Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

MATHEMATICAL FOUNDATIONS OF DIFFERENTIAL PRIVACY

NAOISE HOLOHAN

A THESIS SUBMITTED FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SCHOOL OF COMPUTER SCIENCE AND STATISTICS
TRINITY COLLEGE, DUBLIN

SUPERVISOR: PROF. DOUGLAS J. LEITH
CO-SUPERVISOR: DR. OLIVER MASON (MAYNOOTH UNIVERSITY)

SEPTEMBER 2016 – VERSION 1.1

VERSION HISTORY:

Version 0.1 (25th May 2016): First draft of Chapter 2

Version 0.2 (30th May 2016): Second draft of Chapter 2

Version 0.3 (28th July 2016): Third draft of Chapter 2 and first draft of Chapter 3

Version 0.4 (11th August 2016): First draft of Chapter 4

Version 0.5 (17th August 2016): Minor changes to Chapter 4, and first draft of Chapter 5 and Nomenclature

Version 0.6 (24th August 2016): First draft of Chapter 6 and minor changes elsewhere

Version 0.7 (31st August 2016): First draft of the Abstract and Chapter 1, minor changes to Chapter 6 and updated introductions and conclusions to each other chapter.

Version 0.8 (2nd September 2016): First draft of Chapter 7.

Version 1.0 (7th September 2016): Final changes for submission.

Version 1.1 (13th February 2017): Typos and other edits after viva.

DECLARATION

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Baile Átha Cliath, Meán Fómhair 2016

Naoise Holohan

ABSTRACT

Sensitive personal data collected about us has the potential to give great insight into human behaviour, but it remains a challenge to publish such information while guaranteeing (i) individuals' privacy and (ii) data utility. This problem has received much attention in the computer science research community in recent decades. Since 2006, differential privacy has emerged as a popular paradigm to measure the privacy of sensitive data publications. In this thesis we explore the mathematical foundations of differential privacy, with a particular focus on the tradeoff of privacy and utility.

After reviewing the history of the right to privacy and the literature on differential privacy, we present the cornerstone of our research: a single, unifying rigorous mathematical framework for differential privacy. The abstract approach taken allows for a wide variety of data types, query types and sanitisation methods to be handled in the one framework. This also allows us to prove results in generality. We take a special focus on 1-dimensional mechanisms, and their use in creating more general n-dimensional mechanisms.

The versatility of this unifying framework is demonstrated when we apply it to the special case of categorical data. This also leads us to examine the polytope of differentially private mechanisms, and later to applying the differential privacy protocol to the randomised response technique, a privacy technique first proposed in 1965. We also examine the randomised response technique beyond differential privacy, and how efficiencies can be gained over previous methods.

Throughout this thesis, a special focus is maintained on the privacy/utility tradeoff, and particularly on optimal mechanisms. We present error bounds on differentially private mechanisms, and derive the optimal mechanism for the max-mean error function. By examining the extreme points of the differential privacy polytope, we allow for the determination of the optimal mechanism for more general linear error functions. An optimal randomised response mechanism which satisfies differential privacy is also presented.

CONTENTS

LIST OF FIGURES

ACRONYMS

DPV	Degree of Privacy Violation 124
FOIL	Freedom of Information Law 14
GIC	Group Insurance Commission 12
IMDB	Internet Movie Database 14
MLE	Maximum Likelihood Estimator 107
OECD	Organisation for Economic Co-operation and Development 9
PPDM	Privacy-Preserving Data Mining 11
PPDP	Privacy-Preserving Data Publishing 11
RR	Randomised Response 3
SDC	Statistical Disclosure Control 7
TLC	New York City Taxi and Limousine Commission 14

1 INTRODUCTION

We begin by giving a brief overview of the motivation for the thesis and an explanation of the goals of our research. We also outline the structure of the thesis, give an introduction to the notation used, and give a list of publications that have resulted from the work.

OVERVIEW

1.1 MOTIVATION

The personal data being collected about us by governments and corporations potentially holds huge value for scientific research and commercial purposes. Privacy laws ensure that this data, some of which could be considered sensitive to individuals (e. g. medical records), enjoys considerable protection. From a research perspective, publishing data truthfully, or even answering statistical queries on such data truthfully, would be of greatest benefit, but risks violating privacy. Conversely, people's privacy can easily be protected by not releasing the data, but the value of the data to research would then never be realised.

The problem of protecting user privacy while simultaneously preserving data utility has been attracting the attention of researchers in data publishing for decades. Many models have been proposed to optimise utility while guaranteeing a prescribed level of privacy. Since 2006, differential privacy

has emerged as a popular privacy paradigm, and it is differential privacy with which this thesis is primarily concerned.

Differential privacy gives a quantitative mathematical definition to measure the level of privacy achieved in a given data release. This definition then determines the amount of noise or perturbation that needs to be applied to achieve the desired level of privacy. Under differential privacy, privacy is quantified by how statistically indistinguishable the privacy-preserved outputs from two similar datasets are.

The benefits of differential privacy include its well-defined nature, which allows for mathematical investigation. Additionally, differential privacy has been shown to be resistant to auxiliary/background information attacks, where third-party information is used to compromise the privacy of the data. This has given it an advantage over other privacy-preserving protocols that have been shown to be vulnerable to such attacks.

1.2 CONTRIBUTIONS

We begin by formulating a single, unifying mathematical framework for differential privacy. By taking an abstract approach through probability theory, measure theory and metric spaces, we construct mechanisms for data publication that can be applied to a variety of queries, data types and perturbation techniques. This allows us to fully realise the generality of differential privacy as originally conceived by its creators.

The framework allows us to construct the most general of results. Whatever type of data we're dealing with, whichever queries we seek to ask, or however we wish to perturb the data, these results will continue to hold. The versatility of the framework is demonstrated by applying it to the special case of categorical data. Worked examples are presented for other data types also.

We also highlight the variety of perturbation techniques that can be handled by this framework. Data perturbation and query output perturbation can both easily be accommodated. We also present an implementation of local privacy. Under the local privacy setting, each user perturbs their own

data before supplying it to a central dataset, meaning even the dataset's curator doesn't see the original, truthful version.

As a cornerstone in the study of privacy-preserving data publishing, we give special focus to the privacy/utility tradeoff in this thesis. By calculating and studying the error of a mechanism, we equip data publishers with the tools to offset utility against user privacy, and vice versa.

We examine the utility of response mechanisms with respect to certain error functions, and also present results that can be applied to the complete class of linear error functions for particular mechanisms. Optimal mechanisms are given special attention: given a desired level of privacy, is there a single, best mechanism that will optimise utility? Optimal mechanisms are presented in a number of scenarios in this thesis.

Despite its popularity in the research community, the uptake of differential privacy in practice is proving slow. Evidence of its use in industry is limited at best. Some works have criticised differential privacy for producing high error and low utility. In this respect, we extend our mathematical analysis of differential privacy to Randomised Response (RR).

RR is a widely-used technique in confidential surveying in which data is modified, subject to a given probability distribution, to preserve user privacy. We give a practical implementation to differential privacy by applying it to RR, where data perturbation is an established and accepted practice. We also present optimal RR mechanisms for differential privacy.

Additionally, we extend our analysis of RR beyond differential privacy, with a focus on improved efficiency. We examine the privacy protection of two particular RR models to determine their relative efficiency. We also extend RR to the non-binary case, where the question being asked has more than two (yes/no) answers.

1.3 STRUCTURE

The structure of this thesis is as follows:

- In Chapter 2 we give an outline of the background to data privacy, along with a review of the history and recent innovations in differential privacy and RR.

- In Chapter 3 we formally introduce differential privacy and formulate a framework in the abstract setting of probability on metric spaces. We establish a number of fundamental results, including ones relating to the identity query and local privacy, and introduce the primary types of response mechanisms. We also examine the privacy/utility trade-off and establish general error bounds that apply to all differentially private mechanisms.

- In Chapter 4 we specialise our differential privacy framework to categorical (non-numeric) data. We establish criteria for checking a mechanism for differential privacy in varying degrees of generality and again present results on error. In the context of local privacy, we also present the optimal mechanism for categorical data with respect to a given error function.

- In Chapter 5 we study the polytope of local differential privacy mechanisms and present criteria for identifying and creating extreme points of the polytope. We also fully characterise a number of classes of these extreme points. This gives further insight into utility, given that optimal mechanisms with respect to linear error functions are guaranteed to occur at extreme points of the polytope.

- In Chapter 6 we apply differential privacy to the RR technique, and present criteria to determine the optimal mechanism. We also extend our analysis of RR beyond differential privacy, and quantify efficiency improvements via privacy-protection metrics. Finally, we present a model to extend RR beyond binary responses.

- Concluding remarks and a discussion of possible future work are given in Chapter 7.

1.4 NOTATION

We maintain standard mathematical notation throughout this thesis as much as possible. We denote scalars and vectors by lowercase letters (e. g. a, b, y, z), where the distinction between scalar and vector will typically be clear from context. In some cases where the distinction is not clear, vectors are denoted

by boldface lowercase letters (e. g. $\boldsymbol{d}$). Random variables, matrices and sets are typically denoted by uppercase letters (e. g. A, B, S). Entries of a matrix or vector are denoted by lowercase, indexed letters (e. g. $(A)_{ij} = a_{ij}$). Orders on vectors and matrices are assumed to be entry-wise (e. g. $A \geq 0$ implies $a_{ij} \geq 0$ for each i, j).

A number of standard identities are used throughout the thesis. Examples include the sets of integers ($\mathbb{Z}$) and real numbers ($\mathbb{R}$), the standard basis vectors (e_i), the all-ones vector ($\mathbf{1}$) and the binomial coefficient ($\binom{n}{k} = \frac{n!}{k!(n-k)!}$). Standard functions that are used include the Kronecker delta function ($\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ otherwise), the sign function ($\operatorname{sgn}(x) = 0$ when $x = 0$, $\operatorname{sgn}(x) = \frac{x}{|x|}$ otherwise), the projection function ($\pi_i(v) = v_i$, the ith component of the vector v), the convex hull ($\operatorname{conv}(v_1, \ldots, v_k)$, the set of all convex combinations of the points $v_1, \ldots, v_k$) and the power set ($\mathcal{P}(A)$, the set of all subsets of A). We denote by $\sigma(S)$ the σ-algebra generated by the collection of subsets S.

Given a binary relation R, we write $a \, R \, b \in A$ as shorthand for $a, b \in A$ and $a \, R \, b$ (e. g. $a \leq b \in \mathbb{R}$ implies $a, b \in \mathbb{R}$ and $a \leq b$). We let $[n]$ denote the set of all positive integers less than or equal to n ($[n] = [1, n] \cap \mathbb{Z}$), and $[n]_0$ to denote $[n] \cup \{0\}$.

These and other notational conventions will also typically be introduced on a chapter-by-chapter basis.

1.5 PUBLICATIONS

The following journal papers have been published/submitted reporting the work in this thesis.

1. Holohan, N., Leith, D. J., and Mason, O. Differential privacy in metric spaces: Numerical, categorical and functional data under the one roof. *Information Sciences 305* (2015), 256–268.

2. Holohan, N., Leith, D. J., and Mason, O. Differentially private response mechanisms on categorical data. *Discrete Applied Mathematics 211* (2016), 86–98.

3. HOLOHAN, N., LEITH, D. J., AND MASON, O. Extreme points of the local differential privacy polytope. *arXiv preprint arXiv:1605.05510 [math.CO]* (2016).

2

BACKGROUND

We now examine the field of Data Privacy and Statistical Disclosure Control (SDC). We begin by looking at the emergence of privacy as a human right, and the evolution of data protection in more recent decades. A number of examples are given of significant SDC failures of the recent past. We then give an introduction to SDC, and explain the primary design considerations in constructing a mechanism. We also give an overview of a number of significant SDC models in the context of this thesis.

OVERVIEW

2.1 MOTIVATING DATA PRIVACY

Human beings have been collecting data in various forms for millennia. The recording of numerical information dates from at least 30,000 years ago, an era commonly known as the Late Stone Age [Rud07], while even census-taking has a long history, dating from c. 3800 BC [Aus06].

Nowadays, fast computers and cheap storage have precipitated the creation of a vast mountain of data [MSC13, Pg. 5]. This large bank of digital information has become known as *big data*, and, primarily characterised by its rich detail and large volume [WB13], has been compared to the invention of the microscope in providing new insights to human behaviour [Hig11].

Companies are tapping into big data in the pursuit of greater profits [PF13], while healthcare, fraud detection and crime-fighting can all benefit from big data analysis in the public sector [MCB+11].

Curators of such datasets face many challenges. In particular, the sensitive nature of much of the data being collected means that safeguarding the privacy of the individuals whom the data concerns has become the subject of increasing attention.

2.1.1 *The Right to Privacy*

Privacy is not just an ethical consideration for data curators, but a legal requirement too. The concept of a right to privacy was famously advocated by Warren and Brandeis in their *Harvard Law Review* article of 1890, entitled "The Right to Privacy" [WB90]. Fewer than 60 years later, that same right to privacy was enshrined in the Universal Declaration of Human Rights in 1948.

Warren and Brandeis were reacting to the advances in technology of the time – photography and newspaper publishing – in advocating a right to privacy, and drew inspiration from Judge Thomas Cooley's work a decade earlier [Coo79]. As Warren and Brendeis wrote:

> That the individual shall have full protection in person and in property is a principle as old as the common law; but it has been found necessary from time to time to define anew the exact nature and extent of such protection.
>
> Recent inventions and business methods call attention to the next step which must be taken for the protection of the person, and for securing to the individual what Judge Cooley calls the right "to be let alone". Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life; and numerous mechanical devices threaten to make good the prediction that "what is whispered in the closet shall be proclaimed from the house-tops."

Later, following the Second World War, the Universal Declaration of Human rights was drafted "as a common standard of achievements for all peo-

ples and all nations" [UN 48]. Article 12 of the Declaration protects our right to privacy:

> No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation. Everyone has the right to the protection of the law against such interference or attacks.

Privacy is similarly protected in the International Covenant on Civil and Political Rights [UN 66, Art. 17] and the European Convention on Human Rights [Cou50, Art. 8]. Almost all countries around the world recognise the right to privacy in their constitutions [BD99], however, this is not universal. In Ireland, for example, the constitution makes no direct reference to privacy [Ire37]. In 2006, a working group on privacy noted that

> [t]here is no express constitutional provision relating to privacy per se. Article 40.3 of the Constitution, however, posits a guarantee by the State to defend and vindicate the personal rights of the citizen . . . [Wor06, Par. 2.34]

Various court judgements in the last 50 years have clarified those rights. In 1973 the Supreme Court recognised a right to marital privacy [McG74], and in 1984, acknowledged a general right to privacy [Nor84]. Then, in 1998, the same court ruled in a case that: "There is no doubt but that the Plaintiffs/Appellants enjoy a constitutional right to privacy." [Hau98]

2.1.2 *Data Protection*

While Warren and Brandeis originally advocated the right to privacy in response to the newly-intrusive developments of photography and gossip columns of newspapers, the same right can be logically extended to cover data collected about us.

In 1980, the Organisation for Economic Co-operation and Development (OECD) published a guideline for data protection in an effort to "harmonise national privacy legislation" [OEC80]. It was a pivotal time when many countries were introducing privacy protection legislation. The guideline was up-

dated in 2013 [OEC13], and maintained the original basic principles of data protection:

1. *Collection Limitation* – personal data collection should be limited, lawful and done with the knowledge/consent of the subject;
2. *Quality* – data collected should be relevant to its use and should be accurate, complete and up-to-date;
3. *Purpose Specific* – the purpose for the data should be stated at collection and only used for that purpose;
4. *Use Limited* – data should not be disclosed without the consent of the user;
5. *Security Safeguards* – personal data should be secured against risk of loss, unauthorised access or destruction; and against use, modification or disclosure of data;
6. *Openness* – curators should be open about the existence and purpose of personal data;
7. *Individual Participation* – subjects should be allowed access their data and have any errors relating to them erased/updated;
8. *Accountability* – data curators should be accountable for upholding the above principles.

Data protection legislation is in place across much of the world [Ban14] and the OECD's guidelines have often formed the basis for that legislation [Shioo]. The European Union adopted the guidelines in full in its Data Protection Directive [Cou95] and in the General Data Protection Regulation which is due to come into force across the EU on 25 May 2018 [Eur16].

As the above guidelines illustrate, data protection covers many aspects of the data collection, storage and analysis process. Principles are established for: how it is collected; how it can be used; how it must be stored; and how it can be accessed, and by whom.

In the next section, we draw a distinction between the broad principles of data protection, and protection of one's privacy in the publication of that data.

2.1.3 *Statistical Disclosure Control*

This thesis is concerned with the protection of an individuals' privacy in the publication of data relating to him/her. This is distinct from the broader definition of data protection, as it is only concerned with the controlled dissemination of data from a trusted source to an untrusted third party.

This interpretation of data privacy fits with that of Statistical Disclosure Control (SDC), a term first coined by Tore Dalenius in 1977 [Dal77]. Two primary implementations of SDC exist, whereby statistics on the data are published, known as Privacy-Preserving Data Mining (PPDM), or where the data itself is published, known as Privacy-Preserving Data Publishing (PPDP) [FWCY10].

Data privacy in this context does not relate to the security of the data. Data breaches resulting from hacks, leaks or stolen data are not considered data privacy failures, but rather failures of security. For example, the Ashley Madison data breach in 2015 was not as a result of data privacy failure, but because of lax security in protecting the data [MD15].

For the purpose of SDC, we consider three primary actors:

1. *Data Subject* – an entity described in the data whose privacy we are seeking to protect;
2. *Data Curator* – the trusted collector of the private information;
3. *Data Recipient* – a third-party whose trust cannot be guaranteed, known as an *adversary*, *attacker* or *intruder* in other circumstances.

There are examples in legislation where a distinction is drawn between someone in control of the data (data controller) and someone who has the data for processing or analysing (data processor). We choose not make such a distinction, instead focusing on a framework involving the three actors mentioned above.

As will be detailed in the coming sections, disclosure control can be applied at various points in the life-cycle of the data (Section 2.2), achieved by a variety of methods (Section 2.4), and subject to a variety of models that have been proposed (Section 2.5). Data privacy is complicated further by the huge variety in types of data in existence (Section 2.3).

2.1.4 *Privacy Failures in Data Releases*

As explained previously, unauthorised access to data through leaks, hacks or theft does not fall under the remit of SDC. While less common than data security breaches, there is, however, no dearth of examples of SDC failures in recent times.

We now review a number of common examples regularly cited in the SDC literature, as well as two more recent examples. These will give a flavour of the problem SDC is seeking to address.

2.1.4.1 *Massachusetts Health Records (1990s)*

In the US state of Massachusetts, the Group Insurance Commission (GIC) purchases health insurance for state employees. The GIC is known to have collected patient-specific data on 135,000 employees and their families, and in the mid 1990s decided to release some of this data to researchers [Swe02, Swe05].

Then governor of Massachusetts William Weld claimed the privacy of individuals would be protected because explicit identifiers in the dataset, such as name and address, had been removed [Gre07]. However, a then MIT graduate student, Latanya Sweeney, soon proved those claims to be false. She purchased the voter registration list for Cambridge, Massachusetts for 20 dollars and linked it to the GIC dataset using date of birth, gender and ZIP code. She was able to isolate governor Weld's health records, which she then had delivered to his office [Gre07].

"According to the Cambridge Voter list, six people had his particular birth date, only three of them were men, and, he was the only one in his 5-digit ZIP code," Sweeney told the Pennsylvania House Select Committee on Information Security [Swe05].

A later study revised this to 63 per cent of the population [Golo6]

She also found that those same three attributes were enough to uniquely identify 87 per cent of the US population [Swe00], based on experiments on 1990 US Census data.

2.1.4.2 *AOL Search Logs (2006)*

In 2006, AOL Research publicly released a database of internet search queries covering 650,000 users over a three-month timeframe, amounting to 20 million individual queries. The head of AOL Research at the time, Abdur Chowdhury, said the data had been released to "embrac[e] the vision of an open research community" [Ohm10].

There is some disagreement as to the precise size of the released dataset [HL13]

As with GIC, AOL Research attempted to protect users' privacy by removing identifiers from the dataset. Usernames and IP addresses were replaced by unique identifiers, so searches by the same person could still be linked together but the person's identity kept secret.

Journalists at the *New York Times*, however, were able to identify user No. 4417749 as Thelma Arnold, a sixty-two-year-old widow from Lilburn, Georgia [BZ06]. The hundreds of searches conducted by Arnold in those three months, including searches about Lilburn and people with her surname, were unique enough to identify her.

It was subsequently reported that AOL's chief technology officer resigned and a further two employees were dismissed as a result of the privacy breach [Zel06, Kar06].

2.1.4.3 *Netflix Prize (2006)*

Less than two months after AOL's privacy failings hit the press, DVD rental firm Netflix launched the *Netflix Prize*. Netflix were looking to improve the accuracy of their recommendation system and offered a prize of one million dollars to the first team to achieve an improvement of 10 per cent on Netflix's own system.

Netflix began online streaming of content in 2007 [And07]

To aid researchers, Netflix released a training dataset comprising 100 million ratings from 480,000 randomly-chosen customers on 18,000 movies in a seven year timeframe. Each rating included the rating itself (from 1 to 5 stars), the date of the rating and the movie [Net09].

"To protect customer privacy, all personal information identifying individual customers has been removed and all customer ids have been replaced by randomly-assigned ids," the Netflix Prize rulebook read. Ratings from a single user could still be linked, but the identity of the user would be unknown.

This dataset was not immune to re-identification either. Just two weeks after the dataset was released by Netflix, researchers claimed that a customer's record could be identified using "only a little bit" of information [NS06]. To put their theory into action, the researchers turned to the Internet Movie Database (IMDB). From a random sample of around 50 IMDB users, Narayanan and Shmatikov claimed to have successfully re-identified two of them in the Netflix Prize dataset [NS08].

Although the Netflix Prize proved a success in finding a better recommendation system (the prize was won in 2009), plans for a second contest were cancelled in 2010 because of privacy concerns [Loh09, Loh10].

2.1.4.4 *Facebook Ads (2010)*

More recently, in 2010, a case study by Stanford computer scientist Aleksandra Korolova found that Facebook's advertising system allowed ads to be tailored in such detail as to be targetted at a single person [Kor10].

When selecting which ads to show a user, Facebook uses private and "Friends Only" information. For example, a user is required to provide a date of birth when signing up to Facebook, but need not display their age on their profile.

Korolova demonstrated that this shortcoming could allow an advertiser to infer an individual's private information, such as age or sexual orientation. Facebook modified their systems just one week after being notified of the vulnerability, but Korolova says the fixes are still insufficient to fully protect a user's privacy [Kor10].

2.1.4.5 *New York City Taxi Trips (2014)*

The Freedom of Information Law (FOIL) in New York state allows members of the public to access records of governmental agencies. In March 2014, after seeing a graph tweeted by the New York City Taxi and Limousine Commission (TLC), blogger Chris Whong used FOIL to obtain detailed information for all NYC taxi trips taken in 2013 [Who14b]. The data contained information on fares (total fare, surcharge, toll, and tip) and the trip itself (time, date, distance travelled, and GPS coordinates of pick-up and drop-off) in two tables, linkable by a hashed medallion and hack license identifier. Fares

could be linked to their corresponding trip, but the identity of the taxi was unknown, as was the passenger.

Three months after receiving the data, another contributor noticed that the medallion and hack license identifiers had been created using the `md5` hash function. Once he had calculated the `md5` hash for all possible hack license and medallion numbers (22 million in total), the entire 50GB dataset was re-identified [Pan14]. This re-identified dataset was then used to find how much celebrities tip for taxi journeys, by using photographs to link them to their taxi at the specific date, time and location [Toc14, Tro14].

When the TLC made the same information available for 2014, the medallion and hack license numbers had been removed entirely, making a similar attack much more difficult to implement [Who14a].

2.2 LOCATING THE PRIVACY MECHANISM(S)

Data can have a long, convoluted life-cycle, as it makes its way from the data subject through to the data curator where it can be cleaned, analysed and processed, eventually making its way to the data recipient. From a data privacy perspective however, the points at which privacy-preserving actions can take place are simplified to just two.

As explained in Section 2.1.3, we make no distinction between the collector, controller, owner, processor, etc. of the data. We are only concerned with the act of altering the data for privacy's sake, and not which entity is responsible/liable for it. We therefore use the catch-all term *data curator* to refer to the above.

The two points we consider are the point of collection, and post-collection. The data life-cycle contains other points at which privacy-preserving actions can be implemented, such as authentication privacy at the point of communication with the data curator [DFWBJ15]. These are beyond the scope of SDC and this thesis.

We are not limited in the number of privacy mechanisms we can apply, however. Multiple privacy mechanisms can be added at collection and post-collection, depending on the need.

2.2.1 *Collection*

Also known as *local privacy* or *input perturbation*, applying a privacy mechanism at the collection point of the data means even the data curator may not see the original (truthful) input. The privacy is *local* as the privacy method is implemented locally by the data subject, before it ever reaches the data curator.

This approach is especially useful when the data curator is not trusted by the subjects, since the control for implementing a privacy strategy lies with the data subject.

2.2.2 *Post-Collection*

In the post-collection scenario, data curators apply their privacy-preserving strategies to the data they have collected. This is done before or as the data is delivered to the data recipient. The (trusted) curator can therefore still perform analysis of the data it has collected, although this data will only be truthful if no privacy mechanism is in place at the point of collection.

Most of the examples detailed in Section 2.1.4 implemented a privacy method applied at the post-collection stage.

2.3 TYPES AND STRUCTURE OF DATA

Data, and the format in which it is collected, can be messy and difficult to decipher. It's estimated that as much as 95 per cent of data is unstructured [Meh11, MSC13, Pg. 47], meaning a considerable amount of processing and cleaning must be done before it can be analysed.

SDC deals almost exclusively with structured data, but even here there are difficulties. The many types of data that are available to researchers present challenges when deriving new privacy protocols.

In many instances, structured data is assumed to be constructed of *data points*, also called *observations*. A data point can relate to a single individual, or data from a particular moment in time. Data points consist of *identifiers* (or *explicit identifiers*), so the data point can be uniquely linked to a data subject

(e. g. social security number, passport number, user id, etc.), and *attributes*, which describe the data subject. Attributes are assumed to fall into one of the following categories [Dal86, BS08]:

1. *Sensitive Attribute* – An attribute whose value for each data subject we wish to hide from a data recipient, due to its sensitivity;
2. *Quasi-Identifier* – An attribute which is *non-sensitive* and cannot by itself uniquely identify an observation; however, when combined with other quasi-identifiers and/or another dataset, has the potential to successfully identify an observation.

Structured data is traditionally stored in tabular form, with observations in rows and attributes in columns, as in an Excel spreadsheet. Some data, although stored in tabular form, may require unique techniques to preserve privacy, due to the nature of the data. A number of such data types which we consider in this thesis are listed below:

1. *Location Data* – Location history of a person;
2. *Graph Data* – Data relating to a social network, communication network or physical network;
3. *Time Series* – Data which contains an element of updating in time, such as census information.

Various attempts have been made over the years to define what is meant by a privacy disclosure. What appears to have been the first attempt was made by Tore Dalenius in 1977 [Dal77]. Loosely stated, Dalenius suggested that a disclosure had taken place if an adversary could learn something about a subject which could only be learned with access to the dataset. In 2006, Cynthia Dwork showed that this restrictive standard of privacy protection was almost impossible to attain if any statistics are to be released [Dwo06].

Dalenius conceded that the elimination of disclosure was "not operationally feasible" and may only be possible by the elimination of the statistics themselves

Since then, further attempts have been made to classify various privacy disclosures [Lam93, TMKC14]:

1. *Identity Disclosure/Record Linkage* – when a data recipient can associate a data record with a data subject;

2. *Attribute Disclosure* – when a data recipient can determine a new attribute of a data subject;

3. *Inferential Disclosure* – when a data recipient is able to determine an attribute of a data subject more accurately than before.

Another interpretation of privacy draws a distinction between re-identification and unequivocal/authenticated re-identification, whereby increased uncertainty in successfully re-identifying a data point constitutes privacy [TP12, MW10]. These examples illustrate that there is no 'one size fits all' solution for privacy, and that the interpretation of privacy varies by application [Wu13].

2.4 PRIVACY METHODS

In 1989, Adam and Worthmann completed a detailed and exhaustive study of the methods available at the time [AW89], and many of their techniques are still in widespread use today. Privacy methods fall into two broad categories: *suppressive* and *perturbative* [WDW12]. A number of examples are given for each method below.

2.4.1 *Suppressive*

As the name suggests, suppressive methods make no change to the values of the data, but rather limit the quantity of data that is published. The primary suppressive methods are listed below:

1. *Cell Suppression* – Individual attributes or combinations of attributes of data points are suppressed in order to preserve privacy. In tabular data, examples include the suppression of entire rows and/or columns. An example of this method is shown in Figure 2.1;

2. *Query Restriction* – Where the data recipient has access to a database via a querying mechanism, certain queries can be suppressed if their answers are suspected of violating a given privacy protection requirement, or a limit can be placed on the number of queries that are answered;

4	3	1	3	11
0	1	0	0	1
2	2	1	0	5
1	3	4	2	10
7	9	6	5	27

(a)

4	3	1	3	11
0		0	0	
2	2	1	0	5
1	3	4	2	10
7		6	5	

(b)

4		1		11
0		0		1
2	2	1	0	5
1	3	4	2	10
7	9	6	5	27

(c)

Figure 2.1: Example of cell suppression method on a numeric table with marginals: (a) the input table with sensitive cell in the second row of the second column, (b) the sensitive cell suppressed, with additional suppressions in the marginals, (c) the sensitive cell suppressed, with additional suppressions within the table.

3. *Clustering/Recoding/Partitioning/Aggregating* – Clustering data involves merging one or more groups of data points together;

4. *Sampling* – Instead of releasing an entire dataset, the sampling method randomly chooses a subset of the dataset to release. This differs from the cell suppression method, where the cells are chosen deterministically.

2.4.2 *Perturbative*

In contrast, perturbative methods apply changes to the value of attributes in a data point. The primary perturbative methods are listed below:

1. *Adding Noise* – Random noise is added to the value of the data; this method is generally only applicable to continuous numerical data (e. g. heights, salaries, ages, etc.), although discrete distributions can be used to add noise to discrete numeric values, such as integers;

2. *Rounding* – Original values in this case are replaced by rounded values; this is not entirely different to clustering in the suppressive methods, but is considered a perturbative method as the value of the data is changed.;

3. *Re-Sampling* – A synthetic dataset is generated, sampled from a distribution created from the original dataset; any similarities to an individual are therefore coincidental;

4. *Swapping* – Two variants of swapping exist; (i) *data swapping* is the swapping of individual values in a dataset; (ii) *rank swapping* is generally conducted on numeric values where the full set of values is sorted, and values are then swapped within a restricted range of their own rank.

2.5 PRIVACY MODELS

A more exhaustive survey of SDC models can be found in [FWCY10]

We now examine a number of privacy models that have been developed by the SDC community over recent decades. SDC has been in existence since the late 1970s, and detailing every one of the models is beyond the scope of this thesis. Instead, we focus on those we consider to be of greatest relevance to the work presented in later chapters of this thesis.

We begin by looking at earlier techniques of ad hoc anonymisation, before specialising to the more recent developments of k-anonymity and differential privacy. We also examine randomised response, a special adaptation of local privacy in surveying.

2.5.1 *Ad Hoc Anonymisation*

Anonymising data, also referred to as de-identification in the literature, involves removing all identifiers (e. g. name, social security number, telephone number, address, etc.) that are unique to a person. Anonymisation is still, in some communities, considered an adequate method to preserve privacy in microdata releases [CC14].

As demonstrated by the privacy failings detailed in Section 2.1.4, data curators considered it sufficient to remove these explicit identifiers from the data to preserve the subjects' privacy.

LINKAGE ATTACKS Even with explicit identifiers removed, the data is still vulnerable to a *linkage attack*, whereby an adversary uses the quasi-identifiers to link the dataset with another still containing explicit identifiers. This approach was the one used by Sweeney in re-identifying the GIC dataset, as shown in Figure 2.2.

Section 2.1.4.1

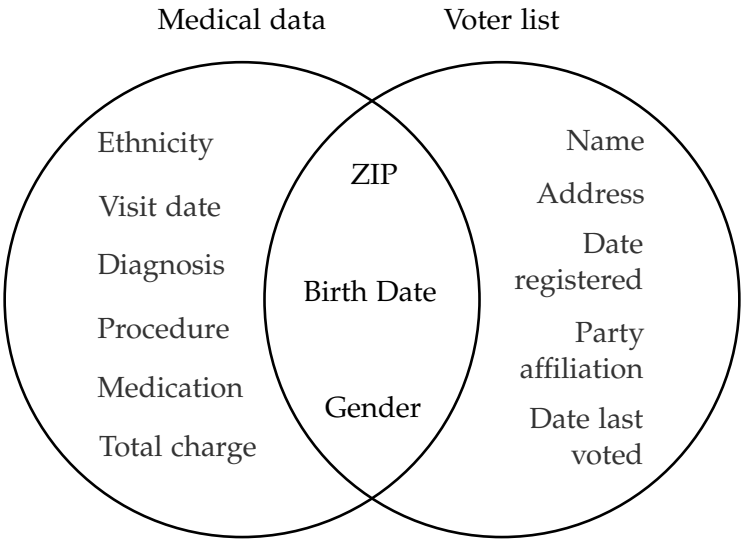

Figure 2.2: An illustration of Sweeney's linkage attack on the GIC dataset. The circle on the left shows the information present in GIC's medical dataset, and the circle on the right shows the information present in the voter registration list. The two dataset could be linked by using the common quasi-identifiers of ZIP code, data of birth and gender.

Linkage attacks in anonymised datasets are possible even when no obvious quasi-identifiers are present. The dataset released as part of the Netflix prize contained no obvious quasi-identifiers, but was still vulnerable to a linkage attack using IMDB data.

Section 2.1.4.3

OTHER ATTACKS Sometimes, the anonymised data is rich enough in detail to allow re-identification. In this case, the data itself acts as a quasi-identifier, as demonstrated in the AOL Research scandal.

Section 2.1.4.2

Anonymisation has also proven ineffective when dealing with other types of data. Releasing the structure of networks (graphs) has proven to be vulnerable to a variety of attacks [BDK07], while human location history has been shown to be non-private even with a small amount of additional information [dMHVB13].

A white paper by Narayanan and Felten in 2014 resolutely rejected claims by Cavoukian and Castro from the same year that anonymisation was an adequate means to preserve privacy [NF14, CC14]. Conversely, a pair of publications in 2015 and 2016 respectively dismissed and advocated the use of anonymisation in publishing credit card metadata [dMRSP15, SMDF16].

2.5.2 *k-Anonymity*

Section 2.1.4.1

In response to her re-identification of the GIC dataset in 1997, Sweeney proposed a new suppressive privacy model, called k-anonymity. Her model ensures that any individuals' quasi-identifiers would be indistinguishable from at least $(k-1)$ others, by implementing a series of suppressions on the data [Swe02].

At first glance, the model appeared to serve its function. The attack she performed on the GIC dataset would not have been possible with k-anonymity. However, flaws were soon found in the model. For example, if a group of k records with the same quasi-identifiers (suppressed in some way) were to each have the same sensitive attribute value, re-identification was possible.

Various enhancements to k-anonymity were subsequently proposed. ℓ-diversity was first developed in [MKGV07] to introduce diversity within the k-anonymity classes. Then, t-closeness was proposed in [LLV07] to ensure the k-anonymity classes fit closely to the distribution of the population. In hindsight however, these were simply reactionary improvements to protect against attacks previously detailed in the literature.

2.5.3 *Differential Privacy*

Around the same time as k-anonymity was proposed, another group of computer scientists were, independently, investigating an alternative approach to privacy in statistical databases.

BUILD-UP In 2003, NEC researchers Dinur and Nissim set about measuring the privacy of data releases [DN03]. Limiting their analysis to count queries, they measured privacy by the ability of an adversary to reconstruct (a portion of) a database. They derived lower bounds for the noise one needed to add to the count queries to maintain privacy.

Dinur and Nissim's proposal was developed in a series of papers up to 2006 [DN04, BDMN05, DMNS06, DKM+06]. Through the papers, the researchers reasoned that privacy was a function of the mechanism, not the

dataset. They also converged on the idea that a mechanism was private if similar databases were indistinguishable through that mechanism.

INCEPTION The concept of privacy pioneered by [BDMN05] was later refined by Dwork, leading to her proposal of *differential privacy* in 2006 [Dwo06]. She reasoned that the distribution of query answers from neighbouring databases should be similar (within a multiplicative fraction). This was simply a rephrasing of the conclusions of the previous works mentioned above.

Dwork assumes each dataset to consist of rows, each of which corresponds to a single data subject's information. If the rows of each dataset lie in D (D can be multi-dimensional), then a dataset $\boldsymbol{d}$ of n rows lies in the set D^n.

We let the randomised output of a query Q on $\boldsymbol{d}$ be denoted by the random variable $X_{Q,\boldsymbol{d}}$ and denote by E_Q the range of Q. Using $\epsilon \geq 0$ as a privacy parameter, then ϵ-differential privacy is satisfied when the following inequality holds for all databases $\boldsymbol{d}, \boldsymbol{d}' \in D^n$ differing in one row and for all measurable subsets $A \subseteq E_Q$:

A formal definition of differential privacy is presented in Chapter 3

$$\mathbb{P}(X_{Q,\boldsymbol{d}} \in A) \leq e^{\epsilon}\mathbb{P}(X_{Q,\boldsymbol{d}'} \in A). \tag{2.1}$$

Under Dwork's definition, an adversary would be unlikely to know if a particular person was in a database from the answer of any query, since answers from similar databases are indistinguishable within a given factor. Such a scenario would hopefully have the added effect of increasing user trust in data collection, as similar knowledge could be gained about an individual even if their data was not included. Phrased another way, a differentially private mechanism should reflect the dataset as a whole and not be sensitive to the values of individual entries.

The notion of *relaxed differential privacy* was first used in [KS08], having previously been noted in [DKM+06] but absent from the original definition of differential privacy. Relaxed differential privacy allows for more flexibility in designing mechanisms over strict differential privacy, especially when small or zero probabilities are involved. For relaxed differential privacy, a second privacy parameter $0 \leq \delta \leq 1$ is included in the original ϵ-differential privacy definition. We say (ϵ, δ)-differential privacy is satisfied if the follow-

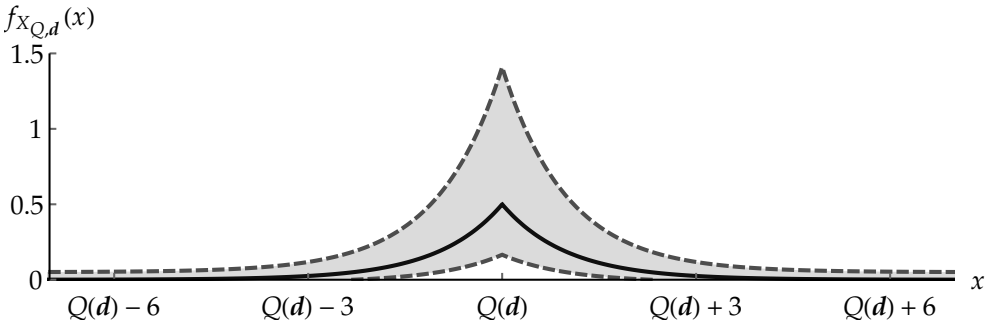

Figure 2.3: An illustration of (ϵ, δ)-differential privacy for $\epsilon = 1$ and $\delta = 0.05$. The probability density function of $X_{Q,d}$, $f_{X_{Q,d}}$, is shown in black. A necessary condition for (ϵ, δ)-differential privacy to hold is that the distributions of neighbouring databases lie in the shaded region.

ing inequality holds for all databases $d, d' \in D^n$ differing in one row and for all measurable subsets $A \subseteq E_Q$:

$$\mathbb{P}(X_{Q,d} \in A) \leq e^{\epsilon}\mathbb{P}(X_{Q,d'} \in A) + \delta. \tag{2.2}$$

A visual illustration of (ϵ, δ)-differential privacy is shown in Figure 2.3.

IMPLEMENTATION Dwork's original differential privacy work included the first implementation of a mechanism on continuous data. The *Laplace mechanism* was proposed as a way to achieve differential privacy, by adding noise sampled from a Laplace distribution.

If we are seeking to achieve ϵ-differential privacy on outputs of a query $Q : D^n \to \mathbb{R}^k$ using the Laplace mechanism, then we must first determine its sensitivity ΔQ,

$$\Delta Q = \max \|Q(d) - Q(d')\|_1, \tag{2.3}$$

where the maximum is taken over all $d, d' \in D^n$ that differ in one row.

It was shown that if we add noise sampled from a Laplace distribution with mean 0 and variance $2\left(\frac{\Delta Q}{\epsilon}\right)^2$, then that mechanism $X_{Q,d}$ will satisfy ϵ-differential privacy:

$$X_{Q,d} = Q(d) + \left(\text{Laplace}\left(0, \frac{\Delta Q}{\epsilon}\right)\right)^k, \tag{2.4}$$

where the second term is shorthand for a k-dimensional Laplace random variable, with mean 0 and scale $\frac{\Delta Q}{\epsilon}$.

In cases where adding numeric noise made no sense, McSherry and Talwar [MT07] proposed the *exponential mechanism* for achieving differential privacy, by using a utility function to select an appropriate output. Namely, suppose that Q maps into a discrete set E_Q. The exponential mechanism chooses a response $q \in E_Q$ based on its utility to the input database $\boldsymbol{d}$. If the utility function is $u : D^n \times E_Q \to \mathbb{R}$, then we define the mechanism as

$$\mathbb{P}(X_{Q,\boldsymbol{d}} = q) = \frac{e^{\epsilon u(\boldsymbol{d},q)}}{\int_{E_Q} e^{\epsilon u(\boldsymbol{d},p)} dp}. \tag{2.5}$$

If we define

$$\Delta u = \max |u(\boldsymbol{d}, q) - u(\boldsymbol{d}', q)|,$$

where the maximum is taken over all $\boldsymbol{d}, \boldsymbol{d}' \in D^n$ that differ in one row and over all $q \in E_Q$, then McSherry and Talwar showed that the exponential mechanism satisfies $(2\epsilon\Delta u)$-differential privacy.

ADOPTION Equipped with a rigorous definition of privacy and simple mechanisms with which to achieve it, differential privacy quickly gained traction in the computer science community. Within a few short years, a wealth of literature had emerged, analysing differential privacy in depth, the highlights of which we now briefly review. A more thorough introduction to the techniques of differential privacy can be found in [DR14].

Early on, the claim of differential privacy's resistance to auxiliary or side information was tested and shown in [KS08, GKS08]. In fact, it was found that more relaxed definitions of differential privacy were resistant to auxiliary information attacks.

[KLN+11] examined the application of differential privacy to learning algorithms and produced results making such an application possible. Differential privacy's utility guarantee was investigated by [GRS12] while its privacy guarantees were examined by [KM11, AAC+11].

The geometry of differential privacy was analysed by [HT10] and produced bounds for the noise to be added on linear queries. Their results effectively put a limit on the number of queries that can be asked of an interactive database, a shortcoming we address in a later chapter. This geometrical work was extended in [NTZ13] to relaxed differential privacy, again providing er-

Chapter 3

ror bounds for differentially private queries. Further work on the high levels of noise required to achieve differential privacy was carried out in [NRS07]. Instead of deciding the noise level for a function based on its sensitivity to all possible databases (global sensitivity), they proposed *smooth sensitivity* of functions to add lower, instance-based noise.

[LC11] examined the question of choosing an appropriate ϵ for differential privacy, a question largely glossed over until then in favour of theoretical investigations.

Optimality of mechanisms has also been of great interest. While the Laplace mechanism has received much attention, the staircase mechanism was derived from it in later work [GV14]. The staircase mechanism was shown to add less noise than the Laplace mechanism while achieving the same privacy standard, thereby giving more accurate results, especially for large ϵ. The optimality of the staircase mechanism was further analysed in [GKOV15].

We build on these findings in Chapter 5

Optimal mechanisms have also been investigated in the relaxed differential privacy and local differential privacy settings [GV13, KOV14, KSS16]. In the local privacy case, [KOV14] derived optimal mechanisms taking inspiration from the staircase mechanism mentioned previously.

The connection between differential privacy and k-anonymity has been shown in multiple papers. [LQS12] showed that differential privacy is obtained from k-anonymity under certain conditions. [DFSC15] went a step further and showed that t-closeness and differential privacy are equivalent. They used a stochastic formulation of t-closeness, the k-anonymity enhancement, for their results.

Statistical sampling was not considered in Dwork's original definition of differential privacy, where datasets were considered to be deterministic. [HRW11] looked at the application of differential privacy to databases whose data is created by randomly sampling from a larger population.

While differential privacy was conceived to block threats from adversaries with infinite computing power, the reality is that computation is limited. [MPRV09] formulated a relaxed form of differential privacy for computationally-bounded adversaries.

Differential privacy has also been used as stepping-stone to formulating new, more general privacy definitions. One example includes [KM14], where

a Bayesian approach to privacy was taken by generalising differential privacy and its notion of indistinguishability. [BKOZB12] also employed a generalisation of differential privacy in developing a framework for verifying the privacy-preserving nature of programmes in practice.

SPECIFIC APPLICATIONS Along with the theoretical investigations of differential privacy, the model has been applied to specific applications in SDC also. Again, we limit our survey to some of the most relevant results.

Graphs, in particular, have received considerable attention in looking for differential privacy applications. As noted in Section 2.5.1, graph structure can be vulnerable to attack. One paper of note is [KRSY11], which looked at releasing differentially private counts of a graph's subgraphs (triangles, stars, etc.). Such subgraph counts are frequently useful for detailed graph analysis, and can also be used to construct synthetic graphs that mimic features of the original graph. The approach of [KRSY11] made use of smooth sensitivity introduced in [NRS07] to satisfy differential privacy while adding less noise than the Laplace mechanism. Their implementation greatly improved on previous work by [RHMS09], which achieved a weaker notion of privacy while adding greater noise.

Releasing the degree distribution in a differentially private manner has been examined in [HLMJ09, KS12]. The former also took a deeper look at how the original definition of differential privacy relates to graphs. They proposed two main adoptions, *edge privacy* and *node privacy*, whereby 'neighbouring' graphs are defined by ones differing in one edge/node. [TC12] later proposed *out-link privacy* for directed graphs, where neighbouring graphs differ in the data provided by a particular individual (i. e. all the directed links *from* that individual to others).

Another interesting approach was taken by [SZW+11], whereby they created synthetic graphs from noisy graph statistics. This was similar to the method briefly mentioned by [KRSY11] in privately recreating graphs from perturbed statistics.

Other interesting applications have also been examined for differential privacy. The release of set-valued data was studied in [CMF+11], while differentially private data mining was discussed in [FS10, MCFY11]. Differential

privacy has also been analysed in the context of principal component analysis in [CSS12], having taken inspiration from McSherry and Talwar's exponential mechanism to derive a more optimal solution for their specific application. Privatising location data with differential privacy was investigated in [ABCP13] by using a planar Laplace distribution to add noise. Finally, differential privacy on time-series data was considered in [RN10, SCR+11].

USE IN PRACTICE While differential privacy has been an interest to researchers for over a decade, uptake within the IT industry has been less prevalent. A paper in 2012 suggested Facebook was using differential privacy in its advertising system to protect its users [CK12]. This was in response to Korolova's revelations that advertisers could infer a user's private information by tailoring their targetting of ads [Kor10].

Section 2.1.4.4

In 2016, Apple announced that the latest version of its operating system, iOS 10, would use differential privacy when collecting usage statistics [Gre16a, Gre16b]. "Starting with iOS 10, Apple is using technology called differential privacy to help discover the usage patterns of a large number of users without compromising individual privacy," Apple explained in a press release [App16].

FUTURE Despite its widespread use within the computer science community, differential privacy is not without its critics. While the original definition in [Dwo06] was useful in its generality, mechanisms are now being developed for specific applications.

A detailed critique was published in [BMS14], highlighting the large amount of noise required to achieve differential privacy in general. Although published 8 years after the first paper on differential privacy, the critique made exclusive use of Dwork's Laplace mechanism, which has since been proven to add more noise than necessary [GV14]. As detailed above, methods are now being developed for specific applications aimed at improving the utility of the results.

However, even more research is needed to fully realise the potential of differential privacy in certain applications. In healthcare for example, existing differential privacy techniques were found to add too much noise for

queries to be useful by [DEE12], and recommended that more research take place before its adoption in the industry.

The original definition of differential privacy as holding over 'neighbouring' databases is also being relaxed. Differential privacy is being used simply as a framework to add noise or perturb data. This is especially evident in the research of local differential privacy, whereby the perturbation to data is conducted by the data subject before supplying that data to the curator. Local differentially private data publishing was examined in [SS14], while optimality of local differential privacy mechanisms was investigated in [DJW13, KOV14].

Achieving differential privacy in randomised response has also been studied [WWH15, WN16], a subject we discuss further below.

2.5.4 *Randomised Response*

Arguable the oldest privacy model in existence, Randomised Response (RR) dates back to 1965 when it was proposed by Stanley L. Warner as a method to eliminate bias in surveying [War65]. Instead of answering a survey question truthfully (or not answering out of fear or embarrassment), subjects would answer in a randomised way.

INCEPTION Warner's original idea related to binary surveys, such as finding the proportion of the population possessing a certain attribute (e. g. people who have cheated on their partner). Each subject would be presented with a spinner which they would spin in private to decide which question to answer, either 'Have you ever cheated on your partner?' or 'Have you always been faithful to your partner?'. Such a set-up affords the subjects *plausible deniability*, as the curator can't know for sure whether each subject possesses the attribute or not. Even without knowing which question each subject answered, the curator would be able to estimate the proportion of the population possessing the attribute from the number of 'Yes' and 'No' responses. The spinner can be replaced by any appropriate randomisation device, such as flipping a coin, rolling a dice or drawing from a pack of cards.

FURTHER RESEARCH A rich body of literature now exists on RR. The inefficiencies of Warner's original RR model have been tackled by a host of authors and many new RR models have been proposed. These include the unrelated question model [GAESH69], the forced response model [Bor71], Moor's procedure [Moo71] and two-stage RR models [NSM90]. Non-binary questioning has received attention, both for numeric values [GJAH71, GS07a] and categorical values [AEGH67, GAESH69]. Privacy protection metrics have also been considered on RR [Lan76, LW76, Loy76, Bos15].

Recently, new randomisation techniques have been proposed for RR. For example, Tian's parallel model uses the inherent randomness of people's attributes (e. g. month or day of birth) to introduce uncertainty [Tia14]. A similar technique was proposed in [MM12] when seeking answers from multiple questions without the need for multiple randomisations. More comprehensive lists of RR models can be found in [Kru13, BIZ15].

One model which we focus on in this thesis is the model proposed by Mangat as an improvement to Warner's in [Man94]. Under the assumption that possession of the attribute is sensitive, only the responses of those not possessing the attribute should be randomised. Members possessing the attribute would therefore answer truthfully, but would still be afforded plausible deniability by the randomised responses of the rest of the population. This deniability is strongest when the proportion of the population possessing the attribute is small. Mangat's model has previously been examined in [Moo97, GS07b].

CRITICISMS Researchers remain divided on the effectiveness of RR however. While some works have shown RR to be an improvement on direct questioning and other methods [vdHvGBH00, GG75, LSOE04, Kru12, TF81], others remain sceptical on its advantage [WS94, WP13, LHG97]. Public trust in RR has also been shown to be lacking [CJ11]. As few as 15 per cent of subjects trusted the RR technique to preserve privacy, and only 21 per cent trusted the method when randomisation was achieved by way of a coin toss

USE IN PRACTICE Nevertheless, RR is actively used in surveying when asking sensitive questions. Examples include surveys on doping and drug

use in elite athletes [SUS10], cognitive-enhancing drug use among university students [DSF+13], faking on a CV [DDH03], corruption [Gin10], sexual behaviour [CDJ+14], and child molestation [FL88]. Google also uses RR to privately collect statistics for web browser settings [EPK14].

2.6 CONCLUDING REMARKS

The goal of this chapter was to introduce the reader to the foundations of Statistical Disclosure Control (SDC) and the history of differential privacy and Randomised Response (RR). The main items considered were as follows:

- A brief look at the legal history of privacy in general and data privacy;
- An introduction to SDC and a review of a number of high-profile data privacy breaches;
- An overview of the primary perturbation locations, data types and perturbation methods considered in the literature;
- An overview of a number of privacy models, with primary focus on the history of differential privacy and RR.

3

FORMULATING DIFFERENTIAL PRIVACY

In this chapter, we formulate a single unifying mathematical framework for differential privacy and establish a number of fundamental results. Through an abstract approach, the framework can be applied to all major data types, query types and perturbation methods considered in the literature thus far. We also establish general utility bounds for differentially private mechanisms.

OVERVIEW

3.1 INTRODUCTION

As was established in Chapter 2, differential privacy has gained traction within the computer science research community as a protocol to measure privacy. Various data types, query types and perturbation types have been considered, but differential privacy has thus far lacked a single generalised framework to deal with this variation of data and methods of data publication.

This chapter is dedicated to developing a single unifying mathematical framework for differential privacy. Under our abstract approach, all major data types, query types and perturbation methods can be handled using a

single framework. This allows us the advantage of deriving results in generality that can be applied uniformly to a wide variety of applications.

We begin in Section 3.2 by formulating the framework, formally defining the concept of differential privacy in this framework, and introducing specialisations to specific methods of perturbation. In Section 3.3 we briefly consider sufficient conditions for differential privacy to hold on a mechanism and consider the special case of the identity query in Section 3.4. In Section 3.5 we formulate a realisation of local privacy in our framework, and present initial results on utility in Section 3.6. Concluding remarks are given in Section 3.7.

3.2 PRELIMINARIES

We first recall some standard concepts and results from probability and measure theory [Rud87, Bil95]. Given any collection $\mathcal{A}$ of subsets of a set Ω, we use $\sigma(\mathcal{A})$ to denote the smallest *σ-algebra* containing $\mathcal{A}$ and refer to $\sigma(\mathcal{A})$ as the σ-algebra generated by $\mathcal{A}$. A set Ω together with a σ-algebra of subsets of Ω is a *measurable space*.

Given a mapping $Q : U \to E$ from a set U to a set E and a subset $A \subseteq E$, the notation $Q^{-1}(A)$ denotes the *pre-image* of A, $Q^{-1}(A) = \{u \in U : Q(u) \in A\}$.

We denote by $[n]$ the set $[1, n] \cap \mathbb{Z}$ for any $n \in \mathbb{Z}$. Note that $[n] = \emptyset$ when $n < 1$. Similarly, we denote by $[n]_0$ the set $[n] \cup \{0\}$ for any $n \in \mathbb{Z}$.

A *monotone class* $\mathcal{M}$ of subsets of some set Ω is defined by the following two properties:

(i) if $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{M}$, and if $A_i \subseteq A_{i+1}$ for all i, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{M}$;

(ii) if $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{M}$, and if $A_i \supseteq A_{i+1}$ for all i, then $\bigcap_{i=1}^{\infty} A_i \in \mathcal{M}$.

The next result, which appears as Theorem 3.4 in [Bil95], characterises $\sigma(\mathcal{A})$ as the smallest monotone class containing $\mathcal{A}$.

Theorem 3.1. *Let $\mathcal{A}$ be an algebra of subsets of some set Ω and let $\mathcal{M}$ be a monotone class such that $\mathcal{A} \subseteq \mathcal{M}$. Then $\sigma(\mathcal{A}) \subseteq \mathcal{M}$.*

Given two measurable spaces $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$, subsets of $X \times Y$ of the form

$$R = \bigcup_{i=1}^{p} X_i \times Y_i,$$

where $X_i \in \mathcal{A}_X$, $Y_i \in \mathcal{A}_Y$ for $i \in [p]$ and $(X_i \times Y_i) \cap (X_j \times Y_j) = \emptyset$ for $i \neq j$, are known as *elementary subsets*. Let $\mathcal{R}$ denote the collection of all elementary subsets and denote by $\mathcal{A}_{X\times Y}$ the usual product σ-algebra on $X \times Y$ generated by $\{A \times B : A \in \mathcal{A}_X, B \in \mathcal{A}_Y\}$. The following result is Theorem 8.3 of [Rud87].

Theorem 3.2. *If $\mathcal{M}$ is a monotone class and $\mathcal{R} \subseteq \mathcal{M}$, then $\mathcal{A}_{X\times Y} \subseteq \mathcal{M}$.*

Finally, for a measure μ on a measurable space $(X, \mathcal{A}_X)$, we recall the following fact.

Proposition 3.1. *Suppose that $\{A_i\}_{i=1}^{\infty} \subseteq \mathcal{A}_X$ satisfies $A_i \subseteq A_{i+1}$ for all i, then $\lim_{i\to\infty} \mu(A_i) = \mu\big(\bigcup_{i=1}^{\infty} A_i\big)$.*

Similarly, if $A_i \supseteq A_{i+1}$ for all i, then $\lim_{i\to\infty} \mu(A_i) = \mu\big(\bigcap_{i=1}^{\infty} A_i\big)$.

3.2.1 *Database Model*

The individual entries of the databases we consider are elements of a set $D \subseteq U$ where U is a metric space with metric ρ. We equip U with the Borel σ-algebra generated by the open sets in U (in the metric topology); D then naturally inherits a σ-algebra $\mathcal{A}_D$. A database $\boldsymbol{d}$ with n rows is given by a vector $\boldsymbol{d} = (d_1, \ldots, d_n) \in D^n$ in which $d_i \in D$ is the ith row. Throughout, we assume that U^n (and D^n) is equipped with the usual product σ-algebra $\mathcal{A}_{U^n}$ generated by $\{A_1 \times \cdots \times A_n : A_i \in \mathcal{A}_U\}$. This ensures that projection maps $\pi_i : U^n \to U$ given by $\pi_i(x_1, \ldots, x_n) = x_i$ are measurable.

Definition 3.1 (Neighbouring databases)**.** *We say that two databases $\boldsymbol{d} = (d_1, \ldots, d_n)$ and $\boldsymbol{d}' = (d'_1, \ldots, d'_n)$ in D^n are* neighbours*, and write $\boldsymbol{d} \sim \boldsymbol{d}'$, if there is some $j \in [n]$ such that $d_j \neq d'_j$ and $d_i = d'_i$ for all $i \in [n] \setminus \{j\}$.*

We will write $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$ as shorthand notation for $\boldsymbol{d}, \boldsymbol{d}' \in D^n$ and $\boldsymbol{d} \sim \boldsymbol{d}'$. More generally, we define the *Hamming distance* between two databases $\boldsymbol{d}$ and $\boldsymbol{d}'$.

Definition 3.2 (Hamming Distance)**.** *The Hamming distance,* $\mathrm{h} : D^n \times D^n \to [n]_0$*, between two databases is the number of rows on which they differ:*

$$\mathrm{h}(\boldsymbol{d}, \boldsymbol{d}') = |\{i : d_i \neq d'_i\}|. \tag{3.1}$$

Hence $\boldsymbol{d} \sim \boldsymbol{d}'$ *if and only if* $\mathrm{h}(\boldsymbol{d}, \boldsymbol{d}') = 1$.

It is worth highlighting the generality of this setting: the metric space D can contain numerical, categorical or functional data; moreover, it can be discrete or continuous.

Example 3.1. If our data concern the hobbies or interests of people, we consider a set of all possible hobbies, denoted by $\mathcal{H}$. For simplicity it is not unreasonable to assume that $\mathcal{H}$ is finite. Our data entries are then drawn from the power set $D := \mathcal{P}(\mathcal{H})$ of $\mathcal{H}$, which will again be a finite set.

There are various natural choices of metric in this case. We could consider the discrete metric on D in which $\rho_1(A, B) = 1$ if $A \neq B$ and 0 otherwise. Alternatively, we could choose the metric given by symmetric distance: $\rho_2(A, B) = |(A \cup B) \backslash (A \cap B)|$. In both of these cases, the Borel σ-algebra consists of all subsets of D. Note that there is no requirement that each entry in a database in D^n have the same size or cardinality, reflecting the fact that not all of us have the same number of interests or hobbies.

Example 3.2. In readings from field deployed sensors, each reading has a time-stamp giving rise to time-course data. A similar example is in smart metering where the supplier collects data from consumers giving electricity consumption over a time-window. Data of this type is naturally represented as either a function or a sequence of real numbers.

In our framework, we can take U to be a sequence space such as ℓ^∞ or ℓ^2, or an appropriate function space such as $C([0, T])$ or $L^2([0, T])$, where T represents the billing period (for instance). All of these spaces have natural norms defined on them (in fact they are all Banach spaces) and can be equipped with the Borel σ-algebra generated from the open sets in the norm topology.

For the most part, we assume that the data space D is compact. This is immediate if D is finite (as in Example 3.1) and is a natural assumption in

most realistic situations. When D is compact, we denote by $\mathrm{diam}(D)$ the diameter of D:

$$\mathrm{diam}(D) = \max_{d,d' \in D} \rho(d, d'). \tag{3.2}$$

3.2.2 *Query Model*

We consider a very general query model. The set of all possible responses is denoted by E_Q while $\mathcal{A}_Q$ is a σ-algebra of subsets of E_Q. In the case where E_Q is a metric space, we assume that $\mathcal{A}_Q$ is the Borel σ-algebra generated by the metric topology. A query Q is then a measurable function, $Q : U^n \to E_Q$ and hence $Q^{-1}(A) \in \mathcal{A}_{U^n}$ for all $A \in \mathcal{A}_Q$.

A query is denoted by $(Q, E_Q, \mathcal{A}_Q)$, which, for simplicity and where there is no ambiguity, we will write simply as Q.

Example 3.3. As with the data in $\boldsymbol{d}$, queries are not restricted to take numerical values in this setting. For instance, if we consider Example 3.1 above, then we could consider a query asking for the number of people in the database who are interested in Classical Music or Football for instance: this would clearly be a numeric query. On the other hand, we could also request the 3 most common hobbies in the database, the output of which would be a categorical set.

3.2.3 *Response Mechanism*

We now formally introduce the concept of a *response mechanism* within this general framework. As is standard in the literature on Probability Theory, we assume that a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is given [Bil95]. Here, Ω is a set (corresponding informally to the set of outcomes of some 'random experiment'), $\mathcal{F}$ is a σ-algebra of subsets of Ω (corresponding to 'events' of interest) while $\mathbb{P}$ is a probability measure defined on $\mathcal{F}$.

We define a response mechanism with respect to a set of queries, $\mathcal{Q}$ which represent the queries of interest for a particular application. A response mechanism with respect to this set is a collection of measurable mappings

$$\{X_{Q,\boldsymbol{d}} : \Omega \to E_Q \mid Q \in \mathcal{Q}, \boldsymbol{d} \in D^n\}. \tag{3.3}$$

Note that $X_{Q,\boldsymbol{d}}$ is an E_Q-valued random variable for each Q and $\boldsymbol{d}$. Response mechanisms are typically described as 'randomised algorithms' in the literature [Dwo06, MT07, KS08], where each query Q is represented by a randomised algorithm as a function of $\boldsymbol{d}$.

We remind the reader that $\mathbb{P}(X_{Q,\boldsymbol{d}} \in A)$ is standard shorthand in probability theory for $\mathbb{P}(\{\omega \in \Omega : X_{Q,\boldsymbol{d}}(\omega) \in A\})$. Throughout the thesis, in cases where there is no ambiguity, the argument ω of $X_{Q,\boldsymbol{d}}(\omega)$ will be suppressed. Furthermore, where there is no ambiguity, the response mechanism will be written as $\{X_{Q,\boldsymbol{d}}\}$.

We now present two specialisations of this general response mechanism, where the perturbation is applied (i) before querying, and, (ii) after querying.

3.2.3.1 *Sanitised Response Mechanisms*

Much of our work in this and subsequent chapters is focused on a particular type of mechanism, the sanitised response mechanism. First of all, we introduce the notation of a *sanitisation* which is a family of measurable mappings

$$\{Y_{\boldsymbol{d}} : \Omega \to U^n \mid \boldsymbol{d} \in D^n\}. \tag{3.4a}$$

This represents 'sanitising' the original database. If the database is sanitised by adding appropriate noise, the mapping $Y_{\boldsymbol{d}}(\omega)$ takes the form $Y_{\boldsymbol{d}}(\omega) = \boldsymbol{d} + N(\omega)$ for some U^n-valued random vector $N(\omega)$. Note that in order to define a mechanism by adding noise, it is necessary for U^n to have a suitable algebraic structure in order for the operation of addition to make sense. However, our framework does not require this extra structure to be present in general.

A response mechanism $\{X_{Q,\boldsymbol{d}}\}$ is said to be a *sanitised response mechanism* if $X_{Q,\boldsymbol{d}}$ takes the particular form

$$X_{Q,\boldsymbol{d}}(\omega) = Q(Y_{\boldsymbol{d}}(\omega)) \tag{3.4b}$$

for all $Q \in \mathcal{Q}$ and $\boldsymbol{d} \in D^n$.

Written in full, the response mechanism is

$$\{X_{Q,\boldsymbol{d}} = Q \circ Y_{\boldsymbol{d}} : \Omega \to E_Q \mid Q \in \mathcal{Q}, \boldsymbol{d} \in D^n\}. \tag{3.4c}$$

Note that $\{Y_{\boldsymbol{d}}\}$ is itself a response mechanism as per (3.3), and corresponds to $\{X_{I,\boldsymbol{d}}\}$.

The motivation behind this choice of terminology is that the mechanism is generated by first sanitising the database via the sanitisation $Y_{\boldsymbol{d}}$ and then answering a query Q using the sanitised database. That is, by answering with $Q(Y_{\boldsymbol{d}}(\omega))$ rather than the true query answer $Q(\boldsymbol{d})$. This corresponds to Privacy-Preserving Data Publishing (PPDP) mentioned in Section 2.1.3.

3.2.3.2 *Output Perturbations*

In the differential privacy literature, output perturbation mechanisms which operate by perturbing or adding noise to the query response have attracted particular interest [Dwo06]. This corresponds to Privacy-Preserving Data Mining (PPDM) mentioned in Section 2.1.3. Such mechanisms can also be expressed in the general framework described here.

Let a query $Q : U^n \to E_Q$ be given. Assume that there is a family of measurable functions

$$\{Z_q : \Omega \to E_Q \mid q \in E_Q\} \tag{3.5a}$$

taking values in the output space E_Q of Q. Informally, for $q \in E_Q$, Z_q can be thought of as the random response of the mechanism when the true query response is q. An output perturbation mechanism is then defined as a response mechanism $\{X_{Q,\boldsymbol{d}}\}$ of the particular form

$$X_{Q,\boldsymbol{d}}(\omega) = Z_{Q(\boldsymbol{d})}(\omega), \tag{3.5b}$$

for each $Q \in \mathcal{Q}$ and $\boldsymbol{d} \in D^n$. Written in full, the response mechanism is

$$\{X_{Q,\boldsymbol{d}} = Z_{Q(\boldsymbol{d})} : \Omega \to E_Q \mid Q \in \mathcal{Q}, \boldsymbol{d} \in D^n\}. \tag{3.5c}$$

For real-valued data, the functions Z_q may take the form $Z_q(\omega) = q + N(\omega)$ where N represents the noise added to the query response as in the well-known Laplace mechanism. For set-valued queries, Z_q can be defined by specifying an appropriate probability mass function on E_Q.

3.2.4 *Differential Privacy*

In the interest of completeness and clarity, we now recall the definition of differential privacy and write it in the setting introduced in this section.

Definition 3.3 (Differential Privacy with respect to a Query)**.** *Let $\epsilon \geq 0$, $0 \leq \delta \leq 1$ be given. A response mechanism is (ϵ, δ)-differentially private with respect to a query $Q^* \in \mathcal{Q}$ if for all $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$ and all $A \in \mathcal{A}_{Q^*}$,*

$$\mathbb{P}(X_{Q^*,\boldsymbol{d}} \in A) \leq e^{\epsilon}\mathbb{P}(X_{Q^*,\boldsymbol{d}'} \in A) + \delta. \tag{3.6}$$

It is important to note that the relation $\boldsymbol{d} \sim \boldsymbol{d}'$ is symmetric, so inequality (3.6) is required to hold when $\boldsymbol{d}$ and $\boldsymbol{d}'$ are swapped.

Definition 3.4 (Differential Privacy)**.** *A response mechanism is (ϵ, δ)-differentially private with respect to set of queries $\mathcal{Q}$ if it is (ϵ, δ)-differentially private with respect to every query $Q^* \in \mathcal{Q}$.*

The above definitions are often referred to as *relaxed* differential privacy; the original notion of differential privacy (now known as *strict* differential privacy) introduced in [Dwo06] considers the case where $\delta = 0$. Relaxed differential privacy was first considered in [KS08].

It follows that a sanitisation $\{Y_{\boldsymbol{d}}\}$ is (ϵ, δ)-differentially private if $\mathbb{P}(Y_{\boldsymbol{d}} \in A) \leq e^{\epsilon}\mathbb{P}(Y_{\boldsymbol{d}'} \in A) + \delta$, for all $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$ and $A \in \mathcal{A}_{U^n}$.

3.3 SUFFICIENT SETS FOR DIFFERENTIAL PRIVACY

In this brief section, we consider the following question: is there a strict subset $\mathcal{S} \subset \mathcal{A}_Q$ such that if (3.6) is satisfied for all $A \in \mathcal{S}$, it is guaranteed to be satisfied for all $A \in \mathcal{A}_Q$? We refer to such a set as a *sufficient set* for differential privacy.

Depending on the application, the query output space may be a subset of $\mathbb{R}^n$ or of a sequence or function space such as $C([0,T])$. A key question for the practical deployment of differentially private mechanisms is how to determine if a mechanism is in fact (ϵ, δ)-differentially private. Testing (3.6) on the entire σ-algebra is a potentially difficult task. Our next result shows that it is sufficient to test this condition on any *algebra* $\mathcal{S} \subset \mathcal{A}_Q$ that generates $\mathcal{A}_Q$.

Theorem 3.3. *Let a query Q and a response mechanism $\{X_{Q,d}\}$ be given and let $\mathcal{S} \subset \mathcal{A}_Q$ be an algebra such that $\sigma(S) = \mathcal{A}_Q$. If (3.6) holds for all sets $A \in \mathcal{S}$, then it holds for all sets $A \in \mathcal{A}_Q$.*

Proof. Let $\mathcal{B}$ denote the collection of sets in E_Q for which (3.6) holds. By assumption, $\mathcal{S} \subseteq \mathcal{B}$. Now let $A_1, A_2, \ldots$ be any collection of sets in $\mathcal{B}$ with $A_i \subseteq A_{i+1}$ for all i. Define $\tilde{A} := \cup_i A_i$ and let $\boldsymbol{d}, \boldsymbol{d}' \in D^n$ with $\boldsymbol{d} \sim \boldsymbol{d}'$ be given. As each $A_i \in \mathcal{B}$, it follows that

$$\mathbb{P}(X_{Q,\boldsymbol{d}} \in A_i) \leq e^{\epsilon}\mathbb{P}(X_{Q,\boldsymbol{d}'} \in A_i) + \delta$$

for all i. As the sequence A_i is increasing, it now follows from Proposition 3.1 that

$$\begin{aligned}\mathbb{P}(X_{Q,\boldsymbol{d}} \in \tilde{A}) &= \lim_{i\to\infty} \mathbb{P}(X_{Q,\boldsymbol{d}} \in A_i) \\ &\leq e^{\epsilon} \lim_{i\to\infty} \mathbb{P}(X_{Q,\boldsymbol{d}'} \in A_i) + \delta \\ &= e^{\epsilon}\mathbb{P}(X_{Q,\boldsymbol{d}'} \in \tilde{A}) + \delta,\end{aligned}$$

and so $\tilde{A} \in \mathcal{B}$. An identical argument shows that for any sequence $\{A_i\}$ of sets in $\mathcal{B}$ with $A_i \supseteq A_{i+1}$ for all i, and $\tilde{A} = \cap_i A_i$, $\tilde{A} \in \mathcal{B}$. Taken together these two observations imply that $\mathcal{B}$ is a *monotone class*. Moreover, $\mathcal{S} \subseteq \mathcal{B}$. The result now follows immediately from Theorem 3.1. □

Using this result, we no longer have to check for differential privacy on countable unions of sets, instead checking on finite unions is sufficient. We show, in the example below, how Theorem 3.3 can be applied to differentially private mechanisms for functional data to obtain results such as those described in Section 3.1 of [HRW13].

Example 3.4. Suppose our query Q takes values in the space $C([0,1])$ of continuous functions on $[0,1]$ equipped with the norm $\|f\|_\infty = \sup\{|f(t)| : t \in [0,1]\}$ and the σ-algebra $\mathcal{A}_Q$ of Borel sets generated by the norm topology. Let a mechanism $\{X_{Q,\boldsymbol{d}}\}$ be given. Then $X_{Q,\boldsymbol{d}}(\omega) \in C([0,1])$ for each $\omega \in \Omega$.

Given a positive integer k and real numbers $0 \leq t_1 < \cdots < t_k \leq 1$, consider the projection $\pi_{t_1,\ldots,t_k} : C([0,1]) \to \mathbb{R}^k$ given by

$$\pi_{t_1,\ldots,t_k}(f) = (f(t_1), \ldots, f(t_k)).$$

These mappings are measurable with respect to the usual Borel σ-algebra on $\mathbb{R}^k$ and hence we can define the $\mathbb{R}^k$-valued mechanism $X^{t_1,\ldots,t_k}_{Q,\boldsymbol{d}} = \pi_{t_1,\ldots,t_k} \circ X_{Q,\boldsymbol{d}}$.

We claim that if the finite-dimensional mechanisms $\{X^{t_1,\ldots,t_k}_{Q,\boldsymbol{d}}\}$ are (ϵ,δ)-differentially private for all k and $t_1,\ldots,t_k$, then the mechanism $\{X_{Q,\boldsymbol{d}}\}$ is (ϵ,δ)-differentially private. The argument to show this is as follows. From the assumption on the finite-dimensional mechanisms, it follows immediately that (3.6) holds for all (so-called cylinder sets) sets A of the form

$$A = \pi^{-1}_{t_1,\ldots,t_k}(B)$$

where B is a Borel set in $\mathbb{R}^k$. These sets form an algebra and it follows from Theorem VII.2.1 of [Par67, Pg. 212] that the σ-algebra they generate is the Borel σ-algebra of $C([0,1])$. It follows immediately from Theorem 3.3 that $\{X_{Q,\boldsymbol{d}}\}$ defines an (ϵ,δ)-differentially private mechanism on $C([0,1])$ as claimed.

3.4 SANITISED RESPONSE MECHANISMS AND THE IDENTITY QUERY

A popular approach to designing differentially private response mechanisms is to add *noise* to the query response $Q(\boldsymbol{d})$. It is known however, that this can

lead to privacy compromises by averaging a large number of responses to an identical query [Dwo08], unless the number or type of queries that can be asked is restricted. We now show that if a sanitised response mechanism (3.4b) is (ϵ, δ)-differentially private with respect to the identity query, then it is (ϵ, δ)-differentially private with respect to *any* query.

It is important to appreciate that we place minimal restrictions on the query Q and its output space. For example, if we consider E_Q to be the space of square summable or bounded real sequences then Q naturally corresponds to an infinite sequence of individual real-valued queries, each of which is a projection of the given Q.

The identity query I is defined by the identity map on the ambient space U^n. Formally, $I(\boldsymbol{x}) = \boldsymbol{x}$ for all $\boldsymbol{x} \in U^n$. Note that an output space E_Q is associated with each query and that these spaces can be different for different queries; in the case of the identity query we select $E_Q = U^n$. In sanitised response mechanisms, the 'sanitised database' $Y_{\boldsymbol{d}}$ can be viewed as a response to the identity query. However, disclosing $Y_{\boldsymbol{d}}$ need *not* disclose the original database $\boldsymbol{d}$ provided that an appropriate privacy-preserving perturbation has been applied.

Importantly, if (ϵ, δ)-differential privacy is achieved with respect to I, then the response to *any* query is (ϵ, δ)-differentially private.

Theorem 3.4 (Identity Query)**.** *Consider a sanitised response mechanism as defined in (3.4b). Suppose that the sanitisation $\{Y_{\boldsymbol{d}}\}$ is (ϵ, δ)-differentially private. Then the mechanism (3.4b) is (ϵ, δ)-differentially private with respect to any query Q.*

Proof. Let $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$. By assumption,

$$\mathbb{P}(Y_{\boldsymbol{d}} \in E) \leq e^{\epsilon}\mathbb{P}(Y_{\boldsymbol{d}'} \in E) + \delta, \tag{3.7}$$

for all $E \in \mathcal{A}_{U^n}$.

Let a query Q and $A \in \mathcal{A}_Q$ be given. As Q is measurable, $Q^{-1}(A) \in \mathcal{A}_{U^n}$. Then, using (3.7),

$$\begin{aligned}\mathbb{P}(X_{Q,\boldsymbol{d}} \in A) &= \mathbb{P}(Q(Y_{\boldsymbol{d}}) \in A) \\ &= \mathbb{P}(Y_{\boldsymbol{d}} \in Q^{-1}(A)) \\ &\leq e^{\epsilon}\mathbb{P}(Y_{\boldsymbol{d}'} \in Q^{-1}(A)) + \delta \\ &= e^{\epsilon}\mathbb{P}(Q(Y_{\boldsymbol{d}'}) \in A) + \delta \\ &= e^{\epsilon}\mathbb{P}(X_{Q,\boldsymbol{d}'} \in A) + \delta.\end{aligned}$$

Thus, the mechanism satisfies (ϵ, δ)-differential privacy with respect to Q also. □

Remark: The previous result shows that the following natural intuition is valid in the abstract setting described here. If we can release a sanitised version of the database in an (ϵ, δ)-differentially private manner, then we can answer any query on this sanitisation in an (ϵ, δ)-differentially private manner also. From a privacy perspective at least, we can do no worse than to release all of the data.

The following worked example highlights a fundamental difference between sanitised response mechanisms and those based on output perturbations. We will see that the conditions for (ϵ, δ)-differential privacy can be violated easily for repeated queries when using output perturbation mechanisms that are (ϵ, δ)-differentially private for a single query.

Example 3.5. Consider the scenario where the same query is asked multiple (k) times on a database. We can easily model this as a single query $Q^{(k)}$ which maps from U^n to $E_Q^k = E_Q \times E_Q \times \ldots \times E_Q$ (the k-fold direct product of E_Q) and is given by $Q^{(k)}(\boldsymbol{d}) = (Q(\boldsymbol{d}), Q(\boldsymbol{d}), \ldots, Q(\boldsymbol{d}))$. Theorem 3.4 shows that if the sanitisation $Y_{\boldsymbol{d}}$ is (ϵ, δ)-differentially private then the sanitised response mechanism given by $Q^{(k)} \circ Y_{\boldsymbol{d}}$ is also (ϵ, δ)-differentially private for all $k \geq 1$.

In contrast, consider the situation for output perturbation mechanisms. For simplicity suppose we have a binary-valued query $Q : U^n \to \{0, 1\}$. So $E_Q = \{0, 1\}$ and to define the output perturbation, we need to specify the distributions of Z_0 and Z_1. If we put $\mathbb{P}(Z_i = i) = 1 - p$, $\mathbb{P}(Z_i \neq i) = p$ for $i = 0, 1$, then it is not difficult to verify that the output perturbation mechanism $X_{Q,\boldsymbol{d}} = Z_{Q(\boldsymbol{d})}$ is (ϵ, δ)-differentially private if and only if $p \geq \frac{1-\delta}{1+e^{\epsilon}}$.

See Example 3.7 in Section 3.5.3 for the details

Let us make the reasonable assumption that there exist two neighbouring databases $\boldsymbol{d}, \boldsymbol{d}'$ in D^n for which the response to Q is different; say $Q(\boldsymbol{d}) = 0$, $Q(\boldsymbol{d}') = 1$. Then for the set $A = \{0\}$, we have $\mathbb{P}(Z_{Q(\boldsymbol{d})} \in A) = \mathbb{P}(Z_0 = 0) = 1 - p$, and similarly $\mathbb{P}(Z_{Q(\boldsymbol{d}')} \in A) = p$.

Now suppose the query Q is asked twice and we wish to use our output perturbation mechanism to answer it privately. The output space is $E_Q \times E_Q$ and the natural way to define the random variables $Z_{(q_1,q_2)}$ for $q_1, q_2 \in E_Q$ is by setting $Z_{(q_1,q_2)} = (Z_1, Z_2)$ where the Z_i are independent and Z_i is identically distributed to Z_{q_i} for $i = 1, 2$. For the scenario described in the last paragraph, if we choose $\epsilon = \ln 2$ and $\delta = 0.2$, then if we choose $p = \frac{1}{3}$ the mechanism is (ϵ, δ)-differentially private with respect to Q. However if we repeat the query twice and consider the set $A = \{0\}$ as before,

$$\mathbb{P}(Z^{(2)}_{Q(\boldsymbol{d})} \in A \times A) = (1-p)^2 = \frac{4}{9},$$

while

$$\mathbb{P}(Z^{(2)}_{Q(\boldsymbol{d}')} \in A \times A) = p^2 = \frac{1}{9}.$$

It is now straightforward to verify by direct calculation that

$$\mathbb{P}(Z^{(2)}_{Q(\boldsymbol{d})} \in A \times A) > e^{\epsilon}\mathbb{P}(Z^{(2)}_{Q(\boldsymbol{d}')} \in A \times A) + \delta$$

for $\epsilon = \ln 2$ and $\delta = 0.2$, showing that applying the output perturbation mechanism twice in this instance will lead to differential privacy being violated.

The following corollary follows directly from Theorem 3.4.

Corollary 3.1. *Consider a sanitised response mechanism as defined in (3.4b) and suppose $I \in \mathcal{Q}$. Then this response mechanism is (ϵ, δ)-differentially private with respect to I if and only if it is (ϵ, δ)-differentially private with respect to every query $Q \in \mathcal{Q}$.*

Proof. "$\Rightarrow$": Theorem 3.4.

"$\Leftarrow$": The response mechanism is (ϵ, δ)-differentially private with respect to every query $Q \in \mathcal{Q}$ by assumption, therefore it must be (ϵ, δ)-differentially private with respect to the identity query I, since $I \in \mathcal{Q}$. □

Observe that the number and details of queries $Q \in \mathcal{Q}$ do not need to be specified in advance for Theorem 3.4 to hold, and so queries may be interactive and unlimited in number. This highlights a fundamental difference between privacy mechanisms that perturb the query response (e. g. by adding Laplacian or Gaussian noise) vs privacy mechanisms that perturb the database itself. Namely, in the former the added noise can be averaged out by an adversary repeating a query multiple times, thereby requiring a limit to be placed on the number of queries allowed, while in the latter an averaging attack of this sort is impossible; a repeated query will simply receive the same answer each time.

3.5 PRODUCT SANITISATIONS

In this section we introduce an implementation of local privacy in our framework. The concept of local privacy involves data being perturbed at the point of collection, which, in our framework, corresponds to a database in which each row is independently sanitised before the entire sanitisation is queried/published.

We call this a *product sanitisation* in our framework, as the resulting sanitisation is a product of n 1-dimensional sanitisation mechanisms. By applying an appropriate privacy-preserving perturbation to each individual input through this 1-dimensional mechanism, we construct the full n-dimensional sanitised database upon which privacy-preserving queries can be answered under certain conditions.

3.5.1 *Preliminary Results*

Before we formally introduce the product sanitisation mechanism, we first establish a number of technical results.

Lemma 3.1. *Let $A_1, \ldots, A_p$, $B_1, \ldots, B_p$ be two collections of non-empty sets. Then the finite union $\bigcup_{i=1}^{p}(A_i \times B_i)$ can be written as*

$$\bigcup_{i=1}^{p}(A_i \times B_i) = \bigcup_{I \subseteq [p]} (\widetilde{A}_I \times \widetilde{B}_I),$$

where $\widetilde{A}_I = \bigcup_{i\in I} A_i$ *and* $\widetilde{B}_I = \bigcap_{i\in I} B_i \setminus \bigcup_{i\notin I} B_i$. *Moreover,* $\widetilde{B}_I \cap \widetilde{B}_J = \emptyset$ *for all* $I \neq J$.

Proof. We need to prove equality and disjointness of the decomposition.

"$\subseteq$": Let $(a,b) \in \bigcup_{i=1}^{p}(A_i \times B_i)$. Then there exists at least one i^* such that $(a,b) \in A_{i^*} \times B_{i^*}$. Let $I_b := \{i : b \in B_i\} \subseteq [p]$ (note $i^* \in I_b$). Then $b \in \bigcap_{i\in I_b} B_i$, but $b \notin B_j$ for any $j \notin I_b$, otherwise j would be an element of I_b. Hence $b \in \bigcap_{i\in I_b} B_i \setminus \bigcup_{i\notin I_b} B_i$. Also $a \in \bigcup_{i\in I_b} A_i$ since $a \in A_{i^*}$. Hence

$$(a,b) \in \bigcup_{I\subseteq[p]} \left(\bigcup_{i\in I} A_i \times \bigcap_{i\in I} B_i \setminus \bigcup_{i\notin I} B_i \right).$$

"$\supseteq$": Let $(a,b) \in \bigcup_{I\subseteq[p]} \left(\bigcup_{i\in I} A_i \times \bigcap_{i\in I} B_i \setminus \bigcup_{i\notin I} B_i\right)$. Then there exists at least one $I^* \subseteq [p]$ such that $(a,b) \in \bigcup_{i\in I^*} A_i \times \bigcap_{i\in I^*} B_i \setminus \bigcup_{i\notin I^*} B_i$. Hence $a \in A_i$ for at least one $i \in I^*$ and $b \in B_i$ for all $i \in I^*$ and so there exists at least one $i \in I^*$ such that $(a,b) \in A_i \times B_i$ and so

$$(a,b) \in \bigcup_{i=1}^{p}(A_i \times B_i).$$

"$\cap = \emptyset$": Finally, we show that $\widetilde{B}_I \cap \widetilde{B}_J = \emptyset$ if $I \neq J$. To see this, note that if $I \neq J$, then we can without loss of generality assume that there is some index $k \in I$ that is not in J. Then any $x \in \widetilde{B}_I$ must be in B_k. However, as $k \in J^C$, it follows that $x \in \cup_{i\notin J} B_k$ and hence that $x \notin B_J$. This shows that the intersection is empty as claimed. □

For future use, we note that an analogous argument to that given above can be used to show the following:

Lemma 3.2. *Let* $A_1, \ldots, A_p$, $B_1, \ldots, B_p$ *be two collections of non-empty sets. Then the finite union* $\bigcup_{i=1}^{p}(A_i \times B_i)$ *can be written as*

$$\bigcup_{j=1}^{q}(\widetilde{A}_j \times \widetilde{B}_j),$$

where $\widetilde{A}_i \cap \widetilde{A}_j = \emptyset$ *for* $i \neq j$.

3.5.2 *Main Results*

We now formulate the product sanitisation response mechanism, which is a special form of the database sanitisation $Y_{\boldsymbol{d}}(\omega)$.

Definition 3.5 (Product Sanitisation)**.** *Given a* 1*-dimensional sanitisation* $\{Y_d : \Omega \to U \mid d \in D\}$ *(the parent mechanism), the* product sanitisation response mechanism *is defined to be the set of measurable mappings*

$$\{Y_{\boldsymbol{d}} : \Omega \to U^n \mid \boldsymbol{d} \in D^n\}$$

given by

$$Y_{\boldsymbol{d}}(\omega) = \left(Y^1_{\boldsymbol{d}}(\omega), \ldots, Y^n_{\boldsymbol{d}}(\omega)\right), \tag{3.8}$$

for $\omega \in \Omega$*, and where the* $Y^i_{\boldsymbol{d}}$ *are independent and* $Y^i_{\boldsymbol{d}}$ *has the same distribution as* Y_{d_i}*, for all* $\boldsymbol{d} \in D^n, i \in [n]$*.*

We say the product sanitisation mechanism $\{Y_{\boldsymbol{d}}\}$ *is generated by the parent mechanism* $\{Y_d\}$*.*

Section 2.2.1

Remark: The overall sanitisation $Y_{\boldsymbol{d}}$ is a realisation of local privacy in our framework: each component is itself a 'random response' of the true answer, taking values in U and where individual components are independent of each other. Finally, the distribution of the random response for the ith component is determined by the value of d_i, the ith component of $\boldsymbol{d}$. For instance, if the database is real-valued then one way of implementing such a sanitisation would be to add independent and identically distributed noise to each entry of the database.

We first note the following lemma concerning such mechanisms.

Lemma 3.3. *Let* $\{Y_{\boldsymbol{d}}\}$ *be a product sanitisation mechanism generated by* $\{Y_d\}$*. If* $\{Y_{\boldsymbol{d}}\}$ *is* (ϵ, δ)*-differentially private, then so too is* $\{Y_d\}$*, i. e.*

$$\mathbb{P}(Y_d \in A) \leq e^{\epsilon}\mathbb{P}(Y_{d'} \in A) + \delta,$$

for all $d, d' \in D, A \in \mathcal{A}_U$*.*

Proof. Let $d, d' \in D$ be given. If $d = d'$, the result is trivial. If $d \neq d'$, take $\boldsymbol{d} = (d, d_2, \ldots, d_n)$, $\boldsymbol{d}' = (d', d_2, \ldots, d_n)$ for any choice of $d_2, \ldots, d_n \in D$. As $Y_{\boldsymbol{d}}$

is (ϵ, δ)-differentially private and the projection $\pi_1 : U^n \to U$ onto the first coordinate is measurable, it follows that for $A \in \mathcal{A}_U$:

$$\begin{aligned}\mathbb{P}(Y_d \in A) &= \mathbb{P}(\pi_1(Y_{\boldsymbol{d}}) \in A) \\ &= \mathbb{P}(Y_{\boldsymbol{d}} \in \pi_1^{-1}(A)) \\ &\leq e^{\epsilon}\mathbb{P}(Y_{\boldsymbol{d}'} \in \pi_1^{-1}(A)) + \delta \\ &= e^{\epsilon}\mathbb{P}(Y_{d'} \in A) + \delta. \end{aligned}$$ □

We now introduce the main result of this section, which states that (ϵ, δ)-differential privacy is satisfied on a product sanitisation mechanism only when it is satisfied on a parent mechanism.

Theorem 3.5. *Consider a family $\{Y_d : \Omega \to U \mid d \in D\}$ of measurable mappings and assume that*

$$\mathbb{P}(Y_d \in A) \leq e^{\epsilon}\mathbb{P}(Y_{d'} \in A) + \delta,$$

for all $d, d' \in D$, $A \in \mathcal{A}_U$. Let $\{Y_{\boldsymbol{d}}\}$ be a product sanitisation mechanism generated by $\{Y_d\}$. Then,

$$\mathbb{P}(Y_{\boldsymbol{d}} \in A) \leq e^{\epsilon}\mathbb{P}(Y_{\boldsymbol{d}'} \in A) + \delta,$$

for all $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$ and all $A \in \mathcal{A}_{U^n}$.

Proof. We shall use induction on n. By assumption, the result is true for $n = 1$. Let $n > 1$ be given and assume that the result is true for all $k \leq n - 1$.

Assume that $\boldsymbol{d}$ and $\boldsymbol{d}'$ differ in the first element so $d_1 \neq d_1'$ but $d_j = d_j'$ for $j \neq 1$. Let

$$R = \bigcup_{i=1}^{p}(A_i \times B_i), \tag{3.9a}$$

where $A_i \in \mathcal{A}_U$, $B_i \in \mathcal{A}_{U^{n-1}}$, be given. It follows from Lemma 3.1 that we can write

$$R = \bigcup_{i=1}^{q}(\widetilde{A}_i \times \widetilde{B}_i), \tag{3.9b}$$

where $\widetilde{A}_i \in \mathcal{A}_U$, $\widetilde{B}_i \in \mathcal{A}_{U_{n-1}}$ for $i \in [q]$ and $\widetilde{B}_i \cap \widetilde{B}_j = \emptyset$ for $i \neq j$. Then, using the fact that the sets $\widetilde{B}_i$ are disjoint and the independence of the components of $Y_{\boldsymbol{d}}$, $Y_{\boldsymbol{d}'}$,

$$\begin{aligned}
\mathbb{P}(Y_{\boldsymbol{d}} \in R) &= \sum_{i=1}^{q} \mathbb{P}(Y_{\boldsymbol{d}} \in \widetilde{A}_i \times \widetilde{B}_i) \\
&= \sum_{i=1}^{q} \mathbb{P}(Y_{d_1} \in \widetilde{A}_i)\, \mathbb{P}(Y_{(d_2,\ldots,d_n)} \in \widetilde{B}_i) \\
&\leq \sum_{i=1}^{q} \Big(e^{\epsilon} \mathbb{P}(Y_{d'_1} \in \widetilde{A}_i) + \delta \Big) \mathbb{P}(Y_{(d'_2,\ldots,d'_n)} \in \widetilde{B}_i) \\
&= e^{\epsilon} \sum_{i=1}^{q} \mathbb{P}(Y_{\boldsymbol{d}'} \in \widetilde{A}_i \times \widetilde{B}_i) + \delta\, \mathbb{P}\left(Y_{(d'_2,\ldots,d'_n)} \in \bigcup_{i=1}^{q} \widetilde{B}_i \right) \\
&\leq e^{\epsilon} \mathbb{P}(Y_{\boldsymbol{d}'} \in R) + \delta.
\end{aligned}$$

If $d_1 = d'_1$, then $(d_2, \ldots, d_n) \sim (d'_2, \ldots, d'_n)$ and we can use Lemma 3.2 and the induction hypothesis to conclude that

$$\begin{aligned}
\mathbb{P}(Y_{\boldsymbol{d}} \in R) &\leq \sum_{i=1}^{q} \mathbb{P}(Y_{d'_1} \in \widetilde{A}_i) \Big(e^{\epsilon} \mathbb{P}(Y_{(d'_2,\ldots,d'_n)} \in \widetilde{B}_i) + \delta \Big) \\
&= e^{\epsilon} \sum_{i=1}^{q} \mathbb{P}(Y_{\boldsymbol{d}'} \in \widetilde{A}_i \times \widetilde{B}_i) + \delta\, \mathbb{P}\left(Y_{d'_1} \in \bigcup_{i=1}^{q} \widetilde{A}_i \right) \\
&\leq e^{\epsilon} \mathbb{P}(Y_{\boldsymbol{d}'} \in R) + \delta.
\end{aligned}$$

Thus for any set R of the form (3.9a), we can conclude that

$$\mathbb{P}(Y_{\boldsymbol{d}} \in R) \leq e^{\epsilon} \mathbb{P}(Y_{\boldsymbol{d}'} \in R) + \delta. \tag{3.9c}$$

In particular, (3.9c) holds for all elementary sets $R \in \mathcal{A}_{U^n}$. A similar argument to that used in the proof of Theorem 3.3 shows that the collection of all sets satisfying (3.9c) is a monotone class; furthermore this collection of subsets contains the elementary sets. The result now follows from Theorem 3.2. □

3.5.3 *Examples*

In this subsection, we describe some simple applications of Theorem 3.5. It is worth noting that the theorem applies to any database space $D \subseteq U$, and to *discrete* spaces in particular. For product sanitisation mechanisms, it can simplify the task of testing the mechanism for differential privacy. For instance, if D is a finite set with $|D|$ elements, then it is only necessary to check (3.6) for all $\binom{|D|}{2}$ pairs of elements of D and all $2^{|D|}$ subsets of D to ensure differential privacy on D^n. In general, we would have $n\binom{|D|}{2}|D|^{n-1}$ pairs of neighbouring elements and $2^{|D|^n}$ subsets to worry about!

Example 3.6. The addition of appropriately scaled Laplacian noise is now a standard approach to the design of differentially private responses. We note here how Theorem 3.5 combined with a simple adaptation of the arguments first given in [Dwo06] (for the case where $\delta = 0$) can be used to construct (ϵ, δ)-differentially private mechanisms for n-dimensional sanitisations with $\delta \neq 0$.

Recall that a Laplacian random variable $X : \Omega \to \mathbb{R}$ with mean zero and variance $2b^2$ has a probability density function given by

$$f_X(x) = \frac{1}{2b} e^{-\frac{|x|}{b}}.$$

Let $D \subset \mathbb{R}$ be bounded; for each $d \in D$, let $Y_d(\omega) = d + L(\omega)$ where $L : \Omega \to \mathbb{R}$ is a Laplacian random variable with mean zero and variance $2b^2$ such that

$$b \geq \frac{\operatorname{diam}(D)}{\epsilon - \log(1-\delta)}.$$

The resulting sanitised response mechanism corresponding to (3.8) is (ϵ, δ)-differentially private for any database in D^n.

To see this, note that by Theorem 3.5, it is sufficient to show that

$$\int_A \frac{e^{-\frac{|x-d|}{b}}}{2b}\, \mathrm{dx} \leq e^{\epsilon} \int_A \frac{e^{-\frac{|x-d'|}{b}}}{2b}\, \mathrm{dx} + \delta,$$

for all $d, d' \in D, A \in \mathcal{B}(\mathbb{R})$, where $\mathcal{B}(\mathbb{R})$ denotes the Borel σ-algebra on the real line $\mathbb{R}$. Using the triangle inequality, $|x - d'| \leq |x - d| + |\Delta|$ where $\Delta = d' - d$, and so it is sufficient to show that

$$\int_A \frac{e^{-\frac{|x-d|}{b}}}{2b}\, dx \leq e^{\epsilon - \frac{|d'-d|}{b}} \int_A \frac{e^{-\frac{|x-d|}{b}}}{2b}\, dx + \delta,$$

for all $d', d \in D, A \in \mathcal{B}(\mathbb{R})$. This last inequality will follow if $1 \leq e^{\epsilon - \frac{|\Delta|}{b}} + \delta$ or $b \geq \frac{|\Delta|}{\epsilon - \log(1-\delta)}$ for all $\Delta \in \{d' - d : d', d \in D\}$.

Of course, keeping in mind that the ℓ^1 sensitivity of the identity query [Dwo08] is precisely given by $\mathrm{diam}(D)$, this example can be seen as an application of the well-known Laplace mechanism to the identity query.

Example 3.7. Consider again our earlier example where $D = \mathcal{P}(\mathcal{H})$ represents the sets of possible hobbies or interests of people. As noted earlier, it is reasonable to assume that D contains finitely many elements; we denote $|D| = m$. Following Theorem 3.5 we will construct a mechanism for 1-dimensional databases: this can then be used to define a mechanism for databases in D^n via (3.8).

For $d \in D$, consider the D-valued random variable Y_d with probability mass function given by

$$\mathbb{P}(Y_d = d) = 1 - p(m-1),$$
$$\mathbb{P}(Y_d = d') = p,$$

where $d \neq d' \in D$. We make the reasonable assumption that $1 - p(m-1) \geq p$ (the true answer is at least as likely to be returned than any single incorrect one).

For (ϵ, δ)-differential privacy, we need the following to hold:

$$\mathbb{P}(Y_d \in A) \leq e^{\epsilon} \mathbb{P}(Y_{d'} \in A) + \delta, \tag{3.10a}$$

where $A \subseteq D$ and $d, d' \in D$.

We claim that (3.10a) will hold if and only if

$$1 - p(m-1) \leq e^{\epsilon} p + \delta. \tag{3.10b}$$

This condition is clearly necessary as can be seen by considering the singleton set $A = \{d\}$. To see that it is also sufficient let $d, d' \in D$ and $A \subseteq D$ be given. There are 4 cases to consider.

1. $d, d' \notin A$: Then $\mathbb{P}(Y_d \in A) = \mathbb{P}(Y_{d'} \in A) = p|A|$ and (ϵ, δ)-differential privacy holds trivially.

2. $d, d' \in A$: Then $\mathbb{P}(Y_d \in A) = \mathbb{P}(Y_{d'} \in A) = p(|A| - 1) + 1 - p(m - 1) = p(|A| - m) + 1$ and (ϵ, δ)-differential privacy holds trivially.

3. $d' \in A, d \notin A$: Then, since $p \leq 1 - p(m-1)$ by hypothesis, $\mathbb{P}(Y_d \in A) \leq \mathbb{P}(Y_{d'} \in A)$ and (ϵ, δ)-differential privacy holds trivially.

4. $d \in A, d' \notin A$: Then

$$\begin{aligned} \mathbb{P}(Y_d \in A) &= p(|A| - m) + 1, \\ \mathbb{P}(Y_{d'} \in A) &= p|A|. \end{aligned}$$

   From (3.10b), we have

$$\begin{aligned} 1 - p(m-1) &\leq e^\epsilon p + \delta \\ &= e^\epsilon (p|A| - p|A| + p) + \delta \\ &\leq e^\epsilon p|A| - p(|A| - 1) + \delta, \end{aligned}$$

   since $|A| \geq 1$ ($d \in A$ by hypothesis). Rearranging the above inequality, we see that

$$p(|A| - m) + 1 \leq e^\epsilon \big(p|A|\big) + \delta.$$

Thus we can construct an (ϵ, δ)-differentially private mechanism of the form (3.8) by choosing $p \geq \frac{1-\delta}{e^\epsilon + m - 1}$.

3.6 UTILITY

In this section, we consider the question of utility for product sanitisations. The literature on the interaction between privacy and utility is considerable and previous work has considered examples such as count queries [Dwo11], contingency table marginals [BCD+07] and spatial data [CPS+12]. As prod-

uct sanitisations are constructed from 1-dimensional mechanisms, we focus on the error of these basic mechanisms here, and a specific error function which we call the *max-mean error*. These results can then be used to derive bounds for data in D^n; the precise form these bounds will take depends on how the metric is constructed on D^n.

We wish to emphasise two points about our work: we are considering a very general class of databases that can incorporate numerical, categorical and functional data; we consider (ϵ, δ)-differential privacy and are not assuming $\delta = 0$.

We first prove the following lemma.

Lemma 3.4. *For a metric $\rho : D \times D \to \mathbb{R}$, the function $\rho(\cdot, d)$ is continuous on D for any fixed $d \in D$.*

Proof. Let $a, b \in D$. By the triangle inequality,

$$|\rho(a, d) - \rho(b, d)| \leq \rho(a, b).$$

Hence, for all $a \in D$ and for all $\varepsilon > 0$, there exists $\delta > 0$ (e. g. let $\delta = \frac{\varepsilon}{2}$) such that for all $b \in D$ where $\rho(a, b) < \delta$, then $|\rho(a, d) - \rho(b, d)| < \varepsilon$. Therefore $\rho(\cdot, d)$ is continuous on D as claimed. □

By Lemma 3.4, $\rho(\cdot, d)$ is measurable with respect to the Borel σ-algebra on D. It follows that $\rho(Y_d, d)$ is a non-negative-valued random variable.

Definition 3.6 (Max-Mean Error)**.** *We define the* max-mean error $\mathcal{E}$ *of a* 1*-dimensional mechanism $\{Y_d\}$ as:*

$$\mathcal{E} = \max_{d \in D} \mathbb{E}\big[\rho(Y_d, d)\big]. \tag{3.11}$$

When the metric ρ is Hamming distance h*, we refer to $\mathcal{E}$ as the* max-mean Hamming error*.*

For $r > 0$ and $x \in D$, $B_r(x)$ denotes the open ball

$$B_r(x) := \{y \in D \mid \rho(y, x) < r\}.$$

We first note that for any differentially private mechanism with $\delta < 1$, $\mathcal{E} > 0$. If $\delta = 1$, then any mechanism is differentially private, even completely truthful ones (in which case $\mathcal{E} = 0$).

Lemma 3.5. *Let a sanitisation $\{Y_d\}$ be given, let $0 \leq \delta < 1$ and assume that*

$$\mathbb{P}(Y_d \in A) \leq e^{\epsilon}\mathbb{P}(Y_{d'} \in A) + \delta, \tag{3.12}$$

for all $d, d' \in D, A \in \mathcal{A}_U$. Then $\mathcal{E} > 0$.

Proof. As D is compact, we can choose u, v in D with $\rho(u, v) = \operatorname{diam}(D)$. Let $r = \frac{\operatorname{diam}(D)}{2}$. Then from (3.12), it follows that

$$\mathbb{P}(Y_u \in B_r(v)) \geq e^{-\epsilon}\big(\mathbb{P}(Y_v \in B_r(v)) - \delta\big).$$

As $\rho(x, u) \geq r > 0$ for all $x \in B_r(v)$, it follows that $\mathbb{E}\big[\rho(Y_u, u)\big] > 0$ unless

$$\mathbb{P}(Y_v \in B_r(v)) \leq \delta. \tag{3.13}$$

However, if this is the case then $\mathbb{P}(\rho(Y_v, v) \geq r) \geq 1 - \delta > 0$ and hence $\mathbb{E}\big[\rho(Y_v, v)\big] \geq r(1 - \delta) > 0$. This completes the proof. □

We now present two simple results giving lower bounds for $\mathcal{E}$. The argument for the following result is inspired by that used to establish Theorem 3.3 of [HT10].

Theorem 3.6. *Let a sanitisation $\{Y_d\}$ satisfying (3.12) be given. Then*

$$\mathcal{E} \geq \frac{\operatorname{diam}(D)}{2(1 + e^{\epsilon})}(1 - \delta). \tag{3.14}$$

Proof. Without loss of generality, we may assume that $\mathcal{E}$ is finite. Moreover, from Lemma 3.5 we know that $\mathcal{E} > 0$. As D is compact, there exist points u, v in D with $\rho(u, v) = \operatorname{diam}(D)$. Set $t = \frac{\operatorname{diam}(D)}{2\mathcal{E}}$; then $t\mathcal{E} = \frac{\operatorname{diam}(D)}{2}$ and the balls $B_{t\mathcal{E}}(u)$, $B_{t\mathcal{E}}(v)$ are disjoint. From Markov's inequality applied to the non-negative random variable $\rho(Y_u, u)$, it follows that

$$\mathbb{P}(Y_u \in B_{t\mathcal{E}}(u)) \geq 1 - \frac{1}{t} = 1 - \frac{2\mathcal{E}}{\operatorname{diam}(D)}. \tag{3.15a}$$

It is now immediate that

$$\mathbb{P}(Y_u \in B_{t\mathcal{E}}(v)) \leq \frac{2\mathcal{E}}{\operatorname{diam}(D)}. \tag{3.15b}$$

From (3.12) we know that

$$\mathbb{P}(Y_u \in B_{t\mathcal{E}}(v)) \geq e^{-\epsilon}(\mathbb{P}(Y_v \in B_{t\mathcal{E}}(v)) - \delta). \tag{3.15c}$$

Combining (3.15b), (3.15c) and noting that (3.15a) also holds with u replaced by v, we see that

$$\frac{2\mathcal{E}}{\operatorname{diam}(D)} \geq e^{-\epsilon}\left(1 - \frac{2\mathcal{E}}{\operatorname{diam}(D)} - \delta\right).$$

Rearranging this completes the proof. □

The previous result applies to any compact metric space D. Now assume that D is discrete so that there exists some $\kappa > 0$ such that

$$\rho(x, y) \geq \kappa \qquad \forall\, x \neq y \in D. \tag{3.16}$$

This combined with D being compact immediately implies that D is finite. A straightforward alteration of the argument of Theorem 3.6 yields the following result.

Theorem 3.7. *Let D be finite with $|D| = m$ and $\kappa = \min_{d,d' \in D} \rho(d, d')$. Let a sanitisation $\{Y_d\}$ satisfying (3.12) be given. Then*

$$\mathcal{E} \geq \frac{\kappa(m-1)}{(m-1+e^{\epsilon})}(1-\delta). \tag{3.17}$$

Proof. It is trivial that the m balls $B_{t\mathcal{E}}(u)$, $u \in D$ are all disjoint where $t = \frac{\kappa}{\mathcal{E}}$. Fix some $u \in D$. By the same reasoning as in the proof of Theorem 3.6,

$$\mathbb{P}(Y_u \in B_{t\mathcal{E}}(u)) \geq 1 - \frac{\mathcal{E}}{\kappa}. \tag{3.18a}$$

Moreover, there must exist some $v \neq u$ such that

$$\mathbb{P}(Y_u \in B_{t\mathcal{E}}(v)) \leq \frac{\mathcal{E}}{\kappa(m-1)}. \tag{3.18b}$$

Choose one such v and apply (3.12) to obtain

$$\mathbb{P}(Y_u \in B_{t\mathcal{E}}(v)) \geq e^{-\epsilon}\big(\mathbb{P}(Y_v \in B_{t\mathcal{E}}(v)) - \delta\big). \tag{3.18c}$$

As in the proof of Theorem 3.6, we can now conclude that

$$\frac{\mathcal{E}}{\kappa(m-1)} \geq e^{-\epsilon}\left(1 - \frac{\mathcal{E}}{\kappa} - \delta\right).$$

Rearranging this inequality completes the proof. □

Example 3.8. Consider again Example 3.7. We have shown that there exists an (ϵ, δ)-differentially private mechanism with $p = \frac{1-\delta}{m-1+e^{\epsilon}}$ where $|D| = m$. If D is equipped with the discrete metric so that $\rho(d, d') = 1$ for all $d \neq d'$, then $\kappa = 1$ and for any d, the expected value of $\rho(Y_d, d)$ for this mechanism is given by

$$\mathcal{E} = \sum_{d \neq d'} p = (m-1)p = \frac{m-1}{m-1+e^{\epsilon}}(1-\delta).$$

So the bound given by Theorem 3.7 is tight in this simple case.

3.7 CONCLUDING REMARKS

In this chapter we formulated a single, unifying, abstract framework for differential privacy in the setting of probability on metric spaces, with mechanisms viewed as measurable functions taking values in query output spaces. We demonstrated the versatility of this general framework by giving examples of query types (Example 3.3), by applying it to functional (Example 3.4), numerical (Example 3.6) and categorical data (Example 3.7), and by formulating mechanisms for database sanitisations (Section 3.2.3.1) and output perturbations (Section 3.2.3.2).

Other main results in this chapter included:

- The introduction of sufficient sets for differential privacy, and proving if differential privacy holds on an algebra of subsets then it is guaranteed to hold on the σ-algebra generated by it (Theorem 3.3);

- A formal proof showing that satisfying differential privacy with respect to the identity query I guarantees differential privacy with respect to any measurable query Q (Theorem 3.4);
- The introduction of product sanitisations as a means to implement local privacy (Definition 3.5), and proving that differential privacy holds an n-dimensional product sanitisation mechanism if and only if it holds on its 1-dimensional parent mechanism (Lemma 3.3 and Theorem 3.5);
- The establishing of lower bounds on the error of product sanitisation mechanisms with respect to the max-mean error (Section 3.6).

4

SPECIALISING TO CATEGORICAL DATA

In this chapter we study differentially private mechanisms on finite datasets. By deriving sufficient sets for neighbouring databases, we obtain necessary and sufficient conditions for differential privacy and a tight lower and upper bound on the max-mean error of a discrete mechanism. We show the equivalence of a product sanitisation mechanism with a database sanitisation mechanism based on Hamming distance, and we also present the optimal product sanitisation mechanism with respect to the max-mean Hamming error.

OVERVIEW

4.1 INTRODUCTION

In this chapter we apply the generalised differential privacy framework, established in Chapter 3, to the special case of categorical data. While adding noise from a continuous probability distribution is useful for numeric data, we require a different approach for categorical data, due to its finite/discrete nature. In this setting, privacy is achieved by swapping values as determined by the probability distribution of the mechanism.

The first results of this chapter concern an adaptation for categorical data of the exponential mechanism introduced by McSherry and Talwar. In particular, we revisit the problem of sufficient sets for differential privacy for

this mechanism, a topic previously considered in Section 3.3. However, in this chapter we examine sufficient sets for fixed pairs of neighbouring databases. In Section 4.3, we present results characterising sufficient sets for the discrete exponential mechanism. From this we are able to give necessary and sufficient conditions for differential privacy for this mechanism.

We also return to the problem of the tradeoff of privacy/utility in this chapter. For categorical data, in the absence of a given metric on the dataset, we measure the error of a sanitisation using Hamming distance; in Theorem 4.4 we derive tight lower/upper bounds on the max-mean error of a discrete exponential mechanism.

In Section 4.4 we consider product sanitisations and prove that the discrete exponential mechanism is equivalent to a product sanitisation mechanism of a certain form. We also establish conditions for differential privacy and the error for these in Theorems 4.6 and 4.7 respectively. Finally in Theorem 4.8 we provide a characterisation of the optimal product sanitisation mechanism, which minimises the max-mean error within the class of product sanitisations (and hence within the class of discrete exponential mechanisms). Concluding remarks are given in Section 4.5.

4.2 PRELIMINARIES

We make use of the differential privacy framework formulated in Section 3.2, but adapt it to the specifics of categorical data.

DATABASE MODEL We consider the data space D to be finite, with m elements ($m \geq 2$). The σ-algebra with which D is equipped is the power set $\mathcal{P}(D)$, and D^n inherits the product σ-algebra, $\mathcal{P}(D^n)$. We are therefore considering all subsets of D and D^n.

Note: In this chapter, we let $U = D$. All the mechanisms we consider are combinatorial in nature, hence there is no additional algebraic structure required.

We consider Hamming distance and neighbouring databases on D^n, as defined in Definitions 3.2 and 3.1.

QUERY MODEL As we are considering the σ-algebra of all subsets of D^n, the measurability of queries in this setting does not come into question. All queries are trivially measurable since $Q^{-1}(A) \in \mathcal{P}(D^n)$ for all $A \in \mathcal{A}_Q$.

RESPONSE MECHANISM We adopt the same general response mechanism from Definition 3.3. However, in this chapter we deal exclusively with sanitised response mechanisms, as defined in Section 3.2.3.1.

One mechanism which we will make use of in this chapter is the exponential mechanism, as described by McSherry and Talwar [MT07]. We now recall the definition of this mechanism for general output spaces; later we will specialise to the case of discrete categorical data.

Definition 4.1 (Exponential Mechanism)**.** *Given $\epsilon \geq 0$, a query Q, a query output space E_Q, a utility function (which measures the utility of all possible query answers to the database being queried) $u : D^n \times E_Q \to \mathbb{R}$, a measure μ on E_Q and a normalisation constant $C_{\boldsymbol{d}}$, the exponential mechanism is defined to be the family of mappings $\{X_{Q,\boldsymbol{d}} : \Omega \to E_Q \mid \boldsymbol{d} \in D^n\}$, with probability density function with respect to μ given by $C_{\boldsymbol{d}}^{-1} e^{\epsilon u(\boldsymbol{d},q)}$ for each $\boldsymbol{d} \in D^n$ and $q \in E_Q$.*

Comment: McSherry and Talwar refer to the measure μ in the above definition as a *base measure* on the output space. As mentioned in [MT07] and seen in [Dwo08, WZ10, CMF+11, HK12, NST12], it is often the uniform probability measure on E_Q and is adjusted by the exponential factor $e^{\epsilon u(\boldsymbol{d},q)}$ to reflect the utility of various pairs of values from D^n and E_Q. The measure can potentially be viewed as describing the underlying distribution of values in E_Q. If E_Q is discrete, as in all cases considered here, we can specify μ by $\{\mu(q) : q \in E_Q\}$ and the exponential mechanism has probability mass function

$$\mathbb{P}(X_{Q,\boldsymbol{d}} = q) = C_{\boldsymbol{d}}^{-1} e^{\epsilon u(\boldsymbol{d},q)} \mu(q), \tag{4.1}$$

where $q \in E_Q$. In fact, we will focus on the situation where Q is the identity map on a finite D^n and μ is given by the uniform measure.

McSherry and Talwar showed that the exponential mechanism (Definition 4.1) satisfies $2\epsilon\Delta u$-differential privacy ($\delta = 0$), where

$$\Delta u = \max_{\substack{\boldsymbol{d} \sim \boldsymbol{d}' \in D^n \\ q \in E_Q}} |u(\boldsymbol{d}, q) - u(\boldsymbol{d}', q)|.$$

Additionally, we make a special note of the conclusion of Theorem 3.4: a sanitised response mechanism which satisfies differential privacy with respect to the identity query I will satisfy differential privacy with respect to any query Q. We therefore need only examine the response mechanism for the identity query, known as the sanitisation in our framework,

$$\{Y_{\boldsymbol{d}} : \Omega \to D^n \mid \boldsymbol{d} \in D^n\}. \tag{4.2}$$

Example 4.1 (Categorical Data I)**.** Suppose the data we are interested in records individuals' favourite hobby. The data set D would contain a list of all possible hobbies $\mathcal{H}$, as in Example 3.1. For simplicity in this example, we restrict answers to the following five hobbies: Sports; Cars; Television; Computer games; and Reading. Hence, $m = 5$. Each database $\boldsymbol{d}$ would contain the favourite hobby of n individuals. If $n = 6$, one possible $\boldsymbol{d}$ could be represented by the following list: Sports; Computer games; Television; Sports; Reading; Television.

Queries on such databases could include counting the number of unique hobbies (4 in the case above) or how many list 'Television' as their favourite hobby (2 in the case above). The identity query would be another valid query.

4.3 SUFFICIENT SETS FOR DISCRETE EXPONENTIAL MECHANISM

In Section 3.3 we considered the question of sufficient sets for differential privacy as the collection of subsets on which (3.6) must hold for (3.6) to hold on all subsets of the σ-algebra. In this section we consider a similar question, that of sufficient sets for differential privacy on neighbouring databases $\boldsymbol{d} \sim \boldsymbol{d}'$.

If we were to check (3.6) for all combinations of subsets and databases, this would require checking $n(m-1)m^n$ pairs of neighbouring databases on $(2^{m^n} - 2)$ subsets of D^n (all subsets except D^n itself and $\emptyset$, both for which (3.6) is satisfied for any mechanism). Therefore, checking a discrete response mechanism for differential privacy requires $n(m-1)m^n(2^{m^n} - 2)$ checks in total.

However, for a given pair of neighbouring databases, it is not always necessary to check (3.6) on all subsets of D^n. So we ask the question: for a given

$\boldsymbol{d} \sim \boldsymbol{d}'$, what is the smallest collection of subsets of D^n that we need to check for (3.6) to hold on all subsets of D^n? We call this collection of subsets the *sufficient sets for differential privacy* on $\boldsymbol{d} \sim \boldsymbol{d}'$.

In this section, we examine the sufficient sets for a class of discrete response mechanisms and show that, even in the most general cases, improvements on workload can be made when checking for differential privacy. We also present conditions that are necessary and sufficient for differential privacy to hold. This compares to the sufficient conditions presented in other differential privacy literature, which can therefore give a conservative estimate on the privacy level achieved.

4.3.1 *General Response Mechanism*

We begin by considering the exponential mechanism described by McSherry and Talwar [MT07], as detailed in Definition 4.1. We wish to assign a probability to each database based on its utility to the input database. This is determined by the utility function, which can be a metric or any other function deemed suitable for a particular application. As discussed in Section 4.2, we are only concerning ourselves with the identity query.

Definition 4.2 (Discrete Exponential Mechanism)**.** *Let $u : D^n \times D^n \to \mathbb{R}$ be given. The* discrete exponential response mechanism *is defined to be a family of measurable mappings $\{Y_{\boldsymbol{d}} : \Omega \to D^n \mid \boldsymbol{d} \in D^n\}$, where each $Y_{\boldsymbol{d}}$ satisfies*

$$\mathbb{P}(Y_{\boldsymbol{d}} = \boldsymbol{d}') = C_{\boldsymbol{d}}^{-1} e^{u(\boldsymbol{d},\boldsymbol{d}')}, \tag{4.3}$$

for all $\boldsymbol{d}, \boldsymbol{d}' \in D^n$. As D^n is finite, we can define the normalisation constant $C_{\boldsymbol{d}}$ as

$$C_{\boldsymbol{d}} = \sum_{\boldsymbol{d}' \in D^n} e^{u(\boldsymbol{d},\boldsymbol{d}')}$$

for each $\boldsymbol{d} \in D^n$.

Remark: While similar to the exponential mechanism (4.1), the discrete exponential mechanism differs by dealing only with the identity query ($Q = I$), by having a uniform measure ($\mu = 1$) and by absorbing ϵ into u, to allow consideration of (ϵ, δ)-differential privacy.

Remark: Ordinarily, to minimise error, we would want to assign the input database itself the highest probability of being returned, with decreasing likelihood the further we move away from the input, as determined by the utility function. This corresponds to $u(\boldsymbol{d}, \boldsymbol{d}) \geq u(\boldsymbol{d}, \boldsymbol{d}')$ for all $\boldsymbol{d}, \boldsymbol{d}' \in D^n$.

Even with this general set-up, we can still make improvements on workload when checking for differential privacy. We begin by defining the following set for a given discrete exponential mechanism $\{Y_{\boldsymbol{d}}\}$ and for each pair of neighbouring databases $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$:

$$\begin{aligned} \mathcal{S} &= \left\{ \boldsymbol{d}^* \in D^n \mid \mathbb{P}(Y_{\boldsymbol{d}} = \boldsymbol{d}^*) > \mathbb{P}(Y_{\boldsymbol{d}'} = \boldsymbol{d}^*) \right\} \\ &= \left\{ \boldsymbol{d}^* \in D^n \mid C_{\boldsymbol{d}}^{-1} e^{u(\boldsymbol{d}, \boldsymbol{d}^*)} > C_{\boldsymbol{d}'}^{-1} e^{u(\boldsymbol{d}', \boldsymbol{d}^*)} \right\} \end{aligned} \tag{4.4}$$

This set is a collection of the 'worst-case' databases for $\boldsymbol{d}$ and $\boldsymbol{d}'$, and, as we show in Theorem 4.1, is the only set of interest when checking for differential privacy.

Theorem 4.1 (Sufficient Sets). *Let $\{Y_{\boldsymbol{d}}\}$ be a discrete exponential mechanism and fix $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$. If (3.6) holds on all $A \subseteq \mathcal{S}$ then it will hold on all $A \subseteq D^n$.*

Proof. We fix $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$, let $A \subseteq D^n$ be given and assume (3.6) holds on all subsets of $\mathcal{S}$. By assumption, (3.6) holds on $A_0 = A \cap \mathcal{S}$, hence $\mathbb{P}(Y_{\boldsymbol{d}} \in A_0) \leq e^{\epsilon} \mathbb{P}(Y_{\boldsymbol{d}'} \in A_0) + \delta$. If $A_0 = A$ we are done, so assume $A \setminus A_0 \neq \emptyset$.

For each $\boldsymbol{d}^* \in A \setminus A_0$, $\mathbb{P}(Y_{\boldsymbol{d}} = \boldsymbol{d}^*) \leq \mathbb{P}(Y_{\boldsymbol{d}'} = \boldsymbol{d}^*)$. Pick one such $\boldsymbol{d}_0^* \in A \setminus A_0$, then

$$\begin{aligned} \mathbb{P}(Y_{\boldsymbol{d}} \in A_0 \cup \{\boldsymbol{d}_0^*\}) &= \mathbb{P}(Y_{\boldsymbol{d}} \in A_0) + \mathbb{P}(Y_{\boldsymbol{d}} = \boldsymbol{d}_0^*) \\ &\leq e^{\epsilon} \mathbb{P}(Y_{\boldsymbol{d}'} \in A_0) + \delta + \mathbb{P}(Y_{\boldsymbol{d}} = \boldsymbol{d}_0^*) \\ &\leq e^{\epsilon} \mathbb{P}(Y_{\boldsymbol{d}'} \in A_0) + \delta + \mathbb{P}(Y_{\boldsymbol{d}'} = \boldsymbol{d}_0^*) \\ &\leq e^{\epsilon} \mathbb{P}(Y_{\boldsymbol{d}'} \in A_0 \cup \{\boldsymbol{d}_0^*\}) + \delta. \end{aligned}$$

Hence, (3.6) holds on $A_1 = A_0 \cup \{\boldsymbol{d}_0^*\}$. We can similarly show that (3.6) holds on $A_2 = A_1 \cup \{\boldsymbol{d}_1^*\}$ for any $\boldsymbol{d}_1^* \in A \setminus A_1$. By repeating this process (picking $\boldsymbol{d}_i^* \in A \setminus A_i$), we can show that (3.6) holds on $A_{i+1} = A_i \cup \{\boldsymbol{d}_i^*\}$ for each i.

Since A is finite (D^n is finite), this process will eventually terminate when $A_i = A$, i. e. $i = |A \setminus A_0|$. Hence, (3.6) will hold on A as required. □

We now look at a simple example to demonstrate the impact of Theorem 4.1.

Example 4.2 (L^1 Norm)**.** For this example, we consider a discrete exponential mechanism where $D = \{0, 1, 2\}, n = 2$ and $u(\boldsymbol{d}, \boldsymbol{d}') = -\|\boldsymbol{d} - \boldsymbol{d}'\|_1$. In this case, $|D^n| = 9$, and we are therefore required to check $2^9 - 2 = 510$ subsets (all subsets of D^n except D^n itself and $\emptyset$) for every pair of neighbouring databases $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$, meaning a total of $510 \times 36 = 18\,360$ checks.

Let $\boldsymbol{d} = \binom{0}{1}$ and $\boldsymbol{d}' = \binom{2}{1}$, then $\mathcal{S} = \left\{\binom{0}{0}, \binom{0}{1}, \binom{0}{2}\right\}$. Hence, for this particular pair of neighbouring databases, it is sufficient to check that (3.6) holds on just $2^3 - 1 = 7$ subsets (all subsets of $\mathcal{S}$ except $\emptyset$).

If we choose $\boldsymbol{d} = \binom{1}{1}$ and $\boldsymbol{d}' = \binom{2}{1}$, then we get $\mathcal{S} = \left\{\binom{0}{0}, \binom{0}{1}, \binom{0}{2}, \binom{1}{0}, \binom{1}{1}, \binom{1}{2}\right\}$. This gives a total of $2^6 - 1 = 63$ subsets to check.

By populating the entire set of databases, we can show that 21 pairs of neighbour databases require 7 subset checks, while the remaining 15 pairs require 63 checks. That leaves us with a total of 1092 subset checks to verify differential privacy, compared with 18 360 without the use of Theorem 4.1.

4.3.2 *Response Mechanism with Fixed C_d*

By Theorem 4.1, we know that, for a discrete exponential mechanism, checking that (3.6) holds on all subsets of $\mathcal{S}$ is equivalent to checking all subsets of D^n. However, if $C_{\boldsymbol{d}} = C$ is fixed for all $\boldsymbol{d} \in D^n$, we can partition $\mathcal{S}$ to reduce our workload further.

To begin, let us present an example of a utility function that gives $C_{\boldsymbol{d}} = C$ for all $\boldsymbol{d} \in D^n$.

Example 4.3. Let $D = D_0 \cup D_1$, where $D_0 \cap D_1 = \emptyset$ and $|D_0| = |D_1| = \hat{m}$ is finite, and define $\rho : D \times D \to \mathbb{R}$ by

$$\rho(d, d') = \begin{cases} 0, & \text{if } d = d', \\ 1, & \text{if } d, d' \in D_i \text{ for some } i \in \{0, 1\}, \\ 2, & \text{if } d \in D_i, d' \in D_{1-i} \text{ for some } i \in \{0, 1\}. \end{cases}$$

We then define the utility function $u : D^n \times D^n \to \mathbb{R}$ by

$$u(\boldsymbol{d}, \boldsymbol{d}') = -\sum_{i=1}^{n} \rho(d_i, d_i').$$

We claim that such a set-up gives $C_{\boldsymbol{d}} = C$.

We will prove this by induction on n. First, let $n = 1$. Given $d \in D$, we have

$$\begin{aligned} C_d &= \sum_{d' \in D} e^{-\rho(d,d')} \\ &= \sum_{\substack{d' \in D \\ \rho(d,d')=0}} 1 + \sum_{\substack{d' \in D \\ \rho(d,d')=1}} e^{-1} + \sum_{\substack{d' \in D \\ \rho(d,d')=2}} e^{-2} \\ &= 1 + (\hat{m} - 1)e^{-1} + \hat{m}e^{-2} \\ &= C^{(1)}. \end{aligned}$$

Hence, $C_d = C^{(1)}$ for each $d \in D$, and the result holds for $n = 1$.

Assume $C_{\boldsymbol{d}} = C^{(k)}$ for each $\boldsymbol{d} \in D^k$. Let $n = k + 1$, and let $\boldsymbol{d}^* \in D^{k+1}$ be given. Define $\boldsymbol{d} \in D^k$ by $d_i = d_i^*$ for each $i \in [k]$, and let $d = d_{k+1}^*$, hence $\boldsymbol{d}^* = \binom{\boldsymbol{d}}{d}$. Then,

$$\begin{aligned} C_{\boldsymbol{d}^*} = C_{\binom{\boldsymbol{d}}{d}} &= \sum_{\boldsymbol{d}' \in D^{k+1}} e^{u(\binom{\boldsymbol{d}}{d}, \boldsymbol{d}')} \\ &= \sum_{\boldsymbol{d}' \in D^k} \sum_{d' \in D} e^{u(\boldsymbol{d}, \boldsymbol{d}') - \rho(d, d')} \\ &= \sum_{d' \in D} e^{-\rho(d,d')} \sum_{\boldsymbol{d}' \in D^k} e^{u(\boldsymbol{d}, \boldsymbol{d}')} \\ &= \sum_{\substack{d' \in D \\ \rho(d,d')=0}} C^{(k)} + \sum_{\substack{d' \in D \\ \rho(d,d')=1}} e^{-1} C^{(k)} + \sum_{\substack{d' \in D \\ \rho(d,d')=2}} e^{-2} C^{(k)} \\ &= C^{(k)} + (\hat{m} - 1)e^{-1} C^{(k)} + \hat{m}e^{-2} C^{(k)} \\ &= C^{(k+1)}, \end{aligned}$$

and the result also holds for $n = k + 1$. By the induction hypothesis, we conclude that $C_{\boldsymbol{d}} = C^{(n)}$ for each $\boldsymbol{d} \in D^n$.

Let us now consider the following set relating to $\mathcal{S}$, where $\boldsymbol{d} \sim \boldsymbol{d}'$ is fixed:

$$\alpha = \{u(\boldsymbol{d}, \boldsymbol{d}^*) - u(\boldsymbol{d}', \boldsymbol{d}^*) \mid \boldsymbol{d}^* \in \mathcal{S}\}. \tag{4.5}$$

There are only finitely many values in α, since $\mathcal{S}$ is finite, so letting $|\alpha| = s$ gives $\alpha \in \mathbb{R}^s$. Note that $\alpha > 0$ by the definition of $\mathcal{S}$ and since $C_{\boldsymbol{d}} = C$ by hypothesis. We label the elements $\alpha_1, \alpha_2, \ldots, \alpha_s \in \alpha$, with $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_s$.

We then partition $\mathcal{S}$ into the collection of subsets $\{\tilde{\mathcal{S}}_1, \tilde{\mathcal{S}}_2, \ldots, \tilde{\mathcal{S}}_s\}$ as follows, for each $i \in [s]$:

$$\tilde{\mathcal{S}}_i = \{\boldsymbol{d}^* \in \mathcal{S} \mid u(\boldsymbol{d}, \boldsymbol{d}^*) - u(\boldsymbol{d}', \boldsymbol{d}^*) = \alpha_i\}. \tag{4.6}$$

Note that for each $i \in [s]$ and for all $\boldsymbol{d}^* \in \tilde{\mathcal{S}}_i$,

$$\begin{aligned} \mathbb{P}(Y_{\boldsymbol{d}'} = \boldsymbol{d}^*) &= C^{-1} e^{u(\boldsymbol{d}', \boldsymbol{d}^*)} \\ &= C^{-1} e^{u(\boldsymbol{d}, \boldsymbol{d}^*) - \alpha_i} \\ &= e^{-\alpha_i} \mathbb{P}(Y_{\boldsymbol{d}} = \boldsymbol{d}^*). \end{aligned} \tag{4.7}$$

We can now show that if (3.6) holds on these partitions, it will hold on all subsets of each partition.

Theorem 4.2. *Fix $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$ and let $\{Y_{\boldsymbol{d}}\}$ be a discrete exponential mechanism with $C_{\boldsymbol{d}} = C$ for all $\boldsymbol{d} \in D^n$. If (3.6) holds on $\tilde{\mathcal{S}}_i$ then it will hold on all $A \subseteq \tilde{\mathcal{S}}_i$ for each $i \in [s]$.*

Proof. Fix $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$ and $i \in [s]$. We assume

$$\mathbb{P}(Y_{\boldsymbol{d}} \in \tilde{\mathcal{S}}_i) \leq e^{\epsilon} \mathbb{P}(Y_{\boldsymbol{d}'} \in \tilde{\mathcal{S}}_i) + \delta,$$

and let $A \subseteq \tilde{\mathcal{S}}_i$. By (4.7) and since A is finite, $\mathbb{P}(Y_{\boldsymbol{d}'} \in A) = e^{-\alpha_i} \mathbb{P}(Y_{\boldsymbol{d}} \in A)$. Since this also holds for the set $\tilde{\mathcal{S}}_i$ itself, we have

$$1 \leq e^{\epsilon - \alpha_i} + \frac{\delta}{\mathbb{P}(Y_{\boldsymbol{d}} \in \tilde{\mathcal{S}}_i)}.$$

Clearly $\mathbb{P}(Y_{\boldsymbol{d}} \in A) \leq \mathbb{P}(Y_{\boldsymbol{d}} \in \tilde{\mathcal{S}}_i)$, which gives

$$1 \leq e^{\epsilon - \alpha_i} + \frac{\delta}{\mathbb{P}(Y_{\boldsymbol{d}} \in A)},$$

or, rewriting,

$$\mathbb{P}(Y_{\boldsymbol{d}} \in A) \leq e^{\epsilon} \mathbb{P}(Y_{\boldsymbol{d}'} \in A) + \delta.$$

Hence, (3.6) holds on all $A \subseteq \tilde{\mathcal{S}}_i$. □

Although Theorem 4.2 tells us that it is sufficient to check each $\tilde{\mathcal{S}}_i$ for (ϵ, δ)-differential privacy to hold on each subset of $\tilde{\mathcal{S}}_i$, we cannot draw any conclusion about subsets of D^n in general (e. g. the set $\tilde{\mathcal{S}}_1 \cup \tilde{\mathcal{S}}_2$). The following corollary shows that the collection $\{\tilde{\mathcal{S}}_1, \ldots, \tilde{\mathcal{S}}_s\}$ are sufficient sets of the discrete exponential mechanism for ϵ-differential privacy to hold (i. e. when $\delta = 0$).

Corollary 4.1 (Sufficient Sets with Fixed $C_{\boldsymbol{d}}$)**.** *Let $\{Y_{\boldsymbol{d}}\}$ be a discrete exponential mechanism, $\delta = 0$ and fix $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$. If (3.6) holds on $\tilde{\mathcal{S}}_i$ for all $i \in [s]$, then it will hold on all subsets of D^n.*

Proof. Let $A \subseteq \mathcal{S}$ and assume that (3.6) holds on $\tilde{\mathcal{S}}_i$ for all $i \in [s]$. Hence, by Theorem 4.2, (3.6) holds on all subsets of $\tilde{\mathcal{S}}_i$ for all $i \in [s]$. As $\{\tilde{\mathcal{S}}_i\}$ partitions $\mathcal{S}$, $A = \bigcup_i A \cap \tilde{\mathcal{S}}_i$, hence $\mathbb{P}(Y_{\boldsymbol{d}} \in A \cap \tilde{\mathcal{S}}_i) \leq e^{\epsilon}\mathbb{P}(Y_{\boldsymbol{d}'} \in A \cap \tilde{\mathcal{S}}_i)$ for all $i \in [s]$.

Since $\tilde{\mathcal{S}}_i \cap \tilde{\mathcal{S}}_j = \emptyset$ when $i \neq j$, $\mathbb{P}(Y_{\boldsymbol{d}} \in \bigcup_i A \cap \tilde{\mathcal{S}}_i) = \sum_i \mathbb{P}(Y_{\boldsymbol{d}} \in A \cap \tilde{\mathcal{S}}_i)$, and so,

$$\begin{aligned}
\mathbb{P}(Y_{\boldsymbol{d}} \in A) &= \mathbb{P}\left(Y_{\boldsymbol{d}} \in \bigcup_{i=1}^{s} A \cap \tilde{\mathcal{S}}_i\right) \\
&= \sum_{i=1}^{s} \mathbb{P}\left(Y_{\boldsymbol{d}} \in A \cap \tilde{\mathcal{S}}_i\right) \\
&\leq e^{\epsilon} \sum_{i=1}^{s} \mathbb{P}\left(Y_{\boldsymbol{d}'} \in A \cap \tilde{\mathcal{S}}_i\right) \\
&= e^{\epsilon}\mathbb{P}(Y_{\boldsymbol{d}'} \in A).
\end{aligned}$$

Therefore (3.6) holds on all subsets of $\mathcal{S}$, and by Theorem 4.1, it holds on all subsets of D^n. □

Example 4.4. Returning to the set-up given in Example 4.3, consider $D_0 = \{A, B\}$ and $D_1 = \{Y, Z\}$, hence $\hat{m} = 2$, and let $n = 2$. Let $\boldsymbol{d} = \binom{A}{Z}$ and $\boldsymbol{d}' = \binom{A}{B}$, and note that $\boldsymbol{d} \sim \boldsymbol{d}'$.

We can now calculate $\mathcal{S}$ for the given $\boldsymbol{d}$ and $\boldsymbol{d}'$:

$$\mathcal{S} = \left\{ \binom{A}{Y}, \binom{A}{Z}, \binom{B}{Y}, \binom{B}{Z}, \binom{Y}{Y}, \binom{Y}{Z}, \binom{Z}{Y}, \binom{Z}{Z} \right\}.$$

By direct calculation, we see that $\alpha = \{1, 2\}$, with which we proceed to calculate

$$\tilde{\mathcal{S}}_1 = \left\{ \binom{A}{Y}, \binom{B}{Y}, \binom{Y}{Y}, \binom{Z}{Y} \right\},$$
$$\tilde{\mathcal{S}}_2 = \left\{ \binom{A}{Z}, \binom{B}{Z}, \binom{Y}{Z}, \binom{Z}{Z} \right\}.$$

Verifying (ϵ, δ)-differential privacy on these two sets is sufficient to prove (ϵ, δ)-differential privacy holds on all of their subsets. Furthermore, $\tilde{\mathcal{S}}_1$ and $\tilde{\mathcal{S}}_2$ are the only sets we must check for ϵ-differential privacy to hold on D^n.

4.3.3 *Discrete Exponential Mechanism with Hamming Distance*

The usefulness of Theorem 4.2 becomes particularly apparent when we restrict our response mechanism to one derived from Hamming distance. For the remainder of this section our utility function $u : D^n \times D^n \to \mathbb{R}$ is defined to be

$$u(\boldsymbol{d}, \boldsymbol{d}') = -k\,\mathrm{h}(\boldsymbol{d}, \boldsymbol{d}'), \tag{4.8}$$

where $k \in \mathbb{R}$ is a privacy parameter. Note that for this set-up, the normalisation constant $C_{\boldsymbol{d}} = C$ is fixed for all $\boldsymbol{d} \in D^n$, as shown in Proposition 4.1 below.

Remark: By enforcing $k \geq 0$, truthful responses are at least as likely as any one incorrect response.

Proposition 4.1. *For a discrete exponential mechanism $\{Y_{\boldsymbol{d}}\}$ whose utility function u satisfies (4.8),*

$$C_{\boldsymbol{d}} = C = \left(1 + \frac{m-1}{e^k}\right)^n. \tag{4.9}$$

Proof. Let $\boldsymbol{d} \in D^n$ be given. Then,

$$\begin{aligned}\sum_{\boldsymbol{d}' \in D^n} \mathbb{P}(Y_{\boldsymbol{d}} = \boldsymbol{d}') &= \sum_{\boldsymbol{d}' \in D^n} C_{\boldsymbol{d}}^{-1}\, e^{-k\,\mathrm{h}(\boldsymbol{d},\boldsymbol{d}')} \\ &= C_{\boldsymbol{d}}^{-1}\, e^0 + C_{\boldsymbol{d}}^{-1}\, n(m-1)e^{-k} + C_{\boldsymbol{d}}^{-1} \binom{n}{2} (m-1)^2 e^{-2k} + \cdots \\ &\quad + C_{\boldsymbol{d}}^{-1}\, (m-1)^n e^{-nk} \\ &= C_{\boldsymbol{d}}^{-1} \sum_{i=0}^{n} \binom{n}{i} (m-1)^i e^{-ik} \\ &= C_{\boldsymbol{d}}^{-1} \left(1 + \frac{m-1}{e^k}\right)^n.\end{aligned}$$

For the probability mass function of $Y_{\boldsymbol{d}}$ to sum to 1, we need

$$C_{\boldsymbol{d}}^{-1} \left(1 + \frac{m-1}{e^k}\right)^n = 1.$$

Rearranging this completes the proof. □

In this case, the sufficient sets for the problem condense down to a single set.

Corollary 4.2 (Sufficient Sets for Hamming Distance)**.** *Consider a discrete exponential mechanism satisfying (4.8) and fix $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$. If (3.6) holds on $\mathcal{S}$, then it will hold on all $A \subseteq D^n$.*

Proof. First, consider the case where $k = 0$. In this case, $\mathbb{P}(Y_{\boldsymbol{d}} = \boldsymbol{d}^*) = C^{-1}$ for all $\boldsymbol{d}, \boldsymbol{d}^* \in D^n$, and so (3.6) holds.

Next, note that, for all $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$ and $\boldsymbol{d}^* \in D^n$,

$$\mathrm{h}(\boldsymbol{d}', \boldsymbol{d}^*) - \mathrm{h}(\boldsymbol{d}, \boldsymbol{d}^*) \in \{-1, 0, 1\},$$

and for each $\boldsymbol{d}^* \in \mathcal{S}$, $-k\,\mathrm{h}(\boldsymbol{d}, \boldsymbol{d}^*) > -k\,\mathrm{h}(\boldsymbol{d}', \boldsymbol{d}^*)$ by definition of $\mathcal{S}$. Hence for all $\boldsymbol{d}^* \in \mathcal{S}$, $\mathrm{h}(\boldsymbol{d}', \boldsymbol{d}^*) - \mathrm{h}(\boldsymbol{d}, \boldsymbol{d}^*) = \mathrm{sgn}(k)$. The set α therefore reduces to a singleton set,

$$\begin{aligned}\alpha &= \{-k\,\mathrm{h}(\boldsymbol{d}, \boldsymbol{d}^*) + k\,\mathrm{h}(\boldsymbol{d}', \boldsymbol{d}^*) \mid \boldsymbol{d}^* \in \mathcal{S}\} \\ &= \{k\,\mathrm{sgn}(k)\} = \{|k|\},\end{aligned}$$

and hence $\mathcal{S} = \tilde{\mathcal{S}}_1$.

We can then conclude that if (3.6) holds on $\mathcal{S}$, it must hold on all subsets of $\mathcal{S}$ (by Theorem 4.1) and also on all subsets of D^n (by Theorem 4.2). □

We have established that, for the discrete exponential mechanism satisfying (4.8) and for a given neighbouring pair of databases $\boldsymbol{d} \sim \boldsymbol{d}'$, to check for differential privacy on all possible subsets of the database space D^n, we need only check a single set $\mathcal{S}$.

It is now a relatively simple task to establish conditions on the response mechanism for differential privacy. For the following theorem, we assume that $\delta < 1$ (note that all mechanisms are trivially (ϵ,1)-differentially private).

Theorem 4.3 (Condition for Differential Privacy)**.** *Let $\delta < 1$. A discrete exponential mechanism $\{Y_{\boldsymbol{d}}\}$ satisfying (4.8) is (ϵ, δ)-differentially private if and only if*

$$\frac{1-(m-1)\delta}{e^{\epsilon}+\delta} \leq e^{k} \leq \frac{e^{\epsilon}+(m-1)\delta}{1-\delta}. \tag{4.10}$$

Proof. We consider two cases, $k > 0$ and $k < 0$. Note that (ϵ, δ)-differential privacy is trivially satisfied when $k = 0$ since $\mathbb{P}(Y_{\boldsymbol{d}} \in A) = \mathbb{P}(Y_{\boldsymbol{d}'} \in A)$ for every $A \subseteq D^n$ and for every $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$.

Fix $\boldsymbol{d} \sim \boldsymbol{d}' \in D^n$. By Corollary 4.2 we need only satisfy (3.6) on $\mathcal{S}$ for it to hold on all $A \subseteq D^n$. Hence we need

$$\mathbb{P}(Y_{\boldsymbol{d}} \in \mathcal{S}) \leq e^{\epsilon}\mathbb{P}(Y_{\boldsymbol{d}'} \in \mathcal{S}) + \delta. \tag{4.11}$$

If $\mathcal{S} = \emptyset$, the result follows trivially, so let's assume that $\mathcal{S} \neq \emptyset$.

Note that by definition, $\mathbb{P}(Y_{\boldsymbol{d}'} \in \mathcal{S}) = e^{-|k|}\mathbb{P}(Y_{\boldsymbol{d}} \in \mathcal{S})$, since for each $\boldsymbol{d}^* \in \mathcal{S}$, we have $\mathrm{h}(\boldsymbol{d}, \boldsymbol{d}^*) - \mathrm{h}(\boldsymbol{d}', \boldsymbol{d}^*) = -\operatorname{sgn}(k)$.

1. $k > 0$: First, let's consider the case where $k > 0$. Since $\mathcal{S} \neq \emptyset$ by assumption, from Corollary 4.2 the mechanism will be (ϵ, δ)-differentially private if and only if

$$1 \leq e^{\epsilon-k} + \frac{\delta}{\mathbb{P}(Y_{\boldsymbol{d}} \in \mathcal{S})}. \tag{4.12}$$

Now consider $\mathbb{P}(Y_{\boldsymbol{d}} \in \mathcal{S})$. By definition, for each $\boldsymbol{d}^* \in \mathcal{S}$, $\mathrm{h}(\boldsymbol{d}', \boldsymbol{d}^*) = \mathrm{h}(\boldsymbol{d}, \boldsymbol{d}^*) + 1$. Therefore, $|\{\boldsymbol{d}^* \mid \mathrm{h}(\boldsymbol{d}, \boldsymbol{d}^*) = c\}| = \binom{n-1}{c}(m-1)^c$, and so,

$$\begin{aligned}\mathbb{P}(Y_{\boldsymbol{d}} \in \mathcal{S}) &= C^{-1} \sum_{i=0}^{n-1} \binom{n-1}{i} (m-1)^i e^{-ik} \\ &= C^{-1} \left(1 + \frac{m-1}{e^k}\right)^{n-1} \\ &= \left(1 + \frac{m-1}{e^k}\right)^{-1},\end{aligned}$$

where the final line follows from Proposition 4.1. Substituting this result into (4.12) gives

$$1 \leq e^{\epsilon - k} + \delta \left(1 + \frac{m-1}{e^k}\right),$$

and solving for e^k gives

$$e^k \leq \frac{e^\epsilon + (m-1)\delta}{1-\delta}.$$

2. $k < 0$: Secondly, let's consider the case where $k < 0$. By symmetry, it can be shown that $\mathbb{P}(Y_{\boldsymbol{d}'} \in \mathcal{S}) = \left(1 + \frac{m-1}{e^k}\right)^{-1}$, giving us $\mathbb{P}(Y_{\boldsymbol{d}} \in \mathcal{S}) = e^{-k}\left(1 + \frac{m-1}{e^k}\right)^{-1}$. From (4.11), we have

$$\begin{aligned}1 &\leq e^{\epsilon + k} + \frac{\delta}{\mathbb{P}(Y_{\boldsymbol{d}} \in \mathcal{S})} \\ &= e^{\epsilon + k} + \delta(e^k + m - 1).\end{aligned}$$

Rewriting gives us

$$1 - (m-1)\delta \leq e^k(e^\epsilon + \delta), \tag{4.13}$$

and solving for e^k completes the proof. □

Remark: For $(\epsilon, 0)$-differential privacy, we require $\frac{1}{e^\epsilon} \leq e^k \leq e^\epsilon$, or equivalently, $|k| \leq \epsilon$. If $\delta \geq \frac{1}{m-1}$, the condition on k simplifies to $e^k \leq \frac{e^\epsilon + (m-1)\delta}{1-\delta}$.

Discussion: If we convert our discrete exponential mechanism back into the form of the exponential mechanism, McSherry and Talwar [MT07] tell us that the mechanism satisfies 2ϵ-differential privacy at worst. However, we have shown in Theorem 4.3 that the mechanism satisfies ϵ-differential

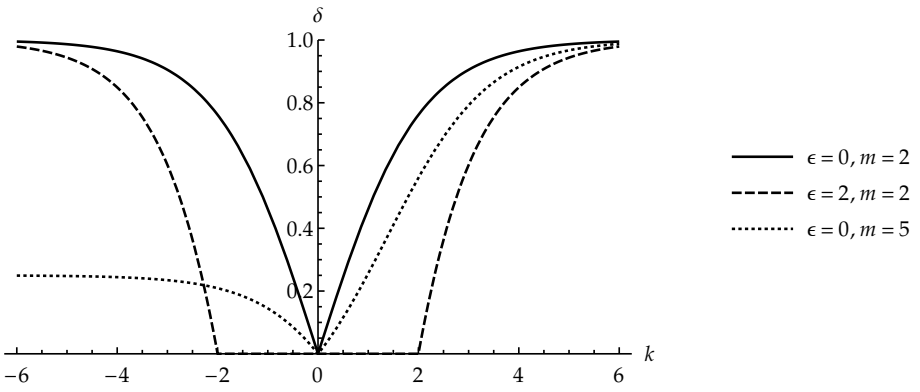

Figure 4.1: A plot of the lower bound for δ versus k for the discrete exponential mechanism satisfying (ϵ, δ)-differential privacy, for varying ϵ and m. Note that $\delta = 0$ is only possible when $|k| \leq \epsilon$. Note also that the bound for δ approaches 1 for large k, but approaches $\frac{1}{m-1}$ for large negative k.

privacy, and that this condition is tight (necessary and sufficient, hence we can do no better). This improves on the looser bound in McSherry and Talwar's proof, as it underestimates the differential privacy achieved by a factor of two. Their mechanism is also limited to ϵ-differential privacy only ($\delta = 0$), whereas we can account for (ϵ, δ)-differential privacy ($\delta > 0$).

Figure 4.1 gives an illustration of Theorem 4.3, and the constraint placed on the selection of δ.

Using the differential privacy constraints established in Theorem 4.3, we can now determine error bounds for the mechanism using the max-mean Hamming error from Chapter 3. Formally, we have the following result.

Theorem 4.4 (Error). *Consider a discrete exponential mechanism satisfying (4.8) which is (ϵ, δ)-differentially private and the max-mean Hamming error $\mathcal{E}$ from Definition 3.6. Then*

$$\frac{1-\delta}{1+\frac{e^{\epsilon}}{m-1}} \leq \frac{\mathcal{E}}{n} \leq \min\left\{1, \frac{e^{\epsilon}+\delta}{e^{\epsilon}+\frac{1}{m-1}}\right\}. \tag{4.14}$$

Proof. Let $\boldsymbol{d} \in D^n$. Then $\mathbb{P}(Y_{\boldsymbol{d}} = \boldsymbol{d}') = Ce^{-k\,\mathrm{h}(\boldsymbol{d},\boldsymbol{d}')}$ and $|\{\boldsymbol{d}' : \mathrm{h}(\boldsymbol{d},\boldsymbol{d}') = c\}| = \binom{n}{c}(m-1)^c$ for all $c \in [n]_0$. Hence,

$$\begin{aligned}
\mathbb{E}[\mathrm{h}(Y_{\boldsymbol{d}},\boldsymbol{d})] &= C^{-1}\sum_{i=0}^{n} i\binom{n}{i}(m-1)^i e^{-ik} \\
&= C^{-1}n\sum_{i=1}^{n}\binom{n-1}{i-1}\left(\frac{m-1}{e^k}\right)^i \\
&= C^{-1}n\frac{m-1}{e^k}\sum_{I=0}^{n-1}\binom{n-1}{I}\left(\frac{m-1}{e^k}\right)^I \\
&= C^{-1}n\frac{m-1}{e^k}\left(1+\frac{m-1}{e^k}\right)^{n-1} \\
&= n\frac{m-1}{e^k}\left(1+\frac{m-1}{e^k}\right)^{-1} \\
&= \frac{n}{1+\frac{e^k}{m-1}}.
\end{aligned}$$

On average, the number of entries that will change is therefore

$$\frac{\mathcal{E}}{n} = \frac{\max_{\boldsymbol{d}\in D^n}\mathbb{E}[\mathrm{h}(Y_{\boldsymbol{d}},\boldsymbol{d})]}{n} = \frac{1}{1+\frac{e^k}{m-1}}.$$

Since the response mechanism is differentially private and noting that $e^k \geq 0$, Theorem 4.3 tells us that $\max\{\frac{1-(m-1)\delta}{e^\epsilon+\delta}, 0\} \leq e^k \leq \frac{e^\epsilon+(m-1)\delta}{1-\delta}$, hence,

$$\frac{1-\delta}{1+\frac{e^\epsilon}{m-1}} \leq \frac{\mathcal{E}}{n} \leq \min\left\{1, \frac{e^\epsilon+\delta}{e^\epsilon+\frac{1}{m-1}}\right\}. \qquad \square$$

Remark: These bounds are tight and can be achieved by setting e^k to the bounds established in Theorem 4.3.

4.4 PRODUCT SANITISATION

The results of Section 4.3 give a clear framework on how to create differentially private mechanisms for discrete data, particularly categorical data using Hamming distance, and to obtain tight (necessary and sufficient) conditions for differential privacy. However, this mechanism requires the creation of a separate probability distribution for each of the m^n unique databases.

In this section, we present a simple method for realising the discrete exponential mechanism by revisiting the product sanitisation mechanism first discussed in Section 3.5.

4.4.1 *Response Mechanism*

We make use of the product sanitisation mechanism from Definition 3.5. Recall that a parent mechanism $\{Y_d\}$ generates a product sanitisation mechanism $\{Y_{\boldsymbol{d}}\}$.

Theorem 3.5 states that differential privacy on a product sanitisation mechanism is guaranteed when its parent mechanism is differentially private; the converse of this result was proven in Lemma 3.3. However, we are able to take a simpler approach to showing this converse by exploiting the finiteness of D.

Corollary 4.3 (Parent Mechanism)**.** *A product sanitisation mechanism $\{Y_{\boldsymbol{d}}\}$ is (ϵ, δ)-differentially private if and only if its parent mechanism $\{Y_d\}$ is (ϵ, δ)-differentially private.*

Proof. "$\Leftarrow$": Theorem 3.5.

"$\Rightarrow$": Let $A \subseteq D$ and $d \neq d' \in D$ be given. Define $A' = A \times D^{n-1} \subseteq D^n$ and $\boldsymbol{d}, \boldsymbol{d}' \in D^n$, such that $d_1 = d$, $d'_1 = d'$ and $d_i = d'_i$ for all $i \in [n] \setminus \{1\}$, hence $\boldsymbol{d} \sim \boldsymbol{d}'$. Since $\{Y_{\boldsymbol{d}}\}$ is differentially private by assumption,

$$\mathbb{P}(Y_{\boldsymbol{d}} \in A') \leq e^{\epsilon} \mathbb{P}(Y_{\boldsymbol{d}'} \in A') + \delta.$$

However, since $\{Y_{\boldsymbol{d}}\}$ is a product sanitisation mechanism,

$$\begin{aligned} \mathbb{P}(Y_{\boldsymbol{d}} \in A') &= \mathbb{P}(Y^1_{\boldsymbol{d}} \in A) \times \prod_{i=2}^{n} \mathbb{P}(Y^i_{\boldsymbol{d}} \in D) \\ &= \mathbb{P}(Y^1_{\boldsymbol{d}} \in A) \\ &= \mathbb{P}(Y_{d_1} \in A). \end{aligned}$$

Hence,

$$\mathbb{P}(Y_d \in A) \leq e^{\epsilon} \mathbb{P}(Y_{d'} \in A) + \delta,$$

for all $d \neq d' \in D$ and $A \subseteq D$, as required. □

For the remainder of this section, given $0 \leq p \leq \frac{1}{m-1}$, the parent mechanism $\{Y_d\}$ of the product sanitisation is defined such that

$$\begin{aligned} \mathbb{P}(Y_d = d) &= 1 - p(m-1), \\ \mathbb{P}(Y_d = d') &= p, \end{aligned} \tag{4.15}$$

for every $d' \in D \setminus \{d\}$.

Remark: If we additionally require $p \leq \frac{1}{m}$, we get $\mathbb{P}(Y_d = d) \geq \mathbb{P}(Y_d = d')$ for every $d' \in D$. Therefore, the mechanism would be at least as likely to return the correct answer as any one incorrect answer, an entirely reasonable assumption.

Remark: $p = \frac{1}{m}$ represents the case of releasing uniform noise (i. e. no information), while decreasing p reduces the error of the mechanism.

Example 4.5 (Categorical Data II)**.** Using the same set-up as in Example 4.1, one such permissable p would be $p = 0.1$, since $0 \leq 0.1 \leq \frac{1}{4}$. Then, if d were to be the value 'Television', and d' the value 'Cars', $\mathbb{P}(Y_d = d) = 0.6$, and $\mathbb{P}(Y_d = d') = 0.1$.

We then sanitise the n-row database $\boldsymbol{d}$ in the same way, working through the database one row at a time.

In the first main result of this section, we show that the product sanitisation mechanism and the discrete exponential mechanism are equivalent, despite being constructed in different ways. This is subject to the discrete exponential mechanism satisfying (4.8) and the product sanitisation mechanism satisfying (4.15).

Theorem 4.5 (Equivalence)**.** *Let $\{Y_{\boldsymbol{d}} \mid \boldsymbol{d} \in D^n\}$ be a discrete exponential mechanism satisfying (4.8), and $\{Y^*_{\boldsymbol{d}} \mid \boldsymbol{d} \in D^n\}$ be a product sanitisation response mechanism, whose parent mechanism satisfies (4.15). Then the probability mass functions of $Y_{\boldsymbol{d}}$ and $Y^*_{\boldsymbol{d}}$ are identical for every $\boldsymbol{d} \in D^n$ when*

$$e^k = 1 + \frac{1}{p} - m. \tag{4.16}$$

Proof. Let $\boldsymbol{d} \in D^n$, then

$$\mathbb{P}(Y_{\boldsymbol{d}} = \boldsymbol{d}') = \left(1 + \frac{m-1}{e^k}\right)^{-n} e^{-k\,\mathrm{h}(\boldsymbol{d},\boldsymbol{d}')},$$

$$\mathbb{P}(Y^*_{\boldsymbol{d}} = \boldsymbol{d}') = (1 - p(m-1))^n \left(\frac{p}{1-p(m-1)}\right)^{\mathrm{h}(\boldsymbol{d},\boldsymbol{d}')},$$

for all $\boldsymbol{d}' \in D^n$.

For the two mechanisms to be equivalent, we need $\frac{p}{1-p(m-1)} = e^{-k}$, or, rewriting, $\frac{1-p(m-1)}{p} = e^k$.

We also need $1 - p(m-1) = \frac{e^k}{e^k+m-1}$, which can be rewritten as $\frac{1}{1-p(m-1)} = 1 + \frac{m-1}{e^k}$, or $\frac{1-p(m-1)}{p} = e^k$.

Hence, the mechanisms $\{Y_{\boldsymbol{d}}\}$ and $\{Y^*_{\boldsymbol{d}}\}$ are identical when $e^k = 1 + \frac{1}{p} - m$ or, equivalently, $p = \frac{1}{e^k+m-1}$. □

Remark: Note that $k \to -\infty$ as $p \to \frac{1}{m-1}$ and $k \to +\infty$ as $p \to 0$.

4.4.2 *Alternative Proofs of Theorems 4.3 and 4.4*

We have already established that the discrete exponential mechanism satisfying (4.8) and the product sanitisation mechanism satisfying (4.15) are equivalent, meaning the results of Theorems 4.3 and 4.4 also apply to the product sanitisation mechanism in this particular set-up. In this sub-section, we provide alternative methods of proof for these theorems that make use of the specific product sanitisation of the mechanism. This may help to broaden understanding of the mechanism and add insight for the reader.

A portion of the proof of the following theorem appears in Section 3.5.3.

Theorem 4.6 (Condition for Differential Privacy)**.** *A product sanitisation mechanism, whose parent mechanism satisfies (4.15), is (ϵ, δ)-differentially private if and only if*

$$\frac{1-\delta}{e^\epsilon + m - 1} \leq p \leq \min\left\{\frac{1}{m-1}, \frac{e^\epsilon + \delta}{1 + (m-1)e^\epsilon}\right\}. \tag{4.17}$$

Proof. We consider two cases, where $p \leq \frac{1}{m}$ and $p > \frac{1}{m}$.

1. $p \leq \frac{1}{m}$: From Example 3.7, a necessary and sufficient condition for (ϵ, δ)-differential privacy is $p \geq \frac{1-\delta}{e^\epsilon+m-1}$.

2. $p > \frac{1}{m}$: We follow the same proof as in Example 3.7. In this case, we have $1-(m-1)p < p$, so $\mathbb{P}(Y_d = d) < \mathbb{P}(Y_{d'} = d)$. If we let $A = \{d\}$, it follows that $p \leq \frac{e^\epsilon+\delta}{1+(m-1)e^\epsilon}$ is a necessary condition for (ϵ, δ)-differential privacy to hold.

   Suppose now that $A \subseteq D$ is given and $p \leq \frac{e^\epsilon+\delta}{1+(m-1)e^\epsilon}$. Similar to Example 3.7, (ϵ, δ)-differential privacy holds for the following cases: (i) $d, d' \notin A$; (ii) $d, d' \in A$; and (iii) $d \in A, d' \notin A$.

   Now suppose $d' \in A$ and $d \notin A$. Then we have

$$\begin{aligned}\mathbb{P}(Y_d \in A) &= p|A|, \\ \mathbb{P}(Y_{d'} \in A) &= p(|A| - m) + 1.\end{aligned}$$

   From the hypothesis, we have $p + (m-1)pe^\epsilon \leq e^\epsilon + \delta$, which gives us

$$\begin{aligned}p &\leq e^\epsilon(1 - (m-1)p + p|A| - p|A|) + \delta \\ &= e^\epsilon p|A| + e^\epsilon(1 - (m-1)p - p|A|) + \delta \\ &< e^\epsilon p|A| + (1 - (m-1)p) - p|A| + \delta,\end{aligned}$$

   since $1-(m-1)p < p$ by assumption. Rewriting the above, we have

$$p(m + |A|) - 1 \leq e^\epsilon p|A| + \delta.$$

   Since $p > \frac{1}{m}$, it can be shown that $pm - 1 > 1 - pm$, which then gives us

$$\begin{aligned}\mathbb{P}(Y_{d'} \in A) &= p(|A| - m) + 1 \\ &< p(m + |A|) - 1 \\ &\leq e^\epsilon p|A| + \delta \\ &= e^\epsilon \mathbb{P}(Y_d \in A) + \delta,\end{aligned}$$

   and so (ϵ, δ)-differential privacy holds as required on all $A \subseteq D$.

   However, when $\delta > \frac{1}{m-1}$, we have $\frac{e^\epsilon+\delta}{1+(m-1)e^\epsilon} > \frac{e^\epsilon+\frac{1}{m-1}}{1+(m-1)e^\epsilon} = \frac{1}{m-1}$. Since we must have $p \leq \frac{1}{m-1}$, it is sufficient to have $p \leq \min\left\{\frac{1}{m-1}, \frac{e^\epsilon+\delta}{1+(m-1)e^\epsilon}\right\}$ for (ϵ, δ)-differential privacy to hold.

We therefore require $p \in \left[\frac{1-\delta}{e^\epsilon+m-1}, \frac{1}{m}\right] \cup \left(\frac{1}{m}, \min\left\{\frac{1}{m-1}, \frac{e^\epsilon+\delta}{1+(m-1)e^\epsilon}\right\}\right]$, or equivalently,

$$\frac{1-\delta}{e^\epsilon+m-1} \leq p \leq \min\left\{\frac{1}{m-1}, \frac{e^\epsilon+\delta}{1+(m-1)e^\epsilon}\right\}. \qquad \square$$

Remark: For $(\epsilon, 0)$-differential privacy, we require $\frac{1}{e^\epsilon+m-1} \leq p \leq \frac{1}{e^{-\epsilon}+m-1}$.

We now look at an alternative proof of Theorem 4.4 which established error bounds on the mechanism.

Theorem 4.7 (Error). *The max-mean Hamming error $\mathcal{E}$ of an (ϵ, δ)-differentially private product sanitisation mechanism, whose parent mechanism satisfies (4.15), satisfies*

$$\frac{1-\delta}{1+\frac{e^\epsilon}{m-1}} \leq \frac{\mathcal{E}}{n} \leq \min\left\{1, \frac{e^\epsilon+\delta}{e^\epsilon+\frac{1}{m-1}}\right\}. \tag{4.18}$$

Proof. Let $\boldsymbol{d} \in D^n$. Then,

$$\begin{aligned}
\mathcal{E} &= \max_{\boldsymbol{d}\in D^n} \mathbb{E}\left[\mathrm{h}(Y_{\boldsymbol{d}}, \boldsymbol{d})\right] \\
&= \max_{\boldsymbol{d}\in D^n} \mathbb{E}\left[\sum_{i=1}^{n} \mathrm{h}(Y^i_{\boldsymbol{d}}, d_i)\right] \\
&= \max_{\boldsymbol{d}\in D^n} \sum_{i=1}^{n} \mathbb{E}\left[\mathrm{h}(Y^i_{\boldsymbol{d}}, d_i)\right] \\
&= \max_{\boldsymbol{d}\in D^n} \sum_{i=1}^{n} \mathbb{P}(Y_{d_i} \neq d_i) \\
&= n\left(\max_{d\in D}\left(1 - \mathbb{P}(Y_d = d)\right)\right) \\
&= np(m-1).
\end{aligned}$$

On average, the number of entries that will change is therefore

$$\frac{\mathcal{E}}{n} = p(m-1).$$

As the mechanism satisfies (ϵ, δ)-differential privacy, $p \geq \frac{1-\delta}{e^\epsilon+m-1}$ and $p \leq \min\left\{\frac{1}{m-1}, \frac{e^\epsilon+\delta}{1+(m-1)e^\epsilon}\right\}$ by Theorem 4.6, and so the result follows. $\square$

Remark: As with the discrete exponential mechanism, these bounds are tight and can be achieved by setting p equal to the upper/lower bound established in Theorem 4.6.

4.4.3 *Optimal Mechanism*

We now present the second main result of this section. Using the same max-mean Hamming error as before, we show how to construct the optimal (ϵ, δ)-differentially private mechanism which produces the minimum error. For the purpose of this subsection, we assume the following labelling of elements in the data set:

$$D = \{\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_m\}. \tag{4.19}$$

Definition 4.3 (Design Matrix)**.** *The parent mechanism $\{Y_d\}$ of a product sanitisation mechanism can be defined by a stochastic matrix*

$$P_{\epsilon,\delta} \in [0,1]^{m\times m},$$

where

$$\mathbb{P}(Y_{\tilde{d}_i} = \tilde{d}_j) = \left[P_{(\epsilon,\delta)}\right]_{ij}. \tag{4.20}$$

We refer to $P_{\epsilon,\delta}$ as a product sanitisation design matrix.

Remark: A parent mechanism which satisfies (4.15) has a design matrix $P_{\epsilon,\delta}$ given as follows:

$$\left[P_{\epsilon,\delta}\right]_{ij} = \begin{cases} 1-(m-1)p & \text{if } i = j, \\ p & \text{if } i \neq j. \end{cases} \tag{4.21}$$

Theorem 4.8 (Optimality)**.** *Let $\{Y_d\}$ be a parent mechanism which satisfies (4.15), with $p = \frac{1-\delta}{e^\epsilon+m-1}$. The max-mean Hamming error, $\mathcal{E} = \max_{\boldsymbol{d}\in D^n} \mathbb{E}[\mathrm{h}(Y_{\boldsymbol{d}}, \boldsymbol{d})]$, produced by its product sanitisation mechanism $\{Y_{\boldsymbol{d}}\}$ is the minimum of all (ϵ, δ)-differentially private product sanitisation mechanisms.*

Proof. Let $A = P_{(\epsilon,\delta)}$, satisfying (4.21) with $p = \frac{1-\delta}{e^\epsilon+m-1}$, be the design matrix of $\{Y_d\}$ (i. e. $\mathbb{P}(Y_{\tilde{d}_i} = \tilde{d}_j) = a_{ij}$). Note that this mechanism has the property that $a_{jj} = e^\epsilon a_{ij} + \delta$, since $a_{jj} = \frac{e^\epsilon+(m-1)\delta}{e^\epsilon+m-1}$ and $a_{ij} = \frac{1-\delta}{e^\epsilon+m-1}$, for all $i \neq j$.

The error of its product sanitisation mechanism $\{Y_{\boldsymbol{d}}\}$, using the same method as in the proof of Theorem 4.7, is

$$\begin{aligned}\max_{\boldsymbol{d}\in D^n} \mathbb{E}[\mathrm{h}(Y_{\boldsymbol{d}}, \boldsymbol{d})] &= n\left(\max_{d\in D}\left(1 - \mathbb{P}(Y_d = d)\right)\right) \\ &= n\left(\max_i (1 - a_{ii})\right) \\ &= n\left(1 - \min_i a_{ii}\right) \\ &= n(1 - a_{ii}),\end{aligned}$$

for all $i \in [m]$, since $a_{jj} = 1 - (m-1)p$ for all $j \in [m]$.

Let B be a design matrix defining the (ϵ, δ)-differentially private parent mechanism $\{Y^*_d\}$, with a corresponding product sanitisation mechanism $\{Y^*_{\boldsymbol{d}}\}$, where $B \neq A$.

Since $A \neq B$, and A and B are stochastic, there exists at least one pair (i^*, j^*), where $b_{i^*j^*} < a_{i^*j^*}$. There are two cases to consider:

1. $i^* = j^*$: The error of $\{Y^*_{\boldsymbol{d}}\}$ is then

$$\begin{aligned}\max_{\boldsymbol{d}\in D^n} \mathbb{E}[\mathrm{h}(Y^*_{\boldsymbol{d}}, \boldsymbol{d})] &= n\left(1 - \min_i b_{ii}\right) \\ &\geq n\left(1 - b_{j^*j^*}\right) \\ &> n(1 - a_{j^*j^*}) \\ &= \max_{\boldsymbol{d}\in D^n} \mathbb{E}[\mathrm{h}(Y_{\boldsymbol{d}}, \boldsymbol{d})].\end{aligned}$$

2. $i^* \neq j^*$: As noted previously, $a_{j^*j^*} = e^\epsilon a_{i^*j^*} + \delta$, and since $\{Y^*_d\}$ is (ϵ, δ)-differentially private, $\mathbb{P}(Y^*_{d_{j^*}} = d_{j^*}) \leq e^\epsilon \mathbb{P}(Y^*_{d_{i^*}} = d_{j^*}) + \delta$, or alternatively, $b_{j^*j^*} \leq e^\epsilon b_{i^*j^*} + \delta$. Therefore,

$$\begin{aligned}b_{j^*j^*} &\leq e^\epsilon b_{i^*j^*} + \delta \\ &< e^\epsilon a_{i^*j^*} + \delta \\ &= a_{j^*j^*}.\end{aligned}$$

Hence, $b_{j^*j^*} < a_{j^*j^*}$, and Case 1 applies.

We therefore conclude that

$$\max_{\boldsymbol{d}\in D^n} \mathbb{E}[\mathrm{h}(Y^*_{\boldsymbol{d}}, \boldsymbol{d})] > \max_{\boldsymbol{d}\in D^n} \mathbb{E}[\mathrm{h}(Y_{\boldsymbol{d}}, \boldsymbol{d})],$$

and that $\{Y_{\boldsymbol{d}}\}$ is the product sanitisation mechanism which produces the optimal error. □

Using Theorem 4.8, we now have a simple method to construct the most accurate (ϵ, δ)-differentially private mechanism possible for discrete data. We have proven that no other product sanitisation mechanism is more accurate than it. It follows from Theorem 4.5 that we now know the optimal discrete exponential mechanism that satisfies (4.8).

Remark: The previous result describes an optimal randomised response mechanism for what is referred to as *local differential privacy* in [KOV14]. Where this paper considered an objective function based on Kullback-Leibler divergence and the case $\delta = 0$, our result addresses the relaxed formulation of differential privacy with $\delta \geq 0$ and an objective function based on Hamming distance.

4.5 CONCLUDING REMARKS

The main goal of this chapter was to adapt the general differential privacy framework from Chapter 3 to the special case of categorical data. In doing so, we established a number of important results:

- Sufficient sets for fixed neighbouring databases $\boldsymbol{d} \sim \boldsymbol{d}'$ (Section 4.3) and for the special case of the discrete exponential mechanism with Hamming distance (Corollary 4.2);
- Necessary and sufficient conditions for differential privacy of the discrete exponential mechanism (Theorem 4.3);
- Equivalence of the discrete exponential mechanism with the product sanitisation approach (Theorem 4.5);
- Proof of the optimal discrete exponential mechanism, and hence product sanitisation mechanism, with respect to the max-mean Hamming error (Theorem 4.8).

5

EXTREME POINTS OF THE LOCAL ϵ-DIFFERENTIAL PRIVACY POLYTOPE

In this chapter we study the convex polytope of $m \times m$ stochastic matrices that define ϵ-differentially private mechanisms. We first present invariance properties of the polytope and results reducing the number of constraints needed to define it. Our main results concern the extreme points of the polytope. In particular, we completely characterise these for matrices with 1, 2 *or* m *non-zero columns.*

OVERVIEW

5.1 INTRODUCTION

The product sanitisation mechanism introduced in Chapter 3 provides us with a simple technique to construct (ϵ, δ)-differentially private mechanisms for data release, as well as giving a practical implementation to the concept of local privacy. In Chapter 4 we further examined the sanitised response mechanism by applying it to categorical data, and presented the optimal product sanitisation mechanism with respect to the max-mean Hamming error in Theorem 4.8.

Section 2.2.1

Each sanitised response mechanism is defined by its parent mechanism, which in turn is defined by its design matrix. The set of all design matri-

ces that satisfy (ϵ, δ)-differential privacy forms a polytope in $\mathbb{R}^{m \times m}$. In this chapter, we study the geometry of this polytope for strict differential privacy ($\delta = 0$), with the primary aim of characterising its extreme points.

Studying the extreme points of this polytope gives further insight into optimal mechanisms. An optimal mechanism with respect to a linear error function is guaranteed to coincide with an extreme point of the polytope. Studying the extreme points may also help to add insight to differentially private mechanisms in general.

We begin by introducing the notation and definitions necessary for the chapter in Section 5.2. We then present invariance properties of the polytope, results reducing the number of constraints needed to define it, and some techniques to help identify and construct extreme points of the polytope in Section 5.3. In the main results of this chapter, we completely characterise the extreme points with 1, 2 or m non-zero columns in Section 5.4. A brief discussion of the results are given in Section 5.5 and concluding remarks are given in Section 5.6.

5.2 NOTATION AND BACKGROUND

To begin, let us introduce the major notation and standard definitions to be used in this chapter. For a matrix $A \in \mathbb{R}^{m \times m}$ and $i \in [m]$, we will use $A^{(i)}$ to denote the ith column of A. A^T denotes the usual matrix transpose. We denote by $\mathbf{1}$ the (column) vector of all ones where the dimension will typically be clear from context. We denote by e_i, $i \in [m]$, the ith standard basis vector of $\mathbb{R}^m$.

5.2.1 *Polyhedra*

Polyhedra and polytopes play a central role in this chapter; we recall their definitions now.

Definition 5.1 (Convex Polyhedron). *Let $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ be an inner product on a real vector space V, and let $\left\{c^{(1)}, \ldots, c^{(q+l)}\right\} \subseteq V$ and $b \in \mathbb{R}^{q+l}$ be given. A convex polyhedron $\mathcal{P} \subseteq V$ is defined as:*

$$\mathcal{P} = \left\{ v \in V : \begin{array}{rl} \langle c^{(i)}, v \rangle = b_i, & \forall\, i \in [q], \\ \langle c^{(q+i)}, v \rangle \leq b_{q+i}, & \forall\, i \in [l]. \end{array} \right\}. \tag{5.1}$$

An inequality constraint, corresponding to $i \in [l]$, is said to be *tight* or *active* on a point v if $\langle c^{(q+i)}, v \rangle = b_{q+i}$.

Definition 5.2 (Convex Polytope). *A convex polytope in a vector space V is the convex hull of a finite collection of points in V.*

$$\mathcal{P} = \operatorname{conv}(v_1, \ldots, v_k), \tag{5.2}$$

where $v_i \in V$ for all $i \in [k]$.

It is well known that all polytopes are polyhedra, but only bounded polyhedra are polytopes [BV04].

We now present the definition of an extreme point of a polyhedron, which is a point that cannot be written as a convex combination of any other points in the polyhedron.

Definition 5.3 (Extreme Point). *For a convex polyhedron $\mathcal{P} \subseteq \mathbb{R}^m$, a point $v \in \mathcal{P}$ is an* extreme point *of $\mathcal{P}$ if there are no two distinct points $w \neq x \in \mathcal{P}$ such that $v = \alpha w + (1 - \alpha)x$ for some $\alpha \in (0, 1)$.*

Equivalently, a point $v \in \mathcal{P}$ is an extreme point of $\mathcal{P}$ if $w, z \in \mathcal{P}$ with $\frac{1}{2}(w + z) = v$ implies $w = z = v$.

We denote by $\operatorname{ex}(\mathcal{P})$ the set of all extreme points of a polyhedron $\mathcal{P}$.

Our primary interest is in characterising the extreme points of a polyhedron relating to local differential privacy, as given in Definition 5.4 below. The following theorem from convex geometry shall prove useful in this regard [Bar02].

Theorem 5.1 (Condition for Extreme Points). *Let $\mathcal{P} \subseteq V$ be a polyhedron, and consider a point $v \in \mathcal{P}$. Denote by the set $I_v \subseteq [l]$ the indices of the inequality*

constraints that are tight on v (i.e. $\langle c^{(q+i)}, v\rangle = b_{q+i}$ for all $i \in I_v$ and $\langle c^{(q+i)}, v\rangle < b_{q+i}$ for all $i \in [l] \setminus I_v$). Then $v \in \mathrm{ex}(\mathcal{P})$ if and only if

$$\mathrm{span}\left(\left\{c^{(1)}, \ldots, c^{(q)}\right\} \cup \left\{c^{(q+i)} : i \in I_v\right\}\right) = V.$$

Equivalently, v is an extreme point of P if and only if there are m linearly independent constraints tight on v, where $\dim(V) = m$.

5.2.2 *Differential Privacy*

As we are working with finite, discrete/categorical data, without loss of generality we let $D = [m]$. For a parent mechanism $\{Y_d\}$ to satisfy ϵ-differential privacy (with $\delta = 0$), we require

$$\mathbb{P}(Y_i = k) \leq e^{\epsilon}\mathbb{P}(Y_j = k),$$

for all $i, j, k \in [m]$.

We let $A \in \mathbb{R}^{m\times m}$ be a design matrix of the mechanism given by

$$a_{ij} = \mathbb{P}(X_i = j).$$

Then A defines an ϵ-differentially private mechanism if and only if the following conditions hold:

$$\sum_{j\in[m]} a_{ij} = 1, \qquad \forall\, i \in [m], \tag{5.3a}$$

$$a_{ij} \geq 0, \qquad \forall\, i, j \in [m], \tag{5.3b}$$

$$a_{ik} \leq e^{\epsilon} a_{jk}, \qquad \forall\, i, j, k \in [m]. \tag{5.3c}$$

We now define the ϵ-differential privacy polytope, comprised of all matrices satisfying the above constraints.

Definition 5.4 (Differential Privacy Polytope). *Fix $m \in \mathbb{N}$ and $\epsilon \geq 0$. The ϵ-differential privacy polytope, $\mathcal{D} \subset \mathbb{R}^{m \times m}$, is defined as follows:*

$$\mathcal{D} = \left\{ A \in \mathbb{R}^{m \times m} : \begin{array}{ll} \sum_j a_{ij} = 1, & \forall\, i \in [m], \\ a_{ij} \geq 0, & \forall\, i, j \in [m], \\ a_{ij} \leq e^{\epsilon} a_{kj}, & \forall\, i, j, k \in [m]. \end{array} \right\}. \tag{5.4}$$

The non-negativity and stochastic constraints ensure $\mathcal{D}$ is bounded; therefore it is a polytope.

Note: As the constraint $a_{ij} \leq e^{\epsilon} a_{kj}$ must hold for all $i, j, k \in [m]$, we require $e^{-\epsilon} a_{kj} \leq a_{ij} \leq e^{\epsilon} a_{kj}$ for each i, j, k. Equivalently, $\max_i a_{ij} \leq e^{\epsilon} \min_i a_{ij}$ for all $j \in [m]$.

Remark: If $\epsilon = 0$, then to have $A \in \mathcal{D}$, we require that $a_{ij} = a_{kj}$ for all $i, j, k \in [m]$.

Using the Hilbert Schmidt inner product on the space of $m \times m$ matrices

$$\langle X, Y \rangle = \operatorname{tr}(X^T Y),$$

Example 5.1 below details how to represent the constraints defining $\mathcal{D}$ in the form of Definition 5.1.

Example 5.1 (Matrix Representation). We wish to represent the mechanism constraints (5.3) in matrix form. Beginning with the non-negativity constraint (5.3b),

$$\langle e_i e_j^T, A \rangle = \operatorname{tr}(e_j e_i^T A) = a_{ij}. \tag{5.5}$$

By defining these constraints as

$$C_{ij}^p = -e_i e_j^T, \tag{5.6a}$$

then we require $\langle C_{ij}^p, A \rangle \leq 0$ for each $i, j \in [m]$.

Since tr is a linear function, we can take linear combinations of (5.5) to form the other constraints.

Now consider the stochastic constraint, (5.3a),

$$\left\langle \sum_{j=1}^{m} e_i e_j^T, A \right\rangle = \sum_{j=1}^{m} \operatorname{tr}(e_j e_i^T A) = \sum_{j=1}^{m} a_{ij}.$$

By defining these constraints as

$$C_i^s = \sum_{j=1}^{m} e_i e_j^T, \tag{5.6b}$$

then we require $\langle C_i^s, A \rangle = 1$ for each $i \in [m]$.

Finally, consider the differential privacy constraints, (5.3c),

$$\begin{aligned} \left\langle e_i e_j^T - e^\epsilon e_k e_j^T, A \right\rangle &= \operatorname{tr}(e_j e_i^T A) - e^\epsilon \operatorname{tr}(e_j e_k^T A) \\ &= a_{ij} - e^\epsilon a_{kj}. \end{aligned}$$

By defining these constraints as

$$C_{ijk}^{dp} = e_i e_j^T - e^\epsilon e_k e_j^T, \tag{5.6c}$$

then we require $\langle C_{ijk}^{dp}, A \rangle \leq 0$ for each $i, j, k \in [m]$:

We can now rewrite the definition of $\mathcal{D}$ in the same form as (5.1) using equations (5.6):

$$\mathcal{D} = \left\{ A \in \mathbb{R}^{m \times m} : \begin{array}{ll} \left\langle C_i^s, A \right\rangle = 1, & \forall\, i \in [m], \\ \left\langle C_{ij}^p, A \right\rangle \leq 0, & \forall\, i, j \in [m], \\ \left\langle C_{ijk}^{dp}, A \right\rangle \leq 0, & \forall\, i, j, k \in [m]. \end{array} \right\}.$$

5.3 PRELIMINARY RESULTS

In this section, we present several preliminary results on the structure of the set $\mathcal{D}$ and its extreme points. We first note that the non-negativity constraint in the definition of $\mathcal{D}$ is redundant in the case where $\epsilon > 0$.

Lemma 5.1 (Non-Negativity). *Fix $\epsilon > 0$. Let $v \in \mathbb{R}^m$ satisfy $v_i \leq e^\epsilon v_j$ for all $i, j \in [m]$. Then $v \geq 0$.*

Proof. Let $v_i < 0$ for some $i \in [m]$. Then, for each $j \in [m]$, we have:

$$\begin{aligned} & e^{-\epsilon} v_i \leq v_j \leq e^{\epsilon} v_i \\ \Rightarrow \quad & e^{-\epsilon} v_i \leq e^{\epsilon} v_i \\ \Rightarrow \quad & e^{-\epsilon} \geq e^{\epsilon} \\ \Rightarrow \quad & \epsilon \leq 0. \end{aligned}$$

By hypothesis $\epsilon > 0$. Hence, we must have $v_i \geq 0$ for each $i \in [m]$. □

Our next lemma notes that if the differential privacy constraint is tight on two elements in a column, then those two elements must be the minimum and maximum entries of that column.

Lemma 5.2 (Min/Max Entries)**.** *Let $v \in \mathbb{R}^m$ be a vector with $v_i \leq e^{\epsilon} v_j$ for all $i, j \in [m]$. Suppose there exists at least one pair i, j where $v_i = e^{\epsilon} v_j$. Then $\min_k v_k = v_j$ and $\max_k v_k = v_i$.*

Proof. Suppose there exists v_l such that $v_l > v_i$. Then, $v_l > e^{\epsilon} v_j$, contradicting the differential privacy constraints. Similarly, if $v_l < v_j$, then $e^{\epsilon} v_l < v_i$. The result follows. □

Several of our results will relate the extreme points A of $\mathcal{D}$ to the non-zero columns in A. With this in mind, we formally define

$$\gamma(A) = \{i \in [m] : A^{(i)} \neq 0\},$$

so that $\gamma(A)$ consists of the indices of the non-zero columns of A and $|\gamma(A)| \in [m]$ gives the number of non-zero columns in A.

Our next result concerns the rank of the extreme points of $\mathcal{D}$; first we note the simple observation that $\text{rank}(A) \leq |\gamma(A)|$ for all $A \in \mathcal{D}$.

Theorem 5.2 (Rank of Extreme Points)**.** *Let $A \in \text{ex}(\mathcal{D})$. Then*

$$\text{rank}(A) = |\gamma(A)|.$$

Proof. Suppose $A \in \text{ex}(\mathcal{D})$. As noted before, $\text{rank}(A) \leq |\gamma(A)|$. If A has only one non-zero column, then clearly $\text{rank}(A) = 1 = |\gamma(A)|$.

Let $|\gamma(A)| > 1$ and suppose $\text{rank}(A) < |\gamma(A)|$. Then there exists $\eta \in \mathbb{R}^m$, $\eta \neq 0$ and $\eta_i = 0$ for all $i \notin \gamma(A)$ (i.e. whenever $A^{(i)} = 0$), such that $\sum_i \eta_i A^{(i)} = 0$.

Let $B = A\,\text{diag}(\eta)$. By construction, $B\mathbf{1} = 0$.

Consider $C = A - \Delta B$, $D = A + \Delta B$, where $0 < \Delta < \frac{1}{\max_i |\eta_i|}$. Then,

1. C and D are stochastic, as A is stochastic and $B\mathbf{1} = 0$;
2. since $a_{ij} \leq e^{\epsilon} a_{kj}$ and $(1 \pm \Delta\eta_j) > 0$ for all $i, j, k \in [m]$, it follows that $c_{ij} \leq e^{\epsilon} c_{kj}$, $d_{ij} \leq e^{\epsilon} d_{kj}$; and
3. $C, D \geq 0$.

Hence, $C, D \in \mathcal{D}$ and $C \neq D$ as $B \neq 0$.

However, $\frac{1}{2}(C + D) = A$, so $A \notin \text{ex}(\mathcal{D})$, a contradiction. Therefore, for every $A \in \text{ex}(\mathcal{D})$, we must have $\text{rank}(A) = |\gamma(A)|$. □

We shall often make implicit use of the following simple corollary to the above result; essentially it states that for an extreme point A with at least 2 non-zero columns, none of these columns can have all their entries equal.

Corollary 5.1. *Let $A \in \text{ex}(\mathcal{D})$ satisfy $|\gamma(A)| \geq 2$. Then there is no $i \in \gamma(A)$, $k \in \mathbb{R}$ with $A^{(i)} = k\mathbf{1}$.*

Proof. Suppose that there is some $k \in \mathbb{R}$, $i_0 \in \gamma(A)$ such that $A^{(i_0)} = k\mathbf{1}$. Clearly, $k \neq 0$ as $i_0 \in \gamma(A)$ and $k \neq 1$ as $|\gamma(A)| \geq 2$. As A is stochastic,

$$\sum_{i \in \gamma(A)} A^{(i)} = \mathbf{1}$$

which implies that

$$\left(1 - \frac{1}{k}\right) A^{(i_0)} + \sum_{\substack{i \in \gamma(A) \\ i \neq i_0}} A^{(i)} = 0.$$

This implies that $\text{rank}(A) < |\gamma(A)|$, contradicting Theorem 5.2. □

It is clear from the definition that $\mathcal{D}$ is closed under row/column permutations. Our next result notes that this same invariance property also holds for extreme points.

Theorem 5.3 (Permutation). *Let $A \in \mathcal{D}$ and let $P_1, P_2 \in \{0,1\}^{m\times m}$ be permutation matrices. Then $P_1AP_2 \in \mathcal{D}$. Furthermore, $A \in \text{ex}(\mathcal{D})$ if and only if $P_1AP_2 \in \text{ex}(\mathcal{D})$.*

Proof. From Definition 5.4, it clearly follows that $P_1AP_2 \in \mathcal{D}$ if $A \in \mathcal{D}$, since permuting a matrix only changes the order of rows and columns, but elements in the same row/column will remain in a common row/column.

Now, let $A \in \text{ex}(\mathcal{D})$ and suppose that $P_1AP_2 \notin \text{ex}(\mathcal{D})$ for some permutation matrices P_1 and P_2. Then, there exist $B \neq C \in \mathcal{D}$ such that,

$$P_1AP_2 = \frac{1}{2}(B + C).$$

However, since $P^{-1} = P^T$ for any permutation matrix P, we have

$$A = \frac{1}{2}(P_1^TBP_2^T + P_1^TCP_2^T),$$

where $P_1^TBP_2^T \neq P_1^TCP_2^T$ since $B \neq C$. This is a contradiction since $A \in \text{ex}(\mathcal{D})$, therefore $P_1AP_2 \in \text{ex}(\mathcal{D})$ also. Thus,

$$A \in \text{ex}(\mathcal{D}) \Rightarrow P_1AP_2 \in \text{ex}(\mathcal{D}),$$

and

$$P_1AP_2 \in \text{ex}(\mathcal{D}) \Rightarrow P_1^TP_1AP_2P_2^T = A \in \text{ex}(\mathcal{D}),$$

hence $A \in \text{ex}(\mathcal{D})$ if and only if $P_1AP_2 \in \text{ex}(\mathcal{D})$ for any permutation matrices P_1 and P_2. □

5.3.1 *Tight constraints*

We now examine the implications of Theorem 5.1 for the extreme points of $\mathcal{D}$. We first note a simple fact concerning the number of linearly independent differential privacy constraints that can be tight on an element of $\mathcal{D}$.

In the next result, we use $\mathcal{C}_j^{dp}$ to denote the set of all tight differential privacy constraints acting on the jth column of a matrix A. Formally, given A, this consists of all constraints such that $a_{ij} - e^{\epsilon}a_{kj} = 0$ where $i, k \in [m]$.

Theorem 5.4 (Constraints on Zero Columns)**.** *Let $A \in \mathcal{D}$ be given. Then,* $\dim(\text{span}(\mathcal{C}_j^{dp})) = m$ *if and only if* $A^{(j)} = 0$.

Proof. If we make the obvious identification of the jth column of A with a column vector, $A^{(j)} \in \mathbb{R}^m$, then each constraint in $\mathcal{C}_j^{dp}$ can be identified with a vector of the form $e_i - e^{\epsilon} e_k = (0, \ldots, 1, 0, \ldots, -e^{\epsilon}, 0, \ldots, 0)^T$ where the 1 occurs in the ith position and e^{ϵ} occurs in the kth position. If $\dim(\text{span}(\mathcal{C}_j^{dp})) = m$, there are m linearly independent vectors $v_1, \ldots, v_m$ such that $v_i^T A^{(j)} = 0$ for $i \in [m]$ so it follows trivially that $A^{(j)} = 0$.

For the converse, it is enough to note that $A^{(j)} = 0$ implies that every differential privacy constraint acting on the jth column is tight and that there are m linearly independent such constraints. To see this consider the matrix T with: $t_{ii} = 1$ for $i \in [m]$; $t_{i+1,i} = -e^{\epsilon}$ for $i \in [m-1]$; $t_{1m} = -e^{\epsilon}$; $t_{jk} = 0$ otherwise. It can readily be verified that T is non-singular. □

Remark: A direct consequence of Theorem 5.4 is that $\dim(\text{span}(\mathcal{C}_j^{dp})) \leq m - 1$ for any $j \in \gamma(A)$.

Our later characterisations of the extreme points of $\mathcal{D}$ shall rely on the following concept of *loose entries*.

Definition 5.5 (Loose Entries of a Matrix)**.** *Given $A \in \mathcal{D}$, define*

$$\lambda(A) = \left\{ (i,j) : a_{ij} \notin \left\{ e^{\epsilon} \min_{k \in [m]} a_{kj}, e^{-\epsilon} \max_{k \in [m]} a_{kj} \right\} \right\}.$$

For a matrix $A \in \mathcal{D}$, we say the entry a_{ij} is loose *if $(i,j) \in \lambda(A)$.*

It follows from Lemma 5.2 that for any (i,j) there exists a k such that $a_{ij} = e^{\pm\epsilon} a_{kj}$ if and only if $(i,j) \notin \lambda(A)$.

Example 5.2 (Loose Entries)**.** Let $\epsilon = \ln(2)$ and

$$A = \frac{1}{7} \begin{pmatrix} 4 & 1 & 2 \\ 3 & 2 & 2 \\ 2 & 1 & 4 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Then $\lambda(A) = \{(2,1)\}$, since $3 \notin \{4,2\}$, and by similar reasoning, $\lambda(B) = \{(1,1),(2,1),(3,1)\}$.

Our next result bounds the number of loose entries of an extreme point in terms of the number of non-zero columns.

Theorem 5.5 (Bound on Loose Entries)**.** *Let $A \in \text{ex}(\mathcal{D})$ with $|\gamma(A)| \geq 2$. Then,*

$$|\lambda(A)| \leq m - |\gamma(A)|.$$

Proof. Let $A \in \text{ex}(\mathcal{D})$ and consider the following sets of constraints active on A. We define

$$\mathcal{C}^{dp} = \bigcup_{j \in \gamma(A)} \mathcal{C}_j^{dp}$$

to be the set of tight differential privacy constraints acting on the columns in $\gamma(A)$. Note the following readily verifiable facts:

(i) for $j \notin \gamma(A)$, every differential privacy constraint acting on column j is tight;

(ii) the m stochastic constraints are tight;

(iii) as $|\gamma(A)| \geq 2$, no non-zero column of A is of the form $k\mathbf{1}$ where $k \in \mathbb{R}$.

It follows from (ii) and Theorem 5.1 that the number of tight, linearly independent differential privacy constraints on A must be $(m^2 - m)$. Furthermore, Theorem 5.4 implies that there are m linearly independent differential privacy constraints active on each of the $(m - |\gamma(A)|)$ zero columns of A. It is not difficult to see that constraints acting on different columns must be linearly independent and hence there are a total of $(m - |\gamma(A)|)m$ linearly independent tight differentially private constraints arising from the zero columns of A. Putting all of this together, we see that there must be

$$m^2 - m - (m - |\gamma(A)|)m = m|\gamma(A)| - m$$

tight differential privacy constraints acting on the non-zero columns of A. Formally:

$$|\mathcal{C}^{dp}| \geq m|\gamma(A)| - m. \tag{5.7}$$

From point (iii) above there are no non-zero columns in which all entries are constant; it follows that for each $j \in \gamma(A)$,

$$|\{i : (i,j) \notin \lambda(A)\}| \geq |\mathcal{C}_j^{dp}| + 1.$$

If we let l_j denote the number of loose entries in column j, the previous inequality can be rewritten as

$$|\mathcal{C}_j^{dp}| \leq m - l_j - 1.$$

Combining this with (5.7) we see that

$$\begin{aligned} m|\gamma(A)| - m &\leq \sum_{j\in\gamma(A)} |\mathcal{C}_j^{dp}| \\ &\leq \sum_{j\in\gamma(A)} m - l_j - 1 \\ &= m|\gamma(A)| - |\lambda(A)| - |\gamma(A)|. \end{aligned}$$

A simple rearrangement now shows that

$$|\lambda(A)| \leq m - |\gamma(A)|$$

as claimed. □

Note: When $|\gamma(A)| = 1$, $|\lambda(A)| = m$.

As we conclude this sub-section, we take a look at the following result for later use, which states that at most one loose entry can appear in any row of an extreme point.

Lemma 5.3 (Loose Entries per Row)**.** *Let $A \in \text{ex}(\mathcal{D})$. No row of A has more than one loose entry (i.e. there exist no two distinct pairs $(i_1, j_1), (i_1, j_2) \in \lambda(A)$ with $j_1 \neq j_2$).*

Proof. Let $A \in \text{ex}(\mathcal{D})$, and assume without loss of generality that $(1,1), (1,2) \in \lambda(A)$. Let

$$\Delta = \min\left\{\max_i a_{i1} - a_{11}, a_{11} - \min_i a_{i1}, \max_i a_{i2} - a_{12}, a_{12} - \min_i a_{i2}\right\}.$$

Hence, $A \pm \Delta(E_{11} - E_{12}) \in \mathcal{D}$.

However, $A = \frac{1}{2}((A + \Delta E_{11} - \Delta E_{12}) + (A - \Delta E_{11} + \Delta E_{12}))$, hence, $A \notin \text{ex}(\mathcal{D})$, a contradiction and so the result follows. □

Finally, for this sub-section we present a number of other results that will add further insight to the behaviour and structure of $\mathcal{D}$ and its extreme points. The next piece of notation will prove useful later.

For $A \in \mathcal{D}$, we define the vector $w' \in \mathbb{R}^m$ where $w'_j = \frac{1}{\min_i a_{ij}}$ for any $j \in \gamma(A)$ and $w'_j = 0$ otherwise. We then denote by $\tilde{A}$ the matrix given by:

$$\tilde{A} = A \operatorname{diag}(w'). \tag{5.8}$$

Then, for any $A \in \mathcal{D}$, $\tilde{a}_{ij} \in [1, e^\epsilon]$ for any $j \in \gamma(A)$, and $\tilde{a}_{ij} = 0$ otherwise. Hence,

$$\tilde{A} \operatorname*{diag}_{1 \le j \le m} \left(\min_i a_{ij} \right) = A.$$

Note: $\gamma(A) = \gamma(\tilde{A})$ and $\lambda(A) = \lambda(\tilde{A})$.

We now show that for any extreme point A, $\tilde{A}$ cannot have a row with equal non-zero values.

Lemma 5.4. *Let $A \in \operatorname{ex}(\mathcal{D})$ with $|\gamma(A)| > 1$. Then for each row $i \in [m]$, there exist two non-zero columns $j, k \in \gamma(A)$ such that $\tilde{a}_{ij} \neq \tilde{a}_{ik}$.*

Proof. We prove this by contradiction. Firstly, suppose there exists a row i such that $\tilde{a}_{ij} = \tilde{a}_{ik}$ for all $j, k \in \gamma(A)$. By Lemma 5.3, each row cannot have more than one loose element, therefore either $\tilde{a}_{ij} = 1$ or $\tilde{a}_{ij} = e^\epsilon$.

Let $w \in \mathbb{R}^m$ be defined by $w_j = \min_i a_{ij}$. Then $A = \tilde{A} \operatorname{diag}(w)$.

Suppose $\tilde{a}_{ij} = 1$, hence $\sum_{k \in \gamma(A)} w_k = 1$. By Theorem 5.5, each column j has at least one pair (i, k) such that $a_{ij} = e^\epsilon a_{kj}$, hence there exists a row i^* such that $\tilde{a}_{i^* j} = e^\epsilon$. However, $\tilde{a}_{i^* k} \geq 1$ for every $k \in \gamma(A)$, so $\sum_{k \in \gamma(A)} \tilde{a}_{i^* k} w_k > 1$, contradicting the stochasticity of A.

A similar argument holds for $\tilde{a}_{ij} = e^\epsilon$. The result follows. □

5.3.2 *Computer Simulations*

The following result is useful when conducting computer simulations, by allowing us to efficiently search for extreme points with specified zero columns. *MATLAB* code making use of this theorem can be found in Appendix A.

Theorem 5.6. *Let $I \subseteq [m]$ and define*

$$\mathcal{D}_I = \left\{ A \in \mathcal{D} : I \cap \gamma(A) = \emptyset \right\}.$$

We therefore have $A^{(i)} = 0$ for each $A \in \mathcal{D}_I$ and $i \in I$.

Then:

(i) $\mathrm{ex}(\mathcal{D}_I) = \{A \in \mathrm{ex}(\mathcal{D}) : I \cap \gamma(A) = \emptyset\}$;

(ii) $\mathcal{D}_\emptyset = \mathcal{D}$;

(iii) $\mathcal{D}_{[m]} = \emptyset$;

(iv) $\mathcal{D}_I \subseteq \mathcal{D}_J$ *for all* $J \subseteq I$;

(v) $\mathrm{ex}(\mathcal{D}_I) \subseteq \mathrm{ex}(\mathcal{D}_J)$ *for all* $J \subseteq I$;

(vi) $\mathcal{D}_{I\cup J} = \mathcal{D}_I \cap \mathcal{D}_J$ *for all* I, J;

(vii) $\mathrm{ex}(\mathcal{D}_{I\cup J}) = \mathrm{ex}(\mathcal{D}_I) \cap \mathrm{ex}(\mathcal{D}_J)$ *for all* I, J.

Proof. Let us begin with (i). Let $A \in \mathcal{D}_I$. Since $\gamma(A)$ gives the indices of non-zero columns, $(A^T\mathbf{1})_i = 0$ for each $i \in I$. Since $A \in \mathcal{D}$ also, if we write $\mathrm{ex}(\mathcal{D}) = \{D_1, \ldots, D_k\}$ we have $A = \sum_{i=1}^k \alpha_i D_i$, where $\alpha_i \geq 0$ for each $i \in [k]$ and $\sum_{i=1}^k \alpha_i = 1$. However, since $A \geq 0$ and $D_i \geq 0$ for each $i \in [k]$, then $\alpha_i = 0$ whenever $I \cap \gamma(D_i) \neq \emptyset$.

Hence, each $A \in \mathcal{D}_I$ can be written as a convex combination of $\{A \in \mathrm{ex}(\mathcal{D}) : I \cap \gamma(A) = \emptyset\}$, and so $\mathrm{ex}(\mathcal{D}_I) \subseteq \{A \in \mathrm{ex}(\mathcal{D}) : I \cap \gamma(A) = \emptyset\}$.

Since each $A \in \mathrm{ex}(\mathcal{D})$ cannot be written as a convex combination of any other elements of $\mathrm{ex}(\mathcal{D})$, we conclude that $\mathrm{ex}(\mathcal{D}_I) = \{A \in \mathrm{ex}(\mathcal{D}) : I \cap \gamma(A) = \emptyset\}$.

(ii)–(vii) follow from (i). □

5.4 EXTREME POINTS FOR FIXED VALUES OF $|\gamma(A)|$

In this section, we characterise extreme points with a specified number of non-zero columns. We note that extreme points with one and two non-zero columns are limited to a specific form, while Section 5.4.3 deals with extreme points with any number of non-zero columns.

5.4.1 *Extreme Points with One Column Non-Zero*

The first case to consider is that of a single non-zero column in the matrix. Due to the stochastic constraints, there are only m such matrices, and as Theorem 5.7 below states, each one of these matrices is an extreme point.

Theorem 5.7 ($|\gamma(A)| = 1$)**.** *Let $E_i \in \mathbb{R}^{m\times m}$ be given by $E_i = \mathbf{1}e_i^T$ for $i \in [m]$ and define the set $\tilde{\mathcal{D}}'$ as:*

$$\tilde{\mathcal{D}}' = \{E_1, \dots, E_m\}.$$

Then $\tilde{\mathcal{D}}' \subseteq \mathrm{ex}(\mathcal{D})$.

Furthermore, $A \in \mathrm{ex}(\mathcal{D})$, $|\gamma(A)| = 1$ implies that $A \in \tilde{\mathcal{D}}'$.

Proof. Suppose $E_i = \frac{1}{2}(B + C)$ for B, C in $\mathcal{D}$. As B, C are both non-negative, it follows immediately that all columns of B and C apart from the ith column are zero. B and C are also both stochastic which immediately implies that $B = C = \mathbf{1}e_i^T$.

Note that if $A \in \mathcal{D}$ with $|\gamma(A)| = 1$, then $A = E_i$ for some i. Hence, if $A \in \mathrm{ex}(\mathcal{D})$ with $|\gamma(A)| = 1$, it follows that $A \in \tilde{\mathcal{D}}'$. □

The points in $\tilde{\mathcal{D}}'$ are extreme points in all cases, regardless of ϵ. Furthermore, in the trivial case of $\epsilon = 0$, the set $\tilde{\mathcal{D}}'$ are the only extreme points.

Corollary 5.2. *Let $\epsilon = 0$. Then,*

$$\mathrm{ex}(\mathcal{D}) = \tilde{\mathcal{D}}'.$$

Proof. Let $\epsilon = 0$. Then, for all $A \in \mathcal{D}$, we have $a_{kj} \leq a_{ij} \leq a_{kj}$, hence $a_{ij} = a_{kj}$ for all $i, j, k \in [m]$, i. e. entries in the same column are equal. It now follows immediately from Corollary 5.1 that if A is an extreme point, $\gamma(A) = 1$ and hence that $A \in \tilde{\mathcal{D}}'$ as claimed. □

5.4.2 *Extreme Points with Two Columns Non-Zero*

Next, we consider the case of two non-zero columns. Although Theorem 5.5 allows for many loose entries to occur in these extreme points, Theorem 5.8 below states that no loose entries are possible.

Theorem 5.8 ($|\gamma(A)| = 2$)**.** *Let $A \in \mathrm{ex}(\mathcal{D})$ where $|\gamma(A)| = 2$. Then A has no loose entries.*

Proof. Without loss of generality, assume that $\gamma(A) = \{1, 2\}$. Define $w \in \mathbb{R}^m$ by $w_j = \min_i a_{ij}$ for $j \in [m]$ and let $\tilde{A}$ be given by (5.8). Then $\tilde{a}_{ij} \in [1, e^\epsilon] \cup \{0\}$ for $i, j \in [m]$.

By Theorem 5.5, $|\lambda(A)| \leq m-2$, so there exist at least two rows with no loose entries. Let row k be one of these rows. Then $\tilde{a}_{k1}, \tilde{a}_{k2} \in \{1, e^\epsilon\}$, but by Lemma 5.4, $\tilde{a}_{k1} \neq \tilde{a}_{k2}$. We can assume that $\tilde{a}_{k1} = e^\epsilon$ and $\tilde{a}_{k2} = 1$ (otherwise just swap columns 1 and 2). As A is stochastic,

$$w_1 e^\epsilon + w_2 = 1. \tag{5.9a}$$

By Lemma 5.3, for all rows $j \in [m]$, at least one of $\tilde{a}_{j1}, \tilde{a}_{j2}$ must be in $\{1, e^\epsilon\}$. Moreover, in order to satisfy (5.9a), $\tilde{a}_{j1} = e^\epsilon$ if and only if $\tilde{a}_{j2} = 1$.

Suppose therefore that there exists a row j where $\tilde{a}_{j1} \in (1, e^\epsilon)$ corresponding to a loose entry in A. It follows from (5.9a) that $\tilde{a}_{j2} = e^\epsilon$. Hence

$$\begin{aligned} 1 &= w_1 \tilde{a}_{j1} + w_2 e^\epsilon \\ &> w_1 + w_2 e^\epsilon. \end{aligned} \tag{5.9b}$$

It follows from Corollary 5.1 that there is some j^* such that $\tilde{a}_{j^*1} = 1$, implying

$$\begin{aligned} 1 &= w_1 + w_2 \tilde{a}_{j^*2} \\ &\leq w_1 + w_2 e^\epsilon, \end{aligned}$$

contradicting (5.9b). Therefore there are no loose entries in the first column.

Now suppose there exists a row j where $\tilde{a}_{j2} \in (1, e^\epsilon)$. As above, it follows that $\tilde{a}_{j1} = 1$. Hence,

$$\begin{aligned} 1 &= w_1 + w_2 \tilde{a}_{j2} \\ &< w_1 + w_2 e^\epsilon. \end{aligned} \tag{5.9c}$$

As before, it follows from Corollary 5.1 that there is some j^* such that $\tilde{a}_{j^*2} = e^\epsilon$, hence,

$$\begin{aligned} 1 &= w_1 \tilde{a}_{j^*1} + w_2 e^\epsilon \\ &\geq w_1 + w_2 e^\epsilon, \end{aligned}$$

contradicting (5.9c). Therefore there are no loose entries in the second column.

Hence $|\lambda(A)| = 0$. □

Using this result along with Lemma 5.4, we can describe the two non-zero columns.

Corollary 5.3. *Let $A \in \text{ex}(\mathcal{D})$ with $|\gamma(A)| = 2$. Let $\gamma(A) = \{j, k\}$ and $\tilde{A}$ be given by (5.8). Then, for every $i \in [m]$, we have*

$$(\tilde{a}_{ij}, \tilde{a}_{ik}) \in \{(1, e^\epsilon), (e^\epsilon, 1)\}. \tag{5.10}$$

Proof. By Theorem 5.8, $\tilde{a}_{ij} \in \{1, e^\epsilon\}$ and $\tilde{a}_{ik} \in \{1, e^\epsilon\}$ for each $i \in [m]$.

By Lemma 5.4, we must have $\tilde{a}_{ij} \neq \tilde{a}_{ik}$ for each $i \in [m]$. So, $\tilde{a}_{ij} = e^\epsilon$ if and only if $\tilde{a}_{ik} = 1$, and $\tilde{a}_{ij} = 1$ if and only if $\tilde{a}_{ik} = e^\epsilon$. □

The follow example illustrates the consequence of Corollary 5.3.

Example 5.3. Every extreme point $A \in \text{ex}(\mathcal{D})$ with $|\gamma(A)| = 2$ must be of the form shown in (5.10), and furthermore both non-zero columns of $\tilde{A}$ must contain at least one 1 and one e^ϵ.

Let $m = 4$ and $A \in \text{ex}(\mathcal{D})$ with $|\gamma(A)| = 2$. One example of such an A is as follows:

$$A = \frac{1}{1+e^\epsilon} \begin{pmatrix} 1 & 0 & e^\epsilon & 0 \\ 1 & 0 & e^\epsilon & 0 \\ e^\epsilon & 0 & 1 & 0 \\ 1 & 0 & e^\epsilon & 0 \end{pmatrix} \in \text{ex}(\mathcal{D}).$$

5.4.3 *Extreme Points with Every Element Constrained*

The next definition is necessary before we can state Theorem 5.9 which is the main result of the chapter.

Definition 5.6. *Let $\tilde{\mathcal{D}} \subset \mathcal{D}$ be defined as follows:*

$$\tilde{\mathcal{D}} = \{A \in \mathcal{D} \mid \text{rank}(A) = |\gamma(A)|, \lambda(A) = \emptyset\}. \tag{5.11}$$

The set $\tilde{\mathcal{D}}$ contains matrices with between 2 and m non-zero columns, which satisfy the rank condition of Theorem 5.2 and have no loose entries (i.e. $\tilde{a}_{ij} \in \{0, 1, e^\epsilon\}$ for each $i, j \in [m]$, where $\tilde{A}$ is given in (5.8)). We now show that every one of these matrices is an extreme point of $\mathcal{D}$.

Theorem 5.9. *Let $\epsilon > 0$. Then,*

$$\tilde{\mathcal{D}} \subset \mathrm{ex}(\mathcal{D}).$$

Proof. Let $A \in \tilde{\mathcal{D}}$ and let $B, C \in \mathcal{D}$ where $\frac{1}{2}(B+C) = A$. Define $w_j = \min_i a_{ij}$ for each $j \in [m]$ (Note that $w_j = 0$ for each $j \notin \gamma(A)$, and $a_{ij} \in \{w_j, e^\epsilon w_j\}$ for each $i, j \in [m]$ since $\lambda(A) = \emptyset$).

Let $\Delta_j = \frac{1}{2}\max|B^{(j)} - C^{(j)}|$ for each $j \in \gamma(A)$. As B and C are non-negative, it is not hard to see that:

$$\Delta_j = 0, \quad \forall\, j \notin \gamma(A). \tag{5.12a}$$

We shall show that the same conclusion must also hold for $j \in \gamma(A)$. To this end, let $j^* \in \gamma(A)$ be given where $\Delta_{j^*} > 0$. Assume without loss of generality that $b_{i_1 j^*} = a_{i_1 j^*} + \Delta_{j^*}$ for some $i_1 \in [m]$ (if not, swap B and C).

We claim that $a_{i_1 j^*} \neq w_{j^*}$. Suppose otherwise. Then there exists $i_2 \in [m]$ where $a_{i_2 j^*} = e^\epsilon w_{j^*}$. However, since $\frac{1}{2}(B+C) = A$, we have $c_{i_1 j^*} = 2a_{i_1 j^*} - b_{i_1 j^*} = a_{i_1 j^*} - \Delta_{j^*}$, and since $C \in \mathcal{D}$, we have

$$\begin{aligned} c_{i_2 j^*} &\leq e^\epsilon c_{i_1 j^*} \\ &= e^\epsilon a_{i_1 j^*} - e^\epsilon \Delta_{j^*} \\ &= a_{i_2 j^*} - e^\epsilon \Delta_{j^*}. \end{aligned}$$

By the definition of Δ_{j^*}, we must have $c_{i_2 j^*} \geq a_{i_2 j^*} - \Delta_{j^*}$. Hence it would follow that $\Delta_{j^*} \geq e^\epsilon \Delta_{j^*}$, a contradiction since $\epsilon > 0$. Thus, $a_{i_1 j^*} = e^\epsilon w_{j^*}$ as claimed (i. e. the max change occurs on the max element of the column).

We now know that $b_{i_1 j^*} = e^\epsilon w_{j^*} + \Delta_{j^*}$. Let

$$I_{j^*} = \{i : a_{ij^*} = w_{j^*}\}.$$

Then for every $i \in I_{j^*}$, since $B \in \mathcal{D}$, we get $e^\epsilon w_{j^*} + \Delta_{j^*} = b_{i_1 j^*} \leq e^\epsilon b_{ij^*}$, hence

$$b_{ij^*} \geq w_{j^*} + e^{-\epsilon}\Delta_{j^*}. \tag{5.12b}$$

Also, for every $i \in I_{j^*}$, since $C \in \mathcal{D}$,

$$\begin{aligned} c_{i_1 j^*} &= e^{\epsilon} w_{j^*} - \Delta_{j^*} \\ &\leq e^{\epsilon} c_{ij^*} \\ &= e^{e} (2a_{ij^*} - b_{ij^*}) \\ &= 2e^{\epsilon} w_{j^*} - e^{\epsilon} b_{ij^*}, \end{aligned}$$

hence $e^{\epsilon} w_{j^*} - \Delta_{j^*} \leq 2e^{\epsilon} w_{j^*} - e^{\epsilon} b_{ij^*}$, or rewriting,

$$b_{ij^*} \leq w_{j^*} + e^{-\epsilon} \Delta_{j^*}. \tag{5.12c}$$

Hence, from (5.12b) and (5.12c),

$$\begin{aligned} b_{ij^*} &= w_{j^*} + e^{-\epsilon} \Delta_{j^*} \\ &= a_{ij^*} + e^{-\epsilon} \Delta_{j^*}, \end{aligned} \tag{5.12d}$$

for every $i \in I_{j^*}$.

It follows readily that for every $i \in I_{j^*}$, $c_{ij^*} = w_{j^*} - e^{-\epsilon} \Delta_{j^*}$.

We next consider indices $i \notin I_{j^*}$. Choose some $i_2 \in I_{j^*}$. For all $i \notin I_{j^*}$, $a_{ij^*} = e^{\epsilon} w_{j^*}$, then

$$b_{ij^*} \leq e^{\epsilon} b_{i_2 j^*} = e^{\epsilon} w_{j^*} + \Delta_{j^*}, \tag{5.12e}$$

and

$$\begin{aligned} c_{ij^*} &= 2a_{ij^*} - b_{ij^*} \\ &= 2e^{\epsilon} w_{j^*} - b_{ij^*} \\ &\leq e^{\epsilon} c_{i_2 j^*} \\ &= e^{\epsilon} w_{j^*} - \Delta_{j^*}, \end{aligned}$$

which can be rewritten as

$$b_{ij^*} \geq e^{\epsilon} w_{j^*} + \Delta_{j^*}. \tag{5.12f}$$

Hence, from (5.12e) and (5.12f),

$$\begin{aligned} b_{ij^*} &= e^{\epsilon} w_{j^*} + \Delta_{j^*} \\ &= a_{ij^*} + \Delta_{j^*}, \end{aligned} \tag{5.12g}$$

for all $i \notin I_{j^*}$.

Putting everything together, it follows from (5.12a), (5.12d) and (5.12g),

$$b_{ij} = \begin{cases} 0, & j \notin \gamma(A), \\ w_j + e^{-\epsilon} g_j \Delta_j, & j \in \gamma(A), i \in I_j \\ e^{\epsilon} w_j + g_j \Delta_j, & j \in \gamma(A), i \notin I_j \end{cases}$$

where $g_j \in \{-1, 1\}$, for all $j \in \gamma(A)$.

Similarly, since $B + C = 2A$,

$$c_{ij} = \begin{cases} 0, & j \notin \gamma(A), \\ w_j - e^{-\epsilon} g_j \Delta_j, & j \in \gamma(A), i \in I_j \\ e^{\epsilon} w_j - g_j \Delta_j, & j \in \gamma(A), i \notin I_j. \end{cases}$$

Rewriting in terms of $\tilde{A}$ (given by (5.8)), $b_{ij} = a_{ij} + g_j e^{-\epsilon} \frac{a_{ij}}{w_j} \Delta_j = a_{ij} + g_j e^{-\epsilon} \tilde{a}_{ij} \Delta_j$ and $c_{ij} = a_{ij} - g_j e^{-\epsilon} \tilde{a}_{ij} \Delta_j$ for all $i, j \in [m]$.

Hence,

$$\begin{aligned} B &= A + e^{-\epsilon} \tilde{A} \operatorname*{diag}_{j \in [m]}(g_j \Delta_j) \\ C &= A - e^{-\epsilon} \tilde{A} \operatorname*{diag}_{j \in [m]}(g_j \Delta_j). \end{aligned}$$

Since A, B are stochastic, we require

$$e^{-\epsilon} \tilde{A} \operatorname*{diag}_{j \in [m]}(g_j \Delta_j) \mathbf{1} = 0.$$

This equation defines a linear relationship between the columns of $\tilde{A}$. Moreover, we know that $\Delta_j = 0$ for $j \notin \gamma(A)$. If $\Delta_j \neq 0$ for any $j \in \gamma(A)$, it would imply that the non-zero columns of $\tilde{A}$ and hence those of A are linearly dependent, contradicting the assumption that $\operatorname{rank}(A) = |\gamma(A)|$. It

follows that $\Delta_j = 0$ for all $j \in [m]$ and hence that $B = C = A$. This completes the proof. □

Furthermore, the set $\tilde{\mathcal{D}}$ contains all extreme points of $\mathcal{D}$ which have no loose entries.

Corollary 5.4. *Let $A \in \mathcal{D}$ with $\lambda(A) = \emptyset$. Then, $A \in \text{ex}(\mathcal{D})$ if and only if $A \in \tilde{\mathcal{D}}$.*

Proof. "⇒": Let $A \in \text{ex}(\mathcal{D})$ with $\lambda(A) = \emptyset$. By Theorem 5.2, $\text{rank}(A) = |\gamma(A)|$, hence $A \in \tilde{\mathcal{D}}$.

"⇐": $A \in \tilde{\mathcal{D}} \Rightarrow A \in \text{ex}(\mathcal{D})$ by Theorem 5.9. □

5.4.4 *Extreme Points with All Columns Non-Zero*

From an application point of view, it is entirely reasonable to only consider matrices (and the resulting response mechanism) with no zero columns.

Having a zero column in a matrix that defines a response mechanism means that the mechanism never releases a particular (or multiple) values as its output. In many circumstances, this feature will not be required of a mechanism.

Using Theorem 5.9, we now present the following corollary, which gives a complete characterisation of extreme points without zero columns.

Corollary 5.5. *Let $A \in \mathcal{D}$, with $|\gamma(A)| = m$. Then, $A \in \text{ex}(\mathcal{D})$ if and only if $A \in \tilde{\mathcal{D}}$*

Equivalently,

$$\{A \in \text{ex}(\mathcal{D}) : |\gamma(A)| = m\} = \{A \in \tilde{\mathcal{D}} : |\gamma(A)| = m\}.$$

Proof. "⇒": Let $A \in \text{ex}(\mathcal{D})$ have m non-zero columns. Then, $\text{rank}(A) = m$ by Theorem 5.2 and $\lambda(A) = \emptyset$ by Theorem 5.5.

"⇐": Let $A \in \mathcal{D}$ such that $\text{rank}(A) = m$ and $\lambda(A) = \emptyset$. Then $A \in \text{ex}(\mathcal{D})$ by Theorem 5.9. □

We now have necessary and sufficient conditions for finding and determining extreme points with m non-zero columns.

5.5 DISCUSSION

We now take a brief look at a number of useful and interesting consequences of the results given in Sections 5.3 and 5.4.

$\mathrm{ex}(\mathcal{D})$ FOR SMALL m From Theorems 5.7 and 5.9, we know that $\tilde{\mathcal{D}}' \cup \tilde{\mathcal{D}} \subseteq \mathrm{ex}(\mathcal{D})$; with the addition of Theorem 5.8 we can make further observations for small m.

Theorem 5.10. *Let $m \leq 3$, then*

$$\mathrm{ex}(\mathcal{D}) = \tilde{\mathcal{D}}' \cup \tilde{\mathcal{D}}.$$

We therefore have a complete characterisation of all extreme points up to $m = 3$.

$\mathrm{ex}(\mathcal{D})$ FOR $m = 4$ While we lack a formal proof, extensive computer simulations suggest the following fact for the case $m = 4$.

Let $m \leq 4$, then

$$\mathrm{ex}(\mathcal{D}) = \tilde{\mathcal{D}}' \cup \tilde{\mathcal{D}}.$$

MATLAB code used for the computer simulations can be found in Appendix A.

$\mathrm{ex}(\mathcal{D})$ FOR $m \geq 5$ When $m = 5$, our previous results allow us to characterise all extreme points A for which $|\gamma(A)| = 1, 2, 5$. However, when $|\gamma(A)| = 4$, we can find extreme points with loose entries.

The following point $A \in \mathcal{D}$ can be shown to be an extreme point of $\mathcal{D}$ by using Theorem 5.1.

$$A = \frac{1}{3 + 2e^{\epsilon}} \begin{pmatrix} 1 & 1 & 2e^{\epsilon} & 1 & 0 \\ e^{\epsilon} & 1 & 2 & e^{\epsilon} & 0 \\ e^{\epsilon} & e^{\epsilon} & 2 & 1 & 0 \\ 1 & e^{\epsilon} & 2 & e^{\epsilon} & 0 \\ 1 & 1 & 1 + e^{\epsilon} & e^{\epsilon} & 0 \end{pmatrix}.$$

Fitting with Theorem 5.5, A has only a single loose entry ($\lambda(A) = \{(5,3)\}$), while we also observe that $\texttt{rank}(A) = 4$, satisfying Theorem 5.2.

We therefore have $A \in \text{ex}(\mathcal{D})$, but $A \notin \tilde{\mathcal{D}}' \cup \tilde{\mathcal{D}}$. Hence, $\tilde{\mathcal{D}}' \cup \tilde{\mathcal{D}} \subset \text{ex}(\mathcal{D})$ in general.

5.6 CONCLUDING REMARKS

The main goal of this chapter was to present results on the extreme points of the ϵ-differential privacy polytope, with the main result being the complete characterisation of all extreme points with no loose entries (Theorem 5.9). Other contributions included:

- Methods to help identify extreme points, such as evaluating the matrix's rank (Theorem 5.2), and the number and position of loose entries (Theorem 5.5 and Lemma 5.3);
- Proving that the set of extreme points is closed under row and column permutation (Theorem 5.3);
- Results to help with the enumeration of extreme points with computer software (Theorem 5.3.2 and Appendix A);
- The complete characterisation of extreme points with 1, 2 and m non-zero columns (Theorems 5.7 and 5.3, and Corollary 5.5 respectively).

6

THE RANDOMISED RESPONSE TECHNIQUE

In this chapter, we study the Randomised Response (RR) technique as a means to achieve privacy in surveying. We examine a generalisation of the original RR technique, and by measuring the error of the statistical estimator, determine the optimal (ϵ, δ)-differentially private RR mechanism for $\delta \leq \frac{1}{2}$. We also examine a privacy-protection metric, apply RR to databases and extend the RR technique to non-binary questioning.

OVERVIEW

6.1 INTRODUCTION

In previous chapters we have studied the sanitised response mechanism as a means to achieve local (ϵ, δ)-differential privacy and studied its privacy/utility tradeoff through its error on an element-by-element basis. In this chapter, we take a different approach to determining the error of such a mechanism, by considering local privacy as a privacy-preserving means to determine the proportion of a population possessing a certain attribute.

For this we turn to the Randomised Response (RR) technique introduced by Stanley L. Warner [War65]. In Section 6.2 we examine a generalisation of Warner's original RR technique, and formulate a Maximum Likelihood Es-

Section 2.5.4

timator (MLE) to calculate a true population estimate from the randomised responses. We then examine this RR model in the context of differential privacy in Section 6.3, and by minimising the error of the MLE, we determine the optimal (ϵ, δ)-differentially private RR mechanism for $\delta \leq \frac{1}{2}$.

We then consider a proposed improvement to Warner's original model, known as Mangat's model [Man94]. By examining a metric of privacy-protection on the RR mechanisms in Section 6.4, we show Mangat's model to be more efficient than Warner's. In Section 6.5 we examine the error introduced by sampling in Mangat's model, and how applying RR to databases reduces the overall error of the mechanism.

Finally in Section 6.6, we extend Mangat's method to categorical variables, where answers are no longer binary. We present closed-form Maximum Likelihood Estimators (MLEs) for the model, and initial results on variance of these estimators. Concluding remarks are given in Section 6.7.

6.2 PRELIMINARIES

We are looking to determine the proportion π of people in the population possessing a particular sensitive attribute, where possession of the attribute is binary. We conduct a survey on n subjects of the population by uniform random sampling with replacement.

A single subject's answer $X_i \in \{0, 1\}$ is a randomised version of their truthful answer $x_i \in \{0, 1\}$, in order to protect their privacy. The randomised response will therefore not definitively reveal a subject's truthful answer. By convention, a value of 1 denotes possession of the sensitive attribute, while 0 denotes that the subject does not possess the attribute. We denote by N the number of randomised responses that return 1, hence $N = \sum_{i \in [n]} X_i$. We are therefore looking to estimate π from $\frac{N}{n}$.

In our framework, this set-up is realised using the product sanitisation mechanism introduced in Section 3.5 and which was studied further for categorical data in Section 4.4. In this case, $D = \{0, 1\}$ (i. e. $m = 2$). The parent mechanism is defined as before, where

$$\mathbb{P}(X_i = k | x_i = j) = p_{jk}, \tag{6.1}$$

which allows us to define the design matrix P of the mechanism, as first given in Definition 4.3.

Definition 6.1 (Design Matrix)**.** *A randomised response mechanism as defined in (6.1) is uniquely determined by its design matrix,*

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}.$$

For the probability mass functions of each X_i to sum to 1, we require $p_{00} + p_{01} = 1$ and $p_{10} + p_{11} = 1$. The design matrix therefore simplifies to

$$P = \begin{pmatrix} p_{00} & 1 - p_{00} \\ 1 - p_{11} & p_{11} \end{pmatrix}, \tag{6.2}$$

where $0 \leq p_{00}, p_{11} \leq 1$.

Remark: For some results in this chapter, we impose additional constraints on p_{00} and p_{11}, such as $p_{00}, p_{11} > \frac{1}{2}$. These assumptions are reasonable to make, as they ensure the mechanism is more likely to return a correct answer than an incorrect one.

As π is the proportion of individuals in the population possessing the sensitive attribute, we can calculate the probability mass function of each X_i:

$$\begin{aligned} \mathbb{P}(X_i = 0) &= (1 - \pi)p_{00} + \pi(1 - p_{11}) \\ &= p_{00} - \pi(p_{00} + p_{11} - 1), \end{aligned} \tag{6.3a}$$

$$\begin{aligned} \mathbb{P}(X_i = 1) &= \pi p_{11} + (1 - \pi)(1 - p_{00}) \\ &= 1 - p_{00} + \pi(p_{00} + p_{11} - 1). \end{aligned} \tag{6.3b}$$

Remark: Direct questioning corresponds to the case where $p_{00} = p_{11} = 1$.

6.2.1 *Estimator, Bias and Error*

Having presented the RR mechanism previously, we now need to establish an estimator of π from the parameters of the mechanism, p_{00} and p_{11}, and

from the distribution of randomised responses, namely $\frac{N}{n}$. We first establish a MLE for the mechanism and, later in this sub-section, look at its bias and error.

Theorem 6.1. *Let $p_{00} + p_{11} \neq 1$. Then the Maximum Likelihood Estimator (MLE) for π of the randomised response mechanism defined in (6.2) is given by*

$$\hat{\Pi}(p_{00}, p_{11}) = \frac{p_{00} - 1}{p_{00} + p_{11} - 1} + \frac{N}{(p_{00} + p_{11} - 1)n}. \tag{6.4}$$

Proof. Let us first index the sample so that $X_i = 1$ for each $i \leq N$, and $X_i = 0$ for each $i > N$. Then the likelihood L of the sample is

$$L = \mathbb{P}(X_i = 1)^N \mathbb{P}(X_i = 0)^{n-N}.$$

The log-likelihood is

$$\log(L) = N \log \mathbb{P}(X_i = 1) + (n - N) \log \mathbb{P}(X_i = 0),$$

the maximum of which occurs when $\frac{\partial \log(L)}{\partial \pi} = 0$. Solving for π completes the proof. □

We also note the following standard identity in probability and statistics,

$$\mathrm{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2, \tag{6.5}$$

for any random variable Y. We now calculate the bias and error of $\hat{\Pi}$. We use the variance of the estimator to characterise error in line with conventional practice. Similarly by convention, we characterise the bias of an estimator as its expected deviation from the quantity it is estimating (i. e. $\mathbb{E}[\hat{\Pi} - \pi]$). We remind the reader of the dependence of $\mathrm{Var}(\hat{\Pi})$ on π by writing $\mathrm{Var}(\hat{\Pi}|\pi)$.

Corollary 6.1. *The MLE $\hat{\Pi}$ constructed in Theorem 6.1 is unbiased and has error*

$$\mathrm{Var}(\hat{\Pi}(p_{00}, p_{11})|\pi) = \frac{\frac{1}{4} - \left(p_{00} - \frac{1}{2} - \pi(p_{00} + p_{11} - 1)\right)^2}{(p_{00} + p_{11} - 1)^2 n}. \tag{6.6}$$

Proof. Since the survey we are conducting is by uniform random sampling with replacement, N is a sum of independent and identically distributed random variables. Therefore, $\mathbb{E}[N] = n\mathbb{E}[X_i]$ and $\mathrm{Var}(N) = n\,\mathrm{Var}(X_i)$.

It can be shown that $\mathbb{E}[X_i] = \mathbb{E}[X_i^2] = \mathbb{P}(X_i = 1) = 1 - p_{00} + \pi(p_{00} + p_{11} - 1)$. Hence,

$$\begin{aligned}\mathbb{E}[\hat{\Pi}] &= \frac{p_{00} - 1}{p_{00} + p_{11} - 1} + \frac{\mathbb{E}[N]}{(p_{00} + p_{11} - 1)n} \\ &= \frac{p_{00} - 1}{p_{00} + p_{11} - 1} + \frac{\mathbb{E}[X_i]}{p_{00} + p_{11} - 1} \\ &= \pi,\end{aligned}$$

and so $\hat{\Pi}$ is unbiased as claimed.

Secondly,

$$\begin{aligned}\mathrm{Var}(\hat{\Pi}|\pi) &= \frac{\mathrm{Var}(N)}{(p_{00} + p_{11} - 1)^2 n^2} \\ &= \frac{\mathrm{Var}(X_i)}{(p_{00} + p_{11} - 1)^2 n} \\ &= \frac{\mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2}{(p_{00} + p_{11} - 1)^2 n} \\ &= \frac{\mathbb{P}(X_i = 1)\mathbb{P}(X_i = 0)}{(p_{00} + p_{11} - 1)^2 n},\end{aligned}$$

which can be simplified to (6.6). □

When conducting a survey on a population, it may be necessary to calculate the margin of error of the estimate. For a confidence level $c \in [0, 1]$, the *margin of error* of a sample is given by $\omega \geq 0$, where

$$\mathbb{P}(|\hat{\Pi} - \pi| \leq \omega) \geq c. \tag{6.7a}$$

In practical applications, a 95% confidence interval is typically used [Jaco5]. Using Chebyshev's inequality, we can calculate the margin of error of a sample to be 4.5σ, where the standard deviation σ is given by $\sqrt{\mathrm{Var}(\hat{\Pi}|\pi)}$, since

$$\mathbb{P}\left(|\hat{\Pi} - \pi| \leq 4.5\sqrt{\mathrm{Var}(\hat{\Pi}|\pi)}\right) \geq 0.95. \tag{6.7b}$$

However, given that this only provides a bound on the probability, it is possible for Chebyshev's inequality to exaggerate the margin of error by providing much greater than 95% confidence.

In many practical situations, the central limit theorem is invoked to determine heuristically a margin of error. For a random variable G that is normally distributed with mean μ and variance σ^2, we have

$$\mathbb{P}(|G - \mu| \leq 1.96\sigma) \approx 0.95, \tag{6.7c}$$

hence 1.96σ is typically taken as the margin of error [Jaco5]. However, this non-rigorous approach only gives a loose representation of the margin of error, given that the guarantee of the central limit theorem only applies in the limit as the sample size n approaches infinity.

Due to this variability in defining the margin of error of a sample, we only focus on determining the error $\operatorname{Var}(\hat{\Pi}|\pi)$ of the estimator in this chapter. This error can easily be used to calculate the margin of error for a particular application, as outlined above.

However, the error is frequently a function of an unknown term (e. g. π in the case of RR). Because of that, in this chapter we also examine $\max_\pi \operatorname{Var}(\hat{\Pi}|\pi)$ to allow the margin of error to be determined independently of π.

Corollary 6.2. *For the MLE $\hat{\Pi}$ constructed in Theorem 6.1,*

$$\max_\pi \operatorname{Var}(\hat{\Pi}|\pi) = \frac{1}{4(p_{00} + p_{11} - 1)^2 n}. \tag{6.8}$$

Proof. From (6.6) it can be shown that $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial \pi} = 0$ when $\pi = \frac{2p_{00}-1}{2(p_{00}+p_{11}-1)}$, and since $\frac{\partial^2 \operatorname{Var}(\hat{\Pi}|\pi)}{\partial \pi^2} = -2$, we have

$$\operatorname{Var}(\hat{\Pi}|\pi) \leq \frac{1}{4(p_{00} + p_{11} - 1)^2 n}. \qquad \square$$

6.2.2 *Warner's RR model*

Warner's model [War65] is a specific case of the generalised model introduced above. Warner proposed that surveyors would present respondents with a spinner which they would spin in private to decide which one of two questions to answer. The spinner would point to a question (e. g. "Have you cheated on your partner?") with probability p_w, and to the complement of

that question (e. g. "Have you been faithful to your partner?") with probability $1 - p_w$. Respondents would then be asked to answer the chosen question truthfully, but without revealing which question they were answering.

Warner's model corresponds to the case where $p_{00} = p_{11} = p_w$. We denote by P_w the design matrix of Warner's model, which follows to be

$$P_w = \begin{pmatrix} p_w & 1 - p_w \\ 1 - p_w & p_w \end{pmatrix},$$

while the probability mass function of each X_i is defined as

$$\begin{aligned} \mathbb{P}(X_i = 0) &= p_w - \pi(2p_w - 1), \\ \mathbb{P}(X_i = 1) &= 1 - p_w + \pi(2p_w - 1). \end{aligned}$$

Using the same unbiased MLE in (6.4), we denote by $\hat{\Pi}_w$ the estimator for Warner's model and, by (6.6), find its error to be

$$\operatorname{Var}(\hat{\Pi}_w(p_w)|\pi) = \frac{\frac{1}{4} - \left(p_w - \frac{1}{2} - \pi(2p_w - 1)\right)^2}{(2p_w - 1)^2 n}, \tag{6.9}$$

and furthermore by (6.8),

$$\max_{\pi} \operatorname{Var}(\hat{\Pi}_w|\pi) = \frac{1}{4(2p_w - 1)^2 n}. \tag{6.10}$$

6.2.3 *Mangat's Improved RR Model*

Mangat proposed an improvement to Warner's model and other RR models in an attempt to improve the efficiency of the RR technique [Man94]. Mangat proposed that subjects possessing the sensitive attribute would always answer truthfully, but those without the attribute would provide a randomised response.

In this case, we have $p_{00} = p_m$ and $p_{11} = 1$, and the design matrix P_m is given by

$$P_m = \begin{pmatrix} p_m & 1 - p_m \\ 0 & 1 \end{pmatrix}.$$

Again we calculate the probability mass function of X_i,

$$\begin{aligned}\mathbb{P}(X_i = 0) &= (1-\pi)p_m,\\ \mathbb{P}(X_i = 1) &= 1-(1-\pi)p_m.\end{aligned}$$

We use the unbiased MLE in (6.4), denoted by $\hat{\Pi}_m$ for Mangat's model, and, by (6.6), find its error to be

$$\begin{aligned}\operatorname{Var}(\hat{\Pi}_m|\pi) &= \frac{\frac{1}{4}-\left(p_m(1-\pi)-\frac{1}{2}\right)^2}{p_m^2 n}\\ &= \frac{(1-\pi)(1-p_m(1-\pi))}{p_m n},\end{aligned} \tag{6.11}$$

and, furthermore, by (6.8),

$$\max_{\pi}\operatorname{Var}(\hat{\Pi}_m|\pi) = \frac{1}{4p_m^2 n}. \tag{6.12}$$

6.3 OPTIMAL DIFFERENTIALLY PRIVATE RR MECHANISM

In this section, we study the generalised RR model, from Section 6.2, in the context of (ϵ,δ)-differential privacy, and determine the mechanism which minimises the MLE error. This question was previously examined in [KBR16] for more general error functions, but only for strict ϵ-differential privacy.

Recall that a mechanism $\{X_i|i\in[n]\}$ with a binary output space is (ϵ,δ)-differentially private, for $\epsilon\geq 0$ and $0\leq\delta\leq 1$, if

$$\mathbb{P}(X_i = j) \leq e^{\epsilon}\mathbb{P}(X_k = j)+\delta, \tag{6.13}$$

for any $i,k\in[n]$ and $j\in\{0,1\}$.

For the RR mechanism given by (6.2) to satisfy (ϵ,δ)-differential privacy, we require the following to hold:

$$p_{11} \leq e^{\epsilon}(1-p_{00})+\delta, \tag{6.14a}$$

$$p_{00} \leq e^{\epsilon}(1-p_{11})+\delta, \tag{6.14b}$$

$$1-p_{00} \leq e^{\epsilon}p_{11}+\delta,$$

$$1-p_{11} \leq e^{\epsilon}p_{00}+\delta.$$

Definition 6.2 (Region of Feasibility). *A RR mechanism, given by (6.2), satisfies (ϵ, δ)-differential privacy if $p_{00}, p_{11} \in \mathcal{R}$, where $\mathcal{R} \subset \mathbb{R}^2$ is defined as*

$$\mathcal{R} = \left\{ (p_{00}, p_{11}) \in \mathbb{R}^2 : \begin{array}{l} p_{00}, p_{11} \in [0,1], \\ p_{00} \leq e^{\epsilon}(1 - p_{11}) + \delta, \\ p_{11} \leq e^{\epsilon}(1 - p_{00}) + \delta, \\ 1 - p_{11} \leq e^{\epsilon} p_{00} + \delta, \\ 1 - p_{00} \leq e^{\epsilon} p_{11} + \delta. \end{array} \right\}. \tag{6.15}$$

For simplicity, we restrict our analysis to $p_{00}, p_{11} > \frac{1}{2}$. As mentioned previously, this ensures that correct answers are at least as likely as incorrect ones in any one randomised response, a reasonable condition to impose. The region of feasibility therefore simplifies to $\mathcal{R}'$ as follows:

$$\begin{aligned} \mathcal{R}' &= \left\{ (p_{00}, p_{11}) \in \mathcal{R} : p_{00}, p_{11} > \frac{1}{2} \right\} \\ &= \left\{ (p_{00}, p_{11}) \in \mathbb{R} : \begin{array}{l} p_{00}, p_{11} \in \left(\frac{1}{2}, 1\right], \\ p_{00} \leq e^{\epsilon}(1 - p_{11}) + \delta, \\ p_{11} \leq e^{\epsilon}(1 - p_{00}) + \delta. \end{array} \right\}. \end{aligned}$$

We are looking to find the RR mechanism which minimises estimator error, while still being (ϵ, δ)-differentially private. Hence, we seek to find

$$\underset{(p_{00}, p_{11}) \in \mathcal{R}'}{\arg\min} \ \operatorname{Var}(\hat{\Pi}\mathrm{I}(p_{00}, p_{11}) | \pi). \tag{6.16}$$

Lemma 6.1. *Suppose $p_{00}, p_{11} > \frac{1}{2}$. Then there exists a RR mechanism which satisfies (ϵ, δ)-differential privacy and which has minimal error when $(p_{00}, p_{11}) \in \partial\mathcal{R}'$ and when at least one of the inequalities (6.14) is tight.*

Furthermore, when $0 < \pi < 1$, the mechanism with minimal error is guaranteed to occur on $\partial\mathcal{R}'$ with at least one of the inequalities (6.14) tight.

Proof. Let's consider $\frac{\partial \operatorname{Var}(\hat{\Pi}\mathrm{I}|\pi)}{\partial p_{00}}$ and $\frac{\partial \operatorname{Var}(\hat{\Pi}\mathrm{I}|\pi)}{\partial p_{11}}$.

Firstly,

$$\frac{\partial \operatorname{Var}(\hat{\Pi}\mathrm{I}|\pi)}{\partial p_{11}} = \frac{2p_{00}(p_{00} - 1) - (p_{00} + p_{11} - 1)(2p_{00} - 1)\pi}{(p_{00} + p_{11} - 1)^3 n},$$

and noting that $p_{00} + p_{11} > 1$, $2p_{00} - 1 > 0$ and $2p_{00}(p_{00} - 1) \leq 0$, we see that $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{11}} \leq 0$. Note that $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{11}} < 0$ when $\pi \neq 0$ or when $p_{00} \neq 1$.

Secondly,

$$\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{00}} = \frac{(p_{00} + p_{11} - 1)(1 - \pi + 2p_{11}\pi) - 2p_{00}p_{11}}{(p_{00} + p_{11} - 1)^3 n}.$$

Since $\pi \leq 1$ and $2p_{11} > 1$ we have $1 - \pi + 2p_{11}\pi \leq 2p_{11}$, and since $p_{00} + p_{11} > 1$, we have

$$\begin{aligned}(p_{00} + p_{11} - 1)(1 - \pi + 2p_{11}\pi) - 2p_{00}p_{11} \\ &\leq (p_{00} + p_{11} - 1)(2p_{11}) - 2p_{00}p_{11} \\ &= 2p_{11}(p_{11} - 1) \\ &\leq 0,\end{aligned}$$

hence $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{00}} \leq 0$. Note that $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{00}} < 0$ when $\pi \neq 1$ or when $p_{11} \neq 1$.

Since $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{00}} \leq 0$ and $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{11}} \leq 0$, there exists a mechanism which minimises the estimator error on the boundary of $\mathcal{R}'$, i. e.

$$\partial\mathcal{R}' \cap \left(\underset{(p_{00},p_{11})\in\mathcal{R}'}{\arg\min} \operatorname{Var}(\hat{\Pi}(p_{00}, p_{11})|\pi) \right) \neq \emptyset. \tag{6.17}$$

However, if $0 < \pi < 1$, we see that $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{00}} < 0$ and $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{11}} < 0$. Hence,

$$\underset{(p_{00},p_{11})\in\mathcal{R}'}{\arg\min} \operatorname{Var}(\hat{\Pi}(p_{00}, p_{11})|\pi) \subseteq \partial\mathcal{R}', \tag{6.18}$$

i. e. the optimal mechanisms *only* occur on the boundary of $\mathcal{R}'$.

Finally, suppose $(p_{00}, p_{11}) \in \partial\mathcal{R}'$, but neither inequalities (6.14) are tight. Then there exist $\Delta_0, \Delta_1 \geq 0$, $\Delta_0 + \Delta_1 > 0$ where $(p_{00} + \Delta_0, p_{11} + \Delta_1) \in \partial\mathcal{R}'$, but because $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{00}} \leq 0$ and $\frac{\partial \operatorname{Var}(\hat{\Pi}|\pi)}{\partial p_{11}} \leq 0$, then $\operatorname{Var}(\hat{\Pi}(p_{00}, p_{11})|\pi) \geq \operatorname{Var}(\hat{\Pi}(p_{00} + \Delta_0, p_{11} + \Delta_1)|\pi)$. Hence minimal error is achieved when at least one of the inequalities (6.14) is tight. □

For the remainder of this section, we assume $\pi \in (0, 1)$. Note that the results on optimal mechanisms still hold for $\pi \in [0, 1]$, however these optima may not be unique.

6.3.1 *Optimal Mechanism for ϵ-Differential Privacy*

We have already established that the parameters for the optimal (ϵ, δ)-differentially private mechanism lie on the boundary of $\mathcal{R}'$. We now examine the case of ϵ-differential privacy, where $\delta = 0$, with the additional assumption that $\epsilon > 0$.

Theorem 6.2. *Let $\pi \in (0,1)$, $p_{00}, p_{11} > \frac{1}{2}$, $\epsilon > 0$ and $\delta = 0$. The ϵ-differentially private RR mechanism which minimises estimator error is given by the design matrix*

$$P_\epsilon = \begin{pmatrix} \frac{e^\epsilon}{e^\epsilon+1} & \frac{1}{e^\epsilon+1} \\ \frac{1}{e^\epsilon+1} & \frac{e^\epsilon}{e^\epsilon+1} \end{pmatrix}.$$

Proof. By Lemma 6.1, we know that the parameters p_{00} and p_{11} of the optimal mechanism exist on the boundary of $\mathcal{R}'$, with one of the inequalities (6.14) tight. We now separately consider the cases where (6.14a) and (6.14b) are tight. By hypothesis, $\delta = 0$ and $\epsilon \neq 0$.

1. (6.14a) tight: $p_{11} = e^\epsilon(1 - p_{00})$, constrained by $p_{11} \geq \frac{1}{2}$ and $p_{00} \leq e^\epsilon(1 - p_{11})$. By (6.14b) and since $p_{00} = 1 - e^{-\epsilon} p_{11}$, we have

$$\begin{aligned} e^\epsilon p_{11} &\leq e^\epsilon - p_{00} \\ &= e^\epsilon - (1 - e^{-\epsilon} p_{11}) \\ &= e^\epsilon - 1 + e^{-\epsilon} p_{11}, \end{aligned}$$

   which we rewrite as

$$p_{11}(e^\epsilon - e^{-\epsilon}) \leq e^\epsilon - 1,$$

   and noting that $e^{2\epsilon} - 1 = (e^\epsilon - 1)(e^\epsilon + 1)$, we see that

$$\begin{aligned} p_{11} &\leq \frac{e^\epsilon - 1}{e^{-\epsilon}(e^{2\epsilon} - 1)} \\ &= \frac{e^\epsilon}{e^\epsilon + 1}. \end{aligned}$$

We are therefore considering $\mathrm{Var}(\hat{\Pi}(p_{00}, p_{11})|\pi)$ on the line $p_{00} = 1 - e^{-\epsilon}p_{11}$ for $\frac{1}{2} \leq p_{11} \leq \frac{e^\epsilon}{e^\epsilon+1}$. We parametrise this line as follows, where $0 \leq t \leq 1$, $p_{00} = r(t)$ and $p_{11} = s(t)$:

$$\begin{aligned} r(t) &= \frac{2e^\epsilon - 1}{2e^\epsilon}(1-t) + t\frac{e^\epsilon}{1+e^\epsilon} = 1 - e^{-\epsilon}s(t), \\ s(t) &= \frac{1}{2}(1-t) + t\frac{e^\epsilon}{1+e^\epsilon}. \end{aligned} \tag{6.19}$$

For simplicity, we let $\hat{\Pi}(r(t), s(t)) = \hat{\Pi}_1(t)$. After some manipulation, we see that

$$\frac{\partial \mathrm{Var}(\hat{\Pi}_1(t)|\pi)}{\partial t} = -\frac{2e^\epsilon(1+e^\epsilon)((e^\epsilon-1)\pi+1))}{(e^\epsilon-1)(1-t+e^\epsilon(1+t))^2 n},$$

and noting that $e^\epsilon > 1$, we see that $\frac{\partial \mathrm{Var}(\hat{\Pi}_1(t)|\pi)}{\partial t} < 0$. Hence,

$$\arg\min_{t\in[0,1]} \mathrm{Var}(\hat{\Pi}_1(t)|\pi) = \{1\}. \tag{6.20a}$$

2. (6.14b) tight: By symmetry of the equations (6.14), we simply let $p_{00} = s(t)$ and $p_{11} = r(t)$. By examining (6.3) and (6.6), we see that

$$\mathrm{Var}(\hat{\Pi}(p_{00}, p_{11})|1-\pi) = \mathrm{Var}(\hat{\Pi}(p_{11}, p_{00})|\pi),$$

and by letting $\hat{\Pi}(s(t), r(t)) = \hat{\Pi}_2(t)$, we get

$$\frac{\partial \mathrm{Var}(\hat{\Pi}_2(t)|\pi)}{\partial t} = -\frac{2e^\epsilon(1+e^\epsilon)((e^\epsilon-1)(1-\pi)+1)}{(e^\epsilon-1)(1-t+e^\epsilon(1+t))^2 n}.$$

Again it follows that $\frac{\partial \mathrm{Var}(\hat{\Pi}_2(t)|\pi)}{\partial t} < 0$, and so

$$\arg\min_{t\in[0,1]} \mathrm{Var}(\hat{\Pi}_2(t)|\pi) = \{1\}. \tag{6.20b}$$

By (6.18), (6.20a) and (6.20b), we can now conclude that

$$\arg\min_{(p_{00},p_{11})\in\mathcal{R}'} \mathrm{Var}(\hat{\Pi}(p_{00}, p_{11})|\pi) = \left\{\left(\frac{e^\epsilon}{e^\epsilon+1}, \frac{e^\epsilon}{e^\epsilon+1}\right)\right\},$$

and so the result follows. □

Remark: This mechanism has been seen before in Sections 4.4.3 and 5.4.2, and also matches the optimal mechanism found by [KBR16].

Remark: When $\epsilon = 0$, all rows of the design matrix must be identical, i. e. $p_{00} = 1 - p_{11}$ and $p_{11} = 1 - p_{00}$. This leads to an unbounded estimator error (6.6), due to a zero denominator. In practical terms, 0-differential privacy enforces the same output distribution for every subject, hence nothing meaningful can be learned.

P_ϵ corresponds to the Warner model with $p_w = \frac{e^\epsilon}{e^\epsilon+1}$. The Mangat model can only satisfy ϵ-differential privacy if $p_m = 0$, resulting in an unbounded estimator error.

6.3.2 *Optimal Mechanism for (ϵ, δ)-Differential Privacy*

Let's now consider the optimal mechanism for (ϵ, δ)-differential privacy. As in (6.19) we can parameterise the boundary of $\mathcal{R}'$ where at least one of inequalities (6.14) is tight. If we let

$$\begin{aligned} r_\delta(t) &= \left(1 + e^{-\epsilon}\left(\delta - \frac{1}{2}\right)\right)(1-t) + \frac{e^\epsilon + \delta}{e^\epsilon + 1}t = 1 - e^{-\epsilon}(s_\delta(t) - \delta), \\ s_\delta(t) &= \frac{1}{2}(1-t) + \frac{e^\epsilon + \delta}{e^\epsilon + 1}t, \end{aligned} \tag{6.21}$$

for $t \in [0,1]$, we parametrise the boundary where (6.14a) holds by $p_{00} = r_\delta(t)$ and $p_{11} = s_\delta(t)$. By symmetry, for (6.14b) to hold, we simply set $p_{00} = s_\delta(t)$ and $p_{11} = r_\delta(t)$.

We note that $t = 1$ denotes an extreme point of $\mathcal{R}'$, the point at which both inequalities (6.14) are tight. Here $p_{00} = p_{11} = r_\delta(1) = s_\delta(1) = \frac{e^\epsilon+\delta}{e^\epsilon+1}$. When $\delta \leq \frac{1}{2}$, $t = 0$ also denotes an extreme point of $\mathcal{R}'$, a point where only one of inequalities (6.14) is tight. Here, $p_{ii} = r_\delta(0) = 1 + e^\epsilon(\delta - \frac{1}{2})$ and $p_{1-i,1-i} = s_\delta(0) = \frac{1}{2}$, for $i \in \{0,1\}$.

We now proceed to the following result which states that the minimal variance of $\hat{\Pi}$ on the boundary of $\mathcal{R}'$, where (6.14a) is tight, will occur at an extreme point of $\mathcal{R}'$.

Lemma 6.2. *Let r_δ and s_δ be given by (6.21), let $\epsilon > 0$ and let $a \leq b \in [0,1]$. Then,*

$$\arg\min_{t\in[a,b]} \operatorname{Var}(\hat{\Pi}(r_\delta(t), s_\delta(t))|\pi) \subseteq \{a, b\}.$$

Proof. For simplicity, we denote $\hat{\Pi}(r_\delta(t), s_\delta(t))$ by $\hat{\Pi}_{1,\delta}(t)$.

By some manipulation, it can be shown that the numerator of $\frac{\partial \operatorname{Var}(\hat{\Pi}_{1,\delta}(t)|\pi)}{\partial t}$ is linear in t, hence it has at most one root which occurs at

$$t = \frac{(1+e^\epsilon)\Big(-e^{2\epsilon}\pi + (1-\pi)(1-2\delta)^2 + e^\epsilon(2\pi(1-2\delta)+4\delta-1)\Big)}{(1+(e^\epsilon-1)\pi)(e^\epsilon-1+2\delta)^2}.$$

By substitution, we find that

$$\frac{\partial^2 \operatorname{Var}(\hat{\Pi}_{1,\delta}(t)|\pi)}{\partial t^2} = -\frac{e^{-2\epsilon}(1+(e^\epsilon-1)\pi)^4(e^\epsilon-1+2\delta)^2}{32(1+e^\epsilon)^2\delta^3(e^\epsilon-1+\delta)^3 n},$$

when $\frac{\partial \operatorname{Var}(\hat{\Pi}_{1,\delta}(t)|\pi)}{\partial t} = 0$.

By inspection and since $\epsilon > 0$, we see that $\frac{\partial^2 \operatorname{Var}(\hat{\Pi}_{1,\delta}(t)|\pi)}{\partial t^2} < 0$ when $\frac{\partial \operatorname{Var}(\hat{\Pi}_{1,\delta}(t)|\pi)}{\partial t} = 0$, and so this point is the maximum of $\operatorname{Var}(\hat{\Pi}_{1,\delta}(t)|\pi)$. Hence, the minimum of $\operatorname{Var}(\hat{\Pi}_{1,\delta}(t)|\pi)$ cannot occur at a mid-point of an interval. The result follows. □

We next show the following results for $\pi \leq \frac{1}{2}$ and $\pi \geq \frac{1}{2}$.

Lemma 6.3. *Let r_δ and s_δ be given by (6.21) and let $\pi \leq \frac{1}{2}$. Then for $t \in [0,1]$,*

$$\operatorname{Var}(\hat{\Pi}(r_\delta(t), s_\delta(t))|\pi) \leq \operatorname{Var}(\hat{\Pi}(s_\delta(t), r_\delta(t))|\pi).$$

Conversely, if $\pi \geq \frac{1}{2}$,

$$\operatorname{Var}(\hat{\Pi}(r_\delta(t), s_\delta(t))|\pi) \geq \operatorname{Var}(\hat{\Pi}(s_\delta(t), r_\delta(t))|\pi).$$

Proof. After manipulation of the terms, we can show that

$$\operatorname{Var}(\hat{\Pi}(r_\delta(t), s_\delta(t))|\pi) - \operatorname{Var}(\hat{\Pi}(s_\delta(t), r_\delta(t))|\pi) = -\frac{(e^\epsilon+1)(1-2\pi)(1-t)}{(e^\epsilon(1+t)+1-t)n}.$$

We see that $1-2\pi \geq 0$ when $\pi \leq \frac{1}{2}$, and $1-2\pi \leq 0$ when $\pi \geq \frac{1}{2}$, and since $t \in [0,1]$, the result follows. □

We now present the main result of this chapter which establishes the optimal (ϵ, δ)-differentially private RR mechanism when $\delta \leq \frac{1}{2}$.

Theorem 6.3. *Let $\epsilon > 0$, $\delta \leq \frac{1}{2}$ and $0 < \pi \leq \frac{1}{2}$, and define $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ by*

$$g(\epsilon, \delta) = \frac{(e^\epsilon - 1)(3\delta - 1) + 3\delta^2}{(e^\epsilon - 1 + 2\delta)^2}. \tag{6.22}$$

Then, for r_δ and s_δ given by (6.21),

$$\underset{(p_{00}, p_{11}) \in \partial\mathcal{R}'}{\arg\min} \operatorname{Var}(\hat{\Pi}(p_{00}, p_{11})|\pi) = \begin{cases} \{(r_\delta(0), s_\delta(0))\}, & \text{if } g(\epsilon, \delta) > \pi, \\ \{(r_\delta(1), s_\delta(1))\}, & \text{if } g(\epsilon, \delta) < \pi, \\ \{(r_\delta(0), s_\delta(0)), (r_\delta(1), s_\delta(1))\}, & \text{if } g(\epsilon, \delta) = \pi. \end{cases}$$

Proof. By Lemmas 6.1, 6.2 and 6.3, we know that when $0 < \pi \leq \frac{1}{2}$ and $\delta \leq \frac{1}{2}$,

$$\underset{(p_{00}, p_{11}) \in \partial\mathcal{R}'}{\arg\min} \operatorname{Var}(\hat{\Pi}(p_{00}, p_{11})|\pi) \subseteq \{(r_\delta(0), s_\delta(0)), (r_\delta(1), s_\delta(1))\}.$$

We note that

$$r_\delta(0) = 1 + e^{-\epsilon}\left(\delta - \frac{1}{2}\right), \qquad s_\delta(0) = \frac{1}{2},$$
$$r_\delta(1) = \frac{e^\epsilon + \delta}{e^\epsilon + 1}, \qquad s_\delta(1) = \frac{e^\epsilon + \delta}{e^\epsilon + 1}.$$

We are therefore seeking to determine the sign of

$$\operatorname{Var}\left(\hat{\Pi}\left(1 + e^{-\epsilon}\left(\delta - \frac{1}{2}\right), \frac{1}{2}\right)\middle|\pi\right) - \operatorname{Var}\left(\hat{\Pi}\left(\frac{e^\epsilon + \delta}{e^\epsilon + 1}, \frac{e^\epsilon + \delta}{e^\epsilon + 1}\right)\middle|\pi\right). \tag{6.23}$$

Again, after some manipulation, we can show that (6.23) simplifies to

$$\frac{(e^\epsilon - 1)(1 - 3\delta) - 3\delta^2 + \pi(e^\epsilon - 1 + 2\delta)^2}{(e^\epsilon - 1 + 2\delta)^2 n},$$

and we note that its denominator is strictly positive since $\epsilon > 0$. Hence, $\operatorname{Var}(\hat{\Pi}(r_\delta(0), s_\delta(0))|\pi) = \operatorname{Var}(\hat{\Pi}(r_\delta(1), s_\delta(1))|\pi)$ when $\pi = g(\epsilon, \delta)$.

Similarly, $\operatorname{Var}(\hat{\Pi}(r_\delta(0), s_\delta(0))|\pi) < \operatorname{Var}(\hat{\Pi}(r_\delta(1), s_\delta(1))|\pi)$ when $\pi < g(\epsilon, \delta)$.

Finally, $\operatorname{Var}(\hat{\Pi}(r_\delta(0), s_\delta(0))|\pi) > \operatorname{Var}(\hat{\Pi}(r_\delta(1), s_\delta(1))|\pi)$ when $\pi > g(\epsilon, \delta)$.

□

Remark: When $g(\epsilon,\delta) \leq \pi$, the optimal mechanism corresponds with that established for ϵ-differential privacy on RR (with an added dependence for δ) and also with the optimal mechanism previously established with respect to the max-mean Hamming error (Theorem 4.8, where $m = 2$). However, when $g(\epsilon,\delta) > \pi$, the optimal mechanism is one which we have not encountered previously in this thesis.

The next corollary follows from the symmetry of $\mathrm{Var}(\hat{\Pi}\mathrm{I}(p_{00},p_{11})|\pi)$ in p_{00} and p_{11}.

Corollary 6.3. *Let $\delta \leq \frac{1}{2}$ and $\frac{1}{2} \leq \pi < 1$. Then, for r_δ and s_δ given by (6.21) and g given by (6.22),*

$$\underset{(p_{00},p_{11})\in\partial\mathcal{R}'}{\arg\min}\ \mathrm{Var}(\hat{\Pi}\mathrm{I}(p_{00},p_{11})|\pi) = \begin{cases} \{(s_\delta(0),r_\delta(0))\}, & \text{if } g(\epsilon,\delta) > 1-\pi, \\ \{(s_\delta(1),r_\delta(1))\}, & \text{if } g(\epsilon,\delta) < 1-\pi, \\ \{(s_\delta(0),r_\delta(0)),(s_\delta(1),r_\delta(1))\}, & \text{if } g(\epsilon,\delta) = 1-\pi. \end{cases}$$

Proof. The result follows from Theorem 6.3 since

$$\mathrm{Var}(\hat{\Pi}\mathrm{I}(p_{00},p_{11})|\pi) = \mathrm{Var}(\hat{\Pi}\mathrm{I}(p_{11},p_{00})|1-\pi). \qquad \square$$

Example 6.1 and Figure 6.1 illustrate the conclusion of Theorem 6.3.

Example 6.1. Consider Theorem 6.3 and Corollary 6.3 for various values of ϵ, δ and π. For simplicity, in each of these examples we set $n = 1$.

1. $\epsilon = 0.1$, $\delta = 0$, $\pi = 0.25$: In this case, we have $g(\epsilon,\delta) = -9.508 < \pi$. Hence, the design matrix of the optimal mechanism is denoted by

$$\begin{pmatrix} \frac{e^\epsilon+\delta}{e^\epsilon+1} & \frac{1-\delta}{e^\epsilon+1} \\ \frac{1-\delta}{e^\epsilon+1} & \frac{e^\epsilon+\delta}{e^\epsilon+1} \end{pmatrix}.$$

   In fact for this pair of (ϵ,δ), since $g(\epsilon,\delta) < 0$, this is the optimal mechanism for any $\pi \in [0,1]$.

   This can be verified by noting that $\mathrm{Var}(\hat{\Pi}\mathrm{I}(r_\delta(1),s_\delta(1))|\pi) = 100.104$ and $\mathrm{Var}(\hat{\Pi}\mathrm{I}(r_\delta(0),s_\delta(0))|\pi) = 109.863$.

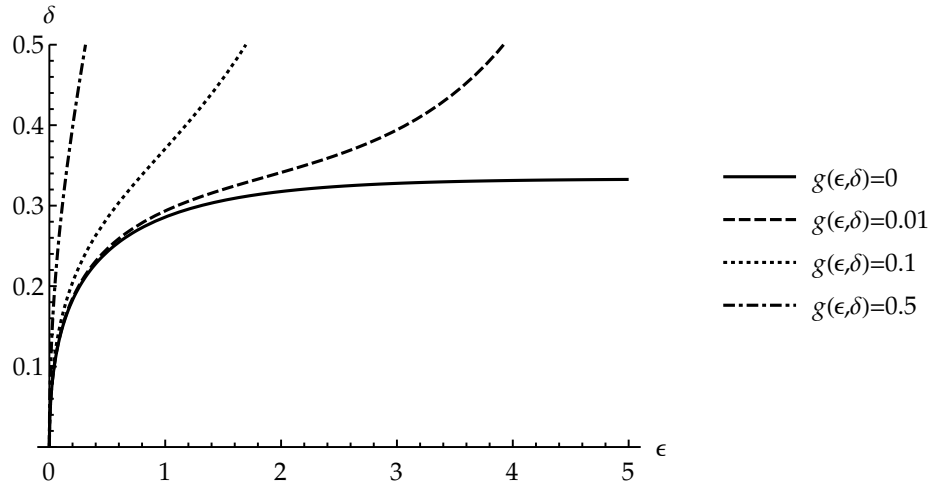

Figure 6.1: A contour plot of various level sets of $g(\epsilon, \delta)$. Given π, ϵ and δ, these level sets can be used to determine the optimal (ϵ, δ)-differentially private RR mechanism.

2. $\epsilon = 1$, $\delta = 0.4$, $\pi = 0.1$: In this case, $g(\epsilon, \delta) = 0.130 > \pi$. Hence, the design matrix of the optimal mechanism is denoted by

$$\begin{pmatrix} 1 + e^{-\epsilon}\left(\delta - \frac{1}{2}\right) & e^{-\epsilon}\left(\frac{1}{2} - \delta\right) \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Again, this can be verified by noting that $\mathrm{Var}(\hat{\Pi}(r_\delta(1), s_\delta(1))|\pi) = 0.385$ and $\mathrm{Var}(\hat{\Pi}(r_\delta(0), s_\delta(0))|\pi) = 0.355$.

3. $\epsilon = 0.5$, $\delta = 0.3$, $\pi = 0.9$: Since $\pi \geq \frac{1}{2}$, we use Corollary 6.3 for this example. We note that $g(\epsilon, \delta) = 0.132 > 1 - \pi$. Hence, the design matrix of the optimal mechanism is denoted by

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ e^{-\epsilon}\left(\frac{1}{2} - \delta\right) & 1 + e^{-\epsilon}\left(\delta - \frac{1}{2}\right) \end{pmatrix}.$$

We see that $\mathrm{Var}(\hat{\Pi}(s_\delta(1), r_\delta(1))|\pi) = 0.965$ and $\mathrm{Var}(\hat{\Pi}(s_\delta(0), r_\delta(0))|\pi) = 0.933$. Note also that $\mathrm{Var}(\hat{\Pi}(r_\delta(0), s_\delta(0))|\pi) = 1.733$, corresponding with the conclusion of Lemma 6.3

6.4 DEGREE OF PRIVACY VIOLATION

We depart from (ϵ, δ)-differential privacy for the remainder of this chapter, and instead look at RR mechanisms in greater generality, with a focus on Warner and Mangat's models. In this section we look to compare the efficiency of the two models.

However, in order to appropriately compare the Warner and the Mangat models, we first need to establish a metric with which to do so. Simply comparing p_w with p_m will not suffice. For example, if $p_w = \frac{1}{2}$, the variance of Warner's model is unbounded, given that the responses are uniformly random, whereas, if $p_m = \frac{1}{2}$, the variance of Mangat's model is bounded, so we can extract meaningful information about the population from the randomised responses.

Instead, we take inspiration from the measure of privacy protection in [Lan76], and define the Degree of Privacy Violation (DPV) to be the probability of a respondent being correctly associated with possessing the sensitive attribute (i. e. by looking at a subject's randomised response, how likely is it that one can correctly guess that the subject possesses the sensitive attribute?).

The Degree of Privacy Violation (DPV) of a mechanism $\{X_i\}$ is given to be α, where

$$\alpha = \max\{\mathbb{P}(x_i = 1|X_i = 1), \mathbb{P}(x_i = 1|X_i = 0)\}, \tag{6.24}$$

hence $\alpha \in [0, 1]$.

We can simplify the DPV further when $p_{00} + p_{11} > 1$.

Theorem 6.4. *Let $p_{00} + p_{11} > 1$, then*

$$DPV = \alpha = \mathbb{P}(x_i = 1|X_i = 1). \tag{6.25}$$

Proof. By Bayes' theorem, we see that

$$\begin{aligned}\mathbb{P}(x_i = 1|X_i = 1) &= \frac{\mathbb{P}(X_i = 1|x_i = 1)\mathbb{P}(x_i = 1)}{\mathbb{P}(X_i = 1)} \\ &= \frac{p_{11}\pi}{1 - p_{00} + \pi(p_{00} + p_{11} - 1)}, \\ \mathbb{P}(x_i = 1|X_i = 0) &= \frac{\mathbb{P}(X_0 = 1|x_i = 1)\mathbb{P}(x_i = 1)}{\mathbb{P}(X_i = 0)} \\ &= \frac{(1 - p_{11})\pi}{p_{00} - \pi(p_{00} + p_{11} - 1)}.\end{aligned}$$

We note that $\mathbb{P}(x_i = 1|X_i = 1) - \mathbb{P}(x_i = 1|X_i = 0) \geq 0$, since the numerator of the difference is

$$\begin{aligned}&p_{11}\pi(p_{00} - \pi(p_{00} + p_{11} - 1)) - (1 - p_{11})\pi(1 - p_{00} + \pi(p_{00} + p_{11} - 1)) \\ &\quad = p_{11}p_{00}\pi - (1 - p_{11})(1 - p_{00})\pi - \pi^2(p_{00} + p_{11} - 1) \\ &\quad = (\pi - \pi^2)(p_{00} + p_{11} - 1) \\ &\quad \geq 0,\end{aligned} \tag{6.26}$$

since $p_{00} + p_{11} > 1$ by hypothesis, and since the denominator is a product of probabilities, $\mathbb{P}(X_i = 0)\mathbb{P}(X_i = 1) > 0$. Hence, $\mathbb{P}(x_i = 1|X_i = 1) \geq \mathbb{P}(x_i = 1|X_i = 0)$. □

The following corollary follows directly from (6.26).

Corollary 6.4. *If* $p_{00} + p_{11} < 1$, *then*

$$DPV = \alpha = \mathbb{P}(x_i = 1|X_i = 0). \tag{6.27}$$

6.4.1 *Warner's Model*

For Warner's model, suppose first that $p_w > \frac{1}{2}$. By Theorem 6.4 we find that

$$\mathbb{P}(x_i = 1|X_i = 1) = \frac{p_w\pi}{1 - p_w + \pi(2p_w - 1)},$$

and to achieve a DPV of α, we require

$$p_w = \frac{\alpha(1 - \pi)}{\alpha(1 - \pi) + \pi(1 - \alpha)}. \tag{6.28}$$

In order to satisfy our assumption that $\frac{1}{2} < p_w \leq 1$, we require $2\alpha(1-\pi) > \alpha(1-\pi) + \pi(1-\alpha)$, which simplifies to $\alpha(1-\pi) > \pi(1-\alpha)$ and is satisfied when

$$\alpha > \pi. \tag{6.29}$$

Note that $p_w \leq 1$, since $\pi(1-\alpha) \geq 0$.

Now, suppose $p_w < \frac{1}{2}$. By Corollary 6.4, we know that

$$DPV = \mathbb{P}(x_i = 1 | X_i = 0),$$

and

$$\mathbb{P}(x_i = 1 | X_i = 0) = \frac{(1-p_w)\pi}{p_w - \pi(2p_w - 1)}.$$

To achieve a DPV of α, we require

$$p_w = \frac{\pi(1-\alpha)}{\alpha(1-\pi) + \pi(1-\alpha)}, \tag{6.30}$$

and to satisfy the assumption that $p_w < \frac{1}{2}$, we additionally require $\alpha > \pi$, as was the case for $p_w > \frac{1}{2}$.

The degree of privacy violation that is incurred using Warner's model is therefore at least the proportion of individuals in the population possessing the sensitive attribute. Hence, the larger the group of people possessing the attribute is, the less protection Warner's model can provide.

Note that from (6.30) we have

$$1 - p_w = \frac{\alpha(1-\pi)}{\alpha(1-\pi) + \pi(1-\alpha)},$$

which corresponds with the value for p_w when $p_w > \frac{1}{2}$ in (6.28). Hence, without loss of generality, for the rest of this section we will assume $p_w > \frac{1}{2}$.

6.4.2 *Mangat's Model*

For Mangat's model with $p_m > 0$, we see that $p_{00} + p_{11} = 1 + p_m > 1$. Hence by Theorem 6.4 its DPV is

$$\alpha = \mathbb{P}(x_i = 1 | X_i = 1) = \frac{\pi}{1 - p_m + \pi p_m},$$

with which we solve for p_m in terms of α,

$$p_m = \frac{\alpha - \pi}{\alpha(1 - \pi)}.$$

We require $0 < p_m \leq 1$, which is achieved when $\alpha > \pi$. Note that $\alpha(1 - \pi) \geq \alpha - \pi$, so $p_m \leq 1$. Similar to Warner's model, the privacy violation of Mangat's model is bounded from below by the proportion of individuals in the population possessing the attribute. In this respect, the two models provide similar protection.

6.4.3 *Error Comparison*

Although the minimum privacy violation incurred by Warner and Mangat's models is identical ($\alpha > \pi$), what can we say about the estimator error incurred when achieving a certain protection? We established the estimator errors for the two models in Sections 6.2.2 and 6.2.3. We now compare these for a fixed DPV α.

Theorem 6.5. *While satisfying a DPV α, Mangat's model is at least as efficient as Warner's model, since*

$$\operatorname{Var}(\hat{\Pi}_w|\pi) \geq \operatorname{Var}(\hat{\Pi}_m|\pi),$$

and furthermore,

$$\max_\pi \operatorname{Var}(\hat{\Pi}_w|\pi) \geq \max_\pi \operatorname{Var}(\hat{\Pi}_m|\pi).$$

Proof. Firstly when satisfying a DPV of α, from (6.9) and after some manipulation, we have

$$\operatorname{Var}(\hat{\Pi}_w|\pi) = \frac{\pi(1 - \pi)(\alpha(1 - \pi) + \pi(\pi - \alpha)}{(\alpha - \pi)^2},$$

and, from (6.11), again after some manipulation, we have

$$\operatorname{Var}(\hat{\Pi}_m|\pi) = \frac{\pi(1 - \pi)^2}{\alpha - \pi}.$$

Therefore,

$$\begin{aligned}\frac{\operatorname{Var}(\hat{\Pi}_w|\pi)}{\operatorname{Var}(\hat{\Pi}_m|\pi)} &= \frac{\pi(1-\pi)(\alpha(1-\pi)+\pi(\pi-\alpha)}{(\alpha-\pi)^2}\cdot\frac{\alpha-\pi}{\pi(1-\pi)^2}\\ &= \frac{\alpha(1-\pi)+\pi(\pi-\alpha)}{(\alpha-\pi)(1-\pi)}\\ &= \frac{(\alpha-\pi)(1-\pi)+\pi(1-\alpha)}{(\alpha-\pi)(1-\pi)}\\ &\geq 1,\end{aligned}$$

and so the Mangat model is more efficient that the Warner model.

Furthermore, from (6.10) and (6.12), we have

$$\begin{aligned}\frac{\max_\pi \operatorname{Var}(\hat{\Pi}_w|\pi)}{\max_\pi \operatorname{Var}(\hat{\Pi}_m|\pi)} &= \left(\frac{p_m}{2p_w-1}\right)^2\\ &= \left(\frac{\frac{\alpha-\pi}{\alpha(1-\pi)}}{\frac{\alpha-\pi}{\alpha(1-\pi)+\pi(1-\alpha)}}\right)^2\\ &= \left(\frac{\alpha(1-\pi)+\pi(1-\alpha)}{\alpha(1-\pi)}\right)^2\\ &= \left(1+\frac{\pi(1-\alpha)}{\alpha(1-\pi)}\right)^2\\ &\geq 1.\end{aligned}$$

□

Remark: At first glance it may seem contradictory for the ratio of the bounded estimator errors to depend on π, when we previously established the bounds to be independent of π. However, the degree of privacy violation is a theoretical, retrospective measure of privacy protection, as demonstrated by (6.29). Without knowing π, we cannot calculate α. Nevertheless, Theorem 6.5 tells us that Mangat's model is more efficient than Warner's, except in the case where $\pi = 0$ or $\alpha = 1$ when they are equally efficient.

As shown in the plot of the ratio of maximum errors, Figure 6.2, the greatest advantage of using Mangat's model is achieved in the high-privacy regime, when the DPV is smallest. The ratio of maximum errors is especially relevant to practical applications, as it allows the margin of error to be determined in independently of π, which is assumed to be unknown.

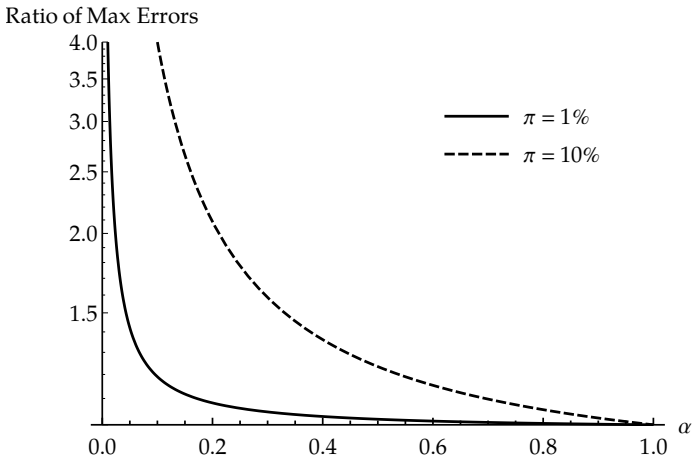

Figure 6.2: Log plot of the ratio of maximum errors $\frac{\max_\pi \mathrm{Var}(\hat{\Pi}_w|\pi)}{\max_\pi \mathrm{Var}(\hat{\Pi}_m|\pi)}$ versus the degree of privacy violation α, for $\pi = 0.01$ and 0.1.

6.5 RANDOMISED RESPONSE WITHOUT SAMPLING

Until now, we have only considered the case where we are sampling from a population to estimate the proportion in the whole population. Instead, what if we were able to survey the entire population using RR?

This approach is particularly relevant to the technique of database sanitisations first discussed in Section 3.2.3.1. Given the randomised responses of the entire database, what is the error of the estimate, given that we are no longer sampling?

6.5.1 *Warner's Model*

Warner noted in [War65] that the error of the estimator $\hat{\Pi}_w$ can be split into two parts: that caused by sampling and that caused by the RR technique. He found that the error due to the RR technique was

$$\frac{1}{n}\left(\frac{1}{4(2p_w-1)^2}-\frac{1}{4}\right).$$

6.5.2 *Mangat's Model*

We now ask what the corresponding error in Mangat's model is due to the RR technique.

Theorem 6.6. *The error of Mangat's model due to the RR technique is given by*

$$\frac{(1-\pi)(1-p_m)}{p_m n}.$$

Proof. We note that the X_i's are independent even when no sampling is involved. We then have

$$\begin{aligned}\mathbb{E}[N] &= \mathbb{E}[\sum X_i] \\ &= \sum_{i:x_i=1} \mathbb{E}[X_i|x_i=1] + \sum_{i:x_i=0} \mathbb{E}[X_i|x_i=0] \\ &= n\pi + n(1-\pi)(1-p_m),\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}(N) &= \mathrm{Var}(\sum X_i) \\ &= \sum_{i:x_i=1} \mathrm{Var}(X_i|x_i=1) + \sum_{i:x_i=0} \mathrm{Var}(X_i|x_i=0) \\ &= 0 + n(1-\pi)\,\mathrm{Var}(X_i|x_i=0).\end{aligned}$$

We also note that

$$\begin{aligned}\mathrm{Var}(X_i|x_i=0) &= \mathbb{E}[X_i^2|x_i=0] - (\mathbb{E}[X_i|x_i=0])^2 \\ &= (1-p_m) - (1-p_m)^2 \\ &= p_m(1-p_m).\end{aligned}$$

Using the same MLE as before, we can show that $\mathbb{E}[\hat{\Pi}] = \pi$ and

$$\begin{aligned}\mathrm{Var}(\hat{\Pi}|\pi) &= \frac{\mathrm{Var}(N)}{p_m^2 n^2} \\ &= \frac{(1-\pi)(1-p_m)}{p_m n}.\end{aligned}$$

□

Remark: From (6.11), we know that the error of Mangat's model when sampling is used is $\frac{(1-\pi)(1-p_m(1-\pi))}{p_m n}$, and it follows that this error is greater than when sampling is not used, since

$$\begin{aligned}\frac{(1-\pi)(1-p_m(1-\pi))}{p_m n} - \frac{(1-\pi)(1-p_m)}{p_m n} &= \frac{\pi(1-\pi)p_m}{p_m n} \\ &\geq 0.\end{aligned}$$

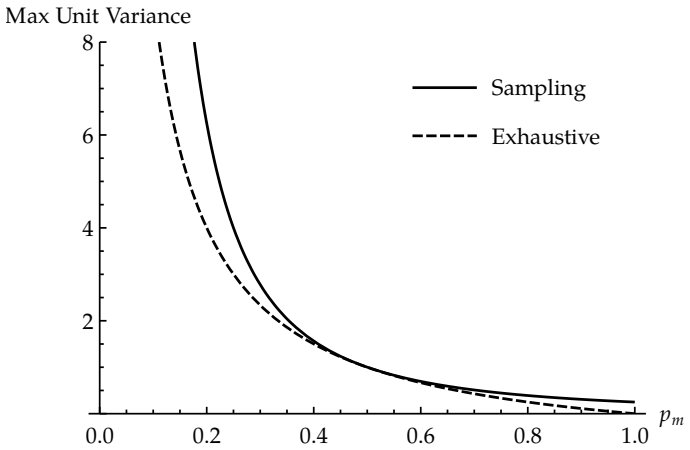

Figure 6.3: An illustration of the maximum unit variance ($n = 1$) for Mangat's RR model when sampling is used (6.8) and when sampling is not used (6.31).

Note: It follows that when there is no sampling,

$$\max_{\pi} \text{Var}(\hat{\text{I}\Pi}|\pi) = \frac{1 - p_m}{p_m n}. \tag{6.31}$$

Remark: By (6.12), we know that the corresponding maximum error for Mangat's model using sampling is $\frac{1}{4p_m^2 n}$, and noting that

$$\begin{aligned} \frac{1}{4p_m^2 n} - \frac{1 - p_m}{p_m n} &= \frac{1}{4p_m^2 n}(1 - 4p_m + 4p_m^2) \\ &= \frac{1}{4p_m^2 n}(2p_m - 1)^2 \\ &\geq 0, \end{aligned}$$

it follows that the estimator error when using RR without sampling is smaller than when sampling is used. However, as can be seen in Figure 6.3, the difference is only profound when p_m is small.

6.6 CATEGORICAL SENSITIVE ATTRIBUTES

We now consider the case where there are multiple categorical sensitive attributes, the proportion in the population of which we seek to learn. In our model, we assume there are m attributes in total, and all but one are sensitive.

One example of this set-up would be a survey asking for people's religion. Assuming that one's religion is a sensitive subject, answers of 'no religion' could be considered non-sensitive, and all other answers considered sensitive.

6.6.1 *Our Model*

We employ a simplified extension of Mangat's model for this problem. As before, subjects with any one of the $(m-1)$ sensitive attributes are asked to respond truthfully, and subjects with the one non-sensitive attribute are asked to give a randomised response. For our model, we randomise uniformly, so every attribute (sensitive and non-sensitive) has equal probability, $\frac{1}{m}$, of being selected as the randomised response.

Definition 6.3 (Super-Binary Mangat Model)**.** *We write the design matrix for the* super-binary Mangat model *as follows:*

$$P = \begin{pmatrix} \frac{1}{m} & \frac{1}{m} & \frac{1}{m} & \cdots & \frac{1}{m} \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}. \tag{6.32}$$

We let $\pi = (\pi_1, \ldots, \pi_m)$ denote the proportion in the population of each of the m attributes, where the attribute denoted by 1 is the non-sensitive attribute. Note that $\sum_i \pi_i = 1$. For each subject $i \in [n]$, the family of variables $\{x_i^1, \ldots, x_i^m\}$ denotes their true attribute, while the family of random variables $\{X_i^1, \ldots, X_i^m\}$ denotes their response, where we denote possession of an attribute as before by $x_i^k = 1$ and $X_i^k = 1$ for $k \in [m]$. Note that the members of the family $\{X_i^1, \ldots, X_i^m\}$ are not independent, since $\sum_k X_i^k = 1$. However, the membersf of different families $\{X_1^k, \ldots, X_n^k\}$ are independent for each $k \in [m]$, since we assume sampling of the population by uniform random sampling with replacement.

Furthermore, we denote by $N_j = \sum_{i \in [n]} X_i^j$ the number of responses possessing attribute j, hence $n = \sum_{j \in [m]} N_j$.

6.6.2 *MLE*

We note from (6.32) that

$$\begin{aligned}\mathbb{P}(X_i^1 = 1) &= \frac{\pi_1}{m},\\ \mathbb{P}(X_i^j = 1) &= \pi_j + \frac{\pi_1}{m}, \qquad\qquad j \in [m] \setminus \{1\}.\end{aligned}$$

We now establish a MLE for each π_j.

Theorem 6.7. *Given a super-binary Mangat model, as defined in Definition 6.3, the MLEs $\hat{\Pi}_j$ for each π_j, $j \geq 2$, solve*

$$\hat{\Pi}_j = \frac{\hat{\Pi}_1}{m} \cdot \frac{N_j(m-1)}{\left(\sum\limits_{k\neq 1,j} \frac{N_k}{m\frac{\hat{\Pi}_k}{\hat{\Pi}_1}+1}\right) + N_1} - \frac{\hat{\Pi}_1}{m}, \tag{6.33a}$$

where $\hat{\Pi}_1$ is given by

$$\hat{\Pi}_1 = 1 - \sum_{k=2}^{m} \hat{\Pi}_k. \tag{6.33b}$$

Proof. We first index the database such that $X_i^1 = 1$ for each $i \in [N_1]$, $X_{N_1+i}^2 = 1$ for each $i \in [N_2]$ and so on (i. e. $X_{N_j+i}^2 = 1$ for each $i \in [N_{j+1}]$ and each $j \in [m-1]$). Then the likelihood is

$$\begin{aligned}L &= \prod_{k=1}^{m} \mathbb{P}(X_i^k = 1)^{N_k}\\ &= \left(\prod_{k=2}^{m}\left(\pi_k + \frac{\pi_1}{m}\right)^{N_k}\right)\left(\frac{\pi_1}{m}\right)^{N_1},\end{aligned}$$

from which we get the log-likelihood,

$$\log L = \sum_{k=2}^{m} N_k \log\left(\pi_k + \frac{\pi_1}{m}\right) + N_1 \log\left(\frac{\pi_1}{m}\right).$$

In order to calculate its partial derivative $\frac{\partial \log L}{\partial \pi_j}$ for $j \in [m] \setminus \{1\}$, we first note that π_1 depends on the other π_j, $\pi_1 = 1 - \sum_{k=2}^{m} \pi_k$. Hence, $\frac{\partial \pi_1}{\partial \pi_j} = -1$. The derivative of the log-likelihood is then

$$\begin{aligned}\frac{\partial \log L}{\partial \pi_j} &= \frac{N_j}{\pi_j + \frac{\pi_1}{m}}\left(1 - \frac{1}{m}\right) + \sum_{k \neq 1,j} \frac{N_k}{\pi_k + \frac{\pi_1}{m}}\left(-\frac{1}{m}\right) + \frac{N_1 m}{\pi_1}\left(-\frac{1}{m}\right) \\ &= \frac{N_j(m-1)}{m\pi_j + \pi_1} - \sum_{k \neq 1,j} \frac{N_k}{m\pi_k + \pi_1} - \frac{N_1}{\pi_1}.\end{aligned}$$

Noting that $\frac{\partial^2 \log L}{\partial \pi_j^2} < 0$ and solving $\frac{\partial \log L}{\partial \pi_j} = 0$ for π_j completes the proof. □

The following result, giving a closed form solution for the MLEs of the model, can be shown by substitution.

Corollary 6.5. *For the super-binary Mangat model, a solution of the system (6.33) is*

$$\begin{aligned}\hat{\Pi}_1 &= \frac{mN_1}{n}, \\ \hat{\Pi}_j &= \frac{N_j - N_1}{n}, \quad j \in [m] \setminus \{1\}.\end{aligned} \tag{6.34}$$

6.6.3 *Estimator Bias and Error*

We now examine the bias and error of the MLEs $\hat{\Pi}_j$. Note that we add π_1 as an argument for the error of $\hat{\Pi}_1$, and we add π_1 and π_j as arguments for the error of $\hat{\Pi}_j$ for each $j \geq 2$. The dependence will become clear in the theorems that follow.

Theorem 6.8. *The estimator $\hat{\Pi}_1$ given in Corollary 6.5 is unbiased and has error*

$$\operatorname{Var}(\hat{\Pi}_1 | \pi_1) = \frac{\pi_1}{n}(m - \pi_1). \tag{6.35}$$

Proof. We note the following for the expected value and variance of $\hat{\Pi}_1$,

$$\begin{aligned}\mathbb{E}[\hat{\Pi}_1] &= \frac{m}{n}\,\mathbb{E}[N_1] \\ &= m\,\mathbb{E}[X_i^1],\end{aligned}$$

and

$$\begin{aligned}\operatorname{Var}(\hat{\Pi}_1|\pi_1) &= \frac{m^2}{n^2}\operatorname{Var}(N_1)\\ &= \frac{m^2}{n}\operatorname{Var}(X_i^1),\end{aligned}$$

since the X_i^1 random variables are independent and identically distributed.

It can be shown that $\mathbb{E}[X_i^1] = \mathbb{E}[(X_i^1)^2] = \frac{\pi_1}{m}$, so by (6.5) we have

$$\begin{aligned}\mathbb{E}[\hat{\Pi}_1] &= \pi_1,\\ \operatorname{Var}(\hat{\Pi}_1|\pi_1) &= \frac{\pi_1}{n}(m-\pi_1),\end{aligned}$$

as claimed. □

We now establish a bound for $\operatorname{Var}(\hat{\Pi}_1|\pi_1)$.

Corollary 6.6. *The estimator $\hat{\Pi}_1$ given in Corollary 6.5 satisfies*

$$\max_{\pi_1}\operatorname{Var}(\hat{\Pi}_1|\pi_1) = \frac{m^2}{4n}. \tag{6.36}$$

Proof. We first note from (6.35) that

$$\frac{\partial\operatorname{Var}(\hat{\Pi}_1|\pi_1)}{\partial\pi_1} = \frac{m-2\pi_1}{n},$$

and so $\frac{\partial\operatorname{Var}(\hat{\Pi}_1|\pi_1)}{\partial\pi_1} = 0$ when $\pi_1 = \frac{m}{2}$. Therefore, since $\frac{\partial^2\operatorname{Var}(\hat{\Pi}_1|\pi_1)}{\partial{\pi_1}^2} = -\frac{2}{n}$, the result follows. □

Note: $\operatorname{Var}(\hat{\Pi}_1|\pi_1)$ and $\max_{\pi_1}\operatorname{Var}(\hat{\Pi}_1|\pi_1)$ increase with m.

Before we examine $\operatorname{Var}(\hat{\Pi}_j|\pi_1,\pi_j)$ for $j \geq 2$, let us first prove the identities in the following lemma.

Lemma 6.4. *The following identities hold for the super-binary Mangat model, where $j \geq 2$,*

$$\begin{aligned}\mathbb{E}[X_i^j - X_i^1] &= \pi_j,\\ \mathbb{E}[(X_i^j - X_i^1)^2] &= \frac{2\pi_1}{m} + \pi_j.\end{aligned}$$

Proof. First, note that

$$\mathbb{P}(X_i^j - X_i^1 = 1) = \mathbb{P}(X_i^j = 1),$$

and

$$\mathbb{P}(X_i^j - X_i^1 = -1) = \mathbb{P}(X_i^1 = 1),$$

since $X_i^j = 1$ only when $X_i^k = 0$ for all $k \neq j \in [m]$.

Hence,

$$\begin{aligned}\mathbb{E}[X_i^j - X_i^1] &= \mathbb{P}(X_i^j - X_i^1 = 1) - \mathbb{P}(X_i^j - X_i^1 = -1) \\ &= \mathbb{P}(X_i^j = 1) - \mathbb{P}(X_i^1 = 1) \\ &= \frac{\pi_1}{m} + \pi_j - \frac{\pi_1}{m} \\ &= \pi_j,\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}[(X_i^j - X_i^1)^2] &= \mathbb{P}(X_i^j - X_i^1 = 1) + \mathbb{P}(X_i^j - X_i^1 = -1) \\ &= \frac{2\pi_1}{m} + \pi_j.\end{aligned}$$

□

We can now consider the error of $\hat{\Pi}_j$ for $j \geq 2$.

Theorem 6.9. *The MLE $\hat{\Pi}_j$, $j \geq 2$ from Corollary 6.5 is unbiased and has error*

$$\operatorname{Var}(\hat{\Pi}_j | \pi_1, \pi_j) = \frac{1}{n}\left(\frac{2\pi_1}{m} + \pi_j(1 - \pi_j)\right). \tag{6.37}$$

Proof. We first note the following identities,

$$\begin{aligned}\mathbb{E}[\hat{\Pi}_j] &= \frac{1}{n}\mathbb{E}[N_j - N_1] \\ &= \mathbb{E}\left[X_i^j - X_i^1\right],\end{aligned}$$

and

$$\begin{aligned}\operatorname{Var}(\hat{\Pi}_j | \pi_1, \pi_j) &= \frac{1}{n^2}\operatorname{Var}(N_j - N_1) \\ &= \frac{1}{n}\operatorname{Var}(X_i^j - X_i^1),\end{aligned}$$

by the independence of the collection $\{X_i^1, \ldots, X_i^m\}$ for each $i \in [n]$.

Making use of (6.5) and substituting for the identities established in Lemma 6.4 completes the proof. □

As before, we now determine a bound for $\mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_1, \pi_j)$.

Corollary 6.7. *The MLE* $\hat{\mathrm{I}}\hat{\mathrm{I}}_j$, $j \geq 2$ *from Corollary 6.5 satisfies*

$$\max_{\pi_1, \pi_j} \mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_1, \pi_j) = \frac{1}{n}\left(\frac{1}{2} + \frac{1}{m}\right)^2. \tag{6.38}$$

Proof. For simplicity, we first rewrite π_1 as

$$\begin{aligned}\pi_1 &= 1 - \pi_j - \sum_{k \neq 1, j} \pi_k \\ &= 1 - \pi_j - \pi^*,\end{aligned}$$

giving us the following representation of the error of $\hat{\mathrm{I}}\hat{\mathrm{I}}_j$ as $\mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_j, \pi^*)$ from (6.37),

$$\mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_j, \pi^*) = \frac{1}{n}\left(\frac{2}{m}(1 - \pi_j - \pi^*) + \pi_j(1 - \pi_j)\right).$$

Note that

$$\frac{\partial \mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_j, \pi^*)}{\partial \pi^*} = \frac{1}{n}\left(-\frac{2}{m}\right),$$

so $\frac{\partial \mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_j, \pi^*)}{\partial \pi^*} < 0$ and, hence, $\mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_j, \pi^*)$ achieves its maximum at $\pi^* = 0$. Furthermore, we see that

$$\frac{\partial \mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_j, \pi^*)}{\partial \pi_j} = \frac{1}{n}\left(-\frac{2}{m} + 1 - 2\pi_j\right),$$

hence $\frac{\partial \mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_j, \pi^*)}{\partial \pi_j} = 0$ at $\pi_j = \frac{1}{2} - \frac{1}{m}$ and, since $\frac{\partial^2 \mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_j, \pi^*)}{\partial {\pi_j}^2} = -\frac{2}{n} < 0$, it is a maximum. Therefore,

$$\begin{aligned}\mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_j, \pi^*) &\leq \mathrm{Var}\left(\hat{\mathrm{I}}\hat{\mathrm{I}}_j \middle| \frac{1}{2} - \frac{1}{m}, 0\right) \\ &= \frac{1}{n}\left(\frac{1}{2} + \frac{1}{m}\right)^2.\end{aligned}$$

□

Note: $\mathrm{Var}(\hat{\mathrm{I}}\hat{\mathrm{I}}_j|\pi_1, \pi_j)$ *decreases* with increasing m.

6.6.4 *Interpreting the Results*

As noted before, the behaviour of $\hat{\Pi}_1$ and $\hat{\Pi}_j$, for $j \geq 2$, varies with m. While $\text{Var}(\hat{\Pi}_1)$ increases as m increases, $\text{Var}(\hat{\Pi}_j)$ decreases. This may seem counter-intuitive at first, as adding more sensitive options decreases the error of the estimate for the sensitive values. To see why this happens, observe that for $\hat{\Pi}_1$, as m increases for a fixed n, there are more options for subjects possessing the non-sensitive attribute to be assigned to by the randomised response method, thereby increasing the variance of N_1 and increasing the error of the estimator. However, for $\hat{\Pi}_j$, $j \geq 2$, increasing m while keeping n fixed results in each sensitive attribute receiving fewer randomised responses from those with the non-sensitive attribute, because the randomisation is uniform. The noise being added to the count of each sensitive attributes is therefore reduced, resulting in a smaller estimator error.

It is important to note that this reduced estimator error for $\hat{\Pi}_j$ comes at the cost of reduced privacy protection for the subjects. With more options for those with the non-sensitive attribute to be randomly assigned, there is a greater likelihood of a subject possessing a sensitive attribute having their privacy violated.

6.7 CONCLUDING REMARKS

The main goal of this chapter was to apply (ϵ, δ)-differential privacy to RR and determine the optimal mechanism that minimises estimator error, which was achieved in Theorem 6.3 for $\delta \leq \frac{1}{2}$. Other contributions include:

- A comparison of the efficiency of the Warner and Mangat RR techniques using the DPV metric (Theorem 6.5);
- The error implications of applying Mangat's model to database sanitisations, when error due to sampling can be eliminated (Section 6.5.2);
- The extension of Mangat's model to non-binary questioning, and the derivation of a closed-form MLE for estimating the proportion of each distinct attribute in the population (Section 6.6).

7

CONCLUSION

We complete this thesis with some brief concluding remarks and present directions for possible future research arising from the problems considered thus far.

OVERVIEW

7.1 SUMMARY

The purpose of this thesis was to conduct a mathematical examination of various aspects of differential privacy. In keeping with that purpose, the primary goals were to (i) develop an abstract mathematical framework for differential privacy, (ii) investigate the privacy/utility tradeoff associated with differential privacy, and (iii) provide for a practical implementation of differential privacy.

We first set about formulating a single, unifying framework for differential privacy, and this allowed us to prove results in a general setting without having to specify the type of data, query or perturbation. Some of those results, such as those relating to the identity query (Theorem 3.4) and the product sanitisation mechanism (Theorem 3.5), opened further avenues for research. For example, many results that we presented have been proven to hold for the identity query, and consequently hold for all other queries too.

We also focused on various error functions on differentially private mechanisms. Our initial focus was on the max-mean error, and the max-mean Hamming error, which represents the worst-case average error we can ex-

pect from such a mechanism. By examining the extreme points of the polytope of locally private mechanisms, we presented results relating to linear error functions. Utility was also considered in the case of Randomised Response (RR).

By examining RR, we also gave a differentially private implementation of a privacy-preserving technique that is already widely used in practice. We presented results that allow for the optimal use of differentially private RR mechanisms, although this analysis was limited to mechanisms defined by diagonally dominant matrices for ease of analysis. We also extended our investigation of RR beyond differential privacy, to include the relative efficiency of the Warner and Mangat RR models. Finally, we extended the Mangat model to non-binary questioning and derived closed-form MLEs.

7.2 CONTRIBUTIONS

The work of this thesis addresses a number of shortcomings in the literature on differential privacy thus far. To the best of our knowledge, our framework is the first that unifies all of the types of data, queries and perturbation methods considered in the literature. This allows for seemingly different implementations of differential privacy to be investigated uniformly.

In the thesis we gave a demonstration of the framework's versatility: (i) data types considered included numerical, categorical and functional; (ii) points in the data lifecycle at which perturbation was applied included local privacy, dataset sanitisation and output perturbation; (iii) perturbation methods included noise addition and value swapping; and (iv) the only limitation on queries was that they are required to be measurable. This adaptability will hopefully be useful in the future to researchers looking to implement differentially private mechanisms for new types of data or perturbation that have not yet been considered in the literature.

We also introduced new techniques for checking mechanisms for differential privacy. Results were presented for sufficient sets for general mechanisms (Theorem 3.3), and also with specialisation to mechanisms on categorical data (Section 4.3). This additionally allowed us to concisely formulate

necessary and sufficient conditions for differential privacy on categorical data, and describe resulting error bounds.

Although the geometry of differential privacy has been studied in the past, we presented new results on the polytope of local ϵ-differential privacy. In particular, we presented unexpected results relating to the extreme points of the polytope, namely that loose entries can occur. This means that the optimal mechanism does not always have to be tight.

One significant contribution was the formulation of the optimal RR mechanism for (ϵ, δ)-differential privacy. We found some unexpected results, including that the optimal mechanism is not guaranteed to correspond with the optimal mechanism for local differential privacy in general.

Beyond differential privacy, our results on the extended Mangat model may prove useful in applying RR to non-binary questioning.

7.3 FUTURE WORK

A number of avenues for further research naturally arise from the results in this thesis.

- Unanswered questions remain on the question of extreme points of the local differential privacy polytope. We do not yet have a complete characterisation for extreme points with loose elements for ϵ-differential privacy. Nor have we investigated the polytope for local (ϵ, δ)-differential privacy. Studying these topics further may give added insight into optimal mechanisms and into the geometry of differential privacy in general.

- In the case of differential privacy in RR, we have presented the optimal mechanism for (ϵ, δ)-differential privacy when $\delta \leq \frac{1}{2}$. Conditions to determine the optimal mechanism for $\delta > \frac{1}{2}$ have yet to be determined.

- Our analysis of RR models only considers Warner and Mangat's models. Using the same mathematical techniques, other models could similarly be analysed and a comparison made on their relative efficiencies.

- The Super-Binary Mangat model prescribed that those not in the sensitive group would randomise their response uniformly at random. It

may be that uniformly randomising their responses does not optimise utility. This question requires some investigation. Also worth investigating is the case where there is more than one non-sensitive attribute.

Nevertheless, it is argued that the work in this thesis provides a firm mathematical foundation upon which future differential privacy techniques, and privacy techniques in general, can be constructed, investigated and implemented.

A

APPENDIX: MATLAB CODE FOR EXTREME POINT ENUMERATION

The following *MATLAB* code is used to enumerate all extreme points of the local (ϵ, δ)-differential privacy polytope. Parameters include (i) the dimension m of the $m \times m$ polytope, (ii) the number c of non-zero columns, (iii) privacy parameters ϵ and δ, and (iv) the number of constraints which we wish to enforce to improve efficiency.

Extreme points are enumerated as vectors in $\mathbb{R}^{m(m-1)}$ using `vtxenum.m`. The function `DPconstraints` populates all (ϵ, δ)-differential privacy constraints in the appropriate vector format as per (5.3c). The vector format of extreme points can be converted to the usual matrix format using `vert2stocmtx`.

The code runs on *MATLAB* version R2015a and makes use of the *Toolbox Manager*, details of which can be found at `www.tbxmanager.com`, with the following packages enabled:

1. `cddmex` (Version 1.0.1);
2. `lcp` (Version 1.0.3);
3. `mpt` (Version 3.1).

A.1 MAIN PROGRAMME

`vtxenum.m` is the main programme for enumerating the extreme points of the local (ϵ, δ)-differential privacy polytope. It makes use of `DPconstraints.m` which follows in Section A.2.

Listing 1: vtxenum.m

```
%% Set-up variables for programme

m = 5; % Dimension of the matrix, m x m
c = 4; % Max number of non-zero columns in the matrix, 1 < c <= m
p = 2; % p = exp(eps)
d = 0; % delta

[A,b] = DPconstraints(m, c, p, d);

% Pick m random constraints to enforce, or enforce all r
% (m takes precedence)
n = 4;
r = [];

if n > 0 || size(r,2) > 0
    if n > 0
        r = ceil(rand(1,n)*size(A,1));
    else
        n = size(r,2);
    end

    Ae = zeros(n, size(A,2));
    be = zeros(n,1);

    for i=1:n
        Ae(i,1:size(A,2)) = A(r(i),1:size(A,2));
        be(i) = b(r(i));
    end
    clear i;
else
    Ae = zeros(1, size(A,2));
    be = zeros(1,1);
end

% Start Toolbox Manager if not already started
if (exist('Polyhedron') == 0)
    startup;
end

% Initialise the polyhedron
PQ = Polyhedron('A', A, 'b', b, 'Ae', Ae, 'be', be);

```

```
43 %% Evaluate polytope for vertices
44
45 if PQ.isEmptySet() > 0
46     disp('Polyhedron is empty');
47 else
48     % Evaluate vertex-representation of polyhedron
49     tic;
50     vert = PQ.V;
51     toc;
52
53     % Add zero columns to vert for standard form
54     if c < m
55         vert(1:size(vert,1),size(vert,2)+1:m*(m-1)) = zeros(size(
               vert,1),m*(m-1)-size(vert,2));
56     end
57
58     % Display results to terminal
59     disp(['Found ', num2str(size(vert,1)), ' vertices satisfying
           ', num2str(n), ' constraints.']);
60 end
```

A.2 DIFFERENTIAL PRIVACY CONSTRAINTS

The following code defines the function `DPconstraints` which populates the (ϵ, δ)-differential privacy constraints for use in `vtxenum.m`.

Listing 2: DPconstraints.m

```
1  function [A,b] = DPconstraints( m, c, p, d )
2      %DPconstraints Returns a matrix of all differential privacy
           constraints
3      %   A and right-hand-side b for an m x m design matrix, where
4      %   exp(epsilon)=p and delta=d.
5
6      % Set defaults
7      if nargin == 1
8          c = m;
9      end
10     if nargin <= 2
```

```
11          p = 2;
12      end
13      if nargin <= 3
14          d = 0;
15      end
16
17      % Do some error-checking
18      if m < 1
19          error('Dimension n of the matrix must be at least 1');
20      elseif c > m
21          warning('Number of non-zero columns c must be at most n (
                resetting c = n)');
22          c = m;
23      end
24      if p < 1
25          warning('Privacy parameter p must be at least 1 (
                resetting p = 1)');
26          p = 1;
27      end
28      if d < 0 || d > 1
29          warning('Privacy parameter d must lie on the unit
                interval (resetting d = 0)');
30          d = 0;
31      end
32
33      A = 0;
34
35      % Add constraints for first column
36      for i=1:m
37          A(1+(i-1)*(m-1):i*(m-1),i) = ones(m-1,1);
38
39          for j=1:m
40              if j<i
41                  A((i-1)*(m-1)+j,j) = -p;
42              elseif j>i
43                  A((i-1)*(m-1)+j-1,j) = -p;
44              end
45          end
46      end
47
48      A2=A;
49
```

```
50      % Add constraints for middle columns
51      if c>2
52          for k=2:c-1
53              A(1+m*(m-1)*(k-1):m*(m-1)*k,1+m*(k-1):m*k)=A2;
54          end
55      end
56
57      % Add constraints for final column
58      for k=1:c-1
59          A(1+m*(m-1)*(c-1):(m-1)*m*c,1+m*(k-1):m*k)=-A2;
60      end
61
62      if d > 0
63          A((m-1)*m*c+1:(m-1)*m*(c+1),:) = -eye(m*(m-1));
64
65          for i=1:m-1
66              A((m-1)*m*(c+1)+1:(m-1)*m*(c+1)+m,1+(i-1)*m:i*m)=eye(
                    m);
67          end
68      end
69
70      % Initialise RHS b
71      b = d*ones((m-1)*m*c,1);
72      b(1+m*(m-1)*(c-1):(m-1)*m*c) = (p-1+d)*ones((m-1)*m*c-m*(m-1)
            *(c-1),1);
73
74      if d > 0
75          b((m-1)*m*c+1:(m-1)*m*(c+1)) = zeros(m*(m-1),1);
76          b((m-1)*m*(c+1)+1:(m-1)*m*(c+1)+m) = ones(m,1);
77      end
78
79  end
```

A.3 VECTOR FORM TO MATRIX FORM

The `vert2stocmtx` function is used to convert an extreme point in vector form to a stochastic matrix for ease of examination.

Listing 3: vert2stocmtx.m

```
function A = vert2stocmtx( vert, m )
%vert2stocmtx Returns a stochastic m x m matrix from vector vert
    listing
%   (m-1) colums

    % Find the dimension of our matrix if not given
    if nargin == 1
        m=round((1+sqrt(1+4*max(size(vert))))/2);
    end

    A=zeros(m);
    for j=1:m-1
        A(1:m,j)=vert(((j-1)*m+1):(j*m))';
    end
    A(1:m,m) = ones(m,1)-sum(A,2);

end
```

BIBLIOGRAPHY

[AAC+11] Alvim, M. S., Andrés, M. E., Chatzikokolakis, K., Degano, P., and Palamidessi, C. Differential privacy: On the trade-off between utility and information leakage. In *Formal Aspects of Security and Trust*. Springer, 2011, pp. 39–54. (Cited on Page 25)

[ABCP13] Andrés, M. E., Bordenabe, N. E., Chatzikokolakis, K., and Palamidessi, C. Geo-indistinguishability: Differential privacy for location-based systems. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security* (New York, NY, USA, 2013), CCS '13, ACM, pp. 901–914. (Cited on Page 28)

[AEGH67] Abul-Ela, A.-L. A., Greenberg, B. G., and Horvitz, D. G. A multi-proportions randomized response model. *Journal of the American Statistical Association 62*, 319 (1967), 990–1008. (Cited on Page 30)

[And07] Anderson, N. Netflix offers streaming movies to subscribers. *Ars Technica*, http://arstechnica.com/uncategorized/2007/01/8627/ [Accessed 2016-05-18], Jan 2007. (Cited on Page 13)

[App16] Apple Inc. Apple previews iOS 10, the biggest iOS release ever. *Apple Newsroom*, http://www.apple.com/newsroom/2016/06/apple-previews-ios-10-biggest-ios-release-ever.html [Accessed: 2016-07-26], June 2016. (Cited on Page 28)

[Aus06] Australian Burean of Statistics. 2006 census: Census through the ages. http://www.abs.gov.au/websitedbs/D3310114.nsf/4a256353001af3ed4b2562bb00121564/

eadaffffb171cab6ca257161000a78d7!OpenDocument [Accessed: 2016-04-13], 2006. (Cited on Page 7)

[AW89] Adam, N. R., and Worthmann, J. C. Security-control methods for statistical databases: A comparative study. *ACM Comput. Surv. 21*, 4 (Dec. 1989), 515–556. (Cited on Page 18)

[Ban14] Banisar, D. National comprehensive data protection/privacy laws and bills 2014 map. *Privacy Laws and Bills* (2014). (Cited on Page 10)

[Bar02] Barvinok, A. *A course in convexity*, vol. 54. American Mathematical Society Providence, 2002. (Cited on Page 85)

[BCD+07] Barak, B., Chaudhuri, K., Dwork, C., Kale, S., McSherry, F., and Talwar, K. Privacy, accuracy, and consistency too: a holistic solution to contingency table release. In *Proceedings of the twenty-sixth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems* (2007), ACM, pp. 273–282. (Cited on Page 53)

[BD99] Banisar, D., and Davies, S. G. Global trends in privacy protection: An international survey of privacy, data protection, and surveillance laws and developments. *John Marshall Journal of Computer & Information Law 18*, 1 (1999). (Cited on Page 9)

[BDK07] Backstrom, L., Dwork, C., and Kleinberg, J. Wherefore art thou R3579X?: Anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of the 16th International Conference on World Wide Web* (New York, NY, USA, 2007), WWW '07, ACM, pp. 181–190. (Cited on Page 21)

[BDMN05] Blum, A., Dwork, C., McSherry, F., and Nissim, K. Practical privacy: The SuLQ framework. In *Proceedings of the Twenty-fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (New York, NY, USA, 2005), PODS '05, ACM, pp. 128–138. (Cited on Pages 22 and 23)

[Bil95] Billingsley, P. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. Wiley New York, 1995. (Cited on Pages 34 and 37)

[BIZ15] Blair, G., Imai, K., and Zhou, Y.-Y. Design and analysis of the randomized response technique. *Journal of the American Statistical Association 110*, 511 (2015), 1304–1319. (Cited on Page 30)

[BKOZB12] Barthe, G., Köpf, B., Olmedo, F., and Zanella Béguelin, S. Probabilistic relational reasoning for differential privacy. *SIGPLAN Not. 47*, 1 (Jan. 2012), 97–110. (Cited on Page 27)

[BMS14] Bambauer, J., Muralidhar, K., and Sarathy, R. Fool's gold: An illustrated critique of differential privacy. *Vanderbilt Journal of Entertainment and Technology Law 16* (2013-2014), 701. (Cited on Page 28)

[Bor71] Boruch, R. F. Assuring confidentiality of responses in social research: A note on strategies. *The American Sociologist 6*, 4 (1971), 308–311. (Cited on Page 30)

[Bos15] Bose, M. Respondent privacy and estimation efficiency in randomized response surveys for discrete-valued sensitive variables. *Statistical Papers 56*, 4 (2015), 1055–1069. (Cited on Page 30)

[BS08] Brickell, J., and Shmatikov, V. The cost of privacy: Destruction of data-mining utility in anonymized data publishing. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2008), KDD '08, ACM, pp. 70–78. (Cited on Page 17)

[BV04] Boyd, S., and Vandenberghe, L. *Convex optimization*. Cambridge University Press, 2004. (Cited on Page 85)

[BZ06] Barbaro, M., and Zeller, T. J. A face is exposed for AOL searcher no. 4417749. *New York Times* (Aug 2006). (Cited on Page 13)

[CC14] Cavoukian, A., and Castro, D. Big data and innovation, setting the record straight: De-identification does work. *White Paper, Jun* (2014). (Cited on Pages 20 and 21)

[CDJ+14] Chen, X., Du, Q., Jin, Z., Xu, T., Shi, J., and Gao, G. The randomized response technique application in the survey of homosexual commercial sex among men in Beijing. *Iran J Public Health 43*, 4 (Apr 2014), 416–422. 26005651[pmid]. (Cited on Page 31)

[CJ11] Coutts, E., and Jann, B. Sensitive questions in online surveys: Experimental results for the randomized response technique (RRT) and the unmatched count technique (UCT). *Sociological Methods & Research 40*, 1 (2011), 169–193. (Cited on Page 30)

[CK12] Chin, A., and Klinefelter, A. Differential privacy as a response to the reidentification threat: The Facebook advertiser case study. *North Carolina Law Review 90*, 5 (2012). (Cited on Page 28)

[CMF+11] Chen, R., Mohammed, N., Fung, B. C., Desai, B. C., and Xiong, L. Publishing set-valued data via differential privacy. *Proceedings of the VLDB Endowment 4*, 11 (2011), 1087–1098. (Cited on Pages 27 and 61)

[Coo79] Cooley, T. Law of torts, 1879. (Cited on Page 8)

[Cou50] Council of Europe. European convention for the protection of human rights and fundamental freedoms, as amended by protocols nos. 11 and 14, November 1950. (Cited on Page 9)

[Cou95] Council of Europe and European Parliament. Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data, October 1995. (Cited on Page 10)

[CPS+12] Cormode, G., Procopiuc, C., Srivastava, D., Shen, E., and Yu, T. Differentially private spatial decompositions. In *Data*

Engineering (ICDE), 2012 IEEE 28th International Conference on (2012), IEEE, pp. 20–31. (Cited on Page 53)

[CSS12] Chaudhuri, K., Sarwate, A., and Sinha, K. Near-optimal differentially private principal components. In *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 989–997. (Cited on Page 28)

[Dal77] Dalenius, T. Towards a methodology for statistical disclosure control. *statistik Tidskrift 15* (1977), 429–444. (Cited on Pages 11 and 17)

[Dal86] Dalenius, T. Finding a needle in a haystack. *Journal of Official Statistics 2*, 3 (1986), 329–336. (Cited on Page 17)

[DDH03] Donovan, J. J., Dwight, S. A., and Hurtz, G. M. An assessment of the prevalence, severity, and verifiability of entry-level applicant faking using the randomized response technique. *Human Performance 16*, 1 (2003), 81–106. (Cited on Page 31)

[DEE12] Dankar, F. K., and El Emam, K. The application of differential privacy to health data. In *Proceedings of the 2012 Joint EDBT/ICDT Workshops* (New York, NY, USA, 2012), EDBT-ICDT '12, ACM, pp. 158–166. (Cited on Page 29)

[DFSC15] Domingo-Ferrer, J., and Soria-Comas, J. From t-closeness to differential privacy and vice versa in data anonymization. *Knowledge-Based Systems 74* (2015), 151–158. (Cited on Page 26)

[DFWBJ15] Domingo-Ferrer, J., Wu, Q., and Blanco-Justicia, A. Flexible and robust privacy-preserving implicit authentication. In *ICT Systems Security and Privacy Protection*. Springer, 2015, pp. 18–34. (Cited on Page 15)

[DJW13] Duchi, J. C., Jordan, M., and Wainwright, M. J. Local privacy and statistical minimax rates. In *Foundations of Com-*

puter Science (FOCS), 2013 IEEE 54th Annual Symposium on (Oct 2013), pp. 429–438. (Cited on Page 29)

[DKM+06] Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., and Naor, M. Our data, ourselves: Privacy via distributed noise generation. In *Advances in Cryptology-EUROCRYPT 2006*. Springer, 2006, pp. 486–503. (Cited on Pages 22 and 23)

[dMHVB13] de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., and Blondel, V. D. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports 3*, 1376 (2013). (Cited on Page 21)

[DMNS06] Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography*. Springer, 2006, pp. 265–284. (Cited on Page 22)

[dMRSP15] de Montjoye, Y.-A., Radaelli, L., Singh, V. K., and Pentland, A. S. Unique in the shopping mall: On the reidentifiability of credit card metadata. *Science 347*, 6221 (2015), 536–539. (Cited on Page 21)

[DN03] Dinur, I., and Nissim, K. Revealing information while preserving privacy. In *Proceedings of the Twenty-second ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (New York, NY, USA, 2003), PODS '03, ACM, pp. 202–210. (Cited on Page 22)

[DN04] Dwork, C., and Nissim, K. Privacy-preserving datamining on vertically partitioned databases. In *Advances in Cryptology–CRYPTO 2004* (2004), Springer, pp. 528–544. (Cited on Page 22)

[DR14] Dwork, C., and Roth, A. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science 9*, 3–4 (2014), 211–407. (Cited on Page 25)

[DSF+13] Dietz, P., Striegel, H., Franke, A. G., Lieb, K., Simon, P., and Ulrich, R. Randomized response estimates for the 12-

month prevalence of cognitive-enhancing drug use in university students. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy 33*, 1 (2013), 44–50. (Cited on Page 31)

[Dwo06] Dwork, C. Differential privacy. In *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006, pp. 1–12. (Cited on Pages 17, 23, 28, 38, 39, 40, and 51)

[Dwo08] Dwork, C. Differential privacy: A survey of results. In *Theory and Applications of Models of Computation*. Springer, 2008, pp. 1–19. (Cited on Pages 43, 52, and 61)

[Dwo11] Dwork, C. A firm foundation for private data analysis. *Communications of the ACM 54*, 1 (2011), 86–95. (Cited on Page 53)

[EPK14] Erlingsson, U., Pihur, V., and Korolova, A. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (New York, NY, USA, 2014), CCS '14, ACM, pp. 1054–1067. (Cited on Page 31)

[Eur16] European Commission. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union 59*, L 119 (4 May 2016), 1–88. (Cited on Page 10)

[FL88] Finkelhor, D., and Lewis, I. A. An epidemiologic approach to the study of child molestationa. *Annals of the New York Academy of Sciences 528*, 1 (1988), 64–78. (Cited on Page 31)

[FS10] Friedman, A., and Schuster, A. Data mining with differential privacy. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*

(New York, NY, USA, 2010), KDD '10, ACM, pp. 493–502. (Cited on Page 27)

[FWCY10] Fung, B. C. M., Wang, K., Chen, R., and Yu, P. S. Privacy-preserving data publishing: A survey of recent developments. *ACM Comput. Surv. 42*, 4 (June 2010), 14:1–14:53. (Cited on Pages 11 and 20)

[GAESH69] Greenberg, B. G., Abul-Ela, A.-L. A., Simmons, W. R., and Horvitz, D. G. The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association 64*, 326 (1969), 520–539. (Cited on Page 30)

[GG75] Goodstadt, M. S., and Gruson, V. The randomized response technique: A test on drug use. *Journal of the American Statistical Association 70*, 352 (1975), 814–818. (Cited on Page 30)

[Gin10] Gingerich, D. W. Understanding off-the-books politics: Conducting inference on the determinants of sensitive behavior with randomized response surveys. *Political Analysis 18*, 3 (2010), 349–380. (Cited on Page 31)

[GJAH71] Greenberg, B. G., Jr., R. R. K., Abernathy, J. R., and Horvitz, D. G. Application of the randomized response technique in obtaining quantitative data. *Journal of the American Statistical Association 66*, 334 (1971), 243–250. (Cited on Page 30)

[GKOV15] Geng, Q., Kairouz, P., Oh, S., and Viswanath, P. The staircase mechanism in differential privacy. *Selected Topics in Signal Processing, IEEE Journal of 9*, 7 (2015), 1176–1184. (Cited on Page 26)

[GKS08] Ganta, S. R., Kasiviswanathan, S. P., and Smith, A. Composition attacks and auxiliary information in data privacy. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining* (2008), ACM, pp. 265–273. (Cited on Page 25)

[Golo6] Golle, P. Revisiting the uniqueness of simple demographics in the US population. In *Proceedings of the 5th ACM Workshop on Privacy in Electronic Society* (New York, NY, USA, 2006), WPES '06, ACM, pp. 77–80. (Cited on Page 12)

[Greo7] Greely, H. T. The uneasy ethical and legal underpinnings of large-scale genomic biobanks. *Annual Review of Genomics and Human Genetics 8*, 1 (2007), 343–364. PMID: 17550341. (Cited on Page 12)

[Gre16a] Green, M. What is differential privacy? *A Few Thoughts on Cryptographic Engineering*, http://blog.cryptographyengineering.com/2016/06/what-is-differential-privacy.html [Accessed: 2016-07-26], June 2016. (Cited on Page 28)

[Gre16b] Greenberg, A. Apple's 'differential privacy' is about collecting your data – but not your data. *Wired.com*, https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/ [Accessed: 2016-07-26], June 2016. (Cited on Page 28)

[GRS12] Ghosh, A., Roughgarden, T., and Sundararajan, M. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing 41*, 6 (2012), 1673–1693. (Cited on Page 25)

[GSo7a] Gjestvang, C. R., and Singh, S. Forced quantitative randomized response model: a new device. *Metrika 66*, 2 (2007), 243–257. (Cited on Page 30)

[GSo7b] Guerriero, M., and Sandri, M. F. A note on the comparison of some randomized response procedures. *Journal of Statistical Planning and Inference 137*, 7 (2007), 2184 – 2190. (Cited on Page 30)

[GV13] Geng, Q., and Viswanath, P. The optimal mechanism in (ϵ,δ)-differential privacy. *CoRR abs/1305.1330* (2013). (Cited on Page 26)

[GV14] Geng, Q., and Viswanath, P. The optimal mechanism in differential privacy. In *Information Theory (ISIT), 2014 IEEE International Symposium on* (2014), IEEE, pp. 2371–2375. (Cited on Pages 26 and 28)

[Hau98] Haughey v. Moriarty. IESC 17, 1998. (Cited on Page 9)

[Hig11] Higginbotham, S. For science, big data is the microscope of the 21st century. *Gigaom*, https://gigaom.com/2011/11/08/for-science-big-data-is-the-microscope-of-the-21st-century/ [Accessed: 2016-08-02], Nov 2011. (Cited on Page 7)

[HK12] Huang, Z., and Kannan, S. The exponential mechanism for social welfare: Private, truthful, and nearly optimal. In *Foundations of Computer Science (FOCS), 2012 IEEE 53rd Annual Symposium on* (Oct 2012), pp. 140–149. (Cited on Page 61)

[HL13] Heffetz, O., and Ligett, K. Privacy and data-based research. Working Paper 19433, National Bureau of Economic Research, September 2013. (Cited on Page 13)

[HLMJ09] Hay, M., Li, C., Miklau, G., and Jensen, D. Accurate estimation of the degree distribution of private networks. In *2009 Ninth IEEE International Conference on Data Mining* (Dec 2009), pp. 169–178. (Cited on Page 27)

[HRW11] Hall, R., Rinaldo, A., and Wasserman, L. Random differential privacy. *arXiv preprint arXiv:1112.2680* (2011). (Cited on Page 26)

[HRW13] Hall, R., Rinaldo, A., and Wasserman, L. Differential privacy for functions and functional data. *The Journal of Machine Learning Research 14*, 1 (2013), 703–727. (Cited on Page 42)

[HT10] Hardt, M., and Talwar, K. On the geometry of differential privacy. In *Proceedings of the forty-second ACM symposium on Theory of computing* (2010), ACM, pp. 705–714. (Cited on Pages 25 and 55)

[Ire37] IRELAND, CONSTITUTION OF. *Bunreacht na hÉireann*. Stationery Office, 1937. (Cited on Page 9)

[Jac05] JACKMAN, S. Pooling the polls over an election campaign. *Australian Journal of Political Science 40*, 4 (2005), 499–517. (Cited on Pages 111 and 112)

[Kar06] KARNITSCHNIG, M. AOL tech chief resigns over issue of released data. *Wall Street Journal* (Aug 2006). (Cited on Page 13)

[KBR16] KAIROUZ, P., BONAWITZ, K., AND RAMAGE, D. Discrete distribution estimation under local privacy. In *Proceedings of The 33rd International Conference on Machine Learning* (2016), pp. 2436–2444. (Cited on Pages 114 and 119)

[KLN+11] KASIVISWANATHAN, S. P., LEE, H. K., NISSIM, K., RASKHODNIKOVA, S., AND SMITH, A. What can we learn privately? *SIAM Journal on Computing 40*, 3 (2011), 793–826. (Cited on Page 25)

[KM11] KIFER, D., AND MACHANAVAJJHALA, A. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data* (2011), ACM, pp. 193–204. (Cited on Page 25)

[KM14] KIFER, D., AND MACHANAVAJJHALA, A. Pufferfish: A framework for mathematical privacy definitions. *ACM Trans. Database Syst. 39*, 1 (Jan. 2014), 3:1–3:36. (Cited on Page 26)

[Kor10] KOROLOVA, A. Privacy violations using microtargeted ads: A case study. In *2010 IEEE International Conference on Data Mining Workshops* (Dec 2010), IEEE, pp. 474–482. (Cited on Pages 14 and 28)

[KOV14] KAIROUZ, P., OH, S., AND VISWANATH, P. Extremal mechanisms for local differential privacy. In *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2879–2887. (Cited on Pages 26, 29, and 82)

[KRSY11] Karwa, V., Raskhodnikova, S., Smith, A., and Yaroslavtsev, G. Private analysis of graph structure. *Proceedings of the VLDB Endowment 4*, 11 (2011), 1146–1157. (Cited on Page 27)

[Kru12] Krumpal, I. Estimating the prevalence of xenophobia and anti-Semitism in Germany: A comparison of randomized response and direct questioning. *Social Science Research 41*, 6 (2012), 1387 – 1403. (Cited on Page 30)

[Kru13] Krumpal, I. Determinants of social desirability bias in sensitive surveys: a literature review. *Quality & Quantity 47*, 4 (2013), 2025–2047. (Cited on Page 30)

[KS08] Kasiviswanathan, S. P., and Smith, A. A note on differential privacy: Defining resistance to arbitrary side information. *CoRR abs/0803.3946* (2008). (Cited on Pages 23, 25, 38, and 40)

[KS12] Karwa, V., and Slavković, A. B. Differentially private graphical degree sequences and synthetic graphs. In *Privacy in Statistical Databases: UNESCO Chair in Data Privacy, International Conference, PSD 2012, Palermo, Italy, September 26-28, 2012. Proceedings* (Berlin, Heidelberg, 2012), Springer Berlin Heidelberg, pp. 273–285. (Cited on Page 27)

[KSS16] Kalantari, K., Sankar, L., and Sarwate, A. Optimal differential privacy mechanisms under Hamming distortion for structured source classes. In *2016 IEEE International Symposium on Information Theory (ISIT)* (July 2016), pp. 2069–2073. (Cited on Page 26)

[Lam93] Lambert, D. Measures of disclosure risk and harm. *Journal of Official Statistics 9*, 2 (1993), 313–331. (Cited on Page 17)

[Lan76] Lanke, J. On the degree of protection in randomized interviews. *International Statistical Review / Revue Internationale de Statistique 44*, 2 (1976), 197–203. (Cited on Pages 30 and 124)

[LC11] Lee, J., and Clifton, C. How much is enough? choosing ε for differential privacy. In *Information Security*. Springer, 2011, pp. 325–340. (Cited on Page 26)

[LHG97] LARKINS, E. R., HUME, E. C., AND GARCHA, B. S. The validity of the randomized response method in tax ethics research. *Journal of Applied Business Research 13*, 3 (1997), 25–32. (Cited on Page 30)

[LLV07] LI, N., LI, T., AND VENKATASUBRAMANIAN, S. *t*-closeness: Privacy beyond *k*-anonymity and ℓ-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on* (2007), IEEE, pp. 106–115. (Cited on Page 22)

[Loh09] LOHR, S. Netflix awards $1 million prize and starts a new contest. *New York Times* (September 2009). (Cited on Page 14)

[Loh10] LOHR, S. Netflix cancels contest plans and settles suit. *New York Times* (March 2010). (Cited on Page 14)

[Loy76] LOYNES, R. M. Asymptotically optimal randomized response procedures. *Journal of the American Statistical Association 71*, 356 (1976), 924–928. (Cited on Page 30)

[LQS12] LI, N., QARDAJI, W., AND SU, D. On sampling, anonymization, and differential privacy or, *k*-anonymization meets differential privacy. In *Proceedings of the 7th ACM Symposium on Information, Computer and Communications Security* (2012), ACM, pp. 32–33. (Cited on Page 26)

[LSOE04] LARA, D., STRICKLER, J., OLAVARRIETA, C. D., AND ELLERTSON, C. Measuring induced abortion in Mexico: A comparison of four methodologies. *Sociological Methods & Research 32*, 4 (2004), 529–558. (Cited on Page 30)

[LW76] LEYSIEFFER, F. W., AND WARNER, S. L. Respondent jeopardy and optimal designs in randomized response models. *Journal of the American Statistical Association 71*, 355 (1976), 649–656. (Cited on Page 30)

[Man94] MANGAT, N. S. An improved randomized response strategy. *Journal of the Royal Statistical Society. Series B (Methodological) 56*, 1 (1994), 93–95. (Cited on Pages 30, 108, and 113)

[MCB+11] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. Big data: The next frontier for innovation, competition, and productivity. *McKinsey & Company*, http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation [Accessed: 2016-05-23], 2011. (Cited on Page 8)

[MCFY11] Mohammed, N., Chen, R., Fung, B. C., and Yu, P. S. Differentially private data release for data mining. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (New York, NY, USA, 2011), KDD '11, ACM, pp. 493–501. (Cited on Page 27)

[McG74] McGee v. Attorney General. IR 284, 1974. (Cited on Page 9)

[MD15] Mansfield-Devine, S. The Ashley Madison affair. *Network Security 2015*, 9 (2015), 8–16. (Cited on Page 11)

[Meh11] Mehta, A. Big data: Powering the next industrial revolution. *Tableau Software White Paper* (2011). (Cited on Page 16)

[MKGV07] Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. ℓ-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD) 1*, 1 (2007), 3. (Cited on Page 22)

[MM12] Moshagen, M., and Musch, J. Surveying multiple sensitive attributes using an extension of the randomized-response technique. *International Journal of Public Opinion Research 24*, 4 (2012), 508–523. (Cited on Page 30)

[Moo71] Moors, J. J. A. Optimization of the unrelated question randomized response model. *Journal of the American Statistical Association 66*, 335 (1971), 627–629. (Cited on Page 30)

[Moo97] Moors, J. J. A. A critical evaluation of Mangat's two-step procedure in randomized response. *CentER Discussion Paper 1997* (1997). (Cited on Page 30)

[MPRV09] Mironov, I., Pandey, O., Reingold, O., and Vadhan, S. Computational differential privacy. In *Advances in Cryptology - CRYPTO 2009: 29th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 16-20, 2009. Proceedings*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009, pp. 126–142. (Cited on Page 26)

[MSC13] Mayer-Schönberger, V., and Cukier, K. *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013. (Cited on Pages 7 and 16)

[MT07] McSherry, F., and Talwar, K. Mechanism design via differential privacy. In *Foundations of Computer Science, 2007. FOCS'07. 48th Annual IEEE Symposium on* (2007), IEEE, pp. 94–103. (Cited on Pages 25, 38, 61, 63, and 72)

[MW10] Masiello, B., and Whitten, A. Engineering privacy in an age of information abundance. In *AAAI Spring Symposium: Intelligent Information Privacy Management* (2010). (Cited on Page 18)

[Net09] Netflix. Netflix prize rules. http://www.netflixprize.com/rules [Accessed: 2016-05-23], 2009. (Cited on Page 13)

[NF14] Narayanan, A., and Felten, E. W. No silver bullet: De-identification still doesn't work. *White Paper* (2014). (Cited on Page 21)

[Nor84] Norris v. Attorney General. IR 36, 1984. (Cited on Page 9)

[NRS07] Nissim, K., Raskhodnikova, S., and Smith, A. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing* (New York, NY, USA, 2007), STOC '07, ACM, pp. 75–84. (Cited on Pages 26 and 27)

[NS06] Narayanan, A., and Shmatikov, V. How to break anonymity of the Netflix prize dataset. *CoRR abs/cs/0610105* (2006). (Cited on Page 14)

[NS08] Narayanan, A., and Shmatikov, V. Robust de-anonymization of large sparse datasets. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on* (2008), IEEE, pp. 111–125. (Cited on Page 14)

[NSM90] N. S. Mangat, R. S. An alternative randomized response procedure. *Biometrika 77*, 2 (1990), 439–442. (Cited on Page 30)

[NST12] Nissim, K., Smorodinsky, R., and Tennenholtz, M. Approximately optimal mechanism design via differential privacy. In *Proceedings of the 3rd innovations in theoretical computer science conference* (2012), ACM, pp. 203–213. (Cited on Page 61)

[NTZ13] Nikolov, A., Talwar, K., and Zhang, L. The geometry of differential privacy: the sparse and approximate cases. In *Proceedings of the forty-fifth annual ACM symposium on Theory of computing* (2013), ACM, pp. 351–360. (Cited on Page 25)

[OEC80] OECD Organisation for Economic Cooperation and Development. Guidelines governing the protection of privacy and transborder flow of personal data, September 1980. (Cited on Page 9)

[OEC13] OECD Organisation for Economic Cooperation and Development. The OECD privacy framework, 2013. (Cited on Page 10)

[Ohm10] Ohm, P. Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review 57*, 6 (2010), 1701–1777. (Cited on Page 13)

[Pan14] Pandurangan, V. On taxis and rainbows: Lessons from NYC's improperly anonymized taxi logs. https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1 [Accessed: 2016-05-11], June 2014. (Cited on Page 15)

[Par67] Parthasarathy, K. *Probability measures on metric spaces*, vol. 352. American Mathematical Soc., 1967. 2005 Reprint. (Cited on Page 42)

[PF13] Provost, F., and Fawcett, T. Data science and its relationship to big data and data-driven decision making. *Big Data 1*, 1 (2013), 51–59. (Cited on Page 8)

[RHMS09] Rastogi, V., Hay, M., Miklau, G., and Suciu, D. Relationship privacy: Output perturbation for queries with joins. In *Proceedings of the Twenty-eighth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems* (New York, NY, USA, 2009), PODS '09, ACM, pp. 107–116. (Cited on Page 27)

[RN10] Rastogi, V., and Nath, S. Differentially private aggregation of distributed time-series with transformation and encryption. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data* (New York, NY, USA, 2010), SIGMOD '10, ACM, pp. 735–746. (Cited on Page 28)

[Rud87] Rudin, W. *Real and Complex Analysis*, third edition ed. Mathematics Series. McGraw-Hill International Editions, 1987. (Cited on Pages 34 and 35)

[Rud07] Rudman, P. *How Mathematics Happened: The First 50,000 Years*. Prometheus Books, 2007. (Cited on Page 7)

[SCR+11] Shi, E., Chan, T. H., Rieffel, E., Chow, R., and Song, D. Privacy-preserving aggregation of time-series data. In *Proc. NDSS* (2011), vol. 2, pp. 1–17. (Cited on Page 28)

[Shi00] Shimanek, A. E. Do you want milk with those cookies: Complying with the safe harbor privacy principles. *J. Corp. L. 26* (2000), 455. (Cited on Page 10)

[SMDF16] Sánchez, D., Martínez, S., and Domingo-Ferrer, J. Comment on "Unique in the shopping mall: On the reidentifiability of credit card metadata". *Science 351*, 6279 (2016), 1274–1274. (Cited on Page 21)

[SS14] Sarwate, A. D., and Sankar, L. A rate-disortion perspective on local differential privacy. In *Communication, Control, and Computing (Allerton), 2014 52nd Annual Allerton Conference on* (Sept 2014), pp. 903–908. (Cited on Page 29)

[SUS10] Striegel, H., Ulrich, R., and Simon, P. Randomized response estimates for doping and illicit drug use in elite athletes. *Drug and Alcohol Dependence 106*, 2–3 (2010), 230 – 232. (Cited on Page 31)

[Swe00] Sweeney, L. Uniqueness of simple demographics in the US population. Tech. rep., Carnegie Mellon University, 2000. (Cited on Page 12)

[Swe02] Sweeney, L. *k*-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10*, 05 (2002), 557–570. (Cited on Pages 12 and 22)

[Swe05] Sweeney, L. Recommendations to identify and combat privacy problems in the Commonwealth, October 2005. Testimony before the Pennsylvania House Select Committee on Information Security (House Resolution 351), Pittsburgh, PA. (Cited on Page 12)

[SZW+11] Sala, A., Zhao, X., Wilson, C., Zheng, H., and Zhao, B. Y. Sharing graphs using differentially private graph models. In *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference* (New York, NY, USA, 2011), IMC '11, ACM, pp. 81–98. (Cited on Page 27)

[TC12] Task, C., and Clifton, C. A guide to differential privacy theory in social network analysis. In *Proceedings of the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012)* (Washington, DC, USA, 2012), ASONAM '12, IEEE Computer Society, pp. 411–417. (Cited on Page 27)

[TF81] Tracy, P. E., and Fox, J. A. The validity of randomized response for sensitive measurements. *American Sociological Review 46*, 2 (1981), 187–200. (Cited on Page 30)

[Tia14] TIAN, G.-L. A new non-randomized response model: The parallel model. *Statistica Neerlandica 68*, 4 (2014), 293–323. (Cited on Page 30)

[TMKC14] TEMPL, M., MEINDL, B., KOWARIK, A., AND CHEN, S. Introduction to statistical disclosure control (SDC). *IHSN Working Paper No. 007*, 2014. (Cited on Page 17)

[Toc14] TOCKAR, A. Riding with the stars: Passenger privacy in the NYC taxicab dataset. *Neustar Research*, https://research.neustar.biz/2014/09/15/riding-with-the-stars/ [Accessed: 2016-05-11], September 2014. (Cited on Page 15)

[TP12] TENE, O., AND POLONETSKY, J. Privacy in the age of big data: a time for big decisions. *Stanford Law Review Online 64* (2012), 63. (Cited on Page 18)

[Tro14] TROTTER, J. Public NYC taxicab database lets you see how celebrities tip. *Gawker.com*, http://gawker.com/1646724546 [Accessed: 2016-05-11], October 2014. (Cited on Page 15)

[UN 48] UN GENERAL ASSEMBLY. Universal declaration of human rights, December 1948. (Cited on Page 9)

[UN 66] UN GENERAL ASSEMBLY. International covenant on civil and political rights, December 1966. (Cited on Page 9)

[vdHvGBH00] VAN DER HEIJDEN, P. G. M., VAN GILS, G., BOUTS, J., AND HOX, J. J. A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning: Eliciting sensitive information in the context of welfare and unemployment benefit. *Sociological Methods & Research 28*, 4 (2000), 505–537. (Cited on Page 30)

[War65] WARNER, S. L. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association 60*, 309 (1965), 63–69. (Cited on Pages 29, 107, 112, and 129)

[WB90] Warren, S. D., and Brandeis, L. D. The right to privacy. *Harvard law review* (1890), 193–220. (Cited on Page 8)

[WB13] Ward, J. S., and Barker, A. Undefined by data: A survey of big data definitions. *CoRR abs/1309.5821* (2013). (Cited on Page 7)

[WDW12] Willenborg, L., and De Waal, T. *Elements of statistical disclosure control*, vol. 155. Springer Science & Business Media, 2012. (Cited on Page 18)

[Who14a] Whong, C. FOILing NYC's *boro* taxi trip data. http://chriswhong.com/open-data/foiling-nycs-boro-taxi-trip-data/ [Accessed: 2016-05-11], December 2014. (Cited on Page 15)

[Who14b] Whong, C. FOILing NYC's taxi trip data. http://chriswhong.com/open-data/foil_nyc_taxi/ [Accessed: 2016-05-11], March 2014. (Cited on Page 14)

[WN16] Waseda, A., and Nojima, R. Analyzing randomized response mechanisms under differential privacy. In *International Conference on Information Security* (2016), Springer, pp. 271–282. (Cited on Page 29)

[Wor06] Working Group on Privacy. Report of working group on privacy, March 2006. (Cited on Page 9)

[WP13] Wolter, F., and Preisendörfer, P. Asking sensitive questions: An evaluation of the randomized response technique versus direct questioning using individual validation data. *Sociological Methods & Research 42*, 3 (2013), 321–353. (Cited on Page 30)

[WS94] Williams, B. L., and Suen, H. A methodological comparison of survey techniques in obtaining self-reports of condom-related behaviors. *Psychological Reports 75*, 3 suppl (1994), 1531–1537. (Cited on Page 30)

[Wu13] Wu, F. T. Defining privacy and utility in data sets. *U. Colo. L. Rev. 84* (2013), 1117. (Cited on Page 18)

[WWH15] Wang, Y., Wu, X., and Hu, D. Using randomized response for differential privacy preserving data collection. *N/A* (2015). (Cited on Page 29)

[WZ10] Wasserman, L., and Zhou, S. A statistical framework for differential privacy. *Journal of the American Statistical Association 105*, 489 (2010), 375–389. (Cited on Page 61)

[Zel06] Zeller, T. J. AOL executive quits after posting of search data. *New York Times* (Aug 2006). (Cited on Page 13)

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

http://code.google.com/p/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Final Version as of February 13, 2017 (`classicthesis` version 1.1).