# An Information-Theoretic Analysis of the Randomized Response Algorithm

Daniel Eftekhari
*Department of Computer Science*
*University of Toronto*
defte@cs.toronto.edu

***Abstract*—We examine and quantify the trade-off between information extraction and privacy preservation in the context of the randomized response (RR) algorithm for survey questionnaires. We achieve this by formulating an optimization problem, defined over transition probability values, that trades off the two complementary goals. We furthermore extend our analysis to the multiuser setting by examining how correlations between multiple RRs' transition probability values can affect the solution of the optimization problem. The findings provide new insights for data-driven applications where there is a need for balancing informativeness with privacy preservation.**

***Index Terms*—multiuser information theory, differential privacy, correlated channels, randomized response**

## I. Introduction

In the analysis of privacy-preserving algorithms, differential privacy (DP) serves as a framework for quantifying the risk of a user's identity being compromised, when statistics are derived from a dataset they have contributed to [1] [2]. The goal of DP is to minimize the risk of any individual's identity being compromised, while still enabling useful aggregate statistics to be reported. An effective technique for increasing privacy is randomization. Here, noise (in some form) is added to data points, so that it is more difficult for a potential attacker to identify a particular user's data point, but without destroying the fidelity of aggregate statistics derived from the dataset.

In contrast, in communication systems noise serves as an obstacle for information transmission. Noise reduces the rate of communication between transmitter and receiver, as measured by their mutual information [3] [4]. This contrast in the role noise plays, as a means to increase privacy on the one hand, and as an obstacle for information extraction on the other, hints at an interesting trade-off between the two objectives.

In this work we explore this trade-off. We do so in the context of the randomized response (RR) algorithm for survey questionnaires, a local[1] DP algorithm, for which we provide a brief background on (together with DP) in Section II. In Section III we show that the competing goals of information extraction and privacy preservation can be used to formulate an optimization problem in which we optimize over the noise rate for the survey question. The solution to the optimization problem maximizes the information extracted from the survey's response and minimizes privacy degradation for individual users. In Section IV we furthermore setup and examine the trade-off between information extraction and privacy preservation, when multiple survey responses are available for the same user. We do so using a multiuser information-theoretic [5] analysis in which we optimize over the noise rates of different survey questions. We review related work in Section V, and briefly conclude in Section VI.

[1]A local DP algorithm is one in which even the data curators do not collect the true responses, but only the noisy versions.

## II. Differential Privacy Background

An algorithm $\mathcal{A}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is $(\epsilon, \delta)$-DP if $\forall \mathcal{S} \subseteq \text{Range}\left(\mathcal{A}\right) \wedge \forall x, x' \in \mathbb{N}^{|\mathcal{X}|}$ s.t. $\|x - x'\| \leq 1$,

$$\mathbb{P}\left(\mathcal{A}\left(x\right) \in \mathcal{S}\right) \leq e^{\epsilon}\, \mathbb{P}\left(\mathcal{A}\left(x'\right) \in \mathcal{S}\right) + \delta. \tag{1}$$

The condition $\|x - x'\| \leq 1$ is often referred to as $(x, x')$ being neighboring datasets. The condition is equivalent to stating that two datasets $x, x'$ are neighboring if at most one row between the two datasets differs.

The prototypical randomization algorithm in DP is randomized response (RR) [6]. Users are presented with a yes/no survey question to which they answer as follows: flip a (possibly biased) coin, if heads, respond truthfully; if tails, flip a second coin and answer "yes" if tails, and "no" if heads. A simple analysis of the setup shows that despite the intrinsically noisy nature of the responses, an accurate estimate of the true proportion of "yes" and "no" responses can be obtained. This algorithm is effective because each user maintains plausible deniability – the ability to claim their answer was purely randomly generated – while the actual summary statistics can be estimated accurately.

Here we review the DP analysis of RR to provide a basis for later analyses.[2] Let $(\alpha_1, \alpha_2)$ be the probabilities of tails for the two coins, and denote $X$ as the true survey response and $Y$ as the outcome of RR. Thus $\mathbb{P}\left(Y = 1 | X = 0\right) = \alpha_1 \alpha_2$

[2]Note that for the remainder of this work, we consider the case where $\delta = 0$ as is common in RR, so that by convention we consider $\epsilon$-DP algorithms for privacy level $\epsilon$.

and $\mathbb{P}(Y=1|X=1) = 1-\alpha_1+\alpha_1\alpha_2$.

The RR algorithm is $(\ln 3)$-DP for $(\alpha_1, \alpha_2) = (1/2, 1/2)$. To show this, it suffices to evaluate $\frac{\mathbb{P}(Y=1|X=1)}{\mathbb{P}(Y=1|X=0)}$, as these values for $X, Y$ maximize the ratio of probabilities in Equation 1:

$$\begin{aligned} e^{\epsilon} &= \frac{1-\alpha_1+\alpha_1\alpha_2}{\alpha_1\alpha_2} \\ \therefore \epsilon &= \log\left(1+\frac{1-\alpha_1}{\alpha_1\alpha_2}\right). \end{aligned} \tag{2}$$

Then letting $(\alpha_1, \alpha_2) = (1/2, 1/2)$, we have $\epsilon = \ln 3$.

Examining a few other special cases is instructive. Clearly if we set $\alpha_1 = 1$, we should obtain perfect privacy, as this corresponds to never reporting the survey respondent's true answer; instead only reporting the result of the second coin flip, where the proportion of the randomly generated 0s and 1s is determined by $\alpha_2$. As expected $\alpha_1 = 1$ indeed leads to perfect privacy preservation ($\epsilon = 0$). In contrast, setting $\alpha_1 = 0$ means always reporting the true survey response, thus destroying privacy completely. This intuition aligns with $\alpha_1 = 0$ leading to $\epsilon = \infty$.

Interestingly, setting $\alpha_2 = 0$ also compromises privacy completely, since $\epsilon = \infty$. The reason is that although an output of $Y = 0$ represents uncertainty in whether the original response was $X = 0$ or $X = 1$, an output of $Y = 1$ is only possible if $X = 1$, which compromises privacy completely.

## III. Information Theory and Differential Privacy

How do the DP and RR settings relate to information theory? To see how, Figure 1 shows a variation of the binary symmetric channel (BSC) [3] which is equivalent to the RR response algorithm for a given user. Multiple uses of the channel are akin to multiple users answering the same survey question. This suggests the insight and machinery of information theory [4] can be used in the analysis of RR.

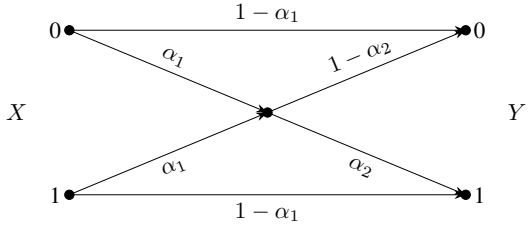


Fig. 1: A variation of the binary symmetric channel, equivalent to the randomized response (RR) algorithm. The transition probabilities of the two coin flips are represented by $\alpha_1, \alpha_2$.

Moreover, we can view the goals of DP as presenting a complementary goal to those of information extraction: here the information extracted from the survey should be informative (as measured by mutual information), but without compromising user privacy.

We next cast this trade-off between information extraction and privacy-preservation as an optimization problem.

### A. Trade-off Between Information Extraction and Privacy Preservation

In information theory, we are conventionally interested in optimizing the input probability distribution for the given error (transition) probabilities of a communication channel. However, in the survey respondent setting, the input is already determined by the user's true response.[3] Thus here we are instead interested in optimizing the *channel transition* probability values. Optimizing either $I(X;Y)$ or $\epsilon$ in isolation over $\alpha_1$ leads immediately to the trivial solutions $\alpha_1 = 0$ and $\alpha_1 = 1$, respectively.[4] Thus only examining the trade-off between information extraction and privacy preservation is of interest.

We cast this trade-off using the following optimization problem:[5][6]

$$G = \max_{\alpha_1} \frac{I(X;Y)}{\epsilon}. \tag{3}$$

Here we have free variables $p_0 = \mathbb{P}(X=0)$ and $\alpha_2$. In practice we will minimize $G^{-1}$, and add a small constant to the denominator for numerical stability. We use the L-BFGS-B algorithm [7] [8] with bound constraint $\alpha_1 \in (0, 0.5)$ for the minimization.

Figure 2 demonstrates that the optimization problem leads to interesting solutions in the trade-off between information extraction and privacy preservation.

## IV. Multiuser Information Theory and Differential Privacy

### A. Multiuser Information Theory Correspondence

Now consider a participant who answers multiple $(K)$ survey questionnaires, each involving yes or no responses. Denote the actual (source) responses as $X_1, X_2, \ldots, X_K$, and the output of the RR mechanism for the survey questions as $(Y_1, Y_2, \ldots, Y_K)$. Consider an adversary who eavesdrops on the RR output of each survey question, so that they collectively have access to the sequence $Y = (Y_1, Y_2, \ldots, Y_K)$. Similar to Section III, the adversary is interested in identifying the true responses $(X_1, X_2, \ldots, X_K)$ from the noisy output $Y$. Figure 3 illustrates the setting schematically.

[3] Furthermore, the probability of the input values may, for example, be derived based on external studies concerning the survey questions.

[4] For the remainder of the work we will only consider optimizing $\alpha_1$, as the optimization problem over $\alpha_2$ is trivial.

[5] Note that defining an additive/subtractive optimization problem, such as $G = \max_{\alpha_1} I(X;Y) - \epsilon$, would not be of value; this is due to the differing scales of $I(X;Y)$ and $\epsilon$ and due to the unboundedness of $\epsilon$.

[6] We measure $I(X;Y)$ in bits rather than nats.

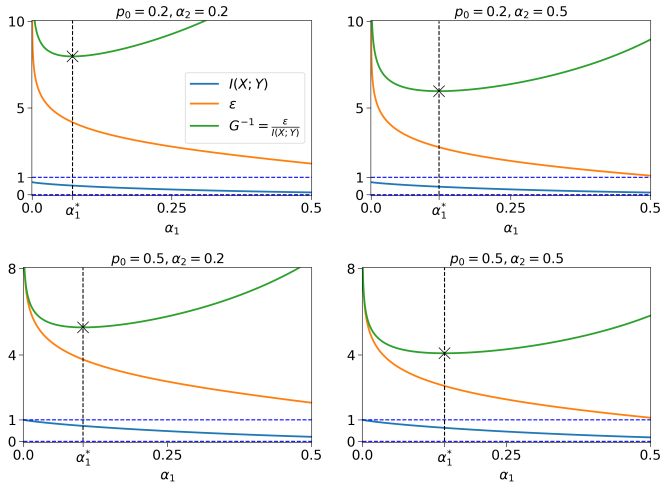


Fig. 2: Plots of $G^{-1}$ vs. $\alpha_1$ for various values of $p_0, \alpha_2$. In each instance, the optimization algorithm's solution $\alpha_1^*$ is marked by $\times$.

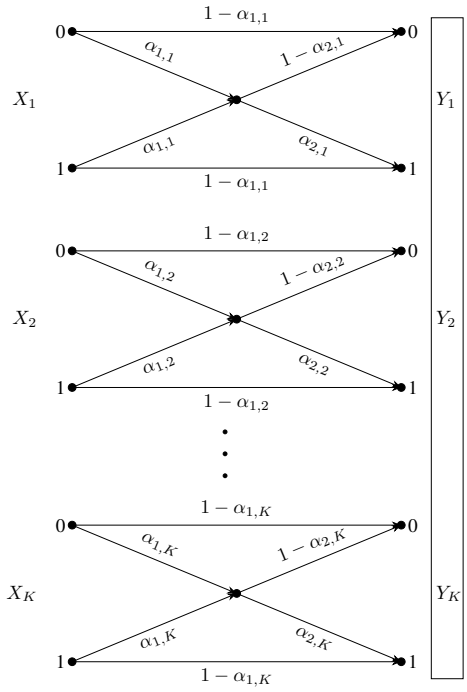


Fig. 3: The multiple randomized response (RR) setting. An adversary uses $Y = (Y_1, Y_2, \ldots, Y_K)$ collectively to estimate $X_1, X_2, \ldots, X_K$. For each survey question indexed by $i \in \{1, 2, \ldots, K\}$, the corresponding channel has transition probabilities $(\alpha_{1,i}, \alpha_{2,i})$, which can be optimized for the best trade-off between information extraction and privacy preservation.

In the context of multiuser information theory, this setting precisely parallels the multiple-access channel (MAC) [9] [4], where we have transmitters $X_1, X_2, \ldots, X_K$ communicating to a common receiver $Y = (Y_1, Y_2, \ldots, Y_K)$. Their mutual information $I(X_1, X_2, \ldots, X_K; Y)$ quantifies the extent to which the receiver can reliably determine the $K$ true responses. Additionally, during decoding the receiver can make use of any known correlations between the survey questions' responses to construct a better estimate of $X_1, X_2, \ldots, X_K$; a setting which parallels the decoding of correlated sources [10] [11] [12].

Here we take care to highlight a few distinguishing factors between the classic multiuser setting and the present one. In the former, correlations between sources, represented by conditional probabilities between the inputs of various channels, are of interest because both the optimization of the joint probability distribution over channel inputs, and the subsequent decoding of the input messages, depends on the correlations. In contrast, because the (marginal) probabilities of inputs are known in the present paper's setting (see Subsection III-A), and because we optimize over the channel transition probabilities rather than over the inputs, input correlations have no impact on the optimization problem. However, as in the classic multiuser setting, using knowledge of the input correlations for subsequent decoding remains possible.

The present work is furthermore interested in correlations between the channels' transition probability values. These correlations affect the optimization over the joint channel transition probabilities values, because the optimal value of $G$ depends on them. However, these correlations do not affect the subsequent decoding of inputs. We discuss the setting of correlated channel transition probability values further in Subsection IV-D.

Finally, we consider how correlations between channel transition probability values might arise in practice.[7] Suppose it is required by some entity that for particularly sensitive survey questions, the probability of tails be higher to further improve privacy on those questions, and suppose it is known when a question is considered sensitive. Then the requirement for higher privacy on particularly sensitive questions, and the knowledge of what constitutes a sensitive question, together leads to a soft constraint on different channels' transition probabilities values – which is equivalent to a correlation between them.

Given the preceding discussion, we will only consider correlations in the context of channel transition probability values, rather than between inputs, for the remainder of the work.

### *B. Differential Privacy Correspondence*

In the context of responses to multiple survey questions, a necessary condition for satisfying $\epsilon_{1:K}$-DP is that

[7]We note that the setting is also of interest in its purely abstract form, removed from the particular application.

for all $\mathcal{S} \subseteq \mathcal{Y}$ with $\mathcal{Y} = \left\{(y_1, \ldots, y_K) \in \{0,1\}^K\right\}$, Equation 1 is satisfied (with $\delta = 0$). Furthermore, in the multiple RR setting each user's set of responses can be considered as the entry of a single-row database; an entry consisting of $K$ responses/attributes. Therefore $(\boldsymbol{x}, \boldsymbol{x}') = ((x_1, \ldots, x_K), (x'_1, \ldots, x'_K))$ are considered neighboring even if $(\boldsymbol{x}, \boldsymbol{x}')$ differ in more than one of the responses/attributes. Thus we must satisfy Equation 1 for all

$$\begin{aligned} &((x_1, \ldots, x_K), (x'_1, \ldots, x'_K)) \in \\ &\Big\{\left((x_1, \ldots, x_K) \in \{0,1\}^K, (x'_1, \ldots, x'_K) \in \{0,1\}^K\right) \\ &\text{s.t.} \sum_{i=1}^{K} \mathbb{1}\left(x_i \neq x'_i\right) \geq 1\Big\}. \end{aligned} \tag{4}$$

We can evaluate $\epsilon_{1:K}$ for this setting as follows. For each of the $i \in \{1, \ldots, K\}$ survey responses, let

$$\begin{aligned} B^i_{0,0} &= \mathbb{P}\left(Y_i = 0 | X_i = 0\right) = 1 - \alpha_{1,i}\alpha_{2,i}, \\ B^i_{0,1} &= \mathbb{P}\left(Y_i = 0 | X_i = 1\right) = \alpha_{1,i} - \alpha_{1,i}\alpha_{2,i}, \\ B^i_{1,0} &= \mathbb{P}\left(Y_i = 1 | X_i = 0\right) = \alpha_{1,i}\alpha_{2,i}, \\ B^i_{1,1} &= \mathbb{P}\left(Y_i = 1 | X_i = 1\right) = 1 - \alpha_{1,i} + \alpha_{1,i}\alpha_{2,i}. \end{aligned} \tag{5}$$

Now consider the two following sets, which serve to index the received responses by their values:

$$\begin{aligned} \mathcal{S}_0 &= \{i : Y_i = 0\}, \\ \mathcal{S}_1 &= \{i : Y_i = 1\}. \end{aligned} \tag{6}$$

Then we have

$$\begin{aligned} &\mathbb{P}\left((Y_1, \ldots, Y_K) | X_1, \ldots, X_K\right) \\ &= \prod_{i \in \mathcal{S}_0} B^i_{0,x_i} \cdot \prod_{i \in \mathcal{S}_1} B^i_{1,x_i} \\ &= \prod_{j \in \{0,1\}} \prod_{i \in \mathcal{S}_j} B^i_{j,x_i} \end{aligned} \tag{7}$$

We next evaluate $\epsilon_{1:K}$ using the following expression:

$$e^{\epsilon_{1:K}} = \max_{\boldsymbol{x}, \boldsymbol{x}'} \frac{\mathbb{P}\left(Y = (1, \ldots, 1) | X_1 = x_1, \ldots, X_K = x_K\right)}{\mathbb{P}\left(Y = (1, \ldots, 1) | X_1 = x'_1, \ldots, X_K = x'_K\right)}, \tag{8}$$

where the maximum is obtained for $\boldsymbol{x} = (1, 1, \ldots, 1)$ and $\boldsymbol{x}' = (0, 0, \ldots, 0)$, so that

$$\begin{aligned} e^{\epsilon_{1:K}} &= \frac{\mathbb{P}\left(Y = (1, 1, \ldots, 1) | X_1 = 1, \ldots, X_K = 1\right)}{\mathbb{P}\left(Y = (1, 1, \ldots, 1) | X_1 = 0, \ldots, X_K = 0\right)} \\ &= \frac{\prod_{i \in \mathcal{S}_1} B^i_{1,1}}{\prod_{i \in \mathcal{S}_1} B^i_{1,0}} \\ &= \frac{\prod_{i \in \mathcal{S}_1} \left(1 - \alpha_{1,i} + \alpha_{1,i}\alpha_{2,i}\right)}{\prod_{i \in \mathcal{S}_1} \left(\alpha_{1,i}\alpha_{2,i}\right)} \\ &= \prod_{i=1}^{K} \frac{1 - \alpha_{1,i} + \alpha_{1,i}\alpha_{2,i}}{\alpha_{1,i}\alpha_{2,i}}. \end{aligned} \tag{9}$$

$$\therefore \epsilon_{1:K} = \sum_{i=1}^{K} \log\left(1 + \frac{1 - \alpha_{1,i}}{\alpha_{1,i}\alpha_{2,i}}\right) = \sum_{i=1}^{K} \epsilon_i.$$

Thus the privacy budget for the multi-response setting corresponds to the sum of the privacy budgets for the individual responses.

### C. Multiple Response Optimization Problem

We now cast the trade-off between information extraction and privacy preservation in the multiple response setting using the following optimization problem:

$$G_{1:K} = \max_{\boldsymbol{\alpha}_1} \frac{I\left(X_1, X_2, \ldots, X_K; (Y_1, Y_2, \ldots, Y_K)\right)}{\epsilon_{1:K}}. \tag{10}$$

Here $\boldsymbol{\alpha}_1 = (\alpha_{1,1}, \alpha_{1,2}, \ldots, \alpha_{1,K})$, and we have free variables $\boldsymbol{p}_0 = (p_{0,1}, p_{0,2}, \ldots, p_{0,K})$ and $\boldsymbol{\alpha}_2 = (\alpha_{2,1}, \alpha_{2,2}, \ldots, \alpha_{2,K})$.

### D. Analysis of the Multiple Response Setting

We analyze two settings: the first where the questionnaire transition probabilities are uncorrelated, and the second where correlations exist. For clarity of exposition in both settings the number of users is set to $K = 2$, and we will only consider correlations between $\alpha_{1,1}, \alpha_{1,2}$ and not between $\alpha_{2,1}, \alpha_{2,2}$.

*1) Example 1: Uncorrelated Channel Transition Probabilities:* Here the values for $\alpha_{1,1}$ and $\alpha_{1,2}$ are independent. Figure 4 shows plots of $G^{-1}_{1:2}$ vs. $\alpha_{1,1}$ for two configurations of the free parameters $(p_{0,1}, p_{0,2}, \alpha_{2,1}, \alpha_{2,2})$. In each plot, $\alpha_{1,2}$ is set to its optimal value $\alpha^*_{1,2}$, derived from the joint optimization over $(\alpha_{1,1}, \alpha_{1,2})$. As in the single response setting, this example demonstrates that the joint optimization problem over the entries of $\boldsymbol{\alpha}_1$ leads to interesting solutions in the trade-off between information extraction and privacy preservation.

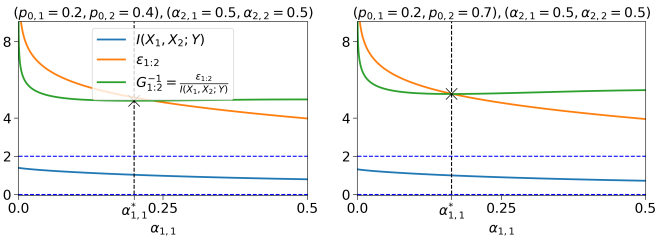


Fig. 4: Plots of $G^{-1}_{1:2}$ vs. $\alpha_{1,1}$ for two configurations of $(p_{0,1}, p_{0,2}, \alpha_{2,1}, \alpha_{2,2})$, and where $\alpha_{1,1}, \alpha_{1,2}$ are uncorrelated. In each case, $\alpha_{1,2}$ was set to its optimal value $\alpha^*_{1,2}$, derived from the joint optimization over $(\alpha_{1,1}, \alpha_{1,2})$.

*2) Example 2: Correlated Channel Transition Probabilities:* To analyze the setting where the $\boldsymbol{\alpha}_1$ are correlated, it is useful to consider the channel shown in Figure 5, which is the equivalent (binary erasure) channel representing the outcome of the first coin flip in RR, given by $T_i \in \{0, 1, e\}$ where $e$ signifies erasure. Clearly $\alpha_{1,i} = \mathbb{P}\left(T_i = e\right)$. We can signify correlations between the values of $\alpha_{1,1}, \alpha_{1,2}$ using the following pairs of conditional probabilities: $\mathbb{P}\left(T_2 \neq e | T_1 \neq e\right) = \beta_1$, $\mathbb{P}\left(T_2 = e | T_1 = e\right) = \beta_2$. Then with $\alpha_{1,1} = \mathbb{P}\left(T_1 = e\right)$ as the only free variable, we can specify

$\alpha_{1,2} = \mathbb{P}(T_2 = e) = (1-\beta_1) + \alpha_{1,1}(\beta_1 + \beta_2 - 1)$. For simplicity let $\beta = \beta_1 = \beta_2$ in the sequel.

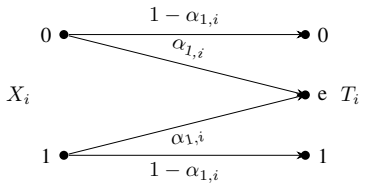


Fig. 5: The binary erasure channel representing the outcome of the first coin flip in RR.

For given $\beta$, denote by $I^\beta, \epsilon^\beta$ the values for the mutual information and the privacy parameter, respectively. Figure 6 demonstrates that for larger values of $\beta$, the information extracted from the survey is higher, and the privacy correspondingly worse. Similarly for lower values of $\beta$, the information extracted from the survey is lower, and the privacy correspondingly better. Thus the effect of correlations in the channels' transition probabilities significantly affects the channel characteristics. Interestingly, as $\alpha_{1,1} \to 0.5$, the value of $\beta$ becomes irrelevant because the channels become completely random.

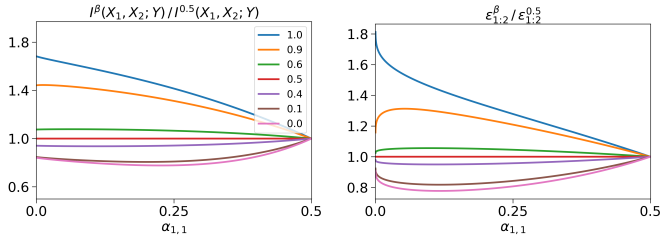


Fig. 6: Plots showing the ratio of $I^\beta(X_1,X_2;Y)/I^{0.5}(X_1,X_2;Y)$ (left) and $\epsilon^\beta_{1:2}/\epsilon^{0.5}_{1:2}$ (right) for various values of $\beta$ (see plot legend). In all cases $(p_{0,1}, p_{0,2}, \alpha_{2,1}, \alpha_{2,2}) = (0.5, 0.5, 0.5, 0.5)$, and $\alpha_{1,2} = (1-\beta) + \alpha_{1,1}(2\beta - 1)$ (see text for details). Higher values of $\beta$ correspond to better information extraction, but correspondingly worse privacy.

## V. Related Work

Several works relate DP and mutual information. In the work of [13], an equivalent definition of DP as a mutual information constraint, positioned between standard $\epsilon$-DP and $(\epsilon, \delta)$-DP in terms of its privacy guarantees, is proposed. The work of [14] develops a framework which characterizes the trade-offs between privacy preservation and accuracy in statistical estimation problems, under the local DP setting. It presents mechanisms that maintain local privacy while achieving minimax optimality in various estimation problems, and investigates the limits of what can be achieved in private data analysis and the trade-off between accuracy and privacy. In the work of [15], DP mechanisms are related to rate-distortion theory, addressing the trade-off between the fidelity of data representation and the amount of data compression. A formal analysis of the trade-off between accuracy and privacy preservation is presented, illustrating that enhancing privacy generally comes at the cost of reduced accuracy. This relationship is modeled through a risk-distortion function that quantifies the trade-off between accuracy and the risk of privacy breaches.

In the present work the trade-off between information extraction and privacy preservation was quantified using an optimization problem, thus offering a principled approach for trading-off the two contrasting goals.

## VI. Conclusion

We explored the relationship between information theory and DP by means of a comprehensive analysis of the RR algorithm in survey questionnaires. By framing the trade-off between information extraction and privacy preservation as an optimization problem, we showed the best possible trade-off between the two can be achieved by tuning the transition probability values. Furthermore, our analysis was extended to the multiuser setting with correlated transition probability values. The optimization problems developed, and their corresponding solutions, offer practical strategies for managing privacy risks in real-world data analysis settings.

## References


[1] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Theory of Cryptography*, S. Halevi and T. Rabin, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 265–284.

[2] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy," *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3–4, p. 211–407, aug 2014. [Online]. Available: https://doi.org/10.1561/0400000042

[3] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006.

[5] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[6] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965, pMID: 12261830. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1965.10480775

[7] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM Journal on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995. [Online]. Available: https://doi.org/10.1137/0916069

[8] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization," *ACM Trans. Math. Softw.*, vol. 23, no. 4, p. 550–560, dec 1997. [Online]. Available: https://doi.org/10.1145/279232.279236

[9] R. Ahlsvvede, "Multi-way communication channels," in *Proc. 2nd Int. Symp. Inf. Theory*, 1971, pp. 23–51.

[10] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[11] D. Slepian and J. K. Wolf†, "A coding theorem for multiple access channels with correlated sources," *The Bell System Technical Journal*, vol. 52, no. 7, pp. 1037–1076, 1973.

[12] T. Cover, A. Gamal, and M. Salehi, "Multiple access channels with arbitrarily correlated sources," *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 648–657, 1980.

[13] P. Cuff and L. Yu, "Differential privacy as a mutual information constraint," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 43–54. [Online]. Available: https://doi.org/10.1145/2976749.2978308

[14] M. I. J. John C. Duchi and M. J. Wainwright, "Minimax optimal procedures for locally private estimation," *Journal of the American Statistical Association*, vol. 113, no. 521, pp. 182–201, 2018. [Online]. Available: https://doi.org/10.1080/01621459.2017.1389735

[15] D. J. Mir, "Information-theoretic foundations of differential privacy," in *Foundations and Practice of Security*, J. Garcia-Alfaro, F. Cuppens, N. Cuppens-Boulahia, A. Miri, and N. Tawbi, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 374–381.