Fundamental Limits in Data Privacy:

From Privacy Measures to Economic Foundations

by

Weina Wang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2016 by the
Graduate Supervisory Committee:

Lei Ying, Co-Chair
Junshan Zhang, Co-Chair
Anna Scaglione
Yanchao Zhang

ARIZONA STATE UNIVERSITY

August 2016

ABSTRACT

Data privacy is emerging as one of the most serious concerns of big data analytics, particularly with the growing use of personal data and the ever-improving capability of data analysis. This dissertation first investigates the relation between different privacy notions, and then puts the main focus on developing economic foundations for a market model of trading private data.

The first part characterizes differential privacy, identifiability and mutual-information privacy by their privacy–distortion functions, which is the optimal achievable privacy level as a function of the maximum allowable distortion. The results show that these notions are fundamentally related and exhibit certain consistency: (1) The gap between the privacy–distortion functions of identifiability and differential privacy is upper bounded by a constant determined by the prior. (2) Identifiability and mutual-information privacy share the same optimal mechanism. (3) The mutual-information optimal mechanism satisfies differential privacy with a level at most a constant away from the optimal level.

The second part studies a market model of trading private data, where a data collector purchases private data from strategic data subjects (individuals) through an incentive mechanism. The value of ϵ units of privacy is measured by the minimum payment such that an individual's equilibrium strategy is to report data in an ϵ-differentially private manner. For the setting with binary private data that represents individuals' knowledge about a common underlying state, asymptotically tight lower and upper bounds on the value of privacy are established as the number of individuals becomes large, and the payment–accuracy tradeoff for learning the state is obtained. The lower bound assures the impossibility of using lower payment to buy ϵ units of privacy, and the upper bound is given by a designed reward mechanism. When the individuals' valuations of privacy are unknown to the data collector, mechanisms with possible negative payments (aiming to penalize individuals with “unacceptably” high privacy valuations) are designed to fulfill the accuracy goal and drive the total payment to zero. For the setting with binary private data following a general joint probability distribution with some symmetry, asymptotically optimal mechanisms are designed in the high data quality regime.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

Figure Page

Chapter 1

INTRODUCTION

From the monetary coupons offered for revealing opinions of a product to the large-scale trade of personal information by data brokers such as Acxiom Kroft (2014), the commoditization of private data has been trending up when big data analytics is playing a more and more critical role in advertising, scientific research, etc. However, in the wake of a number of recent scandals, such as the Netflix data breach and the Veterans Affairs data theft, data privacy is emerging as one of the most serious concerns of big data analytics. This has given rise to a fundamental question: whether big data and privacy can go hand-by-hand or giving up our privacy is inevitable in the big-data era.

1.1 Overview

In this dissertation, we first investigate the relation between different privacy notions. The definition of privacy has been at the center of the research on data privacy, with different notions proposed to capture different perspectives of privacy-sensitive scenarios. Studying the relation between these privacy notions can deepen our understanding of privacy. Privacy concerns become prominent with the ever-improving capability of data analysis. Analyzing personal data results in new discoveries in science and engineering, but also puts individual's privacy at potential risks. Therefore, privacy-preserving data analysis, where the goal is to preserve the accuracy of data analysis while maintaining individual's privacy, has become one of the main challenges of this big data era. The basic idea of privacy-preserving data analysis is to inject a right amount of randomness in the released information to guarantee that an individual's information cannot be inferred. Intuitively, the higher the randomness is, the better privacy protection individual users get, but the less accurate (useful) the output statistical information is. While randomization seems to be inevitable, for the privacy-preserving data analysis it is of great interest to quantitatively define the notion of privacy. Specifically, we need to understand the amount of randomness needed to protect privacy while preserving usefulness of the data. To this end, we consider three different notions: identifiability, differential privacy and mutual-information privacy, where identifiability is concerned with the posteriors of recovering the original data from the released data, differential privacy is concerned with additional disclosures of an individual's information due to the release of the data, and mutual information measures the average amount of information about the original database contained in the released data. While these three different privacy notions are defined from different perspectives, we put these privacy notions under a unified privacy–distortion framework and show that they are fundamentally related.

Next, taking a forward-looking view, we envisage a market model for private data analytics where the data collector uses a *reward mechanism* to incentivize individuals to report informative data, and individuals control their own data privacy by reporting *noisy data* with the randomization algorithms strategically chosen to maximize their payoffs. We quantify the privacy disclosure of an individual's data-reporting strategy by a local variant of differential privacy Dwork *et al.* (2006b); Kasiviswanathan *et al.* (2011); Dwork and Roth (2014), which measures privacy disclosure by the distinguishability between the probability distributions of the reported data for different contents of the private data. A distinctive merit of our approach is that data subjects take full control of their own data privacy and the data collector gets informative data but does not need to bear the responsibility of protecting privacy. One significant challenge, however, is also rooted in this desired merit: the data collector has no direct control or even no information on how individuals would randomize their private data.

We address this challenge by devising a game-theoretic formulation, which allows us to predict how individuals behave to reconcile the conflict between rewards and privacy. To make an individual willing to trade a desired amount of privacy, the data collector needs to incentivize the individual by making sure that doing so benefits her most. Therefore, to grasp the intrinsic characteristics of the market and find the balance point where the data collector and the individuals cut a deal, we focus on individuals' strategies in a Nash equilibrium of the reward mechanism. The fundamental question—"how much is privacy worth"—can then be cast as: *what is the minimum reward to an individual such that her strategy in a Nash equilibrium is to report data with certain units of privacy disclosure?*

1.2 Summary of Contributions

In Chapter 2, we investigate the relation between three different notions of privacy: identifiability, differential privacy and mutual-information privacy. Under a unified privacy–distortion framework, where the distortion is defined to be the expected Hamming distance between the input and output databases, we establish some fundamental connections between these three privacy notions. Given a maximum allowable distortion D, we define the privacy–distortion functions $\epsilon_{\mathrm{i}}^*(D)$, $\epsilon_{\mathrm{d}}^*(D)$, and $\epsilon_{\mathrm{m}}^*(D)$ to be the smallest (most private/best) identifiability level, differential privacy level, and mutual information between the input and output, respectively. We characterize $\epsilon_{\mathrm{i}}^*(D)$ and $\epsilon_{\mathrm{d}}^*(D)$, and prove that $\epsilon_{\mathrm{i}}^*(D) - \epsilon_X \leq \epsilon_{\mathrm{d}}^*(D) \leq \epsilon_{\mathrm{i}}^*(D)$ for D within certain range, where ϵ_X is a constant determined by the prior distribution of the original database X, and diminishes to zero when X is uniformly distributed. Further, we show that $\epsilon_{\mathrm{i}}^*(D)$ and $\epsilon_{\mathrm{m}}^*(D)$ can be achieved by the same mechanism for

D within certain range, i.e., there is a mechanism that simultaneously minimizes the identifiability level and achieves the best mutual-information privacy. Based on these two connections, we prove that this mutual-information optimal mechanism satisfies ϵ-differential privacy with $\epsilon_{\mathrm{d}}^*(D) \leq \epsilon \leq \epsilon_{\mathrm{d}}^*(D) + 2\epsilon_X$. The results in this chapter indicate a consistency between "worst-case" notions of privacy, identifiability and differential privacy, and an "average" notion of privacy, mutual-information privacy.

In Chapter 3–Chapter 5, we study a market model of trading private data and quantify the value of privacy. In Chapter 3, we consider a setting where the private data of each individual represents her knowledge about an underlying state, which is the information that the data collector desires to learn. The value of ϵ units of privacy is measured by the minimum payment of all nonnegative payment mechanisms, under which an individual's best response at a Nash equilibrium is to report the data with a privacy level of ϵ. The higher ϵ is, the less private the reported data is. We establish asymptotically tight lower and upper bounds on the value of ϵ units of privacy as the number of individuals becomes large. The lower bound assures that it is impossible to use a lower amount of reward to obtain ϵ units of privacy from an individual. The upper bound is given by an achievable reward mechanism that we designed, in which the data collector obtains ϵ units of privacy from each individual in a Nash equilibrium, and the expected reward to each individual converges to the lower bound exponentially fast with the number of individuals. We also provide characterizations on the strategies of individuals in a Nash equilibrium, which advance our understanding of the behavior of privacy-aware individuals. Based on these fundamental limits, we further derive lower and upper bounds on the minimum total payment for the data collector to achieve a given learning accuracy target, and show that the total payment of the designed mechanism is at most one individual's payment away from the minimum.

Then in Chapter 4, we consider a setting where the individuals's valuations of privacy are unknown to the data collector. We consider a model where each individual's privacy valuation is characterized by a "cost coefficient", which can be regarded as her type. By allowing possible negative payments (which penalize individuals with "unacceptable" valuations of privacy), we are able to cope with the uncertainty in the cost coefficients and drive down the data analyst's cost. We design a family of payment mechanisms, each of which has a Bayesian Nash equilibrium where the individuals exhibit a threshold behavior: the individuals with cost coefficients above a threshold choose not to participate, and the individuals with cost coefficients below the threshold participate and report data with a guaranteed quality. By choosing appropriate parameters, we obtain a sequence of mechanisms with the number of individuals grows large. Each such mechanism fulfills the accuracy goal at a Bayesian Nash equilibrium, and the corresponding total expected payment goes to zero; i.e., this sequence of mechanisms is asymptotically optimal.

In Chapter 5, we consider a more general model for the private data. The data collector is interested in learning the average of the private data. We design a payment mechanism such that the quality of the collected data is controllable through a parameter ϵ by making sure that each individual's strategy in a Nash equilibrium is to participate and symmetrically randomize her data, while guaranteeing ϵ-differential privacy. With this design, the data collector can achieve any given accuracy objective by using the payment mechanism associated with an appropriate ϵ. The total expected payment of the designed mechanism at equilibrium is asymptotically optimal in the high data quality regime.

1.3 Related Work

1.3.1 Differential Privacy

Differential privacy, as an emerging analytical foundation for privacy-preserving data analysis, was developed by a line of work Dwork *et al.* (2006b); Dwork (2006); Dwork *et al.* (2006a), and since then both interactive model (e.g., Dwork *et al.* (2006b); Nissim *et al.* (2007); Ghosh *et al.* (2009); Roth and Roughgarden (2010); Hardt and Rothblum (2010); Gupta *et al.* (2012); Muthukrishnan and Nikolov (2012)) and non-interactive model (e.g., Blum *et al.* (2008); Dwork *et al.* (2009); Kasiviswanathan *et al.* (2010); Ullman and Vadhan (2011); Gupta *et al.* (2012); Hardt *et al.* (2012); Bun *et al.* (2014)) have been studied in the literature. There is a vast and growing body of work on differential privacy, which we do not attempt to survey but refer interested readers to a comprehensive survey by Dwork and Roth (2014).

1.3.2 Other Notions of Privacy

The privacy guarantee of differential privacy does not depend on the prior distribution of the original database, since it captures the additional disclosure caused by an information releasing mechanism on top of any given disclosure. With the prior taken into account, privacy notions based on the posterior have also been proposed. The seminal work of differential privacy Dwork *et al.* (2006b) also proposed a semantically flavored definition of privacy, named semantic security, and showed its equivalence to differential privacy. This definition measures privacy by the difference between an adversary's prior knowledge of the database and the posterior belief given the output of the mechanism. Differential identifiability Lee and Clifton (2012) and membership privacy Li *et al.* (2013) assume that a database entry can be traced back to the identity of an individual, and the leakage of the information is quantified on whether an individual participates in the database or not. Specifically, differential identifiability is defined to be the posterior probability for any individual to be the only unknown participant of a database given the entries of all the known participants and the output of the mechanism. This probability cannot be directly translated to a differential privacy level. Membership privacy is defined based on the difference between the prior and the posterior probability for an entity to be included in the database. Choosing appropriate prior distribution families makes differential privacy and differential identifiability instantiations of membership privacy under their database model. In this dissertation, the notion of identifiability is defined based on the indistinguishability between the posterior probabilities of neighboring databases given the output of the mechanism, which measures the hardness of identifying the data content of a database entry rather than the identity of the individual who contributes the data.

Information-theoretic privacy measures including mutual information, min-entropy, equivocation, etc, are relatively classical and have a rich history (e.g., Agrawal and Aggarwal (2001); Clark *et al.* (2005); Smith (2009); Zhu and Bettati (2005); Chatzikokolakis *et al.* (2007, 2010); Rebollo-Monedero *et al.* (2010); Alvim *et al.* (2012); du Pin Calmon and Fawaz (2012); Makhdoumi and Fawaz (2013); Mir (2013); Sankar *et al.* (2013); Sarwate and Sankar (2014)). When mutual information is used as the privacy notion, the problem of finding the optimal tradeoff between privacy and distortion can usually be formulated as a rate–distortion problem in the field of information theory (see Cover and Thomas (2006) for an introduction) Rebollo-Monedero *et al.* (2010); du Pin Calmon and Fawaz (2012); Makhdoumi and Fawaz (2013); Mir (2013); Sankar *et al.* (2013); Sarwate and Sankar (2014). In this dissertation, we also utilize results from the celebrated rate–distortion theory to characterize the optimal privacy–distortion tradeoff. However, we are more interested in the relation between the optimal privacy–distortion tradeoffs with different privacy notions:

mutual information, differential privacy, and identifiability, and we quantify the impact of the prior explicitly. The work du Pin Calmon and Fawaz (2012); Makhdoumi and Fawaz (2013) showed that when a mechanism satisfies ϵ-information privacy (defined based on the difference between the prior of the database and the posterior given the output), it is 2ϵ-differentially private, and the mutual information between the database and the output is upper bounded by $\epsilon / \ln 2$. But differential privacy alone does not imply a bound on the mutual information if the possible values and sizes of the database and the output and the prior can be chosen freely. McGregor *et al.* (2010) and De (2012) showed that ϵ-differential privacy implies upper bounds on the mutual information in the order of $O(\epsilon n)$ and $O(\epsilon d)$, respectively, where n is the size of the database and d is the dimension of the data entry. Alvim *et al.* (2012) showed that differential privacy implies a bound on the min-entropy leakage. The above relations between information-theoretic privacy notions and differential privacy, however, are not for the optimal privacy with distortion constraint, although they can contribute to building relations between the optimal tradeoffs. Sarwate and Sankar (2014) showed that the result in McGregor *et al.* (2010) indicates a one direction bound between the optimal differential privacy and the optimal mutual information given the same distortion constraint. Mir (2013) pointed out that the mechanism that achieves the optimal rate–distortion also guarantees a certain level of differential privacy. However, whether this differential privacy level is optimal or how far it is from optimal was not answered.

1.3.3 Market Approaches for Collecting Private Data

Market approaches for collecting data from privacy-aware individuals have led to a fruitful line of work Ghosh and Roth (2011); Fleischer and Lyu (2012); Ligett and Roth (2012); Roth and Schoenebeck (2012); Ghosh and Ligett (2013); Xiao (2013);

Chen *et al.* (2013); Nissim *et al.* (2014); Ghosh *et al.* (2014); Wang *et al.* (2015a, 2016). Our work uniquely studies a data collector/analyst who is not necessarily trustworthy. This results in the procurement of noisy data instead of true data.

There are two primary flavors of mechanism design for collecting data from privacy-aware individuals in the literature, depending on the available actions that the individuals can take. One approach models the scenario where the private data is verifiable, but the privacy costs to individuals incurred by using their data are unknown to the data analyst and individuals have the option to lie about their privacy costs. The goal of the mechanism design is to conduct privacy-preserving analysis on the private data with the privacy costs of individuals properly compensated. In the seminal work Ghosh and Roth (2011), an individual's privacy cost is modeled as a linear function of ϵ if her data is used in an ϵ-differentially private manner. Mechanisms were designed to elicit truthful reporting of the linear coefficients and estimate some statistic cheaply. Subsequent work Fleischer and Lyu (2012); Ligett and Roth (2012); Roth and Schoenebeck (2012); Ghosh and Ligett (2013); Nissim *et al.* (2014) explores various models for individuals' valuation of privacy, especially the correlation between the coefficients and the private bits.

Another line of research Xiao (2013); Chen *et al.* (2013); Ghosh *et al.* (2014) studies the scenario where individuals can lie about their data and will do so if that benefits them, but the data analyst is still trusted—revealing information to the data analyst does not incur privacy costs. In the notable work Ghosh *et al.* (2014), the designed mechanism incentivizes truthful data reporting (without adding any noise) from individuals and satisfies joint differential privacy.

The above work falls into the broad area of the interplay between differential privacy and mechanism design, which was first studied by McSherry and Talwar (2007). They treat differential privacy as a tool to design approximately truthful mechanisms. A comprehensive survey of the development in this area is given by Pai and Roth (2013).

The local model of differential privacy, which is a generalization of randomized response Warner (1965) and is formalized in Kasiviswanathan *et al.* (2011), has been studied in the literature Dwork *et al.* (2006b); Dwork (2006); Hsu *et al.* (2012); Duchi *et al.* (2013); Dwork and Roth (2014); Chen *et al.* (2014); Kairouz *et al.* (2014); Wang *et al.* (2014, 2015b); Bassily and Smith (2015); Shokri (2015). The behavior of individuals with privacy concerns has been studied in Chen *et al.* (2014), which investigates the types of games in which strategic individuals truthfully follow randomized response, rather than sending some arbitrary bit. The hypothesis testing formulation in this dissertation is similar to a setting in Kairouz *et al.* (2014), where the authors find an optimal mechanism that maximizes the statistical discrimination of the hypotheses subject to local differential privacy constraints. In practice, Google's Chrome web browser has implemented the RAPPOR mechanism Erlingsson *et al.* (2014); Fanti *et al.* (2015) to collect users' data, which guarantees that only limited privacy of users is leaked by using randomized response in a novel manner. However, users may still not be willing to report data in the desired way due to the lack of an incentive mechanism.

Chapter 2

RELATION BETWEEN DIFFERENT PRIVACY NOTIONS

2.1 Introduction

We investigate the fundamental connections between these three different privacy notions in the following setting:

- We consider a non-interactive database releasing approach for privacy-preserving data analysis, where a synthetic database is released to the public. The synthetic database is a sanitized version of the original database, on which queries and operations can be carried out as if it was the original database. It is then natural to assume that the synthetic database and the original database are in the same "universe" so the entries have the same interpretation. Therefore we focus on mechanisms that map an input database to an output synthetic database in the same universe. Specifically, we consider a database consisting of n rows, each of which takes values from a finite domain $\mathcal{D}$ of size m. In this dissertation, the database is modeled as a discrete random variable X drawn from $\mathcal{D}^n$ with prior distribution p_X. A mechanism $\mathcal{M}$ takes a database X as input and outputs a database Y, which is also a random variable with alphabet $\mathcal{D}^n$.

- We define the *distortion* between the output database and the input database to be the expected Hamming distance. When the input and output are in the same universe, the Hamming distance measures the number of rows two databases differ in, which directly points to the number of rows that need to be modified in order to guarantee a given privacy level.

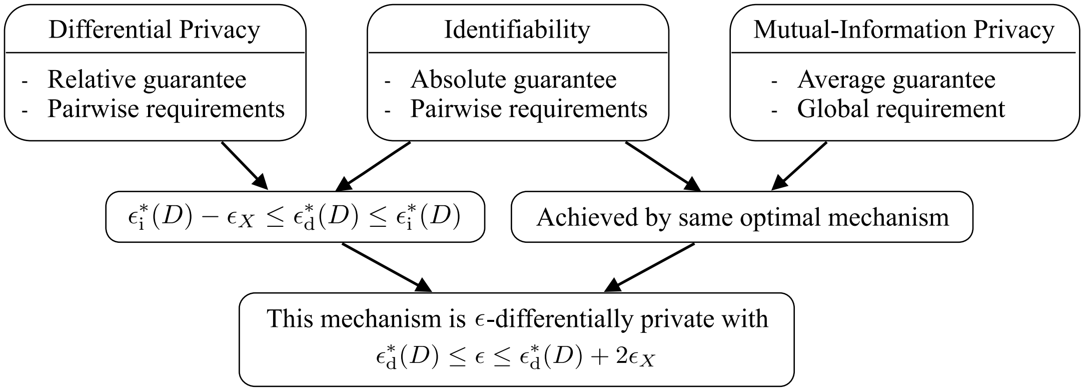

Differential Privacy
- Relative guarantee
- Pairwise requirements
Identifiability
- Absolute guarantee
- Pairwise requirements
Mutual-Information Privacy
- Average guarantee
- Global requirement
$\epsilon_{\mathrm{i}}^*(D) - \epsilon_X \leq \epsilon_{\mathrm{d}}^*(D) \leq \epsilon_{\mathrm{i}}^*(D)$
Achieved by same optimal mechanism
This mechanism is ϵ-differentially private with
$\epsilon_{\mathrm{d}}^*(D) \leq \epsilon \leq \epsilon_{\mathrm{d}}^*(D) + 2\epsilon_X$

Figure 2.1: Relation between identifiability, differential privacy and mutual-information privacy.

In this dissertation, we use a unified *privacy–distortion* framework to understand the relation between the three privacy notions. Given a maximum allowable distortion D, we define the privacy–distortion functions $\epsilon_{\mathrm{i}}^*(D)$, $\epsilon_{\mathrm{d}}^*(D)$, and $\epsilon_{\mathrm{m}}^*(D)$ to be the smallest identifiability level, differential privacy level, and mutual information between the input and output, respectively. Then we have the following main results, which are also summarized in Figure 2.1.

(1) We derive the exact form of the privacy–distortion function $\epsilon_{\mathrm{i}}^*(D)$ under the notion of identifiability, for certain range of the distortion values, by showing that $\epsilon_{\mathrm{i}}^*(D) = h^{-1}(D)$ regardless of the prior distribution, where

$$h^{-1}(D) = \ln\Big(\frac{n}{D} - 1\Big) + \ln(m-1).$$

We further show that for the privacy–distortion function $\epsilon_{\mathrm{d}}^*(D)$ under the notion of differential privacy,

$$\epsilon_{\mathrm{i}}^*(D) - \epsilon_X \leq \epsilon_{\mathrm{d}}^*(D) \leq \epsilon_{\mathrm{i}}^*(D).$$

The constant ϵ_X is determined by the prior distribution p_X only, given by

$$\epsilon_X = \max_{x,x' \in \mathcal{D}^n : x \sim x'} \ln \frac{p_X(x)}{p_X(x')},$$

where $x \sim x'$ denotes that x and x' differ in exactly one row. When the input database has a uniform distribution, we have that $\epsilon_{\mathrm{i}}^* = \epsilon_{\mathrm{d}}^*$, i.e., differential privacy is equivalent to identifiability. Note that for ϵ_X to be finite, the prior p_X needs to have full a support on $\mathcal{D}^n$, i.e., $p_X(x) > 0$ for any $x \in \mathcal{D}^n$. When ϵ_X is large, differential privacy provides only weak guarantee on identifiability. In other words, when ϵ_X is large, it is possible to identify some entries of the database with non-trivial accuracy even if the differential privacy is satisfied. This is because differential privacy provides a *relative* guarantee about disclosures, which ensures that limited *additional* information of an individual is leaked in the released data in addition to the knowledge that an adversary has known. Identifiability, on the other hand, requires an *absolute* guarantee about disclosures when individuals' data is being inferred from the output database assuming that the prior p_X and the mechanism are both known to the adversary.

(2) The privacy–distortion functions $\epsilon_{\mathrm{i}}^*(D)$ and $\epsilon_{\mathrm{m}}^*(D)$ under the notions of identifiability and mutual-information privacy, respectively, can be achieved by the same mechanism for D within certain range, i.e., there is a mechanism that simultaneously minimizes the identifiability level and the mutual information between X and Y. We further prove that this mutual-information optimal mechanism satisfies ϵ-differential privacy that is within a constant difference from the optimal differential privacy level for the given maximum allowable distortion:

$$\epsilon_{\mathrm{d}}^*(D) \leq \epsilon \leq \epsilon_{\mathrm{d}}^*(D) + 2\epsilon_X.$$

These results indicate certain consistency between identifiability and mutual-information privacy, and between differential privacy and mutual-information privacy when the prior p_X is uniform, although identifiability and differential privacy are defined based on "pairwise" requirements on distinguishability and are considered to be "worst-case" notions of privacy, while mutual-information privacy is defined by "global" requirements and is considered to be an "average" notion of privacy. The value of $\epsilon_{\mathrm{m}}^*(D)$ is in bits and thus is not directly comparable with $\epsilon_{\mathrm{i}}^*(D)$ and $\epsilon_{\mathrm{d}}^*(D)$, but the fact that identifiability and mutual-information privacy can be optimized simultaneously in the setting studied in this dissertation reveals the fundamental connections between these three privacy notions.

2.2 Model

Consider a database consisting of n rows, each of which corresponds to the data of a single individual. Each individual's data contains some sensitive information such as the individual's health status. Suppose that each row takes values from a domain $\mathcal{D}$. Then $\mathcal{D}^n$ is the set of all possible values of a database. Two databases, denoted by $x, x' \in \mathcal{D}^n$, are said to be *neighbors* if they differ in exactly one row. Let $x \sim x'$ denote the neighboring relation. In this dissertation, we assume that the domain $\mathcal{D}$ is a finite set and model a database as a discrete random variable X with alphabet $\mathcal{D}^n$ and probability mass function (pmf) p_X. Suppose $|\mathcal{D}| = m$, where m is an integer and $m \geq 2$. A (randomized) mechanism $\mathcal{M}$ takes a database x as the input, and outputs a random variable $\mathcal{M}(x)$.

Definition 1 (Mechanism)**.** A *mechanism* $\mathcal{M}$ is specified by an *associated mapping* $\phi_{\mathcal{M}} \colon \mathcal{D}^n \to \mathcal{F}$, where $\mathcal{F}$ is the set of multivariate cdf's on some range $\mathcal{R}$. Taking database X as the input, the mechanism $\mathcal{M}$ outputs a $\mathcal{R}$-valued random variable Y with $\phi_{\mathcal{M}}(x)$ as the multivariate conditional cdf of Y given $X = x$.

In this dissertation, we focus on mechanisms for which the range is the same as the alphabet of X, i.e., $\mathcal{R} = \mathcal{D}^n$. Then the output Y is also a discrete random variable with alphabet $\mathcal{D}^n$, which can be interpreted as a synthetic database. Denote the conditional pmf of Y given $X = x$ defined by the cdf $\phi_{\mathcal{M}}(x)$ as $p_{Y|X}(\cdot \mid x)$. Then a mechanism in this setting is fully specified by $p_{Y|X}$. When using this mechanism, the database curator samples from $p_{Y|X}(\cdot \mid x)$ to generate a synthetic database Y. The form of the mechanism is assumed to be public since it may be of interest to data analysts.

Throughout this dissertation we use the following basic notation. We denote the set of real numbers by $\mathbb{R}$, the set of nonnegative real numbers by $\mathbb{R}^+$, and the set of nonnegative integers by $\mathbb{N}$. Let $\overline{\mathbb{R}}^+ = \mathbb{R}^+ \cup \{+\infty\}$.

2.2.1 Different Notions of Privacy

In addition to the output database Y, we assume that the adversary also knows the prior distribution p_X, which represents the side information the adversary has, and the privacy-preserving mechanism $\mathcal{M}$. The three notions of privacy studied in this dissertation are defined next.

Definition 2 (Identifiability)**.** A mechanism $\mathcal{M}$ satisfies *ϵ-identifiability* for some $\epsilon \in \overline{\mathbb{R}}^+$ if for any pair of neighboring elements $x, x' \in \mathcal{D}^n$ and any $y \in \mathcal{D}^n$,

$$p_{X|Y}(x \mid y) \leq e^{\epsilon} p_{X|Y}(x' \mid y). \tag{2.1}$$

The notion of identifiability is defined based on the indistinguishability between any two neighboring databases from a Bayesian view. When a mechanism satisfies ϵ-identifiability for a small ϵ, two close (neighboring) databases cannot be distinguished from the posterior probabilities after observing the output database, which makes any individual's data hard to identify. To see the semantic implications of identifiability, we consider the following "worst-case" type of adversaries, who are called *informed adversaries* Dwork *et al.* (2006b). An adversary of this type knows $n - 1$ database entries and tries to identify the value of the remaining one. The notation of identifiability is defined based on neighboring databases to reflect this worst-case scenario. Consider adversaries who know X_{-i}, i.e., all the database entries except X_i. The requirement (2.1) of ϵ-identifiability indicates that for any $x_i, x_i' \in \mathcal{D}$, any $x_{-i} \in \mathcal{D}^{n-1}$ and any $y \in \mathcal{D}^n$,

$$\mathbb{P}\{X_i = x_i \mid X_{-i} = x_{-i}, Y = y\} \leq e^{\epsilon}\mathbb{P}\{X_i = x_i' \mid X_{-i} = x_{-i}, Y = y\}.$$

Therefore, when ϵ-identifiability is satisfied, even for such a worst-case adversary, the probability of correctly identifying the value of X_i is still no greater than $\frac{1}{1+(m-1)e^{-\epsilon}}$, which is close to randomly guessing when ϵ is small. We say that identifiability provides an *absolute* guarantee about disclosures since when it is satisfied, the probability of correctly identifying some individual's data is limited, and thus no bad disclosure can occur. This will become more clear when we discuss the relative guarantee provided by differential privacy.

We remark that in some cases, not all values of ϵ are achievable for ϵ-identifiability. The smallest achievable identifiability level is constrained by the prior p_X, since an adversary can always identify the values of the database entries based on the prior. When the prior itself is very disclosive, no mechanism can make the database entries less identifiable. To illustrate, we give the following example.

Example 1. Consider a database X with a single binary entry, i.e., $\mathcal{D} = \{0, 1\}$ and $n = 1$. Suppose the prior is given by $p_X(0) = 0.55$ and $p_X(1) = 0.45$. Consider the mechanism $\mathcal{M}$ specified by

$$p_{Y|X}(0 \mid 0) = p_{Y|X}(1 \mid 1) = 0.6, \quad p_{Y|X}(1 \mid 0) = p_{Y|X}(0 \mid 1) = 0.4.$$

Then the mechanism $\mathcal{M}$ satisfies ϵ-identifiability for $\epsilon \approx 0.6$. Therefore, the probability of correctly identifying X is guaranteed to be no greater than $\frac{1}{1+e^{-\epsilon}} \approx 0.65$. The smallest identifiability level that can be achieved for this prior is $\epsilon = \ln(0.55/0.45) \approx 0.2$. Now consider another prior that is given by $p_X(0) = 0.9$ and $p_X(1) = 0.1$. Then the mechanism $\mathcal{M}$ satisfies ϵ-identifiability for $\epsilon \approx 2.6$. In this case, no matter what mechanism is used, guessing that $X = 0$ yields a probability of correctness that is no less than 0.9. For an adversary with this prior, which indicates that the adversary has very good knowledge about the entry, no mechanism can achieve ϵ-identifiability for $\epsilon < \ln(0.9/0.1) \approx 2.2$.

Definition 3 (Differential Privacy Dwork *et al.* (2006b); Dwork (2006))**.** A mechanism $\mathcal{M}$ satisfies ϵ*-differential privacy* for some $\epsilon \in \overline{\mathbb{R}}^+$ if for any pair of neighboring elements $x, x' \in \mathcal{D}^n$ and any $y \in \mathcal{D}^n$,

$$p_{Y|X}(y \mid x) \leq e^{\epsilon} p_{Y|X}(y \mid x'). \tag{2.2}$$

Note that Definition 3 is equivalent to the definition of differential privacy in the seminal work Dwork *et al.* (2006b); Dwork (2006) under the model in this dissertation, although the languages used are slightly different. The differential privacy property of a mechanism is only determined by the associated mapping represented by $p_{Y|X}$ and does not depend on the prior.

In contrast to identifiability, differential privacy provides a *relative* guarantee about disclosures Dwork (2006). For any possible given disclosure about an individual, differential privacy ensures that only limited *additional* risk will be caused by the mechanism. To illustrate, we give the following example.

Example 2. We still consider the database X and the mechanism $\mathcal{M}$ in Example 1. The mechanism $\mathcal{M}$ satisfies ϵ-differential privacy for $\epsilon = \ln(0.6/0.4) \approx 0.4$ regardless of the prior p_X. If the prior is given by $p_X(0) = 0.9$ and $p_X(1) = 0.1$, then before seeing the output Y, the probability of correctly identifying X is 0.9. Suppose that the adversary observes an output $Y = 0$. Then the probability of correctly identifying X

becomes $\mathbb{P}(X = 0 \mid Y = 0) \approx 0.93$, which improves by a factor of approximately $e^{0.03}$. In this case, a bad disclosure occurs since the adversary is able to identify X with high probability, but differential privacy is still satisfied as the mechanism $\mathcal{M}$ guarantees that the probability of identification only increases by a bounded multiplicative factor.

Definition 4 (Mutual-Information Privacy)**.** A mechanism $\mathcal{M}$ satisfies ϵ-*mutual-information privacy* for some $\epsilon \in \overline{\mathbb{R}}^{+}$ if the mutual information between X and Y satisfies $I(X;Y) \leq \epsilon$, where

$$I(X;Y) = \sum_{x,y \in \mathcal{D}^n} p_{X,Y}(x,y) \log \frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)}.$$

The notion of mutual information is an information-theoretic notion of privacy, which measures the *average* amount of information about X contained in Y. The mutual information is minimized and equal to 0 when X and Y are independent, and it is maximized and equal to $H(X)$ when $Y = X$.

2.2.2 Distortion

In this dissertation, we measure the usefulness of a mechanism by the distortion between the input database X and the output Y, where smaller distortion corresponds to greater usefulness. Consider the (generalized) Hamming distance $d \colon \mathcal{D}^n \times \mathcal{D}^n \to \mathbb{N}$, where the distance $d(x, x')$ between any two elements $x, x' \in \mathcal{D}^n$ is the number of rows they differ in. We define the distortion between X and Y to be the expected Hamming distance

$$\mathbb{E}[d(X,Y)] = \sum_{x \in \mathcal{D}^n} \sum_{y \in \mathcal{D}^n} p_X(x) p_{Y|X}(y \mid x) d(x,y).$$

The Hamming distance also characterizes the neighboring relation on $\mathcal{D}^n$. Two elements $x, x' \in \mathcal{D}^n$ are neighbors if and only if $d(x,x') = 1$.

2.2.3 Privacy–Distortion Function

A privacy–distortion pair (ϵ, D) is said to be *achievable* if there exists a mechanism $\mathcal{M}$ with output Y such that $\mathcal{M}$ satisfies ϵ-privacy level and $\mathbb{E}[d(X,Y)] \leq D$. The *privacy–distortion function* $\epsilon^*\colon \mathbb{R}^+ \to \overline{\mathbb{R}}^+$ is defined by

$$\epsilon^*(D) = \inf\{\epsilon\colon (\epsilon, D) \text{ is achievable}\},$$

which is the smallest privacy level given the distortion constraint $\mathbb{E}[d(X,Y)] \leq D$. We are only interested in the range $[0,n]$ for D since this is the meaningful range for distortion. The privacy–distortion function depends on the prior p_X, which reflects the impact of the prior on the privacy–distortion tradeoff. To characterize the privacy–distortion function, we also consider the *distortion–privacy function* $D^*\colon \overline{\mathbb{R}}^+ \to \mathbb{R}^+$ defined by

$$D^*(\epsilon) = \inf\{D\colon (\epsilon, D) \text{ is achievable}\},$$

which is the smallest achievable distortion given privacy level ϵ.

In this dissertation we consider three different notions of privacy: identifiability, differential privacy and mutual-information privacy, so we denote the privacy–distortion functions under these three notions by ϵ_{i}^*, ϵ_{d}^* and ϵ_{m}^*, respectively.

2.3 Identifiability versus Differential Privacy

In this section, we establish a fundamental connection between identifiability and differential privacy. We characterize their privacy–distortion functions through studying the distortion–privacy functions. Given privacy level ϵ_{i} and ϵ_{d}, the minimum distortion level is the solution to the following optimization problems.

The Privacy–Distortion Problem under Identifiability (PD-I):

$$\min_{p_{X|Y}, p_Y} \quad \sum_{x \in \mathcal{D}^n} \sum_{y \in \mathcal{D}^n} p_Y(y) p_{X|Y}(x \mid y) d(x,y)$$

$$
\begin{aligned}
\text{subject to} \quad & p_{X|Y}(x \mid y) \le e^{\epsilon_{\mathrm{i}}} p_{X|Y}(x' \mid y), \\
& \qquad \forall x, x' \in \mathcal{D}^n \colon x \sim x', y \in \mathcal{D}^n, && (2.3) \\
& \sum_{x \in \mathcal{D}^n} p_{X|Y}(x \mid y) = 1, \qquad \forall y \in \mathcal{D}^n, && (2.4) \\
& p_{X|Y}(x \mid y) \ge 0, \qquad \forall x, y \in \mathcal{D}^n, && (2.5) \\
& \sum_{y \in \mathcal{D}^n} p_{X|Y}(x \mid y) p_Y(y) = p_X(x), \\
& \qquad \forall x \in \mathcal{D}^n, && (2.6) \\
& p_Y(y) \ge 0, \qquad \forall y \in \mathcal{D}^n. && (2.7)
\end{aligned}
$$

The Privacy–Distortion Problem under Differential Privacy (PD-DP):

$$
\begin{aligned}
\min_{p_{Y|X}} \quad & \sum_{x \in \mathcal{D}^n} \sum_{y \in \mathcal{D}^n} p_X(x) p_{Y|X}(y \mid x) d(x, y) \\
\text{subject to} \quad & p_{Y|X}(y \mid x) \le e^{\epsilon_{\mathrm{d}}} p_{Y|X}(y \mid x'), \\
& \qquad \forall x, x' \in \mathcal{D}^n \colon x \sim x', y \in \mathcal{D}^n, && (2.8) \\
& \sum_{y \in \mathcal{D}^n} p_{Y|X}(y \mid x) = 1, \qquad \forall x \in \mathcal{D}^n, && (2.9) \\
& p_{Y|X}(y \mid x) \ge 0, \qquad \forall x, y \in \mathcal{D}^n. && (2.10)
\end{aligned}
$$

Note that to obtain the distortion–privacy functions, we need to find a mechanism $p_{Y|X}$ to minimize the distortion subject to privacy constraints. However, for identifiability, since it is defined based on $p_{X|Y}$, we change the optimization variable from $p_{Y|X}$ to $(p_{X|Y}, p_Y)$ in PD-I, and the constraints (2.4)–(2.7) ensure that PD-I is equivalent to the original distortion–privacy problem.

For convenience, we first define two constants ϵ_X and $\widetilde{\epsilon}_X$ that are determined by the prior p_X. Let

$$
\epsilon_X = \max_{x, x' \in \mathcal{D}^n : x \sim x'} \ln \frac{p_X(x)}{p_X(x')}, \tag{2.11}
$$

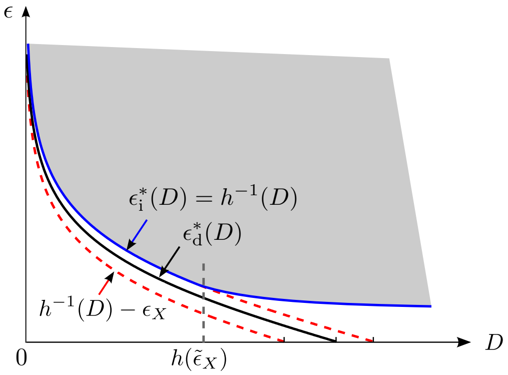


Figure 2.2: The privacy–distortion functions ϵ_{i}^* under identifiability and ϵ_{d}^* under differential privacy satisfy $\epsilon_{\mathrm{i}}^*(D) - \epsilon_X \leq \epsilon_{\mathrm{d}}^*(D) \leq \epsilon_{\mathrm{i}}^*(D)$ for D within certain range.

which is the maximum prior probability difference between two neighboring databases. For ϵ_X to be finite, the prior distribution p_X needs to have full support on $\mathcal{D}^n$, i.e., $p_X(x) > 0$ for any $x \in \mathcal{D}^n$. To define $\widetilde{\epsilon}_X$, note that the prior p_X puts constraints on the posterior probabilities, as given by the constraint (2.6) in PD-I. We say $\{p_{X|Y}(x \mid y), x, y \in \mathcal{D}^n\}$ is *feasible* if there exists a pmf p_Y such that it is the marginal pmf of Y. Let $\widetilde{\epsilon}_X$ be the smallest ϵ such that the following posterior probabilities are feasible:

$$p_{X|Y}(x \mid y) = \frac{e^{-\epsilon d(x,y)}}{\big(1 + (m-1)e^{-\epsilon}\big)^n}, \quad x, y \in \mathcal{D}^n.$$

We will see that the $p_{X|Y}$ in the above form plays an important role in solving PD-I. For any p_X, $\widetilde{\epsilon}_X$ is finite since when $\epsilon \to +\infty$, the pmf $p_Y = p_X$ is the marginal pmf of Y. Finally we define the function

$$h^{-1}(D) = \ln\Big(\frac{n}{D} - 1\Big) + \ln(m-1).$$

Recall that $\epsilon_{\mathrm{i}}^*(D)$ and $\epsilon_{\mathrm{d}}^*(D)$ denote the minimum identifiability level and minimum differential privacy level for a maximum allowable distortion D. The connection between the privacy–distortion functions ϵ_{i}^* and ϵ_{d}^* is established in the following theorem. See Figure 2.2 for an illustration.

Theorem 1. *For identifiability, the privacy–distortion function* ϵ_{i}^* *of a database* X *with* $\epsilon_X < +\infty$ *satisfies*

$$\begin{cases} \epsilon_{\mathrm{i}}^*(D) = h^{-1}(D), & 0 \leq D \leq h(\widetilde{\epsilon}_X), \\ \epsilon_{\mathrm{i}}^*(D) \geq \max\{h^{-1}(D), \epsilon_X\}, & h(\widetilde{\epsilon}_X) < D \leq n. \end{cases} \tag{2.12}$$

For differential privacy, the privacy–distortion function ϵ_{d}^* *of a database* X *satisfies the following bounds for any* D *with* $0 \leq D \leq n$*:*

$$\max\{h^{-1}(D) - \epsilon_X, 0\} \leq \epsilon_{\mathrm{d}}^*(D) \leq \max\{h^{-1}(D), 0\}. \tag{2.13}$$

From the theorem above, we can see that $0 \leq \epsilon_{\mathrm{i}}^*(D) - \epsilon_{\mathrm{d}}^*(D) \leq \epsilon_X$ when $0 \leq D \leq h(\widetilde{\epsilon}_X)$. The lemmas needed in the proof of this theorem can be found in Appendix A. Here we give a sketch of the proof, which consists of the following key steps:

- The first key step is to show that both PD-I and PD-DP, through (respective) relaxations as shown in Figure 2.3, boil down to the same optimization problem.

  Relaxed Privacy–Distortion (R-PD):

$$\begin{aligned} \min_{p_{X|Y}, p_Y} \quad & \sum_{x \in \mathcal{D}^n} \sum_{y \in \mathcal{D}^n} p_Y(y) p_{X|Y}(x \mid y) d(x, y) \\ \text{subject to} \quad & p_{X|Y}(x \mid y) \leq e^{\epsilon} p_{X|Y}(x' \mid y), \\ & \qquad \forall x, x' \in \mathcal{D}^n \colon x \sim x', y \in \mathcal{D}^n, \qquad (2.14) \\ & \sum_{x \in \mathcal{D}^n} p_{X|Y}(x \mid y) = 1, \qquad \forall y \in \mathcal{D}^n, \qquad (2.15) \\ & p_{X|Y}(x \mid y) \geq 0, \qquad \forall x, y \in \mathcal{D}^n, \qquad (2.16) \\ & \sum_{y \in \mathcal{D}^n} p_Y(y) = 1, \qquad (2.17) \\ & p_Y(y) \geq 0, \qquad \forall y \in \mathcal{D}^n. \qquad (2.18) \end{aligned}$$

  Relaxing the constraint (2.6) in PD-I to the constraint (2.17) gives R-PD. Now consider PD-DP. For any neighboring $x, x' \in \mathcal{D}^n$, $p_X(x) \leq e^{\epsilon_X} p_X(x')$ according

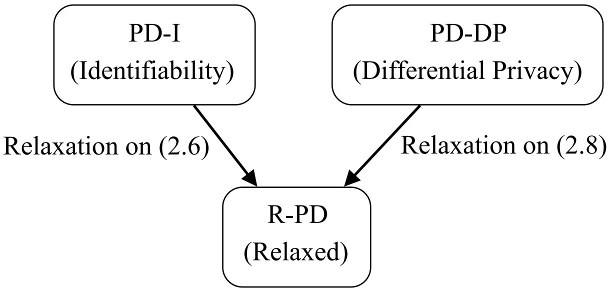


Figure 2.3: Both PD-I and PD-DP boil down to R-PD through different relaxations.

to the definition of ϵ_X, and a necessary condition for the constraint (2.8) to be satisfied is

$$p_X(x)p_{Y|X}(y \mid x) \leq e^{\epsilon_{\mathrm{d}}+\epsilon_X} p_X(x')p_{Y|X}(y \mid x'). \tag{2.19}$$

Therefore, replacing constraint (2.8) with (2.19) and letting $\epsilon = \epsilon_{\mathrm{d}} + \epsilon_X$, we obtain R-PD. So R-PD can be regarded as a relaxation of both PD-I and PD-DP.

- To solve R-PD, it suffices to solve the following optimization problem for any fixed $y \in \mathcal{D}^n$:

$$\begin{aligned}
\min_{p_{X|Y}} \quad & \sum_{x \in \mathcal{D}^n} p_{X|Y}(x \mid y) d(x, y) \\
\text{subject to} \quad & p_{X|Y}(x \mid y) \leq e^{\epsilon} p_{X|Y}(x' \mid y), \\
& \qquad\qquad \forall x, x' \in \mathcal{D}^n \colon x \sim x', \\
& \sum_{x \in \mathcal{D}^n} p_{X|Y}(x \mid y) = 1, \\
& p_{X|Y}(x \mid y) \geq 0, \qquad \forall x \in \mathcal{D}^n.
\end{aligned}$$

Intuitively, to minimize the objective function, which is the average distortion between X and y, we should assign larger probability to $p_{X|Y}(x \mid y)$ with smaller $d(x, y)$, and smaller probability to $p_{X|Y}(x \mid y)$ with larger $d(x, y)$. For the x such that $x = y$, we should assign the largest value to $p_{X|Y}(x \mid y)$ since $d(x, y) = 0$, and as x goes far way from y, we should assign smaller and smaller values to $p_{X|Y}(x \mid y)$. However, the privacy constraint limits the decreasing rate we can use as x goes far away from y due to the neighboring relations. In Lemma 7, we prove that the optimal solution is given by

$$p_{X|Y}(x \mid y) = \frac{e^{-\epsilon d(x,y)}}{\big(1 + (m-1)e^{-\epsilon}\big)^n}, \quad x, y \in \mathcal{D}^n, \tag{2.20}$$

where the probability $p_{X|Y}(x \mid y)$ decreases with rate e^{ϵ} as $d(x, y)$ increases. This is the fastest possible decreasing rate with the privacy constraint, so this solution gives the smallest distortion.

- By Lemma 7, the minimum distortion of R-PD is $D^*_{\text{relaxed}}(\epsilon) = h(\epsilon)$, which gives lower bounds on the distortion–privacy functions under identifiability and under differential privacy. By the connection between distortion–privacy function and privacy–distortion function, Lemma 8 shows that $\epsilon^*_{\text{i}}(D) \geq h^{-1}(D)$ and $\epsilon^*_{\text{d}}(D) \geq h^{-1}(D) - \epsilon_X$ for any D with $0 \leq D \leq n$. Lemma 9 shows another lower bound on ϵ^*_{i}, combining which with the lower bound in Lemma 8 gives the lower bound in Theorem 1.

- Next we design achievable mechanisms to prove the upper bounds in Theorem 1. Notice that when the posterior probabilities given by the solution $p_{X|Y}$ in (2.20) is feasible, the mechanism that corresponds to this $p_{X|Y}$ satisfies ϵ-identifiability. Therefore, the lower bound for identifiability is achievable in this case. Consider the mechanism $\mathcal{E}^{\epsilon}_{\text{i}}$ specified by

$$p_{Y|X}(y \mid x) = \frac{p_Y(y) e^{-\epsilon d(x,y)}}{p_X(x)\big(1 + (m-1)e^{-\epsilon}\big)^n}, \quad x, y \in \mathcal{D}^n, \tag{2.21}$$

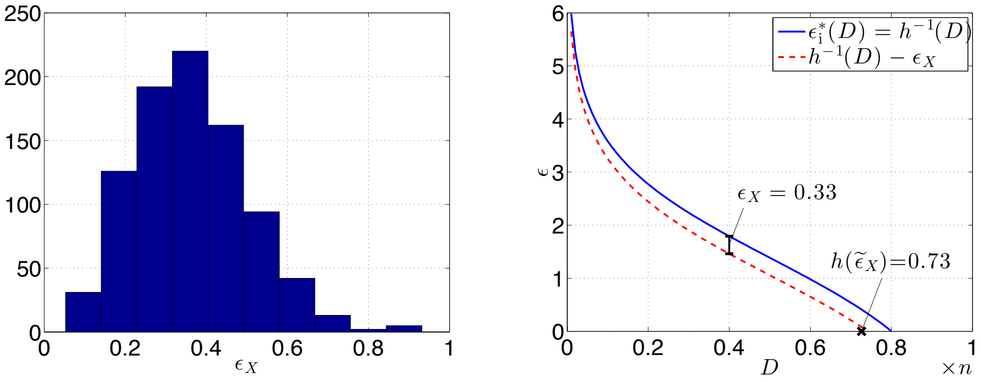
where $\epsilon \geq \widetilde{\epsilon}_X$ and p_Y is the corresponding pmf of Y. The mechanism $\mathcal{E}^{\epsilon}_{\text{i}}$ corresponds to the posterior distributions given by $p_{X|Y}$ in (2.20). Lemma 10 shows that the mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ guarantees an identifiability level of ϵ with distortion $h(\epsilon)$ when $\epsilon \geq \widetilde{\epsilon}_X$, which yields the equality in (2.12) when combining with the lower bound above.

- For differential privacy, consider the mechanism $\mathcal{E}_{\mathrm{d}}^{\epsilon}$ specified by the conditional probabilities

$$p_{Y|X}(y \mid x) = \frac{e^{-\epsilon d(x,y)}}{\big(1+(m-1)e^{-\epsilon}\big)^n}, \quad x, y \in \mathcal{D}^n, \tag{2.22}$$

where $\epsilon \geq 0$. Note that in contrast with the mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$, the mechanism $\mathcal{E}_{\mathrm{d}}^{\epsilon}$ itself has the same form as the solution $p_{X|Y}$ in (2.20). Lemma 11 shows that the mechanism $\mathcal{E}_{\mathrm{d}}^{\epsilon}$ satisfies ϵ-differential privacy with distortion $h(\epsilon)$, which provides the upper bound in (2.13). We remark that the mechanism $\mathcal{E}_{\mathrm{d}}^{\epsilon}$ has the same form as an exponential mechanism with score function $q = -d$ McSherry and Talwar (2007), where the score function has a sensitivity $\Delta q = 1$. In general, an exponential mechanism with parameter ϵ is $2\epsilon\Delta q$-differentially private. However, the mechanism $\mathcal{E}_{\mathrm{d}}^{\epsilon}$ is ϵ-differentially private without the factor 2 since the normalizing term in the denominator of (2.22) does not depend on x.

Illustration. We demonstrate the characterizations of the privacy–distortion functions in Theorem 1 using prior distributions based on a databaset constructed for Netflix Prize. The dataset consists of movie ratings from users, with each rating on a scale from 1 to 5 (integer) stars. We view the ratings of a movie from active users as a database and generate ratings uniformly at random for missing entries. We first calculate the corresponding ϵ_X, assuming that entries of a database are drawn i.i.d. from a distribution. The constant ϵ_X bounds the gap between the upper and lower bounds on $\epsilon_{\mathrm{d}}^*(D)$, and also bounds $\epsilon_{\mathrm{i}}^*(D) - \epsilon_{\mathrm{d}}^*(D)$. In Figure 2.4a, we show the histogram of ϵ_X for 887 most reviewed movies (databases). Next, we pick a database



(a)

(b)

Figure 2.4: Illustration of the characterizations of the privacy–distortion functions in Theorem 1. (a) Histogram of ϵ_X for 887 databases. (b) The privacy–distortion function under identifiability is given by $\epsilon_{\mathrm{i}}^*(D) = h^{-1}(D)$ for $0 \leq D \leq h(\widetilde{\epsilon}_X)$, where $h(\widetilde{\epsilon}_X) = 0.73n$. The privacy–distortion function under differential privacy, $\epsilon_{\mathrm{d}}^*(D)$, lies between $\epsilon_{\mathrm{i}}^*(D) = h^{-1}(D)$ and $h^{-1}(D) - \epsilon_X$, where $\epsilon_X = 0.33$.

whose prior distribution of each entry is given by

$$p_{X_i}(1) = 0.2533, \quad p_{X_i}(2) = 0.1821, \quad p_{X_i}(3) = 0.1821,$$
$$p_{X_i}(4) = 0.1873, \quad p_{X_i}(5) = 0.1953.$$

For this prior, we have $\epsilon_X = 0.33$ and $\widetilde{\epsilon}_X = 0.41$. In Figure 2.4b, we draw the privacy–distortion function $\epsilon_{\mathrm{i}}^*(D) = h^{-1}(D)$ under identifiability for $0 \leq D \leq h(\widetilde{\epsilon}_X)$, where the value $h(\widetilde{\epsilon}_X) = 0.73n$ is displayed in the figure. The curve $\epsilon_{\mathrm{i}}^*(D) = h^{-1}(D)$ gives an upper bound on the privacy–distortion function $\epsilon_{\mathrm{d}}^*(D)$ under differential privacy. We also draw the curve $\max\{h^{-1}(D) - \epsilon_X, 0\}$, which is a lower bound on $\epsilon_{\mathrm{d}}^*(D)$.

2.4 Identifiability, Differential Privacy versus Mutual-Information Privacy

In this section, we first discuss the relation between identifiability and mutual-information privacy. Then we further establish a connection between differential privacy and mutual-information privacy based on this relation between identifiability and mutual-information privacy and the relation between identifiability and differential privacy derived in the last section.

Theorem 2. *For any D with $0 \leq D \leq h(\widetilde{\epsilon}_X)$, the identifiability optimal mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ with $\epsilon = h^{-1}(D)$ is also mutual-information optimal.*

By this theorem, the privacy–distortion functions $\epsilon_{\mathrm{i}}^*(D)$ and $\epsilon_{\mathrm{m}}^*(D)$ under the notions of identifiability and mutual-information privacy, respectively, can be achieved by the same mechanism for D within certain range. This theorem indicates a consistency between identifiability and mutual-information privacy under the privacy–distortion framework since they can be optimized simultaneously.

Recall that given a maximum allowable distortion D, the privacy–distortion function $\epsilon_{\mathrm{m}}^*(D)$ under mutual-information privacy for an input database X with prior p_X is given by the optimal value of the following convex optimization problem.

The Privacy and Distortion Problem under Mutual-Information Privacy (PD-MIP):

$$\begin{aligned}
\min_{p_{Y|X}} \quad & I(X;Y) \\
\text{subject to} \quad & \sum_{x\in\mathcal{D}^n}\sum_{y\in\mathcal{D}^n} p_X(x)p_{Y|X}(y \mid x)d(x,y) \leq D, && (2.23)\\
& \sum_{y\in\mathcal{D}^n} p_{Y|X}(y \mid x) = 1, \qquad \forall x \in \mathcal{D}^n, && (2.24)\\
& p_{Y|X}(y \mid x) \geq 0, \qquad \forall x, y \in \mathcal{D}^n. && (2.25)
\end{aligned}$$

Note that this formulation has the same form as the formulation in the celebrated rate–distortion theory (e.g., see Cover and Thomas (2006)), and thus the privacy–distortion function under mutual-information privacy is identical to the rate–distortion function in this setting. Studies on the rate–distortion function Blahut (1972); Cover and Thomas (2006) have revealed the structure of an optimal solution of PD-MIP using Karush-Kuhn-Tucker (KKT) conditions Boyd and Vandenberghe (2004). We utilize these results to prove Theorem 2.

Proof of Theorem 2. By the KKT conditions for PD-MIP, the mutual information is minimized by

$$p_{Y|X}(y \mid x) = \frac{p_Y(y)e^{-\lambda d(x,y)}}{\sum_{y' \in \mathcal{D}^n} p_Y(y')e^{-\lambda d(x,y')}}, \quad x, y \in \mathcal{D}^n,$$

if there exists a pmf p_Y of Y and $\lambda \geq 0$ such that

$$\sum_{x \in \mathcal{D}^n} \frac{p_X(x)e^{-\lambda d(x,y)}}{\sum_{y' \in \mathcal{D}^n} p_Y(y')e^{-\lambda d(x,y')}} = 1, \quad \text{if } p_Y(y) > 0, \tag{2.26}$$

$$\sum_{x \in \mathcal{D}^n} \frac{p_X(x)e^{-\lambda d(x,y)}}{\sum_{y' \in \mathcal{D}^n} p_Y(y')e^{-\lambda d(x,y')}} \leq 1, \quad \text{if } p_Y(y) = 0, \tag{2.27}$$

$$\lambda\left(\sum_{x \in \mathcal{D}^n} \sum_{y \in \mathcal{D}^n} \frac{p_X(x)p_Y(y)e^{-\lambda d(x,y)}}{\sum_{y' \in \mathcal{D}^n} p_Y(y')e^{-\lambda d(x,y')}} d(x,y) - D\right) = 0, \tag{2.28}$$

where λ is the Lagrange multiplier for the distortion constraint (2.23). This optimal solution has an exponential form. Recall that the identifiability optimal mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ in (2.21) also has an exponential form. In what follows we prove that for properly chosen λ, the conditions (2.26)–(2.28) are satisfied under $\mathcal{E}_{\mathrm{i}}^{\epsilon}$.

For any $0 \leq D \leq h(\widetilde{\epsilon}_X)$, consider the mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ with $\epsilon = h^{-1}(D)$. Let $\lambda = \epsilon$. Recall that under $\mathcal{E}_{\mathrm{i}}^{\epsilon}$,

$$p_{Y|X}(y \mid x) = \frac{p_Y(y)e^{-\epsilon d(x,y)}}{p_X(x)\bigl(1 + (m-1)e^{-\epsilon}\bigr)^n}, \quad x, y \in \mathcal{D}^n.$$

Since $p_{Y|X}$ satisfies that

$$\sum_{y' \in \mathcal{D}^n} p_{Y|X}(y' \mid x) = 1,$$

we have

$$\sum_{y' \in \mathcal{D}^n} p_Y(y') e^{-\epsilon d(x,y')} = p_X(x)\big(1 + (m-1)e^{-\epsilon}\big)^n.$$

Then for any $y \in \mathcal{D}^n$,

$$\begin{aligned} &\sum_{x \in \mathcal{D}^n} \frac{p_X(x) e^{-\epsilon d(x,y)}}{\sum_{y' \in \mathcal{D}^n} p_Y(y') e^{-\epsilon d(x,y')}} \\ &= \sum_{x \in \mathcal{D}^n} \frac{p_X(x) e^{-\epsilon d(x,y)}}{p_X(x)\big(1 + (m-1)e^{-\epsilon}\big)^n} \\ &= 1, \end{aligned}$$

which indicates that (2.26) and (2.27) are satisfied. We can verify that

$$\begin{aligned} &\sum_{x \in \mathcal{D}^n} \sum_{y \in \mathcal{D}^n} \frac{p_X(x) p_Y(y) e^{-\epsilon d(x,y)}}{\sum_{y' \in \mathcal{D}^n} p_Y(y') e^{-\epsilon d(x,y')}} d(x,y) \\ &= \sum_{y \in \mathcal{D}^n} p_Y(y) \sum_{x \in \mathcal{D}^n} \frac{p_X(x) e^{-\epsilon d(x,y)} d(x,y)}{p_X(x)\big(1 + (m-1)e^{-\epsilon}\big)^n} \\ &= h(\epsilon) \\ &= D, \end{aligned}$$

which indicates that (2.28) is satisfied. Therefore, the mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ with $\epsilon = h^{-1}(D)$ gives an optimal solution of PD-MIP, which completes the proof. □

Next, we establish a connection between differential privacy and mutual-information privacy based on Theorem 2 and Theorem 1.

Corollary 1. *For any D with $0 \le D \le h(\widetilde{\epsilon}_X)$, the mutual-information optimal mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ with $\epsilon = h^{-1}(D)$ is ϵ_{d}-differentially private with $\epsilon_{\mathrm{d}}^*(D) \le \epsilon_{\mathrm{d}} \le \epsilon_{\mathrm{d}}^*(D) + 2\epsilon_X$.*

It has been pointed out in Mir (2013) that a mechanism that achieves the optimal rate–distortion also guarantees a certain level of differential privacy. However, whether this differential privacy level is optimal or how far it is from optimal was not answered. Our result in Corollary 1 further shows that the gap between the differential privacy level of the mutual-information optimal mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ and the optimal differential privacy level is no greater than $2\epsilon_X$, which is a constant determined by the prior p_X. Therefore, given a distortion constraint, optimizing for mutual information leads to a differentially private mechanism whose privacy level is close to the optimal differential privacy level. When the prior is uniform, this mutual-information optimal mechanism achieves exactly the optimal differential privacy level. Similar to the relation between identifiability and mutual-information privacy, differential privacy and mutual-information privacy also show a consistency for uniform prior under the privacy–distortion framework, although differential privacy is usually considered to be a “worst-case” notion of privacy and mutual-information is usually considered to be an “average” notion of privacy.

Proof of Corollary 1. By Theorem 2, the mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ with $\epsilon = h^{-1}(D)$ is mutual-information optimal. According to its form, we can verify that $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ with $\epsilon = h^{-1}(D)$ is ϵ_{d}-differentially private with $\epsilon_{\mathrm{d}} = h^{-1}(D) + \epsilon_X$. Since $\epsilon_{\mathrm{d}}^{*}(D)$ is the minimum differential privacy level with distortion constraint given by D, we have $\epsilon_{\mathrm{d}} \geq \epsilon_{\mathrm{d}}^{*}(D)$. By Theorem 1, $h^{-1}(D) \leq \epsilon_{\mathrm{d}}^{*}(D) + \epsilon_X$. Thus $\epsilon_{\mathrm{d}} \leq \epsilon_{\mathrm{d}}^{*}(D) + 2\epsilon_X$, which completes the proof. □

2.5 Conclusions

In this chapter, we investigated the relation between three different notions of privacy: identifiability, differential privacy and mutual-information privacy, where identifiability guarantees indistinguishability between posterior probabilities, differential privacy guarantees limited additional disclosures, and mutual information is an information-theoretic notion. Under a unified privacy–distortion framework, where the distortion is defined to be the expected Hamming distance between the input and output databases, we established some fundamental connections between these three privacy notions. Given a maximum allowable distortion D within certain range, the smallest identifiability level $\epsilon_{\mathrm{i}}^*(D)$ and the smallest differential privacy level $\epsilon_{\mathrm{d}}^*(D)$ are proved to satisfy $\epsilon_{\mathrm{i}}^*(D) - \epsilon_X \leq \epsilon_{\mathrm{d}}^*(D) \leq \epsilon_{\mathrm{i}}^*(D)$, where ϵ_X is a constant determined by the prior of the original database, and diminishes to zero when the prior is uniform. Next, we showed that there is a mechanism that simultaneously minimizes the identifiability level and the mutual information given the same maximum allowable distortion within certain range. We further showed that this mechanism satisfies ϵ-differential privacy with $\epsilon_{\mathrm{d}}^*(D) \leq \epsilon \leq \epsilon_{\mathrm{d}}^*(D) + 2\epsilon_X$.

Our findings in this study reveal some fundamental connections between the three notions of privacy. With these three notions of privacy being defined, many interesting issues deserve further attention. The connections we have established in this work are based on the distortion measure of Hamming distance, which is closely tied with the neighboring relations, and we assume that the output synthetic database and the original database are in the same universe. It would be of great interest to study the connections of these privacy notions under other common distortion measures and other output formats. We remark that our results for Hamming distance can be used to prove lower bounds on the distortion of a differentially private mechanism when the distortion is measured by the distortion at the worst-case query in a query class Wang *et al.* (2015b). Some other interesting directions are as follows. In some cases, the prior p_X is imperfect. Then for privacy notions depending on the prior such as identifiability and mutual-information privacy, it is natural to ask how we can protect privacy with robustness over the prior distribution. Identifiability and differential privacy impose requirements on neighboring databases to protect an individual's privacy. Then are there any practical scenarios that we would desire to generalize this "pairwise" privacy to "group" privacy? The connections between membership privacy and these three notions of privacy also need to be explored, since membership privacy has been proposed as a unifying framework for privacy definitions.

Chapter 3

MARKET MODEL OF TRADING PRIVATE DATA

3.1 Introduction

We consider a game-theoretic model of collecting private data in hypothesis testing, where the data collector is interested in learning information from a population of N individuals. An illustration of our model is shown in Figure 3.1. The information is represented by a binary random variable W, which is called the *state*. Each individual i possesses a binary *signal* S_i, which is her private data, representing her knowledge about the state W. Conditional on the state W, the signals are independently generated such that the probability for each signal S_i to be the same as W is θ, where $0.5 < \theta < 1$. To protect her privacy, an individual reports only a privacy-preserving version of her signal, denoted by X_i, or chooses to not participate after considering both the payment from the data collector and the loss of privacy. The data collector needs to decide the amount of payment and the payment mechanism to get informative reports, i.e., not completely random data. Intuitively, the higher the payment is, the more informative the reported data should be. We will answer the following fundamental questions in this dissertation: *What is the minimum payment needed from the data collector to obtain reported data with a privacy level ϵ? Which payment mechanism can be used to collect private data with minimum cost?* This setting without accounting for data privacy has garnered much attention in the literature (see, e.g., Miller *et al.* (2009); Acemoglu *et al.* (2011); Le *et al.* (2014)), including the application of estimating the underlying value of a new technology by eliciting opinions from individuals.

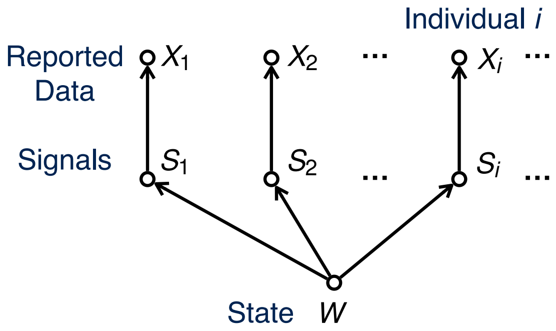


Figure 3.1: Information structure of the model. The data collector is interested in the state W, which is a binary random variable. Each individual i possesses a binary signal S_i, which is her private data. Conditional on W, $S_1, S_2, \ldots, S_N$ are i.i.d. Individual i's reported data is X_i, which is generated based on S_i using a randomized strategy.

Intuitively, the data collector can purchase more informative data (so higher privacy) by offering higher payment. However, the strategic behavior of the privacy-aware individuals makes this more complicated. Due to privacy concerns, an individual's action/strategy is the conditional distributions of the reported data given the realizations of the signal. But the actions of the individuals are not observable to the data collector. Instead, what the data collector receives is the reported data, generated randomly according to the individuals' strategies, so the payments can only be designed based on the reported data. This differs our problem from the conventional mechanism design.

Furthermore, the privacy-aware individuals weigh the privacy loss against the payment to choose the best quantity of privacy to trade. To make an individual willing to trade ϵ level of privacy, the data collector needs to make sure doing this benefits the individual most. We reiterate that the data collector has access only to the reported data instead of the individuals' actions. Note that only compensating the privacy cost incurred is not sufficient. The payment mechanism needs to ensure that ϵ is the best privacy level such that when an individual uses a less-private strategy, the decrease in her payment is faster than the decrease in her privacy cost, and similarly, when an individual uses a more-private strategy, the increase in her payment is slower than the increase in her privacy cost. In other words, with a game-theoretic approach, we consider an individual's best response in a Nash equilibrium, and the value of data privacy is measured by the minimum payment that makes this equilibrium strategy have a privacy level of ϵ, which represents the monetary value of data privacy in a market for private data.

Summary of Results

It is assumed that individuals use the celebrated notion of differential privacy Dwork *et al.* (2006b); Dwork (2006) to evaluate their data privacy. When an individual i uses an ϵ-differentially private randomization strategy to generate X_i, the privacy loss incurred is ϵ, and the individual's cost of privacy loss is a function of ϵ, whose form is assumed to be publicly known. The value of ϵ units of privacy is measured by the minimum payment of all nonnegative payment mechanisms under which an individual's best response in a Nash equilibrium is to report the data with a privacy level of ϵ, where nonnegativity ensures that individuals would not be *charged* for reporting data. Denote this value by $V(\epsilon)$. Our contributions are summarized as follows:

1. We establish a lower bound on $V(\epsilon)$. First we characterize the strategies of individuals at a Nash equilibrium to prove that from a payment perspective, it suffices to focus on nonnegative payment mechanisms under which the best response of an individual in a Nash equilibrium is a symmetric randomized response with a privacy level of ϵ. This strategy generates the reported data by flipping the signal

with probability $\frac{1}{e^{\epsilon}+1}$: for convenience, this is called the ϵ-strategy. Next we prove that the expected payments resulting from any Nash equilibrium of any payment mechanism can be "replicated" by a genie-aided payment mechanism, where the payments are determined with the aid of a genie who knows the underlying state W. This makes the analysis of the Nash equilibria more tractable by decoupling the individuals. The lower bound is then given by necessary conditions for ϵ to be the best privacy level in the genie-aided mechanism. We remark that although the genie-aided mechanism that achieves the lower bound is not implementable, it can be well-approximated, when the number of individuals is large, by the feasible payment mechanism that we design to prove the upper bound.

2. We observe that the equilibrium strategies exhibit some interesting characteristics: the strategy of an individual in a Nash equilibrium is either a symmetric randomized response, which treats the realizations of the private signal symmetrically, or a non-informative strategy, where the reported data is independent of the signal. This characterization holds regardless of the prior distribution of the state, and it also holds for more general probability models of the signals. This characterization advances our understanding of the behavior of privacy-aware individuals. It is worth pointing out that finding an equilibrium strategy of a privacy-aware individual under some payment mechanism involves non-convex optimization.

3. We prove an upper bound on $V(\epsilon)$ by designing a payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$, in which the strategy profile consisting of ϵ-strategies constitutes a Nash equilibrium. The expected payment to each individual at this equilibrium gives an upper bound on $V(\epsilon)$. This upper bound converges to the lower bound exponentially fast as the number of individuals N becomes large, which indicates that the lower and upper bounds are asymptotically tight.

4. The above fundamental bounds on the value of privacy can be further used to study the *payment–accuracy problem*, where the data collector aims to minimize the total payment while achieving an accuracy target in learning the state W. Given an accuracy target τ, which can be regarded as the maximum allowable error, let $F(\tau)$ denote the minimum total payment for achieving τ. We obtain lower and upper bounds on $F(\tau)$ based on the lower and upper bounds on the value of privacy. The upper bound is given by the designed mechanism $\boldsymbol{R}^{(N,\epsilon)}$ with properly chosen parameters, which shows that the total payment of the designed mechanism is at most one individual's payment away from the minimum.

3.2 Model

We consider a single-bit learning problem with privacy-aware individuals as shown in Figure 3.1. Recall that the data collector is interested in learning the state W, which is a binary random variable. For example, the state W can describe the underlying value of some new technology. Let P_W denote the prior PMF of W. We assume that $P_W(1) > 0$, $P_W(0) > 0$, and the prior is common knowledge.

Individuals and Strategies. Consider a population of N individuals and denote the set of individuals by $\mathcal{N} = \{1, 2, \ldots, N\}$. Denote all individuals other than some given individual i by "$-i$." Each individual i possesses a bit S_i, which is her private data, reflecting her knowledge about the state W. For example, the signal S_i can represent individual i's opinion towards the new technology. We call S_i individual i's signal. Let $\boldsymbol{S} = (S_1, S_2, \cdots, S_N)$. Conditional on either value of the state W, the signals $S_1, S_2, \ldots, S_N$ are i.i.d. with the conditional distributions below, where the parameter θ with $0.5 < \theta < 1$ is called *the quality of signals* since larger value of θ

means that each signal is equal to the state with higher probability:

$$\mathbb{P}(S_i = 1 \mid W = 1) = \theta, \quad \mathbb{P}(S_i = 0 \mid W = 1) = 1 - \theta,$$
$$\mathbb{P}(S_i = 0 \mid W = 0) = \theta, \quad \mathbb{P}(S_i = 1 \mid W = 0) = 1 - \theta.$$

Let X_i denote the data reported by individual i and let $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$. The acceptable values for reported data are 0, 1, and "nonparticipation." So X_i takes values in the set $\mathcal{X} = \{0, 1, \perp\}$, where $\perp$ indicates that individual i declines to participate. A strategy of individual i for data reporting is a mapping $\sigma_i \colon \{0,1\} \to \mathcal{D}(\mathcal{X})$, where $\mathcal{D}(\mathcal{X})$ is the set of probability distributions on $\mathcal{X}$. Let $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$. The strategy σ_i prescribes a distribution to X_i for each possible value of S_i, which defines the conditional distribution of X_i given S_i. Since we will discuss different strategies of individual i, we let $\mathbb{P}_{\sigma_i}(X_i = x_i \mid S_i = s_i)$ with $x_i \in \mathcal{X}$ and $s_i \in \{0,1\}$ denote the conditional probabilities defined by strategy σ_i. If a strategy σ_i satisfies that $\mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 1) = \mathbb{P}_{\sigma_i}(X_i = 0 \mid S_i = 0)$ and $\mathbb{P}_{\sigma_i}(X_i = \perp \mid S_i = 1) = \mathbb{P}_{\sigma_i}(X_i = \perp \mid S_i = 0) = 0$, we say σ_i is a *symmetric randomized response*. If a strategy σ_i makes X_i and S_i independent, we say σ_i is *non-informative*; otherwise we say σ_i is *informative*.

Mechanism. The data collector uses a payment mechanism $\boldsymbol{R} \colon \mathcal{X}^N \to \mathbb{R}^N$ to determine the amount of payment to each individual, where $R_i(\boldsymbol{x})$ is the payment to individual i when the reported data is $\boldsymbol{X} = \boldsymbol{x}$. We are interested in payment mechanisms in which the payment to each individual is nonnegative, i.e., $R_i(\boldsymbol{x}) \geq 0$ for any individual i and any $\boldsymbol{x} \in \mathcal{X}^N$, which we call *nonnegative mechanisms*. This constraint is motivated by the fact that in many practical applications such as surveys, the data collector has no means to charge users and can only use payments to incentivize user participation.

Privacy Cost. We quantify the privacy loss incurred when a strategy is in use by the level of (local) differential privacy of the strategy (Dwork *et al.* (2006b); Dwork (2006); Kasiviswanathan *et al.* (2011); Dwork and Roth (2014)), defined as follows.

Definition 5. The level of (local) differential privacy, or simply the privacy level, of a strategy σ_i, denoted by $\zeta(\sigma_i)$, is defined to be

$$\zeta(\sigma_i) = \max\left\{\ln\left(\frac{\mathbb{P}_{\sigma_i}(X_i \in \mathcal{E} \mid S_i = s_i)}{\mathbb{P}_{\sigma_i}(X_i \in \mathcal{E} \mid S_i = 1 - s_i)}\right) : \mathcal{E} \subseteq \{0, 1, \perp\}, s_i \in \{0, 1\}\right\},$$

where we follow the convention that $0/0 = 1$, and the strategy σ_i is said to be $\zeta(\sigma_i)$-differentially private.

The level of differential privacy quantifies the indistinguishablity between the conditional distributions of the reported data given different values of the signal, therefore measuring how disclosive the strategy is. The privacy loss causes a cost to an individual. We assume that when using strategies with the same privacy level, individuals experience the same cost of privacy. Thus, we model each individual's cost of privacy by a function g of the privacy level. We call g the *cost function* and the cost the *privacy cost*. Our results can be extended to the case where the cost functions are heterogeneous (see the discussion in Section 3.3.3). We assume that the form of g is publicly known (Ghosh and Roth Ghosh and Roth (2011) and subsequent work study the scenario that cost functions are private and design truthful mechanisms to elicit them).

We say the cost function g is *proper* if it satisfies the following three conditions:

$$g(\xi) \geq 0, \quad \forall \xi \geq 0, \tag{3.1}$$

$$g(0) = 0, \tag{3.2}$$

$$g \text{ is non-decreasing,} \tag{3.3}$$

where (3.1) follows from the fact that a privacy cost is nonnegative, (3.2) indicates that the privacy cost is 0 when the reported data is independent of the private data, and (3.3) means that the privacy cost will not decrease when the privacy loss becomes larger. In this dissertation, we will focus on a proper cost function that is convex, continuously differentiable, and $g(\xi) = 0$ only for $\xi = 0$. With a little abuse of notation, we also use $g(\sigma_i)$ to denote $g(\zeta(\sigma_i))$, which is the privacy cost to individual i when the strategy σ_i is used.

Nash Equilibrium. The utility of each individual is the difference between her payment and her privacy cost. We assume that the individuals are risk neutral, i.e., they are interested in maximizing their expected utility. We focus on Nash equilibria of a payment mechanism, where each individual has no incentive to unilaterally change her strategy given other individuals' strategies. Formally, a Nash equilibrium in our model is defined as follows.

Definition 6. A strategy profile $\boldsymbol{\sigma}$ is a Nash equilibrium in a payment mechanism $\boldsymbol{R}$ if for any individual i and any strategy σ_i',

$$\mathbb{E}_{\boldsymbol{\sigma}}[R_i(\boldsymbol{X}) - g(\sigma_i)] \geq \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) - g(\sigma_i')],$$

where the expectation is over the reported data $\boldsymbol{X}$, and the subscripts $\boldsymbol{\sigma}$ and $(\sigma_i', \boldsymbol{\sigma}_{-i})$ indicate that $\boldsymbol{X}$ is generated by the strategy profile $\boldsymbol{\sigma}$ and $(\sigma_i', \boldsymbol{\sigma}_{-i})$, respectively.

3.3 The Value of Data Privacy

We say that the data collector *obtains* ϵ units of privacy from an individual i in a payment mechanism if individual i's best response in a Nash equilibrium of the mechanism is to report data with a privacy level of ϵ. Let $\mathcal{R}(i; \epsilon)$ denote the set of nonnegative payment mechanisms in which the data collector obtains ϵ units of privacy from individual i. Then we measure the value of ϵ units of privacy by the minimum payment to individual i of all mechanisms in $\mathcal{R}(i;\epsilon)$. Note that this measure does not depend on the specific identity of i due to the symmetry across individuals. For any mechanism $\boldsymbol{R} \in \mathcal{R}(i;\epsilon)$, let $\boldsymbol{\sigma}^{(\boldsymbol{R};\epsilon)}$ denote the corresponding Nash equilibrium. Then, formally, the value of ϵ units of privacy is measured by

$$V(\epsilon) = \inf_{\boldsymbol{R} \in \mathcal{R}(i;\epsilon)} \mathbb{E}_{\boldsymbol{\sigma}^{(\boldsymbol{R};\epsilon)}}[R_i(\boldsymbol{X})]. \tag{3.4}$$

In this chapter, we first derive a lower bound on $V(\epsilon)$ by characterizing the Nash equilibria and replicating payment mechanisms in $\mathcal{R}(i;\epsilon)$ by genie-aided mechanisms. We then design a payment mechanism in $\mathcal{R}(i;\epsilon)$, and consequently the equilibrium payment to individual i in this mechanism serves as an upper bound of $V(\epsilon)$. The gap between the lower and upper bounds diminishes to zero exponentially fast as the number of individuals N becomes large, which indicates that the lower and upper bounds are asymptotically tight.

3.3.1 Lower Bound

We present a lower bound on $V(\epsilon)$ in Theorem 3 below. For convenience, we define

$$V_{\mathrm{LB}}(\epsilon) = g'(\epsilon)\frac{e^\epsilon + 1}{e^\epsilon}\left(\frac{\theta}{2\theta - 1}(e^\epsilon + 1) - 1\right), \tag{3.5}$$

where g' is the derivative of the privacy cost function of an individual and θ is the quality of signals.

Theorem 3. *The value of ϵ units of privacy measured in* (3.4) *is lower bounded as $V(\epsilon) \geq V_{\mathrm{LB}}(\epsilon)$. Specifically, for any nonnegative payment mechanism $\boldsymbol{R}$, if the strategy of an individual i in a Nash equilibrium has a privacy level of ϵ, then the expected payment to individual i at this equilibrium is lower bounded by $V_{\mathrm{LB}}(\epsilon)$.*

We remark that the lower bound in Theorem 3 can be achieved by a hypothetical payment mechanism in which a genie who knows the realization of the underlying state W guides the data collector on how much to pay each individual. Intuitively, the knowledge of the state W provides more information about the system, which helps the data collector to obtain privacy with less payment. While it may sound like a chicken-and-egg problem as the data collector's sole purpose of paying individuals for their private data is to learn the state W, it will become clear that the philosophy applies and the data collector should utilize the best estimate of W in the payment mechanism to minimize the payment. The insight we gain from this mechanism sheds light on the asymptotically tight upper bound on the value of privacy in Section 3.3.2.

This genie-aided payment mechanism, denoted by $\widehat{\boldsymbol{R}}^{(\epsilon)}$, determines the payment to each individual i based on her own reported data X_i and the state W as follows:

$$\widehat{R}_i^{(\epsilon)}(X_i, W) = \frac{g'(\epsilon)(e^\epsilon + 1)^2}{2e^\epsilon} \widehat{A}_{X_i,W}, \tag{3.6}$$

where

$$\widehat{A}_{1,1} = \frac{1}{(2\theta - 1)P_W(1)}, \quad \widehat{A}_{0,0} = \frac{1}{(2\theta - 1)P_W(0)},$$
$$\widehat{A}_{0,1} = \widehat{A}_{1,0} = 0.$$

In this payment mechanism, it can be proved that the following symmetric randomized response of individual i, which is ϵ-differentially private and is denoted by $\sigma_i^{(\epsilon)}$, is the best response:

$$\begin{aligned}
\mathbb{P}_{\sigma_i^{(\epsilon)}}(X_i = 1 \mid S_i = 1) = \mathbb{P}_{\sigma_i^{(\epsilon)}}(X_i = 0 \mid S_i = 0) &= \frac{e^\epsilon}{e^\epsilon + 1}, \\
\mathbb{P}_{\sigma_i^{(\epsilon)}}(X_i = 1 \mid S_i = 0) = \mathbb{P}_{\sigma_i^{(\epsilon)}}(X_i = 0 \mid S_i = 1) &= \frac{1}{e^\epsilon + 1}, \\
\mathbb{P}_{\sigma_i^{(\epsilon)}}(X_i = \perp \mid S_i = 1) = \mathbb{P}_{\sigma_i^{(\epsilon)}}(X_i = \perp \mid S_i = 0) &= 0.
\end{aligned} \tag{3.7}$$

For convenience, we will refer to this strategy as the ϵ-*strategy*. The expected payment to individual i at this strategy equals to the lower bound in Theorem 3.

Next we sketch the proof of Theorem 3. We first give three lemmas that form the basis of the proof, and then present the proof based on that. The proofs of the lemmas are presented in Appendix B–D.

Characterization of Nash Equilibria

We first characterize individuals' behavior in a Nash equilibrium. In general, an ϵ-differentially private strategy has uncountably many possible forms. However, provided that the strategy is part of a Nash equilibrium (i.e., a best response of an individual), the following lemma substantially reduces the space of possibilities. We remark that a similar phenomenon for privacy-aware individuals has been observed in Chen *et al.* (2014) in a different setting.

Lemma 1. *In any nonnegative payment mechanism, the strategy of an individual in a Nash equilibrium is either a symmetric randomized response, or a non-informative strategy.*

We remark that Lemma 1 holds for more general probability models of the signals. The proof carries over as long as the support of the joint distribution of the signals is the entire domain $\{0,1\}^N$.

By Lemma 1, if an individual's strategy in a Nash equilibrium has a privacy level of ϵ, where $\epsilon > 0$, this equilibrium strategy is either the ϵ-strategy or the $(-\epsilon)$-strategy. The following lemma says that from the payment perspective, it suffices to further focus on the case that it is the ϵ-strategy.

Lemma 2. *For any nonnegative payment mechanism $\boldsymbol{R}$ in which the strategy profile $(\sigma_i^{(-\epsilon)}, \boldsymbol{\sigma}_{-i})$ with some $\epsilon > 0$ is a Nash equilibrium, there exists another nonnegative payment mechanism $\boldsymbol{R}'$ in which $(\sigma_i^{(\epsilon)}, \boldsymbol{\sigma}_{-i})$ is a Nash equilibrium, and the expected payment to each individual at these two equilibria of the two mechanisms are the same.*

This lemma is proved by considering the payment mechanism $\boldsymbol{R}'$ that is constructed by applying $\boldsymbol{R}$ on the reported data after modifying X_i to $1 - X_i$.

Genie-Aided Payment Mechanism

A genie-aided payment mechanism $\widehat{\boldsymbol{R}}$: $\mathcal{X}^N \times \{0,1\} \to \mathbb{R}^N$ determines the payment to an individual based on not only the reported data $\boldsymbol{X}$ but also the underlying state W. Compared with a standard payment mechanism, a genie-aided mechanism is hypothetical since the data collector has access to the underlying state, as if she were aided by a genie. Unless otherwise stated, we consider those nonnegative genie-aided payment mechanisms where $\widehat{R}_i(\boldsymbol{X}, W)$, the payment to individual i, depends on only her own reported data X_i and the underlying state W. Therefore, we will write $\widehat{R}_i(X_i, W)$ to represent $\widehat{R}_i(\boldsymbol{X}, W)$ for conciseness. The following lemma shows that the expected payments resulting from any Nash equilibrium of any payment mechanism can be replicated by a genie-aided payment mechanism with the same Nash equilibrium. Thus we can restrict our attention to genie-aided mechanisms to obtain a lower bound on the value of privacy.

Lemma 3. *For any nonnegative payment mechanism $\boldsymbol{R}$ and any Nash equilibrium $\boldsymbol{\sigma}$ of it, there exists a nonnegative genie-aided mechanism $\widehat{\boldsymbol{R}}$, such that $\boldsymbol{\sigma}$ is also a Nash equilibrium of $\widehat{\boldsymbol{R}}$ and the expected payment to each individual at this equilibrium is the same under $\boldsymbol{R}$ and $\widehat{\boldsymbol{R}}$.*

This lemma is proved by constructing the following genie-aided payment mechanism $\widehat{\boldsymbol{R}}$ according to the desired equilibrium $\boldsymbol{\sigma}$: for any individual i and any $x_i \in \mathcal{X}, w \in \{0,1\}$,

$$\widehat{R}_i(x_i, w) = \overline{R}_i(x_i; w) := \mathbb{E}_{\boldsymbol{\sigma}}[R_i(\boldsymbol{X}) \mid X_i = x_i, W = w].$$

Our intuition is as follows. A genie-aided mechanism can use the state W to generate an incentive to individual i instead of using the reported data $\boldsymbol{X}_{-i}$ of others. The above genie-aided payment mechanism $\widehat{\boldsymbol{R}}$ is constructed such that no matter what strategy individual i uses, her expected utility is the same under $\boldsymbol{R}$ and $\widehat{\boldsymbol{R}}$. Since an individual calculates her best response according to the expected utility, her equilibrium behavior and expected payment are the same under $\widehat{\boldsymbol{R}}$ and $\boldsymbol{R}$. We remark that the Nash equilibria of a genie-aided mechanism are much easier to analyze since the individuals are decoupled in the payments and thus an individual's strategy does not have an influence on other individuals' utility.

Let $\widehat{\mathcal{R}}(i;\epsilon)$ denote the set of nonnegative genie-aided payment mechanisms in which the ϵ-strategy is an individual i's strategy in a Nash equilibrium, and let $\sigma_i^{(\epsilon)}$ denote the ϵ-strategy. Consider

$$\widehat{V}(\epsilon) = \inf_{\widehat{\boldsymbol{R}} \in \widehat{\mathcal{R}}(i;\epsilon)} \mathbb{E}_{\sigma_i^{(\epsilon)}}\Big[\widehat{R}_i(X_i, W)\Big],$$

which is a definition similar to the value of ϵ units of privacy, $V(\epsilon)$, measured in (3.4). Then $\widehat{V}(\epsilon) \le V(\epsilon)$ for the following reasons. Consider any $\boldsymbol{R} \in \mathcal{R}(i;\epsilon)$, i.e., any nonnegative payment mechanism $\boldsymbol{R}$ in which individual i's strategy in a Nash equilibrium has a privacy level of ϵ. With Lemma 1 and 2, we can assume without loss of generality that this equilibrium strategy is the ϵ-strategy. Then by Lemma 3, we can map $\boldsymbol{R}$ to a $\widehat{\boldsymbol{R}} \in \widehat{\mathcal{R}}(i;\epsilon)$, such that

$$\mathbb{E}_{\boldsymbol{\sigma}^{(\boldsymbol{R};\epsilon)}}[R_i(\boldsymbol{X})] = \mathbb{E}_{\sigma_i^{(\epsilon)}}\Big[\widehat{R}_i(X_i, W)\Big].$$

Therefore, the infimum over $\widehat{\mathcal{R}}(i;\epsilon)$ is no greater than the infimum over $\mathcal{R}(i;\epsilon)$, i.e., $\widehat{V}(\epsilon) \le V(\epsilon)$.

Proof of Theorem 3

With Lemma 1, 2 and 3, we can prove the lower bound in Theorem 3 by focusing on the genie-aided mechanisms in $\widehat{\mathcal{R}}(i;\epsilon)$. Then there is no need to consider the strategies of individuals other than individual i since a genie-aided mechanism pays individual i only according to X_i and W. A necessary condition for the ϵ-strategy to be a best response of individual i is that ϵ yields no worse expected payment than other privacy levels. We utilize this necessary condition to obtain a lower bound on the expected payment to individual i, which gives a lower bound on $\widehat{V}(\epsilon)$ and further proves the lower bound in Theorem 3.

Proof of Theorem 3. By Lemma 1, 2 and 3, it suffices to focus on nonnegative genie-aided payment mechanisms in which the ϵ-strategy is an individual i's strategy in a Nash equilibrium, i.e., mechanisms in $\widehat{\mathcal{R}}(i;\epsilon)$. Consider any $\widehat{\boldsymbol{R}} \in \widehat{\mathcal{R}}(i;\epsilon)$ and denote the ϵ-strategy by $\sigma_i^{(\epsilon)}$. Consider the ξ-strategy of individual i with any $\xi \geq 0$ and denote it by $\sigma_i^{(\xi)}$. Then the expected utility of individual i at the strategy $\sigma_i^{(\xi)}$ can be written as

$$\begin{aligned}
&\mathbb{E}_{\sigma_i^{(\xi)}}\left[\widehat{R}_i(X_i, W)\right] - g(\sigma_i^{(\xi)}) \\
={}& \sum_{x_i, s_i, w} \mathbb{P}_{\sigma_i^{(\xi)}}(X_i = x_i \mid S_i = s_i)\mathbb{P}(S_i = s_i, W = w)\widehat{R}_i(x_i, w) \\
&- g(\xi), \\
={}& \overline{K}_1 \frac{e^{\xi}}{e^{\xi}+1} + \overline{K}_0 \frac{1}{e^{\xi}+1} + \overline{K} - g(\xi),
\end{aligned}$$

where

$$\begin{aligned}
\overline{K}_1 &= \{\widehat{R}_i(1,1)P_W(1)\theta + \widehat{R}_i(1,0)P_W(0)(1-\theta)\} \\
&\qquad - \{\widehat{R}_i(0,1)P_W(1)\theta + \widehat{R}_i(0,0)P_W(0)(1-\theta)\}, \\
\overline{K}_0 &= \{\widehat{R}_i(1,1)P_W(1)(1-\theta) + \widehat{R}_i(1,0)P_W(0)\theta\}
\end{aligned}$$

$$- \{\widehat{R}_i(0,1)P_W(1)(1-\theta) + \widehat{R}_i(0,0)P_W(0)\theta\},$$

$$\overline{K} = \widehat{R}_i(0,1)P_W(1) + \widehat{R}_i(0,0)P_W(0).$$

It can be seen that $\overline{K}_1$, $\overline{K}_0$ and $\overline{K}$ do not depend on ξ. Let this expected utility define a function f of ξ; i.e.,

$$f(\xi) = \overline{K}_1 \frac{e^\xi}{e^\xi + 1} + \overline{K}_0 \frac{1}{e^\xi + 1} - g(\xi) + \overline{K}.$$

Then since the ϵ-strategy is individual i's strategy in a Nash equilibrium, the level ϵ maximizes $f(\xi)$. Since

$$f'(\xi) = (\overline{K}_1 - \overline{K}_0)\frac{e^\xi}{(e^\xi + 1)^2} - g'(\xi),$$

and $f'(\epsilon) = 0$, we must have

$$\overline{K}_1 - \overline{K}_0 = g'(\epsilon)\frac{(e^\epsilon + 1)^2}{e^\epsilon}. \tag{3.8}$$

Now we calculate the expected payment to individual i at the ϵ-strategy:

$$\mathbb{E}_{\sigma_i^{(\epsilon)}}\left[\widehat{R}_i(X_i, W)\right] = -(\overline{K}_1 - \overline{K}_0)\frac{1}{e^\epsilon + 1} + (\overline{K}_1 + \overline{K}).$$

By definition,

$$\begin{aligned}\overline{K}_1 + \overline{K} = {} & \widehat{R}_i(1,1)P_W(1)\theta + \widehat{R}_i(1,0)P_W(0)(1-\theta)\\ & + \widehat{R}_i(0,1)P_W(1)(1-\theta) + \widehat{R}_i(0,0)P_W(0)\theta,\end{aligned}$$

and

$$\begin{aligned}\overline{K}_1 - \overline{K}_0 = {} & \big(\widehat{R}_i(1,1) - \widehat{R}_i(0,1)\big)P_W(1)(2\theta - 1)\\ & + \big(\widehat{R}_i(0,0) - \widehat{R}_i(1,0)\big)P_W(0)(2\theta - 1).\end{aligned}$$

Therefore,

$$\overline{K}_1 + \overline{K} = \frac{\theta}{2\theta - 1}(\overline{K}_1 - \overline{K}_0)$$

$$\begin{aligned}&+ \widehat{R}_i(1,0)P_W(0) + \widehat{R}_i(0,1)P_W(1)\\ &\geq \frac{\theta}{2\theta-1}(\overline{K}_1 - \overline{K}_0)\\ &= g'(\epsilon)\frac{(e^\epsilon+1)^2}{e^\epsilon}\frac{\theta}{2\theta-1},\end{aligned}$$

where we have used the nonnegativity of $\widehat{\boldsymbol{R}}$. Then the expected payment to individual i is bounded as follows:

$$\begin{aligned}&\mathbb{E}_{\sigma_i^{(\epsilon)}}\left[\widehat{R}_i(X_i,W)\right]\\ &= -(\overline{K}_1 - \overline{K}_0)\frac{1}{e^\epsilon+1} + (\overline{K}_1 + \overline{K})\\ &\geq g'(\epsilon)\frac{e^\epsilon+1}{e^\epsilon}\left(\frac{\theta}{2\theta-1}(e^\epsilon+1) - 1\right), \qquad (3.9)\end{aligned}$$

which proves the lower bound. □

Now beyond the proof, we take a moment to check when this lower bound can be achieved. To achieve the lower bound, we need the equality in (3.9) to hold and the equation (3.8) to be satisfied, which is equivalent to the following conditions:

$$\widehat{R}_i(1,0) = 0, \qquad (3.10)$$

$$\widehat{R}_i(0,1) = 0, \qquad (3.11)$$

$$\begin{aligned}&(2\theta-1)\Big(\widehat{R}_i(1,1)P_W(1) + \widehat{R}_i(0,0)P_W(0)\Big)\\ &\quad = g'(\epsilon)\frac{(e^\epsilon+1)^2}{e^\epsilon}. \qquad (3.12)\end{aligned}$$

It is easy to check that the genie-aided payment mechanism $\widehat{\boldsymbol{R}}^{(\epsilon)}$ defined in (3.6) is in $\widehat{\mathcal{R}}(i;\epsilon)$ and satisfies (3.10)–(3.12), and therefore achieves the lower bound. Can this lower bound can be achieved by a standard nonnegative payment mechanism? Consider any payment mechanism $\boldsymbol{R} \in \mathcal{R}(i;\epsilon)$. Following similar arguments, we can prove that to achieve the lower bound, $\boldsymbol{R}$ needs to satisfy the following conditions:

$$\overline{R}_i(1;0) = 0, \tag{3.13}$$

$$\overline{R}_i(0;1) = 0, \tag{3.14}$$

$$\begin{aligned}(2\theta - 1)\big(\overline{R}_i(1;1)P_W(1) + \overline{R}_i(0;0)P_W(0)\big)& \\ = g'(\epsilon)\frac{(e^\epsilon + 1)^2}{e^\epsilon},&\end{aligned} \tag{3.15}$$

where recall that $\overline{R}_i(x_i;w) = \mathbb{E}_{\boldsymbol{\sigma}^{(\boldsymbol{R},\epsilon)}}[R_i(\boldsymbol{X}) \mid X_i = x_i, W = w]$ for $x_i, w \in \{0,1\}$. It can be proved that if $\boldsymbol{R}$ satisfies (3.13) and (3.14), then $R_i(\boldsymbol{x}) = 0$ for any $\boldsymbol{x} \in \mathcal{X}^N$, which contradicts (3.15). Therefore, no standard nonnegative payment mechanism can achieve the lower bound. However, as will be shown in the next section, we can design a class of standard nonnegative payment mechanisms such that the expected payment approaches the lower bound as the number of individuals increases. The design follows the insights indicated by the genie-aided mechanism $\widehat{\boldsymbol{R}}^{(\epsilon)}$: to minimize the payment, the data collector should utilize the best estimate of W in the payment mechanism based on the noisy reported data.

3.3.2 Upper Bound

We present an upper bound on $V(\epsilon)$ in Theorem 4 below. For convenience, define

$$d = \frac{1}{2}\ln\frac{(e^\epsilon + 1)^2}{4(\theta e^\epsilon + 1 - \theta)((1-\theta)e^\epsilon + \theta)}, \tag{3.16}$$

where θ is the quality of signal. Note that $d > 0$ for any $\epsilon > 0$. Recall that $V_{\mathrm{LB}}(\epsilon)$ is the lower bound in Theorem 3.

Theorem 4. *The value of ϵ units of privacy measured in* (3.4) *is upper bounded as $V(\epsilon) \le V_{\mathrm{LB}}(\epsilon) + O(e^{-Nd})$, where the $O(\cdot)$ is for $N \to \infty$. Specifically, there exists a nonnegative payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$ in which the strategy profile $\boldsymbol{\sigma}^{(\epsilon)}$ consisting* *of ϵ-strategies is a Nash equilibrium, and the expected payment to each individual i at this equilibrium is upper bounded by $V_{\mathrm{LB}}(\epsilon) + O(e^{-Nd})$.*

Comparing this upper bound with the lower bound $V_{\mathrm{LB}}(\epsilon)$ in Theorem 3 we can see that the gap between the lower and upper bounds is just the term $O(e^{-Nd})$, which diminishes to zero exponentially fast as N goes to infinity.

We present the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$ in the following section. We will show that under $\boldsymbol{R}^{(N,\epsilon)}$, the strategy profile $\boldsymbol{\sigma}^{(\epsilon)}$ consisting of ϵ-strategies is a Nash equilibrium. Therefore, $\boldsymbol{R}^{(N,\epsilon)}$ is a member of $\mathcal{R}(i;\epsilon)$, and the payment to individual i at $\boldsymbol{\sigma}^{(\epsilon)}$ gives an upper bound on the value of privacy.

The design of $\boldsymbol{R}^{(N,\epsilon)}$ is enlightened by the hypothetical payment mechanism $\widehat{\boldsymbol{R}}^{(\epsilon)}$ defined in (3.6). But without direct access to the state W, the mechanism $\boldsymbol{R}^{(N,\epsilon)}$ relies on the reported data from an individual i's peers, i.e., individuals other than individual i, to obtain an estimate of W. We borrow the idea of the peer-prediction method Miller *et al.* (2009), which rewards more for the agreement between an individual and her peers to encourage truthful reporting. However, unlike the peer-prediction method, the individuals here have privacy concerns and they will weigh the privacy cost against the payment to choose the best privacy level. We modify the payments in $\widehat{\boldsymbol{R}}^{(\epsilon)}$ to ensure that the ϵ-strategy is still a best response of each individual in $\boldsymbol{R}^{(N,\epsilon)}$, given that other individuals also follow the ϵ-strategy, which yields the desired Nash equilibrium $\boldsymbol{\sigma}^{(\epsilon)}$.

The equilibrium payment to each individual in $\boldsymbol{R}^{(N,\epsilon)}$ converges to the lower bound in Theorem 3 as the number of individuals N goes to infinity. The intuition behind is that as the number of individuals N goes to infinity, the majority of the reported data from other individuals converges to the underlying state W, and thus $\boldsymbol{R}^{(N,\epsilon)}$ works similar as the genie-aided mechanism $\widehat{\boldsymbol{R}}^{(\epsilon)}$, whose equilibrium payment to each individual equals to the lower bound in Theorem 3.

A Payment Mechanism $\boldsymbol{R}^{(N,\epsilon)}$

The payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$ is designed for purchasing private data from N privacy-aware individuals, parameterized by a privacy parameter ϵ, where $N \geq 2$ and $\epsilon > 0$.

1. Each individual reports her data (which can be the decision of not participating).

2. The data collector counts the number of participants n.

3. For non-participating individuals, the payment is zero.

4. If there is only one participant, pay zero to this participant. Otherwise, for each participating individual i, the data collector computes the variable
$$M_{-i} = \begin{cases} 1 & \text{if} \displaystyle\sum_{j:\, X_j \neq \perp, j \neq i} X_j \geq \left\lfloor \frac{n-1}{2} \right\rfloor + 1, \\ 0 & \text{otherwise,} \end{cases}$$
which is the majority of the other participants' reported data. Then the data collector pays individual i the following amount of payment according to her reported data X_i and M_{-i}:
$$R_i^{(N,\epsilon)}(\boldsymbol{X}) = \frac{g'(\epsilon)(e^\epsilon + 1)^2}{2e^\epsilon} A_{X_i, M_{-i}},$$
where the parameters $A_{1,1}, A_{0,0}, A_{0,1}, A_{1,0}$ are defined in Section 3.3.2.

Payment Parameterization

Let
$$\alpha = \theta \frac{e^\epsilon}{e^\epsilon + 1} + (1 - \theta) \frac{1}{e^\epsilon + 1}.$$
The physical meaning of α can be seen by considering the strategy profile $\boldsymbol{\sigma}^{(\epsilon)}$, where given the state W, the reported data $X_1, X_2, \ldots, X_N$ are i.i.d. with
$$\mathbb{P}_{\sigma_i^{(\epsilon)}}(X_i = 1 \mid W = 1) = \mathbb{P}_{\sigma_i^{(\epsilon)}}(X_i = 0 \mid W = 0) = \alpha.$$

Given that the number of participants is n with $n \geq 2$, define the following quantities. Consider a random variable that follows the binomial distribution with parameters $n-1$ and α. Let $\beta^{(n)}$ denote the probability that this random variable is greater than or equal to $\lfloor \frac{n-1}{2} \rfloor + 1$. Let

$$\gamma^{(n)} = \begin{cases} 1 - \binom{n-1}{\frac{n-1}{2}} \alpha^{\frac{n-1}{2}} (1-\alpha)^{\frac{n-1}{2}} & \text{if } n-1 \text{ is even,} \\ 1 & \text{if } n-1 \text{ is odd.} \end{cases} \tag{3.17}$$

To see the physical meaning of $\beta^{(n)}$ and $\gamma^{(n)}$, still consider $\boldsymbol{\sigma}^{(\epsilon)}$, where the number of participants is $n = N$. Then for an individual i,

$$\begin{aligned} &\mathbb{P}_{\boldsymbol{\sigma}^{(\epsilon)}}(M_{-i} = 1 \mid W = 1) = \beta^{(N)}, \\ &\mathbb{P}_{\boldsymbol{\sigma}^{(\epsilon)}}(M_{-i} = 1 \mid W = 0) = \gamma^{(N)} - \beta^{(N)}. \end{aligned}$$

With the introduced notation, the parameters $A_{1,1}$, $A_{0,0}$, $A_{0,1}$, $A_{1,0}$ used in the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$ are defined as follows:

$$\begin{aligned} A_{1,1} &= \frac{P_W(1)(1-\beta^{(n)}) + P_W(0)(1-(\gamma^{(n)} - \beta^{(n)}))}{(2\beta^{(n)} - \gamma^{(n)})(2\theta - 1)P_W(1)P_W(0)}, \\ A_{0,0} &= \frac{P_W(1)\beta^{(n)} + P_W(0)(\gamma^{(n)} - \beta^{(n)})}{(2\beta^{(n)} - \gamma^{(n)})(2\theta - 1)P_W(1)P_W(0)}, \\ A_{0,1} &= 0, \\ A_{1,0} &= 0. \end{aligned}$$

It is easy to verify that these parameters are nonnegative. Thus $\boldsymbol{R}^{(N,\epsilon)}$ is a nonnegative payment mechanism. The proof of the equilibrium properties of $\boldsymbol{R}^{(N,\epsilon)}$ in Theorem 4 is given below.

Proof of Theorem 4

Proof. It suffices to prove that the strategy profile $\boldsymbol{\sigma}^{(\epsilon)}$ is a Nash equilibrium in $\boldsymbol{R}^{(N,\epsilon)}$ and the expected payment to each individual i at this equilibrium satisfies that

$$\mathbb{E}_{\boldsymbol{\sigma}^{(\epsilon)}}\left[R_i^{(N,\epsilon)}(\boldsymbol{X})\right] \leq V_{\mathrm{LB}}(\epsilon) + O(e^{-Nd}),$$

where recall that

$$V_{\mathrm{LB}}(\epsilon) = \frac{g'(\epsilon)(e^\epsilon + 1)}{e^\epsilon}\left(\frac{\theta}{2\theta - 1}(e^\epsilon + 1) - 1\right).$$

For conciseness, we suppress the explicit dependence on N and ϵ, and write $\boldsymbol{R}$ and $\boldsymbol{\sigma}$ to represent $\boldsymbol{R}^{(N,\epsilon)}$ and $\boldsymbol{\sigma}^{(\epsilon)}$, respectively, in the remainder of this proof.

We first prove that the strategy profile $\boldsymbol{\sigma}$ is a Nash equilibrium in $\boldsymbol{R}$; i.e., for any individual i, the ϵ-strategy is a best response of individual i when other individuals follow $\boldsymbol{\sigma}_{-i}$. Following the notation in the proof of Lemma 1, for any individual i we consider any strategy σ_i' of individual i and let

$$p_1 = \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 1), \quad q_1 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 1),$$
$$p_0 = \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 0), \quad q_0 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 0).$$

Then by the proof of Lemma 1, the best response satisfies either $p_1 = p_0, q_1 = q_0$, or $p_1 = q_0, p_0 = q_1, p_1 + q_1 = 1$, depending on the form of the utility function $U_i(p_1, p_0, q_1, q_0)$, which is the expected utility of individual i at the strategy σ_i' when other individuals follow $\boldsymbol{\sigma}_{-i}$. Thus, we derive the form of $U_i(p_1, p_0, q_1, q_0)$ next. Recall that we let $\overline{R}_i(x_i; w)$ denote $\mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, W = w]$ for $x_i, w \in \{0, 1\}$. Then

$$\begin{aligned}
&U_i(p_1, p_0, q_1, q_0) \\
&= \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) - g(\zeta(\sigma_i'))] \\
&= K_1 p_1 + K_0 p_0 + L_1 q_1 + L_0 q_0 - g(\zeta(p_1, p_0, q_1, q_0)),
\end{aligned}$$

with

$$\begin{aligned}
K_1 &= \{\overline{R}_i(1;1)P_W(1)\theta + \overline{R}_i(1;0)P_W(0)(1-\theta)\},\\
K_0 &= \{\overline{R}_i(1;1)P_W(1)(1-\theta) + \overline{R}_i(1;0)P_W(0)\theta\},\\
L_1 &= \{\overline{R}_i(0;1)P_W(1)\theta + \overline{R}_i(0;0)P_W(0)(1-\theta)\},\\
L_0 &= \{\overline{R}_i(0;1)P_W(1)(1-\theta) + \overline{R}_i(0;0)P_W(0)\theta\}.
\end{aligned}$$

In the designed mechanism $\boldsymbol{R}$, the payment to individual i only depends on X_i and M_{-i}. Thus we can write

$$R_i(X_i; M_{-i}) = R_i(\boldsymbol{X}).$$

Then the value of $\overline{R}_i(x_i; w)$ is calculated as follows:

$$\begin{aligned}
\overline{R}_i(1;1) &= \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = 1, W = 1]\\
&= \beta^{(N)} R_i(1;1) + (1-\beta^{(N)})R_i(1;0),\\
\overline{R}_i(1;0) &= \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = 1, W = 0]\\
&= (\gamma^{(N)} - \beta^{(N)}) R_i(1;1) + (1-(\gamma^{(N)} - \beta^{(N)}))R_i(1;0),\\
\overline{R}_i(0;1) &= \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = 0, W = 1]\\
&= (1-\beta^{(N)}) R_i(0;0) + \beta^{(N)} R_i(0;1),\\
\overline{R}_i(0;0) &= \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = 0, W = 0]\\
&= (1-(\gamma^{(N)} - \beta^{(N)})) R_i(0;0) + (\gamma^{(N)} - \beta^{(N)}) R_i(0;1),
\end{aligned}$$

and it can be verified that K_1, K_0, L_1 and L_0 are all positive. Therefore, by the proof of Lemma 1, the possibility for the best response to be $p_1 = p_0, q_1 = q_0, 0 < p_1 + q_1 < 1$ can be eliminated and the best response strategy must be in one of the following three forms:

$$p_1 = p_0 = q_1 = q_0 = 0, \tag{3.18}$$

$$p_1 = p_0, \quad q_1 = q_0, \quad p_1 + q_1 = 1, \tag{3.19}$$

$$p_1 = q_0, \quad p_0 = q_1, \quad p_1 + q_1 = 1. \tag{3.20}$$

The strategy in (3.18) is to always not participate, which yields an utility of zero. For strategies in the form of (3.19) or (3.20), we can write the expected utility as a function of p_1 and p_0 as follows:

$$\overline{U}_i(p_1, p_0) = \overline{K}_1 p_1 + \overline{K}_0 p_0 + \overline{K} - g(\zeta(p_1, p_0)),$$

where $\overline{K}_1 = K_1 - L_1$, $\overline{K}_0 = K_0 - L_0$, $\overline{K} = L_1 + L_0$, and with a little abuse of notation,

$$\zeta(p_1, p_0) = \max\left\{ \left| \ln \frac{p_1}{p_0} \right|, \left| \ln \frac{1 - p_1}{1 - p_0} \right| \right\}.$$

Inserting the value of $R_i(X_i; M_{-i})$ gives

$$\overline{K}_1 = \frac{g'(\epsilon)(e^\epsilon + 1)^2}{2e^\epsilon}, \quad \overline{K}_0 = -\frac{g'(\epsilon)(e^\epsilon + 1)^2}{2e^\epsilon}.$$

Then a strategy in the form of (3.19) yields an utility of $\overline{K} > 0$. A strategy in the form of (3.20) can be written as

$$p_1 = q_0 = \frac{e^\xi}{e^\xi + 1}, \quad p_0 = q_1 = \frac{1}{e^\xi + 1}.$$

Then the expected utility can be further written as a function f of ξ as follows:

$$f(\xi) = \overline{K}_1 \frac{e^\xi}{e^\xi + 1} + \overline{K}_0 \frac{1}{e^\xi + 1} - g(|\xi|) + \overline{K}.$$

Therefore, to prove that the ϵ-strategy is a best response of individual i, it suffices to prove that ϵ maximizes $f(\xi)$ and $f(\epsilon) \geq \overline{K}$. For any $\xi < 0$, it is easy to see that

$$\overline{K}_1 \frac{e^\xi}{e^\xi + 1} + \overline{K}_0 \frac{1}{e^\xi + 1} < 0 < \overline{K}_1 \frac{e^{-\xi}}{e^{-\xi} + 1} + \overline{K}_0 \frac{1}{e^{-\xi} + 1}.$$

Thus $f(\xi)$ achieves its maximum value at some $\xi \geq 0$. For any $\xi \geq 0$,

$$f'(\xi) = (\overline{K}_1 - \overline{K}_0) \frac{e^\xi}{(e^\xi + 1)^2} - g'(\xi),$$

$$f''(\xi) = -(\overline{K}_1 - \overline{K}_0)\frac{e^{\xi}(e^{\xi}-1)}{(e^{\xi}+1)^3} - g''(\xi) \leq 0,$$

where the second inequality is due to the convexity of the cost function g. Therefore, h is concave. Since $f'(\epsilon) = 0$, ϵ maximizes $f(\xi)$. The optimal value is

$$f(\epsilon) = g'(\epsilon)\frac{e^{\epsilon} - e^{-\epsilon}}{2} - g(\epsilon) + \overline{K}.$$

By the convexity of g,

$$g(\epsilon) \leq g'(\epsilon)\epsilon \leq g'(\epsilon)\frac{e^{\epsilon} - e^{-\epsilon}}{2}.$$

Thus $f(\epsilon) \geq \overline{K}$, which completes the proof for the ϵ-strategy to be a best response of individual i.

Next we calculate the expected payment to individual i at $\boldsymbol{\sigma}$, which can be written as

$$\mathbb{E}_{\boldsymbol{\sigma}}[R_i(\boldsymbol{X})] = -(\overline{K}_1 - \overline{K}_0)\frac{1}{e^{\epsilon}+1} + \overline{K}_1 + \overline{K}.$$

By definitions,

$$\begin{aligned}
&\overline{K}_1 + \overline{K} \\
=& \frac{g'(\epsilon)(e^{\epsilon}+1)^2}{2e^{\epsilon}} \frac{1}{(2\beta^{(N)} - \gamma^{(N)})(2\theta - 1)} \\
&\cdot \Big(2\big(\beta^{(N)}\big)^2 + (4\theta - 2 - 2\gamma^{(N)})\beta^{(N)} \\
&\qquad + 2(1-\theta)\gamma^{(N)} + \beta^{(N)}(1-\beta^{(N)})\frac{P_W(1)}{P_W(0)} \\
&\qquad + (\gamma^{(N)} - \beta^{(N)})(1 - (\gamma^{(N)} - \beta^{(N)}))\frac{P_W(0)}{P_W(1)} \Big) \\
=:& \frac{g'(\epsilon)(e^{\epsilon}+1)^2}{2e^{\epsilon}} h(\beta^{(N)}).
\end{aligned}$$

Then

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\sigma}}[R_i(\boldsymbol{X})] &= \frac{g'(\epsilon)(e^{\epsilon}+1)}{e^{\epsilon}}\left(\frac{1}{2}h(\beta^{(N)})(e^{\epsilon}+1) - 1\right) \\
&= V_{\mathrm{LB}}(\epsilon) + \frac{g'(\epsilon)(e^{\epsilon}+1)^2}{2e^{\epsilon}}\left(h(\beta^{(N)}) - \frac{2\theta}{2\theta - 1}\right).
\end{aligned}$$

To derive an upper bound on the expected payment, we first analyze the function h. Rearranging terms gives

$$\begin{aligned}h(\beta^{(N)}) = \frac{1}{2\theta-1}\frac{1}{2\beta^{(N)}-\gamma^{(N)}} \\ \cdot \Bigg((2-t)\big(\beta^{(N)}\big)^2 + \Bigg(4\theta - 2 - 2\gamma^{(N)} + \frac{P_W(1)}{P_W(0)} \\ + (2\gamma^{(N)}-1)\frac{P_W(0)}{P_W(1)}\Bigg)\beta^{(N)} \\ + 2(1-\theta)\gamma^{(N)} + \gamma^{(N)}(1-\gamma^{(N)})\frac{P_W(0)}{P_W(1)}\Bigg),\end{aligned}$$

where

$$t = \frac{(P_W(1))^2 + (P_W(0))^2}{P_W(1)P_W(0)} \geq 2.$$

Taking derivative yields

$$\begin{aligned}h'(\beta^{(N)}) = \frac{1}{2\theta-1}\frac{1}{(2\beta^{(N)}-\gamma^{(N)})^2} \\ \cdot \Bigg(2(2-t)\Big(\beta^{(N)} - \frac{\gamma^{(N)}}{2}\Big)^2 - \big(\gamma^{(N)}\big)^2 \\ - \frac{\gamma^{(N)}t}{2}(2-\gamma^{(N)}) - 2\gamma^{(N)}(1-\gamma^{(N)})\Bigg).\end{aligned}$$

Therefore, $h'(\beta^{(N)}) \leq 0$ and h is a non-increasing function.

Next we derive a lower bound on $\beta^{(N)}$. Let $Y_1, Y_2, \ldots, Y_{N-1}$ be i.i.d. Bernoulli random variables with parameter α. Then by the definition of $\beta^{(N)}$:

$$\begin{aligned}\beta^{(N)} &= \mathbb{P}\left(\sum_{l=1}^{N-1} Y_l \geq \left\lfloor \frac{N-1}{2} \right\rfloor + 1\right) \\ &= \gamma^{(N)} - \mathbb{P}\left(\sum_{l=1}^{N-1}(1-Y_l) \geq N - 1 - \left\lceil \frac{N-1}{2} \right\rceil + 1\right) \\ &\geq \gamma^{(N)} - \mathbb{P}\left(\sum_{l=1}^{N-1}(1-Y_l) \geq \frac{N-1}{2}\right).\end{aligned}$$

By the Chernoff bound Srikant and Ying (2014),

$$\mathbb{P}\left(\sum_{l=1}^{N-1}(1-Y_l) \geq \frac{N-1}{2}\right) \leq e^{-(N-1)\frac{1}{2}\ln\frac{1}{4\alpha(1-\alpha)}} = e^{-(N-1)d},$$

where $d = \frac{1}{2}\ln\frac{1}{4\alpha(1-\alpha)} > 0$ is the parameter defined in (3.16). Thus,

$$\beta^{(N)} \geq \gamma^{(N)} - e^{-(N-1)d}.$$

By the monotonicity of h,

$$\begin{aligned}
& h(\beta^{(N)}) - \frac{2\theta}{2\theta-1} \\
\leq\, & h\Big(\gamma^{(N)} - e^{-(N-1)d}\Big) - \frac{2\theta}{2\theta-1} \\
=\, & \frac{1}{2\theta-1}\frac{1}{\gamma^{(N)} - 2e^{-(N-1)d}} \\
& \cdot\Bigg((2-t)e^{-2(N-1)d} + \bigg(2(1-\gamma^{(N)}) + 2\gamma^{(N)}t \\
& \qquad - \frac{P_W(1)}{P_W(0)} - (2\gamma^{(N)}-1)\frac{P_W(0)}{P_W(1)}\bigg)e^{-(N-1)d} \\
& \qquad + \gamma^{(N)}\frac{P_W(1)}{P_W(0)} + \big(\gamma^{(N)}\big)^2\frac{P_W(0)}{P_W(1)} - \big(\gamma^{(N)}\big)^2 t\Bigg) \\
\leq\, & \frac{1}{2\theta-1}\frac{1}{\gamma^{(N)} - 2e^{-(N-1)d}} \\
& \cdot\Bigg((2-t)e^{-2(N-1)d} + (2(1-\gamma^{(N)}) + t)e^{-(N-1)d} \\
& \qquad + \gamma^{(N)}(1-\gamma^{(N)})\frac{P_W(1)}{P_W(0)}\Bigg).
\end{aligned}$$

Notice that

$$1 - \gamma^{(N)} = \begin{cases} \binom{N-1}{\frac{N-1}{2}}\alpha^{\frac{N-1}{2}}(1-\alpha)^{\frac{N-1}{2}} & \text{if } N-1 \text{ is even,} \\ 0 & \text{if } N-1 \text{ is odd.} \end{cases}$$

Then when $N-1$ is odd, $\gamma^{(N)} = 1$, and when $N-1$ is even,

$$\begin{aligned}
1 - \gamma^{(N)} &= \binom{N-1}{\frac{N-1}{2}}\alpha^{\frac{N-1}{2}}(1-\alpha)^{\frac{N-1}{2}} \\
&= e^{-(N-1)d}\cdot\binom{N-1}{\frac{N-1}{2}}2^{-(N-1)},
\end{aligned}$$

where

$$\lim_{N\to\infty} \binom{N-1}{\frac{N-1}{2}} 2^{-(N-1)} = 0.$$

Thus $1 - \gamma^{(N)} = O(e^{-Nd})$ as $N \to \infty$.

Therefore,

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\sigma}}[R_i(\boldsymbol{X})] \\
&\le V_{\mathrm{LB}}(\epsilon) + \frac{g'(\epsilon)(e^\epsilon+1)^2}{2e^\epsilon}\left(h\Big(\gamma^{(N)} - e^{-(N-1)d}\Big) - \frac{2\theta}{2\theta-1}\right) \\
&\le V_{\mathrm{LB}}(\epsilon) + \frac{g'(\epsilon)(e^\epsilon+1)^2}{2e^\epsilon}\frac{1}{2\theta-1}\frac{1}{\gamma^{(N)} - 2e^{-(N-1)d}} \\
&\quad \cdot \Big((2-t)e^{-2(N-1)d} + (2(1-\gamma^{(N)}) + t)e^{-(N-1)d} + O(e^{-Nd})\Big) \\
&= V_{\mathrm{LB}}(\epsilon) + O(e^{-Nd}),
\end{aligned}$$

as $N \to \infty$, which completes the proof. □

3.3.3 Extension to Heterogeneous Cost Functions

Our results on the value of privacy are also valid in the scenario where individuals' privacy cost functions are heterogeneous and known. In this case, the value of ϵ units of privacy is still measured by the minimum payment of all nonnegative payment mechanisms under which an individual's best response in a Nash equilibrium is to report the data with a privacy level of ϵ. However, with heterogeneous cost functions, this value differs from individual to individual. Following similar notation, we let $V_i(\epsilon)$ denote the value of ϵ units of privacy to individual i, and let g_i denote the cost function of individual i. Then the following lower and upper bounds, which are almost identical to those in Theorem 3 and 4 except the heterogeneous cost function $g_i(\epsilon)$, hold

$$g_i'(\epsilon)\frac{e^\epsilon+1}{e^\epsilon}\left(\frac{\theta}{2\theta-1}(e^\epsilon+1) - 1\right) \le V_i(\epsilon)$$

$$\leq g_i'(\epsilon)\frac{e^\epsilon + 1}{e^\epsilon}\left(\frac{\theta}{2\theta - 1}(e^\epsilon + 1) - 1\right) + O(e^{-Nd}).$$

The lower bound above can be derived directly from the proof of Theorem 3, since the proof does not depend on whether the cost functions are homogeneous or not. The upper bound above is given by a payment mechanism that works similar to $\boldsymbol{R}^{(N,\epsilon)}$, with the g' in $R_i^{(N,\epsilon)}$ replaced by g_i'. In this mechanism, the strategy profile $\boldsymbol{\sigma}^{(\epsilon)}$ is still a Nash equilibrium, and the expected payment to individual i at this equilibrium can still be upper bounded as in Theorem 4 but again with g' replaced by g_i'.

3.4 Payment vs. Accuracy

In this section, we apply the fundamental bounds on the value of privacy to the payment–accuracy problem, where the data collector aims to minimize the total payment while achieving an accuracy target in learning the state. The solution of this problem can be used to guide the design of review systems. For example, to evaluate the underlying value of a new product, a review system can utilize the results in this section to design a payment mechanism for eliciting informative feedback from testers.

3.4.1 Payment–Accuracy Problem

The data collector learns the state W from the reported data $X_1, X_2, \ldots, X_N$, which is collected through some payment mechanism, by performing hypothesis testing between the following two hypotheses:

$$H_0\colon W = 0,$$

$$H_1\colon W = 1.$$

The conditional distributions of the reported data given the hypotheses are specified by the strategy profile in a Nash equilibrium of the payment mechanism. According to Lemma 1, we can write an equilibrium strategy profile in the form of $(\sigma_i^{(\epsilon_i)}) = (\sigma_1^{(\epsilon_1)}, \sigma_2^{(\epsilon_2)}, \ldots, \sigma_N^{(\epsilon_N)})$ with $\epsilon_i \in \mathbb{R} \backslash \{0\} \cup \{\perp\!\!\!\perp\}$, where recall that $\sigma_i^{(\epsilon_i)}$ is the ϵ_i-strategy. For ease of notation, a non-informative strategy is also called an ϵ-strategy but with $\epsilon = \perp\!\!\!\perp$. Let $\mathcal{R}(\epsilon_1, \epsilon_2, \ldots, \epsilon_N)$ denote the set of nonnegative payment mechanisms in which $(\sigma_i^{(\epsilon_i)})$ is a Nash equilibrium.

We consider an information-theoretic approach based on the Chernoff information Cover and Thomas (2006) to measure the accuracy that can be achieved in hypothesis testing. For each individual i, let $D(\epsilon_i)$ denote the Chernoff information between the conditional distributions of X_i given $W = 1$ and $W = 0$. The larger $D(\epsilon_i)$ is, the more possible that the two hypotheses can be distinguished. In later discussions we will see that the Chernoff information is closely related to the best achievable probability of error.

The data collector aims to minimize the total payment while achieving an accuracy target. The design choices include the number of individuals N, the parameters $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$, and the payment mechanism $\boldsymbol{R}$ in which the strategy profile $(\sigma_i^{(\epsilon_i)})$ is a Nash equilibrium. Then we formulate the mechanism design problem as the following optimization problem, which we call the *payment–accuracy problem*:

$$\begin{aligned} \min_{\substack{N \in \mathbb{N},\, \epsilon_i \in \mathbb{R} \backslash \{0\} \cup \{\perp\!\!\!\perp\}, \forall i \\ \boldsymbol{R} \in \mathcal{R}(\epsilon_1, \epsilon_2, \ldots, \epsilon_N)}} \quad & \sum_{i=1}^{N} \mathbb{E}_{(\sigma_i^{(\epsilon_i)})}[R_i(\boldsymbol{X})] \\ \text{subject to} \quad & e^{-\sum_{i=1}^{N} D(\epsilon_i)} \leq \tau, \end{aligned}$$

where the accuracy target is represented by τ, which is related to the maximum allowable error. We focus on the range $\tau \in (0, 1)$ for nontriviality. Let $F(\tau)$ denote the optimal payment in this problem, i.e., the infimum of the total payment while satisfying the accuracy target τ.

3.4.2 Bounds on the Payment–Accuracy Problem

We present bounds on $F(\tau)$ in Theorem 5 below. For convenience, we define

$$\widetilde{\epsilon} = \inf\left\{\arg\max\left\{\frac{D(\epsilon)}{V_{\mathrm{LB}}(\epsilon)} : \epsilon > 0\right\}\right\}, \quad \widetilde{N} = \left\lceil \frac{\ln(1/\tau)}{D(\widetilde{\epsilon})} \right\rceil, \tag{3.21}$$

where recall that $V_{\mathrm{LB}}(\epsilon)$ is the lower bound in Theorem 3.

Theorem 5. *The optimal payment $F(\tau)$ in the payment–accuracy problem for a given accuracy target $\tau \in (0,1)$ is bounded as: $(\widetilde{N}-1)V_{\mathrm{LB}}(\widetilde{\epsilon}) \leq F(\tau) \leq \widetilde{N}V_{\mathrm{LB}}(\widetilde{\epsilon}) + O(\tau\ln(1/\tau))$, where the $O(\cdot)$ is for $\tau \to 0$.*

The upper bound in Theorem 5 is given by the designed mechanism $\boldsymbol{R}^{(N,\epsilon)}$ with parameters chosen as $\epsilon = \widetilde{\epsilon}$ and $N = \widetilde{N}$. Note that $\widetilde{\epsilon}$ can be proved to have a well-defined finite value independent of τ. By the lower and upper bounds on the value of privacy, the payment to each individual in $\boldsymbol{R}^{(\widetilde{N},\widetilde{\epsilon})}$ is roughly equal to the lower bound $V_{\mathrm{LB}}(\widetilde{\epsilon})$. Then Theorem 5 indicates that the total payment of the designed mechanism $\boldsymbol{R}^{(\widetilde{N},\widetilde{\epsilon})}$ is at most one individual's payment away from the minimum, with the diminishing term $O(\tau\ln(1/\tau))$ omitted. Figure 3.2 shows an illustration of the lower and upper bounds.

Theorem 5 is proved by Lemma 4 and Lemma 5 below, where the lower bound is given by the lower bound on the value of privacy, and the upper bound is given by $\boldsymbol{R}^{(\widetilde{N},\widetilde{\epsilon})}$.

Lower Bound

First, notice that it suffices to limit the choice of each ϵ_i to $(0,+\infty)$ in the payment–accuracy problem, since when $\epsilon_i = \perp\!\!\!\perp$, $D(\epsilon_i) = 0$, and when $\epsilon_i < 0$, $D(\epsilon_i) = D(|\epsilon_i|)$ and there exists another nonnegative payment mechanism with the same payment property and a Nash equilibrium at $(\sigma_i^{(|\epsilon_i|)})$ by Lemma 2.

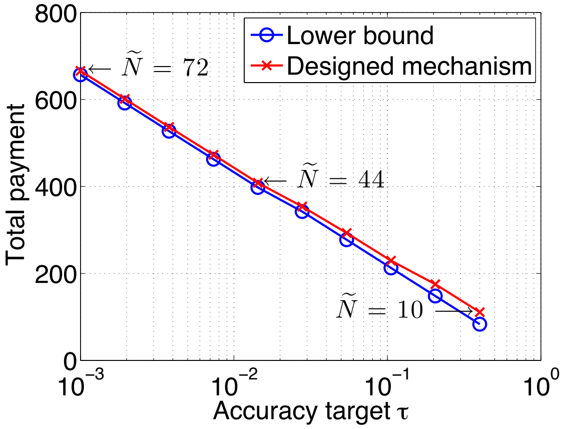


Figure 3.2: Illustration of the lower and upper bounds in Theorem 5 on the minimum total payment for achieving an accuracy target τ, where the upper bound is given by the designed mechanism $\boldsymbol{R}^{(\widetilde{N},\widetilde{\epsilon})}$. In this example, the prior PMF of the state is $P_W(1) = 0.7$, $P_W(0) = 0.3$. The quality of signals is $\theta = 0.8$. The cost function is $g(\epsilon) = \epsilon$. The range of τ shown in the figure is 0.001–0.4.

Now we use the lower bound on the value of privacy to prove the lower bound on $F(\tau)$. By Theorem 3,

$$\inf_{\boldsymbol{R}\in\mathcal{R}(\epsilon_1,\epsilon_2,\ldots,\epsilon_N)} \sum_{i=1}^{N} \mathbb{E}_{(\sigma_i^{(\epsilon_i)})}[R_i(\boldsymbol{X})] \geq \sum_{i=1}^{N} V_{\mathrm{LB}}(\epsilon_i).$$

Therefore, the optimal payment $F(\tau)$ is lower bounded by the optimal value of the following optimization problem (P1):

$$\begin{aligned} \min_{N\in\mathbb{N},\, \epsilon_i\in(0,+\infty),\forall i} \quad & \sum_{i=1}^{N} V_{\mathrm{LB}}(\epsilon_i) \\ \text{subject to} \quad & e^{-\sum_{i=1}^{N} D(\epsilon_i)} \leq \tau. \end{aligned} \tag{P1}$$

Lemma 4. *Any feasible solution* $(N, \epsilon_1, \epsilon_2, \ldots, \epsilon_N)$ *of* (P1) *satisfies*

$$\sum_{i=1}^{N} V_{\mathrm{LB}}(\epsilon_i) \geq (\widetilde{N} - 1) V_{\mathrm{LB}}(\widetilde{\epsilon}),$$

where $\widetilde{\epsilon}$ *and* $\widetilde{N}$ *are defined in* (3.21).

Lemma 4 states that the total expected payment of the data collector is at least $(\widetilde{N}-1)V_{\mathrm{LB}}(\widetilde{\epsilon})$. Note that the value given by the genie-aided payment mechanism $\widehat{\boldsymbol{R}}^{(\widetilde{\epsilon})}$ for $\widetilde{N}$ individuals is $\widetilde{N}V_{\mathrm{LB}}(\widetilde{\epsilon})$, which is at most one $V_{\mathrm{LB}}(\widetilde{\epsilon})$ away from the optimal value of (P1). We can think of $V_{\mathrm{LB}}(\epsilon)$ as the price for ϵ units of privacy and $D(\epsilon)$ as the quality that the data collector gets from ϵ units of privacy due to its contribution to the accuracy. Then the intuition for $(\widetilde{N}, \widetilde{\epsilon}, \dots, \widetilde{\epsilon})$ to be a near-optimal choice is that the privacy level $\widetilde{\epsilon}$ gives the best quality/price ratio and $\widetilde{N}$ is the fewest number of individuals to meet the accuracy target. The proof of Lemma 4 is presented is Appendix E. With this lemma, the lower bound on $F(\tau)$ in Theorem 5 is straightforward.

Upper Bound

Lemma 5. *Choose the parameters in the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$ defined in Section 3.3.2 to be $\epsilon = \widetilde{\epsilon}$ and $N = \widetilde{N}$, where $\widetilde{\epsilon}$ and $\widetilde{N}$ are defined in* (3.21)*. Then in the Nash equilibrium $\boldsymbol{\sigma}^{(\widetilde{\epsilon})}$ of $\boldsymbol{R}^{(\widetilde{N},\widetilde{\epsilon})}$, the accuracy target τ can be achieved, and the total expected payment is upper bounded as*

$$\mathbb{E}_{\boldsymbol{\sigma}^{(\widetilde{\epsilon})}}\left[\sum_{i=1}^{\widetilde{N}} R_i^{(\widetilde{N},\widetilde{\epsilon})}(\boldsymbol{X})\right] \le \widetilde{N}V_{\mathrm{LB}}(\widetilde{\epsilon}) + O(\tau \ln(1/\tau)).$$

This lemma follows from Theorem 4 and we omit the proof here. Since the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$ together with $\epsilon = \widetilde{\epsilon}$ and $N = \widetilde{N}$ is a feasible solution of the payment–accuracy problem, the upper bound in this lemma gives the upper bound on $F(\tau)$ in Theorem 5.

3.5 Discussions on the Accuracy Metric

When we study the relation between payment and accuracy, the accuracy can also be measured by the best achievable probability of error, defined as

$$p_e = \inf_{\psi} \mathbb{P}_{(\sigma_i^{(\epsilon_i)})}(\psi(\boldsymbol{X}) \neq W),$$

where $\psi(\boldsymbol{x})$ is a decision function, with $\psi(\boldsymbol{x}) = 0$ implying that H_0 is accepted and $\psi(\boldsymbol{x}) = 1$ implying that H_1 is accepted. However, p_e is difficult to deal with analytically since its exact form in terms of $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ is intractable.

We measure the accuracy based on the Chernoff information, which is an information-theoretic metric closely related to p_e. It can be proved by the Bhattacharyya bound Kailath (1967) that at the strategy profile $(\sigma_i^{(\epsilon_i)})$,

$$p_e \leq e^{-\sum_{i=1}^{N} D(\epsilon_i)}. \tag{3.22}$$

Therefore, if we want to guarantee that $p_e \leq p_e^{\max}$ for some maximum allowable probability of error $p_e^{\max}$, we can choose $\tau = p_e^{\max}$ in the payment–accuracy problem. In fact, the metric based on the Chernoff information is very close to the metric p_e, since the upper bound (3.22) is tight in exponent when all the ϵ_i are the same, i.e., when the reported data is i.i.d. given the hypothesis.

3.6 Conclusions

In this chapter, we studied “the value of privacy” under a game-theoretic model, where a data collector pays strategic individuals to buy their private data for a learning purpose. The individuals do not consider the data collector to be trustworthy, and thus experience a cost of privacy loss during data reporting. The value of ϵ units of privacy is measured by the minimum payment of all nonnegative payment mechanisms under which an individual's best response in a Nash equilibrium is to report the data with a privacy level of ϵ. We derived asymptotically tight lower and upper bounds on the value of privacy as the number of individuals becomes large, where the upper bound was given by a designed payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$. We further applied these fundamental limits to find the minimum total payment for the data collector to achieve certain learning accuracy target, and derived lower and upper bounds on the minimum payment. The total payment of the designed mechanism $\boldsymbol{R}^{(N,\epsilon)}$ with properly chosen parameters is at most one individual's payment away from the minimum.

Chapter 4

TRADING PRIVATE DATA WITH UNKNOWN VALUATIONS OF PRIVACY: THE EFFECT OF NEGATIVE PAYMENTS

4.1 Introduction

It is natural to expect that different individuals may have different valuations of privacy and their valuations are unknown to the data analyst. Specifically, we study a model where the privacy cost of an individual is a function of her privacy loss. The privacy loss is determined by the individual's data reporting strategy, and the cost function represents the individual's personal valuation of privacy. The exact forms of the privacy cost functions are unknown to the data analyst, which complicates the mechanism design problem. We elaborate further on this in the following. When the cost functions are known to the data analyst, as shown in Chapter 3, she can tune the mechanism such that all the individuals are willing to participate in the market in an equilibrium and the surplus payment is minimal. However, when the cost functions are unknown, the data analyst may need to set the payment high to ensure the participation of an individual in case the individual has high valuation of privacy, but it is also possible that the individual has a low valuation and then the high payment is not a cost-effective choice.

Impact of negative payments

As noted above, because different individuals may have different valuations of privacy, it can be costly for the data analyst to set a payment mechanism which guarantees nonnegative payment to each individual and every data report. With this observation, we consider payment mechanisms where the expected payment of each individual is nonnegative, but the realizations of the payments can be negative. In practice, this can model the scenario where there are repeated data collection (e.g., to learn the ratings of different movies). In some rounds the payments received by the individual may be negative, but in the long run, the total payment will be nonnegative. This is in contrast with the approach that enforces all the realizations of the payments to be nonnegative (which is called a nonnegative mechanism). The constraint of nonnegativity is appealing in practice since paying individuals is more convenient than charging individuals, but it will surely incur higher cost of the data analyst and makes the data analysis more difficult. To see this, let us consider a nonnegative mechanism and an individual whose valuation of privacy is very high. Then participating and reporting only noise to the data analyst is a better strategy for this individual than opting out since she may still receive some nonnegative payment without incurring any privacy cost. Therefore, the data analyst's payment does not buy her any useful information from this data subject, and moreover, the data analyst has to work with these unusable reports during data analysis. To address these difficulties, we utilize negative payments to "filter out" individuals with high privacy costs, i.e., we design the mechanism such that their expected utility is negative if they report only noise. This saves the data analyst's payments on poor quality data and simplifies the data analysis. We will see that we can actually drive the total cost to zero for the data analyst as the population size becomes large.

To implement negative payments in practice, the data analyst can set up an online payment system using virtual currency or credits. Instead of paying real money to an individual every time she reports a data, virtual currency or credits can be added to or reduced from the user's account. An individual can be paid a weekly or monthly with real dollars. Since the expected payment is nonnegative, the real-dollar payment over a long time period is nonnegative with a high probability. We remark that negative payments may not be feasible in many scenarios. The focus of this chapter is to reveal the fundamental benefit of negative payments to the data analyst when feasible.

With the above formulation, the interaction between the data analyst and individuals is clear: an individual acts upon the payment mechanism and her privacy cost function; the data analyst designs the payment mechanism to incentivize the individuals such that they are willing to report data with small enough perturbation that allows the data analyst to achieve the desired learning accuracy. In this formulation, we aim to answer the following questions: (1) How will the individuals behave to reconcile the conflict between privacy and rewards? (2) How should the data analyst design the mechanism such that she can achieve her learning goal cost-effectively?

Summary of Results

We consider the following model for private data, which is the same as the model considered in Chapter 3. A data analyst is interested in learning, from a population of N individuals, an underlying *state* represented by a binary random variable W. As illustrated in Figure 4.1, each individual i possesses a binary *signal* S_i, which is her private data, reflecting her knowledge about the state W. Conditional on the state W, the signals are independently generated such that the probability for each S_i to be the same as W is θ, where $0.5 < \theta < 1$. At the beginning of the data procurement, the data analyst announces a payment mechanism, which determines the amounts of payments according to the reported bits $X_1, X_2, \ldots, X_N$ of individuals.

The model for privacy cost considered in this chapter is different from those in previous chapters. Recall that the privacy cost of an individual is a function of her privacy loss. We measure the privacy loss of an individual's data reporting strategy by the level of (local) differential privacy Dwork *et al.* (2006b); Dwork (2006) of

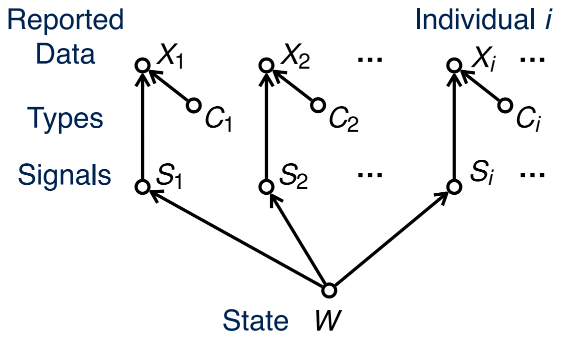


Figure 4.1: Information structure of the model with unknown valuations of privacy. The data analyst is interested in the state W, which is a binary random variable. Each individual i has a private binary signal S_i and a type C_i that characterizes her valuation of privacy. Conditional on W, $S_1, S_2, \ldots, S_N$ are i.i.d. Individual i reports data X_i, which is generated based on S_i and C_i using a randomized strategy.

the strategy. Then the privacy cost function of individual i is characterized by her *type* C_i: when individual i reports data with a (local) differential privacy level of ϵ after observing her type $C_i = c_i$, her privacy loss is ϵ and the corresponding privacy cost is $c_i\epsilon$. The type of an individual is also called her cost coefficient due to this linear form. We assume that an individual's type is independent from her private data, which is applicable to the scenario where an individual's valuation of privacy is intrinsic and thus is not affected by the specific private data she has. The cost coefficients are also illustrated in Figure 4.1. We remark that both settings where an individual's valuation of privacy is independent and correlated with her private data have been studied in the "trustworthy data analyst" model in the literature. In this chapter, we assume that the prior distribution of the state, signals and types is public information. However, the data analyst does not know the private signal and the type of an individual. An individual's utility is the difference between the payment she receives and her privacy cost.

The data analyst learns the state by performing hypothesis testing. The goal of the mechanism design is to elicit data with certain amount of information in a Bayesian Nash equilibrium to fulfill an accuracy goal with minimized total payment. Our main result is on constructing a family of payment mechanisms parameterized by the population size, the prior, and $(c_{\text{th}}, \epsilon)$. These mechanisms provide answers to the proposed questions from the following perspectives.

- *Behavior of individuals with privacy concerns.* We show that the individuals exhibit a threshold behavior in a Bayesian Nash equilibrium of the proposed mechanisms: the individuals with cost coefficients above a threshold c_{th} choose not to participate, and the individuals with cost coefficients below c_{th} participate and report data with a privacy level no smaller than ϵ. Since a larger privacy level means that the data is less perturbed, the data analyst can incentivize the participants to limit the perturbation to a desired extent by choosing an appropriate ϵ. By this result, we can also see that this family of mechanisms resolve the otherwise nuisance that individuals with high privacy costs may participate and report only noise: they are "filtered out", and the "left" participants all report data with quality guarantee.

- *Tradeoff between learning accuracy and cost.* We show that as the population size grows to infinity, the data analyst can learn the underlying state with arbitrarily small probability of error, with the total expected payment at the Bayesian Nash equilibrium going to zero. That is to say, if the data analyst can recruit a large number of individuals, she can choose appropriate parameters to fulfill her learning goal and in the meanwhile drive her cost to zero at a Bayesian Nash equilibrium. Since the total equilibrium expected payment of any mechanism is nonnegative due to individual rationality, this implies that the designed mechanism with properly chosen parameters asymptotically minimizes the cost for achieving any accuracy goal.

4.2 Model

We study the setting in which the data analyst is interested in learning an underlying state W, represented by a binary random variable. Consider a set $\mathcal{N} = \{1, 2, \ldots, N\}$ of individuals. Each individual i possesses a binary signal S_i, which is her private data, and reports data X_i, which takes values in $\mathcal{X} = \{0, 1, \perp\}$, with $\perp$ meaning "to opt out." The data analyst announces a payment mechanism $\boldsymbol{R}\colon \mathcal{X}^N \to \mathbb{R}^N$, which takes the reported data $\boldsymbol{X} = (X_1, \ldots, X_N)$ as input and produces $\boldsymbol{R}(\boldsymbol{X})$, where $R_i(\boldsymbol{X})$ is the payment to individual i. The model is illustrated in Figure 3.1. The payment mechanism induces a game among the individuals. The elements of the game are as follows.

- *Players.* The players in this game are the individuals, who are self-interested, rational and risk-neutral. Following conventional game theory notation, we let "$-i$" denote all the individuals other than some given individual i.

- *Prior Distributions.* The state W follows a probability distribution given by the PMF P_W. We assume that $P_W(1) > 0$ and $P_W(0) > 0$. The individuals' signals $\boldsymbol{S} = (S_1, S_2, \ldots, S_N)$ reflect their knowledge about the state W. Conditional on the state W, the signals $S_1, S_2, \ldots, S_N$ are independently generated according to the following conditional distributions:

$$\mathbb{P}(S_i = 1 \mid W = 1) = \theta, \quad \mathbb{P}(S_i = 0 \mid W = 1) = 1 - \theta,$$
$$\mathbb{P}(S_i = 0 \mid W = 0) = \theta, \quad \mathbb{P}(S_i = 1 \mid W = 0) = 1 - \theta,$$

where the parameter θ with $0.5 < \theta < 1$ is called the quality of signals. We refer to these conditional distributions as the signal structure of the model.

- *Types and Strategies.* An individual i's type C_i, also called her cost coefficient, characterizes her valuation of privacy. The meaning of C_i will become clear after we introduce the privacy cost function. Roughly, an individual with larger C_i experiences more privacy cost for the same privacy loss. A data reporting strategy for individual i is a plan on what to report according to her signal S_i and her type C_i. Thus it is a mapping $\sigma_i \colon \{0,1\} \times (0,+\infty) \to \mathcal{D}(\mathcal{X})$, where $\mathcal{D}(\mathcal{X})$ is the set of probability distributions on $\mathcal{X} = \{0,1,\perp\}$, prescribing a distribution to the reported data X_i for each possible value pair of S_i and C_i. Therefore, the strategy corresponds to the set of conditional distributions of X_i given S_i and C_i. Since we will discuss different strategies of individual i, we denote these conditional probabilities by

$$\mathbb{P}_{\sigma_i}(X_i = x_i \mid S_i = s_i, C_i = c_i), x_i \in \{0,1,\perp\}, s_i \in \{0,1\}, c_i \in (0,+\infty).$$

Let $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_N)$, which is called a strategy profile. A strategy profile is said to be homogeneous if all the strategies in the profile are the same.

- *Utility Functions.* The utility of each individual is the difference between the payment she receives and her privacy cost. An individual experiences a cost due to the privacy loss during data reporting. Recall that we model the privacy cost of an individual as consisting of two components: privacy loss and a privacy cost function, where the privacy loss depends on her data reporting strategy and the privacy cost function represents her valuation of privacy. For an individual i, conditional on her type $C_i = c_i$, we measure individual i's privacy loss for reporting data with strategy σ_i by the privacy level defined as follows:

$$\zeta(c_i, \sigma_i) = \max\left\{ \ln \frac{\mathbb{P}_{\sigma_i}(X_i \in \mathcal{E} \mid S_i = s_i, C_i = c_i)}{\mathbb{P}_{\sigma_i}(X_i \in \mathcal{E} \mid S_i = 1 - s_i, C_i = c_i)} : \mathcal{E} \subseteq \{0,1,\perp\}, s_i \in \{0,1\} \right\},$$

where we follow the convention that $0/0 = 1$. This measure of privacy loss is in the same vein as the local model of differential privacy Kasiviswanathan *et al.* (2011);

Dwork and Roth (2014), which views each individual's data as a database of size 1 and quantifies the privacy guarantee of her local randomizer by the differential privacy level. The difference here is that the strategy σ_i has another input C_i, since an individual can choose the way of perturbing her data according to her cost coefficient. Our measure of privacy loss is the differential privacy level of the strategy σ_i when C_i is given.

Then we model individual i's cost incurred by this privacy loss as a linear function with C_i as the coefficient, i.e., the cost can be written as

$$g(C_i, \sigma_i) = C_i \cdot \zeta(C_i, \sigma_i).$$

We call g the privacy cost function.

We assume that the coefficients $C_1, C_2, \ldots, C_N$ are i.i.d. positive random variables with CDF F_C, and they are independent of W and $\boldsymbol{S}$. The randomness of these coefficients captures the data analyst's uncertainty of individuals' valuations of privacy. The independence assumption is applicable to the scenario where individuals' valuations of privacy are intrinsic and thus are not affected by the specific private data they have. We further assume that F_C is a continuous function and $F_C(c) > 0$ for any $c > 0$, which means that it is possible for individuals to have an arbitrarily low valuation of privacy.

Mechanism design

The data analyst cannot force an individual to report data with a specific strategy. However, the data analyst can design the payment mechanism to impact individuals' strategies to drive the individuals to act in a desired way since the individuals are rational, i.e., they will choose the strategies that benefit them most. We consider the Bayesian Nash equilibria in a payment mechanism, viewing C_i as individual i's type.

Definition 7. A strategy profile $\boldsymbol{\sigma}$ is a *Bayesian Nash equilibrium* of a payment mechanism $\boldsymbol{R}$ if for any individual i, any $c_i > 0$ and any strategy σ_i',

$$\mathbb{E}_{\boldsymbol{\sigma}}[R_i(\boldsymbol{X}) - g(C_i, \sigma_i) \mid C_i = c_i] \geq \mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) - g(C_i, \sigma_i') \mid C_i = c_i],$$

where the subscript $\boldsymbol{\sigma}$ and (σ_i', σ_{-i}) indicate that the distribution of $\boldsymbol{X}$ is determined by the strategy profile $\boldsymbol{\sigma}$ and (σ_i', σ_{-i}), respectively.

The data analyst is interested in learning the state W from the reported data $\boldsymbol{X}$, so she performs hypothesis testing between the following two hypotheses:

$$H_0 \colon W = 0,$$

$$H_1 \colon W = 1.$$

The learning accuracy is measured by the probability of error, denoted by p_e. An accuracy goal can be written as $p_e \leq p_e^{\max}$ for some $p_e^{\max}$.

Then the data analyst aims to design a payment mechanism such that her accuracy goal can be fulfilled at a Bayesian Nash equilibrium and the corresponding total expected payment is minimized. It is easy to see that the equilibrium total expected payment is nonnegative in any mechanism due to the nonnegativity of privacy cost functions and individual rationality. In this mechanism design problem, the joint distribution $\mathcal{P}$ of the state W, the signal $\boldsymbol{S}$ and the cost coefficients, which can be represented by (P_W, θ, F_C), is common knowledge. The data analyst announces the form of the payment mechanism and then the individuals report data simultaneously. The reported data $\boldsymbol{X}$ is public. Each individual i's signal and type, S_i and C_i, are not observable to other individuals or the data analyst. No one has access to the state W.

Theorem 6. *To achieve any accuracy goal of the data analyst, the total expected payment needed at an equilibrium is $o(1)$. Specifically, there exists a sequence of mechanisms, each of which is designed for a different population size N, such that the accuracy goal can be fulfilled at a Bayesian Nash equilibrium of every mechanism in the sequence, and the total expected payment goes to zero as the population size N goes to infinity; i.e., this sequence of mechanisms is asymptotically optimal.*

In the remainder of this section, we present the design of a family of payment mechanisms, parameterized by the population size N, the prior $\mathcal{P}$, a cost coefficient threshold parameter c_{th} and a data quality parameter ϵ. The asymptotically optimal sequence of mechanisms in Theorem 6 is given by a sequence of mechanisms within this family with properly chosen parameters. In particular, c_{th} is a threshold on cost coefficients such that an individual is expected to participate if her coefficient does not exceed the threshold; and ϵ is the target quality which is the level noise expected in the reported data. The formula for calculating c_{th} and ϵ will be presented in (4.5)–(4.11). Theorem 6 is a high level description of Theorem 8, which will be proved in the remainder of this chapter.

Payment Mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$

1. Each individual reports her data (which can also be "to opt out").

2. Compute the number of participants n.

3. For non-participating individuals, the payment is zero.

4. If there is only one participant, the data analyst pays zero to this participant.

Otherwise, for each participating individual i, compute

$$M_{-i} = \begin{cases} 1 & \text{if } \displaystyle\sum_{j\colon X_j \neq \perp, j \neq i} X_j \geq \left\lfloor \frac{n-1}{2} \right\rfloor + 1, \\ 0 & \text{otherwise,} \end{cases}$$

which is the majority of other participants' reported data. Then the data analyst pays individual i according to X_i and M_{-i} as follows:

$$R_i^{(N,\mathcal{P},c_{\text{th}},\epsilon)}(\boldsymbol{X}) = A_{X_i,M_{-i}} \frac{c_{\text{th}}(e^\epsilon+1)^2}{2e^\epsilon} + B_{M_{-i}}\left(\frac{c_{\text{th}}(e^\epsilon+1)}{e^\epsilon} + c_{\text{th}}\epsilon\right),$$

where $A_{1,1}, A_{0,1}, A_{1,0}, A_{0,0}, B_1, B_0$ are given below.

Next we define the coefficients $A_{1,1}$, $A_{0,1}$, $A_{1,0}$, $A_{0,0}$, B_1, B_0 used in the mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$ through a series of calculations. In a nutshell, $A_{1,1}$ and $A_{0,0}$ determine the reward part of the payment to an individual when her reported data matches the majority of others; similarly, $A_{0,1}$ and $A_{1,0}$ determine the penalty part of the payment to an individual when her reported data does not match the majority of others. They incentivize the individuals to report data that reveals certain amount of information about their private signals. The coefficients B_1 and B_0 offset the payments for the cases that the majority of others' reports is 1 and 0, respectively, to discourage the individuals with cost coefficients above threshold parameter c_{th} from participating. We remark that when an individual's reported data does not match with the majority of others, these coefficients make sure that the payment to this individual is negative.

The definition of the coefficients $A_{1,1}$, $A_{0,1}$, $A_{1,0}$, $A_{0,0}$, B_1, B_0 involves some intermediate quantities, the physical meanings of which will be given after we characterize a Bayesian Nash equilibrium of the mechanism in Section 4.4. Given a $c_{\text{th}} \in (0, +\infty)$ and $\epsilon \in (0, +\infty)$, for each $c_i \in (0, c_{\text{th}})$, we consider the following equation with variable ξ:

$$\frac{c_{\text{th}}(e^\epsilon+1)^2}{e^\epsilon} \frac{e^\xi}{(e^\xi+1)^2} - c_i = 0.$$

It can be proved that this equation has a unique solution in $(0, +\infty)$. Let this solution define a function $\xi(c_i)$. Specifically,

$$\xi(c_i) = \ln\left(\frac{1}{\frac{1}{2} - \sqrt{\frac{1}{4} - \frac{c_i}{c_{\text{th}}}\frac{e^\epsilon}{(e^\epsilon+1)^2}}} - 1\right). \tag{4.1}$$

Let

$$\mu = \int_0^{c_{\text{th}}} \frac{e^{\xi(c_i)}}{e^{\xi(c_i)} + 1}\, dF_{C|C_i \le c_{\text{th}}}(c_i),$$

and

$$\alpha = \theta\mu + (1-\theta)(1-\mu), \tag{4.2}$$

where $F_{C|C_i \le c_{\text{th}}}$ is the conditional distribution of C_i given $C_i \le c_{\text{th}}$.

Given that the number of participants is n with $n \ge 2$, we define the following quantities. Consider a random variable that follows the binomial distribution with parameters $n-1$ and α. Let $\beta^{(n)}$ denote the probability that this random variable is greater than or equal to $\lfloor \frac{n-1}{2} \rfloor + 1$. For convenience, we define the following quantity to deal with technical details:

$$\gamma^{(n)} = \begin{cases} 1 - \binom{n-1}{\frac{n-1}{2}} \alpha^{\frac{n-1}{2}} (1-\alpha)^{\frac{n-1}{2}} & \text{if } n-1 \text{ is even,} \\ 1 & \text{if } n-1 \text{ is odd.} \end{cases}$$

Let $P_{\ge 1} = 1 - (1 - F_C(c_{\text{th}}))^{N-1}$, where recall that F_C is the CDF of C_i. We define

$$A_{1,1} = \frac{P_W(1)\theta(1-\beta^{(n)}) + P_W(0)(1-\theta)(1-(\gamma^{(n)} - \beta^{(n)}))}{P_{\ge 1}P_W(1)P_W(0)(2\theta-1)(2\beta^{(n)} - \gamma^{(n)})},$$

$$A_{0,1} = -\frac{P_W(1)(1-\theta)(1-\beta^{(n)}) + P_W(0)\theta(1-(\gamma^{(n)} - \beta^{(n)}))}{P_{\ge 1}P_W(1)P_W(0)(2\theta-1)(2\beta^{(n)} - \gamma^{(n)})},$$

$$A_{1,0} = -\frac{P_W(1)\theta\beta^{(n)} + P_W(0)(1-\theta)(\gamma^{(n)} - \beta^{(n)})}{P_{\ge 1}P_W(1)P_W(0)(2\theta-1)(2\beta^{(n)} - \gamma^{(n)})},$$

$$A_{0,0} = \frac{P_W(1)(1-\theta)\beta^{(n)} + P_W(0)\theta(\gamma^{(n)} - \beta^{(n)})}{P_{\ge 1}P_W(1)P_W(0)(2\theta-1)(2\beta^{(n)} - \gamma^{(n)})},$$

$$B_1 = -\frac{P_W(1)(1-\beta^{(n)}) - P_W(0)(1-(\gamma^{(n)} - \beta^{(n)}))}{2P_{\ge 1}P_W(1)P_W(0)(2\beta^{(n)} - \gamma^{(n)})},$$

$$B_0 = \frac{P_W(1)\beta^{(n)} - P_W(0)(\gamma^{(n)} - \beta^{(n)})}{2P_{\ge 1}P_W(1)P_W(0)(2\beta^{(n)} - \gamma^{(n)})}.$$

4.4 Bayesian Nash Equilibrium

In this section, we first characterize the individuals' behavior at a Bayesian Nash equilibrium of the designed mechanism. The equilibrium behavior affects the quality of the reported data and the payments. Then we leverage the properties of the Bayesian Nash equilibrium to explain the physical meanings of the quantities defined during the construction of the mechanism in Section 4.3.

Theorem 7. *The mechanism* $\boldsymbol{R}^{(N,\mathcal{P},c_{\mathrm{th}},\epsilon)}$ *yields a Bayesian Nash equilibrium* $\boldsymbol{\sigma}$*, in which each individual* i*'s strategy* σ_i *is described as follows:*

- *If* $c_i > c_{\mathrm{th}}$*,*

  $$\mathbb{P}_{\sigma_i}(X_i = \perp \mid S_i = s_i, C_i = c_i) = 1, \quad \textit{for any } s_i \in \{0,1\};$$

  i.e., if individual i*'s cost coefficient is larger than the parameter* c_{th}*, individual* i *declines to participate regardless of her signal.*

- *If* $c_i \le c_{\mathrm{th}}$*,*

  $$\begin{aligned} \mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 1, C_i = c_i) = \mathbb{P}_{\sigma_i}(X_i = 0 \mid S_i = 0, C_i = c_i) &= \frac{e^{\xi(c_i)}}{e^{\xi(c_i)} + 1}, \\ \mathbb{P}_{\sigma_i}(X_i = 0 \mid S_i = 1, C_i = c_i) = \mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 0, C_i = c_i) &= \frac{1}{e^{\xi(c_i)} + 1}, \end{aligned}$$

  where $\xi(c_i)$ *is defined in* (4.1)*; i.e., if individual* i*'s cost coefficient is no larger than the parameter* c_{th}*, individual* i *flips her signal with a probability depending on her cost coefficient to generate her reported data.*

This theorem presents our results on the threshold behavior of individuals and the quality guarantee of the reported data. We sketch the proof of Theorem 7 below. A complete proof is given in Appendix F.

Proof Sketch. We write $\boldsymbol{R}$ to represent the mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$ for conciseness in this proof. Consider any individual i and any strategy σ_i'. Given $C_i = c_i$, let

$$\begin{aligned}
p_1 &= \mathbb{P}_{\sigma_i'}(X_i = 1 \mid C_i = c_i, S_i = 1), \quad p_0 = \mathbb{P}_{\sigma_i'}(X_i = 1 \mid C_i = c_i, S_i = 0),\\
q_1 &= \mathbb{P}_{\sigma_i'}(X_i = 0 \mid C_i = c_i, S_i = 1), \quad q_0 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid C_i = c_i, S_i = 0).
\end{aligned}$$

Then the differential privacy level of σ_i' at c_i can be written as

$$\begin{aligned}
\zeta(p_1, p_0, q_1, q_0) = \max\biggl\{&\left|\ln \frac{p_1}{p_0}\right|, \left|\ln \frac{1-p_1}{1-p_0}\right|, \left|\ln \frac{q_1}{q_0}\right|, \left|\ln \frac{1-q_1}{1-q_0}\right|,\\
&\qquad\qquad\qquad\qquad \left|\ln \frac{1-p_1-q_1}{1-p_0-q_0}\right|, \left|\ln \frac{p_1+q_1}{p_0+q_0}\right|\biggr\}.
\end{aligned}$$

By the design of the mechanism, when other individuals follow σ_{-i}, the expected utility of individual i can be written as

$$\begin{aligned}
&U(p_1, p_0, q_1, q_0)\\
&\triangleq \mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) - g(C_i, \sigma_i') \mid C_i = c_i]\\
&= K_1 p_1 + K_0 p_0 + L_1 q_1 + L_0 q_0 - c_i \zeta(p_1, p_0, q_1, q_0),
\end{aligned}$$

where

$$\begin{aligned}
K_1 = L_0 &= \frac{1}{2}\biggl(\frac{c_{\text{th}}(e^\epsilon + 1)}{e^\epsilon} + c_{\text{th}}\epsilon\biggr),\\
K_0 = L_1 &= \frac{1}{2}\biggl(\frac{c_{\text{th}}(e^\epsilon + 1)}{e^\epsilon} + c_{\text{th}}\epsilon - \frac{c_{\text{th}}(e^\epsilon + 1)^2}{e^\epsilon}\biggr).
\end{aligned}$$

Now we find the best response of individual i, i.e., an optimal solution of the optimization problem below, by three steps:

$$\begin{aligned}
\max_{p_1, p_0, q_1, q_0} \quad & U(p_1, p_0, q_1, q_0)\\
\text{subject to} \quad & 0 \le p_1 \le 1, 0 \le q_1 \le 1,\\
& 0 \le p_1 + q_1 \le 1,
\end{aligned}$$

$$0 \le p_0 \le 1, 0 \le q_0 \le 1,$$
$$0 \le p_0 + q_0 \le 1.$$

Step 1: First, by the symmetry that $K_1 = L_0$ and $K_0 = L_1$, we can focus on an optimal solution $(p_1^*, p_0^*, q_1^*, q_0^*)$ such that $p_1^* = q_0^*$ and $p_0^* = q_1^*$, since for any feasible solution (p_1, p_0, q_1, q_0), the solution (p_1', p_0', q_1', q_0') given by $p_1' = q_0' = \frac{p_1+q_0}{2}, p_0' = q_1' = \frac{p_0+q_1}{2}$ yields $U(p_1', p_0', q_1', q_0') \ge U(p_1, p_0, q_1, q_0)$. Further, an optimal solution $(p_1^*, p_0^*, q_1^*, q_0^*)$ such that $p_1^* = q_0^*$ and $p_0^* = q_1^*$ must satisfy that $p_1^* \ge q_1^*$, since otherwise by swapping p_1^* and p_0^* with q_1^* and q_0^*, respectively, the utility is increased, which contradicts with the optimality of $(p_1^*, p_0^*, q_1^*, q_0^*)$.

Step 2: Next, for any such an optimal solution, i.e., $(p_1^*, p_0^*, q_1^*, q_0^*)$ with $p_1^* = q_0^*$ and $p_0^* = q_1^*$, we can prove that one of the following two holds: $p_1^* = q_0^* = p_0^* = q_1^* = 0$ or $p_1^* + q_1^* = p_0^* + q_0^* = 1, p_1^* > q_1^*$.

Step 3: According to Step 1 and Step 2, we can find an optimal solution among those feasible solutions, say (p_1, p_0, q_1, q_0), with $p_1 = q_0$ and $p_0 = q_1$, and satisfy either

$$p_1 = q_0 = p_0 = q_1 = 0, \text{ or} \tag{4.3}$$

$$p_1 + q_1 = p_0 + q_0 = 1, p_1 > q_1. \tag{4.4}$$

Consider any feasible solution (p_1, p_0, q_1, q_0) with $p_1 = q_0$ and $p_0 = q_1$ and satisfies (4.4), which can be written as

$$p_1 = q_0 = \frac{e^{\epsilon_i}}{e^{\epsilon_i} + 1}, \quad p_0 = q_1 = \frac{1}{e^{\epsilon_i} + 1},$$

for some $\epsilon_i > 0$. Then

$$U(p_1, p_0, q_1, q_0) = -\frac{c_{\text{th}}(e^\epsilon + 1)^2}{e^\epsilon} \frac{1}{e^{\epsilon_i} + 1} - c_i\epsilon_i + \frac{c_{\text{th}}(e^\epsilon + 1)}{e^\epsilon} + c_{\text{th}}\epsilon.$$

Consider a function $h\colon (0, +\infty) \to \mathbb{R}$ defined as

$$h(\epsilon_i) = -\frac{c_{\text{th}}(e^\epsilon + 1)^2}{e^\epsilon} \frac{1}{e^{\epsilon_i} + 1} - c_i\epsilon_i.$$

Then h is a strictly concave function, and thus ϵ_i^* that satisfies

$$h'(\epsilon_i^*) = \frac{c_{\text{th}}(e^\epsilon + 1)^2}{e^\epsilon} \frac{e^{\epsilon_i^*}}{(e^{\epsilon_i^*} + 1)^2} - c_i = 0,$$

i.e., $\epsilon_i^* = \xi(c_i)$ defined in (4.1), maximizes $h(\cdot)$, and hence maximizes the utility. Therefore, among those feasible solutions that satisfy (4.4), the solution $(\widetilde{p}_1^*, \widetilde{p}_0^*, \widetilde{q}_1^*, \widetilde{q}_0^*)$ with

$$\widetilde{p}_1^* = \widetilde{q}_0^* = \frac{e^{\xi(c_i)}}{e^{\xi(c_i)} + 1}, \quad \widetilde{p}_0^* = \widetilde{q}_1^* = \frac{1}{e^{\xi(c_i)} + 1}$$

maximizes the utility. This implies that an optimal solution is either $(0,0,0,0)$ or $(\widetilde{p}_1^*, \widetilde{p}_0^*, \widetilde{q}_1^*, \widetilde{q}_0^*)$. Next, we can prove that if $c_i > c_{\text{th}}$, we have $\xi(c_i) < \epsilon$ and $U(\widetilde{p}_1^*, \widetilde{p}_0^*, \widetilde{q}_1^*, \widetilde{q}_0^*) < 0 = U(0,0,0,0)$, so $(0,0,0,0)$ is an optimal solution. For the other case that $c_i \leq c_{\text{th}}$, we can prove that $U(\widetilde{p}_1^*, \widetilde{p}_0^*, \widetilde{q}_1^*, \widetilde{q}_0^*) \geq 0 = U(0,0,0,0)$, so $(\widetilde{p}_1^*, \widetilde{p}_0^*, \widetilde{q}_1^*, \widetilde{q}_0^*)$ is an optimal solution.

In summary, by the three steps above, a best response of individual i is the strategy σ_i described in the theorem, which completes the proof that $\boldsymbol{\sigma}$ is a Bayesian Nash equilibrium of the mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$. □

The following corollary describes the quality of the reported data and the expected payment to each participant at the Bayesian Nash equilibrium in Theorem 7.

Corollary 1. *For the mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$, consider the Bayesian Nash equilibrium $\boldsymbol{\sigma}$ given in Theorem 7.*

- *For each participating individual i,*

$$\begin{aligned} &\mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 1, \text{ individual } i \text{ participates}) \\ &= \mathbb{P}_{\sigma_i}(X_i = 0 \mid S_i = 0, \text{ individual } i \text{ participates}) \\ &= \mu \\ &\geq \frac{e^\epsilon}{e^\epsilon + 1}. \end{aligned}$$

- *The expected payment to each participating individual i is bounded as*

$$\mathbb{E}_{\boldsymbol{\sigma}}[R_i^{(N,\mathcal{P},c_{\text{th}},\epsilon)}(\boldsymbol{X}) \mid \textit{individual } i \textit{ participates}] \leq c_{\text{th}}(1 + e^{-\epsilon} + \epsilon).$$

Proof. The proof for these two results is intuitive once we have Theorem 7. In the equilibrium $\boldsymbol{\sigma}$, the event $\{$individual i participates$\}$ is equivalent to the event $\{C_i \leq c_{\text{th}}\}$. Thus

$$\begin{aligned}\mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 1, \text{ individual } i \text{ participates}) &= \mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 1, C_i \leq c_{\text{th}}) \\ &= \int_0^{c_{\text{th}}} \frac{e^{\xi(c_i)}}{e^{\xi(c_i)} + 1}\, dF_{C \mid C_i \leq c_{\text{th}}}(c_i) \\ &= \mu.\end{aligned}$$

By the definition of $\xi(c_i)$ in (4.1), $\xi(c_i) \geq \epsilon$ when $c_i \leq c_{\text{th}}$. Hence

$$\mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 1, \text{ individual } i \text{ participates}) \geq \frac{e^{\epsilon}}{e^{\epsilon} + 1}.$$

Since for any $c_i \leq c_{\text{th}}$, $\mathbb{P}_{\sigma_i}(X_i = 0 \mid C_i = c_i, S_i = 0) = \mathbb{P}_{\sigma_i}(X_i = 1 \mid C_i = c_i, S_i = 1)$, we have

$$\begin{aligned}&\mathbb{P}_{\sigma_i}(X_i = 0 \mid S_i = 0, \text{ individual } i \text{ participates}) \\ &= \mathbb{P}_{\sigma_i}(X_i = 1 \mid S_i = 1, \text{ individual } i \text{ participates}).\end{aligned}$$

By the calculations in the proof of Theorem 7, the expected payment of individual i given $C_i = c_i$ with $c_i \leq c_{\text{th}}$ satisfies

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\sigma}}[R_i^{(N,\mathcal{P},c_{\text{th}},\epsilon)}(\boldsymbol{X}) \mid C_i = c_i] &= -\frac{c_{\text{th}}(e^{\epsilon} + 1)^2}{e^{\epsilon}} \frac{1}{e^{\xi(c_i)} + 1} + \frac{c_{\text{th}}(e^{\epsilon} + 1)}{e^{\epsilon}} + c_{\text{th}}\epsilon \\ &\leq c_{\text{th}}(1 + e^{-\epsilon} + \epsilon).\end{aligned}$$

Hence

$$\mathbb{E}_{\boldsymbol{\sigma}}[R_i^{(N,\mathcal{P},c_{\text{th}},\epsilon)}(\boldsymbol{X}) \mid \text{individual } i \text{ participates}]$$

$$
\begin{aligned}
&= \int_0^{c_{\text{th}}} \mathbb{E}_{\boldsymbol{\sigma}}[R_i(\boldsymbol{X}) \mid C_i = c_i] \, dF_{C|C_i \le c_{\text{th}}(c_i)} \\
&\le c_{\text{th}}(1 + e^{-\epsilon} + \epsilon).
\end{aligned}
$$

□

Theorem 7 and Corollary 1 show how individuals with high privacy costs are "filtered out" in the equilibrium by negative payments. In other words, they will decide not to participate because the expected payment is negative, which is a result of possible negative payments in the proposed mechanism. The "left" individuals, i.e., participants, all report data with quality guarantee. The roles of the parameters c_{th} and ϵ in the designed mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$ are as follows: The parameter c_{th} works as a threshold on the cost coefficients for participation; The parameter ϵ gives a guarantee on the probability that a participant's reported data is the same as the signal, which measures the quality of the reported data. We remark that in this equilibrium, each individual's exact cost coefficient is not revealed to other.

The physical meanings of the quantities $\xi(c_i)$, μ, α, $\beta^{(n)}$, $\gamma^{(n)}$ and $P_{\ge 1}$ defined during the construction of the mechanism in Section 4.3 can be well explained at the Bayesian Nash equilibrium given in Theorem 7. The quantity $\xi(c_i)$ shows up in Theorem 7, characterizing the strategy σ_i of individual i when $c_i \le c_{\text{th}}$. It is the differential privacy level of σ_i given $C_i = c_i$ when $c_i \le c_{\text{th}}$. Now let us condition on the event that individual i participates, which, by Theorem 7, is equivalent to the event $C_i \le c_{\text{th}}$. The quantity μ shows up in Corollary 1, and it is the probability that individual i truthfully reports her signal, given whatever the signal is. Then the quantity α is the probability that the reported data X_i is consistent with the state W, given whatever the state is. Conditional on the event that there are $n-1$ participants among the individuals other than individual i, where $n \ge 2$, the quantities β_n and $1-(\gamma_n-\beta_n)$ are the probabilities that the majority of these participants' reported data agrees with the state, given that the state is 1 and 0, respectively. Finally, the quantity $P_{\geq 1}$ is the probability that at least one individual among the individuals other than individual i participates.

4.5 Accuracy and Payment

In this section, we show that the data analyst can achieve any accuracy goal in the Bayesian Nash equilibrium with proper choice of parameters N, c_{th} and ϵ. The cost of the data analyst, which is the total expected payment at the equilibrium, goes to zero as the number of individuals goes to infinity. Since the privacy cost of an individual is always nonnegative, the total expected payment at an equilibrium of any mechanism is nonnegative due to individual rationality. Therefore, the designed mechanism asymptotically minimizes the cost for the data analyst to achieve any accuracy goal.

Recall that with the purchased data $\boldsymbol{X}$, the data analyst learns the state W by performing hypothesis testing between the following two hypotheses:

$$H_0 \colon W = 0,$$
$$H_1 \colon W = 1.$$

An accuracy goal can be written as $p_e \leq p_e^{\max}$ for some $p_e^{\max}$, where p_e is the probability of error for hypothesis testing. We consider the decision function of maximum likelihood and choose the values for $N, c_{\text{th}}, \epsilon$ as follows. First pick any ϵ such that

$$\epsilon \in (0, +\infty). \tag{4.5}$$

Let

$$D(\epsilon) = \frac{1}{2} \ln \frac{(e^\epsilon + 1)^2}{4(\theta e^\epsilon + 1 - \theta)((1-\theta)e^\epsilon + \theta)}, \tag{4.6}$$

$$n_e(\epsilon) = \frac{-\ln(\frac{1}{2}p_e^{\max})}{D(\epsilon)}, \tag{4.7}$$

$$\rho(\epsilon) = \frac{1}{n_e(\epsilon)p_e^{\max}} + 2 + \sqrt{\frac{1}{(n_e(\epsilon))^2(p_e^{\max})^2} + \frac{2}{n_e(\epsilon)p_e^{\max}}}. \tag{4.8}$$

Then pick any integer N such that

$$N > \rho(\epsilon)n_e(\epsilon). \tag{4.9}$$

For the selected N, let

$$p_{\text{th}}(N,\epsilon) = \frac{\rho(\epsilon)n_e(\epsilon)}{N}, \tag{4.10}$$

which is roughly the participation percentage, and

$$c_{\text{th}}(N,\epsilon) = \inf\{c\colon F_C(c) = p_{\text{th}}(N,\epsilon)\}. \tag{4.11}$$

Recall that we assume F_C to be a continuous function, so the set $\{c\colon F_C(c) = p_{\text{th}}(N,\epsilon)\}$ is nonempty and thus $c_{\text{th}}(N,\epsilon) \geq 0$ is finite. An example of this procedure of parameter selection (4.5)–(4.11) (and the resulted upper bound on total expected payment) is shown in Figure 4.2.

The choices in (4.5)–(4.11) first fix the quality that the data analyst expects to obtain from each participant and the types of individuals the data analyst would like to collect data from, and the accuracy goal can be met when the population size is large enough to make sure that there are enough participants.

Theorem 8. *For the mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$, consider the Bayesian Nash equilibrium $\boldsymbol{\sigma}$ given in Theorem 7. Given an accuracy goal $p_e \leq p_e^{\max}$, let the parameter tuple $(N, c_{\text{th}}, \epsilon)$ be chosen according to* (4.5)–(4.11) *and the data analyst performs hypothesis testing using the maximum likelihood approach.*

- *The decision function ψ has the following form:*

$$\psi(\boldsymbol{X}) = \begin{cases} 1 & \text{if } \sum_i \mathbb{1}_{\{X_i=1\}} \geq \sum_i \mathbb{1}_{\{X_i=0\}}, \\ 0 & \text{otherwise;} \end{cases} \tag{4.12}$$

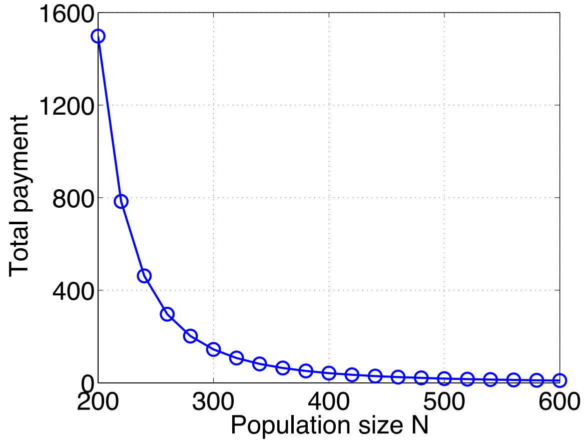


Figure 4.2: Illustration of the upper bound in Theorem 8 on the total expected payment of the designed mechanism. In this example, the quality of signals is $\theta = 0.8$, the maximum allowable probability of error is $p_e^{\max} = 0.05$, and the CDF F_C is a log-normal distribution. Pick $\epsilon = 2$, and then $D(\epsilon), n_e(\epsilon)$ and $\rho(\epsilon)$ are calculated according to (4.6)–(4.8). We consider the population size N within the range 200–600, which satisfies (4.9). As shown in the figure, tripling the population size from 200 to 600 drives the cost down by more than 99%.

- *The probability of error, p_e, meets the accuracy goal $p_e \leq p_e^{\max}$;*
- *The total expected payment is bounded as*

$$\mathbb{E}_{\boldsymbol{\sigma}}\left[\sum_{i=1}^{N} R_i^{(N,\mathcal{P},c_{\text{th}},\epsilon)}(\boldsymbol{X})\right] \leq c_{\text{th}}(\epsilon, N)\rho(\epsilon)n_e(\epsilon) \cdot (1 + e^{-\epsilon} + \epsilon). \tag{4.13}$$

Since $\rho(\epsilon)$ and $n_e(\epsilon)$ are constants for given ϵ, and $c_{\text{th}}(\epsilon, N)$ goes to 0 as $N \to \infty$, this total expected payment goes to zero, with the accuracy goal met, as $N \to \infty$.

This theorem shows that the approach of choosing parameters according to (4.5)–(4.11) for the designed family of mechanisms not only meets the accuracy goal of the data analyst but is also cost-effective. The intuition is that as N becomes large, the requirement on the participation percentage becomes lower, which allows the mechanism to collect data from individuals with lower privacy costs and thus drives down the data analyst's cost. Fix an $\epsilon \in (0, +\infty)$ and consider a sequence of mechanisms, each of which is designed for a different population size N that satisfies (4.9) and has parameter c_{th} chosen according to (4.11). Then this sequence of mechanisms gives the asymptotically optimal sequence in Theorem 6.

Figure 4.2 shows an illustration of the upper bound on the total expected payment in (4.13). In this example, tripling the population size from 200 to 600 drives the cost down by more than 99%. The rate at which the total expected payment converges to 0 depends on the distribution F_C. An interesting question is which distribution can better characterize individuals' valuations of privacy in real world. We will evaluation the convergence rates for different F_C's in our full report.

To prove Theorem 8, we need the following lemma.

Lemma 6. *For an accuracy goal* $p_e \le p_e^{\max}$*, let the parameter tuple* $(N, c_{\mathrm{th}}, \epsilon)$ *be chosen according to* (4.5)–(4.11)*. Then for any fixed* $\epsilon \in (0, +\infty)$*,*

$$\lim_{N \to +\infty} c_{\mathrm{th}}(\epsilon, N) = 0. \tag{4.14}$$

Proof. Recall that we assume that F_C is a continuous function and for any $c > 0$, $F_C(c) > 0$. For any $\delta > 0$, let $N_0 = \frac{\rho(\epsilon) n_e(\epsilon)}{F_C(\delta)}$, where $F_C(\delta) > 0$ due to our assumption. Then for any $N \ge N_0$, any c such that $F_C(c) = \frac{\rho(\epsilon) n_e(\epsilon)}{N}$ satisfies that $c \le \delta$, since a CDF is a non-decreasing function. Thus, we have $c_{\mathrm{th}}(\epsilon, N) = \inf\{c\colon F_C(c) = \frac{\rho(\epsilon) n_e(\epsilon)}{N}\} \le \delta$, which implies that $\lim_{N \to +\infty} c_{\mathrm{th}}(\epsilon, N) = 0$. □

We sketch the proof of Theorem 8 below. A complete proof is given in Appendix G.

Proof Sketch of Theorem 8. Let the parameter tuple $(N, c_{\mathrm{th}}, \epsilon)$ be chosen according to (4.5)–(4.11). Then c_{th} is a function of N and ϵ. We write $n_e, \rho, p_{\mathrm{th}}, c_{\mathrm{th}}$ to represent $n_e(\epsilon), \rho(\epsilon), p_{\mathrm{th}}(N, \epsilon), c_{\mathrm{th}}(N, \epsilon)$ and keep their dependence on N, ϵ in mind for conciseness in this proof.

Theorem 7 describes the form of $\boldsymbol{\sigma}$, which determines the distribution of each X_i given S_i and C_i. For any realization $\boldsymbol{X} = x$, since ψ uses maximum likelihood,

$$\psi(x) = \begin{cases} 1 & \text{if } \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1) \geq \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0), \\ 0 & \text{otherwise.} \end{cases} \tag{4.15}$$

The probabilities $\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1)$ and $\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0)$ can be calculated according to the form of $\boldsymbol{\sigma}$. Then the condition $\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1) \geq \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0)$ can be proved to be equivalent to the condition that the number of 1's is larger than or equal to the number of 0's in x, and the form of ψ in (4.12) is derived.

Next we calculate the probability of error, p_e. Let

$$k = \sqrt{\frac{2}{p_e^{\max}}}, \quad d = \sqrt{Np_{\text{th}}(1 - p_{\text{th}})}.$$

Then we split p_e into two parts as follows

$$\begin{aligned} p_e &= \mathbb{P}_{\boldsymbol{\sigma}}(\psi(\boldsymbol{X}) \neq W) \\ &= \mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\text{th}}\right| \geq kd, \psi(\boldsymbol{X}) \neq W\right) \\ &\quad + \mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\text{th}}\right| < kd, \psi(\boldsymbol{X}) \neq W\right). \end{aligned}$$

Since the random variables $\mathbb{1}_{\{X_i \neq \perp\}} = \mathbb{1}_{\{C_i \leq c_{\text{th}}\}}$ are i.i.d. with mean p_{th} and variance $\frac{d^2}{N}$, the first term can be bounded by Chebyshev's inequality as

$$\mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\text{th}}\right| \geq kd, \psi(\boldsymbol{X}) \neq W\right) \leq \frac{p_e^{\max}}{2}.$$

For the second term of p_e, we have

$$\begin{aligned} &\mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\text{th}}\right| < kd, \psi(\boldsymbol{X}) \neq W\right) \\ \leq\ &\mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\text{th}}\right| < kd, \psi(\boldsymbol{X}) \neq W \,\middle|\, W = 1\right) \end{aligned}$$

$$+\,\mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N}\mathbb{1}_{\{X_i\neq\perp\}}-Np_{\text{th}}\right|<kd,\psi(\boldsymbol{X})\neq W\;\middle|\;W=0\right)$$
$$=\sum_{x\in\mathcal{B}\cap\mathcal{R}_1}\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X}=x\mid W=1)+\sum_{x\in\mathcal{B}\cap\mathcal{R}_0}\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X}=x\mid W=0),$$

where

$$\mathcal{B}=\left\{x\in\mathcal{X}^N\colon\big||\mathcal{A}(x)|-Np_{\text{th}}\big|<kd\right\},$$
$$\mathcal{R}_1=\left\{x\in\mathcal{X}^N\colon\psi(x)\neq1\right\},\quad\mathcal{R}_0=\left\{x\in\mathcal{X}^N\colon\psi(x)\neq0\right\},$$

and $|\mathcal{A}(x)|$ is the number of participants, i.e., the cardinality of the set $\mathcal{A}(x)=\{i\in\mathcal{N}\colon x_i\neq\perp\}$. Then $\mathcal{B}\cap\mathcal{R}_1$ consists of the reported data such that the number of participants departs from Np_{th} by at most kd and the maximum likelihood decision rejects $W=1$. Similar explanation applies to $\mathcal{B}\cap\mathcal{R}_0$. By the choice of n_e, ρ, p_{th} and N, such number of participants is large enough to make sure that with maximum likelihood decision, the sum of the two types of error satisfies

$$\sum_{x\in\mathcal{B}\cap\mathcal{R}_1}\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X}=x\mid W=1)+\sum_{x\in\mathcal{B}\cap\mathcal{R}_0}\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X}=x\mid W=0)\leq e^{-n_eD(\epsilon)}=\frac{p_e^{\max}}{2}.$$

This gives an upper bound on the second term of p_e; i.e.,

$$\mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N}\mathbb{1}_{\{X_i\neq\perp\}}-Np_{\text{th}}\right|<kd,\psi(\boldsymbol{X})\neq W\right)\leq\frac{p_e^{\max}}{2}.$$

Therefore, $p_e\leq p_e^{\max}$.

Finally, we bound the total expected payment. Let J be the number of participants. By Corollary 1,

$$\mathbb{E}_{\boldsymbol{\sigma}}\left[\sum_{i=1}^{N}R_i^{(N,\mathcal{P},c_{\text{th}},\epsilon)}(\boldsymbol{X})\;\middle|\;J\right]\leq Jc_{\text{th}}(1+e^{-\epsilon}+\epsilon).$$

By Theorem 7, $J=\sum_{i=1}^{N}\mathbb{1}_{\{C_i\leq c_{\text{th}}\}}$. Then $\mathbb{E}_{\boldsymbol{\sigma}}[J]=Np_{\text{th}}=\rho n_e$. Therefore,

$$\mathbb{E}_{\boldsymbol{\sigma}}\left[\sum_{i=1}^{N}R_i^{(N,\mathcal{P},c_{\text{th}},\epsilon)}(\boldsymbol{X})\right]=\mathbb{E}_{\boldsymbol{\sigma}}\left[\mathbb{E}_{\boldsymbol{\sigma}}\left[\sum_{i=1}^{N}R_i^{(N,\mathcal{P},c_{\text{th}},\epsilon)}(\boldsymbol{X})\;\middle|\;J\right]\right]$$

$$\leq \mathbb{E}_{\boldsymbol{\sigma}}[J]c_{\text{th}}(1 + e^{-\epsilon} + \epsilon)$$
$$= \rho n_e c_{\text{th}}(1 + e^{-\epsilon} + \epsilon).$$

The parameters ρ and n_e do not depend on the choice of N. However, by Lemma 6, $\lim_{N\to+\infty} c_{\text{th}} = 0$. Therefore, the total expected payment goes to zero as the chosen N goes to infinity. □

4.6 Conclusions

We considered incentive mechanisms for collecting private data from strategic, privacy-aware individuals, whose valuations of privacy are unknown. The data analyst is interested in learning an underlying state from the private data of individuals with minimum overall payment. We considered a model where a data analyst is not necessarily trustworthy, and data subjects are endowed with the ability to control their own privacy, which frees the data analyst from the responsibility of privacy protection. We designed a family of payment mechanisms for the data analyst, which utilize negative payments to prevent individuals with high privacy valuations from reporting only noise and cut down the cost of the data analyst. In each designed mechanism, the individuals exhibit a threshold behavior at a Bayesian Nash equilibrium: only those with cost coefficients below some threshold participate, and they report data with certain quality guarantee, where the threshold and the quality guarantee are both parameters of the mechanism. With appropriate choices of parameters, the data analyst can fulfill any accuracy goal with diminishing cost at the equilibrium as the number of individuals grows to infinity.

Chapter 5

TRADING PRIVATE DATA WITH GENERAL PRIOR DISTRIBUTION

5.1 Introduction

We consider the following model, which is illustrated in Figure 5.1. There are N individuals and each individual i has a private bit S_i, e.g., her rating of a movie, which is either "good" or "bad" like in the rotten tomatoes website. The joint probability distribution of $S_1, S_2, \ldots, S_N$ is common knowledge. The data collector is interested in learning the proportion of 1's in the private bits, which can be viewed as the popularity of a movie. The data collector uses a payment mechanism to determine the amount of payment to each individual based on their reported data $X_1, X_2, \ldots, X_N$. When an individual i uses an ϵ-differentially private randomization algorithm to generate her reported data X_i, the privacy loss incurred is ϵ, and her cost of privacy is a function of ϵ. The form of this function is also publicly known.

We study this problem with a game-theoretic approach, where we assume the individuals are strategic and hence the quality of data an individual reports is determined by her best response that takes into account both the payment and the privacy loss. A primary goal of the data collector is to design a payment mechanism in which an individual's best response (or the Nash equilibrium of the game) has the desired level of quality. To design such a payment mechanism, we borrow ideas from the peer prediction method Miller *et al.* (2009), which makes use of the correlation among private data (which is called signals in their context) to induce truthful reporting from individuals who have no privacy concern. We should caution that different from the peer prediction method, the privacy concern of individuals in

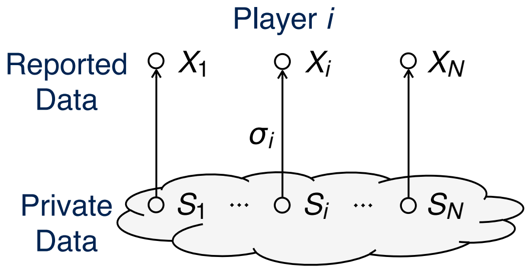


Figure 5.1: Information structure of the model with general prior distribution. Each individual i has a private bit S_i and reports X_i, which is generated based on S_i using a randomized strategy.

this study fundamentally changes the structure of the game and gives the following distinctive features to our problem. First, since the notion of differential privacy is adopted, the privacy loss of an individual i is determined by both the strategy for $S_i = 1$ and that for $S_i = 0$. Therefore, when choosing the randomization strategy, an individual needs to perform joint optimization over the two possibilities and make a contingent plan. Second, the mechanism in this dissertation is not intended to elicit truthful data reporting. The data collector is satisfied with the data quality as long as the accuracy objective can be achieved. In fact, truthful reporting may even not be preferred since it would otherwise cost the data collector unnecessary additional payments. Consequently, when we build this study upon the peer prediction method, the prediction should be made on the randomized data instead of the original data.

Taking these features into consideration, we design a payment mechanism in which the randomized response strategy Warner (1965) that generates the reported data by flipping the private bit with probability $\frac{1}{e^{\epsilon}+1}$, where $\epsilon > 0$, proves to be an equilibrium. This equilibrium strategy is ϵ-differentially private, so the collected data itself is privacy preserving. By adjusting the corresponding parameter in the mechanism, the data collector can control the privacy level ϵ and thus control the data quality to achieve any given accuracy objective. In contrast to most of the existing work, which considers a trusted data collector and thus focuses on designing truthful mechanisms, our designed mechanism addresses individuals' privacy concern where the data collector may not be trusted, and is the first one that considers quality control in such a scenario to suit the principle's accuracy objective.

5.2 Model

In this chapter, the model for private data is more general than Chapter 3. We still let S_i denote player i's private bit, and let $\boldsymbol{S} = (S_1, S_2, \cdots, S_N)$. The joint probability distribution of $S_1, S_2, \cdots, S_N$ is common knowledge. We assume that this distribution is symmetric over players; i.e., for any binary sequence $\boldsymbol{s} \in \{0,1\}^N$ and any of its permutations $\boldsymbol{s}'$, $\mathbb{P}(\boldsymbol{S} = \boldsymbol{s}) = \mathbb{P}(\boldsymbol{S} = \boldsymbol{s}')$. Other notation is the same as that in Chapter 3.

The data collector is interested in learning the proportion of 1's in $S_1, S_2, \ldots, S_N$, i.e., $\bar{S} = \frac{1}{N}\sum_{i=1}^{N} S_i$. Let $\hat{\mu}$ be an estimate of $\bar{S}$ from the reported data $X_1, X_2, \ldots, X_N$. Then we measure the accuracy of $\hat{\mu}$ by the following definition, which has been used in the literature (e.g., Ghosh and Roth (2011), where a fixed number $\frac{1}{3}$ is used instead of δ).

Definition 8. An estimate $\hat{\mu}$ of $\bar{S}$ is (α, δ)-*accurate* if $|\bar{S} - \hat{\mu}| \leq \alpha$ holds with probability at least $1 - \delta$.

5.3 A Payment Mechanism for Quality Control

We wish to design mechanisms such that the quality of the collected data in equilibrium is controllable. Then the data collector can achieve her accuracy objective by adjusting parameters in the mechanism. In this section, we present our design of the payment mechanism. Consider the following payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$ for collecting privacy-preserving data from N players, parameterized by a data quality parameter ϵ, where $N \geq 2$ and $\epsilon > 0$.

The payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$

1. Each player reports her data (which can also be the decision of not participating).
2. For non-participating players, the payment is zero.
3. If there is only one participant, pay zero to this participant. Otherwise, for each participating player i, arbitrarily choose another participating player j and pay player i according to X_i and X_j as follows:
$$R_i^{(N,\epsilon)}(X) = \frac{g'(\epsilon)(e^\epsilon + 1)^2}{2e^\epsilon} A_{X_i,X_j}, \tag{5.1}$$
where parameters $A_{1,1}, A_{0,0}, A_{0,1}, A_{1,0}$ are calculated in the next section.

After the collection of data, the data collector estimates $\bar{S} = \frac{1}{N}\sum_{i=1}^{N} S_i$ by

$$\hat{\mu} = \frac{e^\epsilon + 1}{e^\epsilon - 1}\left(\frac{1}{n}\sum_{i\colon X_i \neq \perp} X_i\right) - \frac{1}{e^\epsilon - 1}, \tag{5.2}$$

where n is the number of participants.

5.3.1 Payment Parameterization

Recall that we assume that the joint distribution of $S_1, S_2, \ldots, S_N$ is symmetric over players. As a consequence, the private bits of the players have the same marginal distribution. Denote this marginal distribution as follows:

$$P_1 = \mathbb{P}(S_i = 1), \quad P_0 = \mathbb{P}(S_i = 0). \tag{5.3}$$

Due to symmetry, the marginal distribution of any two private bits S_i and S_j with $i \neq j$ does not depend on the specific identities i and j either. Denote the marginal distribution of S_i and S_j with $i \neq j$ as follows:

$$\begin{aligned} &P_{1,1} = \mathbb{P}(S_i = 1, S_j = 1), \quad P_{0,0} = \mathbb{P}(S_i = 0, S_j = 0), \\ &P_{0,1} = \mathbb{P}(S_i = 0, S_j = 1) = P_{1,0} = \mathbb{P}(S_i = 1, S_j = 0). \end{aligned} \tag{5.4}$$

We further define a constant D as follows:

$$\begin{aligned} D &= \mathbb{P}(S_j = 1, S_i = 1)\mathbb{P}(S_j = 0, S_i = 0) \\ &\quad - \mathbb{P}(S_j = 0, S_i = 1)\mathbb{P}(S_j = 1, S_i = 0) \\ &= P_{1,1}P_{0,0} - P_{0,1}P_{1,0}, \end{aligned} \tag{5.5}$$

which can be verified to equal to the covariance of S_i and S_j. We assume that $D \neq 0$, which is equivalent to the case that S_i and S_j are not independent for any two distinct players i and j (See Appendix H for the proof of the equivalence).

The parameters $A_{1,1}, A_{0,0}, A_{0,1}, A_{1,0}$ used in the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$ are defined as follows:

- If $D > 0$,

$$A_{1,1} = \frac{(e^\epsilon + 1)^2}{e^{2\epsilon} - 1} \frac{1}{D} \left(\frac{1}{e^\epsilon + 1} P_1 + \frac{e^\epsilon}{e^\epsilon + 1} P_0 \right), \tag{5.6}$$

$$A_{0,0} = \frac{(e^\epsilon + 1)^2}{e^{2\epsilon} - 1} \frac{1}{D} \left(\frac{e^\epsilon}{e^\epsilon + 1} P_1 + \frac{1}{e^\epsilon + 1} P_0 \right), \tag{5.7}$$

$$A_{0,1} = 0, \tag{5.8}$$

$$A_{1,0} = 0. \tag{5.9}$$

- If $D < 0$,

$$A_{1,1} = 0, \tag{5.10}$$

$$A_{0,0} = 0, \tag{5.11}$$

$$A_{0,1} = -\frac{(e^\epsilon + 1)^2}{e^{2\epsilon} - 1} \frac{1}{D} \left(\frac{1}{e^\epsilon + 1} P_1 + \frac{e^\epsilon}{e^\epsilon + 1} P_0 \right), \tag{5.12}$$

$$A_{1,0} = -\frac{(e^\epsilon + 1)^2}{e^{2\epsilon} - 1} \frac{1}{D} \left(\frac{e^\epsilon}{e^\epsilon + 1} P_1 + \frac{1}{e^\epsilon + 1} P_0 \right). \tag{5.13}$$

From the above definition of these parameters we can see the intuition behind the design of mechanism $\boldsymbol{R}^{(N,\epsilon)}$. When the private bits of two players are positively correlated ($D > 0$), they tend to be the same. Thus, the mechanism rewards agreement on the reported data to encourage informative data reporting. Similarly, when the private bits of two players are negatively correlated ($D < 0$), they tend to be different, and thus correspondingly, the mechanism rewards disagreement to encourage informative data reporting. However, the more informative the reported data is, the more privacy cost a player will experience. This tension will make each player choose a compromise, which is telling truth to some extent.

5.3.2 Nash Equilibrium

Theorem 9. *The strategy profile, consisting of the following strategy of player i that is denoted by σ_i^*, is a Nash equilibrium under the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$:*

$$\begin{aligned}
&\mathbb{P}_{\sigma_i^*}(X_i = 1 \mid S_i = 1) = \mathbb{P}_{\sigma_i^*}(X_i = 0 \mid S_i = 0) = \frac{e^\epsilon}{e^\epsilon + 1},\\
&\mathbb{P}_{\sigma_i^*}(X_i = 0 \mid S_i = 1) = \mathbb{P}_{\sigma_i^*}(X_i = 1 \mid S_i = 0) = \frac{1}{e^\epsilon + 1},\\
&\mathbb{P}_{\sigma_i^*}(X_i = \perp \mid S_i = 1) = \mathbb{P}_{\sigma_i^*}(X_i = \perp \mid S_i = 0) = 0,
\end{aligned} \tag{5.14}$$

i.e., each player generates her reported data by flipping the private bit with probability $\frac{1}{e^\epsilon+1}$.

Proof. See Appendix I. □

By Theorem 9, the parameter ϵ of the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$ plays two roles in the equilibrium σ^*. On one hand, the strategy each player uses to randomize her data is ϵ-differentially private. Therefore, the parameter ϵ controls how much privacy each player is willing to trade for payment. On the other hand, the parameter ϵ describes the quality of the reported data of each player i, since ϵ controls the probability that the reported data is the same as the true private data as follows:

$$\mathbb{P}_{\sigma_i^*}(X_i = S_i) = \frac{e^\epsilon}{e^\epsilon + 1}. \tag{5.15}$$

Therefore, the larger ϵ is, the more privacy each player is willing to sell, and the higher data quality the data collector obtains. With the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$, the data collector is not only able to know how the data has been randomized, but also able to control the quality of the collected data.

5.3.3 Estimation Accuracy

In this section, we discuss how the data collector should choose the parameter ϵ to achieve the accuracy objective of estimating $\bar{S}$.

Theorem 10. *For any α, δ with $\alpha > 0$ and $0 < \delta < 1$, if*

$$\epsilon \geq \ln\left(2 + \frac{1}{N\alpha^2\delta}\right), \tag{5.16}$$

then in the equilibrium σ^* of the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$, the estimate $\hat{\mu}$ given in (5.2) is (α, δ)-accurate.

Proof. See Appendix J. □

Since the parameter ϵ of the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$ describes the quality of the collected data in the equilibrium σ^*, intuitively, the data collector can achieve higher accuracy objective by increasing ϵ. Theorem 10 confirms this intuition. For an accuracy objective (α, δ), the smaller α and δ are, the higher accuracy is required to achieve according to the definition of accuracy in Definition 8. However, no matter how high the accuracy objective is, by Theorem 10, the data collector can always achieve it by choosing large enough ϵ, i.e., good enough data quality.

5.3.4 *Asymptotic Optimality in the High Quality Regime*

From the principal's perspective, the strategy profile σ^* given in Theorem 9 is very attractive. When players follow σ^*, the quality of the collected data can be controlled by a single parameter ϵ, and $\bar{S}$ can be estimated by the simple estimator $\hat{\mu}$. In this section, we focus on nonnegative payment mechanisms in which σ^* forms a Nash equilibrium. We study the optimality of the proposed mechanism in terms of the total expected payment needed to collect data with a given quality level ϵ. We first derive an lower bound on the total expected payment of a nonnegative payment mechanism in which σ^* is an equilibrium. Then we compare the expected payment of the proposed mechanism with this lower bound and show that the proposed mechanism is asymptotically optimal in the high quality regime, i.e., as ϵ goes to infinity.

Proposition 1. *For any nonnegative payment mechanism R in which σ^* is a Nash equilibrium, the total expected payment at σ^* is lower bounded, given as follows:*

$$\mathbb{E}_{\sigma^*}\left[\sum_{i=1}^{N} R_i(X)\right] \geq N g'(\epsilon)(e^{\epsilon}+1). \tag{5.17}$$

Proof. See Appendix K. □

Therefore, to have an equilibrium at σ^*, a nonnegative payment mechanism needs to pay at least $Ng'(\epsilon)(e^{\epsilon}+1)$ to the players. In the asymptotic regime that ϵ goes to infinity, this lower bound is on the order of $\mathcal{O}(g'(\epsilon)e^{\epsilon})$. In the equilibrium σ^* of the payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$, the total expected payment is given by

$$\begin{aligned}&\mathbb{E}_{\sigma^*}\left[\sum_{i=1}^{N} R_i^{(N,\epsilon)}(X)\right]\\ &= Ng'(\epsilon)(e^{\epsilon}+1) && (5.18)\\ &\quad + \frac{Ng'(\epsilon)(e^{\epsilon}+1)^2}{2e^{\epsilon}}\frac{(e^{\epsilon}+1)^2}{e^{2\epsilon}-1}\frac{1}{|D|} && (5.19)\end{aligned}$$

$$\cdot \left(\frac{e^{2\epsilon}}{(e^\epsilon+1)^2} P_{0,1} + \frac{e^\epsilon}{(e^\epsilon+1)^2}(P_1^2 + P_0^2) \right. \tag{5.20}$$

$$\left. + \frac{1}{(e^\epsilon+1)^2}(P_1 P_{1,1} + P_0 P_{0,0}) \right), \tag{5.21}$$

which can be obtained from the proof of Theorem 9.

In the asymptotic regime that ϵ goes to infinity, the total expected payment of mechanism $\boldsymbol{R}^{(N,\epsilon)}$ is dominated by the first term, which is identical to the lower bound $Ng'(\epsilon)(e^\epsilon+1)$, so the mechanism is asymptotically optimal in the high-quality regime.

5.4 Conclusions

In this chapter we showed how to design the payment mechanism to achieve quality control when collecting data from privacy-sensitive individuals. We considered a model in which individuals do not trust the data collector and take into account a privacy cost that depends on the level of the (local) differential privacy of the data reporting strategy. Due to privacy concerns, an individual may be only willing to report a noisy version of the private data, which degrades the quality of the collected data. Our proposed mechanism incentives individuals to use a randomized response strategy with a desired noise level in the Nash equilibrium. This strategy generates the reported data by flipping the private data with probability $\frac{1}{e^\epsilon+1}$, where $\epsilon > 0$ is a parameter of the mechanism. Therefore, the quality of the collected data is controllable by adjusting ϵ. With properly selected parameters, any accuracy goal can be fulfilled at the Nash equilibrium, and the total expected payment of the designed mechanism is asymptotically optimal in the high quality regime. Note that the model of the private data in this work is a very general one. Considering some specific but well motivated structure for the model of the private data, such as the model we considered in Chapter 3 and 4, to find better mechanisms is an exciting direction for future work.

Chapter 6

CONCLUSIONS

In this dissertation, data privacy was studied from two perspectives: the relation between different privacy notions (Chapter 2) and the economic foundations for a market model of trading private data (Chapter 3–Chapter 5).

Chapter 2 investigated the relation between three different notions of privacy: identifiability, differential privacy and mutual-information privacy, where identifiability guarantees indistinguishability between posterior probabilities, differential privacy guarantees limited additional disclosures, and mutual information is an information-theoretic notion. Under a unified privacy–distortion framework, where the distortion is defined to be the expected Hamming distance between the input and output databases, we established some fundamental connections between these three privacy notions. Given a maximum allowable distortion D within certain range, the smallest identifiability level $\epsilon_{\mathrm{i}}^*(D)$ and the smallest differential privacy level $\epsilon_{\mathrm{d}}^*(D)$ are proved to satisfy $\epsilon_{\mathrm{i}}^*(D) - \epsilon_X \leq \epsilon_{\mathrm{d}}^*(D) \leq \epsilon_{\mathrm{i}}^*(D)$, where ϵ_X is a constant determined by the prior of the original database, and diminishes to zero when the prior is uniform. Next, we showed that there is a mechanism that simultaneously minimizes the identifiability level and the mutual information given the same maximum allowable distortion within certain range. We further showed that this mechanism satisfies ϵ-differential privacy with $\epsilon_{\mathrm{d}}^*(D) \leq \epsilon \leq \epsilon_{\mathrm{d}}^*(D) + 2\epsilon_X$.

Our findings in this part reveal some fundamental connections between the three notions of privacy. With these three notions of privacy being defined, many interesting issues deserve further attention. The connections we have established in this work are based on the distortion measure of Hamming distance, which is closely tied with the neighboring relations, and we assume that the output synthetic database and the original database are in the same universe. It would be of great interest to study the connections of these privacy notions under other common distortion measures and other output formats. We remark that our results for Hamming distance can be used to prove lower bounds on the distortion of a differentially private mechanism when the distortion is measured by the distortion at the worst-case query in a query class Wang *et al.* (2015b). Some other interesting directions are as follows. In some cases, the prior p_X is imperfect. Then for privacy notions depending on the prior such as identifiability and mutual-information privacy, it is natural to ask how we can protect privacy with robustness over the prior distribution. Identifiability and differential privacy impose requirements on neighboring databases to protect an individual's privacy. Then are there any practical scenarios that we would desire to generalize this "pairwise" privacy to "group" privacy? The connections between membership privacy and these three notions of privacy also need to be explored, since membership privacy has been proposed as a unifying framework for privacy definitions.

Starting from Chapter 3, we studied a market for trading private data, where a data collector purchases private data from strategic data subjects (individuals) through an incentive mechanism. The data subjects do not consider the data collector to be trustworthy, and thus experience a cost incurred by the privacy loss during data reporting. The data subjects are endowed with the ability to control their own privacy, which also frees the data collector from the responsibility of privacy protection.

Chapter 3 studied "the value of privacy" under a setting where the private data of individuals is binary data and represents their knowledge about a common underlying state. The value of ϵ units of privacy is measured by the minimum payment of all nonnegative payment mechanisms under which an individual's best response in a Nash equilibrium is to report the data with a privacy level of ϵ. We derived asymptotically tight lower and upper bounds on the value of privacy as the number of individuals becomes large, where the upper bound was given by a designed payment mechanism $\boldsymbol{R}^{(N,\epsilon)}$. We further applied these fundamental limits to find the minimum total payment for the data collector to achieve certain accuracy target for learning the underlying state, and derived lower and upper bounds on the minimum payment. The total payment of the designed mechanism $\boldsymbol{R}^{(N,\epsilon)}$ with properly chosen parameters is at most one individual's payment away from the minimum.

Chapter 4 considered a setting where the individuals's valuations of privacy are unknown to the data collector/analyst. The data analyst is interested in learning the underlying state from the reported data with minimum overall payment. We designed a family of payment mechanisms for the data analyst, which utilize negative payments to prevent individuals with high privacy valuations from reporting only noise and cut down the cost of the data analyst. In each designed mechanism, the individuals exhibit a threshold behavior at a Bayesian Nash equilibrium: only those with cost coefficients below some threshold participate, and they report data with certain quality guarantee, where the threshold and the quality guarantee are both parameters of the mechanism. With appropriate choices of parameters, the data analyst can fulfill any accuracy goal with diminishing cost at the equilibrium as the number of individuals grows to infinity.

Chapter 5 showed how to design the payment mechanism to achieve quality control when individuals' binary private data follows a general joint probability distribution with some symmetry. The data collector is interested in learning the average of the private data. Our proposed mechanism incentives individuals to use a randomized response strategy with a desired noise level in the Nash equilibrium. This strategy generates the reported data by flipping the private data with probability $\frac{1}{e^{\epsilon}+1}$, where $\epsilon > 0$ is a parameter of the mechanism. Therefore, the quality of the collected data is controllable by adjusting ϵ. With properly selected parameters, any accuracy goal can be fulfilled at the Nash equilibrium, and the total expected payment of the designed mechanism is asymptotically optimal in the high quality regime.

Our study from the economic perspective suggested a market-based model to address the privacy concerns in data collection for big data analytics. Under this market model, many interesting directions can be further explored. We just list a few of them below. We can consider more general but still structured models for private data. We have considered the structure where the private data is binary and is correlated through a common underlying state, which is also binary. We have also considered a general distribution for the private data where the distribution is symmetric over individuals. To broaden the range of applications, it would be of great interest to study models with larger alphabets for the private data and the underlying state, or with a more complicated correlation structure among the private data. The learning goal of the data collector should also be set accordingly. For example, people's opinions can have impacts on each other and form certain dynamics (see, e.g., Acemoglu and Ozdaglar (2010)) in reality. Models that capture the structure of the dynamics would be very appealing. We can also consider a market where there are multiple rounds of interactions between the data collector and the data subjects. For example, in a crowdsourcing scenario, a worker (data subject) may be interested in participating in multiple tasks released by the same crowdsourcer (data collector). The multi-round setting provides an opportunity of learning for the data collector. Some characteristics of the data subjects may be learnable during the process and provide useful information. The data collector may adjust the design of the payment mechanism after seeing the results from previous rounds. Another direction worth further investigations goes back to the privacy notions. We have studied the relations between three different privacy notions: identifiability, differential privacy and mutual-information privacy. We used local differential privacy as the privacy measure in the market model. However, other privacy notions also deserve comprehensive investigations in a market model. Using different notions of privacy will change the structure of incentives, thereby resulting in new fundamental tradeoffs. To tackle different privacy notions in a market model, we may leverage the relations we have established. This is by no means a complete list of the problems worth studying. Incorporating privacy protection into big data analytics is a complicated problem that needs persistent efforts from various aspects.

REFERENCES

Acemoglu, D., M. A. Dahleh, I. Lobel and A. Ozdaglar, "Bayesian learning in social networks", Review of Econ. Stud. **78**, 4, 1201–1236 (2011).

Acemoglu, D. and A. Ozdaglar, "Opinion dynamics and learning in social networks", Dyn. Games and Applicat. **1**, 1, 3–49 (2010).

Agrawal, D. and C. C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms", in "Symp. Principles Database Systems (PODS)", pp. 247–255 (Santa Barbara, CA, 2001).

Alvim, M. S., M. E. Andrés, K. Chatzikokolakis, P. Degano and C. Palamidessi, "Differential privacy: On the trade-off between utility and information leakage", in "Formal Aspects of Security and Trust", vol. 7140 of *Lecture Notes in Comput. Sci.*, pp. 39–54 (2012).

Bassily, R. and A. Smith, "Local, private, efficient protocols for succinct histograms", in "Proc. Ann. ACM Symp. Theory of Computing (STOC)", pp. 127–135 (Portland, OR, 2015).

Blahut, R. E., "Computation of channel capacity and rate-distortion functions", IEEE Trans. Inf. Theory **18**, 460–473 (1972).

Blum, A., K. Ligett and A. Roth, "A learning theory approach to non-interactive database privacy", in "Proc. Ann. ACM Symp. Theory of Computing (STOC)", pp. 609–618 (Victoria, Canada, 2008).

Boyd, S. and L. Vandenberghe, *Convex Optimization* (Cambridge Univ. Press, New York, NY, 2004).

Bun, M., J. Ullman and S. Vadhan, "Fingerprinting codes and the price of approximate differential privacy", in "Proc. Ann. ACM Symp. Theory of Computing (STOC)", pp. 1–10 (New York, NY, 2014).

Chatzikokolakis, K., T. Chothia and A. Guha, "Statistical measurement of information leakage", in "Proc. Int. Conf. Tools and Algorithms for the Construction and Analysis of Systems (TACAS)", pp. 390–404 (Paphos, Cyprus, 2010).

Chatzikokolakis, K., C. Palamidessi and P. Panangaden, "Anonymity protocols as noisy channels", in "Proc. Int. Conf. Trustworthy Global Computing (TGC)", pp. 281–300 (Lucca, Italy, 2007).

Chen, Y., S. Chong, I. A. Kash, T. Moran and S. Vadhan, "Truthful mechanisms for agents that value privacy", in "Proc. ACM Conf. Electronic Commerce (EC)", pp. 215–232 (Philadelphia, PA, 2013).

Chen, Y., O. Sheffet and S. Vadhan, "Privacy games", in "Int. Conf. Web and Internet Economics (WINE)", vol. 8877, pp. 371–385 (2014).

Clark, D., S. Hunt and P. Malacaria, "Quantitative information flow, relations and polymorphic types", J. Log. Comput. **15**, 2, 181–199 (2005).

Cover, T. M. and J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, Hoboken, NJ, 2006), 2nd edn.

De, A., "Lower bounds in differential privacy", in "Theory of Cryptography", vol. 7194 of *Lecture Notes in Comput. Sci.*, pp. 321–338 (2012).

du Pin Calmon, F. and N. Fawaz, "Privacy against statistical inference", in "Proc. Ann. Allerton Conf. Communication, Control and Computing", pp. 1401–1408 (Monticello, IL, 2012).

Duchi, J. C., M. I. Jordan and M. J. Wainwright, "Local privacy and minimax bounds: Sharp rates for probability estimation", in "Advances Neural Information Processing Systems (NIPS)", pp. 1529–1537 (Lake Tahoe, NV, 2013).

Dwork, C., "Differential privacy", in "Proc. Int. Conf. Automata, Languages and Programming (ICALP)", pp. 1–12 (Venice, Italy, 2006).

Dwork, C., K. Kenthapadi, F. McSherry, I. Mironov and M. Naor, "Our data, ourselves: privacy via distributed noise generation", in "Proc. Annu. Int. Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT)", pp. 486–503 (St. Petersburg, Russia, 2006a).

Dwork, C., F. McSherry, K. Nissim and A. Smith, "Calibrating noise to sensitivity in private data analysis", in "Proc. Conf. Theory of Cryptography (TCC)", pp. 265–284 (New York, NY, 2006b).

Dwork, C., M. Naor, O. Reingold, G. N. Rothblum and S. Vadhan, "On the complexity of differentially private data release: efficient algorithms and hardness results", in "Proc. Ann. ACM Symp. Theory of Computing (STOC)", pp. 381–390 (Bethesda, MD, 2009).

Dwork, C. and A. Roth, "The algorithmic foundations of differential privacy", Found. Trends Theor. Comput. Sci. **9**, 3–4, 211–407 (2014).

Erlingsson, Ú., V. Pihur and A. Korolova, "RAPPOR: Randomized aggregatable privacy-preserving ordinal response", in "Proc. ACM SIGSAC Conf. Computer and Communication Security (CCS)", pp. 1054–1067 (Scottsdale, AZ, 2014).

Fanti, G. C., V. Pihur and Ú. Erlingsson, "Building a RAPPOR with the unknown: Privacy-preserving learning of associations and data dictionaries", arXiv:1503.01214 [cs.CR] (2015).

Fleischer, L. K. and Y. Lyu, "Approximately optimal auctions for selling privacy when costs are correlated with data", in "Proc. ACM Conf. Electronic Commerce (EC)", pp. 568–585 (Valencia, Spain, 2012).

Ghosh, A. and K. Ligett, "Privacy and coordination: Computing on databases with endogenous participation", in "Proc. ACM Conf. Electronic Commerce (EC)", pp. 543–560 (Philadelphia, PA, 2013).

Ghosh, A., K. Ligett, A. Roth and G. Schoenebeck, "Buying private data without verification", in "Proc. ACM Conf. Economics and Computation (EC)", pp. 931–948 (Palo Alto, CA, 2014).

Ghosh, A. and A. Roth, "Selling privacy at auction", in "Proc. ACM Conf. Electronic Commerce (EC)", pp. 199–208 (San Jose, CA, 2011).

Ghosh, A., T. Roughgarden and M. Sundararajan, "Universally utility-maximizing privacy mechanisms", in "Proc. Ann. ACM Symp. Theory of Computing (STOC)", pp. 351–360 (Bethesda, MD, 2009).

Gupta, A., A. Roth and J. Ullman, "Iterative constructions and private data release", in "Proc. Conf. Theory of Cryptography (TCC)", pp. 339–356 (Sicily, Italy, 2012).

Hardt, M., K. Ligett and F. McSherry, "A simple and practical algorithm for differentially private data release", in "Advances Neural Information Processing Systems (NIPS)", pp. 2348–2356 (Lake Tahoe, NV, 2012).

Hardt, M. and G. N. Rothblum, "A multiplicative weights mechanism for privacy-preserving data analysis", in "Proc. Ann. IEEE Symp. Found. Comput. Sci. (FOCS)", pp. 61–70 (Las Vegas, NV, 2010).

Hsu, J., S. Khanna and A. Roth, "Distributed private heavy hitters", in "Proc. Int. Conf. Automata, Languages and Programming (ICALP)", pp. 461–472 (Warwick, UK, 2012).

Kailath, T., "The divergence and Bhattacharyya distance measures in signal selection", IEEE Trans. Commun. Technol. **15**, 1, 52–60 (1967).

Kairouz, P., S. Oh and P. Viswanath, "Extremal mechanisms for local differential privacy", in "Advances Neural Information Processing Systems (NIPS)", pp. 2879–2887 (Montreal, Canada, 2014).

Kasiviswanathan, S., M. Rudelson, A. Smith and J. Ullman, "The price of privately releasing contingency tables and the spectra of random matrices with correlated rows", in "Proc. Ann. ACM Symp. Theory of Computing (STOC)", pp. 775–784 (Cambridge, MA, 2010).

Kasiviswanathan, S. P., H. K. Lee, K. Nissim, S. Raskhodnikova and A. Smith, "What can we learn privately?", SIAM J. Comput. **40**, 3, 793–826 (2011).

Kroft, S., "The data brokers: selling your personal information", CBS News (2014).

Le, T. N., V. G. Subramanian and R. A. Berry, "The value of noise for informational cascades", in "Proc. IEEE Int. Symp. Information Theory (ISIT)", pp. 1101–1105 (Honolulu, HI, 2014).

Lee, J. and C. Clifton, "Differential identifiability", in "Proc. Ann. ACM SIGKDD Conf. Knowledge Discovery and Data Mining (KDD)", pp. 1041–1049 (Beijing, China, 2012).

Li, N., W. Qardaji, D. Su, Y. Wu and W. Yang, "Membership privacy: A unifying framework for privacy definitions", in "Proc. ACM SIGSAC Conf. Computer and Communication Security (CCS)", pp. 889–900 (Berlin, Germany, 2013).

Ligett, K. and A. Roth, "Take it or leave it: Running a survey when privacy comes at a cost", in "Proc. Int. Workshop Internet and Network Economics (WINE)", pp. 378–391 (Liverpool, UK, 2012).

Makhdoumi, A. and N. Fawaz, "Privacy-utility tradeoff under statistical uncertainty", in "Proc. Ann. Allerton Conf. Communication, Control and Computing", pp. 1627–1634 (Monticello, IL, 2013).

McGregor, A., I. Mironov, T. Pitassi, O. Reingold, K. Talwar and S. Vadhan, "The limits of two-party differential privacy", in "Proc. Ann. IEEE Symp. Found. Comput. Sci. (FOCS)", pp. 81–90 (Las Vegas, NV, 2010).

McSherry, F. and K. Talwar, "Mechanism design via differential privacy", in "Proc. Ann. IEEE Symp. Found. Comput. Sci. (FOCS)", pp. 94–103 (Providence, RI, 2007).

Miller, N., P. Resnick and R. Zeckhauser, "Eliciting informative feedback: The peer-prediction method", in "Computing with Social Trust", Human–Computer Interaction Series, pp. 185–212 (Springer London, 2009).

Mir, D. J., "Information-theoretic foundations of differential privacy", in "Found. and Practice of Security", vol. 7743 of *Lecture Notes in Comput. Sci.*, pp. 374–381 (2013).

Muthukrishnan, S. and A. Nikolov, "Optimal private halfspace counting via discrepancy", in "Proc. Ann. ACM Symp. Theory of Computing (STOC)", pp. 1285–1292 (New York, NY, 2012).

Nissim, K., S. Raskhodnikova and A. Smith, "Smooth sensitivity and sampling in private data analysis", in "Proc. Ann. ACM Symp. Theory of Computing (STOC)", pp. 75–84 (San Diego, CA, 2007).

Nissim, K., S. Vadhan and D. Xiao, "Redrawing the boundaries on purchasing data from privacy-sensitive individuals", in "Proc. Conf. Innovations in Theoretical Computer Science (ITCS)", pp. 411–422 (Princeton, NJ, 2014).

Pai, M. M. and A. Roth, "Privacy and mechanism design", SIGecom Exch. **12**, 1, 8–29 (2013).

Rebollo-Monedero, D., J. Forné and J. Domingo-Ferrer, "From t-closeness-like privacy to postrandomization via information theory", IEEE Trans. Knowl. Data Eng. **22**, 11, 1623–1636 (2010).

Roth, A. and T. Roughgarden, “Interactive privacy via the median mechanism”, in “Proc. Ann. ACM Symp. Theory of Computing (STOC)”, pp. 765–774 (Cambridge, MA, 2010).

Roth, A. and G. Schoenebeck, “Conducting truthful surveys, cheaply”, in “Proc. ACM Conf. Electronic Commerce (EC)”, pp. 826–843 (Valencia, Spain, 2012).

Sankar, L., S. R. Rajagopalan and H. V. Poor, “Utility-privacy tradeoffs in databases: An information-theoretic approach”, IEEE Trans. Inf. Forens. Security **8**, 6, 838–852 (2013).

Sarwate, A. D. and L. Sankar, “A rate-disortion perspective on local differential privacy”, in “Proc. Ann. Allerton Conf. Communication, Control and Computing”, pp. 903–908 (Monticello, IL, 2014).

Shokri, R., “Privacy games: Optimal user-centric data obfuscation”, in “Proc. Privacy Enhancing Technologies (PETS)”, pp. 299–315 (Philadelphia, PA, 2015).

Smith, G., “On the foundations of quantitative information flow”, in “Proc. Int. Conf. Foundations of Software Science and Computational Structures (FSSACS)”, pp. 288–302 (York, UK, 2009).

Srikant, R. and L. Ying, *Communication Networks: An Optimization, Control and Stochastic Networks Perspective* (Cambridge Univ. Press, New York, 2014).

Ullman, J. and S. Vadhan, “PCPs and the hardness of generating private synthetic data”, in “Proc. Conf. Theory of Cryptography (TCC)”, pp. 400–416 (Providence, RI, 2011).

Wang, W., L. Ying and J. Zhang, “On the relation between identifiability, differential privacy, and mutual-information privacy”, in “Proc. Ann. Allerton Conf. Communication, Control and Computing”, pp. 1086–1092 (Monticello, IL, 2014).

Wang, W., L. Ying and J. Zhang, “A game-theoretic approach to quality control for collecting privacy-preserving data”, in “Proc. Ann. Allerton Conf. Communication, Control and Computing”, pp. 474–479 (Monticello, IL, 2015a).

Wang, W., L. Ying and J. Zhang, “A minimax distortion view of differentially private query release”, in “Proc. Asilomar Conf. Signals, Systems, and Computers”, pp. 1046–1050 (Pacific Grove, CA, 2015b).

Wang, W., L. Ying and J. Zhang, “The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits”, in “Proc. Ann. ACM SIGMETRICS Conf.”, (Antibes Juan-les-Pins, France, 2016).

Warner, S. L., “Randomized response: A survey technique for eliminating evasive answer bias”, J. Amer. Stat. Assoc. **60**, 309, 63–69 (1965).

Xiao, D., “Is privacy compatible with truthfulness?”, in “Proc. Conf. Innovations in Theoretical Computer Science (ITCS)”, pp. 67–86 (Berkeley, CA, 2013).

Zhu, Y. and R. Bettati, “Anonymity vs. information leakage in anonymity systems”, in “Proc. IEEE Int. Conf. Distributed Computing Systems (ICDCS)”, pp. 514–524 (Columbus, OH, 2005).

APPENDIX A

PROOF OF THEOREM 1

Lemma 7. *The minimum distortion $D^*_{\text{relaxed}}(\epsilon)$ of the relaxed optimization problem R-PD satisfies*

$$D^*_{\text{relaxed}}(\epsilon) = h(\epsilon), \tag{A.1}$$

where

$$h(\epsilon) = \frac{n}{1 + \frac{e^\epsilon}{m-1}}.$$

Proof. We first prove the following claim, which gives a lower bound on the minimum distortion $D^*_{\text{relaxed}}(\epsilon)$.

Claim. *Any feasible solution $\{p_{X|Y}(x \mid y), x, y \in \mathcal{D}^n\}$ of R-PD satisfies*

$$\sum_{x \in \mathcal{D}^n} p_{X|Y}(x \mid y) d(x, y) \geq h(\epsilon).$$

Proof of the Claim. Consider any feasible $\{p_{X|Y}(x \mid y), x, y \in \mathcal{D}^n\}$. For any $y \in \mathcal{D}^n$ and any integer l with $0 \leq l \leq n$, let $\mathcal{N}_l(y)$ be the set of elements with distance l to y, i.e.,

$$\mathcal{N}_l(y) = \{v \in \mathcal{D}^n \colon d(v, y) = l\}. \tag{A.2}$$

Denote $P_l = \mathbb{P}\{X \in \mathcal{N}_l(y) \mid Y = y\}$. Then

$$\sum_{x \in \mathcal{D}^n} p_{X|Y}(x \mid y) d(x, y) = \sum_{l=0}^{n} l P_l.$$

We first derive a lower bound on P_n. For any $u \in \mathcal{N}_{l-1}(y)$, $\mathcal{N}_1(u) \cap \mathcal{N}_l(y)$ consists of the neighbors of u that are in $\mathcal{N}_l(y)$. By the constraint (2.14), for any $v \in \mathcal{N}_1(u) \cap \mathcal{N}_l(y)$,

$$p_{X|Y}(u \mid y) \leq e^\epsilon p_{X|Y}(v \mid y). \tag{A.3}$$

Each $u \in \mathcal{N}_{l-1}(y)$ has $n - (l-1)$ rows that are the same with the corresponding rows of y. Each neighbor of u in $\mathcal{N}_l(y)$ can be obtained by changing one of these $n - (l-1)$ rows to a different element in $\mathcal{D}$, which is left with $m-1$ choices. Therefore, each $u \in \mathcal{N}_{l-1}(y)$ has $(n-l+1)(m-1)$ neighbors in $\mathcal{N}_l(y)$. By similar arguments, each $v \in \mathcal{N}_l(y)$ has l neighbors in $\mathcal{N}_{l-1}(y)$. Taking summation of (A.3) over $u \in \mathcal{N}_{l-1}(y), v \in \mathcal{N}_l(y)$ with $u \sim v$ yields

$$\begin{aligned} &\sum_{u \in \mathcal{N}_{l-1}(y)} (n-l+1)(m-1) p_{X|Y}(u \mid y) \\ &\leq e^\epsilon \sum_{u \in \mathcal{N}_{l-1}(y)} \sum_{v \in \mathcal{N}_1(u) \cap \mathcal{N}_l(y)} p_{X|Y}(v \mid y). \end{aligned}$$

Thus

$$\begin{aligned} &(n-l+1)(m-1) P_{l-1} \\ &\leq e^\epsilon \sum_{v \in \mathcal{N}_l(y)} \sum_{u \in \mathcal{N}_1(v) \cap \mathcal{N}_{l-1}(y)} p_{X|Y}(v \mid y) \end{aligned} \tag{A.4}$$

$$= e^{\epsilon} l P_l. \tag{A.5}$$

Recall that $N_l \triangleq |\mathcal{N}_l(x)| = \binom{n}{l}(m-1)^l$. Then by (A.5) we obtain that, for any l with $1 \le l \le n$,

$$\frac{P_{l-1}}{N_{l-1}} \le \frac{P_l}{N_l} e^{\epsilon}.$$

As a consequence, for any l with $0 \le l \le n$,

$$P_l \le \frac{N_l}{N_n} e^{(n-l)\epsilon} P_n. \tag{A.6}$$

Since $\sum_{l=0}^{n} P_l = 1$, taking summation over l in (A.6) yields

$$\begin{aligned} 1 &\le P_n \frac{1}{N_n e^{-n\epsilon}} \sum_{l=0}^{n} N_l e^{-l\epsilon} \\ &= P_n \frac{\big(1+(m-1)e^{-\epsilon}\big)^n}{N_n e^{-n\epsilon}}, \end{aligned}$$

i.e.,

$$P_n \ge \frac{N_n e^{-n\epsilon}}{\big(1+(m-1)e^{-\epsilon}\big)^n}.$$

This lower bound on P_n gives the following lower bound:

$$\begin{aligned} \sum_{l=0}^{n} l P_l \ge \sum_{l=0}^{n} l \bigg(P_l + a \frac{N_l e^{-l\epsilon}}{\sum_{k=0}^{n-1} N_k e^{-k\epsilon}} \bigg) \\ + \frac{n N_n e^{-n\epsilon}}{\big(1+(m-1)e^{-\epsilon}\big)^n}, \end{aligned}$$

where $a = P_n - \frac{N_n e^{-n\epsilon}}{\big(1+(m-1)e^{-\epsilon}\big)^n}$.

Consider the following optimization problem:

$$\begin{aligned} \min \quad & \sum_{l=0}^{n-1} l Q_l \\ \text{subject to} \quad & Q_l \ge 0, \qquad l = 0, 1, \dots, n-1, \\ & \frac{Q_{l-1}}{N_{l-1}} \le \frac{Q_l}{N_l} e^{\epsilon}, \quad l = 1, 2, \dots, n-1, \\ & \sum_{l=0}^{n-1} Q_l = 1 - \frac{N_n e^{-n\epsilon}}{\big(1+(m-1)e^{-\epsilon}\big)^n}. \end{aligned}$$

Suppose the optimal solution of this problem is $\{Q_0^*, Q_1^*, \dots, Q_{n-1}^*\}$. Then

$$\sum_{l=0}^{n-1} l \bigg(P_l + a \frac{N_l e^{-l\epsilon}}{\sum_{k=0}^{n-1} N_k e^{-k\epsilon}} \bigg) \ge \sum_{l=0}^{n-1} l Q_l^*$$

as $\left\{P_l + a\frac{N_l e^{-l\epsilon}}{\sum_{k=0}^{n-1} N_k e^{-k\epsilon}}, l = 0, 1, \dots, n-1\right\}$ is a feasible solution. Therefore,

$$\sum_{l=0}^{n} lP_l \geq \sum_{l=0}^{n-1} lQ_l^* + \frac{nN_n e^{-n\epsilon}}{\left(1 + (m-1)e^{-\epsilon}\right)^n}.$$

Similar to $\{P_l, l = 0, \dots, n\}$, $\{Q_l^*, l = 0, \dots, n-1\}$ satisfies

$$Q_l^* \leq \frac{N_l}{N_{n-1}} e^{(n-1-l)\epsilon} Q_{n-1}^*. \tag{A.7}$$

Since $\sum_{l=0}^{n-1} Q_l^* = 1 - \frac{N_n e^{-n\epsilon}}{\left(1+(m-1)e^{-\epsilon}\right)^n}$, taking summation over l in (A.7) yields

$$Q_{n-1}^* \geq \frac{N_{n-1} e^{-(n-1)\epsilon}}{\left(1 + (m-1)e^{-\epsilon}\right)^n}.$$

Using similar arguments we have

$$\sum_{l=0}^{n-1} lQ_l^* \geq \sum_{l=0}^{n-2} lC_l^* + \frac{(n-1)N_{n-1} e^{-(n-1)\epsilon}}{\left(1 + (m-1)e^{-\epsilon}\right)^n},$$

where $\{C_l^*, l = 0, \dots, n-2\}$ is the optimal solution of

$$\begin{aligned}
\text{min} \quad & \sum_{l=0}^{n-2} lC_l \\
\text{subject to} \quad & C_l \geq 0, \qquad l = 0, 1, \dots, n-2, \\
& \frac{C_{l-1}}{N_{l-1}} \leq \frac{C_l}{N_l} e^{\epsilon}, \quad l = 1, 2, \dots, n-2, \\
& \sum_{l=0}^{n-2} C_l = 1 - \frac{N_{n-1} e^{-(n-1)\epsilon}}{\left(1 + (m-1)e^{-\epsilon}\right)^n} - \frac{N_n e^{-n\epsilon}}{\left(1 + (m-1)e^{-\epsilon}\right)^n}.
\end{aligned}$$

Continue this procedure we obtain

$$\sum_{l=0}^{n} lP_l \geq \sum_{l=0}^{n} \frac{lN_l e^{-(n-l)\epsilon}}{\left(1 + (m-1)e^{-\epsilon}\right)^n} = \frac{n}{1 + \frac{e^{\epsilon}}{m-1}} = h(\epsilon).$$

Therefore, for any feasible $\{p_{X|Y}(x \mid y), x, y \in \mathcal{D}^n\}$,

$$\sum_{x \in \mathcal{D}^n} p_{X|Y}(x \mid y) d(x, y) = \sum_{l=0}^{n} lP_l \geq h(\epsilon),$$

which completes the proof of the claim.

By this claim, any feasible solution satisfies

$$\sum_{x\in\mathcal{D}^n}\sum_{y\in\mathcal{D}^n} p_Y(y)p_{X|Y}(x \mid y)d(x,y) \geq h(\epsilon).$$

Therefore

$$D^*_{\text{relaxed}}(\epsilon) \geq h(\epsilon). \tag{A.8}$$

Next we prove the following claim, which gives an upper bound on the minimum distortion $D^*_{\text{relaxed}}(\epsilon)$.

Claim. *Consider*

$$p_{X|Y}(x \mid y) = \frac{e^{-\epsilon d(x,y)}}{\big(1+(m-1)e^{-\epsilon}\big)^n}, \quad x, y \in \mathcal{D}^n,$$

and any $\{p_Y(y), y \in \mathcal{D}^n\}$ *with*

$$\sum_{y\in\mathcal{D}^n} p_Y(y) = 1, \qquad p_Y(y) \geq 0, \quad \forall y \in \mathcal{D}^n.$$

Then $\{p_{X|Y}(x \mid y), x, y \in \mathcal{D}^n\}$ *and* $\{p_Y(y), y \in \mathcal{D}^n\}$ *form a feasible solution of R-PD, and*

$$\sum_{x\in\mathcal{D}^n}\sum_{y\in\mathcal{D}^n} p_Y(y)p_{X|Y}(x \mid y)d(x,y) = h(\epsilon).$$

Proof of the Claim. Obviously the considered $\{p_{X|Y}(x \mid y), x, y \in \mathcal{D}^n\}$ and $\{p_Y(y), y \in \mathcal{D}^n\}$ satisfy constraints (2.16)–(2.18). Therefore to prove the feasibility, we are left with constraint (2.14) and (2.15). We first verify constraint (2.14). Consider any pair of neighboring elements $x, x' \in \mathcal{D}^n$ and any $y \in \mathcal{D}^n$. Then by the triangle inequality,

$$d(x,y) \leq d(x',y) - d(x',x) = d(x',y) - 1.$$

Therefore,

$$\begin{aligned} p_{X|Y}(x \mid y) &= \frac{e^{-\epsilon d(x,y)}}{\big(1+(m-1)e^{-\epsilon}\big)^n} \\ &\leq \frac{e^{-\epsilon(d(x',y)-1)}}{\big(1+(m-1)e^{-\epsilon}\big)^n} \\ &= e^{\epsilon} p_{X|Y}(x' \mid y). \end{aligned}$$

Next we verify constraint (2.15). For any $y \in \mathcal{D}^n$ and any integer l with $0 \leq l \leq n$, let $\mathcal{N}_l(x)$ be the set of elements with distance l to y as defined in (A.2). Then it is easy to see that $N_l \triangleq |\mathcal{N}_l(y)| = \binom{n}{l}(m-1)^l$, and for any $y \in \mathcal{D}^n$,

$$\mathcal{D}^n = \bigcup_{l=0}^{n} \mathcal{N}_l(y).$$

Therefore, for any $y \in \mathcal{D}^n$,

$$\begin{aligned}
&\sum_{x \in \mathcal{D}^n} p_{X|Y}(x \mid y) \\
&= \sum_{x \in \mathcal{D}^n} \frac{e^{-\epsilon d(x,y)}}{\bigl(1+(m-1)e^{-\epsilon}\bigr)^n} \\
&= \frac{1}{\bigl(1+(m-1)e^{-\epsilon}\bigr)^n} \sum_{l=0}^{n} \sum_{x \in \mathcal{N}_l(y)} e^{-\epsilon d(x,y)} \\
&= \frac{1}{\bigl(1+(m-1)e^{-\epsilon}\bigr)^n} \sum_{l=0}^{n} \binom{n}{l} (m-1)^l e^{-\epsilon l} \\
&= 1.
\end{aligned}$$

With feasibility verified, we can proceed to calculate the distortion. Let $g_\epsilon = 1 + (m-1)e^{-\epsilon}$. Then

$$\begin{aligned}
&\sum_{x \in \mathcal{D}^n} \sum_{y \in \mathcal{D}^n} p_Y(y) p_{X|Y}(x \mid y) d(x,y) \\
&= \frac{1}{(g_\epsilon)^n} \sum_{y \in \mathcal{D}^n} p_Y(y) \sum_{l=0}^{n} \sum_{x \in \mathcal{N}_l(y)} e^{-\epsilon d(x,y)} d(x,y) \\
&= \frac{1}{(g_\epsilon)^n} \sum_{y \in \mathcal{D}^n} p_Y(y) \sum_{l=0}^{n} \binom{n}{l} (m-1)^l e^{-\epsilon l} l \\
&= \frac{n(m-1)e^{-\epsilon}\bigl(1+(m-1)e^{-\epsilon}\bigr)^{n-1}}{(g_\epsilon)^n} \sum_{y \in \mathcal{D}^n} p_Y(y) \\
&= \frac{n}{1+\frac{e^\epsilon}{m-1}} \\
&= h(\epsilon),
\end{aligned}$$

which completes the proof of the claim.

By this claim, there exists a feasible solution such that

$$\sum_{x \in \mathcal{D}^n} \sum_{y \in \mathcal{D}^n} p_Y(y) p_{X|Y}(x \mid y) d(x,y) = h(\epsilon),$$

which implies

$$D^*_{\text{relaxed}}(\epsilon) \leq h(\epsilon).$$

Combining this upper bound with the lower bound (A.8) gives

$$D^*_{\text{relaxed}}(\epsilon) = h(\epsilon).$$

□

Lemma 8. *The optimal value $D^*_{\text{relaxed}}(\epsilon) = h(\epsilon)$ of R-PD implies the following lower bounds for any D with $0 \le D \le n$:*

$$\epsilon^*_{\text{i}}(D) \ge h^{-1}(D), \tag{A.9}$$
$$\epsilon^*_{\text{d}}(D) \ge \max\{h^{-1}(D) - \epsilon_X, 0\}. \tag{A.10}$$

Proof. First we derive the lower bound on $\epsilon^*_{\text{i}}(D)$. Let δ be an arbitrary positive number. For any D with $0 \le D \le n$, let $\epsilon_{D,\delta} = \epsilon^*_{\text{i}}(D) + \delta$. Then by the definition of ϵ^*_{i}, we have that $(\epsilon_{D,\delta}, D)$ is achievable under identifiability. Therefore

$$D \ge D^*_{\text{i}}(\epsilon_{D,\delta}) \ge D^*_{\text{relaxed}}(\epsilon_{D,\delta}) = h(\epsilon_{D,\delta}),$$

where $D^*_{\text{i}}(\cdot)$ is the optimal value of PD-I. Since h is a decreasing function, this implies that $\epsilon_{D,\delta} \ge h^{-1}(D)$. Therefore

$$\epsilon^*_{\text{i}}(D) \ge h^{-1}(D) - \delta.$$

Letting $\delta \to 0$ yields

$$\epsilon^*_{\text{i}}(D) \ge h^{-1}(D).$$

Next we derive the lower bound on $\epsilon^*_{\text{d}}(D)$ using arguments similar to those in the proof of the lower bound on $\epsilon^*_{\text{i}}(D)$. Let δ be an arbitrary positive number. For any D with $0 \le D \le n$, let $\epsilon_{D,\delta} = \epsilon^*_{\text{d}}(D) + \delta$. Then by the definition of ϵ^*_{d}, we have that $(\epsilon_{D,\delta}, D)$ is achievable under differential privacy. Therefore

$$D \ge D^*_{\text{d}}(\epsilon_{D,\delta}) \ge D^*_{\text{relaxed}}(\epsilon_{D,\delta} + \epsilon_X) = h(\epsilon_{D,\delta} + \epsilon_X),$$

where $D^*_{\text{d}}(\cdot)$ is the optimal value of PD-DP. Since h is a decreasing function, this implies that $\epsilon_{D,\delta} + \epsilon_X \ge h^{-1}(D)$. Therefore

$$\epsilon^*_{\text{d}}(D) \ge h^{-1}(D) - \epsilon_X - \delta.$$

Letting $\delta \to 0$ yields

$$\epsilon^*_{\text{d}}(D) \ge h^{-1}(D) - \epsilon_X.$$

Since the privacy level is nonnegative, we obtain the lower bound in (A.10). □

Lemma 9. *The privacy–distortion function ϵ^*_{i} of a database X is bounded from below as*

$$\epsilon^*_{\text{i}}(D) \ge \epsilon_X$$

for any D with $0 \le D \le n$, where ϵ_X is the constant defined in (2.11)*.*

Proof. Suppose by contradiction that there exists a D with $0 \le D \le n$ such that $\epsilon^*_{\text{i}}(D) < \epsilon_X$. Let δ be an arbitrary positive number with $0 < \delta < \epsilon_X - \epsilon^*_{\text{i}}(D)$, and let $\epsilon = \epsilon^*_{\text{i}}(D) + \delta$. Then $\epsilon < \epsilon_X$ and (ϵ, D) is achievable under identifiability. Consider the mechanism that achieves (ϵ, D). Then by the requirement of identifiability, for any neighboring $x, x' \in \mathcal{D}^n$ and any $y \in \mathcal{D}^n$,

$$p_{X|Y}(x \mid y) \le e^{\epsilon} p_{X|Y}(x' \mid y). \tag{A.11}$$

Let $p_Y(\cdot)$ be the pmf of the output Y. Then $p_Y(y) \geq 0$ for any $y \in \mathcal{D}^n$. Therefore, multiplying both sides of (A.11) by $p_Y(y)$ and taking summation over $y \in \mathcal{D}^n$ yield

$$\sum_{y \in \mathcal{D}^n} p_{X|Y}(x \mid y) p_Y(y) \leq \sum_{y \in \mathcal{D}^n} e^{\epsilon} p_{X|Y}(x' \mid y) p_Y(y),$$

which implies

$$p_X(x) \leq e^{\epsilon} p_X(x').$$

Then there do not exist neighboring $x, x' \in \mathcal{D}^n$ with $p_X(x) = e^{\epsilon_X} p_X(x')$ since $\epsilon < \epsilon_X$, which contradicts with the definition of ϵ_X in (2.11). □

Lemma 10. *For $\epsilon \geq \widetilde{\epsilon}_X$, the mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ defined in* (2.21) *satisfies ϵ-identifiability, and the distortion of $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ is given by $\mathbb{E}[d(X,Y)] = h(\epsilon)$.*

Proof. Consider any $\epsilon \geq \widetilde{\epsilon}_X$. Then under the mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$, the posterior probability for any $x, y \in \mathcal{D}^n$ is given by

$$p_{X|Y}(x \mid y) = \frac{p_{Y|X}(y \mid x) p_X(x)}{p_Y(y)} = \frac{e^{-\epsilon d(x,y)}}{\left(1 + (m-1)e^{-\epsilon}\right)^n}.$$

As shown in the proof of Lemma 7, this $\{p_{X|Y}(x \mid y), x, y \in \mathcal{D}^n\}$ and the corresponding $\{p_Y(y), y \in \mathcal{D}^n\}$ form an optimal solution of the relaxed optimization problem R-PD. Following the same arguments as in the proof of Lemma 7 we can conclude that $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ satisfies ϵ-identifiability, and the distortion of $\mathcal{E}_{\mathrm{i}}^{\epsilon}$ is given by $\mathbb{E}[d(X,Y)] = h(\epsilon)$. □

Lemma 11. *The mechanism $\mathcal{E}_{\mathrm{d}}^{\epsilon}$ defined in* (2.22) *satisfies ϵ-differential privacy, and the distortion of $\mathcal{E}_{\mathrm{d}}^{\epsilon}$ is given by $\mathbb{E}[d(X,Y)] = h(\epsilon)$.*

Proof. Under mechanism $\mathcal{E}_{\mathrm{d}}^{\epsilon}$, $\{p_{Y|X}(y \mid x), x, y \in \mathcal{D}^n\}$ has the same form as the posteriors under mechanism $\mathcal{E}_{\mathrm{i}}^{\epsilon}$. Therefore still by similar arguments as in the proof of Lemma 7, $\mathcal{E}_{\mathrm{d}}^{\epsilon}$ satisfies ϵ-differential privacy, and the distortion of $\mathcal{E}_{\mathrm{d}}^{\epsilon}$ is given by $\mathbb{E}[d(X,Y)] = h(\epsilon)$. □

APPENDIX B

PROOF OF LEMMA 1

Proof. Consider any nonnegative payment mechanism $\boldsymbol{R}$ and a Nash equilibrium of it, denoted by $\boldsymbol{\sigma}$. For an individual i, consider any strategy σ_i' of individual i and let

$$p_1 = \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 1), \quad q_1 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 1),$$
$$p_0 = \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 0), \quad q_0 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 0).$$

When other individuals follow $\boldsymbol{\sigma}_{-i}$, the expected utility of individual i at the strategy σ_i' is a function of (p_1, p_0, q_1, q_0), denoted by $U_i(p_1, p_0, q_1, q_0)$. We derive the form of this function below. The expected payment to individual i can be written as

$$\begin{aligned}
&\mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X})] \\
&= \sum_{x_i, \boldsymbol{s}} \Big\{ \mathbb{P}_{\sigma_i'}(X_i = x_i, S_i = s_i, \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}) \\
&\qquad \cdot \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, S_i = s_i, \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}] \Big\} \\
&= \sum_{x_i, \boldsymbol{s}} \Big\{ \mathbb{P}_{\sigma_i'}(X_i = x_i \mid S_i = s_i) \mathbb{P}(S_i = s_i, \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}) \\
&\qquad \cdot \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}] \Big\},
\end{aligned}$$

where we have used the fact that X_i is independent from $\boldsymbol{S}_{-i}$ given S_i, and $\boldsymbol{X}_{-i}$ is independent from S_i given X_i and $\boldsymbol{S}_{-i}$. The term $\mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}]$ does not depend on the strategy of individual i since

$$\begin{aligned}
&\mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}] \\
&= \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(x_i, \boldsymbol{X}_{-i}) \mid X_i = x_i, \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}] \\
&= \mathbb{E}_{\boldsymbol{\sigma}_{-i}}[R_i(x_i, \boldsymbol{X}_{-i}) \mid \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}],
\end{aligned}$$

where the last equality follow from the conditional independence between X_i and $\boldsymbol{X}_{-i}$ given $\boldsymbol{S}_{-i}$. Then

$$\begin{aligned}
&\mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X})] \\
&= \sum_{x_i, s_i} \Big\{ \mathbb{P}_{\sigma_i'}(X_i = x_i \mid S_i = s_i) \\
&\qquad \cdot \sum_{\boldsymbol{s}_{-i}} \Big(\mathbb{P}(\boldsymbol{S} = \boldsymbol{s}) \mathbb{E}_{\boldsymbol{\sigma}_{-i}}[R_i(x_i, \boldsymbol{X}_{-i}) \mid \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}] \Big) \Big\} \\
&= K_1 p_1 + K_0 p_0 + L_1 q_1 + L_0 q_0,
\end{aligned}$$

where

$$\begin{aligned}
K_{s_i} = \sum_{\boldsymbol{s}_{-i}} &\Big(\mathbb{P}(S_i = s_i, \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}) \\
&\cdot \mathbb{E}_{\boldsymbol{\sigma}_{-i}}[R_i(1, \boldsymbol{X}_{-i}) \mid \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}] \Big), \ s_i \in \{0, 1\},
\end{aligned}$$

are the expected payment received by individual i when she reports 1, weighted by $\mathbb{P}(S_i = 1)$ and $\mathbb{P}(S_i = 0)$ when her private signal is 1 and 0, respectively, and

$$L_{s_i} = \sum_{\boldsymbol{s}_{-i}} \Big(\mathbb{P}(S_i = s_i, \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}) \\ \cdot \mathbb{E}_{\boldsymbol{\sigma}_{-i}}[R_i(0, \boldsymbol{X}_{-i}) \mid \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}] \Big),\ s_i \in \{0, 1\},$$

are the expected payment received by individual i when she reports 0, weighted by $\mathbb{P}(S_i = 1)$ and $\mathbb{P}(S_i = 0)$ when her private signal is 1 and 0, respectively. Note that K_1, K_0, L_1 and L_0 do not depend on p_1, p_0, q_1, and q_0. The privacy level of the reported data at strategy σ_i' is

$$\zeta(\sigma_i') = \max\bigg\{ \bigg|\ln\frac{p_1}{p_0}\bigg|, \bigg|\ln\frac{1-p_1}{1-p_0}\bigg|, \bigg|\ln\frac{q_1}{q_0}\bigg|, \bigg|\ln\frac{1-q_1}{1-q_0}\bigg|, \\ \bigg|\ln\frac{1-p_1-q_1}{1-p_0-q_0}\bigg|, \bigg|\ln\frac{p_1+q_1}{p_0+q_0}\bigg| \bigg\}.$$

With a little abuse of notation, we regard $\zeta(\sigma_i')$ as a function $\zeta(p_1, p_0, q_1, q_0)$. The expected utility of individual i can thus be written as

$$\begin{aligned} & U_i(p_1, p_0, q_1, q_0) \\ &= \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) - g(\zeta(\sigma_i'))] \\ &= K_1 p_1 + K_0 p_0 + L_1 q_1 + L_0 q_0 - g(\zeta(p_1, p_0, q_1, q_0)). \end{aligned}$$

Next we discuss the best response of individual i for different cases of the values of K_1, K_0, L_1 and L_0. Since $\boldsymbol{R}$ is a nonnegative payment mechanism, these values are all nonnegative. Notice that for any $s_i \in \{0, 1\}$, $\boldsymbol{s}_{-i} \in \{0, 1\}^{N-1}$, $\mathbb{P}(S_i = s_i, \boldsymbol{S}_{-i} = \boldsymbol{s}_{-i}) > 0$. Therefore, K_1 and K_0 are either both equal to zero or both positive. The same argument also applies to L_1 and L_0. (1) When all of K_1, K_0, L_1 and L_0 are zero, a best response of individual i should minimize the privacy cost. Thus the strategy of individual i in a Nash equilibrium is to report X_i that is independent of S_i so the privacy cost is zero. (2) When K_1 and K_0 are positive but L_1 and L_0 are zero, the best response of individual i is to always report $X_i = 1$. (3) Similarly, when K_1 and K_0 are zero but L_1 and L_0 are positive, the best response of individual i is to always report $X_i = 0$. We can see that the strategy of individual i in a Nash equilibrium is non-informative in all the three cases above. (4) In the remainder of this proof, we focus on the case that all of K_1, K_0, L_1 and L_0 are positive.

If a best response of individual i is to always not participate, then it is a non-informative strategy. Otherwise, a best response of individual i is specified by an optimal solution of the following optimization problem:

$$\begin{aligned} \max_{p_1, p_0, q_1, q_0} \quad & U_i(p_1, p_0, q_1, q_0) \\ \text{subject to} \quad & 0 \le p_1 \le 1, 0 \le q_1 \le 1, \\ & 0 \le p_1 + q_1 \le 1, \\ & 0 \le p_0 \le 1, 0 \le q_0 \le 1, \\ & 0 \le p_0 + q_0 \le 1, \\ & p_1 + q_1 + p_0 + q_0 > 0. \end{aligned} \tag{P}$$

First, we prove that an optimal solution $(p_1^*, p_0^*, q_1^*, q_0^*)$ of (P) must satisfy that $p_1^* + q_1^* = p_0^* + q_0^*$. Suppose not. Without loss of generality we assume that $p_1^* + q_1^* < p_0^* + q_0^*$. We will find another solution $(p_1', p_0^*, q_1', q_0^*)$ that yields better utility, which contradicts the optimality of $(p_1^*, p_0^*, q_1^*, q_0^*)$.

Since we assume that $p_1^* + q_1^* < p_0^* + q_0^*$, then at least one of the following two inequality holds: $p_1^* < p_0^*$, $q_1^* < q_0^*$. Still without loss of generality we assume that $p_1^* < p_0^*$. Then if $q_1^* < q_0^*$, let $p_1' = p_0^*$ and $q_1' = q_0^*$. Since K_1 and L_1 are positive, $(p_1', p_0^*, q_1', q_0^*)$ yields higher payment. It is easy to verify that $\zeta(p_1', p_0^*, q_1', q_0^*) < \zeta(p_1^*, p_0^*, q_1^*, q_0^*)$. Thus $(p_1', p_0^*, q_1', q_0^*)$ yields better utility. For the other case that $q_1^* \geq q_0^*$, let $p_1' = p_0^* + q_0^* - q_1^*$ and $q_1' = q_1^*$. Then $p_1^* < p_1' \leq p_0^*$. Since K_1 is positive, $(p_1', p_0^*, q_1', q_0^*)$ yields higher payment. To check the privacy cost, notice that

$$\zeta(p_1^*, p_0^*, q_1^*, q_0^*) = \max\left\{\ln\frac{p_0^*}{p_1^*}, \ln\frac{1-p_1^*}{1-p_0^*}, \ln\frac{q_1^*}{q_0^*}, \ln\frac{1-q_0^*}{1-q_1^*},\right.$$
$$\left.\ln\frac{1-p_1^*-q_1^*}{1-p_0^*-q_0^*}, \ln\frac{p_0^*+q_0^*}{p_1^*+q_1^*}\right\},$$

and

$$\zeta(p_1', p_0^*, q_1', q_0^*) = \max\left\{\ln\frac{p_0^*}{p_1'}, \ln\frac{1-p_1'}{1-p_0^*}, \ln\frac{q_1'}{q_0^*}, \ln\frac{1-q_0^*}{1-q_1'}\right\}.$$

Since $p_1' > p_1^*$ and $q_1' = q_1^*$, $\zeta(p_1', p_0^*, q_1', q_0^*) \leq \zeta(p_1^*, p_0^*, q_1^*, q_0^*)$. Thus $(p_1', p_0^*, q_1', q_0^*)$ yields better utility. Therefore, by contradiction, we must have $p_1^* + q_1^* = p_0^* + q_0^*$.

Next, we prove that an optimal solution $(p_1^*, p_0^*, q_1^*, q_0^*)$ must satisfy that $p_1^* + q_1^* = p_0^* + q_0^* = 1$. Still, suppose not. Then we will find another solution (p_1', p_0', q_1', q_0') that yields better utility. Let

$$p_1' = \frac{p_1^*}{p_1^* + q_1^*}, \quad q_1' = \frac{q_1^*}{p_1^* + q_1^*},$$
$$p_0' = \frac{p_0^*}{p_0^* + q_0^*}, \quad q_0' = \frac{q_0^*}{p_0^* + q_0^*}.$$

We have seen that $p_1^* + q_1^* = p_0^* + q_0^*$. By the last constraint of (P), $p_1^* + q_1^* = p_0^* + q_0^* > 0$. Since we assume that $p_1^* + q_1^*$ and $p_0^* + q_0^*$ are not equal to 1, they must be less than 1. Since K_1, K_0, L_1 and L_0 are positive, (p_1', p_0', q_1', q_0') yields higher payment. It is easy to verify that $\zeta(p_1', p_0', q_1', q_0') \leq \zeta(p_1^*, p_0^*, q_1^*, q_0^*)$. Thus (p_1', p_0', q_1', q_0') yields better utility, which contradicts the optimality of $(p_1^*, p_0^*, q_1^*, q_0^*)$.

By the results above, to find an optimal solution of (P), we can focus on feasible (p_1, p_0, q_1, q_0) such that $q_1 = 1 - p_1$ and $q_0 = 1 - p_0$. Let

$$\overline{U}_i(p_1, p_0) = \overline{K}_1 p_1 + \overline{K}_0 p_0 + \overline{K} - g(\zeta(p_1, p_0)),$$

where $\overline{K}_1 = K_1 - L_1$, $\overline{K}_0 = K_0 - L_0$, $\overline{K} = L_1 + L_0$, and with a little abuse of notation,

$$\zeta(p_1, p_0) = \max\left\{\left|\ln\frac{p_1}{p_0}\right|, \left|\ln\frac{1-p_1}{1-p_0}\right|\right\}.$$

Then $(p_1^*, p_0^*, q_1^*, q_0^*)$ is an optimal solution of (P) if and only if (p_1^*, p_0^*) is an optimal solution of the following optimization problem P':

$$\max_{0\le p_1\le 1, 0\le p_0\le 1} \quad \overline{U}_i(p_1, p_0) \tag{P'}$$

Let (p_1^*, p_0^*) be an optimal solution of (P'). The strategy specified by $(p_1^*, p_0^*, 1-p_1^*, 1-p_0^*)$ is a symmetric randomized response if $p_1^*+p_0^*=1$, and is non-informative if $p_1^*=p_0^*$. Thus it suffices to prove that if $p_1^*+p_0^*\neq 1$, then $p_1^*=p_0^*$. We divide the case that $p_1^*+p_0^*\neq 1$ into two cases: $p_1^*+p_0^*>1$ and $p_1^*+p_0^*<1$, and prove that $p_1^*=p_0^*$ in both cases.

Case 1: $p_1^*+p_0^*>1$. Suppose, for contradiction, $p_1^*\neq p_0^*$.

If $p_1^*=1$, then

$$\max\left\{\left|\ln\frac{p_1^*}{p_0^*}\right|, \left|\ln\frac{1-p_1^*}{1-p_0^*}\right|\right\} = +\infty.$$

Consider $p_1=1$ and $p_0=1$. Then by the convention

$$\max\left\{\left|\ln\frac{p_1}{p_0}\right|, \left|\ln\frac{1-p_1}{1-p_0}\right|\right\} = 0.$$

Since $\overline{U}_i(p_1^*, p_0^*) \ge \overline{U}_i(p_1, p_0)$, then $\overline{K}_1 + \overline{K}_0 p_0^* - g(+\infty) \ge \overline{K}_1 + \overline{K}_0 - g(0)$. Thus

$$g(+\infty) \le -\overline{K}_0(1-p_0^*) < +\infty. \tag{B.1}$$

Since $g(+\infty)\ge 0$, this also indicates that $\overline{K}_0\le 0$. Next consider $p_1=1$ and $p_0=0$. Then

$$\max\left\{\left|\ln\frac{p_1}{p_0}\right|, \left|\ln\frac{1-p_1}{1-p_0}\right|\right\} = +\infty.$$

Since $\overline{U}_i(p_1^*, p_0^*) \ge \overline{U}_i(p_1, p_0)$, then $\overline{K}_1+\overline{K}_0 p_0^*-g(+\infty) \ge \overline{K}_1 - g(+\infty)$. Thus $\overline{K}_0\ge 0$, where we have used the fact that $g(+\infty)<+\infty$. Combining the above arguments we have $\overline{K}_0=0$. However, by (B.1), this indicates that $g(+\infty)=0$, which contradicts the assumption that $g(\xi)=0$ only for $\xi=0$. Therefore, $p_1^*\neq 1$. Following similar arguments we have $p_0^*\neq 1$, either.

If $p_1^*>p_0^*$, then noticing that $p_1^*+p_0^*>1$ we have

$$\max\left\{\left|\ln\frac{p_1^*}{p_0^*}\right|, \left|\ln\frac{1-p_1^*}{1-p_0^*}\right|\right\} = \ln\frac{1-p_0^*}{1-p_1^*}.$$

Consider

$$p_1 = \frac{\frac{1-p_0^*}{1-p_1^*}}{\frac{1-p_0^*}{1-p_1^*}+1}, \quad p_0 = \frac{1}{\frac{1-p_0^*}{1-p_1^*}+1}.$$

Then

$$\max\left\{\left|\ln\frac{p_1}{p_0}\right|, \left|\ln\frac{1-p_1}{1-p_0}\right|\right\} = \ln\frac{1-p_0^*}{1-p_1^*}.$$

Since $\overline{U}_i(p_1^*, p_0^*) \geq \overline{U}_i(p_1, p_0)$, then

$$\overline{K}_1 p_1^* + \overline{K}_0 p_0^* - g\left(\ln \frac{1-p_0^*}{1-p_1^*}\right) + \overline{K}$$
$$\geq \overline{K}_1 p_1 + \overline{K}_0 p_0 - g\left(\ln \frac{1-p_0^*}{1-p_1^*}\right) + \overline{K}.$$

Thus, inserting p_1 and p_0 we obtain

$$\overline{K}_1(1-p_1^*) + \overline{K}_0(1-p_0^*) \geq 0,$$

where we have used the condition $p_1^* + p_0^* > 1$. Next still consider $p_1 = p_0 = 1$. Since $\overline{U}_i(p_1^*, p_0^*) \geq \overline{U}_i(p_1, p_0)$, then

$$\overline{K}_1 p_1^* + \overline{K}_0 p_0^* - g\left(\ln \frac{1-p_0^*}{1-p_1^*}\right) \geq \overline{K}_1 + \overline{K}_0 - g(0).$$

Thus

$$-g\left(\ln \frac{1-p_0^*}{1-p_1^*}\right) \geq \overline{K}_1(1-p_1^*) + \overline{K}_0(1-p_0^*) \geq 0.$$

which indicates that $\ln \frac{1-p_0^*}{1-p_1^*} = 0$. Thus $p_1^* = p_0^*$, which contradicts the assumption.

If $p_1^* < p_0^*$, then noticing that $p_1^* + p_0^* > 1$ we have

$$\max\left\{\left|\ln \frac{p_1^*}{p_0^*}\right|, \left|\ln \frac{1-p_1^*}{1-p_0^*}\right|\right\} = \ln \frac{1-p_1^*}{1-p_0^*}.$$

We use similar arguments to obtain contradiction. Consider

$$p_1 = \frac{1}{\frac{1-p_1^*}{1-p_0^*} + 1}, \quad p_0 = \frac{\frac{1-p_1^*}{1-p_0^*}}{\frac{1-p_1^*}{1-p_0^*} + 1}.$$

Then since $\overline{U}_i(p_1^*, p_0^*) \geq \overline{U}_i(p_1, p_0)$, we have $\overline{K}_1(1-p_1^*) + \overline{K}_0(1-p_0^*) \geq 0$. Next still consider $p_1 = p_0 = 1$. Then since $\overline{U}_i(p_1^*, p_0^*) \geq \overline{U}_i(p_1, p_0)$, we have

$$-g\left(\ln \frac{1-p_1^*}{1-p_0^*}\right) \geq \overline{K}_1(1-p_1^*) + \overline{K}_0(1-p_0^*) \geq 0.$$

which again indicates that $\ln \frac{1-p_1^*}{1-p_0^*} = 0$. So $p_1^* = p_0^*$, which contradicts the assumption.

In summary, for the case that $p_1^* + p_0^* > 1$, $p_1^* = p_0^*$.

Case 2: $p_1^* + p_0^* < 1$. Suppose, for contradiction, $p_1^* \neq p_0^*$. Then we obtain contradictions by similar arguments as used in Case 1.

First, by comparing $\overline{U}_i(p_1^*, p_0^*)$ with $\overline{U}_i(0,0)$ and $\overline{U}_i(0,1)$ we can prove that $p_1^* \neq 0, p_0^* \neq 0$. If $p_1^* > p_0^*$, then by comparing $\overline{U}_i(p_1^*, p_0^*)$ with the expected utility at

$$p_1 = \frac{\frac{p_1^*}{p_0^*}}{\frac{p_1^*}{p_0^*} + 1}, \quad p_0 = \frac{1}{\frac{p_1^*}{p_0^*} + 1},$$

we have $\overline{K}_1 p_1^* + \overline{K}_0 p_0^* \leq 0$. By comparing $\overline{U}_i(p_1^*, p_0^*)$ with $\overline{U}_i(0,0)$, we have

$$g\left(\ln \frac{1 - p_1^*}{1 - p_0^*}\right) \leq \overline{K}_1 p_1^* + \overline{K}_0 p_0^* \leq 0.$$

Therefore, $p_1^* = p_0^*$, which contradicts the assumption. If $p_1^* < p_0^*$, then by comparing $\overline{U}_i(p_1^*, p_0^*)$ with the expected utility at

$$p_1 = \frac{1}{\frac{p_0^*}{p_1^*} + 1}, \quad p_0 = \frac{\frac{p_0^*}{p_1^*}}{\frac{p_0^*}{p_1^*} + 1},$$

we have $\overline{K}_1 p_1^* + \overline{K}_0 p_0^* \leq 0$. By comparing $\overline{U}_i(p_1^*, p_0^*)$ with $\overline{U}_i(0,0)$, we have

$$g\left(\ln \frac{1 - p_1^*}{1 - p_0^*}\right) \leq \overline{K}_1 p_1^* + \overline{K}_0 p_0^* \leq 0.$$

Therefore, $p_1^* = p_0^*$, which contradicts the assumption. In summary, for the case that $p_1^* + p_0^* < 1$, we also have $p_1^* = p_0^*$ by similar arguments as used in Case 1. This completes the proof.

□

APPENDIX C

PROOF OF LEMMA 2

Proof. For any nonnegative payment mechanism $\boldsymbol{R}$ in which the strategy profile $(\sigma_i^{(-\epsilon)}, \boldsymbol{\sigma}_{-i})$ is a Nash equilibrium, consider the payment mechanism $\boldsymbol{R}'$ defined by

$$\boldsymbol{R}'(x_i, \boldsymbol{x}_{-i}) = \boldsymbol{R}(1 - x_i, \boldsymbol{x}_{-i}).$$

We first prove that $(\sigma_i^{(\epsilon)}, \boldsymbol{\sigma}_{-i})$ is a Nash equilibrium in $\boldsymbol{R}'$. For an individual i, consider any strategy σ_i' of individual i and let

$$\begin{aligned} p_1 &= \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 1), \quad q_1 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 1), \\ p_0 &= \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 0), \quad q_0 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 0). \end{aligned}$$

We say (p_1, p_0, q_1, q_0) is feasible if it satisfies that

$$\begin{aligned} &0 \le p_1 \le 1, \quad 0 \le q_1 \le 1, \quad 0 \le p_1 + q_1 \le 1, \\ &0 \le p_0 \le 1, \quad 0 \le q_0 \le 1, \quad 0 \le p_0 + q_0 \le 1. \end{aligned}$$

Then following the notation in the proof of Lemma 1, in the mechanism $\boldsymbol{R}'$ and $\boldsymbol{R}$, we denote the expected utility of individual i at σ_i' when other individuals follow $\boldsymbol{\sigma}_{-i}$ by $U_i'(p_1, p_0, q_1, q_0)$ and $U_i(p_1, p_0, q_1, q_0)$, respectively, and they can be written as follows:

$$\begin{aligned} U_i'(p_1, p_0, q_1, q_0) &= K_{1,i}' p_1 + K_{0,i}' p_0 + L_{1,i}' q_1 + L_{0,i}' q_0 \\ &\quad - g(\zeta(p_1, p_0, q_1, q_0)), \\ U_i(p_1, p_0, q_1, q_0) &= K_{1,i} p_1 + K_{0,i} p_0 + L_{1,i} q_1 + L_{0,i} q_0 \\ &\quad - g(\zeta(p_1, p_0, q_1, q_0)). \end{aligned}$$

We derive the relations between $K_{1,i}'$, $K_{0,i}'$, $L_{1,i}'$, $L_{0,i}'$ and $K_{1,i}$, $K_{0,i}$, $L_{1,i}$, $L_{0,i}$. By definition,

$$\begin{aligned} K_{1,i}' &= \sum_{\boldsymbol{x}_{-i}} R_i'(1, \boldsymbol{x}_{-i}) \mathbb{P}_{\boldsymbol{\sigma}_{-i}}(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}, S_i = 1) \\ &= \sum_{\boldsymbol{x}_{-i}} R_i(0, \boldsymbol{x}_{-i}) \mathbb{P}_{\boldsymbol{\sigma}_{-i}}(\boldsymbol{X}_{-i} = \boldsymbol{x}_{-i}, S_i = 1) \\ &= L_{1,i}. \end{aligned}$$

Similarly, $K_{0,i}' = L_{0,i}$, $L_{1,i}' = K_{1,i}$ and $L_{0,i}' = K_{0,i}$. Since $(\sigma_i^{(-\epsilon)}, \boldsymbol{\sigma}_{-i})$ is a Nash equilibrium in $\boldsymbol{R}$, for any feasible (p_1, p_0, q_1, q_0),

$$U_i\left(\frac{1}{e^\epsilon + 1}, \frac{e^\epsilon}{e^\epsilon + 1}, \frac{e^\epsilon}{e^\epsilon + 1}, \frac{1}{e^\epsilon + 1}\right) \ge U_i(p_1, p_0, q_1, q_0).$$

Therefore, for any feasible (p_1, p_0, q_1, q_0),

$$U_i'\left(\frac{e^\epsilon}{e^\epsilon + 1}, \frac{1}{e^\epsilon + 1}, \frac{1}{e^\epsilon + 1}, \frac{e^\epsilon}{e^\epsilon + 1}\right) \ge U_i'(p_1, p_0, q_1, q_0),$$

where we have used the symmetry property of the cost function g. This implies that $\sigma_i^{(\epsilon)}$ is a best response of individual i in $\boldsymbol{R}'$ when other individuals follow $\boldsymbol{\sigma}_{-i}$. Now consider any individual j with $j \neq i$ and any strategy σ_j'. Let

$$\begin{aligned} p_1 &= \mathbb{P}_{\sigma_j'}(X_j = 1 \mid S_j = 1), \quad q_1 = \mathbb{P}_{\sigma_j'}(X_j = 0 \mid S_j = 1), \\ p_0 &= \mathbb{P}_{\sigma_j'}(X_j = 1 \mid S_j = 0), \quad q_0 = \mathbb{P}_{\sigma_j'}(X_j = 0 \mid S_j = 0). \end{aligned}$$

Let $\sigma_{-j}^{(\epsilon)} = (\sigma_i^{(\epsilon)}, \boldsymbol{\sigma}_{-i,j})$ and $\sigma_{-j}^{(-\epsilon)} = (\sigma_i^{(-\epsilon)}, \boldsymbol{\sigma}_{-i,j})$. Then similarly, in the mechanism $\boldsymbol{R}'$ and $\boldsymbol{R}$, we denote the expected utility of individual j at σ_j' when other individuals follow $\sigma_{-j}^{(\epsilon)}$ and $\sigma_{-j}^{(-\epsilon)}$ by $U_j'(p_1, p_0, q_1, q_0)$ and $U_j(p_1, p_0, q_1, q_0)$, respectively, and they can be written as follows:

$$\begin{aligned} U_j'(p_1, p_0, q_1, q_0) &= K_{1,j}'p_1 + K_{0,j}'p_0 + L_{1,j}'q_1 + L_{0,j}'q_0 \\ &\quad - g(\zeta(p_1, p_0, q_1, q_0)), \\ U_j(p_1, p_0, q_1, q_0) &= K_{1,j}p_1 + K_{0,j}p_0 + L_{1,j}q_1 + L_{0,j}q_0 \\ &\quad - g(\zeta(p_1, p_0, q_1, q_0)). \end{aligned}$$

We derive the relations between $K_{1,j}'$, $K_{0,j}'$, $L_{1,j}'$, $L_{0,j}'$ and $K_{1,j}$, $K_{0,j}$, $L_{1,j}$, $L_{0,j}$. By definition,

$$\begin{aligned} K_{1,j}' = &\sum_{\boldsymbol{x}_{-i,j}} \sum_{x_i} R_i'(x_i, 1, \boldsymbol{x}_{-i,j}) \\ &\cdot \sum_{s_i} \mathbb{P}_{\sigma_i^{(\epsilon)}}(X_i = x_i \mid S_i = s_i) \\ &\qquad \cdot \mathbb{P}_{\boldsymbol{\sigma}_{-i,j}}(\boldsymbol{X}_{-i,j} = \boldsymbol{x}_{-i,j}, S_i = s_i, S_j = 1) \\ = &\sum_{\boldsymbol{x}_{-i,j}} \sum_{x_i} R_i(1 - x_i, 1, \boldsymbol{x}_{-i,j}) \\ &\cdot \sum_{s_i} \mathbb{P}_{\sigma_i^{(-\epsilon)}}(X_i = 1 - x_i \mid S_i = s_i) \\ &\qquad \cdot \mathbb{P}_{\boldsymbol{\sigma}_{-i,j}}(\boldsymbol{X}_{-i,j} = \boldsymbol{x}_{-i,j}, S_i = s_i, S_j = 1) \\ = &\, K_{1,j}. \end{aligned}$$

Similarly, $K_{0,j}' = K_{0,j}$, $L_{1,j}' = L_{1,j}$, and $L_{0,j}' = L_{0,j}$. Therefore, for any feasible (p_1, p_0, q_1, q_0),

$$U_j'(p_1, p_0, q_1, q_0) = U_j(p_1, p_0, q_1, q_0).$$

Thus σ_j is a best response of individual j in $\boldsymbol{R}'$ when other individuals follow $\sigma_{-j}^{(\epsilon)}$. This completes the proof for $(\sigma_i^{(\epsilon)}, \boldsymbol{\sigma}_{-i})$ to be a Nash equilibrium in $\boldsymbol{R}'$.

With the above proof, it is not hard to verify that the expected payment to each individual at these two equilibria of the two mechanisms are the same. □

APPENDIX D

PROOF OF LEMMA 3

Proof. Consider any payment mechanism $\boldsymbol{R}$ and any Nash equilibrium $\boldsymbol{\sigma}$ of it. We will construct a genie-aided mechanism $\widehat{\boldsymbol{R}}$ such that $\boldsymbol{\sigma}$ is also a Nash equilibrium of $\widehat{\boldsymbol{R}}$ and the expected payment to each individual at this equilibrium is the same under $\boldsymbol{R}$ and $\widehat{\boldsymbol{R}}$.

As in the proof of Lemma 1, for any individual i, consider any strategy σ_i' of individual i and let

$$\begin{aligned} p_1 &= \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 1), \quad q_1 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 1), \\ p_0 &= \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 0), \quad q_0 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 0). \end{aligned}$$

Then we will first derive the expected utility of individual i at the strategy σ_i' as a function of (p_1, p_0, q_1, q_0), denoted by $U_i(p_1, p_0, q_1, q_0)$, but using a slightly different expression from the form in Lemma 1. When other individuals follow $\boldsymbol{\sigma}_{-i}$, the expected payment to individual i at the strategy σ_i' can be written as

$$\begin{aligned} &\mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X})] \\ &= \sum_{x_i, s_i, w} \Big\{ \mathbb{P}_{\sigma_i'}(X_i = x_i, S_i = s_i, W = w) \\ &\qquad\qquad \cdot \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, S_i = s_i, W = w] \Big\} \\ &= \sum_{x_i, s_i, w} \Big\{ \mathbb{P}_{\sigma_i'}(X_i = x_i \mid S_i = s_i) \mathbb{P}(S_i = s_i, W = w) \\ &\qquad\qquad \cdot \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, W = w] \Big\}, \end{aligned}$$

where we have used the fact that X_i is independent from W given S_i, and $\boldsymbol{X}_{-i}$ is independent from S_i given X_i and W. Let $\overline{R}_i(x_i; w)$ denote $\mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, W = w]$ for $x_i, w \in \{0, 1\}$. Then $\overline{R}_i(x_i; w)$ does not depend on the strategy of individual i since

$$\begin{aligned} &\mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, W = w] \\ &= \mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(x_i, \boldsymbol{X}_{-i}) \mid X_i = x_i, W = w] \\ &= \mathbb{E}_{\boldsymbol{\sigma}_{-i}}[R_i(x_i, \boldsymbol{X}_{-i}) \mid W = w], \end{aligned}$$

where the last equality follows from the conditional independence between X_i and $\boldsymbol{X}_{-i}$ given W. With this notation,

$$\begin{aligned} &\mathbb{E}_{(\sigma_i', \boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X})] \\ &= \sum_{x_i, s_i} \Big\{ \mathbb{P}_{\sigma_i'}(X_i = x_i \mid S_i = s_i) \\ &\qquad\qquad \cdot \sum_{w} \mathbb{P}(S_i = s_i, W = w) \overline{R}_i(x_i; w) \Big\}. \end{aligned}$$

Therefore, the expected utility of individual i is given by

$$\begin{aligned}&U_i(p_1,p_0,q_1,q_0)\\ &=\mathbb{E}_{(\sigma_i',\boldsymbol{\sigma}_{-i})}[R_i(\boldsymbol{X})-g(\zeta(\sigma_i'))]\\ &=K_1p_1+K_0p_0+L_1q_1+L_0q_0-g(\zeta(p_1,p_0,q_1,q_0)),\end{aligned}$$

where

$$\begin{aligned}K_{s_i}&=\sum_w \mathbb{P}(S_i=s_i,W=w)\overline{R}_i(1;w),\ \ s_i\in\{0,1\},\\ L_{s_i}&=\sum_w \mathbb{P}(S_i=s_i,W=w)\overline{R}_i(0;w),\ \ s_i\in\{0,1\}.\end{aligned}$$

Consider a genie-aided mechanism $\widehat{\boldsymbol{R}}$ defined as follows: for any individual i,

$$\widehat{R}_i(x_i,w)=\overline{R}_i(x_i;w),\ \ x_i\in\mathcal{X},w\in\{0,1\}.$$

Still consider any individual i and any strategy σ_i' of individual i. Let $\widehat{U}_i(p_1,p_0,q_1,q_0)$ denote the expected utility of individual i at the strategy σ_i' when other individuals follow $\boldsymbol{\sigma}_{-i}$. Then

$$\begin{aligned}&\widehat{U}_i(p_1,p_0,q_1,q_0)\\ &=\mathbb{E}_{(\sigma_i',\boldsymbol{\sigma}_{-i})}\left[\widehat{R}_i(X_i,W)-g(\zeta(p_1,p_0,q_1,q_0))\right]\\ &=\sum_{x_i,s_i,w}\mathbb{P}_{\sigma_i'}(X_i=x_i\mid S_i=s_i)\mathbb{P}(S_i=s_i,W=w)\widehat{R}_i(x_i,w)\\ &\quad -g(\zeta(p_1,p_0,q_1,q_0))\\ &=K_1p_1+K_0p_0+L_1q_1+L_0q_0-g(\zeta(p_1,p_0,q_1,q_0)),\end{aligned}$$

where the last equality follows from the definition of $\widehat{R}_i(x_i,w)$. Thus, $\widehat{U}_i(p_1,p_0,q_1,q_0)=U_i(p_1,p_0,q_1,q_0)$. Since $\boldsymbol{\sigma}$ is a Nash equilibrium of $\boldsymbol{R}$, the (p_1,p_0,q_1,q_0) that corresponds to σ_i maximizes $U_i(p_1,p_0,q_1,q_0)$, which implies that σ_i is also a best response of individual i under the genie-aided mechanism $\widehat{\boldsymbol{R}}$ when other individuals follow $\boldsymbol{\sigma}_{-i}$. Therefore, $\boldsymbol{\sigma}$ is also a Nash equilibrium of $\widehat{\boldsymbol{R}}$, and the expected payment to each individual at this equilibrium is the same under $\boldsymbol{R}$ and $\widehat{\boldsymbol{R}}$. □

APPENDIX E

PROOF OF LEMMA 4

Proof. We first prove that $\widetilde{\epsilon}$ is well-defined. Let a function $r\colon (0,+\infty) \to \mathbb{R}$ be defined as

$$r(\epsilon) = \frac{D(\epsilon)}{V_{\mathrm{LB}}(\epsilon)}.$$

Let $P_1^{(\epsilon)}$ and $P_0^{(\epsilon)}$ be the conditional distributions of the reported X_i at the ϵ-strategy given $W = 1$ and $W = 0$, respectively, and let P_{U} be the uniform distribution on $\{0,1\}$. Then note that

$$\begin{aligned} D(\epsilon) &= D_{\mathrm{KL}}(P_{\mathrm{U}}||P_1^{(\epsilon)}) = D_{\mathrm{KL}}(P_{\mathrm{U}}||P_0^{(\epsilon)}) \\ &= \frac{1}{2}\ln\frac{(e^{\epsilon}+1)^2}{4(\theta e^{\epsilon}+1-\theta)((1-\theta)e^{\epsilon}+\theta)}. \end{aligned}$$

Therefore, the function r is continuous on $(0,+\infty)$. Further, the function r attains its maximum value in a bounded subset of $(0,+\infty)$ since for any $\epsilon \in (0,+\infty)$, $r(\epsilon) > 0$, and

$$\begin{aligned} \lim_{\epsilon\to 0} r(\epsilon) &= 0 \\ \lim_{\epsilon\to +\infty} r(\epsilon) &= 0. \end{aligned}$$

The set $\arg\max r(\epsilon)$ is a closed set since it is the inverse image of one point. Therefore, $\widetilde{\epsilon} = \inf\{\arg\max r(\epsilon)\}$ is well-defined.

Now consider any feasible $(N, \epsilon_1, \epsilon_2, \ldots, \epsilon_N)$ of (P1). By the construction of $\widetilde{\epsilon}$, for any individual i,

$$V_{\mathrm{LB}}(\epsilon_i) \geq \frac{V_{\mathrm{LB}}(\widetilde{\epsilon})}{D(\widetilde{\epsilon})} D(\epsilon_i).$$

Then

$$\begin{aligned} \sum_{i=1}^{N} V_{\mathrm{LB}}(\epsilon_i) &\geq \frac{V_{\mathrm{LB}}(\widetilde{\epsilon})}{D(\widetilde{\epsilon})} \sum_{i=1}^{N} D(\epsilon_i) \\ &\geq \frac{V_{\mathrm{LB}}(\widetilde{\epsilon})}{D(\widetilde{\epsilon})} \ln(1/\tau), \end{aligned}$$

where the second inequality follows from the feasibility of $(N, \epsilon_1, \epsilon_2, \ldots, \epsilon_N)$. By the construction of $\widetilde{N}$,

$$\widetilde{N} < \frac{\ln(1/\tau)}{D(\widetilde{\epsilon})} + 1.$$

Therefore,

$$\sum_{i=1}^{N} V_{\mathrm{LB}}(\epsilon_i) \geq (\widetilde{N} - 1) V_{\mathrm{LB}}(\widetilde{\epsilon}),$$

which completes the proof. □

APPENDIX F

PROOF OF THEOREM 7

Proof. We write $\boldsymbol{R}$ to represent the mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\text{th}},\epsilon)}$ for conciseness in this proof. Consider any individual i and any strategy σ_i'. Given $C_i = c_i$, let

$$\begin{aligned} p_1 &= \mathbb{P}_{\sigma_i'}(X_i = 1 \mid C_i = c_i, S_i = 1), \quad p_0 = \mathbb{P}_{\sigma_i'}(X_i = 1 \mid C_i = c_i, S_i = 0), \\ q_1 &= \mathbb{P}_{\sigma_i'}(X_i = 0 \mid C_i = c_i, S_i = 1), \quad q_0 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid C_i = c_i, S_i = 0). \end{aligned}$$

Consider the function

$$\begin{aligned} \zeta(p_1, p_0, q_1, q_0) = \max\Bigg\{ & \left|\ln\frac{p_1}{p_0}\right|, \left|\ln\frac{1-p_1}{1-p_0}\right|, \left|\ln\frac{q_1}{q_0}\right|, \left|\ln\frac{1-q_1}{1-q_0}\right|, \\ & \left|\ln\frac{1-p_1-q_1}{1-p_0-q_0}\right|, \left|\ln\frac{p_1+q_1}{p_0+q_0}\right|\Bigg\}. \end{aligned}$$

Then $\zeta(p_1, p_0, q_1, q_0)$ is the differential privacy level of σ_i' at c_i. When other individuals follow σ_{-i}, let J_{-i} be the number of participants among individuals other than individual i. Then the expected utility of individual i can be written as

$$\begin{aligned} &\mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) - g(C_i, \sigma_i') \mid C_i = c_i] \\ &= \sum_{n=1}^{N} \mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) \mid J_{-i} = n-1, C_i = c_i] \cdot \mathbb{P}_{\boldsymbol{\sigma}}(J_{-i} = n-1 \mid C_i = c_i) \\ &\quad - c_i \zeta(p_1, p_0, q_1, q_0). \end{aligned}$$

Next we derive the form of the utility as a function of p_1, p_0, q_1 and q_0. According to σ_{-i},

$$\{J_{-i} = n-1\} = \left\{\sum_{j \neq i} \mathbb{1}_{\{C_j \leq c_{\text{th}}\}} = n-1\right\}, \tag{F.1}$$

where $\mathbb{1}_E$ with E, an arbitrary event in the probability space, is the indicator function of E. Therefore,

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\sigma}}(J_{-i} = n-1 \mid C_i = c_i) &= \mathbb{P}_{\boldsymbol{\sigma}}\left(\sum_{j \neq i} \mathbb{1}_{\{C_j \leq c_{\text{th}}\}} = n-1 \mid C_i = c_i\right) \\ &= \mathbb{P}\left(\sum_{j \neq i} \mathbb{1}_{\{C_j \leq c_{\text{th}}\}} = n-1\right), \end{aligned}$$

which does not depend on the strategy of individual i. When $n = 1$, $\mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) \mid J_{-i} = n-1, C_i = c_i] = 0$, and $\mathbb{P}_{\boldsymbol{\sigma}}(J_{-i} = n-1 \mid C_i = c_i) = 1 - P_{\geq 1}$, where recall that $P_{\geq 1} = 1 - (1 - F_C(c_{\text{th}}))^{N-1}$. When $n > 1$, utilizing the equivalence relation (F.1) and the prior,

$$\begin{aligned} &\mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) \mid J_{-i} = n-1, C_i = c_i] \\ = &\sum_{x_i \in \{0,1,\perp\}, v \in \{0,1\}} \mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, W = w, J_{-i} = n-1, C_i = c_i] \end{aligned}$$

$$
\begin{aligned}
&\qquad\qquad\quad \cdot \mathbb{P}_{\sigma_i'}(X_i = x_i, W = w \mid C_i = c_i) \\
&= \sum_{x_i \in \{0,1,\perp\}, v \in \{0,1\}} \Bigg\{ \mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, W = w, J_{-i} = n-1] \cdot \mathbb{P}(W = w) \\
&\qquad\qquad \cdot \sum_{S_i \in \{0,1\}} \mathbb{P}(S_i = s_i \mid W = w) \mathbb{P}_{\sigma_i'}(X_i = x_i \mid C_i = c_i, S_i = s_i) \Bigg\}.
\end{aligned}
$$

Let

$$
\begin{aligned}
\widehat{\boldsymbol{R}}_i(x_i, v, n) = \mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, W = w, J_{-i} = n-1], \\
x_i \in \{0,1,\perp\}, v \in \{0,1\}, 1 < n \leq N.
\end{aligned}
$$

Then $\widehat{\boldsymbol{R}}_i(x_i, v, n)$ does not depend on the strategy of individual i since

$$
\begin{aligned}
&\mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) \mid X_i = x_i, W = w, J_{-i} = n-1] \\
&= \mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(x_i, X_{-i}) \mid X_i = x_i, W = w, J_{-i} = n-1] \\
&= \mathbb{E}_{\sigma_{-i}}[R_i(x_i, X_{-i}) \mid W = w, J_{-i} = n-1].
\end{aligned}
$$

The value of $\widehat{\boldsymbol{R}}_i(x_i, v, n)$ can be calculated from the description of the mechanism. With this notation,

$$
\begin{aligned}
&\mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) \mid J_{-i} = n-1, C_i = c_i] \\
&= \widehat{\boldsymbol{R}}_i(1,1,n) P_W(1) \big(\theta p_1 + (1-\theta) p_0\big) \\
&\quad + \widehat{\boldsymbol{R}}_i(0,1,n) P_W(1) \big(\theta q_1 + (1-\theta) q_0\big) \\
&\quad + \widehat{\boldsymbol{R}}_i(1,0,n) P_W(0) \big((1-\theta) p_1 + \theta p_0\big) \\
&\quad + \widehat{\boldsymbol{R}}_i(0,0,n) P_W(0) \big((1-\theta) q_1 + \theta q_0\big) \\
&= \frac{1}{P_{\geq 1}} (K_1 p_1 + K_0 p_0 + L_1 q_1 + L_0 q_0),
\end{aligned} \tag{F.2}
$$

where

$$
\begin{aligned}
K_1 &= P_{\geq 1} \Big(\widehat{\boldsymbol{R}}_i(1,1,n) P_W(1) \theta + \widehat{\boldsymbol{R}}_i(1,0,n) P_W(0)(1-\theta) \Big), \\
K_0 &= P_{\geq 1} \Big(\widehat{\boldsymbol{R}}_i(1,1,n) P_W(1)(1-\theta) + \widehat{\boldsymbol{R}}_i(1,0,n) P_W(0) \theta \Big), \\
L_1 &= P_{\geq 1} \Big(\widehat{\boldsymbol{R}}_i(0,1,n) P_W(1) \theta + \widehat{\boldsymbol{R}}_i(0,0,n) P_W(0)(1-\theta) \Big), \\
L_0 &= P_{\geq 1} \Big(\widehat{\boldsymbol{R}}_i(0,1,n) P_W(1)(1-\theta) + \widehat{\boldsymbol{R}}_i(0,0,n) P_W(0) \theta \Big),
\end{aligned}
$$

and we have used the fact that $\widehat{\boldsymbol{R}}_i(\perp, v, n) = 0$ for any v and n. Note that K_1, K_0, L_1 and L_0 do not depend on p_1, p_0, q_1 and q_0. By the description of the mechanism,

$$
K_1 = L_0 = \frac{1}{2} \bigg(\frac{c_{\text{th}}(e^{\epsilon} + 1)}{e^{\epsilon}} + c_{\text{th}} \epsilon \bigg),
$$

$$K_0 = L_1 = \frac{1}{2}\biggl(\frac{c_{\text{th}}(e^\epsilon + 1)}{e^\epsilon} + c_{\text{th}}\epsilon - \frac{c_{\text{th}}(e^\epsilon + 1)^2}{e^\epsilon}\biggr).$$

We can see that K_1, K_0, L_1 and L_0 do not depend on n, either. Therefore, combining the case that $n = 1$ and (F.2), the expected utility of individual i can be written as

$$\begin{aligned}&\mathbb{E}_{(\sigma_i', \sigma_{-i})}[R_i(\boldsymbol{X}) - g(C_i, \sigma_i') \mid C_i = c_i]\\ &= K_1 p_1 + K_0 p_0 + L_1 q_1 + L_0 q_0 - c_i \zeta(p_1, p_0, q_1, q_0).\end{aligned}$$

Let this utility define a function U of p_1, p_0, q_1, and q_0; i.e.,

$$U(p_1, p_0, q_1, q_0) = K_1 p_1 + K_0 p_0 + L_1 q_1 + L_0 q_0 - c_i \zeta(p_1, p_0, q_1, q_0).$$

Now we find the best response of individual i, i.e., an optimal solution of the following optimization problem:

$$\begin{aligned}\max_{p_1, p_0, q_1, q_0} \quad & U(p_1, p_0, q_1, q_0)\\ \text{subject to} \quad & 0 \le p_1 \le 1, 0 \le q_1 \le 1,\\ & 0 \le p_1 + q_1 \le 1,\\ & 0 \le p_0 \le 1, 0 \le q_0 \le 1,\\ & 0 \le p_0 + q_0 \le 1,\end{aligned}$$

by the following three steps.

Step 1: First we can focus on an optimal solution $(p_1^*, p_0^*, q_1^*, q_0^*)$ such that $p_1^* = q_0^*$ and $p_0^* = q_1^*$ for the following reasons. For any feasible solution (p_1, p_0, q_1, q_0), consider the solution (p_1', p_0', q_1', q_0') given by

$$p_1' = q_0' = \frac{p_1 + q_0}{2}, \quad p_0' = q_1' = \frac{p_0 + q_1}{2}.$$

Then since $K_1 = L_0$ and $K_0 = L_1$,

$$K_1 p_1' + K_0 p_0' + L_1 q_1' + L_0 q_0' = K_1 p_1 + K_0 p_0 + L_1 q_1 + L_0 q_0.$$

By the definition of the function ζ,

$$\begin{aligned} p_0 e^{-\zeta(p_1, p_0, q_1, q_0)} &\le p_1 \le p_0 e^{\zeta(p_1, p_0, q_1, q_0)},\\ (1 - p_0) e^{-\zeta(p_1, p_0, q_1, q_0)} &\le 1 - p_1 \le (1 - p_0) e^{\zeta(p_1, p_0, q_1, q_0)},\\ q_0 e^{-\zeta(p_1, p_0, q_1, q_0)} &\le q_1 \le q_0 e^{\zeta(p_1, p_0, q_1, q_0)},\\ (1 - q_0) e^{-\zeta(p_1, p_0, q_1, q_0)} &\le 1 - q_1 \le (1 - q_0) e^{\zeta(p_1, p_0, q_1, q_0)}.\end{aligned}$$

Then it is not hard to verify that

$$\begin{aligned} p_0' e^{-\zeta(p_1, p_0, q_1, q_0)} &\le p_1' \le p_0' e^{\zeta(p_1, p_0, q_1, q_0)},\\ (1 - p_0') e^{-\zeta(p_1, p_0, q_1, q_0)} &\le 1 - p_1' \le (1 - p_0') e^{\zeta(p_1, p_0, q_1, q_0)},\end{aligned}$$

$$q_0' e^{-\zeta(p_1,p_0,q_1,q_0)} \le q_1' \le q_0' e^{\zeta(p_1,p_0,q_1,q_0)},$$
$$(1-q_0') e^{-\zeta(p_1,p_0,q_1,q_0)} \le 1-q_1' \le (1-q_0') e^{\zeta(p_1,p_0,q_1,q_0)}.$$

Besides,

$$1-p_1'-q_1' = 1-p_0'-q_0', \quad p_1'+q_1' = p_0'+q_0'.$$

Thus

$$\zeta(p_1',p_0',q_1',q_0') \le \zeta(p_1,p_0,q_1,q_0),$$

and

$$U(p_1',p_0',q_1',q_0') \ge U(p_1,p_0,q_1,q_0).$$

Further, an optimal solution $(p_1^*,p_0^*,q_1^*,q_0^*)$ such that $p_1^*=q_0^*$ and $p_0^*=q_1^*$ must satisfy that $p_1^* \ge q_1^*$, since otherwise by swapping p_1^* and p_0^* with q_1^* and q_0^*, respectively, the utility is increased, which contradicts with the optimality of $(p_1^*,p_0^*,q_1^*,q_0^*)$.

Step 2: Next, for any such an optimal solution, i.e., $(p_1^*,p_0^*,q_1^*,q_0^*)$ with $p_1^*=q_0^*$ and $p_0^*=q_1^*$, we prove that one of the following two holds

$$p_1^* = q_0^* = p_0^* = q_1^* = 0, \text{ or} \tag{F.3}$$
$$p_1^*+q_1^* = p_0^*+q_0^* = 1, p_1^* > q_1^*. \tag{F.4}$$

Suppose not. Since $(p_1,p_0,q_1,q_0)=(0,0,0,0)$ is a feasible solution, $U(p_1^*,p_0^*,q_1^*,q_0^*) \ge U(p_1,p_0,q_1,q_0) \ge 0$, and thus $K_1p_1^*+K_0p_0^*+L_1q_1^*+L_0q_0^* \ge 0$. Suppose that $K_1p_1^*+K_0p_0^*+L_1q_1^*+L_0q_0^* = 0$. Then since $U(p_1^*,p_0^*,q_1^*,q_0^*) \ge 0$, we have $\zeta(p_1^*,p_0^*,q_1^*,q_0^*)=0$, which implies that $p_1^*=p_0^*$ and $q_1^*=q_0^*$. Thus,

$$K_1p_1^*+K_0p_0^*+L_1q_1^*+L_0q_0^* = \frac{c_{\text{th}}}{2}(e^{-\epsilon}+2\epsilon-e^{\epsilon})(p_1^*+q_1^*).$$

Since $e^{-\epsilon}+2\epsilon-e^{\epsilon}<0$ for any $\epsilon>0$, it must be that $p_1^*+q_1^*=0$, which contradicts with the assumption that $(p_1^*,p_0^*,q_1^*,q_0^*)$ does not satisfy (F.3). Therefore, $K_1p_1^*+K_0p_0^*+L_1q_1^*+L_0q_0^*>0$. Since $(p_1^*,p_0^*,q_1^*,q_0^*)$ does not satisfy (F.4), either $p_1^*+q_1^*=p_0^*+q_0^*<1$ or $p_1^*=q_1^*$. If $p_1^*+q_1^*=p_0^*+q_0^*<1$, consider the solution (p_1',p_0',q_1',q_0') given by

$$p_1' = \frac{p_1^*}{p_1^*+q_1^*}, \quad p_0' = \frac{p_0^*}{p_0^*+q_0^*},$$
$$q_1' = \frac{q_1^*}{p_1^*+q_1^*}, \quad q_0' = \frac{q_0^*}{p_0^*+q_0^*}.$$

Then

$$\begin{aligned} K_1p_1'+K_0p_0'+L_1q_1'+L_0q_0' &= \frac{1}{p_1^*+q_1^*}(K_1p_1^*+K_0p_0^*+L_1q_1^*+L_0q_0^*) \\ &> K_1p_1^*+K_0p_0^*+L_1q_1^*+L_0q_0^*. \end{aligned}$$

However, $\zeta(p_1',p_0',q_1',q_0') = \zeta(p_1^*,p_0^*,q_1^*,q_0^*)$. Thus $U(p_1',p_0',q_1',q_0') > U(p_1^*,p_0^*,q_1^*,q_0^*)$, which contradicts with the optimality of $(p_1^*,p_0^*,q_1^*,q_0^*)$. If $p_1^*=q_1^*$, then $p_1^*=q_0^*=p_0^*=q_1^*$ and

$$K_1p_1^*+K_0p_0^*+L_1q_1^*+L_0q_0^* = c_{\text{th}}(e^{-\epsilon}+2\epsilon-e^{\epsilon})p_1^* < 0,$$

which contradicts with the fact that $K_1p_1^* + K_0p_0^* + L_1q_1^* + L_0q_0^* > 0$. In summary, for any optimal solution $(p_1^*, p_0^*, q_1^*, q_0^*)$ with $p_1^* = q_0^*$ and $p_0^* = q_1^*$, either (F.3) or (F.4) holds.

Step 3: According to Step 1 and Step 2, we can find an optimal solution among those feasible solutions, say (p_1, p_0, q_1, q_0), with $p_1 = q_0$ and $p_0 = q_1$, that satisfy either

$$p_1 = q_0 = p_0 = q_1 = 0, \text{ or} \tag{F.5}$$

$$p_1 + q_1 = p_0 + q_0 = 1, p_1 > q_1. \tag{F.6}$$

Consider any feasible solution (p_1, p_0, q_1, q_0) with $p_1 = q_0$ and $p_0 = q_1$ and satisfies (F.6), which can be written as

$$p_1 = q_0 = \frac{e^{\epsilon_i}}{e^{\epsilon_i} + 1}, \quad p_0 = q_1 = \frac{1}{e^{\epsilon_i} + 1},$$

for some $\epsilon_i > 0$. Then

$$U(p_1, p_0, q_1, q_0) = -\frac{c_{\text{th}}(e^{\epsilon} + 1)^2}{e^{\epsilon}} \frac{1}{e^{\epsilon_i} + 1} - c_i\epsilon_i + \frac{c_{\text{th}}(e^{\epsilon} + 1)}{e^{\epsilon}} + c_{\text{th}}\epsilon.$$

Consider a function $h\colon (0, +\infty) \to \mathbb{R}$ defined as

$$h(\epsilon_i) = -\frac{c_{\text{th}}(e^{\epsilon} + 1)^2}{e^{\epsilon}} \frac{1}{e^{\epsilon_i} + 1} - c_i\epsilon_i. \tag{F.7}$$

Then

$$h'(\epsilon_i) = \frac{c_{\text{th}}(e^{\epsilon} + 1)^2}{e^{\epsilon}} \frac{e^{\epsilon_i}}{(e^{\epsilon_i} + 1)^2} - c_i,$$

$$h''(\epsilon_i) = -\frac{c_{\text{th}}(e^{\epsilon} + 1)^2}{e^{\epsilon}} \frac{e^{\epsilon_i}(e^{\epsilon_i} - 1)}{(e^{\epsilon_i} + 1)^3} < 0.$$

Thus, ϵ_i^* that satisfies

$$\frac{c_{\text{th}}(e^{\epsilon} + 1)^2}{e^{\epsilon}} \frac{e^{\epsilon_i^*}}{(e^{\epsilon_i^*} + 1)^2} - c_i = 0,$$

i.e., $\epsilon_i^* = \xi(c_i)$ defined in (4.1), maximizes $h(\cdot)$, and hence maximizes the utility. Therefore, among those feasible solutions that satisfy (F.6), the solution $(\widetilde{p}_1^*, \widetilde{p}_0^*, \widetilde{q}_1^*, \widetilde{q}_0^*)$ with

$$\widetilde{p}_1^* = \widetilde{q}_0^* = \frac{e^{\xi(c_i)}}{e^{\xi(c_i)} + 1}, \quad \widetilde{p}_0^* = \widetilde{q}_1^* = \frac{1}{e^{\xi(c_i)} + 1}$$

maximizes the utility. This implies that an optimal solution is either $(0, 0, 0, 0)$ or $(\widetilde{p}_1^*, \widetilde{p}_0^*, \widetilde{q}_1^*, \widetilde{q}_0^*)$. In the remainder of this step, we prove that if $c_i > c_{\text{th}}$, $(0, 0, 0, 0)$ is an optimal solution, and otherwise, i.e., $c_i \le c_{\text{th}}$, $(\widetilde{p}_1^*, \widetilde{p}_0^*, \widetilde{q}_1^*, \widetilde{q}_0^*)$ is an optimal solution. First consider the case that $c_i > c_{\text{th}}$. Then

$$U(\widetilde{p}_1^*, \widetilde{p}_0^*, \widetilde{q}_1^*, \widetilde{q}_0^*) = -\frac{c_{\text{th}}(e^{\epsilon} + 1)^2}{e^{\epsilon}} \frac{1}{e^{\xi(c_i)} + 1} + \frac{c_{\text{th}}(e^{\epsilon} + 1)}{e^{\epsilon}} + c_{\text{th}}\epsilon$$

$$
\begin{aligned}
&- \frac{c_{\mathrm{th}}(e^\epsilon+1)^2}{e^\epsilon}\frac{\xi(c_i)e^{\xi(c_i)}}{(e^{\xi(c_i)}+1)^2}\\
&= \frac{c_{\mathrm{th}}(e^\epsilon+1)^2}{e^\epsilon}\left(\frac{1}{e^\epsilon+1}+\frac{\epsilon e^\epsilon}{(e^\epsilon+1)^2}-\frac{1}{e^{\xi(c_i)}+1}-\frac{\xi(c_i)e^{\xi(c_i)}}{(e^{\xi(c_i)}+1)^2}\right).
\end{aligned}
$$

Consider a function $z\colon (0,+\infty)\to\mathbb{R}$ defined as

$$
z(\epsilon_i)=\frac{1}{e^\epsilon+1}+\frac{\epsilon e^\epsilon}{(e^\epsilon+1)^2}-\frac{1}{e^{\epsilon_i}+1}-\frac{\epsilon_i e^{\epsilon_i}}{(e^{\epsilon_i}+1)^2}.
$$

Then

$$
\begin{aligned}
&z(\epsilon)=0,\\
&z'(\epsilon_i)=\frac{\epsilon_i e^{\epsilon_i}(e^{\epsilon_i}-1)}{(e^{\epsilon_i}+1)^3}>0.
\end{aligned}
$$

Thus, for any $\epsilon_i<\epsilon$, $z(\epsilon_i)<0$. Since

$$
U(\widetilde{p}_1^*,\widetilde{p}_0^*,\widetilde{q}_1^*,\widetilde{q}_0^*)=\frac{c_{\mathrm{th}}(e^\epsilon+1)^2}{e^\epsilon}z(\xi(c_i)),
$$

and $\xi(c_i)<\epsilon$ due to $c_i>c_{\mathrm{th}}$, we have $U(\widetilde{p}_1^*,\widetilde{p}_0^*,\widetilde{q}_1^*,\widetilde{q}_0^*)<0=U(0,0,0,0)$. Therefore, for the case that $c_i>c_{\mathrm{th}}$, $(0,0,0,0)$ is an optimal solution. Next consider the case that $c_i\le c_{\mathrm{th}}$. Write the utility as

$$
U(\widetilde{p}_1^*,\widetilde{p}_0^*,\widetilde{q}_1^*,\widetilde{q}_0^*)=h(\xi(c_i))+\frac{c_{\mathrm{th}}(e^\epsilon+1)}{e^\epsilon}+c_{\mathrm{th}}\epsilon,
$$

where the function h is defined in (F.7). Since

$$
h(\epsilon)+\frac{c_{\mathrm{th}}(e^\epsilon+1)}{e^\epsilon}+c_{\mathrm{th}}\epsilon=(c_{\mathrm{th}}-c_i)\epsilon\ge 0,
$$

and $\xi(c_i)$ maximizes $h(\cdot)$, we have $U(\widetilde{p}_1^*,\widetilde{p}_0^*,\widetilde{q}_1^*,\widetilde{q}_0^*)\ge 0=U(0,0,0,0)$. Therefore, for the case that $c_i\le c_{\mathrm{th}}$, $(\widetilde{p}_1^*,\widetilde{p}_0^*,\widetilde{q}_1^*,\widetilde{q}_0^*)$ is an optimal solution.

In summary, by the three steps above, a best response of individual i is described as follows:

- If $c_i>c_{\mathrm{th}}$,

$$
\mathbb{P}_{\sigma_i}(X_i=\perp\mid C_i=c_i,S_i=s_i)=1,\quad\text{for any } S_i\in\{0,1\}.
$$

- If $c_i\le c_{\mathrm{th}}$,

$$
\begin{aligned}
&\mathbb{P}_{\sigma_i}(X_i=1\mid C_i=c_i,S_i=1)=\mathbb{P}_{\sigma_i}(X_i=0\mid C_i=c_i,S_i=0)=\frac{e^{\xi(c_i)}}{e^{\xi(c_i)}+1},\\
&\mathbb{P}_{\sigma_i}(X_i=0\mid C_i=c_i,S_i=1)=\mathbb{P}_{\sigma_i}(X_i=1\mid C_i=c_i,S_i=0)=\frac{1}{e^{\xi(c_i)}+1},
\end{aligned}
$$

  where $\xi(c_i)$ is defined in (4.1).

This completes the proof that $\boldsymbol{\sigma}$ is a Bayesian Nash equilibrium of the mechanism $\boldsymbol{R}^{(N,\mathcal{P},c_{\mathrm{th}},\epsilon)}$. □

APPENDIX G

PROOF OF THEOREM 8

Proof. Let the parameter tuple $(N, c_{\text{th}}, \epsilon)$ be chosen according to (4.5)–(4.11). Then c_{th} is a function of N and ϵ. We write $n_e, \rho, p_{\text{th}}, c_{\text{th}}$ to represent $n_e(\epsilon)$, $\rho(\epsilon)$, $p_{\text{th}}(N, \epsilon)$, $c_{\text{th}}(N, \epsilon)$ and keep their dependence on N, ϵ in mind for conciseness in this proof.

We first derive the form of the maximum likelihood decision function ψ. For any realization $\boldsymbol{X} = x$, since ψ uses maximum likelihood,

$$\psi(x) = \begin{cases} 1 & \text{if } \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1) \geq \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0), \\ 0 & \text{otherwise.} \end{cases}$$

Let $\mathcal{A}(x) = \{i \in \mathcal{N} \colon x_i \neq \perp\}$. By Theorem 7,

$$\begin{aligned} &\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1) \\ &= \prod_{i \in \mathcal{A}(x)} \mathbb{P}_{\sigma_i}(X_i = x_i, C_i \leq c_{\text{th}} \mid W = 1) \cdot \prod_{j \notin \mathcal{A}(x)} \mathbb{P}(C_j > c_{\text{th}}) \\ &= \prod_{i \in \mathcal{A}(x)} \mathbb{P}_{\sigma_i}(X_i = x_i \mid C_i \leq c_{\text{th}}, W = 1) \mathbb{P}(C_i \leq c_{\text{th}}) \cdot \prod_{j \notin \mathcal{A}(x)} \mathbb{P}(C_j > c_{\text{th}}) \\ &= \prod_{i \in \mathcal{A}(x)} \alpha^{x_i} (1 - \alpha)^{1 - x_i} \mathbb{P}(C_i \leq c_{\text{th}}) \cdot \prod_{j \notin \mathcal{A}(x)} \mathbb{P}(C_j > c_{\text{th}}), \end{aligned}$$

where recall that α is defined as in (4.2). Similarly,

$$\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0) = \prod_{i \in \mathcal{A}(x)} (1 - \alpha)^{x_i} \alpha^{1 - x_i} \mathbb{P}(C_i \leq c_{\text{th}}) \cdot \prod_{j \notin \mathcal{A}(x)} \mathbb{P}(C_j > c_{\text{th}}).$$

By Corollary 1,

$$\alpha = \theta\mu + (1 - \theta)(1 - \mu) \geq \theta \frac{e^{\epsilon}}{e^{\epsilon} + 1} + (1 - \theta) \frac{1}{e^{\epsilon} + 1} > \frac{1}{2}.$$

Thus, the condition $\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1) \geq \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0)$ is equivalent to the condition that the number of 1's is larger than or equal to the number of 0's in x. Therefore,

$$\psi(\boldsymbol{X}) = \begin{cases} 1 & \text{if } \sum_i \mathbb{1}_{\{X_i = 1\}} \geq \sum_i \mathbb{1}_{\{X_i = 0\}}, \\ 0 & \text{otherwise.} \end{cases}$$

Next we calculate the probability of error, p_e. Let

$$k = \sqrt{\frac{2}{p_e^{\max}}}, \quad d = \sqrt{N p_{\text{th}} (1 - p_{\text{th}})}.$$

By definition,

$$\begin{aligned} p_e &= \mathbb{P}_{\boldsymbol{\sigma}}(\psi(\boldsymbol{X}) \neq W) \\ &= \mathbb{P}_{\boldsymbol{\sigma}}\left(\left| \sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - N p_{\text{th}} \right| \geq kd, \psi(\boldsymbol{X}) \neq W \right) \end{aligned}$$

$$+ \mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\mathrm{th}}\right| < kd, \psi(\boldsymbol{X}) \neq W\right).$$

Since the random variables $\mathbb{1}_{\{X_i \neq \perp\}} = \mathbb{1}_{\{C_i \leq c_{\mathrm{th}}\}}$ are i.i.d. with mean p_{th} and variance $\frac{d^2}{N}$, by Chebyshev's inequality,

$$\begin{aligned}
\mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\mathrm{th}}\right| \geq kd, \psi(\boldsymbol{X}) \neq W\right) &\leq \mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\mathrm{th}}\right| \geq kd\right) \\
&\leq \frac{1}{k^2} \\
&= \frac{p_e^{\max}}{2}.
\end{aligned}$$

For the second part of p_e, we have

$$\begin{aligned}
&\mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\mathrm{th}}\right| < kd, \psi(\boldsymbol{X}) \neq W\right) \\
&\leq \mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\mathrm{th}}\right| < kd, \psi(\boldsymbol{X}) \neq W \;\middle|\; W = 1\right) \\
&\quad + \mathbb{P}_{\boldsymbol{\sigma}}\left(\left|\sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\mathrm{th}}\right| < kd, \psi(\boldsymbol{X}) \neq W \;\middle|\; W = 0\right) \\
&= \sum_{x \in \mathcal{B} \cap \mathcal{R}_1} \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1) + \sum_{x \in \mathcal{B} \cap \mathcal{R}_0} \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0),
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{B} &= \left\{x \in \mathcal{X}^N \colon \left||\mathcal{A}(x)| - Np_{\mathrm{th}}\right| < kd\right\}, \\
\mathcal{R}_1 &= \left\{x \in \mathcal{X}^N \colon \psi(x) \neq 1\right\}, \\
\mathcal{R}_0 &= \left\{x \in \mathcal{X}^N \colon \psi(x) \neq 0\right\},
\end{aligned}$$

and $|\mathcal{A}(x)|$ is the cardinality of the set $\mathcal{A}(x) = \{i \in \mathcal{N} \colon x_i \neq \perp\}$. Since ψ uses maximum likelihood,

$$\begin{aligned}
&\sum_{x \in \mathcal{B} \cap \mathcal{R}_1} \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1) + \sum_{x \in \mathcal{B} \cap \mathcal{R}_0} \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0) \\
&= \sum_{x \in \mathcal{B}} \min\left\{\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1), \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0)\right\} \\
&\leq \sum_{x \in \mathcal{B}} \sqrt{\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1)\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0)} \\
&= \sum_{x \in \mathcal{B}} \left\{\left(\sqrt{\alpha(1-\alpha)}\right)^{|\mathcal{A}(x)|} \cdot \prod_{i \in \mathcal{A}(x)} \mathbb{P}(C_i \leq c_{\mathrm{th}}) \cdot \prod_{j \neq \mathcal{A}(x)} \mathbb{P}(C_j > c_{\mathrm{th}})\right\}.
\end{aligned}$$

Combining the x's with the same $\mathcal{A}(x)$ yields

$$\begin{aligned}
&\sum_{x\in\mathcal{B}\cap\mathcal{R}_1}\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X}=x\mid W=1)+\sum_{x\in\mathcal{B}\cap\mathcal{R}_0}\mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X}=x\mid W=0)\\
&\le \sum_{\mathcal{I}\subseteq\mathcal{N}:||\mathcal{I}|-Np_{\text{th}}|<kd}\left\{2^{|\mathcal{I}|}\left(\sqrt{\alpha(1-\alpha)}\right)^{|\mathcal{I}|}\cdot\prod_{i\in\mathcal{I}}\mathbb{P}(C_i\le c_{\text{th}})\cdot\prod_{j\neq\mathcal{I}}\mathbb{P}(C_j>c_{\text{th}})\right\}\\
&= \sum_{\mathcal{I}\subseteq\mathcal{N}:||\mathcal{I}|-Np_{\text{th}}|<kd}\left\{e^{-\frac{|\mathcal{I}|}{2}\ln\frac{1}{4\alpha(1-\alpha)}}\cdot\prod_{i\in\mathcal{I}}\mathbb{P}(C_i\le c_{\text{th}})\cdot\prod_{j\neq\mathcal{I}}\mathbb{P}(C_j>c_{\text{th}})\right\}.
\end{aligned}$$

For any $\mathcal{I}\subseteq\mathcal{N}$ such that $\left||\mathcal{I}|-Np_{\text{th}}\right|<kd$,

$$\begin{aligned}
|\mathcal{I}| &> Np_{\text{th}}-kd\\
&=\rho n_e-\sqrt{\frac{2}{p_e^{\max}}}\sqrt{\rho n_e(1-p_{\text{th}})}\\
&>\rho n_e-\sqrt{\frac{2}{p_e^{\max}}}\sqrt{\rho n_e}.
\end{aligned}$$

By the choice of ρ,

$$\begin{aligned}
\sqrt{\rho n_e} &> \sqrt{\frac{1}{p_e^{\max}}+n_e+\sqrt{\frac{1}{(p_e^{\max})^2}+\frac{2n_e}{p_e^{\max}}}}\\
&=\sqrt{\frac{1}{2p_e^{\max}}}+\sqrt{\frac{1}{2p_e^{\max}}+n_e}.
\end{aligned}$$

Thus

$$|\mathcal{I}|>n_e.$$

We have known that

$$\alpha\ge\theta\frac{e^\epsilon}{e^\epsilon+1}+(1-\theta)\frac{1}{e^\epsilon+1}.$$

Combining the two inequalities above yields

$$e^{-\frac{|\mathcal{I}|}{2}\ln\frac{1}{4\alpha(1-\alpha)}}<e^{-\frac{n_e}{2}\ln\frac{(e^\epsilon+1)^2}{4(\theta e^\epsilon+1-\theta)((1-\theta)e^\epsilon+\theta)}}=e^{-n_eD(\epsilon)}.$$

By the choice of n_e,

$$e^{-n_eD(\epsilon)}=\frac{p_e^{\max}}{2}.$$

Hence,

$$\begin{aligned}
&\sum_{x \in \mathcal{B} \cap \mathcal{R}_1} \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 1) + \sum_{x \in \mathcal{B} \cap \mathcal{R}_0} \mathbb{P}_{\boldsymbol{\sigma}}(\boldsymbol{X} = x \mid W = 0) \\
&\le \frac{p_e^{\max}}{2} \sum_{\mathcal{I} \subseteq \mathcal{N} : ||\mathcal{I}| - Np_{\mathrm{th}}| < kd} \left\{ \prod_{i \in \mathcal{I}} \mathbb{P}(C_i \le c_{\mathrm{th}}) \cdot \prod_{j \neq \mathcal{I}} \mathbb{P}(C_j > c_{\mathrm{th}}) \right\} \\
&\le \frac{p_e^{\max}}{2} \sum_{\mathcal{I} \subseteq \mathcal{N}} \left\{ \prod_{i \in \mathcal{I}} \mathbb{P}(C_i \le c_{\mathrm{th}}) \cdot \prod_{j \neq \mathcal{I}} \mathbb{P}(C_j > c_{\mathrm{th}}) \right\} \\
&= \frac{p_e^{\max}}{2}.
\end{aligned}$$

This gives an upper bound on the second part of p_e; i.e.,

$$\mathbb{P}_{\boldsymbol{\sigma}}\left(\left| \sum_{i=1}^{N} \mathbb{1}_{\{X_i \neq \perp\}} - Np_{\mathrm{th}} \right| < kd, \psi(\boldsymbol{X}) \neq W \right) \le \frac{p_e^{\max}}{2}.$$

Therefore,

$$p_e \le p_e^{\max}.$$

Finally, we bound the total expected payment. Let J be the number of participants. By Corollary 1,

$$\mathbb{E}_{\boldsymbol{\sigma}}\left[\sum_{i=1}^{N} R_i^{(N,\mathcal{P},c_{\mathrm{th}},\epsilon)}(\boldsymbol{X}) \;\middle|\; J \right] \le Jc_{\mathrm{th}}(1 + e^{-\epsilon} + \epsilon).$$

By Theorem 7,

$$J = \sum_{i=1}^{N} \mathbb{1}_{\{C_i \le c_{\mathrm{th}}\}}.$$

Then $\mathbb{E}_{\boldsymbol{\sigma}}[J] = Np_{\mathrm{th}} = \rho n_e$. Therefore,

$$\begin{aligned}
\mathbb{E}_{\boldsymbol{\sigma}}\left[\sum_{i=1}^{N} R_i^{(N,\mathcal{P},c_{\mathrm{th}},\epsilon)}(\boldsymbol{X}) \right] &= \mathbb{E}_{\boldsymbol{\sigma}}\left[\mathbb{E}_{\boldsymbol{\sigma}}\left[\sum_{i=1}^{N} R_i^{(N,\mathcal{P},c_{\mathrm{th}},\epsilon)}(\boldsymbol{X}) \;\middle|\; J \right] \right] \\
&\le \mathbb{E}_{\boldsymbol{\sigma}}[J] c_{\mathrm{th}}(1 + e^{-\epsilon} + \epsilon) \\
&= \rho n_e c_{\mathrm{th}}(1 + e^{-\epsilon} + \epsilon).
\end{aligned}$$

The parameters ρ and n_e do not depend on the choice of N. However, by Lemma 6,

$$\lim_{N \to +\infty} c_{\mathrm{th}} = 0.$$

Therefore, the total expected payment goes to zero as the chosen N goes to infinity.
□

APPENDIX H

EQUIVALENCE OF $D \neq 0$ AND DEPENDENCE

In this section we prove that $D \neq 0$ is equivalent to the statement S_i and S_j are not independent for any two distinct players i and j. The direction that $D \neq 0$ implies dependence is obvious since D is the covariance of S_i and S_j.

For the other direction, suppose by contradiction that $D = 0$. Consider any two distinct players i and j. Recall that

$$P_1 = \mathbb{P}(S_i = 1), \quad P_0 = \mathbb{P}(S_i = 0).$$

First notice that $P_1 \neq 0$ and $P_0 \neq 0$ since otherwise S_i and S_j are independent. Then $D = 0$ implies that

$$\begin{aligned} &\mathbb{P}(S_j = 1 \mid S_i = 1)\mathbb{P}(S_j = 0 \mid S_i = 0) \\ &= \mathbb{P}(S_j = 0 \mid S_i = 1)\mathbb{P}(S_j = 1 \mid S_i = 0). \end{aligned} \tag{H.1}$$

Since $\mathbb{P}(S_j = 0 \mid S_i = 1) = 1 - \mathbb{P}(S_j = 1 \mid S_i = 1)$ and $\mathbb{P}(S_j = 1 \mid S_i = 0) = 1 - \mathbb{P}(S_j = 0 \mid S_i = 0)$, (H.1) further implies that

$$\begin{aligned} \mathbb{P}(S_j = 1 \mid S_i = 1) &= 1 - \mathbb{P}(S_j = 0 \mid S_i = 0) \\ &= \mathbb{P}(S_j = 1 \mid S_i = 0). \end{aligned}$$

Similarly,

$$\mathbb{P}(S_j = 0 \mid S_i = 1) = \mathbb{P}(S_j = 0 \mid S_i = 0).$$

Therefore, S_i and S_j are independent, which contradicts with the assumption that they are not independent. This completes the proof.

APPENDIX I

PROOF OF THEOREM 9

Proof. We write $\boldsymbol{R}$ to represent the mechanism $\boldsymbol{R}^{(N,\epsilon)}$ for conciseness in this proof. For any player i and any strategy σ_i', let

$$\begin{aligned} p_1 &= \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 1), \quad p_0 = \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 0), \\ q_1 &= \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 1), \quad q_0 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 0). \end{aligned}$$

We consider the case that $D > 0$. The proof for the case that $D < 0$ is similar.

Suppose that other players follow σ_{-i}^*. Let the payment of player i be computed using the reported data X_j of some other player j. Then the expected payment of player i can be written as

$$\begin{aligned} &\mathbb{E}_{(\sigma_i',\sigma_{-i}^*)}[R_i(X)] \\ &= \sum_{x_i,x_j\in\{0,1\}} R_i(x_i,x_j)\mathbb{P}_{(\sigma_i',\sigma_{-i}^*)}(X_i = x_i, X_j = x_j) \\ &= \sum_{x_i,S_i\in\{0,1\}} \Bigg(\mathbb{P}_{\sigma_i'}(X_i = x_i \mid S_i = s_i) \\ &\qquad\qquad \cdot \sum_{x_j\in\{0,1\}} R_i(x_i,x_j)\mathbb{P}_{\sigma_j^*}(X_j = x_j, S_i = s_i)\Bigg) \\ &= K_1p_1 + K_0p_0 + L_1q_1 + L_0q_0, \end{aligned}$$

where

$$\begin{aligned} K_1 &= \frac{g'(\epsilon)(e^\epsilon+1)^2}{2e^\epsilon}A_{1,1}\cdot\mathbb{P}_{\sigma_j^*}(X_j = 1, S_i = 1), \\ K_0 &= \frac{g'(\epsilon)(e^\epsilon+1)^2}{2e^\epsilon}A_{1,1}\cdot\mathbb{P}_{\sigma_j^*}(X_j = 1, S_i = 0), \\ L_1 &= \frac{g'(\epsilon)(e^\epsilon+1)^2}{2e^\epsilon}A_{0,0}\cdot\mathbb{P}_{\sigma_j^*}(X_j = 0, S_i = 1), \\ L_0 &= \frac{g'(\epsilon)(e^\epsilon+1)^2}{2e^\epsilon}A_{0,0}\cdot\mathbb{P}_{\sigma_j^*}(X_j = 0, S_i = 0). \end{aligned}$$

Note that K_1, K_0, L_1 and L_0 are all positive and they do not depend on p_1, p_0, q_1 and q_0.

The privacy level of σ_i' can be written as

$$\begin{aligned} \zeta(\sigma_i') = \max\Bigg\{&\left|\ln\frac{p_1}{p_0}\right|, \left|\ln\frac{1-p_1}{1-p_0}\right|, \left|\ln\frac{q_1}{q_0}\right|, \left|\ln\frac{1-q_1}{1-q_0}\right|, \\ &\left|\ln\frac{1-p_1-q_1}{1-p_0-q_0}\right|, \left|\ln\frac{p_1+q_1}{p_0+q_0}\right|\Bigg\}. \end{aligned}$$

With a little abuse of notation, we consider $\zeta(\sigma_i')$ as a function $\zeta(p_1,p_0,q_1,q_0)$.

The expected utility of player i can thus be written as

$$\begin{aligned} &\mathbb{E}_{(\sigma_i',\sigma_{-i}^*)}[R_i(X) - g(\zeta(\sigma_i'))] \\ &= K_1p_1 + K_0p_0 + L_1q_1 + L_0q_0 - g(\zeta(p_1,p_0,q_1,q_0)). \end{aligned}$$

Let this utility define a function $U(p_1, p_0, q_1, q_0)$. Now we find the best response of player i, i.e., the (p_1, p_0, q_1, q_0) that maximizes $U(p_1, p_0, q_1, q_0)$. If player i does not participate, then $p_1 = p_0 = q_1 = q_0 = 0$ and $U(0, 0, 0, 0) = 0$. Otherwise, we find an optimal solution of the following optimization problem:

$$\begin{aligned} \max_{p_1,p_0,q_1,q_0} \quad & U(p_1, p_0, q_1, q_0) && \text{(P)} \\ \text{subject to} \quad & 0 \le p_1 \le 1, 0 \le q_1 \le 1, \\ & 0 \le p_1 + q_1 \le 1, \\ & 0 \le p_0 \le 1, 0 \le q_0 \le 1, \\ & 0 \le p_0 + q_0 \le 1, \\ & p_1 + q_1 + p_0 + q_0 > 0, && \text{(I.1)} \end{aligned}$$

by the following three steps.

Step 1. First, we prove that an optimal solution $(p_1^*, p_0^*, q_1^*, q_0^*)$ must satisfy that $p_1^* + q_1^* = p_0^* + q_0^*$. Suppose not. Without loss of generality we assume that $p_1^* + q_1^* < p_0^* + q_0^*$. We will find another solution $(p_1', p_0^*, q_1', q_0^*)$ that yields better utility, which contradicts the optimality of $(p_1^*, p_0^*, q_1^*, q_0^*)$.

Since we assume that $p_1^* + q_1^* < p_0^* + q_0^*$, then at least one of the following two inequality holds: $p_1^* < p_0^*$, $q_1^* < q_0^*$. Still without loss of generality we assume that $p_1^* < p_0^*$. Then if $q_1^* < q_0^*$, let $p_1' = p_0^*$ and $q_1' = q_0^*$. Since K_1 and L_1 are positive, $(p_1', p_0^*, q_1', q_0^*)$ yields higher payment. It is easy to verify that $\zeta(p_1', p_0^*, q_1', q_0^*) < \zeta(p_1^*, p_0^*, q_1^*, q_0^*)$. Thus $(p_1', p_0^*, q_1', q_0^*)$ yields better utility. For the other case that $q_1^* \ge q_0^*$, let $p_1' = p_0^* + q_0^* - q_1^*$ and $q_1' = q_1^*$. Then $p_1^* < p_1' \le p_0^*$. Since K_1 is positive, $(p_1', p_0^*, q_1', q_0^*)$ yields higher payment. To check the privacy cost, notice that

$$\begin{aligned} \zeta(p_1^*, p_0^*, q_1^*, q_0^*) = \max\bigg\{ & \ln\frac{p_0^*}{p_1^*}, \ln\frac{1-p_1^*}{1-p_0^*}, \ln\frac{q_1^*}{q_0^*}, \ln\frac{1-q_0^*}{1-q_1^*}, \\ & \ln\frac{1-p_1^*-q_1^*}{1-p_0^*-q_0^*}, \ln\frac{p_0^*+q_0^*}{p_1^*+q_1^*}\bigg\}, \end{aligned}$$

and

$$\zeta(p_1', p_0^*, q_1', q_0^*) = \max\left\{\ln\frac{p_0^*}{p_1'}, \ln\frac{1-p_1'}{1-p_0^*}, \ln\frac{q_1'}{q_0^*}, \ln\frac{1-q_0^*}{1-q_1'}\right\}.$$

Since $p_1' > p_1^*$ and $q_1' = q_1^*$, $\zeta(p_1', p_0^*, q_1', q_0^*) \le \zeta(p_1^*, p_0^*, q_1^*, q_0^*)$. Thus $(p_1', p_0^*, q_1', q_0^*)$ yields better utility. Therefore, by contradiction, we must have $p_1^* + q_1^* = p_0^* + q_0^*$.

Step 2. Next, we prove that an optimal solution $(p_1^*, p_0^*, q_1^*, q_0^*)$ must satisfy that $p_1^* + q_1^* = p_0^* + q_0^* = 1$. Still, suppose not. Then we will find another solution (p_1', p_0', q_1', q_0') that yields better utility.

Let

$$\begin{aligned} p_1' &= \frac{p_1^*}{p_1^* + q_1^*}, \quad & q_1' &= \frac{q_1^*}{p_1^* + q_1^*}, \\ p_0' &= \frac{p_0^*}{p_0^* + q_0^*}, \quad & q_0' &= \frac{q_0^*}{p_0^* + q_0^*}. \end{aligned}$$

By Step 1, $p_1^* + q_1^* = p_0^* + q_0^*$. By constraint (I.1), $p_1^* + q_1^* = p_0^* + q_0^* > 0$. Since we assume that $p_1^* + q_1^*$ and $p_0^* + q_0^*$ are not equal to 1, they must be less than 1. Since K_1, K_0, L_1 and L_0 are positive, (p_1', p_0', q_1', q_0') yields higher payment. It is easy to verify that $\zeta(p_1', p_0', q_1', q_0') \leq \zeta(p_1^*, p_0^*, q_1^*, q_0^*)$. Thus (p_1', p_0', q_1', q_0') yields better utility, which contradicts the optimality of $(p_1^*, p_0^*, q_1^*, q_0^*)$.

Step 3. By Step 1 and Step 2, the optimization problem (P) can be written as:

$$\max_{p_1, p_0 \in [0,1]} \bar{K}_1 p_1 + \bar{K}_0 p_0 - g(\zeta(p_1, p_0, 1-p_1, 1-p_0)) + \bar{K}, \tag{P1}$$

where

$$\begin{aligned}
\bar{K}_1 &= K_1 - L_1 = \frac{g'(\epsilon)(e^\epsilon + 1)^2}{2e^\epsilon}, \\
\bar{K}_0 &= K_0 - L_0 = -\frac{g'(\epsilon)(e^\epsilon + 1)^2}{2e^\epsilon}, \\
\bar{K} &= L_1 + L_0 \\
&= \frac{g'(\epsilon)(e^\epsilon + 1)^2}{2e^\epsilon} \frac{(e^\epsilon + 1)^2}{e^{2\epsilon} - 1} \frac{1}{D} \\
&\quad \cdot \left(\frac{e^\epsilon}{e^\epsilon + 1} P_1 + \frac{1}{e^\epsilon + 1} P_0 \right) \left(\frac{1}{e^\epsilon + 1} P_1 + \frac{e^\epsilon}{e^\epsilon + 1} P_0 \right).
\end{aligned}$$

The above calculation is done by noticing that

$$\begin{aligned}
\frac{e^{2\epsilon} - 1}{(e^\epsilon + 1)^2} D = &\ \mathbb{P}_{\sigma_j}(X_j = 1, S_i = 1) \mathbb{P}_{\sigma_j}(X_j = 0, S_i = 0) \\
&- \mathbb{P}_{\sigma_j}(X_j = 0, S_i = 1) \mathbb{P}_{\sigma_j}(X_j = 1, S_i = 0).
\end{aligned}$$

Solving (P1) is equivalent to solving the following optimization problem

$$\begin{aligned}
\max_{p_1, p_0, \xi} \quad & \bar{K}_1 p_1 + \bar{K}_0 p_0 - g(\xi) + \bar{K} \qquad \text{(P2)} \\
\text{subject to} \quad & \ln p_1 - \ln p_0 - \xi \leq 0 \\
& \ln p_1 - \ln p_0 + \xi \geq 0 \\
& \ln(1 - p_1) - \ln(1 - p_0) - \xi \leq 0 \\
& \ln(1 - p_1) - \ln(1 - p_0) + \xi \geq 0 \\
& p_1 \in [0, 1], p_0 \in [0, 1], \xi \in [0, +\infty].
\end{aligned}$$

The problem (P2) can be solved as follows: we first fix a $\xi \in [0, +\infty]$ and maximize the objective function with respect to p_1 and p_0; then we find an optimal ξ. For $\xi = 0$, the objective function always equals to $\bar{K}$ for feasible (p_1, p_0). For $\xi = +\infty$, the objective function always equal to $-\infty$. For any fixed $0 < \xi < +\infty$, the problem (P2) is a linear programming problem. The optimal solution is

$$(p_1^{(\xi)}, p_0^{(\xi)}) = \left(\frac{e^\xi}{e^\xi + 1}, \frac{1}{e^\xi + 1} \right),$$

and the optimal value is

$$-\frac{g'(\epsilon)(e^\epsilon+1)^2}{e^\epsilon}\frac{1}{e^\xi+1}-g(\xi)+\bar{K}_1+\bar{K}.$$

Let this optimal value defines a function f of ξ; i.e.,

$$f(\xi)=-\frac{g'(\epsilon)(e^\epsilon+1)^2}{e^\epsilon}\frac{1}{e^\xi+1}-g(\xi)+\bar{K}_1+\bar{K}. \tag{I.2}$$

To find the optimal ξ of $f(\xi)$, we calculate the derivatives of f as follows:

$$\begin{aligned} f'(\xi)&=\frac{g'(\epsilon)(e^\epsilon+1)^2}{e^\epsilon}\frac{e^\xi}{(e^\xi+1)^2}-g'(\xi),\\ f''(\xi)&=-\frac{g'(\epsilon)(e^\epsilon+1)^2}{e^\epsilon}\frac{e^\xi(e^\xi-1)}{(e^\xi+1)^3}-g''(\xi)\leq 0, \end{aligned}$$

where the second inequality is due to the convexity of the cost function g. Therefore, f is concave. Since $f'(\epsilon)=0$, the maximum value of f is achieved at ϵ. The optimal value is given by

$$\begin{aligned} f(\epsilon)&=-\frac{g'(\epsilon)(e^\epsilon+1)^2}{e^\epsilon}\frac{1}{e^\epsilon+1}-g(\epsilon)+\bar{K}_1+\bar{K}\\ &=g'(\epsilon)\frac{e^\epsilon-e^{-\epsilon}}{2}-g(\epsilon)+\bar{K}. \end{aligned}$$

By the convexity of g,

$$g(\epsilon)\leq g'(\epsilon)\epsilon\leq g'(\epsilon)\frac{e^\epsilon-e^{-\epsilon}}{2}.$$

Therefore, the optimal value satisfies that $f(\epsilon)\geq\bar{K}$, which is greater than 0, and the optimal solution of (P1) is given by

$$p_1^*=\frac{e^\epsilon}{e^\epsilon+1},\quad p_0^*=\frac{1}{e^\epsilon+1}.$$

According to the three steps above, the optimal solution of (P) is given by

$$\begin{aligned} p_1^*&=\frac{e^\epsilon}{e^\epsilon+1},\quad p_0^*=\frac{1}{e^\epsilon+1},\\ q_1^*&=\frac{1}{e^\epsilon+1},\quad q_0^*=\frac{e^\epsilon}{e^\epsilon+1}, \end{aligned}$$

and the optimal value is greater than 0. Therefore, the best response of player i is σ_i^*, which implies that σ^* is a Nash equilibrium of the mechanism $\boldsymbol{R}^{(N,\epsilon)}$. □

APPENDIX J

PROOF OF THEOREM 10

Proof. In the equilibrium σ^* of the mechanism $\boldsymbol{R}^{(N,\epsilon)}$, for each player i, given the private bit S_i, the reported data X_i is independent of S_{-i} and X_{-i}, and for any $s_i \in \{0,1\}$,

$$\begin{aligned}\mathbb{P}_{\sigma_i^*}(X_i = s_i \mid S_i = s_i) &= \frac{e^\epsilon}{e^\epsilon + 1},\\ \mathbb{P}_{\sigma_i^*}(X_i = 1 - s_i \mid S_i = s_i) &= \frac{1}{e^\epsilon + 1}.\end{aligned}$$

Therefore, given $S = s$ for any $s \in \{0,1\}^N$, $X_1, X_2, \ldots, X_N$ are independent random variables and each X_i has the distribution:

$$\begin{aligned}\mathbb{P}_{\sigma_i^*}(X_i = 1 \mid S = s) &= \frac{e^{s_i\epsilon}}{e^\epsilon + 1},\\ \mathbb{P}_{\sigma_i^*}(X_i = 0 \mid S = s) &= \frac{e^{(1-s_i)\epsilon}}{e^\epsilon + 1}.\end{aligned}$$

Recall that the principal is interested in estimating $\bar{S} = \frac{1}{N}\sum_{i=1}^N S_i$. The mechanism $\boldsymbol{R}^{(N,\epsilon)}$ estimates $\bar{S}$ by $\hat{\mu}$, which can be written as follows in the equilibrium σ^*:

$$\hat{\mu} = \frac{e^\epsilon + 1}{e^\epsilon - 1}\frac{1}{N}\sum_{i=1}^N X_i - \frac{1}{e^\epsilon - 1}.$$

We bound the probability for $|\bar{S} - \hat{\mu}| > \alpha$ in the equilibrium σ^*. First we write this probability as follows:

$$\begin{aligned}&\mathbb{P}_{\sigma^*}(|\bar{S} - \hat{\mu}| > \alpha)\\ =&\sum_{s\in\{0,1\}^N} \mathbb{P}_{\sigma^*}(|\bar{S} - \hat{\mu}| > \alpha \mid S = s)\mathbb{P}(S = s)\end{aligned}$$

Given any $s \in \{0,1\}^N$, notice that

$$\bar{S} - \hat{\mu} = \frac{1}{N}\sum_{i=1}^N \left(\frac{e^\epsilon + 1}{e^\epsilon - 1}X_i - S_i - \frac{1}{e^\epsilon - 1}\right)$$

is the average of N independent random variables. The expectation and variance of $\bar{S} - \hat{\mu}$ can be calculated as

$$\begin{aligned}\mathbb{E}_{\sigma^*}[\bar{S} - \hat{\mu}] &= 0,\\ \mathbb{V}_{\sigma^*}(\bar{S} - \hat{\mu}) &= \frac{1}{N}\frac{e^\epsilon}{(e^\epsilon - 1)^2}.\end{aligned}$$

Then by Chebyshev's inequality,

$$\mathbb{P}_{\sigma^*}(|\bar{S} - \hat{\mu}| > \alpha \mid S = s) \le \frac{1}{\alpha^2 N}\frac{e^\epsilon}{(e^\epsilon - 1)^2}.$$

Therefore,

$$\mathbb{P}_{\sigma^*}(|\bar{S} - \hat{\mu}| > \alpha) \leq \frac{1}{\alpha^2 N} \frac{e^\epsilon}{(e^\epsilon - 1)^2}.$$

Since we choose

$$\epsilon \geq \ln\left(2 + \frac{1}{N\alpha^2\delta}\right),$$

we have

$$\begin{aligned} \frac{1}{\alpha^2 N} \frac{e^\epsilon}{(e^\epsilon - 1)^2} &= \frac{1}{\alpha^2 N} \frac{1}{e^\epsilon + e^{-\epsilon} - 2} \\ &\leq \frac{1}{\alpha^2 N} \frac{1}{e^\epsilon - 2} \\ &\leq \delta. \end{aligned}$$

Therefore, $\mathbb{P}_{\sigma^*}(|\bar{S} - \hat{\mu}| > \alpha) \leq \delta$ and thus $\mathbb{P}_{\sigma^*}(|\bar{S} - \hat{\mu}| \leq \alpha) \geq 1 - \delta$, which indicates that the estimate $\hat{\mu}$ is (α, δ)-accurate in the equilibrium σ^*. □

APPENDIX K

PROOF OF PROPOSITION 1

Proof. Consider any nonnegative payment mechanism $\boldsymbol{R}$. For any player i and any strategy σ_i', let

$$\begin{aligned} p_1 &= \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 1), \quad p_0 = \mathbb{P}_{\sigma_i'}(X_i = 1 \mid S_i = 0), \\ q_1 &= \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 1), \quad q_0 = \mathbb{P}_{\sigma_i'}(X_i = 0 \mid S_i = 0). \end{aligned}$$

Consider the strategy profile σ^*. Similar to the proof of Theorem 9, we write the expected payment of player i as

$$\begin{aligned} &\mathbb{E}_{(\sigma_i', \sigma_{-i}^*)}[R_i(X)] \\ &= \sum_{x \in \{0,1\}^N} R_i(x_i, x_{-i}) \mathbb{P}_{(\sigma_i', \sigma_{-i}^*)}(X_i, X_{-i}) \\ &= \sum_{x_i, s_i \in \{0,1\}} \Bigg(\mathbb{P}_{\sigma_i'}(X_i = x_i \mid S_i = s_i) \\ &\quad \cdot \sum_{x_{-i} \in \{0,1\}^{N-1}} R_i(x_i, x_{-i}) \mathbb{P}_{\sigma_{-i}^*}(X_{-i} = x_{-i}, S_i = s_i) \Bigg) \\ &= K_1 p_1 + K_0 p_0 + L_1 q_1 + L_0 q_0, \end{aligned}$$

where

$$\begin{aligned} K_1 &= \sum_{x_{-i} \in \{0,1\}^{N-1}} R_i(1, x_{-i}) \mathbb{P}_{\sigma_{-i}}(X_{-i} = x_{-i}, S_i = 1), \\ K_0 &= \sum_{x_{-i} \in \{0,1\}^{N-1}} R_i(1, x_{-i}) \mathbb{P}_{\sigma_{-i}}(X_{-i} = x_{-i}, S_i = 0), \\ L_1 &= \sum_{x_{-i} \in \{0,1\}^{N-1}} R_i(0, x_{-i}) \mathbb{P}_{\sigma_{-i}}(X_{-i} = x_{-i}, S_i = 1), \\ L_0 &= \sum_{x_{-i} \in \{0,1\}^{N-1}} R_i(0, x_{-i}) \mathbb{P}_{\sigma_{-i}}(X_{-i} = x_{-i}, S_i = 0). \end{aligned}$$

Note that K_1, K_0, L_1 and L_0 are all nonnegative and they do not depend on σ_i'. Then the expected utility of player i can be written as

$$\begin{aligned} &\mathbb{E}_{(\sigma_i', \sigma_{-i}^*)}[R_i(X) - g(\zeta(\sigma_i'))] \\ &= K_1 p_1 + K_0 p_0 + L_1 q_1 + L_0 q_0 - g(\zeta(p_1, p_0, q_1, q_0)). \end{aligned}$$

Consider the strategy $\sigma_i^{(\xi)}$ of player i defined as follows

$$\begin{aligned} \mathbb{P}_{\sigma_i^{(\xi)}}(X_i = 1 \mid S_i = 1) = \mathbb{P}_{\sigma_i^{(\xi)}}(X_i = 0 \mid S_i = 0) = \frac{e^\xi}{e^\xi + 1}, \\ \mathbb{P}_{\sigma_i^{(\xi)}}(X_i = 0 \mid S_i = 1) = \mathbb{P}_{\sigma_i^{(\xi)}}(X_i = 1 \mid S_i = 0) = \frac{1}{e^\xi + 1}. \end{aligned}$$

Then the expected utility of player i can be further written as

$$\begin{aligned}
&\mathbb{E}_{(\sigma_i^{(\xi)},\sigma_{-i}^*)}[R_i(X) - g(\zeta(\sigma_i^{(\xi)}))] \\
&= (K_1 - L_1)\frac{e^\xi}{e^\xi + 1} + (K_0 - L_0)\frac{1}{e^\xi + 1} + L_1 + L_0 - g(\xi) \\
&= \bar{K}_1\frac{e^\xi}{e^\xi + 1} + \bar{K}_0\frac{1}{e^\xi + 1} + \bar{K} - g(\xi) \\
&= -(\bar{K}_1 - \bar{K}_0)\frac{1}{e^\xi + 1} - g(\xi) + \bar{K}_1 + \bar{K},
\end{aligned}$$

where

$$\bar{K}_1 = K_1 - L_1, \ \ \bar{K}_0 = K_0 - L_0, \ \ \bar{K} = L_1 + L_0.$$

Let this expected utility define a function h of ξ; i.e.,

$$h(\xi) = -(\bar{K}_1 - \bar{K}_0)\frac{1}{e^\xi + 1} - g(\xi) + \bar{K}_1 + \bar{K}.$$

Then a necessary condition for σ^* to be a Nash equilibrium is that the level ϵ in σ_i^* maximizes $h(\xi)$. Since

$$h'(\xi) = (\bar{K}_1 - \bar{K}_0)\frac{e^\xi}{(e^\xi + 1)^2} - g'(\xi),$$

we must have

$$\bar{K}_1 - \bar{K}_0 = \frac{g'(\epsilon)(e^\epsilon + 1)^2}{e^\epsilon}.$$

Next let us bound $\bar{K}_1 + \bar{K}$. By definitions,

$$\begin{aligned}
\bar{K}_1 - \bar{K}_0 = \sum_{x_{-i}} &\Big(R_i(1, x_{-i}) - R_i(0, x_{-i})\Big) \\
&\cdot \Big(\mathbb{P}_{\sigma_{-i}^*}(X_{-i} = x_{-i}, S_i = 1) \\
&\quad - \mathbb{P}_{\sigma_{-i}^*}(X_{-i} = x_{-i}, S_i = 0)\Big).
\end{aligned}$$

Let $\mathcal{A} = \{x_{-i} \in \{0,1\}^{N-1} \colon R_i(1, x_{-i}) \geq R_i(0, x_{-i})\}$. Then

$$\begin{aligned}
\bar{K}_1 + \bar{K} &= \sum_{x_{-i}}\Big(R_i(1, x_{-i})\mathbb{P}_{\sigma_{-i}^*}(X_{-i} = x_{-i}, S_i = 1) \\
&\qquad + R_i(0, x_{-i})\mathbb{P}_{\sigma_{-i}^*}(X_{-i} = x_{-i}, S_i = 0)\Big) \\
&\geq \sum_{x_{-i} \in \mathcal{A}}\Big(R_i(1, x_{-i}) - R_i(0, x_{-i})\Big) \\
&\qquad \cdot \mathbb{P}_{\sigma_{-i}^*}(X_{-i} = x_{-i}, S_i = 1)
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{x_{-i} \in \mathcal{A}^c} \Big(R_i(0, x_{-i}) - R_i(1, x_{-i}) \Big) \\
&\qquad\qquad \cdot \mathbb{P}_{\sigma^*_{-i}}(X_{-i} = x_{-i}, S_i = 0) \\
&\geq \bar{K}_1 - \bar{K}_0 \\
&= \frac{g'(\epsilon)(e^\epsilon + 1)^2}{e^\epsilon}.
\end{aligned}
$$

Therefore, the expected payment to player i at σ^* is lower bounded as

$$
\begin{aligned}
\mathbb{E}_{\sigma^*}[R_i(X)] &= -(\bar{K}_1 - \bar{K}_0) \frac{1}{e^\xi + 1} + \bar{K}_1 + \bar{K} \\
&\geq g'(\epsilon)(e^\epsilon + 1),
\end{aligned}
$$

and thus the total expected payment at σ^* is lower bounded as

$$
\mathbb{E}_{\sigma^*}\left[\sum_{i=1}^N R_i(X) \right] \geq N g'(\epsilon)(e^\epsilon + 1).
$$

□