$

were Equation (9) follows from Hölder's inequality and Equation (10) follows from our assumption. This implies that, if one wants Equation (10) to be bounded by the confidence δ, denoting $\frac{\alpha-1}{\alpha} = \frac{1}{\beta}$, the following has to happen:

$$n \geq \frac{D_\alpha(\mu K \| \mu) - \log(\delta^\beta)}{2\eta^2}. \tag{11}$$

Taking the limit of $\alpha \to \infty$, one obtains Equation (6). □

This characterisation leads to an explicit analysis of the so-called "privacy-accuracy" trade-off: $D_\alpha(\mu K \| \mu)$ measures how far is the distribution induced by $\mathcal{M}$ from the original distribution μ. The larger this quantity is, the more "private" $\mathcal{M}$ will be. However, this will require more samples to estimate $\bar{X}$ with accuracy η and confidence δ. Equations (5) and (6) show how the number of samples grows as a function of η, δ and of our information-theoretic measure of "privacy". Another important observation is that $D_\alpha(\mu K \| \mu) \geq 0$ for every $\alpha > 0$ and is equal to 0 if and only if $\mu K = \mu$. This is intuitive: if the Mechanism is not altering the probabilities (and thus, not enforcing privacy at all) one should fall back on the non-private setting (cf. Equation (4)) and since $D_\alpha(\mu K \| \mu)$ would be 0 for every $\alpha > 0$ (including $\alpha = \infty$) that would be the case. The take-home message from this analysis is: given a query Q and an acceptable level of accuracy η and confidence δ for this query, when the samples are drawn at random and observed without any privacy/obfuscation mechanism one can deduce (under mild assumptions on Q) the minimum number of samples required to achieve η and δ. The number of samples will naturally grow whenever a privacy-enforcing mechanism $\mathcal{M}$ is introduced (with a corresponding Markov Kernel K). From Theorem 1, analysing the algorithm $\mathcal{M}$, one can then deduce (through $D_\alpha(\mu K \| \mu)$) the increase in the number of samples as a function of the mechanism itself. Notice that, even though D_α is increasing in α for a given pair of measures, picking a smaller α will reduce D_α however, it will also imply an increase in the $-\log(\delta^{\frac{\alpha}{\alpha-1}})$ leading thus a trade-off between the two quantities. This is also exemplified in Equation (6) as $D_\infty(\mu K \| \mu) > D_\alpha(\mu K \| \mu)$ for every $1 < \alpha < \infty$ while the additive term $-\log(\delta)$ is the smallest one can achieve in the family $-\frac{\alpha}{\alpha-1}\log(\delta)$. The minimum sample complexity for the query Q would thus be

$$\min_{1<\alpha\leq\infty} \frac{D_\alpha(\mu K \| \mu) - \log(\delta^{\frac{\alpha}{\alpha-1}})}{2\eta^2}. \tag{12}$$

Using D_α or other information measures to quantify the information leakage is not new, and, as we have seen in Section 1, it dates back to Shannon himself. The Rényi's divergences are strongly connected to Differential Privacy [12]. Moreover, they have also been used to provide a relaxation of DP itself [13]. For instance, in the context of Theorem 1 and assuming ϵ-DP one can prove the following.

Theorem 2. *Let $X^n \sim \mu$ and let $\nu = \mu K$ with K the Markov Kernel induced by a private mechanism $\mathcal{M}$. Assume that $\mathcal{M}$ is $\epsilon-$DP. Then, given any input vector, $\hat{x}^n \in \mathcal{X}^n$ and any $\alpha > 1$ one has that*

$$D_\alpha(\mu K \| \mu) \leq (n\epsilon)\frac{\alpha}{\alpha-1} \cdot D_\alpha(K(\cdot|\hat{x}^n) \| \mu). \tag{13}$$

To conclude the discussion on the sample complexity of private queries, let us state a more general setting. The most general statements of concentration of measures usually involve a sequence of iid random variables X^n and smooth (generally Lipschitz, bounded or with bounded differences) functions of these random variables. Given that we are trying to compute a function (query) of X^n and, considering again that the best we can do is in the absence of privacy, let us assume that Q is a separately convex Lipschitz function. One can then show that if the norm of the gradient of Q is such that $\|\nabla Q\| \leq \frac{1}{n}$, then the following holds true [11]:

$$\mu(Q(X^n) - \mu(Q(X^n)) \geq \eta) \leq \exp\left(-\frac{n\eta^2}{2}\right). \tag{14}$$

Like before, this represents the golden standard *i.e.*, the best convergence one can hope for in the absence of privacy and for a given family of queries. Enforcing privacy through $\mathcal{M}$ can only give worse performances. Thus we can provide the following result, along the lines of Theorem 1

Corollary 1. *Assume that $Q : \mathcal{X}^n \to \mathbb{R}$ is a separately convex Lipschitz function and such that $\|\nabla Q\| \leq \frac{1}{n}$. Let $\mathcal{M} : \mathcal{X}^n \to \mathcal{X}^n$ be a privacy-enforcing mechanism. Then one has that for every $\alpha \in [1, +\infty]$*

$$n \geq \frac{2(D_\alpha(\mu K \| \mu) - \log(\delta^{\frac{\alpha}{\alpha-1}}))}{\eta^2}. \tag{15}$$

3.3. What about fictitious data sets?

A common theme in the literature is the testing of the performances of a private mechanism on fictitious data sets. Several, more or less elaborated models exist with this purpose. Usually, the data is created using a vast number of samples and considering, quite often, more or less correlated Gaussian samples (*e.g.*, [14, Section 5]). This could, perhaps, represent a starting benchmark but it cannot be the only one. For instance, given a Gaussian Random variable with an unknown mean, the problem of estimating the mean from random samples has been amply studied [15; 16]. It is well known that if the observations are noisy and the noise is Gaussian itself, then the empirical mean of the samples is a sufficient statistic and also the optimal estimator. One can easily characterise the optimal number of samples required to estimate the mean with arbitrary accuracy and then compare it with the (empirical) number of the samples needed to estimate it, if the noise is constructed (or added on top of the Gaussian) so that the observations are "private". However, when doing this type of validation experiments, one incurs two dangers:

- access to an unlimited number of samples;
- concentration of measure for Gaussians and (smooth) functions of Gaussians is, essentially, as good as it gets.

Searching the concentration of measure or the estimation theory literature, most of the settings that are theoretically analysable consist in the concentration of (Lipschitz functions, with bounded gradient or differences, of) Gaussian random variables around their mean/median. Moreover, said concentration is usually of the fastest kind: exponential with the number of samples. This, together with the fact that one can generate an unlimited number of samples, will typically lead to an underestimate of the adequate number of samples needed in practice. In turn, this leads to non-representative performances of the privacy-enhancing mechanism in real-world settings. While a good starting point, synthetic data sets should not be the only tool to compare different mechanisms. This reinforces the need for standard (and large enough) data sets to test the performances of privacy-enhancing mechanisms.

4. CONCLUSIONS

The problem of the the privacy-accuracy trade-off in open data was considered. Open data are meant to be interrogated. Rendering them too private leads to terrible performances in terms of queries. Researchers need to be able to test the performances of their algorithms as they try to ensure both the accuracy of the queries and privacy. In order for this to happen, the creation of benchmark data sets, similarly to MNIST in Machine Learning, is envisioned. A preliminary theoretical analysis of the characteristics of such a data set is presented with an argument on why fictitious samples are generally not representative enough of the real performances of the algorithms.

REFERENCES

[1] C. E. Shannon, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

[2] I. Issa, A. B. Wagner, and S. Kamath, "An operational approach to information leakage," *IEEE Transactions on Information Theory*, vol. 66, no. 3, pp. 1625–1657, 2020.

[3] Y. Y. Shkel, R. S. Blum, and H. V. Poor, "Secrecy by design with applications to privacy and compression," *IEEE Trans. Inf. Theory*, vol. 67, no. 2, pp. 824–843, 2021.

[4] M. Alvim, K. Chatzikokolakis, C. Palamidessi, and G. Smith, "Measuring information leakage using generalized gain functions," in *Computer Security Foundations Symposium (CSF), 2012 IEEE 25th*, June 2012, pp. 265–279.

[5] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, 2014.

[6] W. Zhang, O. Ohrimenko, and R. Cummings, "Attribute privacy: Framework and mechanisms," *CoRR*, vol. abs/2009.04013, 2020.

[7] D. Kifer and A. Machanavajjhala, "Pufferfish: A framework for mathematical privacy definitions," *ACM Trans. Database Syst.*, vol. 39, no. 1, jan 2014.

[8] S. Verdú, "α-mutual information," in *2015 Information Theory and Applications Workshop, ITA 2015, San Diego, CA, USA, February 1-6, 2015*, 2015, pp. 1–6.

[9] T. van Erven and P. Harremoës, "Rényi divergence and kullback-keibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 3797–3820, July 2014.

[10] I. Csiszar, "Generalized cutoff rates and Rényi's information measures," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 26–34, Jan 1995.

[11] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[12] P. Cuff and L. Yu, "Differential privacy as a mutual information constraint," *CoRR*, vol. abs/1608.03677, 2016.

[13] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, 2017, pp. 263–275.

[14] C. Dwork, V. Feldman, M. Hardt, T. Pitassi, O. Reingold, and A. Roth, "Generalization in adaptive data analysis and holdout reuse," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*. Cambridge, MA, USA: MIT Pressf, 2015.

[15] D. Guo, Y. Wu, S. S. Shitz, and S. Verdú, "Estimation in gaussian noise: Properties of the minimum mean-square error," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 2371–2385, 2011.

[16] L. Chisci and A. Farina, *Survey on Estimation*. London: Springer London, 2013, pp. 1–24.