



Doctoral Thesis in Electrical Engineering

# Pointwise Maximal Leakage: Robust, Flexible and Explainable Privacy

SARA SAEIDIAN

# Pointwise Maximal Leakage: Robust, Flexible and Explainable Privacy

SARA SAEIDIAN

Academic Dissertation which, with due permission of the KTH Royal Institute of Technology, is submitted for public defence for the Degree of Doctor of Philosophy on Friday the 9th February 2024, at 2:00 p.m. in D3, Lindstedtsvägen 9, Stockholm.

Doctoral Thesis in Electrical Engineering  
KTH Royal Institute of Technology  
Stockholm, Sweden 2024

© Sara Saeidian

ISBN 978-91-8040-804-2

TRITA-EECS-AVL-2024:7

Printed by: Universitetservice US-AB, Sweden 2024

---

## Abstract

For several decades now, safeguarding sensitive information from disclosure has been a key focus in computer science and information theory. Especially, in the past two decades, the subject of *privacy* has received significant attention due to the widespread collection and processing of data in various facets of society. A central question in this area is “What can be inferred about individuals from the data collected from them?”

This doctoral thesis delves into a foundational and application-agnostic exploration of the theory of privacy. The overarching objective is to construct a comprehensive framework for evaluating and designing privacy-preserving data processing systems that adhere to three essential criteria:

- *Explainability.* The notion of information leakage (or privacy loss) employed in this framework should be operationally meaningful. That is, it should naturally emerge from the analysis of adversarial attack scenarios. Privacy guarantees within this framework should be comprehensible to stakeholders and the associated privacy parameters should be meaningful and interpretable.
- *Robustness.* The notion of information leakage employed should demonstrate resilience against a diverse array of potential adversaries, accommodating a broad range of attack scenarios while refraining from making restrictive assumptions about adversarial capabilities.
- *Flexibility.* The framework should offer value in a variety of contexts, catering to both highly privacy-sensitive applications and those with more relaxed privacy requirements. The notion of information leakage employed should also be applicable to various data types.

The privacy notion proposed in this thesis that aligns with all the above criteria is called *pointwise maximal leakage* (PML). PML is a random variable that measures the amount of information leaking about a secret random variable  $X$  to a publicly available related random variable  $Y$ . We first develop PML for finite random variables by studying two seemingly different but mathematically equivalent adversarial setups: the *randomized function model* and the *gain function model*. We then extend the gain function model to random variables on arbitrary probability spaces to obtain a more general form of PML. Furthermore, we study the properties of PML in terms of pre and post-processing inequalities and composition, define various privacy guarantees, and compare PML with existing privacy notions from the literature including differential privacy and its local variant.

PML, by definition, is an *inferential* privacy measure in the sense that it compares an adversary’s posterior knowledge about  $X$  with her prior knowledge. However, a prevalent misconception in the area suggests that meaningful inferential privacy guarantees are unattainable, due to an over-interpretation of a result called the *impossibility of absolute disclosure prevention*. Through a pivotal shift in perspective, we characterize precisely the types of disclosures

---

that can be prevented through privacy guarantees and those that remain inevitable. In this way, we argue in favor of inferential privacy measures.

On the more application-oriented front, we examine a common machine learning framework for privacy-preserving learning called Private Aggregation of Teacher Ensembles (PATE) using an information-theoretic privacy measure. Specifically, we propose a conditional form of the notion of *maximal leakage* to quantify the amount of information leaking about individual data entries and prove that the leakage is Schur-concave when the injected noise has a log-concave probability density. The Schur-concavity of the leakage implies that increased classification accuracy improves privacy. We also derive upper bounds on the information leakage when the injected noise has Laplace distribution.

Finally, we design optimal privacy mechanisms that minimize Hamming distortion subject to maximal leakage constraints assuming that (i) the data-generating distribution (i.e., the prior) is known, or (ii) the prior belongs to a certain set of possible distributions. We prove that sets of priors that contain more “uniform” distributions generate larger distortion. We also prove that privacy mechanisms that distribute the privacy budget more uniformly over the outcomes create smaller worst-case distortion.

**Keywords:** Privacy, information leakage, pointwise maximal leakage, disclosure prevention, inferential privacy, mechanism design.

---

## Sammanfattning

Att skydda känslig information mot oavsiktligt avslöjande har varit ett viktigt forskningsmål inom datavetenskap och informationsteori under de senaste decennierna. I synnerhet under de senaste två decennierna har ämnet dataintegritet fått stor uppmärksamhet, inte minst på grund av den omfattande datainsamlingen som pågår i stora delar av samhället. En central fråga inom området är "Vilka slutsatser kan dras om individer från de data som samlas in från dem?"

Denna avhandling fördjupar sig i teorin bakom dataintegritet från ett fundamentalt och tillämpningsoberoende perspektiv. Det övergripande målet är att skapa ett allsidigt ramverk för att designa och utvärdera dataintegritetsbevarande databehandlingssystem som följer tre essentiella kriterier:

- *Förklarbarhet.* Definitionen av informationsläckage (eller minskningen av dataintegritet) i detta ramverk bör ha en operationell betydelse, det vill säga att definitionen uppkommer naturligt från en analys av potentiella fientliga attacker. Dataintegritetsgarantier inom detta ramverk bör också vara förstäneliga för intressenter, och motsvarande dataintegritetsparametrar bör vara meningsfulla och tolkningsbara.
- *Robusthet.* Definitionen av informationsläckage bör uppvisa motståndskraft mot en mångfald av potentiella fientliga attacker: definitionen bör vara tillämpbar på ett brett spektrum av fientliga attacker och undvika att göra restriktiva antaganden om den fientliga förmågan.
- *Flexibilitet.* Ramverket bör vara användbart i ett brett spektrum av tillämpningar; både i situationer där dataintegritet är av yttersta vikt, och där kraven inte är lika strikta. Definitionen av informationsläckage bör också vara applicerbart på olika datatyper.

Definitionen av dataintegritet som presenteras i denna avhandling följer kriterierna ovan och kallas *punktvist maximalt läckage* (PML). PML är en stokastisk variabel som mäter mängden informationsläckage från en hemlig stokastisk variabel  $X$  till en relaterad, men publik, stokastisk variabel  $Y$ . Vi börjar med att definiera PML för diskreta stokastiska variabler genom studier av två till synes olika, men matematiskt ekvivalenta, attackscenarier: den *slumpmässiga funktionsmodellen* och *vinstfunktionsmodellen*. Vi vidareutvecklar vinstfunktionsmodellen till stokastiska variabler i godtyckliga sannolikhetsrum, vilket resulterar i en mer generell form av PML. Vidare studerar vi egenskaperna för PML före och efter databehandling och funktionskomposition; definierar flera dataintegritetsgarantier; samt jämför PML med existerande dataintegritetsdefinitioner, såsom differentiell dataintegritet och dess lokala variant.

Per definition är PML ett *inferentiellt* dataintegritetsmått, i bemärkelsen att det jämför en fiendes information om  $X$  före och efter databehandling. En vanlig missuppfattning inom forskningsfältet är dock att meningsfulla inferentiella dataintegritetsgarantier är ouppnåeliga. Detta beror på en övertolkning

---

av ett resultat som kallas *omöjligheten att helt förebygga informationsutlämnande*. Genom en grundläggande perspektivförändring kan vi precisera karaktärisera de typerna av informationsutlämnande som kan förebyggas genom dataintegritetsgarantier, och de som förblir oundvikliga. Med bakgrund av detta argumenterar vi för användandet av inferentiella dataintegritetsmått.

En tillämpning vi undersöker är ett vanligt maskininlärningsramverk för dataintegritetsbevarande inläring som kallas Privat Aggregation av Lärarens-embler (eng: *Private Aggregation of Teacher Ensembles* (PATE)), genom ett informationsteoretiskt dataintegritetsmått. Specifikt föreslår vi en betingad form av maximalt läckage för att kvantifiera mängden informationsläckage från individuella datapunkter, och visar att läckaget är Schur-konkavt när det tillagda bruset har en log-konkav sannolikhetsfördelning. Läckagets Schur-konkavitet innebär att ökad klassificeringsprestanda stärker dataintegriteten. Vi härleder också övre gränser på informationsläckaget när det tillagda bruset följer en Laplacefördelning.

Till sist designar vi optimala dataintegritetsmekanismer som minimerar Hammingdistorsionen i situationer där det maximala läckaget är begränsat, under antagande att (i) a-priori-fördelningen är känd, (ii) a-priori-fördelningen tillhör en given mängd av möjliga sannolikhetsfördelningar. Vi visar att de mängder av a-priori-fördelningar som innehåller fler *uniforma* sannolikhetsfördelningar genererar större distorsion. Vi visar också att dataintegritetsmekanismer som distribuerar dataintegritetsbudgeten mer uniformt över utfallen ger upphov till mindre distorsion i värsta fall.

**Nyckelord:** Dataintegritet, informationsläckage, punktvis maximalt läckage, avslöjningsprevention, inferentiell dataintegritet, mekanismdesign.

*To the fearless women defying norms, the champions of science and freedom.*



---

# Acknowledgments

---

My sincere gratitude goes to my main supervisor, Professor Tobias Oechtering, for his support and guidance during the past four and a half years. Your constant availability for discussions and the opportunity to work on this interesting topic have been invaluable to my academic journey. I am also thankful for your encouraging words which served as a motivation during challenging times. I extend my thanks to my co-supervisors, Professor Mikael Skoglund and Assistant Professor Giulia Cervia. Mikael's composed and always timely wisdom proves that academic excellence does not need to be loud and flashy. To Giulia, I am grateful for her mentorship as well as friendship. A special note of thanks for hosting my research visit at Lille, an experience that turned out to be both highly educational and enjoyable.

I extend my appreciation to the opponent for this thesis, Professor Catuscia Palamidessi as well as the members of the grading committee, Professor Fady Alajaji, Professor Simone Fischer-Hübner, and Professor Daniel Kifer. I am also grateful to Professor Mikael Johansson for acting as the defense chair and Professor Joakim Jaldén for the advance review of the thesis. Special thanks go to Professor Parastoo Sadeghi for many insightful research discussions, and to Professor Cecilia Magnusson Sjöberg for her patient explanations about the GDPR and the lawyer's perspective on privacy. I would also like to thank the Digital Futures centre for supporting my position via the project DataLEASH.

Now, to the individuals whose impact on my life is beyond description. Dimitris, your constant presence in my life has been a reliable source of calm. Your top-class humor combined with your remarkable culinary skills (notably, in crafting pizzas, steaks, and noodle soups) has often rejuvenated my tired soul. Sina, your kindness and caring nature truly make you stand out. You have consistently been my number one source of information (and gossip) within KTH or otherwise. Rezvaneh, my deepest thanks go to you for being the strongest person that I know. You are my role model of a no-nonsense, badass, and stubborn woman.

Next, to the esteemed members of the ISE PhD student/post-doc brigade, including past and present members: Samie, Borja, Wendi, Martin, Anubhab, Antoine, Jaume, Aris, Sangwon, Amaury, Leo, Steven, Movitz, Shudi, Xuechun, Javier, Vishnu, Mengyuan, Maryam, Jeannie, Lissy, Neel, Cheng, Yusen, Ramana, Linghui, Baptiste, Hamid, Germán, Prakash, and others. Our lunch hours and afterworks have been some of the most enjoyable moments of my PhD experience. These occasions not only allowed for the exchange of pleasantries but also served as an opportunity for discussing many intriguing research ideas (e.g.,

extensions of the impactful work “machine learning for detecting attractive people”). Special thanks to Martin for the fantastic translation of my thesis abstract to Swedish, and to Antoine for answering all my defense-related questions patiently. Double special thanks to Borja for our many many discussions over the years. Whenever I found myself stuck on a problem, yours was always the first door I would knock. Even more special thanks is reserved for Wendi and Xuechun whose feminine camaraderie has been a valuable resource in our extremely male-dominated workplace. And, of course, a special nod to the elusive “banana thief” whose cunning, mischievous, and professional-quality acts of theft left us all bewildered for many months, and gave us a mystery to solve.

I am incredibly thankful to my friends Parmiss, Samie, Shahab, Amir, Parastu, Albin, Aida, Oskar, Shima, Farhad, Avenia, Mehdi, and Saba. You have been my lifeline since the moment I moved away from home. The times spent together with you (notably, our Persian dance parties) brought me warmth, especially during the dark and cold months of Stockholm. My deepest appreciation goes to Arash, Pegah, and Ehsan, whose friendship and support transcend the passage of years and the vast geographical distances that separate us. A special shout-out to the members of the PhD student council, especially Saumey and Susanna. Thank you for caring, and thank you for the time and energy you invested in making the PhD experience more enjoyable for everyone. Also, a big thank you to ChatGPT for proofreading my texts during the past months.

Last, but most certainly not least, my profound and heartfelt thanks go to my family: Niloufar, Morteza, Alireza, and Yasmin. I owe you everything I have ever achieved. Thank you for your unwavering support, and for believing in me. Your love and encouragement have allowed me to dream without limits.

*Sara Saeidian*  
Stockholm, Jan. 2024

---

# List of Papers

---

Papers included in the thesis:

- I Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Quantifying membership privacy via information leakage. *IEEE Transactions on Information Forensics and Security*, 16:3096–3108, 2021
- II Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Optimal maximal leakage-distortion tradeoff. In *2021 IEEE Information Theory Workshop (ITW)*, pages 1–6. IEEE, 2021
- III Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Pointwise maximal leakage. *IEEE Transactions on Information Theory*, 69 (12):8054–8080, 2023
- IV Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Pointwise maximal leakage on general alphabets. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 388–393, 2023
- V Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Rethinking disclosure prevention with pointwise maximal leakage. *Submitted to: Journal of Privacy and Confidentiality*, 2023. URL <https://people.kth.se/~oech/JPC23.pdf>

Other contributions by the author not included in the thesis:

- VI Leonhard Grosse, Sara Saeidian, and Tobias J. Oechtering. Extremal mechanisms for pointwise maximal leakage. *Submitted to: IEEE Transactions on Information Forensics and Security*, 2023. URL <https://arxiv.org/pdf/2310.07381.pdf>
- VII Leonhard Grosse, Sara Saeidian, Parastoo Sadeghi, Tobias J. Oechtering, and Mikael Skoglund. Quantifying privacy via information density. *To be Submitted to: 2024 IEEE International Symposium on Information Theory (ISIT)*, 2024



---

# Contents

---

<b>Acknowledgments</b>	<b>vii</b>
<b>List of Papers</b>	<b>ix</b>
<b>Contents</b>	<b>1</b>
<b>1 Introduction</b>	<b>5</b>
1.1 In Pursuit of Privacy . . . . .	5
1.2 Motivation and Scope . . . . .	6
1.3 Outline and Contributions . . . . .	7
<b>2 Preliminaries</b>	<b>11</b>
2.1 Notations and Assumptions . . . . .	11
2.2 Shannon Entropy, Relative Entropy, and Mutual Information . . . . .	13
2.3 Rényi Entropy and Rényi Divergence . . . . .	15
2.4 $f$ -divergence and $f$ -information . . . . .	17
2.5 Sibson and Arimoto Mutual Information . . . . .	18
2.6 Majorization Theory . . . . .	19
<b>3 An Overview of Existing Privacy Notions</b>	<b>23</b>
3.1 Syntactic Privacy . . . . .	23
3.2 Differential Privacy . . . . .	23
3.3 Information-theoretic Privacy . . . . .	29
3.4 Quantitative Information Flow and Maximal Leakage . . . . .	31
3.5 Other Notions . . . . .	36
<b>4 Pointwise Maximal Leakage: Definitions</b>	<b>39</b>
4.1 Randomized Function View of Leakage . . . . .	39
4.2 Gain Function View of Leakage . . . . .	43
4.3 PML vs Dynamic Min-entropy Leakage . . . . .	46
4.4 PML on General Alphabets . . . . .	47
<b>Appendices</b>	<b>55</b>
4.A Proof of Theorem 4.10 . . . . .	55
4.B Proof of Propostion 4.18 . . . . .	60
4.C Proof of Lemma 4.25 . . . . .	62

<b>5</b>	<b>PML: Properties, Privacy Guarantees, and Comparisons</b>	<b>63</b>
5.1	Properties of PML	63
5.2	Basic Privacy Guarantees	66
5.3	Privacy Guarantees: Data-processing Properties	68
5.4	Privacy Guarantees: Composition Properties	78
5.5	Relationship to Local Privacy Measures and Statistical Notions	79
5.6	Relationship to Notions of Database Privacy	86
	<b>Appendices</b>	<b>95</b>
5.A	Proof of Lemma 5.2	95
5.B	Proof of Lemma 5.5	97
5.C	Proof of Proposition 5.9	97
5.D	Proof of Lemma 5.16	98
5.E	Proof of Proposition 5.21	98
5.F	Proof of Proposition 5.23	99
5.G	Proof of Theorem 5.24	99
5.H	Proof of Proposition 5.28	102
5.I	Proof of Proposition 5.30	103
5.J	Proof of Proposition 5.32	103
5.K	Proof of Proposition 5.35	104
5.L	Proof of Proposition 5.34	104
5.M	Proof of Proposition 5.36	105
5.N	Proof of Proposition 5.42	106
<b>6</b>	<b>Disclosure Prevention with PML</b>	<b>109</b>
6.1	Notes on Notation and Terminology	110
6.2	Impossibility of Absolute Disclosure Prevention	112
6.3	What Is Privacy and What Is Utility?	113
6.4	Entropy-based Disclosure Prevention	115
6.5	How to Pick $\epsilon$ ?	122
	<b>Appendices</b>	<b>127</b>
6.A	Proof of Proposition 6.8	127
6.B	Proof of Proposition 6.10	127
<b>7</b>	<b>Application: Privacy Risk Assessment</b>	<b>131</b>
7.1	Pointwise Conditional Maximal Leakage	132
7.2	System Model	133
7.3	Privacy-accuracy Synergy	135
7.4	Leakage Bounds	137
	<b>Appendices</b>	<b>141</b>
7.A	Proof of Theorem 7.5	141
7.B	Proof of Proposition 7.6	145

7.C Proof of Proposition 7.10 . . . . .	147
<b>8 Application: Maximal Leakage-Distortion Tradeoff</b>	<b>149</b>
8.1 Problem Setup . . . . .	151
8.2 Known Prior Distribution . . . . .	152
8.3 A Set of Possible Priors . . . . .	154
8.4 Ordering Privacy Mechanisms by Maximum Distortion . . . . .	158
<b>Appendices</b>	<b>161</b>
8.A Proof of Proposition 8.4 . . . . .	161
8.B Proof of Lemma 8.5 . . . . .	162
8.C Proof of Proposition 8.7 . . . . .	162
<b>9 Conclusions and Future Directions</b>	<b>165</b>
<b>Bibliography</b>	<b>167</b>



---

# 1. Introduction

---

Data is the new currency of today’s information age [50]. Private and public organizations increasingly rely on data collected from individuals for making decisions and providing services. While the insights derived from data undoubtedly contribute value to societies, the corresponding *privacy* risks must not be overlooked. For instance, in the U.S., data obtained from the decennial census plays a crucial role in determining congressional apportionment and allocating federal funding. However, if not appropriately privatized, such demographically rich data can also be used to identify individuals, particularly outliers, subjecting them to risk of fraud, disinformation, or similar [138]. This poses a fundamental challenge: how can we harness the utility provided by the data while ensuring that sensitive information is not disclosed?

Attempts to answer the above question have led to extensive research in various fields, including computer science and information theory. Generally speaking, the bulk of the (technical) research on privacy aims to answer one of two fundamental questions: (i) What is privacy, and how can we quantify it? (ii) What algorithms (i.e., *privacy mechanisms*) provide good utility while satisfying a certain definition of privacy? This thesis mainly focuses on the first question.

## 1.1 In Pursuit of Privacy

To define privacy, inspiration can be drawn from the field of cryptography, in particular, *semantic security* [54]. In this context, privacy may be articulated as the principle that “nothing should be learnable about individuals from the data that cannot be learned without the data.” Unfortunately, this definition is too restrictive to be satisfied by any useful privacy mechanism. That is, a privacy mechanism guaranteeing *absolute disclosure prevention* cannot provide any utility [33]. Consequently, a less restrictive definition is required.

One widely accepted definition is *differential privacy* (DP) [32, 36] adopted both by public agencies (e.g., the U.S. Census Bureau [2]) and big data collectors from industry (e.g., Apple [135], Google [42], and Microsoft [28]). DP assumes that data collected from individuals is stored in a database that returns answers to queries in a privacy-preserving manner. The objective is to disclose properties of the population as a whole while preserving the privacy of each individual. Specifically, DP ensures that two databases differing in a single entry, presumably information pertaining to a single individual, cannot be distinguished based on their corresponding query responses. This approach aligns with the principle that

“nothing should be learnable about an individual participating in a database that could not be learned without participation” [33].

Despite its widespread success, some works have argued that DP may not provide sufficient protection for databases containing *correlated* data [75, 76, 58, 92, 149, 86, 156]. Informally, this is because there may be no one-to-one mapping between individuals and entries in the database, and each person’s information may contribute to multiple entries. Differential privacy enthusiasts counter this view by arguing that correlations can always lead to disclosing more information than intended by a privacy mechanism [102, 103]. They further argue that attempting to ensure that a database discloses no information about individuals aligns with the principle of absolute disclosure prevention, which was shown to be unattainable by Dwork and Naor [33]. To end this divide, Tschantz et al. [137] suggest that DP should be understood through the lens of *causality* [110] rather than Bayesian probability theory, without assuming an underlying distribution on databases.

In parallel to these developments, in the computer security and information theory communities, several quantitative notions of information leakage have been proposed [131, 20, 44, 4–6, 63, 90, 53, 82]. These notions assess an information system’s vulnerability by considering a threat model where an adversary pursues a certain objective. One notable privacy definition arising from this approach is *maximal leakage* [63]. While maximal leakage has a strong operational meaning and satisfies useful properties, it is an “average-case” measure, that is, it characterizes privacy for the average outcome of a privacy mechanism. This average-case characterization may be insufficient in privacy-critical applications with strict requirements.

Other works still explore a diverse array of information measures and statistical quantities as privacy measures. These include mutual information [10, 9, 143, 95, 89, 115], probability of correctly guessing [11], *f*-information [27], and information density [22, 65, 66] (see also [141, 18, 62] for an extensive list of various privacy measures). Nevertheless, the adoption of most of these quantities as privacy measures is axiomatic, and it remains unclear whether or not they have any operational significance in a privacy context.

## 1.2 Motivation and Scope

This thesis is motivated by the observation that even though privacy is a well-studied topic in many disciplines, relatively few (technical) works give a precise definition of privacy or discuss what type of privacy is captured by the various measures proposed in the literature. In response to this observation, here we establish an application-agnostic and theoretical framework for the assessment and design of privacy-preserving data processing systems. Central to the framework is a concept of *information leakage* or *privacy loss* that satisfies three fundamental properties:

- *Explainability.* The notion of information leakage employed is operationally meaningful, emerging organically from the analysis of adversarial attack scenarios. This ensures that the privacy guarantees within this framework can be explained to stakeholders and that the associated privacy parameters are meaningful and interpretable.
- *Robustness.* The notion of information leakage employed is resilient against a diverse array of potential adversaries, accommodating a broad range of attack scenarios while refraining from making restrictive assumptions about adversarial capabilities.
- *Flexibility.* The framework caters to both highly privacy-sensitive applications and those with more relaxed privacy requirements. The notion of information leakage employed is also applicable to various data types, enhancing the versatility of the framework.

The concept of information leakage introduced in this thesis is called *point-wise maximal leakage* (PML). PML, denoted by  $\ell(X \rightarrow Y)$ , is a random variable that measures the amount of information leaking about a secret random variable  $X$  to a publicly available and related random variable  $Y$ . PML is introduced by drawing inspiration from multiple existing privacy definitions and harnesses the strengths of each framework. First, our aspirations and objectives closely align with those that motivated the definition of differential privacy. Specifically, PML privacy guarantees enable disclosing aggregate properties of an entire population while concealing the intricate details of the data. Moreover, PML demonstrates flexibility akin to differential privacy since it is a random variable that can be bounded and manipulated in various ways. Second, our framework is rooted in quantitative information flow whose well-established problem formulations provide explainability and robustness for our approach. To define PML, we build upon the adversarial models of Alvim et al. [4] and Issa et al. [63] which were used to define maximal leakage. However, we redirect our attention from the “average outcome” characterization of the previous works to individual outcomes of  $Y$  to strengthen the definition. Finally, our framework is equipped with information-theoretic tools, language, and analysis methods.

It is crucial to highlight that our primary focus lies in defining, interpreting, and justifying PML, and establishing connections with other frameworks. Nevertheless, in the later chapters of the thesis, we briefly explore PML as a tool for assessing information leakage in existing systems and designing optimal data release mechanisms.

### 1.3 Outline and Contributions

The rest of this thesis is divided into eight chapters. We briefly summarize the contents of each chapter in the following.

In Chapter 2, we discuss notation and define fundamental concepts used in the subsequent chapters, such as Rényi entropy and divergences,  $f$ -divergences, mutual information,  $f$ -information, and so on.

In Chapter 3, we provide a summary of several existing privacy frameworks with an emphasis on differential privacy, maximal leakage, and their variants or extensions. We also briefly discuss other measures such as mutual information,  $f$ -information and (local) information privacy.

In Chapter 4, we introduce pointwise maximal leakage.<sup>1</sup> First, we assume that both the sensitive information  $X$  and the outcome of the privacy mechanism  $Y$  are finite random variables. We then define PML by analyzing two adversarial scenarios: the *randomized function* model of leakage introduced by Issa et al. [63], and the *gain function* model of leakage introduced by Alvim et al. [4]. Interestingly, we also establish that, despite their apparent distinctions, the randomized function model and the gain function model are mathematically equivalent. Next, we relax the assumption of finite random variables and extend the gain function model to encompass arbitrary probability spaces, resulting in a highly general form of PML. This chapter contains results (and, possibly, verbatim copied text) from the following two papers:

- Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Pointwise maximal leakage. *IEEE Transactions on Information Theory*, 69 (12):8054–8080, 2023,<sup>1</sup>
- Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Pointwise maximal leakage on general alphabets. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 388–393, 2023.<sup>1</sup>

In Chapter 5, we further develop the theory of PML in three directions. First, we study the properties of PML, e.g., how it composes when several outcomes are observed, how the leakage is affected by pre- and post-processing, and so on. Second, we define several privacy guarantees by imposing different restrictions on  $\ell(X \rightarrow Y)$ . For example, we can require  $\ell(X \rightarrow Y)$  to be a bounded random variable. Here, we also study the data-processing and composition properties of our privacy guarantees. Third, we examine how PML relates to other privacy/statistical notions. These include max-information [39, 118], differential privacy, local differential privacy,  $f$ -information [27], and so on. We derive bounds between the different notions and discuss their implications. This chapter contains results (and, possibly, verbatim copied text) from the following two papers:

---

<sup>1</sup>Conceptualization and derivation of all results related to PML was done by the candidate. The candidate together with the supervisors has contributed to writing (including reviewing and editing) the papers on PML.

- Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Pointwise maximal leakage. *IEEE Transactions on Information Theory*, 69 (12):8054–8080, 2023,
- Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Rethinking disclosure prevention with pointwise maximal leakage. *Submitted to: Journal of Privacy and Confidentiality*, 2023. URL <https://people.kth.se/~oech/JPC23.pdf>.<sup>1</sup>

In Chapter 6, we discuss the impossibility of absolute disclosure prevention, and ask: If privacy is guaranteed in the sense of PML, then what kind of information about  $X$  can be disclosed, and what remains concealed? We show that a privacy mechanism satisfying a PML guarantee allows disclosing features of  $X$  that have small entropy, while protecting features with large entropy. Moreover, we argue that low-entropy features of  $X$  capture properties of the population as a whole while high-entropy features of  $X$  describe instance-dependent properties. Consequently, guaranteeing privacy in the sense of PML aligns with the principle that “nothing should be learnable about  $X$  that could not be learned from the prior distribution.” Our discussions also yield a precise meaning for the privacy parameter in a PML guarantee. This chapter contains results (and, possibly, verbatim copied text) from the following paper:

- Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Rethinking disclosure prevention with pointwise maximal leakage. *Submitted to: Journal of Privacy and Confidentiality*, 2023. URL <https://people.kth.se/~oech/JPC23.pdf>.

On the more application-oriented front, in Chapter 7 we examine a machine learning architecture for privacy-preserving classification called *Private Aggregation of Teacher Ensembles* (PATE) [108, 109]. Specifically, we use a conditional form of maximal leakage to quantify the amount of information leaking about each data entry and prove that the leakage is Schur-concave when the injected noise has a log-concave probability density. The Schur-concavity of the leakage suggests a synergy between privacy and accuracy in the framework. We also derive upper bounds on the information leakage when the injected noise has Laplace distribution. This chapter contains results (and, possibly, verbatim copied text) from the following paper:

- Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Quantifying membership privacy via information leakage. *IEEE Transactions on Information Forensics and Security*, 16:3096–3108, 2021.<sup>2</sup>

---

<sup>2</sup>This paper was conceptualized by Tobias J. Oechtering. The results were derived by the candidate. The candidate together with the supervisors has contributed to writing (including reviewing and editing) the paper.

In Chapter 8, we study the design of optimal privacy mechanisms. More precisely, we formulate a privacy-utility tradeoff problem using maximal leakage as the privacy measure and the expected Hamming distortion as the utility measure. We study three different but related problems. First, we assume that the prior distribution is known and we find the optimal privacy mechanism that achieves the smallest distortion subject to a maximal leakage constraint. Second, we assume that the prior belongs to some set of possible distributions and formulate a min-max problem for finding the smallest distortion achievable for the worst-case prior in the set. Third, we define a partial order on privacy mechanisms based on the largest distortion they generate. We show that sets of priors that contain more uniform distributions lead to larger distortion, while privacy mechanisms that distribute the privacy budget more uniformly over the symbols create smaller worst-case distortion. This chapter contains results (and, possibly, verbatim copied text) from the following paper:

- Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Optimal maximal leakage-distortion tradeoff. In *2021 IEEE Information Theory Workshop (ITW)*, pages 1–6. IEEE, 2021.<sup>3</sup>

In Chapter 9, we present our conclusions and discuss possible future work.

---

<sup>3</sup>The candidate has conceptualized and derived the results in this paper. The candidate together with the supervisors has contributed to writing (including reviewing and editing) this paper.

---

## 2. Preliminaries

---

In this chapter, we discuss the notation used throughout the thesis. We also define some information-theoretic quantities integral to our work. Some of these quantities, like mutual information, have been directly employed as privacy measures, while others, such as the Rényi divergence, are used to formulate privacy measures. In the final section of this chapter, we present a few definitions and results from *majorization theory*. These will be prove instrumental in Chapters 7 and 8.

### 2.1 Notations and Assumptions

We adopt the following notational conventions:  $\mathbb{R} = (-\infty, \infty)$ ,  $\mathbb{R}_+ = [0, \infty)$ ,  $\bar{\mathbb{R}}_+ = [0, \infty]$ ,  $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ ,  $\mathbb{N} = \{0, 1, \dots\}$ ,  $\mathbb{N}^* = \{1, 2, \dots\}$ ,  $[n] = \{1, \dots, n\}$  with  $n \in \mathbb{N}^*$ .  $\log(\cdot)$  denotes the natural logarithm. We also use the conventions that  $0/0 = 1$ ,  $x/0 = \infty$  if  $x > 0$ , and  $0 \cdot \log 0 = 0$ .

### Probability Notations

Suppose  $(\Omega, \mathcal{H}, \mathbb{P})$  is an abstract probability space fixed in the background, where  $\Omega$  is the sample space,  $\mathcal{H}$  is the event space, and  $\mathbb{P}$  is a probability measure on the measurable space  $(\Omega, \mathcal{H})$ . We use  $\mathcal{H}_+$  to denote the set of all functions that are measurable relative to  $\mathcal{H}$  and  $\mathcal{B}_{\bar{\mathbb{R}}_+}$ , where  $\mathcal{B}_{\bar{\mathbb{R}}_+}$  denotes the Borel  $\sigma$ -algebra on  $\bar{\mathbb{R}}_+$ . Given  $f \in \mathcal{H}_+$ , the essential supremum of  $f$  with respect to  $\mathbb{P}$  is  $\text{ess sup}_{\mathbb{P}} f = \sup\{c \in \mathbb{R}_+ : \mathbb{P}(f > c) > 0\}$ . Given  $A \in \mathcal{H}$ ,  $\mathbf{1}_A$  denotes the indicator function of the set  $A$ , that is,

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \omega \notin A, \end{cases} \quad \omega \in \Omega.$$

Suppose  $\mu$  and  $\nu$  are measures on  $(\Omega, \mathcal{H})$  and assume that  $\mu$  is  $\sigma$ -finite. If  $\nu$  is absolutely continuous with respect to  $\mu$ , denoted by  $\nu \ll \mu$ , then we write  $p = \frac{d\nu}{d\mu}$ , or alternatively,  $\nu(d\omega) = p(\omega) \mu(d\omega)$  to imply that

$$\int_{\Omega} f(\omega) \nu(d\omega) = \int_{\Omega} f(\omega) p(\omega) \mu(d\omega),$$

for all  $f \in \mathcal{H}_+$ , where  $p \in \mathcal{H}_+$  is the Radon-Nikodym derivative of  $\nu$  with respect to  $\mu$ .

Let  $\mathcal{X}$  be a set and  $\mathcal{S}_{\mathcal{X}}$  a  $\sigma$ -algebra on  $\mathcal{X}$ . A mapping  $X : \Omega \rightarrow \mathcal{X}$  is called a random variable taking values in  $(\mathcal{X}, \mathcal{S}_{\mathcal{X}})$  if  $X$  is measurable relative to  $\mathcal{H}$  and  $\mathcal{S}_{\mathcal{X}}$ . In this thesis, we exclusively use  $X$  to denote some data containing sensitive information, i.e., the *secret*. We use  $P_X$  to denote the distribution of  $X$ .

Often, we assume that  $\mathcal{X}$  is a discrete set and use some notations specific to this case. In particular, with a slight abuse of notation, we use  $P_X$  to also denote the probability mass function (pmf) of  $X$  and write  $P_X(x) := P_X(\{x\})$  for  $x \in \mathcal{X}$ . In this case,  $\mathcal{S}_{\mathcal{X}}$  is the discrete  $\sigma$ -algebra on  $\mathcal{X}$ . Furthermore, we use  $\text{supp}(P_X) := \{x \in \mathcal{X} : P_X(x) > 0\}$  to represent the support set of  $P_X$  and  $\mathcal{P}_{\mathcal{X}}$  to denote the set of all distributions with full support on  $\mathcal{X}$ .

Let  $(\mathcal{Y}, \mathcal{S}_{\mathcal{Y}})$  be a measurable space. A mapping  $P_{Y|X} : \mathcal{X} \times \mathcal{S}_{\mathcal{Y}} \rightarrow [0, 1]$  is called a transition probability kernel (or simply kernel) from  $(\mathcal{X}, \mathcal{S}_{\mathcal{X}})$  into  $(\mathcal{Y}, \mathcal{S}_{\mathcal{Y}})$  if the mapping  $x \mapsto P_{Y|X=x}(B)$  is in  $\mathcal{S}_{\mathcal{X}_+}$  for all  $B \in \mathcal{S}_{\mathcal{Y}}$ , and  $P_{Y|X=x}(\cdot)$  is a probability measure on  $(\mathcal{Y}, \mathcal{S}_{\mathcal{Y}})$  for all  $x \in \mathcal{X}$ . We sometimes use  $P_{Y|X}(B | x)$  instead of  $P_{Y|X=x}(B)$  for  $B \in \mathcal{S}_{\mathcal{Y}}$  and  $x \in \mathcal{X}$ . This notation is less awkward when we do not want to specify the outcome of  $X$  but leave it as a random variable. The kernel  $P_{Y|X}$  induces a random variable  $Y$  taking values in  $(\mathcal{Y}, \mathcal{S}_{\mathcal{Y}})$  with distribution  $P_Y$ , where

$$P_Y(B) = \int_{\mathcal{X}} P_{Y|X=x}(B) P_X(dx), \quad (2.1)$$

for all  $B \in \mathcal{S}_{\mathcal{Y}}$ . We write  $P_Y = P_{Y|X} \circ P_X$  to represent *marginalization* over  $X$  described by (2.1). In this thesis, we exclusively use  $Y$  to denote some publicly available data that contains information about  $X$ . In that sense, we also use the terms *channel* and *privacy mechanism* to refer to  $P_{Y|X}$ . When  $\mathcal{Y}$  is a discrete set, we use a similar notation described above in the case of discrete  $\mathcal{X}$ . For example, we write  $P_{Y|X=x}(y) := P_{Y|X=x}(\{y\})$  with  $y \in \mathcal{Y}$ .

Let  $P_{XY}$  be a probability measure on the product space  $(\mathcal{X} \times \mathcal{Y}, \mathcal{S}_{\mathcal{X}} \otimes \mathcal{S}_{\mathcal{Y}})$  with marginals  $P_X$  and  $P_Y$ . Then, we write  $P_{XY}(dx, dy) = P_X(dx) P_{Y|X=x}(dy)$  to imply that

$$\mathbb{E}[f] = \int_{\mathcal{X} \times \mathcal{Y}} f(x, y) P_{XY}(dx, dy) = \int_{\mathcal{X}} P_X(dx) \int_{\mathcal{Y}} f(x, y) P_{Y|X=x}(dy),$$

for all  $f \in (\mathcal{S}_{\mathcal{X}} \otimes \mathcal{S}_{\mathcal{Y}})_+$ . We may also use the more succinct notation  $P_{XY} = P_X \times P_{Y|X}$ . Let  $P_X \times P_Y$  denote the product measure of  $P_X$  and  $P_Y$ . If  $P_{XY} \ll P_X \times P_Y$ , then we call the Radon-Nikodym derivative

$$i_{P_{XY}} := \frac{dP_{XY}}{d(P_X \times P_Y)}, \quad (2.2)$$

the *information density* of  $X$  and  $Y$ , which is a jointly measurable function.

Let  $\sigma Y$  denote the  $\sigma$ -algebra generated by  $Y$  on  $\Omega$ . We use  $\mathbb{E}[f | Y]$  to denote the conditional expectation of  $f \in \mathcal{H}_+$  given  $\sigma Y$ . Since  $\mathbb{E}[f | Y] \in (\sigma Y)_+$ , then there exists  $\phi \in \mathcal{S}_{\mathcal{Y}_+}$  such that  $\mathbb{E}[f | Y] = \phi \circ Y$ . Hence, we use the notation  $\mathbb{E}[f | Y = y]$  to represent  $\phi(y)$  for each  $y \in \mathcal{Y}$ .

## Standard Borel Assumption

Unless stated otherwise, all measurable spaces are assumed to be *standard Borel* [79, Def. 8.35]. Standard Borel spaces have many convenient properties. Most prominently, joint distributions on standard Borel spaces can always be *disintegrated* into a kernel and a marginal distribution [24, Thm. IV.2.18]. That is, if  $P_{XY}$  is a distribution on  $(\mathcal{X} \times \mathcal{Y}, \mathcal{S}_X \otimes \mathcal{S}_Y)$  with marginal  $P_X$  on  $(\mathcal{X}, \mathcal{S}_X)$ , then there exists a transition probability kernel from  $(\mathcal{X}, \mathcal{S}_X)$  into  $(\mathcal{Y}, \mathcal{S}_Y)$  such that  $P_{XY}(dx, dy) = P_X(dx) P_{Y|X=x}(dy)$ . Another advantage is that if  $P_{Y|X}$  and  $Q_{Y|X}$  are both kernels and  $P_{Y|X=x} \ll Q_{Y|X=x}$  for all  $x \in \mathcal{X}$ , then we may invoke Doob's version of the Radon-Nikodym theorem to obtain a Radon-Nikodym derivative  $\frac{dP_{Y|X}}{dQ_{Y|X}}$  which is jointly measurable in  $(x, y)$  [24, Thm. V.4.44]. This has the effect that we may alternatively use  $\frac{dP_{Y|X}}{dP_Y}$  as the information density when  $P_{Y|X=x} \ll P_Y$  for all  $x \in \mathcal{X}$ .

Note that all measurable spaces of practical interest are standard Borel, e.g., countable alphabets,  $\mathbb{R}^n$ , or complete separable metric spaces endowed with Borel  $\sigma$ -algebras.

## 2.2 Shannon Entropy, Relative Entropy, and Mutual Information

Now, we recall the most fundamental concepts in information theory, namely, *Shannon entropy*<sup>1</sup>, *relative entropy* or Kullback-Leibler (KL) divergence, and *mutual information*.

**Definition 2.1** (Shannon entropy). Suppose  $X \sim P_X$  is a random variable taking values in the discrete alphabet  $\mathcal{X}$ . The Shannon entropy of  $X$  is defined as

$$H(X) := - \sum_{x \in \mathcal{X}} P_X(x) \log P_X(x).$$

Note that  $H(X)$  is a function of the distribution of  $X$ , so we could alternatively write  $H(P_X)$ .

Shannon entropy is a measure of the information content of a random variable. It is an operationally meaningful quantity and describes the length of the shortest code that (losslessly) compresses data whose source is an i.i.d random variable [128],[25, Chapter 5]. Moreover,  $H(X) \geq 0$  with equality if and only if  $P_X$  is degenerate, i.e., if  $\text{supp}(P_X)$  is a singleton. When  $\mathcal{X}$  is finite, we have  $H(X) \leq \log|\mathcal{X}|$  with equality if and only if  $P_X$  is the uniform distribution on  $\mathcal{X}$ .

---

<sup>1</sup>Shannon entropy is usually just called entropy. In this thesis, however, our default notion of the uncertainty of a probability distribution is *min-entropy*, and we always use the term Shannon entropy to refer to  $H(X)$ .

Below, we define the conditional Shannon entropy, which is a measure of the uncertainty in the value of a random variable when another random variable is known.

**Definition 2.2** (Conditional Shannon entropy). Suppose  $X$  and  $Y$  are discrete random variables. The conditional Shannon entropy of  $X$  given  $Y$  is defined as

$$\begin{aligned} H(X | Y) &:= - \sum_{y \in \mathcal{Y}} P_Y(y) H(X | Y = y) \\ &= - \sum_{y \in \mathcal{Y}} \sum_{x \in \mathcal{X}} P_{XY}(x, y) \log P_{X|Y=y}(x). \end{aligned}$$

It is well-known that *conditioning reduces Shannon entropy*, that is,  $H(X | Y) \leq H(X)$  with equality if and only if  $X$  and  $Y$  are independent. However, it is worth emphasizing that the Shannon entropy need not be reduced for all outcomes of  $Y$ . In other words, there may exist a distribution  $P_{XY}$  and an outcome  $y \in \mathcal{Y}$  such that  $H(X | Y = y) > H(X)$ . Moreover, the *chain rule* for Shannon entropy states that

$$H(X, Y) = H(X) + H(Y | X).$$

**Definition 2.3** (Relative entropy). Let  $P$  and  $Q$  be probability measures on  $(\Omega, \mathcal{H})$ . Suppose  $P, Q \ll \mu$ , where  $\mu$  is a  $\sigma$ -finite measure on  $(\Omega, \mathcal{H})$ . The relative entropy between  $P$  and  $Q$  is defined as

$$D(P \| Q) := \int_{\Omega} p \log \frac{p}{q} d\mu,$$

where  $p := \frac{dP}{d\mu}$  and  $q := \frac{dQ}{d\mu}$ .

*Remark 2.4.* If  $P \ll Q$ , we could alternatively write

$$\begin{aligned} D(P \| Q) &= \mathbb{E}_Q \left[ \frac{dP}{dQ} \log \left( \frac{dP}{dQ} \right) \right] \\ &= \mathbb{E}_P \left[ \log \left( \frac{dP}{dQ} \right) \right], \end{aligned}$$

and if  $P \not\ll Q$ , then  $D(P \| Q) = \infty$ .

Relative entropy is a measure of the difference of two distributions  $P$  and  $Q$ , and has many applications in statistics and information theory. For example,  $D(P \| Q)$  describes the increase in the expected length of a code for data with source distribution  $P$  using a code designed for source distribution  $Q$  [25, Chapter 5]. It is easy to see that  $D(P \| Q) \geq 0$  with equality if and only if  $P = Q$ .

A key property of relative entropy is an inequality known as the *data-processing inequality*. It states that if  $P_Y = P_{Y|X} \circ P_X$  and  $Q_Y = P_{Y|X} \circ Q_X$ , then

$D(P_Y \| Q_Y) \leq D(P_X \| Q_X)$ . Intuitively, the data-processing inequality implies that distinguishing “noisy” distributions is more difficult than distinguishing the original distributions.

**Definition 2.5** (Mutual information). Given a pair of random variables  $(X, Y) \sim P_{XY}$  with marginals  $P_X$  and  $P_Y$ , their mutual information is defined as

$$I(X; Y) := D(P_{XY} \| P_X \times P_Y).$$

*Remark 2.6.* Since  $(\mathcal{X}, \mathcal{S}_X)$  and  $(\mathcal{Y}, \mathcal{S}_Y)$  are assumed to be standard Borel, we can alternatively write mutual information in the following forms:

$$\begin{aligned} I(X; Y) &= D(P_{Y|X} \times P_X \| P_X \times P_Y) \\ &= D(P_{Y|X} \| P_Y | P_X) \\ &= \mathbb{E}_{X \sim P_X} [D(P_{Y|X}(\cdot | X) \| P_Y)]. \end{aligned}$$

Furthermore, if  $P_{XY} \ll P_X \times P_Y$ , then

$$I(X; Y) = \mathbb{E}_{P_{XY}} [i_{P_{XY}}],$$

where  $i_{P_{XY}}$  is the information density defined in (2.2), and if  $P_{XY} \not\ll P_X \times P_Y$ , then  $I(X; Y) = \infty$ . An example of a case where  $P_{XY} \not\ll P_X \times P_Y$  is  $X \sim \mathcal{N}(0, \sigma^2)$  and  $Y = -X$ . Moreover, when both  $X$  and  $Y$  are finite we have

$$\begin{aligned} I(X; Y) &= H(X) - H(X | Y) \\ &= H(X) + H(Y) - H(X, Y), \end{aligned} \tag{2.3}$$

where the second equality follows from the chain rule.

$I(X; Y)$  measures the dependence between  $X$  and  $Y$  by comparing the joint distribution with the product of the marginals using relative entropy. Mutual information also satisfies a data-processing inequality. In particular, if  $X - Y - Z$  is a Markov chain<sup>2</sup>, then  $I(X; Z) \leq \min\{I(X; Y), I(Y; Z)\}$ . Intuitively, this inequality states that no clever manipulation of  $Y$  can lead to increasing the amount of information available about  $X$ .

## 2.3 Rényi Entropy and Rényi Divergence

*Rényi entropy* is an information measure generalizing Shannon entropy that preserves the additivity of independent events [116]. It includes *min-entropy* or Rényi entropy of order infinity as a special case, denoted by  $H_\infty(\cdot)$ . Min-entropy is our designated quantifier of the uncertainty of a probability distribution in this thesis. We usually use the terms entropy and min-entropy interchangeably.

<sup>2</sup>We say that  $X - Y - Z$  is a Markov chain if  $P_{XYZ} = P_X \times P_{Y|X} \times P_{Z|Y}$ . In other words,  $X$  and  $Z$  are conditionally independent given  $Y$ . This implies that  $X$  depends on  $Z$  only through  $Y$  and vice versa.

**Definition 2.7** (Rényi entropy). Let  $\alpha \in (0, 1) \cup (1, \infty)$ . Suppose  $X \sim P_X$  is a random variable taking values in the discrete alphabet  $\mathcal{X}$ . The Rényi entropy of  $X$  is defined as

$$H_\alpha(X) := \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P_X(x)^\alpha.$$

Furthermore, the Rényi entropies of orders  $\alpha = 1, \infty$  are defined by continuity as

$$H_1(X) := \lim_{\alpha \rightarrow 1} H_\alpha(X) = - \sum_{x \in \mathcal{X}} P_X(x) \log P_X(x) = H(X),$$

$$H_\infty(X) := \lim_{\alpha \rightarrow \infty} H_\alpha(X) = - \log \left( \max_{x \in \mathcal{X}} P_X(x) \right).$$

Rényi entropy is decreasing in  $\alpha$  [116], and is a Schur-concave function of  $P_X$  (Definition 2.15) for all  $\alpha \in (0, \infty]$  [60]. Therefore, when  $\mathcal{X}$  is a finite set we have

$$H_\alpha(X) \leq \log |\mathcal{X}|,$$

with equality if and only if  $X$  is uniformly distributed on  $\mathcal{X}$ . Furthermore,  $H_\alpha(X) \geq 0$  with equality if and only if  $P_X$  is degenerate.

Unlike Shannon entropy, there is no commonly accepted notion of *conditional* Rényi entropy even though several possible definitions have been proposed in the literature [48, 134]. Fehr and Berens [48] argue in favor of Arimoto's definition due to its suitable properties, e.g., monotonicity under conditioning.

**Definition 2.8** (Arimoto conditional Rényi entropy). Let  $\alpha \in (0, 1) \cup (1, \infty)$ . Suppose  $X$  and  $Y$  are discrete random variables. The conditional Rényi entropy of  $X$  given  $Y$  is defined as

$$H_\alpha^A(X | Y) := \frac{\alpha}{1-\alpha} \log \sum_{y \in \mathcal{Y}} \left( \sum_{x \in \mathcal{X}} P_{XY}(x, y)^\alpha \right)^{\frac{1}{\alpha}}.$$

Furthermore, the conditional Rényi entropies of orders  $\alpha = 1, \infty$  are defined by continuity as

$$H_1^A(X | Y) := \lim_{\alpha \rightarrow 1} H_\alpha^A(X | Y) = H(X | Y),$$

$$H_\infty^A(X | Y) := \lim_{\alpha \rightarrow \infty} H_\alpha^A(X | Y) = - \log \sum_{y \in \mathcal{Y}} P_Y(y) \max_{x \in \mathcal{X}} P_{X|Y=y}(x).$$

It can be verified that  $H_\alpha^A(X | Y)$  is non-negative and decreasing under conditioning, that is,  $H_\alpha^A(X | Y) \leq H_\alpha(X)$  for all  $\alpha \in (0, \infty]$  [48]. More properties of  $H_\alpha^A(X | Y)$  are given in [134, 48].

Similarly to how Rényi entropy generalizes Shannon entropy, *Rényi divergence* generalizes the relative entropy [116]. Here, we are especially interested in the

Rényi divergence of order  $\infty$ , denoted by  $D_\infty(\cdot\|\cdot)$ . As we show in Chapter 4, PML can be written as the Rényi divergence of order  $\infty$  of the posterior distribution from the prior.

**Definition 2.9** (Rényi divergence). Let  $\alpha \in (0, 1) \cup (1, \infty)$ . Suppose  $P$  and  $Q$  be probability measures on  $(\Omega, \mathcal{H})$ . Suppose  $P, Q \ll \mu$ , where  $\mu$  is a  $\sigma$ -finite measure on  $(\Omega, \mathcal{H})$ . The Rényi divergence of  $P$  from  $Q$  is defined as

$$D_\alpha(P\|Q) := \frac{1}{\alpha - 1} \log \int_\Omega p^\alpha q^{1-\alpha} d\mu,$$

where  $p := \frac{dP}{d\mu}$  and  $q := \frac{dQ}{d\mu}$ . Furthermore, the Rényi divergences of orders  $\alpha = 1, \infty$  are defined by continuity as

$$\begin{aligned} D_1(P\|Q) &:= \lim_{\alpha \rightarrow 1} D_\alpha(P\|Q) = \int_\Omega p \log \frac{p}{q} d\mu = D(P\|Q), \\ D_\infty(P\|Q) &:= \lim_{\alpha \rightarrow \infty} D_\alpha(P\|Q) = \log \sup_{A \in \mathcal{H}} \frac{P(A)}{Q(A)} = \log \left( \operatorname{ess\,sup}_P \frac{p}{q} \right). \end{aligned}$$

Rényi divergence is non-negative and increasing in  $\alpha$ , therefore,  $0 \leq D_\alpha(P\|Q) \leq D_\infty(P\|Q)$  for all  $\alpha \in (0, \infty)$ . Moreover, if  $\alpha \geq 1$  and  $P \not\ll Q$ , then  $D_\alpha(P\|Q) = \infty$ . Rényi divergence also satisfies a data-processing inequality: If  $P_Y = P_{Y|X} \circ P_X$  and  $Q_Y = P_{Y|X} \circ Q_X$ , then  $D_\alpha(P_Y\|Q_Y) \leq D_\alpha(P_X\|Q_X)$  for all  $\alpha \in (0, \infty]$  [139].

## 2.4 $f$ -divergence and $f$ -information

Introduced by Csizár [26],  $f$ -divergences are yet another class of divergence measures that generalize the relative entropy.  $f$ -divergences maintain some of the key properties of the relative entropy, e.g., the data-processing inequality.

**Definition 2.10** ( $f$ -divergence). Let  $f : (0, \infty) \rightarrow \mathbb{R}$  be a convex function with  $f(1) = 0$ . Suppose  $P$  and  $Q$  are two probability distributions on  $(\Omega, \mathcal{H})$  with  $P, Q \ll \mu$ , where  $\mu$  is a  $\sigma$ -finite measure on  $(\Omega, \mathcal{H})$ . The  $f$ -divergence between  $P$  and  $Q$  is defined as

$$D_f(P\|Q) := \int_{\{q>0\}} q f\left(\frac{p}{q}\right) d\mu + f^*(\infty)P(\{q=0\}),$$

where  $p := \frac{dP}{d\mu}$  and  $q := \frac{dQ}{d\mu}$ ,  $f(0) := \lim_{x \downarrow 0} f(x)$  and  $f^*(\infty) := \lim_{x \downarrow 0} x f\left(\frac{1}{x}\right)$ . Specifically, if  $P \ll Q$ , then

$$D_f(P\|Q) := \mathbb{E}_Q \left[ f\left(\frac{dP}{dQ}\right) \right].$$

Similarly to relative entropy,  $f$ -divergences are non-negative and satisfy a data-processing inequality: If  $P_Y = P_{Y|X} \circ P_X$  and  $Q_Y = P_{Y|X} \circ Q_X$ , then  $D_f(P_Y\|Q_Y) \leq D_f(P_X\|Q_X)$ . Some examples of  $f$ -divergences are

- relative entropy:

$$D(P\|Q) = \int p \log \frac{p}{q} d\mu,$$

obtained by  $f(x) = x \log x$ ,

- total variation distance:

$$\text{TV}(P, Q) = \frac{1}{2} \int |p - q| d\mu,$$

obtained by  $f(x) = \frac{1}{2}|x - 1|$ , and

- $\chi^2$ -divergence:

$$\chi^2(P\|Q) = \int \left( \frac{p^2}{q} - 1 \right) d\mu,$$

obtained by  $f(x) = (x - 1)^2$ .

We may also define an information measure generalizing mutual information as the  $f$ -divergence between a joint distribution  $P_{XY}$  and the product of marginals  $P_X \times P_Y$ .

**Definition 2.11** ( $f$ -information [27, Def. 7]). Let  $f : (0, \infty) \rightarrow \mathbb{R}$  be a convex function satisfying  $f(1) = 0$ . Given a pair of random variables  $(X, Y) \sim P_{XY}$  with marginals  $P_X$  and  $P_Y$ , their  $f$ -information is defined as

$$I_f(X; Y) := D_f(P_{XY} \| P_X \times P_Y).$$

Some works have used  $f$ -information as an information leakage measure, for example, [27, 114, 23, 142]. We discuss  $f$ -information as a privacy measure in Chapter 3.

## 2.5 Sibson and Arimoto Mutual Information

While Rényi entropy and Rényi divergence are well-established as generalizations of Shannon entropy and relative entropy, obtaining an agreed-upon generalization of mutual information has proved to be more challenging. Two noteworthy definitions are due to Sibson [130] and Arimoto, and both are operationally meaningful in a privacy context [90]. Here, we define Sibson and Arimoto mutual information but we discuss them as privacy measures later in Chapter 3. Both definitions are restricted to discrete random variables.

**Definition 2.12** (Sibson mutual information). Given a pair of discrete random variables  $(X, Y) \sim P_{XY}$  with marginals  $P_X$  and  $P_Y$  the Sibson mutual information of order  $\alpha \in (0, 1) \cup (1, \infty)$  is defined as

$$I_\alpha^S(X; Y) := \inf_{Q_Y} D_\alpha(P_{XY} \| P_X \times Q_Y)$$

$$= \frac{\alpha}{\alpha - 1} \log \sum_{y \in \mathcal{Y}} \left( \sum_{x \in \mathcal{X}} P_X(x) P_{Y|X=x}(y)^\alpha \right)^{\frac{1}{\alpha}},$$

where the infimum is over all distributions with  $\text{supp}(Q_Y) = \text{supp}(P_Y)$ . Furthermore, the Sibson mutual information of orders  $\alpha = 1, \infty$  are defined by continuity as

$$\begin{aligned} I_1^S(X; Y) &:= \lim_{\alpha \rightarrow 1} I_\alpha^S(X; Y) = \inf_{Q_Y} D(P_{XY} \| P_X \times Q_Y) = I(X; Y), \\ I_\infty^S(X; Y) &:= \lim_{\alpha \rightarrow \infty} I_\alpha^S(X; Y) = \log \sum_{y \in \mathcal{Y}} \sup_{x \in \text{supp}(P_X)} P_{Y|X=x}(y). \end{aligned} \quad (2.4)$$

Similarly to mutual information, Sibson mutual information is non-negative and satisfies a data-processing inequality. That is, given a Markov chain  $X - Y - Z$  we have  $I_\alpha^S(X; Z) \leq \min\{I_\alpha^S(X; Y), I_\alpha^S(Y; Z)\}$  for all  $\alpha \in (0, \infty]$  [112].

**Definition 2.13** (Arimoto mutual information). Given a pair of discrete random variables  $(X, Y) \sim P_{XY}$  with marginals  $P_X$  and  $P_Y$  the Arimoto mutual information of order  $\alpha \in (0, 1) \cup (1, \infty)$  is defined as

$$\begin{aligned} I_\alpha^A(X; Y) &:= H_\alpha(X) - H_\alpha(X | Y) \\ &= \frac{\alpha}{\alpha - 1} \log \frac{\sum_{y \in \mathcal{Y}} \left( \sum_{x \in \mathcal{X}} P_{XY}(x, y)^\alpha \right)^{\frac{1}{\alpha}}}{\left( \sum_{x \in \mathcal{X}} P_X(x)^\alpha \right)^{\frac{1}{\alpha}}}. \end{aligned}$$

Furthermore, the Arimoto mutual information of orders  $\alpha = 1, \infty$  are defined by continuity as

$$\begin{aligned} I_1^A(X; Y) &:= \lim_{\alpha \rightarrow 1} I_\alpha^A(X; Y) = I(X; Y), \\ I_\infty^A(X; Y) &:= \lim_{\alpha \rightarrow \infty} I_\alpha^A(X; Y) = \log \frac{\sum_{y \in \mathcal{Y}} P_Y(y) \max_{x \in \mathcal{X}} P_{X|Y=y}(x)}{\max_{x \in \mathcal{X}} P_X(x)}. \end{aligned}$$

It follows directly from the monotonicity of conditional Rényi entropy that  $I_\alpha^A(X; Y) \geq 0$  for all  $\alpha \in (0, \infty]$ . In general, Arimoto mutual information does *not* satisfy a data-processing inequality [90].

## 2.6 Majorization Theory

In the final section of this chapter, we give a few definitions and results from *majorization theory* that will prove useful in Chapters 7 and 8. The theory of

majorization formalizes the intuitive idea that the components of a vector  $p \in \mathbb{R}^n$  can be more or less “spread out” compared to the components of another vector  $q \in \mathbb{R}^n$ . All the definitions and results presented here can be found in [98].

Let  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$  be a vector. We use  $x_{[k]}$  to denote the  $k$ th largest element in  $x$  and  $x_{(k)}$  to denote the  $k$ th smallest element of  $x$ . The following notations represent the sum of various elements of  $x$ :

- $\tilde{x}_k = \sum_{j=1}^k x_j$  denotes the sum of the first  $k$  elements of  $x$ ,
- $\tilde{x}_{[k]} = \sum_{j=1}^k x_{[j]}$  denotes the sum of the  $k$  largest elements of  $x$ ,
- $\tilde{x}_{(k)} = \sum_{j=1}^k x_{(j)}$  denotes the sum of the  $k$  smallest elements of  $x$ .

**Definition 2.14** (Majorization). Given  $p, q \in \mathbb{R}^n$ , we say that  $p$  majorizes  $q$ , denoted by  $p \succ q$ , if

$$\tilde{q}_{[m]} \leq \tilde{p}_{[m]} \quad \text{for } m = 1, \dots, n-1 \quad \text{and} \quad \tilde{p}_n = \tilde{q}_n, \quad (2.5)$$

or alternatively, if

$$\tilde{q}_{(m)} \geq \tilde{p}_{(m)} \quad \text{for } m = 1, \dots, n-1 \quad \text{and} \quad \tilde{p}_n = \tilde{q}_n. \quad (2.6)$$

Majorization defines a partial order on vectors in  $\mathbb{R}^n$ , i.e., a relation that is reflexive, transitive, and anti-symmetric. Note that not all  $n$ -dimensional vectors can be compared in terms of majorization, e.g.,  $(4, 4, 1)$  and  $(5, 2, 2)$  cannot be compared. On the other hand, if we define  $\mathcal{Q} = \{(q_1, q_2, q_3) \in \mathbb{R}_+^3 : \sum_{i=1}^3 q_i = 9\}$ , then  $(3, 3, 3)$  is majorized by all  $q \in \mathcal{Q}$  while  $(9, 0, 0)$ ,  $(0, 9, 0)$  and  $(0, 0, 9)$  majorize all  $q \in \mathcal{Q}$ . A graphical illustration of majorization is given in Figure 2.1.

Now, we introduce Schur-convex functions which are order-preserving (i.e., increasing) with respect to the majorization partial order.

**Definition 2.15** (Schur-convex/concave function). Let  $\mathcal{J} \subseteq \mathbb{R}$  and  $p, q \in \mathcal{J}^n$ . A real-valued function  $\Phi : \mathcal{J}^n \rightarrow \mathbb{R}$  is said to be Schur-convex if  $q \prec p$  implies  $\Phi(q) \leq \Phi(p)$ . Similarly,  $\Phi$  is said to be Schur-concave if  $q \prec p$  implies  $\Phi(q) \geq \Phi(p)$ .

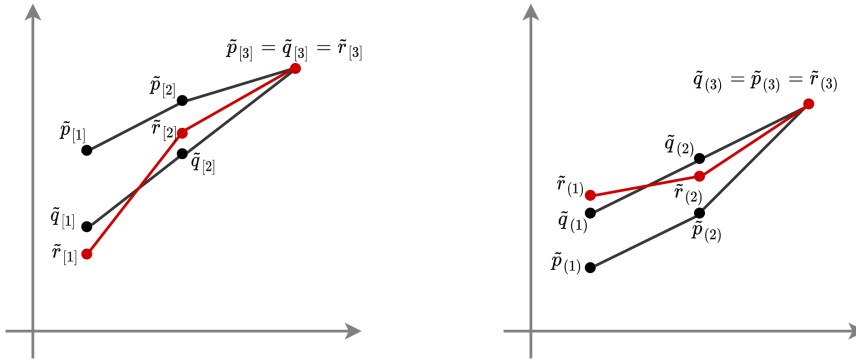
The following result is a direct consequence of Definitions 2.14 and 2.15.

**Proposition 2.16** ([67, Thm 2.21]). Let  $p \in \mathbb{R}_+^n$  and  $S \geq 0$ . Suppose  $\Phi : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$  is a Schur-convex function. Consider the following problems:

$$\underset{p \in \mathbb{R}_+^n}{\text{maximize}} \Phi(p) \quad \text{subject to} \quad \sum_{i=1}^n p_i = S,$$

and

$$\underset{p \in \mathbb{R}_+^n}{\text{minimize}} \Phi(p) \quad \text{subject to} \quad \sum_{i=1}^n p_i = S.$$



(a)  $\tilde{p}_{[k]} > \tilde{q}_{[k]}, \tilde{r}_{[k]}$  for  $k = 1, 2$  and  $\tilde{p}_{[3]} = \tilde{q}_{[3]} = \tilde{r}_{[3]}$ . Since  $\tilde{q}_{[1]} > \tilde{r}_{[1]}$  but  $\tilde{q}_{[2]} < \tilde{r}_{[2]}$ ,  $q$  and  $r$  cannot be compared.

(b)  $\tilde{p}_{(k)} < \tilde{q}_{(k)}, \tilde{r}_{(k)}$  for  $k = 1, 2$  and  $\tilde{p}_{(3)} = \tilde{q}_{(3)} = \tilde{r}_{(3)}$ . Since  $\tilde{q}_{(1)} < \tilde{r}_{(1)}$  but  $\tilde{q}_{(2)} > \tilde{r}_{(2)}$ ,  $q$  and  $r$  cannot be compared.

Figure 2.1: Illustration of majorization using three vectors  $p, q, r \in \mathbb{R}_+^3$ , where we have  $q, r \prec p$ , but  $q$  and  $r$  cannot be compared in terms of majorization.

Then, the global maximum is achieved by  $p_{max} = (0, \dots, 0, S, 0, \dots, 0)$ , and the global minimum is achieved by  $p_{min} = \frac{S}{n}(1, \dots, 1)$ .

Next, we discuss the conditions for a function to be Schur-convex. A function  $\Phi(p)$  is said to be *symmetric* if  $p \in \mathbb{R}^n$  can be arbitrarily permuted without changing the value of  $\Phi(p)$ .

**Theorem 2.17** (Schur's condition). *Let  $J \subseteq \mathbb{R}$  be an open interval and suppose  $\Phi : J^n \rightarrow \mathbb{R}$  is a continuously differentiable function. Then, necessary and sufficient conditions for  $\Phi$  to be Schur-convex are*

$$\Phi \text{ is symmetric on } J^n$$

and

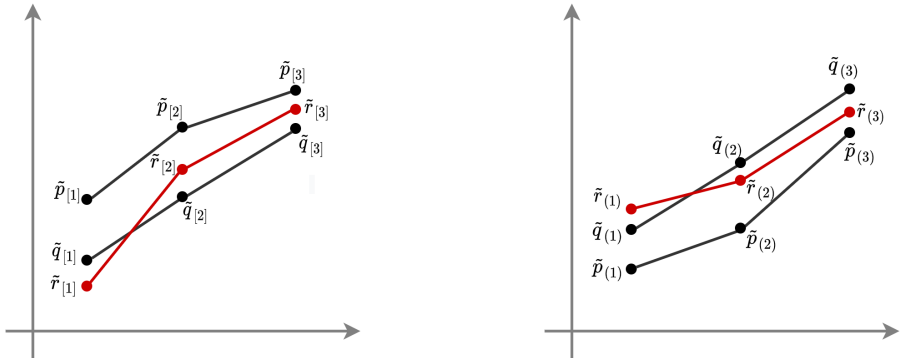
$$(p_i - p_j) \left( \frac{\partial \Phi}{\partial p_i} - \frac{\partial \Phi}{\partial p_j} \right) \geq 0,$$

for all  $p \in J^n$  and  $i, j \in [n]$ .

*Remark 2.18.* Since  $\Phi$  is symmetric, it is sufficient to verify the reduced condition

$$(p_1 - p_2) \left( \frac{\partial \Phi}{\partial p_1} - \frac{\partial \Phi}{\partial p_2} \right) \geq 0. \quad (2.7)$$

In Definition 2.14, the sum of the elements in vectors  $p$  and  $q$  are required to be equal. If we remove this condition, we arrive at the concept of *weak* majorization, which also defines a partial order on vectors.



(a)  $q, r \prec_w p$  but  $q$  and  $r$  cannot be compared.

(b)  $q, r \prec^w p$  but  $q$  and  $r$  cannot be compared.

Figure 2.2: Illustration of weak majorization using three vectors  $p, q, r \in \mathbb{R}_+^3$ .

**Definition 2.19** (Weak majorization). Given two vectors  $p, q \in \mathbb{R}^n$ , we say that  $p$  weakly *sub-majorizes*  $q$ , denoted by  $q \prec_w p$ , if

$$\tilde{q}_{[m]} \leq \tilde{p}_{[m]} \quad \text{for all } m = 1, \dots, n.$$

Furthermore, we say that  $p$  weakly *super-majorizes*  $q$ , denoted by  $q \prec^w p$ , if

$$\tilde{q}_{(m)} \geq \tilde{p}_{(m)} \quad \text{for all } m = 1, \dots, n.$$

Weak majorization is illustrated in Figure 2.2.

In order for a Schur-convex to be order-preserving with respect to weak majorization, we need to specify an extra condition on the function.

**Theorem 2.20.** *Let  $\mathcal{J} \subseteq \mathbb{R}$  and  $p, q \in \mathcal{J}^n$ . Suppose  $\Phi : \mathcal{J}^n \rightarrow \mathbb{R}$  is a Schur-convex function. If  $\Phi$  is increasing in each coordinate, then  $q \prec_w p$  implies  $\Phi(q) \leq \Phi(p)$ . Conversely, if  $\Phi$  is decreasing in each coordinate, then  $q \prec^w p$  implies  $\Phi(q) \leq \Phi(p)$ .*

---

## 3. An Overview of Existing Privacy Notions

---

This chapter gives a concise overview of the key advancements in the field of privacy, aiming to discern the strengths and weaknesses of each existing framework. This account also elucidates how PML fits within the broader context of existing work.

### 3.1 Syntactic Privacy

The term *syntactic privacy* refers to a class of data anonymization techniques that define privacy as a property of a dataset [133, 93, 84]. These methods were extensively researched in the pre-differential privacy era and aimed to generate tabular datasets that are immune against specific privacy attacks such as *re-identification* or *attribute disclosure*. A notable example of a syntactic privacy definition is *k-anonymity* [133] which divides all attributes of a dataset into *quasi-identifiers* and *sensitive attributes*. The goal of *k-anonymity* is then to prevent re-identification by ensuring that all combinations of quasi-identifiers appear at least  $k$  times. This method, however, remains vulnerable to attribute disclosure attacks, whereby sensitive attributes may still be learnable from a *k-anonymized* dataset.

The emergence of differential privacy prompted a shift in perspective and researchers began defining privacy as a property of the algorithm generating the output data, rather than an inherent characteristic of the output data itself. Nevertheless, it has been demonstrated that some syntactic privacy algorithms offer protection comparable to differential privacy [85].

### 3.2 Differential Privacy

Often called the gold standard of privacy, *differential privacy* (DP) is by far the most widely adopted notion of privacy. Introduced by Dwork et al. [36], DP in its original form considers the *centralized setting* where a trusted curator collects data containing sensitive information in a database. The objective of DP is then to provide responses to queries posed to the database in a privacy-preserving manner. This is achieved by ensuring that two databases differing in a single entry, termed *neighboring* databases, remain *indistinguishable* [36]. Let  $\mathcal{X}$  denote the set of possible databases and  $\mathcal{Y}$  be an arbitrary output space. Suppose  $M : \mathcal{X} \rightarrow \mathcal{Y}$

is a (randomized) function. Informally, if  $M$  satisfies DP then  $M(x)$  is not much affected by changing a single entry in  $x \in \mathcal{X}$ . Assuming that each entry in the database contains information about an individual, then we could say that outcomes of differentially private computations do not change considerably whether or not each individual is present in the database. Put differently, DP aligns with the principle that “nothing should be learnable about an individual participating in a database that could not be learned without participation.” Thus, DP provides a utilitarian view of privacy by guaranteeing that if an individual is harmed because of knowledge extracted from a database, then this was not caused by their presence in the database and could not have been avoided even if had they opted out.

**Definition 3.1** ( $\epsilon$ -DP). Let  $\epsilon \geq 0$ . A privacy mechanism  $M : \mathcal{X} \rightarrow \mathcal{Y}$  is said to satisfy  $\epsilon$ -DP if for all pairs of neighboring datasets  $x_1, x_2 \in \mathcal{X}$  and all measurable sets  $\mathcal{E} \subseteq \mathcal{Y}$  we have

$$\mathbb{P}[M(x_1) \in \mathcal{E}] \leq e^\epsilon \mathbb{P}[M(x_2) \in \mathcal{E}].$$

*Remark 3.2.* In this section, we diverge from our usual notation for the sake of consistency with the DP literature. Formally,  $M$  is a transition probability kernel from  $\mathcal{X}$  into  $\mathcal{Y}$ . Thus,  $\mathbb{P}[M(x_1) \in \mathcal{E}]$  should be understood as  $\mathbb{P}[\mathcal{E} \mid X = x]$  where the probability is over the randomness of  $M$ .

The concept of “neighboring” databases can be formalized in different ways leading to various DP notions (see [111, Chapter 4]). For example, if Definition 3.1 applies to  $x_1, x_2$  such that  $x_1$  can be obtained from  $x_2$  by adding/removing one entry, then  $M$  is said to satisfy  $\epsilon$ -unbounded DP. On the other hand, if Definition 3.1 applies to  $x_1, x_2$  such that  $x_1$  can be obtained from  $x_2$  by changing the value of one record, then  $M$  is said to satisfy  $\epsilon$ -bounded DP [75]. The strongest possible variant is called *free-lunch privacy* [75] in which any two databases are assumed to be neighbors.

**Definition 3.3** ( $\epsilon$ -free lunch privacy). Let  $\epsilon \geq 0$ . A privacy mechanism  $M : \mathcal{X} \rightarrow \mathcal{Y}$  is said to satisfy  $\epsilon$ -DP if for all pairs of datasets  $x_1, x_2 \in \mathcal{X}$  and all measurable sets  $\mathcal{E} \subseteq \mathcal{Y}$  we have

$$\mathbb{P}[M(x_1) \in \mathcal{E}] \leq e^\epsilon \mathbb{P}[M(x_2) \in \mathcal{E}].$$

The strength of differential privacy lies in simple and out-of-the-box privacy mechanisms that satisfy its various definitions. One such commonly used mechanism is the *Laplace mechanism* used to answer numerical queries with bounded  $\ell_1$ -sensitivity [36]. Let  $\text{Lap}(b)$  denote the zero mean Laplace distribution with scale parameter  $b > 0$  (i.e., variance  $2b^2$ ).

**Definition 3.4** (Laplace mechanism). Suppose  $f : \mathcal{X} \rightarrow \mathbb{R}^k$  is a function with  $\ell_1$ -sensitivity

$$\Delta_1(f) := \sup_{x_1, x_2 \in \mathcal{X} : x_1 \sim x_2} \|f(x_1) - f(x_2)\|_1.$$

Given  $\epsilon > 0$ , let  $N = (N_1, \dots, N_k)$  be an i.i.d random vector with  $N_i \sim \text{Lap}\left(\frac{\Delta_1(f)}{\epsilon}\right)$  for  $i \in [k]$ . The Laplace mechanism is defined as

$$M_f^{\text{Lap}}(x) := f(x) + N,$$

where  $x \in \mathcal{X}$ .

It can be shown that the Laplace mechanism satisfies  $\epsilon$ -DP [36].

The  $\ell_1$ -sensitivity of a query  $f$  describes the largest change in  $f(x)$  upon altering the value of one entry in  $x$ . The Laplace mechanism then computes  $f$  and perturbs each coordinate of  $f$  with Laplace noise scaled according to  $\Delta_1(f)$ . Examples of queries that can be answered via the Laplace mechanism include *counting queries*, that is, queries of the form “How many entries in the database satisfy property  $A$ ?” and *histogram queries* [38]. Another popular  $\epsilon$ -DP mechanism is the *exponential mechanism* for answering categorical queries [104]. The Laplace mechanism together with the exponential mechanism forms the basis for nearly all  $\epsilon$ -DP algorithms (see [87] for many practical examples and implementations.)

Today, differential privacy has evolved from Definition 3.1 into an extensive framework incorporating various new definitions that relax the notion of  $\epsilon$ -DP, diverse privacy mechanisms satisfying these definitions, as well as algorithms and software implementations. Most DP definitions are expressed in terms of the *privacy loss random variable* (PLRV).

**Definition 3.5** (Privacy loss random variable). Let  $M : \mathcal{X} \rightarrow \mathcal{Y}$  be a privacy mechanism and  $x_1 \in \mathcal{X}$  and  $x_2 \in \mathcal{X}$  two neighboring datasets. The privacy loss random variable between  $M(x_1)$  and  $M(x_2)$  is

$$\mathcal{L}_{M(x_1)/M(x_2)}(Y) = \log \frac{\mathbb{P}[M(x_1) = Y]}{\mathbb{P}[M(x_2) = Y]}.$$

The privacy loss random variable is essentially the log-likelihood ratio of the outcomes of  $M$  given two neighboring databases. According to the Neyman-Pearson lemma [107], PLRV captures everything we need to know about the (in)distinguishability of  $M(x_1)$  and  $M(x_2)$ . Thus, PLRV turns DP into a highly flexible framework where many definitions are conceived by controlling different statistics of the privacy loss. For example,  $\epsilon$ -DP requires that

$$\mathbb{P}_{Y \sim M(x_1)}[\mathcal{L}_{M(x_1)/M(x_2)}(Y) \leq \epsilon] = 1,$$

for all neighboring databases  $x_1, x_2 \in \mathcal{X}$ . Note that in Definitions 3.1 and 3.5 databases  $x_1, x_2$  are fixed and probabilities are calculated only according to the randomness of  $M$ .

## Approximate Differential Privacy

In some privacy mechanisms such as the *Gaussian mechanism*, the privacy loss is not uniformly bounded by  $\epsilon$ , but exceeds it on rare occasions [35]. The concept of approximate DP or  $(\epsilon, \delta)$ -DP extends Definition 3.1 to include these mechanisms.

**Definition 3.6** ( $(\epsilon, \delta)$ -DP). Let  $\epsilon \geq 0$  and  $\delta \in [0, 1]$ . A privacy mechanism  $M : \mathcal{X} \rightarrow \mathcal{Y}$  is said to satisfy  $(\epsilon, \delta)$ -DP if for all pairs of neighboring datasets  $x_1, x_2 \in \mathcal{X}$  and all measurable sets  $\mathcal{E} \subseteq \mathcal{Y}$  we have

$$\mathbb{P}[M(x_1) \in \mathcal{E}] \leq e^\epsilon \mathbb{P}[M(x_2) \in \mathcal{E}] + \delta,$$

or alternatively, if

$$\mathbb{E}_{Y \sim M(x_1)} \left[ \max \{0, 1 - e^{\epsilon - \mathcal{L}_{M(x_1)/M(x_2)}(Y)}\} \right] \leq \delta.$$

Below, we define the Gaussian mechanism which is the most prominent example of a mechanism satisfying  $(\epsilon, \delta)$ -DP [38].

**Definition 3.7** (Gaussian mechanism). Suppose  $f : \mathcal{X} \rightarrow \mathbb{R}^k$  is a function with  $\ell_2$ -sensitivity

$$\Delta_2(f) := \sup_{x_1, x_2 \in \mathcal{X} : x_1 \sim x_2} \|f(x_1) - f(x_2)\|_2.$$

Let  $\epsilon \in (0, 1)$ ,  $\delta \in (0, 1)$  and  $c^2 > 2 \log\left(\frac{1.25}{\delta}\right)$ . Given  $\sigma \geq \frac{c \cdot \Delta_2(f)}{\epsilon}$ , let  $N = (N_1, \dots, N_k)$  be an i.i.d random vector with  $N_i \sim \mathcal{N}(0, \sigma^2)$  for  $i \in [k]$ . The Gaussian mechanism is defined as

$$M_f^{\text{Gaus}}(x) := f(x) + N,$$

where  $x \in \mathcal{X}$ .

Since the Gaussian distribution has lighter tails compared to the Laplace distribution, it allows for more accurate data release and higher utility at the cost of the extra additive parameter  $\delta$ .

A common intuitive understanding of  $(\epsilon, \delta)$ -DP is that  $\epsilon$ -DP could fail with probability at most  $\delta$ . This intuition is, however, imprecise because it has been argued that  $\delta$  cannot be exactly mapped to the probability of failure [105]. Formally, if we want to allow  $\epsilon$ -DP to fail with probability at most  $\delta$ , then we arrive at the following definition first proposed by Machanavajjhala et al. [94].

**Definition 3.8** ( $(\epsilon, \delta)$ -probabilistic DP). Let  $\epsilon \geq 0$  and  $\delta \in [0, 1]$ . A privacy mechanism  $M : \mathcal{X} \rightarrow \mathcal{Y}$  is said to satisfy  $(\epsilon, \delta)$ -probabilistic DP if for all pairs of neighboring datasets  $x_1, x_2 \in \mathcal{X}$  we have

$$\mathbb{P}_{Y \sim M(x_1)} [\mathcal{L}_{M(x_1)/M(x_2)}(Y) > \epsilon] \leq \delta.$$

While the definition of probabilistic DP is mathematically more elegant than approximate DP, it is hardly ever used while approximate DP is omnipresent in the literature. This is because probabilistic DP is not *closed under post-processing* [74, 105], whereas approximate DP is.

## Post-processing and Composition

There are two properties of privacy definitions that have received significant attention: *closedness under post-processing* and *composition*. Informally, if a privacy guarantee is closed under post-processing, then it is not possible to manipulate the output of a mechanism to make it less private. For example,  $\epsilon$ -DP is closed under post-processing. Hence, if mechanism  $M_1$  satisfies  $\epsilon$ -DP, then  $M_2 = g \circ M_1$  also satisfies  $\epsilon$ -DP, where  $g$  is any function that does not depend on the data.

Composition, on the other hand, describes if and how outcomes of multiple privacy mechanisms satisfy a privacy definition. This property is crucial for two primary reasons. First, each person's data can be included in several databases. Second, composition facilitates the modular design of privacy-preserving data processing systems and analysis of iterative algorithms. Notably, the ability to accurately track the privacy loss has played a pivotal role in enabling differentially private deep learning where the data is accessed in many iterations [1].

The following result, called the *advanced composition theorem* is yet another reason for the popularity of  $(\epsilon, \delta)$ -DP [37]. It describes how to strike a balance between the two parameters  $\epsilon$  and  $\delta$  resulting from the composition of  $k$  mechanisms, where we may be willing to accept a larger  $\delta$  to achieve a smaller  $\epsilon$ .

**Theorem 3.9** (Advanced composition [132, Thm. 22]). *For  $j \in [k]$ , let  $M_j : \mathcal{X} \times \mathcal{Y}_{j-1} \rightarrow \mathcal{Y}_j$  be privacy mechanisms. Suppose  $M_j$  satisfies  $(\epsilon_j, \delta_j)$ -DP for each  $j \in [k]$ . For  $j \in [k]$ , define privacy mechanisms  $M_{1:j} : \mathcal{X} \rightarrow \mathcal{Y}_j$  inductively by  $M_{1:j}(d) := M_j(d, M_{1:j-1}(d))$ . Then,  $M_{1:k}$  satisfies  $(\epsilon, \delta)$ -DP for any  $\delta > \sum_{j=1}^k \delta_j$  and*

$$\epsilon = \min \left\{ \sum_{j=1}^k \epsilon_j, \frac{1}{2} \sum_{j=1}^k \epsilon_j^2 + \sqrt{2 \log \left( \frac{1}{\delta'} \right) \sum_{j=1}^k \epsilon_j^2} \right\},$$

where  $\delta' = \delta - \sum_{j=1}^k \delta_j$ .

## Other DP Variants

According to Pejó and Desfontaines [111], the literature now contains around 200 variations and extensions of the original DP definition. These modifications often tailor DP to diverse contexts or assumptions, for instance, by altering the definition of neighboring databases. Specifically during the past decade, the two primary drivers for introducing new definitions and algorithmic tools have been (i) the need for the precise accounting of the privacy loss, and (ii) exploiting the randomness inherent in algorithms (due to shuffling data points [43] or subsampling batches of data [14]) for improving privacy. These have given rise to formulations such as Rényi differential privacy [106], concentrated differential privacy [34, 21],  $f$ -differential privacy [30], privacy profiles [15], (analytical) Fourier accountant [80, 157], and so on.

## Semantics of Differential Privacy

As discussed earlier in this section, the goal of differential privacy is to ensure that individuals face no greater risk of harm by participating in a database than they would without participation. While one might intuitively assume that this goal is realized through Definition 3.1, it is not immediately clear from this definition alone what information an adversary can learn from the outcome of a differentially private mechanism. For this reason, several works have attempted to explain the guarantees of DP more precisely. Notably, two approaches stand out: the *frequentist* approach, formulated as a hypothesis test between two neighboring datasets, and the *Bayesian* approach using *posterior-to-posterior* comparisons. Suppose an adversary’s objective is to infer information about Alice. In the frequentist approach to understanding DP, a hypothesis test is conducted between two databases that differ only in Alice’s data. It can then be shown that DP imposes a tradeoff between the Type I and Type II error probabilities in such a hypothesis test [146, 68, 30]. On the other hand, the Bayesian approach involves comparing the adversary’s posterior distribution in the *actual* world with that in a *counterfactual* world, where the actual and counterfactual worlds differ only in Alice’s data [71, 78].

A few other works, such as [36, 51], discuss the Bayesian semantics of differential privacy through *prior-to-posterior* comparisons. In particular, Dwork et al. [36, Def. 6] define *semantic privacy* by comparing the prior and the posterior distributions of an adversary who knows all the entries in the database except for Alice’s data.<sup>1</sup> They then show that semantic privacy is equivalent to differential privacy [36, Claim 3]. Nevertheless, it is commonly believed that prior-to-posterior comparisons are unsuitable for understanding DP (and privacy in general). The concern is that such a comparison fails to distinguish between changes in the adversary’s distribution due to learning about the whole population and changes that are due to the participation of an individual in the database. The following example from [71] illustrates this point. Consider a clinical study that establishes that smoking increases the chances of developing heart disease. If an adversary initially believed that smoking reduces the risk of heart disease, and also knows that Alice smokes, then the study forces a significant adjustment in their belief about Alice. This adjustment, however, is only a *perceived* privacy privacy, not an *actual* one. This is because the knowledge gained from the study is about the whole population and not Alice specifically.

As a side note, later in the thesis, we challenge the idea that prior-to-posterior comparisons cannot distinguish between perceived privacy breaches and actual ones. The key lies in distinguishing between the adversary’s prior distribution encoding her prior belief, and the true data-generating distribution encoding properties of the population. We discuss this topic in detail in Chapter 6.

---

<sup>1</sup>An adversary who knows all the entries in a database except for one is often called an *informed* adversary.

## Local Differential Privacy

So far we have assumed that all sensitive data is collected in a database managed by a trusted data curator. The *local* model of privacy presents an alternative, eliminating the need for such a central database. In this model, each user's data point is perturbed while being collected, ensuring that only the user has access to their original, unperturbed information. This leads to the concept of local differential privacy (LDP) [72, 31].

**Definition 3.10** ( $\epsilon$ -LDP). Let  $\epsilon \geq 0$ . A privacy mechanism  $M : \mathcal{X} \rightarrow \mathcal{Y}$  is said to satisfy  $\epsilon$ -LDP if for all  $x, x' \in \mathcal{X}$  and all measurable sets  $\mathcal{E} \subseteq \mathcal{Y}$  we have

$$\mathbb{P}[M(x) \in \mathcal{E}] \leq e^\epsilon \mathbb{P}[M(x') \in \mathcal{E}].$$

Note that in the above definition,  $\mathcal{X}$  represents the set of possible values for a single user's data, as opposed to the set of all possible databases in central DP.

Below, we define the (generalized) randomized response mechanism [145, 69] which is the simplest perturbation method satisfying LDP. A comprehensive survey on LDP can be found in [150].

**Definition 3.11** (Randomized response). Suppose  $\mathcal{X} = \mathcal{Y} = [n]$ . Given  $\epsilon > 0$ , the randomized response mechanism  $M^{\text{RR}} : [n] \rightarrow [n]$  is defined as

$$\mathbb{P}[M^{\text{RR}}(x) = y] = \begin{cases} \frac{e^\epsilon}{n-1+e^\epsilon} & \text{if } x = y, \\ \frac{1}{n-1+e^\epsilon} & \text{if } x \neq y, \end{cases}$$

where  $x, y \in [n]$ .

## 3.3 Information-theoretic Privacy

The study of secrecy (and subsequently, privacy) in information theory originates from the seminal work of Shannon [127], where the goal was to hide the data from unauthorized parties. Let  $X$  denote some data containing sensitive information and  $Y$  denote the transmitted information. Shannon [127] defined the concept of *perfect secrecy* as the condition that  $I(X; Y) = 0$ , implying the independence of  $X$  and  $Y$ . Perfect secrecy was later weakened to allow for a small amount of information leakage, leading to the concepts of *weak secrecy* [148] and *strong secrecy* [100]. In these secrecy scenarios, we distinguish between the authorized and unauthorized parties, where a secret key is shared between the data holder and the authorized party.

More recently, and inspired by these early works, mutual information has been adopted as a privacy measure in a large number of works such as [8, 10, 9, 143, 95, 89, 115]. These works usually assume that a specific feature of  $X$ , denoted by

$U$ , captures the private information and should be kept hidden. Otherwise, the goal is to maintain as much information about  $X$  in  $Y$  as possible. For instance, in [8],  $I(U; Y)$  represents the leaked information while  $I(X; Y)$  represents the utility. Then, given  $\epsilon \geq 0$ , the *rate-privacy function*  $g_\epsilon(U; X)$  is defined as

$$g_\epsilon(U; X) := \sup_{P_{Y|X}: I(U; Y) \leq \epsilon} I(X; Y), \quad (3.1)$$

and the goal is to find the optimal privacy mechanism  $P_{Y|X}$  achieving the highest utility.

Note that despite its prevalence as a privacy measure, it has been argued that mutual information may misrepresent information leakage in certain systems. The following illustrating example is due to Smith [131].

**Example 3.12.** Given  $k \in \mathbb{N}^*$ , let  $\mathcal{X} = \{0, 1\}^{8k}$  and suppose  $X = (X_1, \dots, X_{8k})$  is uniformly sampled from  $\mathcal{X}$ . Consider the following two candidates for the disclosed information:

$$Y_1 = \begin{cases} X & \text{if } X \bmod 8 = 0, \\ 1 & \text{otherwise,} \end{cases} \quad \text{and} \quad Y_2 = (X_1, \dots, X_{k+1}).$$

Then,  $I(X; Y_1) \approx (k + 0.169) \log 2 \leq I(X; Y_2) = (k + 1) \log 2$ . However, the probability of correctly guessing  $X$  from  $Y_1$  is at least  $1/8$  since in one out of eight cases  $X$  is fully disclosed, whereas the probability of correctly guessing  $X$  from  $Y_2$  is only  $2^{-7k+1}$ . Here, if one considers the full disclosure  $X$  as a catastrophic privacy breach, then the use of mutual information as a privacy measure may be counter-intuitive.

Problem formulations similar to (3.1) have also been studied using generalized information measures such as *f-information* (see Definition 2.11). Specific instances of *f-information* used as privacy measures include mutual information (associated with KL-divergence),  $\chi^2$ -information [23, 142] (associated with  $\chi^2$ -divergence), and total variation privacy [114] (associated with total variation distance). Some works have also used per-letter *f*-divergences (as opposed to the average-case divergence in *f-information*) as privacy measures [151–153].

It is instructive to highlight some key aspects shared by the above information-theoretic approaches to privacy:

- These studies mainly focus on understanding the fundamental tradeoffs between privacy and utility, rather than developing simple and computationally efficient mechanisms.
- Random variables  $X$ ,  $U$ , and  $Y$  are usually assumed to be finite, and more general setups are not as extensively studied.
- *f-information* is symmetric in  $X$  and  $Y$ , indicating that the amount of information leaking about  $X$  to  $Y$  is equal to the amount of information leaking from  $Y$  to  $X$ .

- $f$ -information is a function of the joint distribution  $P_{XY} = P_{Y|X} \times P_X$ ; hence, it depends both on the privacy mechanism  $P_{Y|X}$  as well as the prior distribution  $P_X$ .
- $f$ -information offers an average-case measure of information leakage, making it a weaker notion of privacy compared to, say, LDP.
- Composition and closedness under pre or post-processing are examined via chain rules and data-processing inequalities. For example, it follows from the data-processing inequality for  $f$ -divergences that if  $X - Y - Z$  is a Markov chain, then  $I_f(X; Z) \leq I_f(Y; Z)$ .

Other extensions of mutual information introduced by Arimoto and Sibson [140, 130] are also used as privacy measures. We discuss these in the next section as generalizations of the notion of maximal leakage.

### 3.4 Quantitative Information Flow and Maximal Leakage

In much of the literature, the prevalent approach to tackling privacy problems has been to start with a particular definition of privacy, study the properties that follow from the definition, and design/optimize mechanisms that guarantee a certain level of privacy and utility. An alternative approach is to start from a *threat model* describing an adversary with specific objectives and study the system's vulnerability as a result of this adversarial model. This approach has several advantages. First, it encourages us to make our assumptions about the capabilities of the adversary (e.g., in terms of computational power or prior knowledge of the system) and her objectives explicit. Second, the privacy definition obtained by studying a threat model is operationally meaningful and easier to interpret. Third, the discussions around the advantages and limitations of different privacy measures become more transparent and objective.

The threat-model approach to privacy has been adopted in a line of work termed *quantitative information flow* [131, 20, 44, 4–6], in which several notions of information leakage are motivated, defined and studied. A central notion in quantitative information flow is *min-entropy leakage* [131, 20] (later called *multiplicative Bayes leakage* [6]) which considers a passive but computationally unbounded adversary who tries to guess the value of the secret  $X$  in one try. Min-entropy leakage, denoted by  $L(X \rightarrow Y)$ , quantifies (the log of) the increase in the probability of correctly guessing  $X$  having observed the output  $Y$ , compared to guessing  $X$  with no observations.

**Definition 3.13** (Min-entropy leakage). Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$ , the min-entropy leakage from  $X$  to  $Y$  is defined as

$$L(X \rightarrow Y) := \log \frac{\sup_{P_{\hat{X}|Y}} \mathbb{P} \left[ X = \hat{X}(Y) \right]}{\max_{x \in \mathcal{X}} P_X(x)}.$$

More simply, min-entropy leakage can be expressed as

$$\begin{aligned} L(X \rightarrow Y) &= \log \frac{\sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} P_{X,Y}(x, y)}{\max_{x \in \mathcal{X}} P_X(x)} \\ &= H_\infty(X) - H_\infty^A(X | Y) \\ &= I_\infty^A(X; Y). \end{aligned}$$

That is, min-entropy leakage is equal to Arimoto mutual information of order  $\infty$  (Definition 2.13). Clearly, min-entropy leakage depends on both the prior distribution  $P_X$  and the privacy mechanism  $P_{Y|X}$ . Therefore, to obtain a privacy measure that depends only on  $P_{Y|X}$ , Braun et al. [20] maximize min-entropy leakage over all possible priors, which leads to a quantity called *min-capacity* (later called *multiplicative Bayes capacity* [6]) expressed as

$$\begin{aligned} \sup_{P_X} L(X \rightarrow Y) &= \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}: P_X(x) > 0} P_{Y|X=x}(y) \\ &= I_\infty^S(X; Y). \end{aligned} \tag{3.2}$$

That is, min-capacity is equal to Sibson mutual information of order infinity (Definition 2.12). Interestingly, the supremum is attained by the uniform distribution [20].

Subsequent works in this area have extended the adversarial model assumed by min-entropy leakage [4, 44, 5, 63, 6]. Among these, we find two works particularly interesting: the *g-leakage* framework introduced by Alvim et al. [4] and the *maximal leakage* definition of Issa et al. [63].

### *g*-leakage

The *g*-leakage framework [4] generalizes the threat model of min-entropy leakage by considering an adversary whose goal is to construct a guess of  $X$  that maximizes a non-negative *gain* function  $g : \mathcal{X} \times \mathcal{W} \rightarrow \mathbb{R}_+$ , representing her objective. Having observed  $Y = y$ , the adversary makes a guess  $w \in \mathcal{W}$  such that the expected gain  $\mathbb{E}[g(X, w) | Y = y]$  is maximized. Then, *g*-leakage, denoted by  $L_g(X \rightarrow Y)$ , is defined as (the log of) the ratio of the expected adversarial gain having access to  $Y$ , and the expected adversarial gain without access.

**Definition 3.14** (*g*-leakage). Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$  and a gain function  $g : \mathcal{X} \times \mathcal{W} \rightarrow \mathbb{R}_+$ , the *g*-leakage from  $X$  to  $Y$  is defined as

$$L_g(X \rightarrow Y) = \log \frac{\sum_{y \in \mathcal{Y}} P_Y(y) \max_{w \in \mathcal{W}} \mathbb{E}[g(X, w) | Y = y]}{\max_{w \in \mathcal{W}} \mathbb{E}[g(X, w)]}.$$

*g*-leakage is a very useful tool for modeling a variety of adversarial goals, such as guessing the secret  $X$  in  $k \geq 1$  attempts or approximately guessing the

secret [4]. Moreover, it has been shown that for all prior distributions, maximizing  $g$ -leakage over all possible gain functions yields the same quantity as min-capacity [5]. That is,

$$\sup_g L_g(X \rightarrow Y) = \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}: P_X(x) > 0} P_{Y|X=x}(y), \quad (3.3)$$

for all  $P_X$ .

## Maximal Leakage

The setup put forward by Issa et al. [63] generalizes the threat model of min-entropy leakage by considering adversaries who are interested in guessing the value of a (randomized) discrete function of  $X$ , denoted by  $U$ . Maximal leakage, denoted by  $\mathcal{L}(X \rightarrow Y)$ , is then defined as (the log of) the increase in the probability of correctly guessing an arbitrary  $U$  having access to  $Y$ , compared to the probability of correctly guessing  $U$  without access.

**Definition 3.15** (Maximal leakage). Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$  the maximal leakage from  $X$  to  $Y$  is defined as

$$\begin{aligned} \mathcal{L}(X \rightarrow Y) &:= \sup_{U: U-X-Y} \log \frac{\sup_{P_{\hat{U}|Y}} \mathbb{P} [U = \hat{U}(Y)]}{\max_{u \in \mathcal{U}} P_U(u)} \\ &= \sup_{U: U-X-Y} L(U \rightarrow Y), \end{aligned}$$

where  $U$  and  $\hat{U}$  take values in the same arbitrary but finite alphabet  $\mathcal{U}$ .

In the above definition, a supremum is taken over all  $U$ 's satisfying the Markov chain  $U - X - Y$ . This serves two purposes. First, it adds robustness to the model since we may not know what function of  $X$  the adversary is interested in. Second, it ensures that maximal leakage satisfies a pre-processing inequality. That is, given a Markov chain  $Z - X - Y$  we directly get  $\mathcal{L}(Z \rightarrow Y) \leq \mathcal{L}(X \rightarrow Y)$ . Note that the supremum is essentially over all kernels  $P_{U|X}$  so that we could equivalently write

$$\mathcal{L}(X \rightarrow Y) := \sup_{P_{U|X}} \log \frac{\sup_{P_{\hat{U}|Y}} \mathbb{P} [U = \hat{U}(Y)]}{\max_{u \in \mathcal{U}} P_U(u)}.$$

Issa et al. [63] showed that for all  $P_X$ , maximal leakage takes a simple form and is also equal to min-capacity.

**Theorem 3.16** ([63, Thm. 1]). *Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$  the maximal leakage from  $X$  to  $Y$  can be expressed as*

$$\mathcal{L}(X \rightarrow Y) = \log \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}: P_X(x) > 0} P_{Y|X=x}(y). \quad (3.4)$$

Thus, (3.2), (3.3) and (3.4) are all equal to Sibson mutual information of order  $\infty$ . In what follows, we use the common term maximal leakage to refer to this quantity. Note that maximal leakage inherits its non-negativity and data processing inequality from Sibson mutual information. Therefore,  $\mathcal{L}(X \rightarrow Y) \geq 0$  and given a Markov chain  $X - Y - Z$ , it holds that  $\mathcal{L}(X \rightarrow Z) \leq \min\{\mathcal{L}(X \rightarrow Y), \mathcal{L}(Y \rightarrow Z)\}$ .

Issa et al. [63] also define several other forms of leakage as variations of their basic threat model. Here, we mention two definitions: *conditional maximal leakage* and *maximal realizable leakage*. Conditional maximal leakage considers an adversary who possesses some side information about  $X$  and is useful for understanding how maximal leakage composes. Maximal realizable leakage, on the other hand, characterizes the information leakage for the worst-case outcome of  $Y$ . Issa et al. [63] introduce maximal realizable leakage by observing that maximal leakage can remain small even if some outcomes of  $Y$  have large leakage as long as those outcomes have a sufficiently small probability. Here, our interest in maximal realizable leakage mainly stems from a result linking the leakage framework to local differential privacy.

**Definition 3.17** (Conditional maximal leakage [63, Def. 6]). Given a joint distribution  $P_{XYZ}$  on the finite set  $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$  the maximal leakage from  $X$  to  $Y$  given  $Z$  is defined as

$$\mathcal{L}(X \rightarrow Y | Z) := \sup_{U:U-(X,Z)-Y} \log \frac{\sup_{P_{\hat{U}|Y,Z}} \mathbb{P} \left[ U = \hat{U}(Y) \right]}{\sup_{P_{\tilde{U}|Z}} \mathbb{P} \left[ U = \tilde{U}(Z) \right]},$$

where  $U, \hat{U}, \tilde{U}$  take values in the same arbitrary but finite alphabet  $\mathcal{U}$ .

**Theorem 3.18** ([63, Thm. 6]). Given a joint distribution  $P_{XYZ}$  on the finite set  $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$  the conditional maximal leakage from  $X$  to  $Y$  given  $Z$  can be expressed as

$$\mathcal{L}(X \rightarrow Y | Z) = \log \max_{z \in \mathcal{Z}: P_Z(z) > 0} \sum_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}: P_{X|Z=z}(x) > 0} P_{Y|Z=z, X=x}(y).$$

The composition inequality for maximal leakage states that  $\mathcal{L}(X \rightarrow Y, Z) \leq \mathcal{L}(X \rightarrow Z) + \mathcal{L}(X \rightarrow Y | Z)$  [63, Corollary 2].

**Definition 3.19** (Maximal realizable leakage [63, Def. 8]). Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$  the maximal realizable leakage from  $X$  to  $Y$  is defined as

$$\mathcal{L}^r(X \rightarrow Y) := \sup_{U:U-X-Y} \log \frac{\max_{y \in \mathcal{Y}} \max_{u \in \mathcal{U}} P_{U|Y=y}(u)}{\max_{u \in \mathcal{U}} P_U(u)}.$$

**Theorem 3.20** ([63, Thm. 13]). *Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$  the maximal leakage realizable from  $X$  to  $Y$  can be expressed as*

$$\mathcal{L}^r(X \rightarrow Y) = D_\infty(P_{XY} \| P_X \times P_Y).$$

The following conceptually important result establishes a link between the divergence  $D_\infty(P_{XY} \| P_X \times P_Y)$  and the log-likelihood ratio in the expression of LDP.

**Theorem 3.21** ([63, Thm. 14]). *Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$ , it holds that*

$$\sup_{P_X \in \mathcal{P}_X} D_\infty(P_{XY} \| P_X \times P_Y) = \max_{\substack{x, x' \in \mathcal{X}, \\ y \in \mathcal{Y}}} \log \frac{P_{Y|X=x}(y)}{P_{Y|X=x'}(y)},$$

where  $\mathcal{P}_X$  denotes the set of all distributions with full support on  $\mathcal{X}$ .

In Chapter 5, we use a slightly more general form of Theorem 3.21 to connect PML and DP. We also use this result in Chapter 6 to discuss the no-free-lunch theorem of Kifer and Machanavajjhala [75].

## Maximal $\alpha$ -leakage

Maximal  $\alpha$ -leakage is an extension of maximal leakage motivated by the fact that an adversary may have a different objective than maximizing the probability of correctly guessing the value of a random variable. To address this, Liao et al. [90] introduce a tunable loss function called  $\alpha$ -loss with  $\alpha \in [1, \infty]$ , and assume that the adversary's objective is to minimize this loss. In particular, for  $\alpha = \infty$ ,  $\alpha$ -loss reduces to the probability of error, which retrieves the setup of maximal leakage. Maximal  $\alpha$ -leakage is then defined by assuming that the adversary's optimal action, captured by  $P_{\hat{U}|Y}$  and  $P_{\tilde{U}}$ , minimizes the  $\alpha$ -loss.

**Definition 3.22** (Maximal  $\alpha$ -leakage [90]). *Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$  and  $\alpha \in (1, \infty)$ , the maximal  $\alpha$ -leakage from  $X$  to  $Y$  is defined as<sup>2</sup>*

$$\mathcal{L}_\alpha(X \rightarrow Y) := \sup_{U: \mathcal{U} \rightarrow \mathcal{X} \times \mathcal{Y}} \frac{\alpha}{\alpha - 1} \log \frac{\sup_{P_{\hat{U}|Y}} \mathbb{E} \left[ P_{\hat{U}|Y}(U | Y)^{\frac{\alpha-1}{\alpha}} \right]}{\sup_{P_{\tilde{U}}} \mathbb{E} \left[ P_{\tilde{U}}(U)^{\frac{\alpha-1}{\alpha}} \right]},$$

where  $U$ ,  $\hat{U}$ , and  $\tilde{U}$  take values in the same arbitrary but finite alphabet  $\mathcal{U}$ .

Furthermore, the maximal  $\alpha$ -leakage at  $\alpha = 1, \infty$  is defined by continuous extension as

$$\mathcal{L}_1(X \rightarrow Y) := \lim_{\alpha \rightarrow 1} \mathcal{L}_\alpha(X \rightarrow Y),$$

<sup>2</sup>In [90], maximal  $\alpha$ -leakage is denoted by  $\mathcal{L}_\alpha^{\max}(X \rightarrow Y)$  whereas here we use the notation  $\mathcal{L}_\alpha(X \rightarrow Y)$ .

$$\mathcal{L}_\infty(X \rightarrow Y) := \lim_{\alpha \rightarrow \infty} \mathcal{L}_\alpha(X \rightarrow Y) = \mathcal{L}(X \rightarrow Y).$$

Liao et al. [90] proved that for  $\alpha \in (1, \infty]$  maximal  $\alpha$ -leakage simplifies to Arimoto channel capacity and for  $\alpha = 1$  it is equal to mutual information.

**Theorem 3.23** ([90, Thm. 2]). *Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$  the maximal  $\alpha$ -leakage from  $X$  to  $Y$  can be expressed as*

$$\mathcal{L}_\alpha(X \rightarrow Y) = \begin{cases} \sup_{\tilde{P}_X \in \mathcal{P}_X} I_\alpha^S(\tilde{X}; Y) = \sup_{\tilde{P}_X \in \mathcal{P}_X} I_\alpha^A(\tilde{X}; Y) & \text{if } \alpha \in (1, \infty], \\ I(X; Y), & \text{if } \alpha = 1. \end{cases}$$

It is worth emphasizing that maximal  $\alpha$ -leakage is increasing in  $\alpha$ , hence,  $\mathcal{L}_\alpha(X \rightarrow Y) \leq \mathcal{L}(X \rightarrow Y)$  for all  $\alpha \in [1, \infty)$  [90, Thm. 3]. Moreover, for  $\alpha \in (1, \infty]$  maximal  $\alpha$ -leakage is a function of the mechanism  $P_{Y|X}$  alone and does not depend on the prior distribution  $P_X$ .

In [53], maximal  $\alpha$ -leakage is further extended into *maximal  $(\alpha, \beta)$ -leakage*. Similarly to maximal  $\alpha$ -leakage, maximal  $(\alpha, \beta)$ -leakage uses the parameter  $\alpha \in [1, \infty]$  to encode various adversarial objectives. It also uses a parameter  $\beta \in [1, \infty]$  to describe a transition between simply averaging over all the outcomes of  $Y$  at  $\beta = 1$  and taking the maximum over  $y \in \mathcal{Y}$  at  $\beta = \infty$ .

### 3.5 Other Notions

In addition to the notions discussed above, many other privacy definitions and frameworks have been proposed in the literature, for instance, probability of correctly guessing [11], guessing entropy [99, 97], membership privacy [86], Pufferfish privacy [77], Bayesian differential privacy [149], coupled-worlds privacy [16], and noiseless privacy [17], to name a few. Below, we state two more definitions: (local) *information privacy* [22, 65, 66] and (local) *differential identifiability* [83, 143].

Information privacy [22] and its local variant, local information privacy (LIP) are privacy notions that bound the information density  $i_{P_{XY}}(X; Y)$ . Jiang et al. [66] motivate the definition of LIP by arguing that it is a *context-aware*<sup>3</sup> privacy notion whose privacy guarantees are comparable with that of LDP.

**Definition 3.24** (Local information privacy [66]). *Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$  the privacy mechanism  $P_{Y|X}$  is said to satisfy  $\epsilon$ -LIP with  $\epsilon \geq 0$  if*

$$-\epsilon \leq i_{P_{XY}}(x; y) \leq \epsilon,$$

for all  $y \in \mathcal{Y}$  and all  $x \in \mathcal{X}$ .

---

<sup>3</sup>A privacy measure is said to be context-aware if it depends on the prior distribution  $P_X$  in addition to the privacy mechanism  $P_{Y|X}$ .

LIP has been studied in several privacy-utility tradeoff problems. For instance, Jiang et al. [65] derive optimal mechanisms that minimize the expected distortion subject to an LIP constraint. Other works study and design *watch-dog* mechanisms that merge outcomes with large information density [61, 29, 120, 154, 155].

Moreover, differential identifiability and its local variant, local differential identifiability (LDI) have emerged from legal definitions of privacy [83, 143]. These concepts are formulated directly in terms of the adversary's gained knowledge (i.e., posterior distribution) and are also context-aware.

**Definition 3.25** (Local differential identifiability [83, 143]). Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$  the privacy mechanism  $P_{Y|X}$  is said to satisfy  $\epsilon$ -LDI with  $\epsilon \geq 0$  if

$$\log \frac{P_{X|Y=y}(x)}{P_{X|Y=y}(x')} \leq \epsilon.$$

for all  $y \in \mathcal{Y}$  and all  $x, x' \in \text{supp}(P_X)$ .

Note that the notion of identifiability was originally introduced in a centralized setting [83, 143], where  $x$  and  $x'$  denote neighboring databases. Here, we have given a local version of the definition, where the ratio  $\frac{P_{X|Y=y}(x)}{P_{X|Y=y}(x')}$  is bounded by  $e^\epsilon$  for all  $x, x' \in \text{supp}(P_X)$ .



---

## 4. Pointwise Maximal Leakage: Definitions

---

This chapter introduces pointwise maximal leakage (PML). We start by assuming that  $X$  and  $Y$  are finite random variables and define PML as an extension of maximal leakage. For this, we study two adversarial scenarios: the randomized function model of leakage [63] and the gain function model of leakage [4]. Later in the chapter, we extend the gain function model to  $X$  and  $Y$  on arbitrary probability spaces.

### 4.1 Randomized Function View of Leakage

We begin by describing our first threat model, which is a pointwise extension of the model proposed by Issa et al. [63]. Suppose  $X$  is a random variable distributed according to  $P_X$  over a finite alphabet  $\mathcal{X}$ . We use  $X$  to represent some data containing sensitive information. Further, suppose  $Y$  is a random variable taking values in a finite alphabet  $\mathcal{Y}$  which is the output of a privacy mechanism  $P_{Y|X}$  with input  $X$ . Consider an adversary who is interested in guessing the value of a randomized function of  $X$ , called  $U$ , induced by  $P_{U|X}$  and satisfying the Markov chain  $U - X - Y$ . The adversary, who is computationally unbounded, observes an outcome  $y \in \text{supp}(P_Y)$  (where  $P_Y = P_{Y|X} \circ P_X$  is the output distribution) and constructs a guess of  $U$  called  $\hat{U}$  according to a kernel  $P_{\hat{U}|Y}$ . The adversary is passive in the sense that she cannot affect the outcomes of the system. Furthermore, the adversary has *white-box* knowledge about the system. That is, she knows the joint distribution  $P_{UXY}$ , and therefore, can optimize her choice of guessing kernel  $P_{\hat{U}|Y}$  to maximize her chances of correctly guessing  $U$ .

To measure the information leakage of a disclosed outcome  $y$ , we consider the ratio of the probability of correctly guessing  $U$  having observed  $y$ , and the probability of correctly guessing  $U$  with no observations (in this case, the best guess is the most probable outcome according to  $P_U$ ). Accordingly, we define the pointwise  $U$ -leakage of  $X$  as follows:

$$\ell_U(X \rightarrow y) := \log \frac{\sup_{P_{\hat{U}|Y}} \mathbb{P} [U = \hat{U} \mid Y = y]}{\max_{u \in \mathcal{U}} P_U(u)}, \quad (4.1)$$

where  $\mathcal{U}$  denotes the alphabet of  $U$ . As we may not know what  $U$  the adversary is interested in, or different adversaries may be interested in guessing different  $U$ 's, we consider the worst-case scenario by taking the supremum of (4.1) over all possible randomized functions of  $X$ . Considering this setup, we define pointwise maximal leakage denoted by  $\ell_{P_{XY}}(X \rightarrow y)$  as follows.

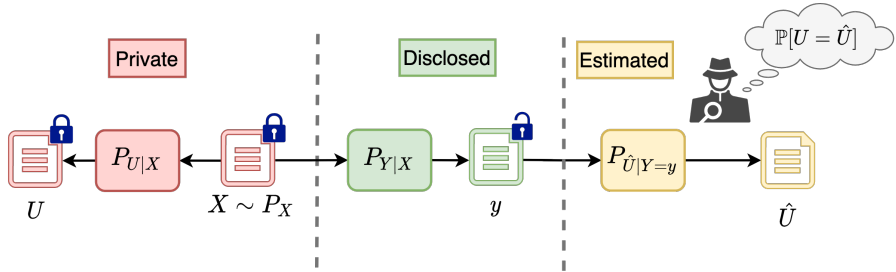


Figure 4.1: System model for the randomized function view of leakage: An adversary observes an outcome  $y$  of the channel  $P_{Y|X}$ , and tries to guess the value of a randomized function of  $X$ , denoted by  $U$ .

**Definition 4.1** (Pointwise maximal leakage). Let  $P_{XY}$  denote the joint distribution of  $X$  and  $Y$ . The pointwise maximal leakage from  $X$  to  $y \in \mathcal{Y}$  is<sup>1</sup>

$$\begin{aligned} \ell_{P_{XY}}(X \rightarrow y) &:= \sup_{P_{U|X}} \ell_U(X \rightarrow y) \\ &= \log \sup_{P_{U|X}} \frac{\sup_{P_{\hat{U}|Y}} \mathbb{P}[U = \hat{U} \mid Y = y]}{\max_{u \in \mathcal{U}} P_U(u)}. \end{aligned} \quad (4.2)$$

Below, we show that  $\ell_{P_{XY}}(X \rightarrow y)$  takes a simple form and can be expressed as the Rényi divergence of order  $\infty$  of the posterior distribution from the prior.

**Theorem 4.2.** *Given a joint distribution  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$ , the pointwise maximal leakage from  $X$  to  $y \in \mathcal{Y}$  is*

$$\ell_{P_{XY}}(X \rightarrow y) = D_\infty(P_{X|Y=y} \| P_X). \quad (4.3)$$

*Proof.* Fix an arbitrary random variable  $U$  satisfying the Markov chain  $U - X - Y$ . The numerator of (4.1) can be written as<sup>2</sup>

$$\begin{aligned} \sup_{P_{\hat{U}|Y}} \mathbb{P}[U = \hat{U} \mid Y = y] &= \sup_{P_{\hat{U}|Y}} \sum_{u, \hat{u}} \mathbf{1}[u = \hat{u}] P_{U\hat{U}|Y=y}(u, \hat{u}) \\ &= \sup_{P_{\hat{U}|Y}} \sum_{u, \hat{u}} \mathbf{1}[u = \hat{u}] P_{U|Y=y}(u) P_{\hat{U}|Y=y}(\hat{u}) \\ &= \sup_{P_{\hat{U}|Y}} \sum_u P_{U|Y=y}(u) P_{\hat{U}|Y=y}(u) \\ &= \max_{u \in \mathcal{U}} P_{U|Y=y}(u), \end{aligned}$$

<sup>1</sup>To be able to define the leakage for all  $y \in \mathcal{Y}$ , we may assume that  $\mathbb{P}(\cdot \mid Y = y) = \mathbb{P}(\cdot)$  if  $P_Y(y) = 0$ . That is, conditioning on events with probability zero equals no conditioning.

<sup>2</sup>With a slight abuse of notation, we write  $\mathbf{1}[u = \hat{u}]$  instead of  $\mathbf{1}_{\{u\}}(\hat{u})$ .

where the last equality follows from the fact that the optimal estimator  $P_{\hat{U}|Y}^*$  in the above problem satisfies

$$P_{\hat{U}|Y=y}^*(u) = \begin{cases} 1 & \text{for some } u \in \arg \max_{u \in \mathcal{U}} P_{U|Y=y}(u), \\ 0 & \text{otherwise.} \end{cases}$$

Thus, we can write

$$\begin{aligned} \exp(\ell_U(X \rightarrow y)) &= \frac{\sup_{P_{\hat{U}|Y}} \mathbb{P}[U = \hat{U} \mid Y = y]}{\max_{u' \in \mathcal{U}} P_U(u')} \\ &= \frac{\max_{u \in \mathcal{U}} P_{U|Y=y}(u)}{\max_{u' \in \mathcal{U}} P_U(u')} \\ &= \frac{\max_{u \in \mathcal{U}} \sum_{x \in \text{supp}(P_X)} P_{UX|Y=y}(u, x)}{\max_{u' \in \mathcal{U}} P_U(u')} \\ &= \frac{\max_{u \in \mathcal{U}} \sum_{x \in \text{supp}(P_X)} P_{X|Y=y}(x) P_{U|X=x}(u)}{\max_{u' \in \mathcal{U}} P_U(u')} \\ &= \max_{u \in \mathcal{U}} \sum_{x \in \text{supp}(P_X)} \frac{P_{X|Y=y}(x)}{P_X(x)} P_{X|U=u}(x) \frac{P_U(u)}{\max_{u' \in \mathcal{U}} P_U(u')} \\ &\leq \max_{u \in \mathcal{U}} \sum_{x \in \text{supp}(P_X)} \frac{P_{X|Y=y}(x)}{P_X(x)} P_{X|U=u}(x) \end{aligned} \quad (4.4a)$$

$$\begin{aligned} &\leq \max_{x \in \text{supp}(P_X)} \frac{P_{X|Y=y}(x)}{P_X(x)}. \\ &= D_\infty(P_{X|Y=y} \| P_X). \end{aligned} \quad (4.4b)$$

Taking the supremum over all  $U$ 's satisfying  $U - X - Y$  we obtain

$$\ell_{P_{XY}}(X \rightarrow y) \leq D_\infty(P_{X|Y=y} \| P_X). \quad (4.5)$$

To prove the reverse inequality, we construct a  $U$  achieving the bound in (4.5). Note that inequality (4.4b) holds with equality if there exists  $u^* \in \mathcal{U}$  such that

$$P_{X|U=u^*}(x) = \begin{cases} 1 & \text{for some } x \in \arg \max_{x \in \text{supp}(P_X)} \frac{P_{X|Y=y}(x)}{P_X(x)}, \\ 0 & \text{otherwise.} \end{cases} \quad (4.6)$$

Furthermore,  $u^*$  will also satisfy (4.4a) with equality if

$$P_U(u^*) = \max_{u \in \mathcal{U}} P_U(u). \quad (4.7)$$

An example of  $U$  satisfying both of the above conditions can be obtained through the ‘‘shattering’’ channel defined in the proof of [63, Thm. 1]. Roughly speaking,

the shattering channel breaks down each  $x \in \text{supp}(P_X)$  with probability  $P_X(x)$  into  $k(x)$  corresponding elements with probability  $\min_{x \in \text{supp}(P_X)} P(x)$ , thus creating a random variable  $U_\zeta$  with an (almost) uniform distribution. We recall the definition of the shattering channel  $P_{U_\zeta|X}$  for completeness.

**Definition 4.3** (Shattering channel [63]). Let  $p^* := \min_{x \in \text{supp}(P_X)} P_X(x)$ . Given  $x \in \text{supp}(P_X)$ , let  $k(x) := \frac{P_X(x)}{p^*}$  and  $\mathcal{U}_\zeta = \bigcup_{x \in \text{supp}(P_X)} \{(x, 1), \dots, (x, \lceil k(x) \rceil)\}$ , where  $\lceil k(x) \rceil$  denotes the smallest integer greater than or equal to  $k(x)$ . The shattering channel  $P_{U_\zeta|X}$  is defined as

$$P_{U_\zeta|X=x}(i_u, j_u) = \begin{cases} \frac{p^*}{P_X(x)} & \text{if } i_u = x, j_u = 1, \dots, \lceil k(x) \rceil, \\ 1 - \frac{\lceil k(x) \rceil p^*}{P_X(x)} & \text{if } i_u = x, j_u = \lceil k(x) \rceil, \\ 0 & \text{otherwise,} \end{cases}$$

with  $u = (i_u, j_u) \in \mathcal{U}_\zeta$  and  $x \in \text{supp}(P_X)$ , where  $\lfloor k(x) \rfloor$  denotes the largest integer smaller than or equal to  $k(x)$ .

The shattering channel induces the joint distribution  $P_{U_\zeta X}$ :

$$P_{U_\zeta X}((i_u, j_u), x) = \begin{cases} p^* & \text{if } i_u = x, j_u = 1, \dots, \lceil k(x) \rceil, \\ P_X(x) - \lceil k(x) \rceil p^* & \text{if } i_u = x, j_u = \lceil k(x) \rceil, \\ 0 & \text{otherwise,} \end{cases}$$

and  $P_{U_\zeta}$  is obtained as

$$P_{U_\zeta}(i_u, j_u) = \begin{cases} p^* & \text{if } i_u = x, j_u = 1, \dots, \lceil k(x) \rceil, \\ P_X(x) - \lceil k(x) \rceil p^* & \text{if } i_u = x, j_u = \lceil k(x) \rceil, \end{cases}$$

for  $(i_u, j_u) \in \mathcal{U}_\zeta$ . Clearly, each  $u$  is mapped to exactly one  $x$ , so condition (4.6) holds. Furthermore, each  $x$  corresponds to at least one  $u$  with probability  $P_{U_\zeta}(u) = \max_{u'} P_{U_\zeta}(u') = p^*$ , so condition (4.7) also holds. Consequently, the random variable  $U_\zeta$  obtained through the shattering channel satisfies both (4.6) and (4.7), and attains the bound in (4.5).  $\square$

*Remark 4.4.* PML can alternatively be written as

$$\begin{aligned} \ell_{P_{XY}}(X \rightarrow y) &= \max_{x \in \text{supp}(P_X)} \log \frac{P_{X|Y=y}(x)}{P_X(x)} \\ &= \max_{x \in \text{supp}(P_X)} \log \frac{P_{Y|X=x}(y)}{P_Y(y)} \\ &= \max_{x \in \text{supp}(P_X)} i_{P_{XY}}(x; y). \end{aligned}$$

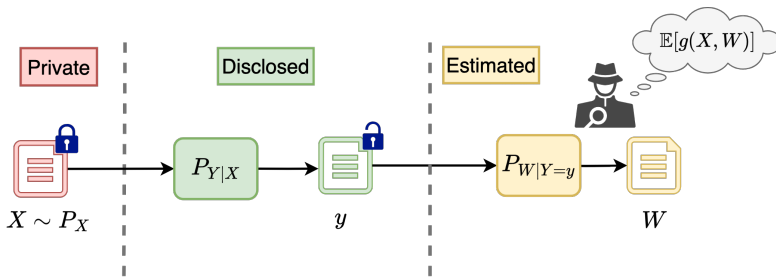


Figure 4.2: System model for the gain function view of leakage: An adversary observes an outcome  $y$  of the channel  $P_{Y|X}$ , and tries to construct a guess  $w$  of  $X$  in order to maximize a gain function  $g$ .

Note that unlike privacy measures such as maximal leakage and (local) differential privacy that depend only on the mechanism  $P_{Y|X}$ , PML depends on both the mechanism  $P_{Y|X}$  and the prior  $P_X$ , i.e., it is a property of the joint distribution  $P_{XY}$ . In what follows, we often assume that the prior  $P_X$  is arbitrary but fixed, and study PML as a function of the mechanism  $P_{Y|X}$ . When the joint distribution used to calculate PML or information density is clear from the context, we do not specify it as a subscript and write  $\ell(X \rightarrow y)$  or  $i(x; y)$ .

## 4.2 Gain Function View of Leakage

The threat model assumed in Theorem 4.2 considers an adversary who is interested in guessing the value of a randomized function of  $X$ . In this section, we argue that pointwise maximal leakage can be obtained using an alternative threat model based on (a pointwise extension of) the  $g$ -leakage framework introduced in [4]. Below, we describe this alternative threat model.

Suppose a passive and computationally unbounded adversary observes  $y \in \text{supp}(P_Y)$ , an outcome of the channel  $P_{Y|X}$ , and constructs a guess  $W$  of  $X$  using a kernel  $P_{W|Y}$  in order to maximize her expected *gain*. The adversary selects her guess from a non-empty finite set  $\mathcal{W}$  (not necessarily equal to  $\mathcal{X}$ ), and her gain is captured by a function  $g : \mathcal{X} \times \mathcal{W} \rightarrow \mathbb{R}_+$ . In order to measure the amount of information leaking from  $y$ , we consider the ratio of the expected adversarial gain having observed  $y$ , and the expected adversarial gain with no observations. As such, we define the pointwise  $g$ -leakage of  $X$  as follows:

$$\ell_g(X \rightarrow y) := \log \frac{\sup_{P_{W|Y}} \mathbb{E}[g(X, W) \mid Y = y]}{\max_{w \in \mathcal{W}} \mathbb{E}[g(X, w)]}. \quad (4.8)$$

In Theorem 4.10, we will show that the randomized function view and the gain function view of leakage are equivalent in the sense that for every gain function  $g$ , there exists a corresponding randomized function of  $X$ ,  $U_g$ , such that  $\ell_g(X \rightarrow$

$y) = \ell_{U_g}(X \rightarrow y)$ , and conversely, for every randomized function of  $X, U$ , there exists a corresponding gain function  $g_U$  such that  $\ell_U(X \rightarrow y) = \ell_{g_U}(X \rightarrow y)$ . Before presenting this result, let us demonstrate through a few examples how gain functions can be used to model different adversarial objectives.

**Example 4.5** (Identity gain function [4, Def. 3.5]). The simplest type of gain function is the identity gain which models an adversary interested in guessing the secret  $X$  itself, who is only rewarded for correct guesses. Here, the guessing space of the adversary is  $\mathcal{W} = \mathcal{X}$ , and her gain function is given by  $g^{\text{identity}}(x, w) = \mathbf{1}[x = w]$ . The  $g$ -leakage for the identity gain is

$$\ell_{g^{\text{identity}}}(X \rightarrow y) = \frac{\max_{x \in \mathcal{X}} P_{X|Y=y}(x)}{\max_{x \in \mathcal{X}} P_X(x)}, \quad (4.9)$$

which is equal to the *dynamic min-entropy leakage* defined in [44, Def. 3]. We will further discuss the identity gain and its associated  $g$ -leakage in Section 4.3.

**Example 4.6** (Membership/group privacy). Consider a centralized setting where  $x \in \mathcal{X}$  represents a database whose rows constitute data collected from individuals. Suppose an adversary is interested in guessing whether or not Alice's data is included in  $X$ , and is rewarded with a binary gain depending on whether or not her guess is correct. We can model this problem as follows: Let  $\mathcal{X}_1 = \{x \in \mathcal{X} : x \text{ contains Alice's data}\}$  denote the set of databases that contain Alice's data, and let  $\mathcal{X}_0 = \mathcal{X} \setminus \mathcal{X}_1$  be the set of databases that do not contain Alice's data. An adversary who is interested in finding out Alice's membership makes a binary guess from  $\mathcal{W} = \{0, 1\}$ , and is rewarded according to  $g(x, w) = \mathbf{1}[w = i]$  with  $i \in \{0, 1\}$  and  $x \in \mathcal{X}_i$ .

More generally, suppose the adversary has a list of  $k$  individuals and wants to find out if any one of their data points is included in  $X$ . Further, suppose the adversary is rewarded based on the number of correct guesses that she makes. To model this problem, we bi-partition the set of all databases in  $k$  different ways, one for each individual on the list. Let  $\mathcal{X}_{j_1} = \{x \in \mathcal{X} : x \text{ contains the } j\text{-th individual's data}\}$  and  $\mathcal{X}_{j_0} = \mathcal{X} \setminus \mathcal{X}_{j_1}$  for  $j = 1, \dots, k$ . Then,  $\mathcal{W} = \{0, 1\}^k$  is the guessing space of the adversary, and  $g(x, w) = \sum_{j=1}^k \mathbf{1}[w_j = j_i]$  with  $x \in \mathcal{X}_{j_i}$  and  $i \in \{0, 1\}$  is her gain function. This example can be easily extended to model cases where different individuals signify different gains for the adversary.

*Remark 4.7.* We emphasize that in the membership privacy example above we do *not* assume that the adversary is *informed* [36] (an informed adversary knows all the entries in the database except for a single entry which may be Alice's). In our setup, we assume that the adversary knows the joint distribution  $P_{XY}$ , while any other side information should be explicitly modeled as such. The concept of an informed adversary was originally proposed as a model for a very powerful adversary. However, it is now well-known that more side information does not necessarily make an adversary more effective [75, 149]. For example, [75] provides

three definitions of privacy against adversaries that either (i) know all the entries in a database except for a single entry, (ii) know all the attributes in a database except for a single attribute of a single entry, (iii) know all the bits in a database except for a single bit of a single entry. Then, it is shown that the privacy definition that seeks to limit the inference of the more knowledgeable adversary (i.e., the third adversary) may actually leak more sensitive information to the less knowledgeable adversaries.

In Chapter 5, we define a conditional form of pointwise maximal leakage that models an adversary who possesses some side information about the secret. There, we will see that side information can both increase and decrease the information leakage due to observing an outcome.

**Example 4.8** (Multiple guesses (the  $k$ -tries gain function in [4])). Consider a side-channel setting in which  $X$  represents a password and  $Y$  represents some information leaking about the password, for example, through the inter-keystroke delays. Suppose an adversary is allowed  $k \geq 1$  attempts at guessing the password correctly before getting cut off from the system. Let  $\mathcal{X}$  be the set of all possible passwords and let  $\mathcal{W} = \{w \subset \mathcal{X} : |w| \leq k\}$  denote the collection of subsets of  $\mathcal{X}$  containing  $k$  or less passwords. Then, we can model the adversary's gain through the function  $g(x, w) = \mathbf{1}_w(x)$ , where  $w$  represents the set of  $k$  or fewer attempts that the adversary makes at guessing the correct password  $x$ .

**Example 4.9** (Metric spaces [4]). Suppose  $(\mathcal{X}, \rho)$  is a metric space, where  $\mathcal{X}$  is a finite set, and  $\rho$  is a metric on  $\mathcal{X}$ . Assume that the goal of the adversary is to construct a guess  $w$  of  $x$  that minimizes  $\rho(x, w)$ . This scenario can be modeled by taking  $\mathcal{W} = \mathcal{X}$  and some non-negative gain function that is decreasing in  $\rho(x, w)$ , for example,  $g(x, w) = \exp(-\rho(x, w))$ . Many problems can be modeled as metric spaces. A simple example is in geo-location applications where the goal of an adversary may be to locate a user as accurately as possible based on partial or noisy measurements.

We have now seen how a variety of adversarial objectives can be modeled using gain functions. In the following result, we show that the definition of  $g$ -leakage in (4.8) is equivalent to the definition of  $U$ -leakage in (4.1). Thus, we unify two seemingly different ways of defining (pointwise) maximal leakage. The proof of the theorem is deferred to Appendix 4.A.

**Theorem 4.10.** *For all joint distributions  $P_{XY}$  on the finite set  $\mathcal{X} \times \mathcal{Y}$ , the randomized function view and the gain function view of leakage are equivalent. That is, for every randomized function of  $X$ , denoted by  $U$ , there exists a space  $\mathcal{W}_U$  and a gain function  $g_U : \mathcal{X} \times \mathcal{W}_U \rightarrow \mathbb{R}_+$  such that  $\ell_U(X \rightarrow y) = \ell_{g_U}(X \rightarrow y)$ . Conversely, for every gain function  $g : \mathcal{X} \times \mathcal{W} \rightarrow \mathbb{R}_+$ , there exists a randomized function of  $X$ , denoted by  $U_g$ , such that  $\ell_g(X \rightarrow y) = \ell_{U_g}(X \rightarrow y)$ .*

Note that while the above result establishes the equivalence of the gain function view and the randomized function view for pointwise leakages, it generalizes

readily to the average-case leakages of [5] and [63]. Furthermore, we obtain the following corollary which provides an alternative definition of PML.

**Corollary 4.11.** *Given  $y \in \mathcal{Y}$ , pointwise maximal leakage can be defined as*

$$\begin{aligned} \ell_{P_{XY}}(X \rightarrow y) &:= \sup_g \ell_g(X \rightarrow y) \\ &= \log \sup_g \frac{\sup_{P_{W|Y}} \mathbb{E}[g(X, W) \mid Y = y]}{\max_{w \in \mathcal{W}} \mathbb{E}[g(X, w)]}, \end{aligned}$$

where the supremum is over all gain functions with non-negative and finite ranges.

*Remark 4.12.* Corollary 4.11 is related to a recent result of Kurri et al. [81], where a variational formula for Rényi divergence of order infinity is derived as the ratio of the expected gains for guessing a randomized function of  $X$ .

### 4.3 PML vs Dynamic Min-entropy Leakage

We now take a quick detour and discuss a definition introduced by Espinoza and Smith [44] that similarly to PML, aims to measure the information leakage associated with individual observations. Espinoza and Smith [44] define the *dynamic min-entropy leakage* of an outcome  $y \in \mathcal{Y}$  as

$$\ell^{\text{dynamic}}(X \rightarrow y) := \log \frac{\max_{x \in \mathcal{X}} P_{X|Y=y}(x)}{\max_{x \in \mathcal{X}} P_X(x)},$$

which is equal to the  $g$ -leakage associated with the identity gain in (4.9). That is, the dynamic leakage is derived under the assumption that the adversary is trying to guess the secret  $X$  itself, but does not consider other gain functions.

Espinoza and Smith [44] withdraw from further developing the idea of measuring the pointwise information leakage based on  $\ell^{\text{dynamic}}(X \rightarrow y)$  by arguing that the above privacy measure suffers from two drawbacks. First, they argue that the above definition cannot be axiomatically justified as it is shown that  $\ell^{\text{dynamic}}(X \rightarrow y)$  may be negative (see [44, Example 4]), that is, the adversary's certainty about the secret may actually *decrease* by observing an outcome  $y$ . Second, they believe that dynamic policy enforcement based on individual outcomes (for example, discarding high-leakage outcomes) may reveal information about the secret  $X$ .

Note that the first issue mentioned above does not apply to  $\ell(X \rightarrow y)$ . It is easy to see that  $\ell^{\text{dynamic}}(X \rightarrow y) \leq \ell(X \rightarrow y)$ , and we have shown in Lemma 5.2 that  $\ell(X \rightarrow y) \geq 0$ , which implies that in our current setup, observations can never decrease certainty about a secret  $X$ . This is because  $\ell(X \rightarrow y)$  is defined by considering all possible gain functions an adversary may be interested in, while  $\ell^{\text{dynamic}}(X \rightarrow y)$  is defined only for the identity gain of Example 4.5.

Furthermore, contrary to [44], we believe that effective policy enforcement depends crucially on the ability to quantify the information leaking from individual outcomes as this allows us to treat information leakage as a random variable. Viewing information leakage as a random variable, we have the flexibility to define different types of privacy guarantees by specifying requirements on various statistics of the information leakage. The resulting framework is then versatile enough to be applied to a wide range of problems. We develop this idea in Chapter 5.

## 4.4 PML on General Alphabets

In this section, we extend the gain function model of leakage to obtain a universal definition of PML that requires no assumptions about the underlying probability spaces.

Let  $X$  and  $Y$  be random variables taking values in measurable spaces  $(\mathcal{X}, \mathcal{S}_X)$  and  $(\mathcal{Y}, \mathcal{S}_Y)$ . Let  $P_{XY}$  denote the joint distribution of  $X$  and  $Y$ , which is a probability measure on the product space  $(\mathcal{X} \times \mathcal{Y}, \mathcal{S}_X \otimes \mathcal{S}_Y)$ .

**Definition 4.13.** Given a joint distribution  $P_{XY}$  on  $(\mathcal{X} \times \mathcal{Y}, \mathcal{S}_X \otimes \mathcal{S}_Y)$ , we define the pointwise maximal leakage from  $X$  to  $y \in \mathcal{Y}$  as

$$\ell_{P_{XY}}(X \rightarrow y) := \log \sup_{\substack{(\mathcal{W}, \mathcal{S}_W), \\ g \in \mathcal{G}}} \frac{\sup_{P_{W|Y}} \mathbb{E}[g(X, W) \mid Y = y]}{\sup_{w \in \mathcal{W}} \mathbb{E}[g(X, w)]}, \quad (4.10)$$

where the supremum in the numerator is over all transition probability kernels  $P_{W|Y}$  from  $(\mathcal{Y}, \mathcal{S}_Y)$  into  $(\mathcal{W}, \mathcal{S}_W)$ , and  $\mathcal{G}$  denotes the set of all gain functions defined as

$$\mathcal{G} := \left\{ g \in (\mathcal{S}_X \otimes \mathcal{S}_W)_+ \mid \sup_{w \in \mathcal{W}} \mathbb{E}[g(X, w)] < \infty \right\}.$$

The above definition models an adversary who is interested in constructing an estimate of  $X$ , denoted by  $W$ , maximizing the expected value of her gain function.  $W$  is a random variable taking values in  $(\mathcal{W}, \mathcal{S}_W)$  and gain functions are picked from the set  $\mathcal{G}$ . Then, to measure the amount of information leaking about  $X$  by disclosing an outcome  $Y = y$ , we evaluate the ratio of the adversary's expected gain given observation  $y \in \mathcal{Y}$ , and her expected gain without any observations. PML is then defined by taking the supremum of this ratio over all possible measurable spaces  $(\mathcal{W}, \mathcal{S}_W)$  and all  $g \in \mathcal{G}$ . Note that the requirement  $\sup_{w \in \mathcal{W}} \mathbb{E}[g(X, w)] < \infty$  implies that the adversary chooses a function such that her expected gain can potentially be improved upon observing  $y \in \mathcal{Y}$ .

Below, we provide two examples of gain functions that describe typical attack scenarios. The first one concerns an adversary who wishes to guess the value of a discrete function of  $X$ , denoted by  $U$ , which retrieves the setup of [63, Thm. 7]. This setup can be used to model a hypothesis-testing adversary. For example, we

may take  $U = \mathbf{1}_{A^*}(X)$  to model a binary hypothesis test for determining whether or not  $X$  is in the set  $A^* \in \mathcal{S}_X$ . The second example describes an adversary who aims to approximate the value of a random variable on a separable metric space.

**Example 4.14** (Guessing a discrete function of  $X$ ). Suppose  $U$  is a discrete random variable taking values in the set  $A$  and induced by the kernel  $P_{U|X}$ . To model an adversary who is interested in guessing the value of  $U$  we let  $\mathcal{W} = A$ , define  $\mathcal{S}_\mathcal{W}$  to be the discrete  $\sigma$ -algebra on  $A$  (i.e., its power set), and express the gain function  $g_\bullet$  as follows:

$$g_\bullet(x, w) = P_{U|X=x}(w), \quad x \in \mathcal{X}, w \in \mathcal{W}.$$

In this case, the denominator of (4.10) describes the prior probability of correctly guessing the value of  $U$  whereas the numerator represents the posterior probability of correctly guessing  $U$  given  $Y = y$ .

**Example 4.15** (Approximate guessing in metric spaces). Let  $(A, \rho)$  be a complete and separable metric space. Suppose  $U$  is a random variable taking values in  $(A, \mathcal{B}_A)$  induced by a kernel  $P_{U|X}$ , where  $\mathcal{B}_A$  denotes the Borel  $\sigma$ -algebra on  $A$ . Fix  $\varepsilon > 0$ . Our goal is to model an adversary who attempts to guess the value of  $U$  within a radius of  $\varepsilon$ . Suppose  $\mathcal{W}$  is a countable dense subset of  $A$  and  $\mathcal{S}_\mathcal{W}$  is the discrete  $\sigma$ -algebra on  $\mathcal{W}$ . Let  $B_\varepsilon(w) = \{a \in A : \rho(a, w) < \varepsilon\}$  denote the open ball of radius  $\varepsilon$  centered at  $w \in \mathcal{W}$ . Consider the gain function  $g_\sim$  defined as

$$g_\sim(x, w) = P_{U|X=x}(B_\varepsilon(w)), \quad x \in \mathcal{X}, w \in \mathcal{W}.$$

Note that the countability of  $\mathcal{W}$  ensures that  $g_\sim$  defined above is  $\mathcal{S}_X \otimes \mathcal{S}_\mathcal{W}$ -measurable. Then, for fixed  $w \in \mathcal{W}$  we have

$$\begin{aligned} \mathbb{E}[g_\sim(X, w)] &= \int P_{U|X=x}(B_\varepsilon(w)) P_X(dx) = P_U(B_\varepsilon(w)) \\ &= \mathbb{P}[U \in B_\varepsilon(w)]. \end{aligned}$$

Hence, evaluating the denominator of (4.10) with  $g_\sim$  yields the prior probability of approximately guessing  $U$ . Similarly, it can be verified that the numerator of (4.10) evaluated with  $g_\sim$  describes the posterior probability of approximately guessing  $U$  given  $Y = y$ .

Now, we show that PML in the form described by Definition 4.13 can too be written as the Rényi divergence of order infinity of the posterior distribution of  $X$  from the prior distribution of  $X$ . The proof is inspired by a result of van Erven and Harremoës [139, Thm. 2] where it is shown that the general expression for Rényi divergence can be written as the supremum of the divergence evaluated over all finite and measurable partitions of the underlying  $\sigma$ -algebra.

Theorem 4.16 requires a single assumption: that the joint distribution  $P_{XY}$  can be disintegrated into the marginal  $P_Y$  and a transition probability kernel  $P_{X|Y}$  from  $(\mathcal{Y}, \mathcal{S}_\mathcal{Y})$  into  $(\mathcal{X}, \mathcal{S}_X)$ . We discuss this assumption in Remark 4.17.

**Theorem 4.16.** *Suppose there exists a transition probability kernel  $P_{X|Y}$  from  $(\mathcal{Y}, \mathcal{S}_y)$  into  $(\mathcal{X}, \mathcal{S}_x)$  such that  $P_{XY}(dx, dy) = P_Y(dy)P_{X|Y=y}(dx)$ . Then, the point-wise maximal leakage from  $X$  to  $y \in \mathcal{Y}$  is*

$$\ell_{P_{XY}}(X \rightarrow y) = D_\infty(P_{X|Y=y} \| P_X).$$

*Proof.* Fix  $y \in \mathcal{Y}$ . We begin by simplifying the numerator of (4.10) by showing that

$$\sup_{P_{W|Y}} \mathbb{E}[g(X, W) | Y = y] = \sup_{w \in \mathcal{W}} \int_{\mathcal{X}} g(x, w) P_{X|Y=y}(dx), \quad (4.11)$$

where the supremum in the LHS of (4.11) is over all kernels from  $(\mathcal{Y}, \mathcal{S}_y)$  into  $(\mathcal{W}, \mathcal{S}_w)$ . Fix an arbitrary kernel  $P_{W|Y}$ . We write

$$\begin{aligned} \mathbb{E}[g(X, W) | Y = y] &= \int_{\mathcal{X} \times \mathcal{W}} g(x, w) P_{XW|Y=y}(dx, dw) \\ &= \int_{\mathcal{W}} P_{W|Y=y}(dw) \int_{\mathcal{X}} g(x, w) P_{X|Y=y}(dx) \\ &\leq \int_{\mathcal{W}} P_{W|Y=y}(dw) \left( \sup_{w \in \mathcal{W}} \int_{\mathcal{X}} g(x, w) P_{X|Y=y}(dx) \right) \\ &= \sup_{w \in \mathcal{W}} \int_{\mathcal{X}} g(x, w) P_{X|Y=y}(dx). \end{aligned}$$

Taking the supremum over all kernels  $P_{W|Y}$  we get

$$\sup_{P_{W|Y}} \mathbb{E}[g(X, W) | Y = y] \leq \sup_{w \in \mathcal{W}} \int_{\mathcal{X}} g(x, w) P_{X|Y=y}(dx).$$

To show the reverse inequality, fix an arbitrary  $a < \sup_{w \in \mathcal{W}} \int_{\mathcal{X}} g(x, w) P_{X|Y=y}(dx)$ . Then, there exists  $w' \in \mathcal{W}$  such that  $\int_{\mathcal{X}} g(x, w') P_{X|Y=y}(dx) \geq a$ . Let  $\delta_w$  denote the Dirac measure defined by

$$\delta_w(A) = \begin{cases} 1 & \text{if } w \in A, \\ 0 & \text{if } w \notin A, \end{cases}$$

for each  $A \in \mathcal{S}_w$ . We can write

$$\begin{aligned} \sup_{P_{W|Y}} \mathbb{E}[g(X, W) | Y = y] &\geq \int_{\mathcal{X}} P_{X|Y=y}(dx) \int_{\mathcal{W}} g(x, w) \delta_{w'}(dw) \\ &= \int_{\mathcal{X}} g(x, w') P_{X|Y=y}(dx) \\ &\geq a. \end{aligned}$$

Then, letting  $a \rightarrow \sup_{w \in \mathcal{W}} \int_{\mathcal{X}} g(x, w) P_{X|Y=y}(dx)$  we obtain (4.11).

Suppose  $P_{X|Y=y} \ll P_X$ . For notational convenience, let  $f := \frac{dP_{X|Y=y}}{dP_X}$  denote the Radon-Nikodym derivative of  $P_{X|Y=y}$  with respect to  $P_X$ . First, we show that  $\ell_{P_{XY}}(X \rightarrow y) \leq D_\infty(P_{X|Y=y} \| P_X)$ . Fix an arbitrary measurable space  $(\mathcal{W}, \mathcal{S}_\mathcal{W})$ , and a gain function  $g \in \mathcal{G}$ . We can write

$$\begin{aligned} \frac{\sup_{P_{W|Y}} \mathbb{E}[g(X, W) \mid Y = y]}{\sup_{w \in \mathcal{W}} \mathbb{E}[g(X, w)]} &= \frac{\sup_{w \in \mathcal{W}} \int_{\mathcal{X}} g(x, w) P_{X|Y=y}(dx)}{\sup_{w \in \mathcal{W}} \int_{\mathcal{X}} g(x, w) P_X(dx)} \\ &\leq \sup_{w \in \mathcal{W}} \frac{\int_{\mathcal{X}} g(x, w) P_{X|Y=y}(dx)}{\int_{\mathcal{X}} g(x, w) P_X(dx)} \\ &= \sup_{w \in \mathcal{W}} \frac{\int_{\mathcal{X}} g(x, w) f(x) P_X(dx)}{\int_{\mathcal{X}} g(x, w) P_X(dx)} \\ &\leq \operatorname{ess\,sup}_{P_X} f \\ &= \exp\left(D_\infty(P_{X|Y=y} \| P_X)\right). \end{aligned}$$

Thus,  $\ell_{P_{XY}}(X \rightarrow y) \leq D_\infty(P_{X|Y=y} \| P_X)$ .

Now, we show that  $\ell_{P_{XY}}(X \rightarrow y) \geq D_\infty(P_{X|Y=y} \| P_X)$ . Since  $P_{X|Y=y} \ll P_X$  we may, without loss of generality, assume that  $f(x) < \infty$  for all  $x \in \mathcal{X}$ . Let  $\mathcal{W} = \mathbb{Z} \cup \{-\infty\}$ , and suppose  $\mathcal{S}_\mathcal{W}$  is the discrete  $\sigma$ -algebra on  $\mathcal{W}$ . Fix  $\varepsilon > 0$  and consider the following (countable and disjoint) partition of  $\mathcal{X}$ :

$$B_w^\varepsilon = \{x \in \mathcal{X} : e^{w\varepsilon} \leq f(x) < e^{(w+1)\varepsilon}\}, \quad w \in \mathcal{W}, \quad (4.12)$$

which is indexed by  $\mathcal{W}$ . Note that since  $f$  is  $\mathcal{S}_\mathcal{X}$ -measurable, then  $B_w^\varepsilon \in \mathcal{S}_\mathcal{X}$  for all  $w \in \mathcal{W}$ . Let us define the gain function  $g^* : \mathcal{X} \times \mathcal{W} \rightarrow \mathbb{R}_+$  as follows:

$$g^*(x, w) = \begin{cases} \frac{1}{P_X(B_w^\varepsilon)} \mathbf{1}_{B_w^\varepsilon}(x) & \text{if } P_X(B_w^\varepsilon) > 0, \\ 0 & \text{if } P_X(B_w^\varepsilon) = 0. \end{cases}$$

Then, we can write

$$\begin{aligned} \exp\left(\ell_{P_{XY}}(X \rightarrow y)\right) &\geq \frac{\sup_{P_{W|Y}} \mathbb{E}[g^*(X, W) \mid Y = y]}{\sup_{w \in \mathcal{W}} \mathbb{E}[g^*(X, w)]} \\ &= \frac{\sup_{w \in \mathcal{W}} \int_{\mathcal{X}} g^*(x, w) P_{X|Y=y}(dx)}{\sup_{w \in \mathcal{W}} \int_{\mathcal{X}} g^*(x, w) P_X(dx)} \\ &= \frac{\sup_{w \in \mathcal{W}: P_X(B_w^\varepsilon) > 0} \int_{\mathcal{X}} \frac{1}{P_X(B_w^\varepsilon)} \mathbf{1}_{B_w^\varepsilon}(x) P_{X|Y=y}(dx)}{\sup_{w \in \mathcal{W}: P_X(B_w^\varepsilon) > 0} \int_{\mathcal{X}} \frac{1}{P_X(B_w^\varepsilon)} \mathbf{1}_{B_w^\varepsilon}(x) P_X(dx)} \\ &= \frac{\sup_{w \in \mathcal{W}: P_X(B_w^\varepsilon) > 0} \frac{P_{X|Y=y}(B_w^\varepsilon)}{P_X(B_w^\varepsilon)}}{\sup_{w \in \mathcal{W}: P_X(B_w^\varepsilon) > 0} \frac{P_X(B_w^\varepsilon)}{P_X(B_w^\varepsilon)}} \end{aligned}$$

$$= \sup_{w \in \mathcal{W}: P_X(B_w^\varepsilon) > 0} \frac{P_{X|Y=y}(B_w^\varepsilon)}{P_X(B_w^\varepsilon)} \quad (4.13a)$$

$$= \operatorname{ess\,sup}_{P_X} \bar{f}, \quad (4.13b)$$

where

$$\bar{f}(x) := \sum_{w \in \mathcal{W}} \frac{P_{X|Y=y}(B_w^\varepsilon)}{P_X(B_w^\varepsilon)} \mathbf{1}_{B_w^\varepsilon}(x), \quad x \in \mathcal{X}.$$

In (4.13b), we have written (4.13a) as a function of  $x$ . We have replaced the supremum over  $w$  in (4.13a) with the essential supremum over  $x$  in (4.13b) because  $\bar{f}$  is constant on each set  $B_w^\varepsilon$ . Note that  $\bar{f}(x) < \infty$  for all  $x \in \mathcal{X}$  even if there exists  $w \in \mathcal{W}$  such that  $P_X(B_w^\varepsilon) = 0$ . This is because  $P_{X|Y=y} \ll P_X$  and we use the convention that  $0/0 = 1$ .

Let  $\mathcal{F} := \sigma\{B_w^\varepsilon\}$  denote the  $\sigma$ -algebra on  $\mathcal{X}$  generated by the collection of sets  $\{B_w^\varepsilon\}$ . We argue that  $\bar{f}$  is (a version of) the conditional expectation of  $f$  given  $\mathcal{F}$ , that is,  $\bar{f} = \mathbb{E}[f | \mathcal{F}]$ . Clearly,  $\bar{f}$  is  $\mathcal{F}$ -measurable, so we should verify that  $\int_A f dP_X = \int_A \bar{f} dP_X$  for all  $A \in \mathcal{F}$ . It is, however, sufficient to verify this equality for  $A = B_w^\varepsilon$  because each non-empty set in  $\mathcal{F}$  can be written as a countable union of sets in  $\{B_w^\varepsilon\}$  and the monotone convergence theorem ensures that  $\int_{\cup_i C_i} f dP_X = \sum_i \int_{C_i} f dP_X$  for each countable collection of disjoint sets  $\{C_i\}$  in  $\mathcal{F}$ . Thus, by noting that

$$\int_{B_w^\varepsilon} \bar{f} dP_X = P_{X|Y=y}(B_w^\varepsilon) = \int_{B_w^\varepsilon} f dP_X,$$

for all  $w \in \mathcal{W}$  we conclude that  $\bar{f} = \mathbb{E}[f | \mathcal{F}]$ .

Finally, we can write

$$\begin{aligned} \ell_{P_{XY}}(X \rightarrow y) &\geq \log \operatorname{ess\,sup}_{P_X} \mathbb{E}[f | \mathcal{F}] \\ &\geq \log \left( \left( \operatorname{ess\,sup}_{P_X} f \right) e^{-\varepsilon} \right) \\ &= \log \operatorname{ess\,sup}_{P_X} f - \varepsilon \\ &= D_\infty(P_{X|Y=y} \| P_X) - \varepsilon, \end{aligned} \quad (4.14a)$$

where (4.14a) is due to the fact that by the definition of the sets  $\{B_w^\varepsilon\}$  in (4.12),  $\mathbb{E}[f | \mathcal{F}]$  never differs from  $f$  by more than a factor of  $e^\varepsilon$ . Then, letting  $\varepsilon \rightarrow 0$ , we obtain  $\ell_{P_{XY}}(X \rightarrow y) \geq D_\infty(P_{X|Y=y} \| P_X)$ , which completes the proof for the case  $P_{X|Y=y} \ll P_X$ .

On the other hand, if  $P_{X|Y=y} \not\ll P_X$  then there exists  $A_0 \in \mathcal{S}_X$  such that  $P_X(A_0) = 0$  and  $P_{X|Y=y}(A_0) > 0$ . Let  $(\mathcal{W}, \mathcal{S}_W)$  be an arbitrary measurable space, and consider the gain function  $g(x, w) = \mathbf{1}_{A_0}(x)$  for all  $w \in \mathcal{W}$ . Then, it is easy

to see that  $\mathbb{E}[g(X, W) \mid Y = y] = P_{X|Y=y}(A_0) > 0$  for all kernels  $P_{W|Y}$  while  $\sup_{w \in \mathcal{W}} \mathbb{E}[g(X, w)] = 0$ . Hence,  $\ell_{P_{XY}}(X \rightarrow y) = D_\infty(P_{X|Y=y} \| P_X) = \infty$ , as desired.  $\square$

*Remark 4.17.* Theorem 4.16 assumes that the joint distribution  $P_{XY}$  can be disintegrated into the marginal  $P_Y$  and a kernel  $P_{X|Y}$ . This can be achieved in different ways. For example, we may start with a distribution  $P_Y$  and a kernel  $P_{X|Y}$  and construct  $P_{XY}$  such that it satisfies  $P_{XY}(dx, dy) = P_Y(dy)P_{X|Y=y}(dx)$ . Otherwise, we may assume that  $(\mathcal{X}, \mathcal{S}_X)$  is a standard Borel space [79, Def. 8.35] in which case the existence of a regular version of the conditional probability  $\mathbb{P}[\cdot \mid Y]$  restricted to  $\sigma X \subset \mathcal{H}$  is guaranteed [24, Thm. IV.2.18]. In this latter case,  $P_{X|Y}$  is any kernel satisfying  $\mathbb{P}[X \in A \mid Y](\omega) = P_{X|Y=Y(\omega)}(A)$  for  $\mathbb{P}$ -almost all  $\omega \in \Omega$  and all  $A \in \mathcal{S}_X$ . Hence, Theorem 4.16 requires that  $P_{XY}$  can be disintegrated into a marginal distribution and a kernel, though it is immaterial how this is actually achieved. For a detailed discussion on disintegration theorems and the existence of regular conditional probabilities see [46].

Equipped with Theorem 4.16, we can calculate the information leaking from an observation  $y \in \mathcal{Y}$ . However, this result alone is insufficient for obtaining an information leakage random variable  $\ell_{P_{XY}}(X \rightarrow Y)$ . The difficulty is that the mapping  $y \mapsto \ell_{P_{XY}}(X \rightarrow y)$  must be  $\mathcal{S}_Y$ -measurable and there are certain nuances associated with this task. For example, we need to ensure that if  $P_{X|Y=y} \ll P_X$ , then the Radon-Nikodym derivative  $\frac{dP_{X|Y=y}}{dP_X}$  is jointly measurable in  $(x, y)$ , or that the set  $\{y \in \mathcal{Y} : P_{X|Y=y} \ll P_X\}$  is measurable.

To obtain a measurable version of  $\ell_{P_{XY}}(X \rightarrow y)$  we use the pragmatic assumption that the joint distribution  $P_{XY}$  is absolutely continuous with respect to the product of two  $\sigma$ -finite measures on  $(\mathcal{X}, \mathcal{S}_X)$  and  $(\mathcal{Y}, \mathcal{S}_Y)$  [113, Sec. 2.6]. This assumption also has the advantage of guaranteeing that  $P_{XY}(dx, dy) = P_Y(dy)P_{X|Y=y}(dx)$  holds.

**Proposition 4.18** (Information leakage random variable). *Suppose  $P_{XY}$  is a probability measure on  $(\mathcal{X} \times \mathcal{Y}, \mathcal{S}_X \otimes \mathcal{S}_Y)$  satisfying*

$$P_{XY}(dx, dy) = p(x, y) \mu(dx) \nu(dy), \quad x \in \mathcal{X}, y \in \mathcal{Y},$$

where  $\mu$  and  $\nu$  are  $\sigma$ -finite measures on  $(\mathcal{X}, \mathcal{S}_X)$  and  $(\mathcal{Y}, \mathcal{S}_Y)$ , respectively, and  $p \in (\mathcal{S}_X \otimes \mathcal{S}_Y)_+$ . Then, there exists a transition probability kernel  $P_{X|Y}$  from  $(\mathcal{Y}, \mathcal{S}_Y)$  into  $(\mathcal{X}, \mathcal{S}_X)$  such that

- (i)  $P_{XY}(dx, dy) = P_Y(dy) P_{X|Y=y}(dx)$ ;
- (ii)  $P_{X|Y=y} \ll P_X$  for  $\nu$ -almost all  $y \in \mathcal{Y}$ ; and
- (iii) the mapping  $y \mapsto \ell_{P_{XY}}(X \rightarrow y)$  is  $\mathcal{S}_Y$ -measurable.

Proposition 4.18 is proved in Appendix 4.B.

*Remark 4.19.* The assumption of Proposition 4.18 guarantees that  $P_{XY} \ll P_X \times P_Y$ , where  $P_X \times P_Y$  is the product of  $P_X$  and  $P_Y$ . This assumption also allows us to write PML in different forms using densities:

$$\begin{aligned} \ell_{P_{XY}}(X \rightarrow y) &= \operatorname{ess\,sup}_{P_X} i_{P_{XY}}(X; y) \\ &= \log \left( \operatorname{ess\,sup}_{P_X} \frac{f_{X|Y}(X, y)}{f_X(X)} \right) \\ &= \log \left( \operatorname{ess\,sup}_{P_X} \frac{f_{Y|X}(y, X)}{f_Y(y)} \right), \end{aligned}$$

where  $i_{P_{XY}}(X; Y) = \log \frac{dP_{XY}}{d(P_X \times P_Y)}(X, Y)$  is the information density,  $f_{X|Y} \in (\mathcal{S}_X \otimes \mathcal{S}_Y)_+$  denotes the density of  $P_{X|Y}$  with respect to  $\mu$ , and  $f_X \in \mathcal{S}_X$  denotes the density of  $P_X$  with respect to  $\mu$ . Densities  $f_{Y|X}$  and  $f_Y$  are defined similarly.

Finally, we calculate PML in some examples involving discrete or absolutely continuous random variables.

**Example 4.20** (Poisson and Binomial distributions). Suppose  $X \sim \operatorname{Pois}(\lambda p)$ , where  $\lambda > 1$ ,  $p \in (0, 1)$ , and  $\lambda(1 - p) < 1$ . Assume  $Y$  is defined through the kernel

$$P_{Y|X=x}(y) = \begin{cases} \frac{(\lambda(1-p))^{y-x} e^{-\lambda(1-p)}}{(y-x)!} & \text{if } y \geq x, \\ 0 & \text{if } y < x, \end{cases}$$

where  $x \in \mathbb{N}$ . It can be easily verified that  $X | Y = y \sim \operatorname{Binom}(y, p)$ . Hence, the PML from  $X$  to  $y \in \mathbb{N}$  is given by

$$\begin{aligned} \ell_{P_{XY}}(X \rightarrow y) &= \log \sup_{x \in \mathbb{N}} \frac{P_{X|Y=y}(x)}{P_X(x)} \\ &= \log \max_{x \in \{0, \dots, y\}} \frac{P_{X|Y=y}(x)}{P_X(x)} \\ &= \log (e^{\lambda p} \lambda^{-y} y!). \end{aligned}$$

**Example 4.21** (Geometric distribution). Suppose  $X \sim \operatorname{Geom}(p)$  with  $p \in (0, 1)$ . Let  $Y$  be a binary random variable defined through the kernel  $P_{Y|X=x}(0) = 1 - P_{Y|X=x}(1) = q^x$  with  $q \in (0, 1)$  and  $x \in \mathbb{N}^*$ . Then,  $P_Y(0) = 1 - P_Y(1) = \frac{pq}{1-q+pq}$ , and

$$\begin{aligned} \ell_{P_{XY}}(X \rightarrow 0) &= \log \sup_{x \in \mathbb{N}^*} \frac{P_{Y|X=x}(0)}{P_Y(0)} = \log \frac{1-q+pq}{p}, \\ \ell_{P_{XY}}(X \rightarrow 1) &= \log \sup_{x \in \mathbb{N}^*} \frac{P_{Y|X=x}(1)}{P_Y(1)} = \log \frac{1-q+pq}{1-q}. \end{aligned}$$

**Example 4.22** (Additive Gaussian Noise). Suppose  $Y = X + N$  where  $X \sim \mathcal{N}(0, \sigma_X^2)$ ,  $N \sim \mathcal{N}(0, \sigma_N^2)$ , and  $X$  and  $N$  are independent. The PML from  $X$  to  $y \in \mathbb{R}$  is given by

$$\begin{aligned} \ell_{P_{XY}}(X \rightarrow y) &= \log \sup_{x \in \mathbb{R}} \frac{f_{Y|X}(y, x)}{f_Y(y)} \\ &= \frac{1}{2} \log \left( 1 + \frac{\sigma_X^2}{\sigma_N^2} \right) + \frac{y^2}{2(\sigma_X^2 + \sigma_N^2)}. \end{aligned}$$

As expected, for fixed  $y \in \mathbb{R}$  and  $\sigma_X^2$ , taking  $\sigma_N^2 \rightarrow \infty$  implies  $\ell_{P_{XY}}(X \rightarrow y) \rightarrow 0$ .

**Example 4.23** (Gaussian mixtures). Suppose  $X \sim \text{Ber}(\frac{1}{2})$  is an equiprobable Bernoulli random variable, and  $Y | X = x \sim \mathcal{N}(x, \sigma^2)$  has Gaussian distribution with mean  $x \in \{0, 1\}$  and variance  $\sigma^2$ . The PML from  $X$  to each  $y \in \mathbb{R}$  can be computed as

$$\begin{aligned} \ell_{P_{XY}}(X \rightarrow y) &= \log \max_{x \in \{0, 1\}} \frac{f_{Y|X}(y, x)}{f_Y(y)} \\ &= \log \frac{2}{\exp\left(-\frac{|y-\frac{1}{2}|}{\sigma^2}\right) + 1}. \end{aligned}$$

Specifically,  $\ell_{P_{XY}}(X \rightarrow \frac{1}{2}) = 0$  and  $\lim_{y \rightarrow \infty} \ell_{P_{XY}}(X \rightarrow y) = \lim_{y \rightarrow -\infty} \ell_{P_{XY}}(X \rightarrow y) = \log 2$ .

**Example 4.24** (Bivariate Gaussian). Suppose  $X$  and  $Y$  are zero-mean jointly Gaussian random variables with variances  $\sigma_X^2$  and  $\sigma_Y^2$ , respectively, and correlation coefficient  $\rho \in (-1, 1)$ . Then,  $Y | X = x \sim \mathcal{N}(\frac{\sigma_Y}{\sigma_X} \rho x, (1 - \rho^2)\sigma_Y^2)$ , and the PML from  $X$  to  $y \in \mathbb{R}$  is

$$\ell_{P_{XY}}(X \rightarrow y) = \begin{cases} \frac{y^2}{2\sigma_Y^2} - \frac{1}{2} \log(1 - \rho^2) & \text{if } \rho \neq 0, \\ 0 & \text{if } \rho = 0. \end{cases}$$

---

# Appendices

---

## 4.A Proof of Theorem 4.10

Suppose without loss of generality that  $P_X$  has full support. Given an arbitrary gain function  $g$ , the  $g$ -leakage of  $X$  can be written as

$$\begin{aligned}
 \ell_g(X \rightarrow y) &= \log \frac{\sup_{P_{W|Y}} \mathbb{E}[g(X, W) \mid Y = y]}{\max_{w \in \mathcal{W}} \mathbb{E}[g(X, w)]} \\
 &= \log \frac{\sup_{P_{W|Y}} \sum_{x \in \mathcal{X}} \sum_{w \in \mathcal{W}} g(x, w) P_{X|Y=y}(x) P_{W|Y=y}(w)}{\max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} g(x, w) P_X(x)} \\
 &= \log \frac{\max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} g(x, w) P_{X|Y=y}(x)}{\max_{w \in \mathcal{W}} \sum_{x \in \mathcal{X}} g(x, w) P_X(x)}, \tag{4.15}
 \end{aligned}$$

where the last equality follows by plugging in  $P_{W|Y}^*$  satisfying

$$P_{W|Y=y}^*(w) := \begin{cases} 1 & \text{for some } w \in \arg \max_w \sum_{x \in \mathcal{X}} g(x, w) P_{X|Y=y}(x) \\ 0 & \text{otherwise,} \end{cases}$$

in the numerator. Furthermore, as shown in the proof of Theorem 4.2, given a randomized function of  $X$  denoted by  $U$ , the  $U$ -leakage of  $X$  can be expressed as

$$\begin{aligned}
 \ell_U(X \rightarrow y) &= \log \frac{\max_{u \in \mathcal{U}} P_{U|Y=y}(u)}{\max_{u \in \mathcal{U}} P_U(u)} \\
 &= \log \frac{\max_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} P_{U|X=x}(u) P_{X|Y=y}(x)}{\max_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} P_{U|X=x}(u) P_X(x)}. \tag{4.16}
 \end{aligned}$$

To show the equivalence, first, we prove the simpler direction by showing that each  $U$ -leakage can be written as a  $g$ -leakage. Given an arbitrary randomized function of  $X$  denoted by  $U$ , define  $\mathcal{W}_U := \mathcal{U}$  such that each  $u \in \mathcal{U}$  corresponds uniquely to some  $w_u \in \mathcal{W}_U$ , and let  $g_U(x, w_u) := P_{U|X=x}(u)$  for all  $x \in \mathcal{X}$  and  $u \in \mathcal{U}$ . By computing expressions (4.15) and (4.16), it is easy to see that  $\ell_U(X \rightarrow y) = \ell_{g_U}(X \rightarrow y)$ . This construction implies that a randomized function of  $X$  is simply a gain function that satisfies  $\sum_w g_U(x, w) = 1$ , for all  $x \in \mathcal{X}$ , that is, the total gain associated with each secret  $x \in \mathcal{X}$  is a constant.

Now, we show that each  $g$ -leakage can be written as a  $U$ -leakage. Fix an arbitrary gain function  $g$ . Without loss of generality, suppose  $g(x, w) \leq 1$  for all

$x \in \mathcal{X}$  and  $w \in \mathcal{W}$  (this can be achieved by normalizing the gain function by  $\max_{x,w} g(x, w)$ ). In what follows, we construct a randomized function of  $X$  using a channel that generalizes the shattering channel of Definition 4.3. We need to consider the following two cases:

**Case 1: The same  $w$  maximizes the numerator and the denominator in (4.15).**

Here, we construct a randomized function of  $X$ , denoted by  $S$ , which is described by the kernel  $P_{S|X}$  and satisfies  $\ell_S(X \rightarrow y) = \ell_g(X \rightarrow y)$ . Let  $w_S \in \arg \max_w \sum_x g(x, w) P_{X|Y=y}(x)$  which (following from the definition of Case 1) also satisfies  $w_S \in \arg \max_w \sum_x g(x, w) P_X(x)$ . Informally,  $w_S$  denotes the adversary's best guess both after observing  $Y = y$ , and with no observations.

Define the set  $\mathcal{X}_S := \{x \in \mathcal{X} : g(x, w_S) > 0\}$ . For now, let us assume that  $\mathcal{X}_S = \mathcal{X}$  but later we will discuss how the proof can be adapted if  $\mathcal{X}_S$  is a proper subset of  $\mathcal{X}$ . For each  $x \in \mathcal{X}_S$ , let  $k_S(x) := 1/g(x, w_S)$ , and define  $k_S := \max_{x \in \mathcal{X}_S} k_S(x)$ . Roughly speaking,  $k_S(x)$  is the cardinality of the support of  $P_{S|X=x}$  while  $k_S$  is the cardinality of the support of  $P_S$ .

Now, we construct  $S$  taking values in an alphabet  $\mathcal{S}$  such that  $|\mathcal{S}| = \lceil k_S \rceil$ . For all  $x \in \mathcal{X}_S$ , the kernel  $P_{S|X=x}$  is defined by

$$P_{S|X=x}(s_i) := \begin{cases} g(x, w_S) & \text{if } 1 \leq i \leq \lfloor k_S(x) \rfloor, \\ 1 - \lfloor k_S(x) \rfloor g(x, w_S) & \text{if } i = \lceil k_S(x) \rceil, \\ 0 & \text{if } \lceil k_S(x) \rceil + 1 \leq i \leq \lceil k_S \rceil. \end{cases}$$

Informally, for each  $x$ , the above kernel allocates chunks of probability equal to  $g(x, w_S)$  to the first  $\lfloor k_S(x) \rfloor$  letters, and the  $\lceil k_S(x) \rceil$ -th letter is used to contain the remaining probability  $1 - \lfloor k_S(x) \rfloor g(x, w_S)$ . Note that the above kernel indeed satisfies

$$\sum_{i=1}^{\lceil k_S \rceil} P_{S|X=x}(s_i) = 1,$$

for all  $x \in \mathcal{X}_S$ .

Renaming  $S$  to  $U_g$ , we verify that the random variable constructed above satisfies  $\ell_{U_g}(X \rightarrow y) = \ell_g(X \rightarrow y)$ :

$$\begin{aligned} \max_{u \in \mathcal{U}_g} \sum_x P_{U_g|X=x}(u) P_{X|Y=y}(x) &= \sum_x g(x, w_{U_g}) P_{X|Y=y}(x) \\ &= \max_w \sum_x g(x, w) P_{X|Y=y}(x), \end{aligned}$$

where the last equality follows from the definition of  $w_{U_g}$ . Similarly, we have

$$\max_{u \in \mathcal{U}_g} \sum_x P_{U_g|X=x}(u) P_X(x) = \sum_x g(x, w_{U_g}) P_X(x)$$

$$= \max_w \sum_x g(x, w) P_X(x).$$

**Case 2: The maximizing  $w$ 's in the numerator and the denominator of (4.15) are different.**

We construct two randomized functions of  $X$  denoted by  $S$  and  $T$ , one for the numerator of (4.15) and one for the denominator. Let

$$w_S \in \arg \max_w \sum_x g(x, w) P_{X|Y=y}(x),$$

$$w_T \in \arg \max_w \sum_x g(x, w) P_X(x),$$

and define

$$\mathcal{X}_S := \{x \in \mathcal{X}: g(x, w_S) > 0\},$$

$$\mathcal{X}_T := \{x \in \mathcal{X}: g(x, w_T) > 0\},$$

where  $w_S$  denotes the adversary's best guess having observed  $Y = y$ , and  $w_T$  denotes the adversary's best guess without an observation. We need to consider the following two cases:

**Case 2.1:**  $\mathcal{X}_S = \mathcal{X}_T$ .

Let  $\mathcal{X}_{U_g} := \mathcal{X}_S = \mathcal{X}_T$ . Once again, let us assume that  $\mathcal{X}_{U_g} = \mathcal{X}$ . Similarly to what we had in Case 1, for all  $x \in \mathcal{X}_{U_g}$  we define

$$k_S(x) := \frac{1}{g(x, w_S)}, \quad k_S := \max_{x \in \mathcal{X}_{U_g}} k_S(x),$$

$$k_T(x) := \frac{1}{g(x, w_T)}, \quad k_T := \max_{x \in \mathcal{X}_{U_g}} k_T(x).$$

Let  $\mathcal{S}$  denote the support set of random variable  $S$ , and  $\mathcal{T}$  denote the support set of random variable  $T$ , where  $|\mathcal{S}| = \lceil k_S \rceil$  and  $|\mathcal{T}| = \lceil k_T \rceil$ . For all  $x \in \mathcal{X}_{U_g}$ , the kernels  $P_{S|X=x}$  and  $P_{T|X=x}$  are defined as

$$P_{S|X=x}(s_i) := \begin{cases} g(x, w_S) & \text{if } 1 \leq i \leq \lfloor k_S(x) \rfloor, \\ 1 - \lfloor k_S(x) \rfloor g(x, w_S) & \text{if } i = \lceil k_S(x) \rceil, \\ 0 & \text{if } \lceil k_S(x) \rceil + 1 \leq i \leq \lfloor k_S \rfloor, \end{cases}$$

and

$$P_{T|X=x}(t_j) := \begin{cases} g(x, w_T) & \text{if } 1 \leq j \leq \lfloor k_T(x) \rfloor, \\ 1 - \lfloor k_T(x) \rfloor g(x, w_T) & \text{if } j = \lceil k_T(x) \rceil, \\ 0 & \text{if } \lceil k_T(x) \rceil + 1 \leq j \leq \lceil k_T \rceil. \end{cases}$$

Finally, we define  $U_g$  as the Bernoulli mixture of  $S$  and  $T$ . Let  $\mathcal{U}_g := \mathcal{S} \cup \mathcal{T}$  denote the alphabet of  $U_g$ . For all  $x \in \mathcal{X}_{U_g}$ , we define  $P_{U_g|X=x}(u) := \frac{1}{2}P_{S|X=x}(u) + \frac{1}{2}P_{T|X=x}(u)$ , where  $P_{S|X=x}(u) = 0$  for  $u \in \mathcal{T}$  and  $P_{T|X=x}(u) = 0$  for  $u \in \mathcal{S}$ .<sup>3</sup> Let us verify that  $U_g$  satisfies  $\ell_{U_g}(X \rightarrow y) = \ell_g(X \rightarrow y)$ :

$$\begin{aligned} \max_{u \in \mathcal{U}_g} \sum_x P_{U_g|X=x}(u) P_{X|Y=y}(x) &= \sum_x \frac{1}{2} g(x, w_S) P_{X|Y=y}(x) \\ &= \frac{1}{2} \max_w \sum_x g(x, w) P_{X|Y=y}(x), \end{aligned}$$

and also,

$$\begin{aligned} \max_{u \in \mathcal{U}_g} \sum_x P_{U_g|X=x}(u) P_X(x) &= \sum_x \frac{1}{2} g(x, w_T) P_X(x) \\ &= \frac{1}{2} \max_w \sum_x g(x, w) P_X(x). \end{aligned}$$

Thus, we have

$$\begin{aligned} \ell_{U_g}(X \rightarrow y) &= \log \frac{\max_{u \in \mathcal{U}_g} \sum_x P_{U_g|X=x}(u) P_{X|Y=y}(x)}{\max_{u \in \mathcal{U}_g} \sum_x P_{U_g|X=x}(u) P_X(x)} \\ &= \log \frac{\frac{1}{2} \max_w \sum_x g(x, w) P_{X|Y=y}(x)}{\frac{1}{2} \max_w \sum_x g(x, w) P_X(x)} \\ &= \ell_g(X \rightarrow y). \end{aligned}$$

**Case 2.2:**  $\mathcal{X}_S \neq \mathcal{X}_T$ .

Let  $n_S$  and  $n_T$  be positive integers. Here, the idea is that we increase the sizes of the sets  $\mathcal{S}$  and  $\mathcal{T}$  by  $n_S$  and  $n_T$ , respectively, where these extra letters are used to support those  $x$ 's for which we either have  $g(x, w_S) = 0$  or  $g(x, w_T) = 0$ . We need to distinguish between three types of  $x$ 's:

(i) For  $x \in \mathcal{X}_S \cap \mathcal{X}_T$  we define

$$P_{S|X=x}(s_i) := \begin{cases} g(x, w_S) & \text{if } 1 \leq i \leq \lfloor k_S(x) \rfloor, \\ 1 - \lfloor k_S(x) \rfloor g(x, w_S) & \text{if } i = \lceil k_S(x) \rceil, \\ 0 & \text{if } \lceil k_S(x) \rceil + 1 \leq i \leq \lceil k_S(x) \rceil + n_S, \end{cases}$$

and

$$P_{T|X=x}(t_j) := \begin{cases} g(x, w_T) & \text{if } 1 \leq j \leq \lfloor k_T(x) \rfloor, \\ 1 - \lfloor k_T(x) \rfloor g(x, w_T) & \text{if } j = \lceil k_T(x) \rceil, \\ 0 & \text{if } \lceil k_T(x) \rceil + 1 \leq j \leq \lceil k_T(x) \rceil + n_T. \end{cases}$$

---

<sup>3</sup>This is a slight abuse of notation. Strictly speaking,  $P_{S|X=x}(u)$  is defined only for  $u \in \mathcal{S}$  and  $P_{T|X=x}(u)$  is defined only for  $u \in \mathcal{T}$ .

(ii) For  $x \in \mathcal{X}_S \setminus \mathcal{X}_T$  we let

$$P_{S|X=x}(s_i) := \begin{cases} g(x, w_S) & \text{if } 1 \leq i \leq \lfloor k_S(x) \rfloor, \\ 1 - \lfloor k_S(x) \rfloor g(x, w_S) & \text{if } i = \lfloor k_S(x) \rfloor, \\ 0 & \text{if } \lfloor k_S(x) \rfloor + 1 \leq i \leq \lfloor k_S \rfloor + n_S, \end{cases}$$

and

$$P_{T|X=x}(t_j) := \begin{cases} 0 & \text{if } 1 \leq j \leq \lfloor k_T \rfloor, \\ \frac{1}{n_T} & \text{if } \lfloor k_T \rfloor + 1 \leq j \leq \lfloor k_T \rfloor + n_T. \end{cases}$$

(iii) For  $x \in \mathcal{X}_T \setminus \mathcal{X}_S$  we let

$$P_{S|X=x}(s_i) := \begin{cases} 0 & \text{if } 1 \leq i \leq \lfloor k_S \rfloor, \\ \frac{1}{n_S} & \text{if } \lfloor k_S \rfloor + 1 \leq i \leq \lfloor k_S \rfloor + n_S, \end{cases}$$

and

$$P_{T|X=x}(t_j) := \begin{cases} g(x, w_T) & \text{if } 1 \leq j \leq \lfloor k_T(x) \rfloor, \\ 1 - \lfloor k_T(x) \rfloor g(x, w_T) & \text{if } j = \lfloor k_T(x) \rfloor, \\ 0 & \text{if } \lfloor k_T(x) \rfloor + 1 \leq j \leq \lfloor k_T \rfloor + n_T. \end{cases}$$

Now, we define  $U_g$  as before. Suppose  $\mathcal{U}_g = \mathcal{S} \cup \mathcal{T}$  is the alphabet of  $U_g$ . For  $x \in \mathcal{X}_S \cup \mathcal{X}_T$  (where we are assuming  $\mathcal{X} = \mathcal{X}_S \cup \mathcal{X}_T$ ), let  $P_{U_g|X=x}(u) = \frac{1}{2}P_{S|X=x}(u) + \frac{1}{2}P_{T|X=x}(u)$ , where  $P_{S|X=x}(u) = 0$  for  $u \in \mathcal{T}$  and  $P_{T|X=x}(u) = 0$  for  $u \in \mathcal{S}$ . Then, we can write

$$\begin{aligned} & \max_{u \in \mathcal{U}_g} \sum_x P_{U_g|X=x}(u) P_{X|Y=y}(x) = \\ & \max \left\{ \sum_{x \in \mathcal{X}_S} \frac{g(x, w_S)}{2} P_{X|Y=y}(x), \sum_{x \in \mathcal{X}_T \setminus \mathcal{X}_S} \frac{1}{2n_S} P_{X|Y=y}(x) \right\}. \end{aligned}$$

By taking  $n_S$  to be large enough, we can ensure that

$$\sum_{x \in \mathcal{X}_T \setminus \mathcal{X}_S} \frac{1}{2n_S} P_{X|Y=y}(x) \leq \sum_{x \in \mathcal{X}_S} \frac{1}{2} g(x, w_S) P_{X|Y=y}(x),$$

which yields

$$\begin{aligned} \max_{u \in \mathcal{U}_g} \sum_x P_{U_g|X=x}(u) P_{X|Y=y}(x) &= \sum_{x \in \mathcal{X}_S} \frac{1}{2} g(x, w_S) P_{X|Y=y}(x) \\ &= \frac{1}{2} \max_w \sum_{x \in \mathcal{X}} g(x, w) P_{X|Y=y}(x). \end{aligned}$$

Similarly, for  $n_T$  large enough we have

$$\begin{aligned}
 & \max_{u \in \mathcal{U}_g} \sum_x P_{U_g|X=x}(u) P_X(x) \\
 &= \max \left\{ \sum_{x \in \mathcal{X}_T} \frac{1}{2} g(x, w_T) P_X(x), \sum_{x \in \mathcal{X}_S \setminus \mathcal{X}_T} \frac{1}{2n_T} P_X(x) \right\} \\
 &= \sum_{x \in \mathcal{X}_T} \frac{1}{2} g(x, w_T) P_X(x) \\
 &= \frac{1}{2} \max_w \sum_{x \in \mathcal{X}} g(x, w) P_X(x).
 \end{aligned}$$

Hence, we conclude that  $\ell_{U_g}(X \rightarrow y) = \ell_g(X \rightarrow y)$ .

The only point left to discuss is regarding the case where  $\mathcal{X}_S \cup \mathcal{X}_T$  is a proper subset of  $\mathcal{X}$ . Let  $n_O$  be a positive integer. Once again, we increase the size of the alphabet  $\mathcal{U}_g$  by  $n_O$  letters, where these extra letters are used to support the  $x$ 's in  $\mathcal{X} \setminus (\mathcal{X}_S \cup \mathcal{X}_T)$ . Hence, we let  $\mathcal{U}_g = \mathcal{S} \cup \mathcal{T} \cup \mathcal{O}$ , where  $\mathcal{O}$  is a finite set containing  $n_O$  elements. For  $x \in \mathcal{X} \setminus (\mathcal{X}_S \cup \mathcal{X}_T)$  we define the channel  $P_{U_g|X=x}$  as

$$P_{U_g|X=x}(u) = \begin{cases} \frac{1}{n_O} & \text{if } u \in \mathcal{O}, \\ 0 & \text{otherwise.} \end{cases}$$

For  $x \in \mathcal{X}_S \cup \mathcal{X}_T$ , we let  $P_{U_g|X=x}(u) = 0$  when  $u \in \mathcal{O}$ ; otherwise  $P_{U_g|X=x}(u)$  is defined as in Case 2.2. It is straightforward to verify that for  $n_O$  large enough ( $\frac{1}{n_O}$  small enough), the values of the numerator and the denominator in the expression of  $\ell_{U_g}(X \rightarrow y)$  remain as before, from which we conclude that  $\ell_{U_g}(X \rightarrow y) = \ell_g(X \rightarrow y)$ .

## 4.B Proof of Propostion 4.18

Define the functions

$$\begin{aligned}
 q(y) &:= \int_{\mathcal{X}} p(x, y) \mu(dx), \quad y \in \mathcal{Y}, \\
 r(x) &:= \int_{\mathcal{Y}} p(x, y) \nu(dy), \quad x \in \mathcal{X}, \\
 k(x, y) &:= \begin{cases} \frac{p(x, y)}{q(y)} & \text{if } q(y) > 0, \\ r(x) & \text{if } q(y) = 0, \end{cases} \quad x \in \mathcal{X}, y \in \mathcal{Y}.
 \end{aligned}$$

Let

$$P_{X|Y=y}(A) := \int_A k(x, y) \mu(dx), \quad A \in \mathcal{S}_x, y \in \mathcal{Y}.$$

It can be verified that  $P_{X|Y}$  defined above is a transition probability kernel from  $(\mathcal{Y}, \mathcal{S}_Y)$  into  $(\mathcal{X}, \mathcal{S}_X)$ .

First, we show that  $P_{X|Y=y} \ll P_X$  holds  $\nu$ -almost everywhere. Suppose  $A_0 \in \mathcal{S}_X$  satisfies  $P_X(A_0) = 0$ . Noting that  $P_X(dx) = r(x)\mu(dx)$ , we have

$$P_X(A_0) = \int_{A_0} r(x) \mu(dx) = \int_{\mathcal{X}} \mathbf{1}_{A_0}(x) r(x) \mu(dx) = 0,$$

i.e.,  $\mathbf{1}_{A_0}(x) r(x) = 0$   $\mu$ -almost everywhere. In other words,

$$\mu(A_0 \cap \{x \in \mathcal{X} : r(x) > 0\}) = 0. \quad (4.17)$$

Now, if  $q(y) = 0$ , then  $P_{X|Y=y}(A_0) = P_X(A_0) = 0$  by construction. So, suppose  $q(y) > 0$ . In this case, we have

$$\begin{aligned} P_{X|Y=y}(A_0) &= \frac{1}{q(y)} \int_{A_0} p(x, y) \mu(dx) \\ &= \frac{1}{q(y)} \left( \int_{A_0 \cap \{r>0\}} p(x, y) \mu(dx) + \int_{A_0 \cap \{r=0\}} p(x, y) \mu(dx) \right). \end{aligned}$$

The first integral is zero due to (4.17). Moreover, for each  $x \in \mathcal{X}$ ,  $r(x) = 0$  implies that  $p(x, y) = 0$   $\nu$ -almost everywhere; thus, the second integral is also zero  $\nu$ -almost everywhere. We conclude that  $P_{X|Y=y} \ll P_X$  for  $\nu$ -almost all  $y \in \mathcal{Y}$ .

Next, we argue that  $P_{XY}(dx, dy) = P_Y(dy) P_{X|Y=y}(dx)$ . Noting that  $P_Y(dy) = q(y) \nu(dy)$  and  $P_{X|Y=y}(dx) = k(x, y)\mu(dx)$  we write

$$\begin{aligned} P_Y(dy) P_{X|Y=y}(dx) &= q(y) k(x, y) \mu(dx) \nu(dy) \\ &= \begin{cases} p(x, y) \mu(dx) \nu(dy) & \text{if } q(y) > 0 \\ 0 & \text{if } q(y) = 0 \end{cases} \\ &= p(x, y) \mu(dx) \nu(dy) \\ &= P_{XY}(dx, dy), \end{aligned} \quad (4.18a)$$

where (4.18a) is due to the fact that for each  $y \in \mathcal{Y}$ ,  $q(y) = 0$  implies  $p(x, y) = 0$   $\mu$ -almost everywhere and a Radon-Nikodym derivative is specified uniquely up to almost everywhere equivalence.

Finally, we show that the mapping  $y \mapsto \ell_{P_{XY}}(X \rightarrow y)$  is  $\mathcal{S}_Y$ -measurable. Define the set  $B_0 = \{y \in \mathcal{Y} : \int_{\{r=0\}} k(x, y) \mu(dx) = 0\}$  which is guaranteed to be in  $\mathcal{S}_Y$  by Fubini's theorem [24, Thm. I.6.14]. The leakage  $\ell_{P_{XY}}(X \rightarrow y)$  can be expressed as

$$\ell_{P_{XY}}(X \rightarrow y) = \begin{cases} \text{ess sup}_{P_X} \left( \frac{k(x, y)}{r(x)} \right) & \text{if } y \in B_0, \\ \infty & \text{if } y \notin B_0. \end{cases}$$

Note that  $\frac{k(x,y)}{r(x)}$ , which is an  $(\mathcal{S}_X \otimes \mathcal{S}_Y)$ -measurable function, is used as the Radon-Nikodym derivative  $\frac{dP_{X|Y=y}}{dP_X}$ . It remains to show that the essential supremum of a jointly measurable function is measurable. We state this in the form of a lemma, proved in Appendix 4.C.

**Lemma 4.25.** *Given measurable spaces  $(X, \mathcal{S}_X)$  and  $(Y, \mathcal{S}_Y)$ , suppose  $s \in (\mathcal{S}_X \otimes \mathcal{S}_Y)_+$ . Let  $P_X$  be a probability measure on  $(X, \mathcal{S}_X)$ . Then, the function  $t : Y \rightarrow \bar{\mathbb{R}}_+$  defined as  $t(y) = \text{ess sup}_{P_X} s(x, y)$  is  $\mathcal{S}_Y$ -measurable.*

Equipped with Lemma 4.25, we conclude that the mapping  $y \mapsto \ell_{P_{XY}}(X \rightarrow y)$  is  $\mathcal{S}_Y$ -measurable, as desired.

#### 4.C Proof of Lemma 4.25

To show that  $t$  is  $\mathcal{S}_Y$ -measurable it suffices to show that the inverse image  $t^{-1}(c, \infty]$  is in  $\mathcal{S}_Y$  for each  $c \in \mathbb{R}_+$ . Fix an arbitrary  $c \in \mathbb{R}_+$ . Given  $y \in Y$ , define the set  $C_y = \{x \in X : s(x, y) > c\}$  which is in  $\mathcal{S}_X$  by the measurability of the mapping  $x \mapsto s(x, y)$  for fixed  $y \in Y$ . Now, we write

$$\begin{aligned} t^{-1}(c, \infty] &= \{y \in Y : t(y) > c\} \\ &= \{y \in Y : P_X(\{x \in X : s(x, y) > c\}) > 0\} \\ &= \{y \in Y : P_X(C_y) > 0\} \\ &= \left\{ y \in Y : \left( \int_X \mathbf{1}_{C_y}(x) P_X(dx) \right) > 0 \right\}. \end{aligned}$$

The mapping  $(x, y) \mapsto \mathbf{1}_{C_y}(x)$  is  $\mathcal{S}_X \otimes \mathcal{S}_Y$ -measurable since  $\{(x, y) \in X \times Y : \mathbf{1}_{C_y}(x) = 1\} = \{(x, y) \in X \times Y : s(x, y) > c\} \in \mathcal{S}_X \otimes \mathcal{S}_Y$ . Then, Fubini's theorem ensures that  $y \mapsto \int_X \mathbf{1}_{C_y}(x) P_X(dx)$  is  $\mathcal{S}_Y$ -measurable, which in turn, implies that  $t^{-1}(c, \infty]$  belongs to  $\mathcal{S}_Y$ .

---

## 5. PML: Properties, Privacy Guarantees, and Comparisons

---

Having introduced PML in Chapter 4, here we further develop the theory surrounding PML by discussing its properties, defining privacy guarantees, and establishing connections with common privacy measures from the literature. For simplicity, in this chapter, we assume that both  $X$  and  $Y$  are finite random variables. Most statements can be readily extended to the general case.

### 5.1 Properties of PML

In this section, we recount several useful properties of  $\ell_{P_{XY}}(X \rightarrow y)$ . For instance, we discuss how pointwise maximal leakage increases by observing multiple outcomes, how it is affected by pre- and post-processing, and so on. Before we discuss these properties, let us first define a conditional form of PML which allows us to model adversaries who possess some side information about the secret  $X$ .

**Definition 5.1** (Conditional pointwise maximal leakage). Let  $P_{XYZ}$  denote the joint distribution of  $X$ ,  $Y$ , and  $Z$ . Given  $z \in \mathcal{Z}$ , the conditional pointwise maximal leakage from  $X$  to  $y \in \mathcal{Y}$  is defined as<sup>1</sup>

$$\ell_{P_{XY|Z}}(X \rightarrow y | z) := \log \sup_{P_{U|X,Z}} \frac{\sup_{P_{\hat{U}|Y,Z}} \mathbb{P} [U = \hat{U} | Y = y, Z = z]}{\sup_{P_{\tilde{U}|Z}} \mathbb{P} [U = \tilde{U} | Z = z]}.$$

To obtain a simpler expression for  $\ell_{P_{XY|Z}}(X \rightarrow y | z)$ , we may condition all the distributions in the proof of Theorem 4.2 on  $Z = z$  and get

$$\begin{aligned} \ell_{P_{XY|Z}}(X \rightarrow y | z) &= D_{\infty} (P_{X|Y=y, Z=z} \| P_{X|Z=z}) \\ &= \log \max_{x \in \text{supp}(P_{X|Z=z})} \frac{P_{X|Y=y, Z=z}(x)}{P_{X|Z=z}(x)} \\ &= \log \max_{x \in \text{supp}(P_{X|Z=z})} \frac{P_{Y|X=x, Z=z}(y)}{P_{Y|Z=z}(y)} \\ &= \max_{x \in \text{supp}(P_{X|Z=z})} i_{P_{XY|Z}}(x; y | z), \end{aligned} \tag{5.1}$$

---

<sup>1</sup>We assume that the Markov chain  $U - (X, Z) - Y$  holds.

where

$$i_{P_{XY|Z}}(x; y | z) := \log \frac{P_{XY|Z=z}(x, y)}{P_{X|Z=z}(x)P_{Y|Z=z}(y)},$$

denotes the conditional information density.

The following lemma (proved in Appendix 5.A) collects several useful properties of PML.

**Lemma 5.2.** *Suppose  $X$  is distributed according to  $P_X$ . Pointwise maximal leakage satisfies the following properties:*

(i) (Upper/lower bounds). *For all mechanisms  $P_{Y|X}$  and all  $y \in \mathcal{Y}$  it holds that*

$$0 \leq \ell(X \rightarrow y) \leq -\log \left( \min_{x \in \text{supp}(P_X)} P_X(x) \right),$$

*where the left-hand side inequality holds with equality if and only if  $P_{Y|X=x}(y) = P_{Y|X=x'}(y)$  for all  $x, x' \in \text{supp}(P_X)$ , and the right-hand side inequality holds with equality if and only if  $P_{X|Y=y}(x^*) = 1$  for some  $x^* \in \arg \min_{x \in \text{supp}(P_X)} P_X(x)$ .*

(ii) (Independence/deterministic mappings). *If  $X$  and  $Y$  are independent random variables, then  $\ell(X \rightarrow y) = 0$  for all  $y \in \mathcal{Y}$ . If  $Y$  is the output of a deterministic mapping with input  $X$ , then  $\ell(X \rightarrow y) = -\log P_Y(y)$ .*

(iii) (Pre-processing). *Suppose the Markov chain  $X - Y - Z$  holds, where  $Y$  represents some processing of the secret  $X$ , and  $Z$  denotes the observable outcome of a channel  $P_{Z|Y}$  with input  $Y$ . Then, for all  $z \in \text{supp}(P_Z)$  we have*

$$\ell(X \rightarrow z) \leq \ell(Y \rightarrow z),$$

*with equality if there exists  $x^* \in \arg \max_{x \in \text{supp}(P_X)} i(x; z)$  such that  $i(y; z) = \ell(Y \rightarrow z)$*

*for all  $y \in \text{supp}(P_{Y|X=x^*})$ .*

(iv) (Post-processing). *Suppose the Markov chain  $X - Y - Z$  holds, where  $Y$  represents the observable outcome of a channel with input  $X$ , and  $Z$  denotes some post-processing of  $Y$ . Then, for all  $z \in \text{supp}(P_Z)$  we have*

$$\ell(X \rightarrow z) \leq \max_{y \in \mathcal{Y}} \ell(X \rightarrow y),$$

*with equality if either of the following conditions is satisfied:*

- a)  $X$  and  $Y$  are independent, or
- b) there exists  $y_z \in \mathcal{Y}$  such that  $P_{Y|Z=z}(y_z) = 1$  and

$$\ell(X \rightarrow y_z) = \max_{y \in \mathcal{Y}} \ell(X \rightarrow y).$$

(v) (Conditionally-independent side information). If the Markov chain  $Z - X - Y$  holds, where  $Z$  represents some side information about  $X$ , then,

$$\ell(X \rightarrow y | z) = \ell(X \rightarrow y) - i(y; z),$$

for all  $z \in \text{supp}(P_Z)$  and  $y \in \text{supp}(P_{Y|Z=z})$ .

(vi) (Composition). Given a mechanism  $P_{YZ|X}$ , for all  $(y, z) \in \text{supp}(P_{YZ})$  it holds that

$$\begin{aligned} \ell(X \rightarrow y, z) &= \max_{x \in \text{supp}(P_X)} i(x; y, z) \\ &\leq \ell(X \rightarrow z) + \ell(X \rightarrow y | z), \end{aligned}$$

with equality if the sets  $\arg \max_{x \in \text{supp}(P_X)} i(x; y | z)$  and  $\arg \max_{x \in \text{supp}(P_X)} i(x; z)$  have non-empty intersection.

*Remark 5.3.* A few points are worth emphasizing about the above properties:

- (i) The second property in the lemma describes the information leakage of deterministic mechanisms. Surprisingly, not all deterministic outcomes leak the same amount of information, and outcomes with lower probabilities have higher leakage. This is because PML is a *relative* or *inferential* privacy measure obtained by comparing the adversary's posterior distribution with her prior distribution. Hence, the information leaked to an adversary depends on how consistent the observed outcome is with the adversary's prior belief captured by  $P_X$ . As such, deterministic outcomes with smaller probabilities leak more information since an adversary would be "more surprised" by observing them.
- (ii) Concerning the post-processing property, one may hope for the stronger statement  $\ell(X \rightarrow z) \leq \ell(X \rightarrow y)$  for all  $y \in \mathcal{Y}$ . To see why this statement is not true, suppose  $Z = Y$ . In this case, the best bound we can have is indeed  $\ell(X \rightarrow y) \leq \max_y \ell(X \rightarrow y)$ .
- (iii) In general, side information can both increase and decrease information leakage. As an example, suppose  $X, Y, Z$  are binary random variables where  $X$  is uniformly distributed and the joint distribution  $P_{XYZ}$  is induced by the following channels:

$$P_{Z|X=x}(0) = \begin{cases} \frac{2}{5} & \text{if } x = 0, \\ \frac{3}{5} & \text{if } x = 1, \end{cases} \quad P_{Y|X=x, Z=z}(0) = \begin{cases} \frac{1}{2} & \text{if } x = 0, z = 0, \\ \frac{1}{3} & \text{if } x = 0, z = 1, \\ \frac{2}{3} & \text{if } x = 1, z = 0, \\ \frac{1}{2} & \text{if } x = 1, z = 1, \end{cases}$$

with  $P_{Z|X=x}(1) = 1 - P_{Z|X=x}(0)$  and  $P_{Y|X=x,Z=z}(1) = 1 - P_{Y|X=x,Z=z}(0)$ . Then, it can be verified that  $\ell_{P_{XY|Z}}(X \rightarrow 0 | 0) = \log \frac{10}{9}$ ,  $\ell_{P_{XY|Z}}(X \rightarrow 0 | 1) = \log \frac{5}{4}$  and  $\ell_{P_{XY}}(X \rightarrow 0) = \log \frac{6}{5}$ . Therefore,

$$\ell_{P_{XY|Z}}(X \rightarrow 0 | 0) < \ell_{P_{XY}}(X \rightarrow 0) < \ell_{P_{XY|Z}}(X \rightarrow 0 | 1).$$

## 5.2 Basic Privacy Guarantees

In Theorem 4.2, we showed that  $\ell(X \rightarrow y)$  is a function of  $y$ . Since  $Y$  is a random variable distributed according to  $P_Y$ , this in turn allows us to define a random variable  $\ell(X \rightarrow Y)$  with a distribution induced by  $P_Y$ . From this point of view, a privacy guarantee is essentially a requirement imposed on statistics of  $\ell(X \rightarrow Y)$ ; thus, we have the flexibility to define different types of privacy guarantees depending on how strict requirements we need to meet. This section contains several examples of such guarantees. We start with the simplest and most conservative definition.

**Definition 5.4** (Almost-sure guarantee). Suppose  $X$  is distributed according to  $P_X$ . We say that a privacy mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -PML with  $\epsilon \geq 0$  if

$$\mathbb{P}_{Y \sim P_Y}[\ell(X \rightarrow Y) \leq \epsilon] = 1. \quad (5.2)$$

As we are assuming that the random variables  $X$  and  $Y$  are finite, the above condition can also be expressed as

$$\max_{(x,y) \in \text{supp}(P_{XY})} i_{P_{XY}}(x; y) = D_\infty(P_{XY} \| P_X \times P_Y) \leq \epsilon.$$

Note that  $D_\infty(P_{XY} \| P_X \times P_Y)$  coincides with the definition of maximal realizable leakage (Definition 3.19) and also max-information [39].

The following lemma establishes some basic facts about  $\epsilon$ -PML guarantees, and is proved in Appendix 5.B.

**Lemma 5.5.** *Suppose  $X$  is distributed according to  $P_X$ .*

(i) *All privacy mechanisms  $P_{Y|X}$  satisfy  $\epsilon_{\max}$ -PML, where*

$$\epsilon_{\max} := -\log \min_{x \in \text{supp}(P_X)} P_X(x),$$

*and  $\epsilon_{\max} \geq \log 2$ . Furthermore, we have*

$$\inf\{\epsilon \geq 0: \mathbb{P}_{Y \sim P_Y}[\ell(X \rightarrow Y) \leq \epsilon] = 1\} = \epsilon_{\max},$$

*if and only if there exists  $y \in \text{supp}(P_Y)$  and  $x^* \in \arg \min_x P_X(x)$  such that  $P_{X|Y=y}(x^*) = 1$ , or equivalently,  $P_{Y|X=x}(y) = 0$  for all  $x \neq x^*$ .*

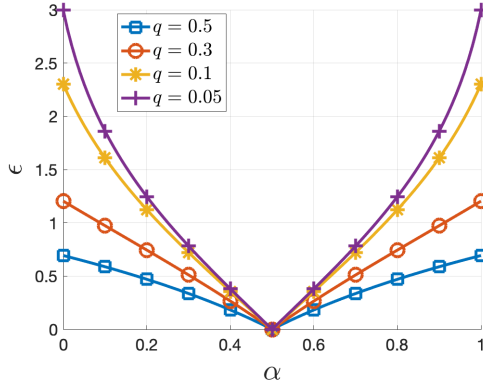


Figure 5.2.1: Leakage of the binary symmetric channel with  $q \in \{0.05, 0.1, 0.3, 0.5\}$ .

(ii) A privacy mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -PML with  $\epsilon = 0$  if and only if  $X$  and  $Y$  are independent random variables.

Below, we give an example of a channel satisfying  $\epsilon$ -PML.

**Example 5.6** (Binary symmetric channel). Suppose  $P_{Y|X}$  is a binary symmetric channel with crossover probability  $0 \leq \alpha \leq 1$ . The input distribution is described by  $P_X(0) = 1 - P_X(1) = q$ , where  $0 < q \leq 0.5$ . Then,  $P_{Y|X}$  satisfies  $\epsilon$ -PML with

$$\begin{aligned} \epsilon &= \begin{cases} \log \frac{1 - \alpha}{q(1 - 2\alpha) + \alpha} & \text{if } \alpha \leq \frac{1}{2}, \\ \log \frac{\alpha}{q(2\alpha - 1) + 1 - \alpha} & \text{if } \alpha > \frac{1}{2}, \end{cases} \\ &= \log \frac{|\alpha - \frac{1}{2}| + \frac{1}{2}}{\frac{1}{2} - |\alpha - \frac{1}{2}|(1 - 2q)}. \end{aligned}$$

Figure 5.2.1 depicts the leakage of the channel as a function of  $\alpha$  for different values of  $q$ . As expected, at  $\alpha = \frac{1}{2}$  no information is leaked because  $X$  and  $Y$  are independent. It is also interesting to note that assuming a fixed  $\alpha$ ,  $\epsilon$  is decreasing in  $q$ , implying that the binary symmetric channel leaks more information about skewed priors.

In order for an  $\epsilon$ -PML guarantee to hold, all  $y \in \text{supp}(P_Y)$  must satisfy  $\ell(X \rightarrow y) \leq \epsilon$ . As this condition may prove to be too restrictive in practice, in what follows we discuss two possible relaxations of the almost-sure guarantee: We either bound the information leakage by  $\epsilon$  with high probability, or we bound the expected leakage (i.e., maximal leakage) by  $\epsilon$ .

**Definition 5.7** (Tail-bound guarantee). Suppose  $X$  is distributed according to  $P_X$ . We say that a mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -PML with  $\epsilon \geq 0$  and  $0 \leq \delta \leq 1$  if

$$\mathbb{P}_{Y \sim P_Y}[\ell(X \rightarrow Y) \leq \epsilon] \geq 1 - \delta. \quad (5.3)$$

Clearly,  $\epsilon$ -PML and  $(\epsilon, 0)$ -PML are equivalent. Also, note that given an arbitrary but fixed prior  $P_X$ , if a channel  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -PML, then it also satisfies  $(\epsilon', \delta')$ -PML for all  $\epsilon \leq \epsilon'$  and all  $\delta \leq \delta' \leq 1$ .

**Definition 5.8** (Average-case guarantee). Suppose  $X$  is distributed according to  $P_X$ . We say that the expected information leakage of a mechanism  $P_{Y|X}$  is bounded by  $\epsilon \geq 0$  if

$$\mathbb{E}_{Y \sim P_Y} \left[ \exp(\ell(X \rightarrow Y)) \right] \leq e^\epsilon,$$

or equivalently, if  $\mathcal{L}(P_{Y|X}) \leq \epsilon$ , where  $\mathcal{L}(P_{Y|X})$  denotes maximal leakage (Definition 3.15).

Note that here we have denoted maximal leakage by  $\mathcal{L}(P_{Y|X})$  instead of  $\mathcal{L}(X \rightarrow Y)$  to emphasize that maximal leakage is a property of the channel  $P_{Y|X}$  and does not depend on the prior  $P_X$ .<sup>2</sup>

### 5.3 Privacy Guarantees: Data-processing Properties

Data-processing inequalities are often used while analyzing the end-to-end information leakage in larger systems. While the properties presented in Lemma 5.2 allow us to assess pointwise maximal leakage for the outcomes of a pre- or post-processed random variable, it is also of practical benefit to understand how different privacy guarantees are affected by pre- and post-processing. This type of characterization is useful when we do not have access to the distribution of the leakage over the outcomes, but know that a privacy mechanism satisfies a certain privacy guarantee.

What we are specifically interested in is to understand whether or not different privacy guarantees are closed under pre- and post-processing (in [144], a privacy guarantee that is closed under pre-processing is said to satisfy the *linkage* inequality). Suppose  $P_{Y|X}$  satisfies some privacy guarantee, say,  $\epsilon$ -PML. If the  $\epsilon$ -PML guarantee is closed under post-processing, then we can rest assured that for all post-processing channels  $P_{Z|Y}$ , the overall channel  $P_{Z|X}$  also satisfies  $\epsilon$ -PML. In other words, there exists no computation that an adversary could use to undermine the original guarantee. Similarly, if the  $\epsilon$ -PML guarantee is closed under pre-processing, then all (randomized) functions of  $X$  would be at least as well-protected as  $X$ .

---

<sup>2</sup>Technically, maximal leakage depends on the support set of the prior  $P_X$ , but we can without loss of generality assume that  $P_X$  has full support.

The following result collects the data-processing properties satisfied by the privacy guarantees defined above. Part (iv) of the result concerning maximal leakage is the data-processing inequality for Sibson mutual information of order  $\infty$  and we re-state it here for completeness. The proposition is proved in Appendix 5.C.

**Proposition 5.9.** *Suppose  $X$  is distributed according to  $P_X$  and assume that the Markov chain  $X - Y - Z$  holds. Given  $\epsilon \geq 0$  and  $0 \leq \delta \leq 1$ , we have*

- (i) *If  $P_{Z|Y}$  satisfies  $\epsilon$ -PML, then  $P_{Z|X}$  also satisfies  $\epsilon$ -PML.*
- (ii) *If  $P_{Y|X}$  satisfies  $\epsilon$ -PML, then  $P_{Z|X}$  also satisfies  $\epsilon$ -PML.*
- (iii) *If  $P_{Z|Y}$  satisfies  $(\epsilon, \delta)$ -PML, then  $P_{Z|X}$  also satisfies  $(\epsilon, \delta)$ -PML.*
- (iv)  $\mathcal{L}(P_{Z|X}) \leq \min\{\mathcal{L}(P_{Y|X}), \mathcal{L}(P_{Z|Y})\}$ .

Below, we examine the truncated geometric mechanism [52, 69] which is a common perturbation method satisfying LDP. This mechanism behaves interestingly under PML because the truncation, which is a form of post-processing, does not further decrease the leakage.

**Example 5.10** (Truncated geometric mechanism). Suppose  $\mathcal{X} = [k]$  with  $k \geq 3$ ,  $\mathcal{Y} = \mathbb{Z}$ , and let  $Y$  be the output of the geometric mechanism described by

$$P_{Y|X=x}(y) = \frac{1 - \exp(-\frac{\alpha}{k-1})}{1 + \exp(-\frac{\alpha}{k-1})} \cdot \exp(-\frac{\alpha|y-x|}{k-1}),$$

where  $x \in \mathcal{X}$ ,  $y \in \mathcal{Y}$ , and  $\alpha \geq 0$ . Clearly,  $P_{Y|X}$  satisfies  $\alpha$ -LDP. Furthermore, the leakage from  $X$  to each  $y \in \mathbb{Z}$  is

$$\ell_{P_{XY}}(X \rightarrow y) = \begin{cases} -\log \mathbb{E}[\exp(-\frac{\alpha|y-X|}{k-1})] & \text{if } y \in \mathcal{X}, \\ -\log \mathbb{E}[\exp(-\frac{\alpha(k-X)}{k-1})] & \text{if } y > k, \\ -\log \mathbb{E}[\exp(-\frac{\alpha(X-1)}{k-1})] & \text{if } y < 1. \end{cases}$$

Note that  $\ell_{P_{XY}}(X \rightarrow y)$  does not depend on the actual value of  $y$  when  $y > k$  or  $y < 1$ .

Let  $Z$  denote the truncated version of  $Y$ , that is,

$$Z = \begin{cases} Y & \text{if } Y \in \mathcal{X}, \\ 1 & \text{if } Y < 1, \\ k & \text{if } Y > k. \end{cases}$$

Then,  $\ell_{P_{XZ}}(X \rightarrow i) = \ell_{P_{XY}}(X \rightarrow i)$  for  $i \in \{2, \dots, k-1\}$ , and

$$\ell_{P_{XZ}}(X \rightarrow 1) = \log \frac{\max_x \sum_{y=-\infty}^1 P_{Y|X=x}(y)}{\sum_{y=-\infty}^1 P_Y(y)}$$

$$\begin{aligned}
 &= \log \frac{\sum_{y=-\infty}^1 P_{Y|X=1}(y)}{\sum_{y=-\infty}^1 P_Y(y)} \\
 &= \ell_{P_{XY}}(X \rightarrow 1),
 \end{aligned}$$

and

$$\begin{aligned}
 \ell_{P_{XZ}}(X \rightarrow k) &= \log \frac{\max_x \sum_{y=k}^{\infty} P_{Y|X=x}(y)}{\sum_{y=k}^{\infty} P_Y(y)} \\
 &= \log \frac{\sum_{y=k}^{\infty} P_{Y|X=k}(y)}{\sum_{y=k}^{\infty} P_Y(y)} \\
 &= \ell_{P_{XY}}(X \rightarrow k).
 \end{aligned}$$

Now, since  $|y - x| \leq \max\{y - 1, k - y\}$  for all  $x, y \in \mathcal{X}$ , both  $P_{Y|X}$  and  $P_{Z|X}$  satisfy  $\epsilon$ -PML with

$$\begin{aligned}
 \epsilon &= \max_{y \in \mathcal{Y}} \ell_{P_{XY}}(X \rightarrow y) \\
 &= \max_{z \in \mathcal{Z}} \ell_{P_{XZ}}(X \rightarrow z) \\
 &= \max \left\{ -\log \mathbb{E} \left[ \exp \left( -\frac{\alpha(k - X)}{k - 1} \right) \right], -\log \mathbb{E} \left[ \exp \left( -\frac{\alpha(X - 1)}{k - 1} \right) \right] \right\} \\
 &\leq \frac{\alpha}{k - 1} \cdot \max \left\{ k - \mathbb{E}[X], \mathbb{E}[X] - 1 \right\},
 \end{aligned}$$

where the last line is due to Jensen's inequality.

Interestingly, when the distribution  $P_X$  is highly skewed with either  $\mathbb{E}[X] \rightarrow 1$  or  $\mathbb{E}[X] \rightarrow k$  the gap between  $\epsilon$  and  $\alpha$  vanishes and  $\epsilon \rightarrow \alpha$  since Jensen's inequality holds with equality for degenerate distributions. On the other hand, when  $X$  is uniformly distributed, the bound reduces to  $\epsilon \leq \frac{\alpha}{2}$  and the gap between  $\epsilon$  and  $\alpha$  is maximized.

Conspicuous by its absence in Proposition 5.9 is the post-processing property for the  $(\epsilon, \delta)$ -PML privacy guarantee. It turns out that, in general,  $(\epsilon, \delta)$ -PML is not closed under post-processing. To see why, let us consider the following example.

**Example 5.11.** Suppose  $X$  is a uniformly distributed random variable defined over an alphabet with four elements, and that the Markov chain  $X - Y - Z$  holds. Suppose the channels  $P_{Y|X}$  and  $P_{Z|Y}$  are defined as

$$P_{Y|X} = \begin{bmatrix} 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \end{bmatrix}, \quad P_{Z|Y} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

It can be easily verified that  $\ell(X \rightarrow y_1) = \ell(X \rightarrow y_2) = \log 4$ , and  $\ell(X \rightarrow y_3) = \ell(X \rightarrow y_4) = \log \frac{6}{5}$ . Since  $P_Y(y_1) = P_Y(y_2) = \frac{1}{12}$ ,  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta_1)$ -PML with  $\epsilon_1 = \log \frac{6}{5}$  and  $\delta_1 = \frac{1}{6}$ . On the other hand, one may also verify that  $P_Z(z_1) = P_Z(z_2) = \frac{1}{2}$  and  $\ell(X \rightarrow z_1) = \ell(X \rightarrow z_2) = \log \frac{4}{3}$ . Hence,  $P_{Z|X}$  does not satisfy  $(\epsilon_1, \delta_1)$ -PML; instead, it satisfies  $\epsilon_2$ -PML with  $\epsilon_2 = \log \frac{4}{3} > \epsilon_1$  (and  $\delta_2 = 0$ ). Note that the outcome  $z_1$  is equivalent to the event  $\{y_1, y_3\}$ ,  $z_2$  is equivalent to the event  $\{y_2, y_4\}$ , and both outcomes have probability greater than  $\delta_1$ .

Informally speaking, when a mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -PML this implies that  $\text{supp}(P_Y)$  can be partitioned into two sets: a set of “good”  $y$ ’s with probability at least  $1 - \delta$  whose members satisfy  $\ell(X \rightarrow y) \leq \epsilon$ , and a set of “bad”  $y$ ’s with probability at most  $\delta$  and  $\ell(X \rightarrow y) > \epsilon$  (see Definition 5.7). However, through a post-processing channel  $P_{Z|Y}$ , we may define new outcomes as a combination of the members of the good and bad sets of  $y$  (as in Example 5.11). As a result, the probability of the set whose members satisfy  $\ell(X \rightarrow z) > \epsilon$  (that is, the set of “bad”  $z$ ’s) may no longer be bounded by  $\delta$ . Also, note that while in Example 5.11 we have  $\epsilon_2 > \epsilon_1$  and  $\delta_2 < \delta_1$ , this need not always be the case; one may come up with examples where both  $\epsilon$  and  $\delta$  increase by post-processing.

*Remark 5.12.* As discussed in Chapter 3, a similar behavior has been observed in the case of differential privacy. Specifically, it has been shown that probabilistic DP (Definition 3.8) is not closed under post-processing [74, 105]. Differential privacy solves this problem by introducing approximate DP (Definition 3.6) using an additive parameter  $\delta$  which is closed under post-processing. Note that approximate DP is a strictly weaker guarantee compared to probabilistic DP in the sense that probabilistic DP implies approximate DP but the reverse direction does not necessarily hold [105]. Here, we take a different approach and propose a new privacy guarantee that maintains its probabilistic flavor.

Below, we define a new probabilistic privacy guarantee that is similar to  $(\epsilon, \delta)$ -PML, but is closed under post-processing. Drawing on Example 5.11, our new definition ensures that all post-processed outcomes with probability at least  $\delta$  have their PML bounded by  $\epsilon$ . We provide two alternative formulations of our new privacy guarantee: The first one in Definition 5.13 describes a somewhat technical condition, so we re-state it in a more intuitive form in Definition 5.18.

**Definition 5.13.** Given an arbitrary but fixed prior  $P_X$ , we say that a privacy mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -closedness with  $\epsilon \geq 0$  and  $0 \leq \delta \leq 1$ , if for all post-processing channels  $P_{Z|Y}$  and all  $z \in \text{supp}(P_Z)$ ,  $P_Z(z) \geq \delta$  implies  $\ell_{P_{XZ}}(X \rightarrow z) \leq \epsilon$ .

Based on Definition 5.13, to check whether or not a certain mechanism  $P_{Y|X}$  satisfies the desired closedness property, one needs to examine all possible post-processing channels  $P_{Z|Y}$ . This raises the question of whether it is possible to come up with a definition equivalent to Definition 5.13, which can be stated as a

property of the channel  $P_{Y|X}$  itself. In what follows, we show that this is indeed possible, but we need a few other ingredients before we are ready to state this alternative definition. First, we recall two concepts from [101].

**Definition 5.14** (Similar outcomes [101]). Given a channel  $P_{Y|X}$ , we say that the outcomes  $y, y' \in \mathcal{Y}$  are similar if their corresponding columns in the matrix form of  $P_{Y|X}$  are scalar multiples of each other, or equivalently, if  $P_{X|Y=y}(x) = P_{X|Y=y'}(x)$  for all  $x \in \text{supp}(P_X)$ .

Note that if the outcomes  $y, y' \in \text{supp}(P_Y)$  are similar, then  $i(x; y) = i(x; y')$  for all  $x \in \text{supp}(P_X)$  and  $\ell(X \rightarrow y) = \ell(X \rightarrow y')$ .

**Definition 5.15** (Reduced channel [101, Def. 3]). Given a channel  $P_{Y|X}$ , its reduced channel denoted by  $P_{\bar{Y}|X}$  is formed by removing all-zero columns from  $P_{Y|X}$ , and merging (i.e., adding) the columns corresponding to similar outcomes.

Let  $P_{\bar{Y}|X}$  denote the reduced channel corresponding to the mechanism  $P_{Y|X}$ . We define an equivalence relation  $P_{\bar{Y}|X} \sim P_{Y|X}$  if  $P_{\bar{Y}|X}$  has  $P_{Y|X}$  as its reduced channel. Then, the equivalence class of  $P_{Y|X}$ , denoted by  $\mathcal{C}(P_{Y|X})$ , is the collection of all mechanisms whose reduced channel is  $P_{\bar{Y}|X}$ . Suppose  $P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})$ . We will use  $\bar{Y}$  to denote the (output) random variable induced by the channel  $P_{\bar{Y}|X}$  whose alphabet is represented by  $\bar{\mathcal{Y}}$ , and whose marginal distribution is denoted by  $P_{\bar{Y}}$ .

Similar outcomes lead to the same posterior distribution, information density, and PML. Thus, the channels in a class  $\mathcal{C}(P_{Y|X})$  behave identically with respect to information measures that are defined based on the information density, such as mutual information and maximal leakage. The following result states that if  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -closedness, then all  $P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})$  also satisfy  $(\epsilon, \delta)$ -closedness. The lemma is proved in Appendix 5.D.

**Lemma 5.16.** *Given an arbitrary but fixed prior  $P_X$  and an  $(\epsilon, \delta)$  pair with  $\epsilon \geq 0$  and  $0 \leq \delta \leq 1$ , if a mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -closedness then all  $P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})$  also satisfy  $(\epsilon, \delta)$ -closedness.*

One last concept that we need to introduce is the notion of the maximal leakage associated with arbitrary events (that is, subsets) of  $\mathcal{Y}$ . We call this new form of leakage *event maximal leakage* (EML), which is defined fairly similarly to PML. We use EML to present an alternative formulation of Definition 5.13.

**Definition 5.17** (Event maximal leakage (EML)). Let  $P_{XY}$  denote the joint distribution of  $X$  and  $Y$ . Given an event  $\mathcal{E} \subseteq \mathcal{Y}$ , the maximal leakage from  $X$  to  $\mathcal{E}$  is defined as

$$\ell_{P_{XY}}(X \rightarrow \mathcal{E}) := \log \max_{x \in \text{supp}(P_X)} \frac{P_{Y|X=x}(\mathcal{E})}{P_Y(\mathcal{E})}$$

$$\begin{aligned}
 &= \log \max_{x \in \text{supp}(P_X)} \frac{\sum_{y \in \mathcal{E}} P_{Y|X=x}(y)}{\sum_{y' \in \mathcal{E}} P_Y(y')} \\
 &= \log \max_{x \in \text{supp}(P_X)} \sum_{y \in \mathcal{E}} \frac{P_Y(y)}{\sum_{y' \in \mathcal{E}} P_Y(y')} \exp(i_{P_{XY}}(x; y)).
 \end{aligned}$$

EML essentially quantifies the information leaking from the outcomes of a deterministic function of  $Y$ . To see why, suppose  $Z$  is a random variable produced by some deterministic post-processing of  $Y$ . Then, outcomes of  $Z$  are either the re-labeled outcomes of  $Y$  or they result from merging several outcomes of  $Y$  into a single symbol. If  $z \in \mathcal{Z}$  is a re-labeling of  $y \in \mathcal{Y}$ , then  $\ell_{P_{XZ}}(X \rightarrow z) = \ell_{P_{XY}}(X \rightarrow y)$ . On the other hand, if  $z' \in \mathcal{Z}$  is a result of combining  $y \in \mathcal{E} \subseteq \mathcal{Y}$  into a single symbol, then  $\ell_{P_{XZ}}(X \rightarrow z') = \ell_{P_{XY}}(X \rightarrow \mathcal{E})$ .

Having defined the maximal leakage associated with arbitrary subsets of  $\mathcal{Y}$ , we are now ready to provide an alternative form for Definition 5.13.

**Definition 5.18.** Suppose  $X$  is distributed according to  $P_X$ . We say that a privacy mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -EML with  $\epsilon \geq 0$  and  $0 \leq \delta \leq 1$  if for all  $P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})$  and all events  $\mathcal{E} \subseteq \bar{\mathcal{Y}}$ ,  $P_{\bar{Y}}(\mathcal{E}) \geq \delta$  implies  $\ell_{P_{X\bar{Y}}}(X \rightarrow \mathcal{E}) \leq \epsilon$ .

Clearly,  $(\epsilon, 0)$ -EML and  $\epsilon$ -PML are equivalent. Furthermore, given an arbitrary but fixed prior  $P_X$ , if a channel  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -EML, then it also satisfies  $(\epsilon', \delta')$ -EML for all  $\epsilon \leq \epsilon'$ , and all  $\delta \leq \delta' \leq 1$ .

Next, we show that a privacy mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -closedness if and only if it satisfies  $(\epsilon, \delta)$ -EML. That is, Definitions 5.13 and 5.18 are equivalent.

**Theorem 5.19.** *Given an arbitrary but fixed prior  $P_X$ , and a pair  $(\epsilon, \delta)$  with  $\epsilon \geq 0$  and  $0 \leq \delta \leq 1$ , a privacy mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -closedness if and only if it satisfies  $(\epsilon, \delta)$ -EML.*

*Proof.* Suppose without loss of generality that  $P_X$  has full support. We first show that if a privacy mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -closedness, then it satisfies  $(\epsilon, \delta)$ -EML. Informally, this result follows from the fact that for all  $P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})$ , optimizing over the events in  $\text{supp}(P_{\bar{Y}})$  with probability at least  $\delta$  is equivalent to optimizing over the outcomes of all deterministic mappings  $P_{Z|\bar{Y}}$  with probability at least  $\delta$ . More concretely, suppose  $P_{Z|\bar{Y}}$  is a deterministic channel, that is, the matrix form of  $P_{Z|\bar{Y}}$  consists only of zeros and ones. Then, each outcome  $z \in \text{supp}(P_Z)$  corresponds to some event  $\mathcal{E}_z \subseteq \bar{\mathcal{Y}}$  such that  $P_{Z|\bar{Y}=y}(z) = 1$  for all  $y \in \mathcal{E}_z$ , and  $P_{\bar{Y}}(\mathcal{E}_z) = P_Z(z)$ . Let  $\mathcal{D}_{\bar{\mathcal{Y}}}$  denote the set of all deterministic mappings with domain  $\bar{\mathcal{Y}}$ . We can write

$$\begin{aligned}
 \epsilon &\geq \sup_{P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})} \sup_{P_{Z|\bar{Y}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \ell_{P_{XZ}}(X \rightarrow z) \\
 &\geq \sup_{P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})} \sup_{P_{Z|\bar{Y}} \in \mathcal{D}_{\bar{\mathcal{Y}}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \ell_{P_{XZ}}(X \rightarrow z)
 \end{aligned}$$

$$\begin{aligned}
 &= \sup_{P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})} \sup_{P_{Z|\bar{Y}} \in \mathcal{D}_{\bar{y}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \log \max_x \frac{P_{Z|X=x}(z)}{P_Z(z)} \\
 &= \sup_{P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})} \sup_{P_{Z|\bar{Y}} \in \mathcal{D}_{\bar{y}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \log \max_x \frac{P_{\bar{Y}|X=x}(\mathcal{E}_z)}{P_{\bar{Y}}(\mathcal{E}_z)} \\
 &= \sup_{P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})} \max_{\substack{\mathcal{E} \subseteq \text{supp}(P_{\bar{Y}}): \\ P_{\bar{Y}}(\mathcal{E}) \geq \delta}} \log \max_x \frac{P_{\bar{Y}|X=x}(\mathcal{E})}{P_{\bar{Y}}(\mathcal{E})} \\
 &= \sup_{P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})} \max_{\substack{\mathcal{E} \subseteq \text{supp}(P_{\bar{Y}}): \\ P_{\bar{Y}}(\mathcal{E}) \geq \delta}} \ell_{P_{X\bar{Y}}}(X \rightarrow \mathcal{E}),
 \end{aligned}$$

where the first inequality follows from Lemma 5.16. Thus,  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -EML.

Now, we show that if a mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -EML, then it satisfies  $(\epsilon, \delta)$ -closedness. Let the function  $h : \mathcal{C}(P_{Y|X}) \times [0, 1] \rightarrow [1, \infty)$  be defined as

$$h(P_{\bar{Y}|X}, \delta) = \sup_{P_{Z|\bar{Y}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \exp\left(\ell_{P_{XZ}}(X \rightarrow z)\right). \quad (5.4)$$

We can write  $h$  as

$$h(P_{\bar{Y}|X}, \delta) = \max_x h_x(P_{\bar{Y}|X}, \delta),$$

where

$$\begin{aligned}
 h_x(P_{\bar{Y}|X}, \delta) &:= \sup_{P_{Z|\bar{Y}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \frac{P_{Z|X=x}(z)}{P_Z(z)} \\
 &= \sup_{P_{Z|\bar{Y}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \frac{\sum_{y \in \text{supp}(P_{\bar{Y}})} P_{Z|\bar{Y}=y}(z) P_{\bar{Y}|X=x}(y)}{\sum_{y' \in \text{supp}(P_{\bar{Y}})} P_{Z|\bar{Y}=y'}(z) P_{\bar{Y}}(y')} \\
 &= \sup_{P_{Z|\bar{Y}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \sum_{y \in \text{supp}(P_{\bar{Y}})} \frac{P_{Z|\bar{Y}=y}(z) P_{\bar{Y}}(y)}{\sum_{y' \in \text{supp}(P_{\bar{Y}})} P_{Z|\bar{Y}=y'}(z) P_{\bar{Y}}(y')} \left( \frac{P_{\bar{Y}|X=x}(y)}{P_{\bar{Y}}(y)} \right). \quad (5.5)
 \end{aligned}$$

By Lemma 5.16, it suffices to show that for each  $x$ , there exists  $P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})$  satisfying  $h_x(P_{\bar{Y}|X}, \delta) \leq \exp(\epsilon)$ . Hence, we solve the above optimization problem for the reduced channel associated with the class  $\mathcal{C}(P_{Y|X})$ , denoted by  $P_{Y_r|X}$ .

Fix some  $x \in \mathcal{X}$ . Let  $n_r := |\mathcal{Y}_r|$  denote the cardinality of  $\mathcal{Y}_r$  (recall that  $P_{Y_r}$  has full support). We re-write (5.5) for  $P_{Y_r|X}$  as

$$\max_{a_1, \dots, a_{n_r}} \sum_{j=1}^{n_r} \frac{a_j P_{Y_r}(y_j)}{\sum_{j'=1}^{n_r} a_{j'} P_{Y_r}(y_{j'})} \exp(i_{P_{X Y_r}}(x; y_j)),$$

$$\begin{aligned} \text{subject to } \quad & \sum_{j=1}^{n_r} a_j P_{Y_r}(y_j) \geq \delta, \\ & 0 \leq a_j \leq 1, \quad \forall j \in [n_r], \end{aligned} \quad (5.6)$$

where  $\{a_j\}$  specify  $P_{Z|Y_r=y_j}(z)$  for the  $z \in \text{supp}(P_Z)$  with the largest PML which also satisfies  $P_Z(z) \geq \delta$ .

Suppose the elements in  $\mathcal{Y}_r$  are labelled such that  $i_{P_{X Y_r}}(x; y_1) \geq i_{P_{X Y_r}}(x; y_2) \geq \dots \geq i_{P_{X Y_r}}(x; y_{n_r})$ . Given an integer  $k \in [n_r]$ , let  $\mathcal{F}_k := \{y_1, \dots, y_k\}$  be the set containing  $k$  elements from  $\mathcal{Y}_r$  that have the largest information density with  $x$ . Let  $k^* \in [n_r]$  be the smallest integer such that  $P_{Y_r}(\mathcal{F}_{k^*}) \geq \delta$ . The objective function in problem (5.6) is a linear-fractional function which is quasi-convex (in fact, quasi-linear) [19, Section 3.4], and the feasible region is a convex polytope. Therefore, the optimal solution is an extreme point of the feasible region given by

$$a_j^* = \begin{cases} 1, & \text{if } j = 1, \dots, k^* - 1, \\ \zeta, & \text{if } j = k^*, \\ 0, & \text{otherwise,} \end{cases}$$

where the parameter  $0 < \zeta \leq 1$  can be calculated by

$$P_{Y_r}(\mathcal{F}_{k^*-1}) + \zeta P_{Y_r}(y_{k^*}) = \delta. \quad (5.7)$$

We also obtain  $h_x(P_{Y_r|X}, \delta)$  as

$$h_x(P_{Y_r|X}, \delta) = \frac{1}{\delta} \left( P_{Y_r|X=x}(\mathcal{F}_{k^*-1}) + \zeta P_{Y_r|X=x}(y_{k^*}) \right). \quad (5.8)$$

Since  $k^*$  is the smallest integer such that  $P_{Y_r}(\mathcal{F}_{k^*}) \geq \delta$ , we need to consider the following two possibilities: We either have  $P_{Y_r}(\mathcal{F}_{k^*}) = \delta$  or  $P_{Y_r}(\mathcal{F}_{k^*}) > \delta$ . First, suppose  $\mathcal{F}_{k^*} = \delta$ . In this case, the optimal parameters become

$$a_j^* = \begin{cases} 1, & \text{if } j = 1, \dots, k^*, \\ 0, & \text{otherwise,} \end{cases}$$

that is,  $\zeta = 1$  in (5.7). Since  $\{a_j^*\}$  consist of only zeros and ones, it in fact specifies a deterministic outcome  $z^* \in \text{supp}(P_{Z^*})$  for some channel  $P_{Z^*|Y_r}$ . This outcome corresponds to the event  $\mathcal{F}_{k^*}$  in the sense that  $P_{Z|Y_r=y}(z^*) = 1$  for all  $y \in \mathcal{F}_{k^*}$ , and  $P_{Y_r}(\mathcal{F}_{k^*}) = P_Z(z^*)$ . Thus, we get

$$\begin{aligned} h_x(P_{Y_r|X}, \delta) &= \sup_{P_{Z|Y_r}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \frac{P_{Z|X=x}(z)}{P_Z(z)} \\ &= \frac{P_{Z^*|X=x}(z^*)}{P_{Z^*}(z^*)} \\ &= \frac{P_{Y_r|X=x}(\mathcal{F}_{k^*})}{P_{Y_r}(\mathcal{F}_{k^*})} \\ &\leq \exp(\epsilon), \end{aligned} \quad (5.9)$$

where the last inequality follows from the fact that  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -EML.

Now, suppose  $P_{Y_r}(\mathcal{F}_{k^*}) > \delta$  which implies that  $0 < \zeta < 1$  in (5.7). In this case, we construct  $P_{\hat{Y}|X} \in \mathcal{C}(P_{Y|X})$  whose columns are identical to the columns of  $P_{Y_r|X}$ , except that the  $k^*$ -th column of  $P_{Y_r|X}$  is split into two corresponding columns in  $P_{\hat{Y}|X}$  given by

$$P_{\hat{Y}|X=x}(y_{k^*_{(1)}}) = \zeta P_{Y_r|X=x}(y_{k^*}),$$

and

$$P_{\hat{Y}|X=x}(y_{k^*_{(2)}}) = (1 - \zeta) P_{Y_r|X=x}(y_{k^*}),$$

for all  $x \in \mathcal{X}$ . Note that the outcomes  $y_{k^*_{(1)}}, y_{k^*_{(2)}} \in \text{supp}(P_{\hat{Y}})$  defined above are similar, and satisfy

$$i_{P_{X\hat{Y}}}(x; y_{k^*_{(1)}}) = i_{P_{X\hat{Y}}}(x; y_{k^*_{(2)}}) = i_{P_{XY_r}}(x; y_{k^*}).$$

Now, we find  $h_x(P_{\hat{Y}|X}, \delta)$ . Forming the optimization problem (5.5) for  $P_{\hat{Y}|X}$ , it is easy to see that the optimal parameters are

$$a_j^* = \begin{cases} 1, & \text{if } j = 1, \dots, k^* - 1, k^*_{(1)} \\ 0, & \text{otherwise,} \end{cases}$$

which, once again, specifies a deterministic outcome. Using arguments similar to (5.9), we get  $h_x(P_{\hat{Y}|X}, \delta) \leq \exp(\epsilon)$ .

Finally, as  $x$  was chosen arbitrarily, we conclude that  $h(P_{Y|X}, \delta) \leq \exp(\epsilon)$ , that is,  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -closedness.  $\square$

*Remark 5.20.* The proof of Theorem 5.19 sheds light on the role of the class  $\mathcal{C}(P_{Y|X})$ : For each  $0 \leq \delta \leq 1$ , there exists  $P_{\hat{Y}|X} \in \mathcal{C}(P_{Y|X})$  and a “least private” event  $\mathcal{E}^* \subseteq \text{supp}(P_{\hat{Y}})$  satisfying  $P_{\hat{Y}}(\mathcal{E}^*) = \delta$ . As such, without loss of generality, we unify the channels in the equivalence class  $\mathcal{C}(P_{Y|X})$  and assume that  $\mathcal{E}^* \subseteq \text{supp}(P_Y)$ . Then, to show that  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -EML, it suffices to show that  $\ell_{P_{XY}}(X \rightarrow \mathcal{E}^*) \leq \epsilon$ .

Recall that our motivation for introducing the notion of  $(\epsilon, \delta)$ -EML was to obtain a probabilistic privacy guarantee which is closed under both pre- and post-processing. The following result formally shows that this is indeed the case, and is proved in Appendix 5.E.

**Proposition 5.21.** *Suppose the Markov chain  $X - Y - Z$  holds. Given  $\epsilon \geq 0$  and  $0 \leq \delta \leq 1$ , it holds that*

- (i) (Pre-processing) *If  $P_{Z|Y}$  satisfies  $(\epsilon, \delta)$ -EML, then  $P_{Z|X}$  satisfies  $(\epsilon, \delta)$ -EML.*
- (ii) (Post-processing) *If  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -EML, then  $P_{Z|X}$  satisfies  $(\epsilon, \delta)$ -EML.*

Now, let us re-visit Example 5.11 and analyze it through the lens of event maximal leakage.

**Example 5.22.** Suppose  $P_X$ ,  $P_{Y|X}$  and,  $P_{Z|Y}$  are defined as in Example 5.11, and let  $\delta = \frac{1}{6}$ . Our goal is to find the smallest  $\epsilon_1 \geq 0$  such that  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta)$ -EML, and the smallest  $\epsilon_2 \geq 0$  such that  $P_{Z|X}$  satisfies  $(\epsilon_2, \delta)$ -EML. First, note that the outcomes  $y_3$  and  $y_4$  are similar; hence, by merging them, we obtain the reduced channel  $P_{Y_r|X}$  as

$$P_{Y_r|X} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} \end{bmatrix}.$$

Now, for each  $x$ , we find  $h_x(P_{Y_r|X}, \delta)$  defined in (5.8):

$$\begin{aligned} h_{x_1}(P_{Y_r|X}, \delta) &= h_{x_2}(P_{Y_r|X}, \delta) = \frac{6}{5}, \\ h_{x_3}(P_{Y_r|X}, \delta) &= h_{x_4}(P_{Y_r|X}, \delta) = \frac{12}{5}, \end{aligned}$$

which implies that  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta)$ -EML with

$$\epsilon_1 = \log \max_x h_x(P_{Y_r|X}, \delta) = \log \frac{12}{5}.$$

Furthermore, the channel  $P_{Z|X}$  is given by

$$P_{Z|X} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

Since  $P_Z(z_1) = P_Z(z_2) = \frac{1}{2}$  and  $\ell(X \rightarrow z_1) = \ell(X \rightarrow z_2) = \log \frac{4}{3}$ , it follows that  $P_{Z|X}$  satisfies  $(\epsilon_2, \delta)$ -EML with  $\epsilon_2 = \log \frac{4}{3}$ . Note that since  $\epsilon_2 < \epsilon_1$ ,  $P_{Z|X}$  also satisfies  $(\epsilon_1, \delta)$ -EML which was expected from Proposition 5.21.

As the final topic in this section, we discuss the relationship between the  $(\epsilon, \delta)$ -EML and  $(\epsilon, \delta)$ -PML privacy guarantees. By Examples 5.11 and 5.22, it is clear that  $(\epsilon, \delta)$ -PML does not imply  $(\epsilon, \delta)$ -EML. In general,  $(\epsilon, \delta)$ -EML does not imply  $(\epsilon, \delta)$ -PML either. For example, consider the binary symmetric channel

$$P_{Y|X} = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix},$$

with uniform  $P_X$ . It is straightforward to verify that  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta_1)$ -EML with  $\epsilon_1 = \log \frac{34}{30}$  and  $\delta_1 = 0.6$ . However, due to the symmetry of the channel,

$P_{Y|X}$  satisfies  $(\epsilon_2, \delta_2)$ -PML with  $\epsilon_2 = \log \frac{6}{5}$  and all  $0 \leq \delta_2 < 1$ . Note that for  $\delta_1 = \delta_2 = 0.6$  we have  $\epsilon_1 < \epsilon_2$ .

While in general  $(\epsilon, \delta)$ -EML does not imply  $(\epsilon, \delta)$ -PML, there exists a condition under which  $(\epsilon, \delta)$ -EML does in fact imply  $(\epsilon, \delta)$ -PML. This condition is described below, and is proved in Appendix 5.F.

**Proposition 5.23.** *Given an arbitrary but fixed prior  $P_X$ , suppose the privacy mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -EML. Let  $\mathcal{A} = \{y \in \text{supp}(P_Y) : \ell(X \rightarrow y) > \epsilon\}$ . If there exists  $x^* \in \text{supp}(P_X)$  satisfying  $x^* \in \arg \max_{x \in \text{supp}(P_X)} i(x; y)$  for all  $y \in \mathcal{A}$ , then  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -PML.*

## 5.4 Privacy Guarantees: Composition Properties

A second group of properties that are helpful while assessing the privacy levels of more complicated systems are composition properties. Let  $X$  denote the sensitive data, and let  $Y$  be the output of a channel  $P_{Y|X}$ . Suppose  $X$  and  $Y$  are inputs to the channel  $P_{Z|XY}$  inducing a random variable  $Z$  in its output. What we are interested in is to find out how much information the overall channel  $P_{ZY|X}$  leaks about  $X$ .

The composition property stated in Lemma 5.2 describes an upper bound on the leakage resulting from composing the two channels  $P_{Y|X}$  and  $P_{Z|XY}$ . In this section, our goal is to understand how different privacy guarantees, namely,  $\epsilon$ -PML,  $(\epsilon, \delta)$ -PML and  $(\epsilon, \delta)$ -EML are affected by adaptively composing two channels. Naturally, one can formulate various problems by making different assumptions about the channels  $P_{Y|X}$  and  $P_{Z|XY}$ . The following result contains several such problem formulations and results, and its proof is deferred to Appendix 5.G.

**Theorem 5.24.** *Consider three random variables  $X$ ,  $Y$  and  $Z$  where  $X$  denotes the secret,  $Y$  is the output of a channel  $P_{Y|X}$ , and  $Z$  is the output of a channel  $P_{Z|XY}$ .*

- (i) *Suppose  $P_{Y|X}$  satisfies  $\epsilon_1$ -PML, and for all  $y \in \mathcal{Y}$ ,  $P_{Z|X, Y=y}$  satisfies  $\epsilon_2$ -PML with  $\epsilon_1, \epsilon_2 \geq 0$ . Then,  $P_{YZ|X}$  satisfies  $\epsilon$ -PML with  $\epsilon = \epsilon_1 + \epsilon_2$ .*
- (ii) *Suppose  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta_1)$ -PML, and for all  $y \in \mathcal{Y}$ ,  $P_{Z|X, Y=y}$  satisfies  $(\epsilon_2, \delta_2)$ -PML. Then,  $P_{YZ|X}$  satisfies  $(\epsilon, \delta)$ -PML with  $\epsilon = \epsilon_1 + \epsilon_2$  and  $\delta = \delta_1 + \delta_2 - \delta_1 \delta_2$ .*
- (iii) *Suppose  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta_1)$ -PML, and*

$$\mathbb{P}_{(Y,Z) \sim P_{YZ}} \left[ \ell(X \rightarrow Z | Y) \leq \epsilon_2 \right] \geq 1 - \delta_2.$$

*Then, the channel  $P_{YZ|X}$  satisfies  $(\epsilon, \delta)$ -PML with  $\epsilon = \epsilon_1 + \epsilon_2$  and  $\delta = \delta_1 + \delta_2$ .*

(iv) Suppose  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta_1)$ -EML, and for all  $y \in \mathcal{Y}$ ,  $P_{Z|X, Y=y}$  satisfies  $(\epsilon_2, \delta_2)$ -EML. Given an event  $\mathcal{E} \subseteq \mathcal{Y} \times \mathcal{Z}$ , define the sets

$$\begin{aligned}\mathcal{E}_Y &:= \{y \in \mathcal{Y} : (y, z) \in \mathcal{E} \text{ for some } z \in \mathcal{Z}\} \\ \mathcal{E}_Z(y) &:= \{z \in \mathcal{Z} : (y, z) \in \mathcal{E}\}.\end{aligned}$$

If  $0 \leq \delta_2 \leq \min_{y \in \mathcal{E}_Y} P_{Z|Y=y}(\mathcal{E}_Z(y))$ , then  $P_{YZ}(\mathcal{E}) \geq \delta_1$  implies  $\ell(X \rightarrow \mathcal{E}) \leq \epsilon_1 + \epsilon_2$ .

Specifically, if  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta_1)$ -EML, and for all  $y \in \mathcal{Y}$ ,  $P_{Z|X, Y=y}$  satisfies  $\epsilon_2$ -PML, then  $P_{YZ|X}$  satisfies  $(\epsilon_1 + \epsilon_2, \delta_1)$ -EML.

(v) Suppose  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta_1)$ -EML, and for all  $y \in \mathcal{Y}$ ,  $P_{Z|X, Y=y}$  satisfies  $(\epsilon_2, \delta_2)$ -EML. Then,  $P_{YZ|X}$  satisfies  $(\epsilon, \delta)$ -EML with

$$\epsilon = \log \left( \frac{\delta_2}{\delta_1 + \delta_2} \cdot \exp(\epsilon_{\max}) + \exp(\epsilon_1 + \epsilon_2) \right),$$

and  $\delta = \delta_1 + \delta_2$ , where  $\epsilon_{\max} := -\log \min_{x \in \text{supp}(P_X)} P_X(x)$ .

## 5.5 Relationship to Local Privacy Measures and Statistical Notions

In this section, we discuss how pointwise maximal leakage and the privacy guarantees defined in the previous section relate to several existing privacy/statistical notions from the literature. More specifically, we discuss max-information [39, 118], local differential privacy [72, 31], local information privacy [22, 66], local differential identifiability [83, 143], mutual information,  $f$ -information [27], and total variation privacy [114].

Here, we consider the local setup, i.e., we assume that the goal of the privacy mechanism  $P_{Y|X}$  is to protect the whole  $X$ , presumably the data of a single individual. The results of this section are summarized in Table 5.5.1.

### Max-information

Max-information is a statistical quantity that was introduced as a tool for studying generalization in adaptive data analysis [39, 118]. Note that while max-information has not been developed as a notion of privacy, it is defined similarly to pointwise maximal leakage, and therefore, their comparison is appropriate. Before we state the definition of (approximate) max-information, we need the following definition of *approximate max-divergence* which is a weakening of *max-divergence*, i.e., Rényi divergence of order infinity.

**Definition 5.25** (Approximate max-divergence [39]). Let  $P$  and  $Q$  be probability measures on the finite set  $\Omega$  and suppose  $P \ll Q$ . Given  $0 \leq \delta \leq 1$ , the  $\delta$ -

Table 5.5.1: Summary of the Results of Section 5.5

Privacy/Statistical Notion	Relation/Bound	Ref.
Max-information	$I_\infty(X; Y) = \max_y \ell(X \rightarrow y)$	Def. 5.26
Approximate max-information	$(\epsilon, \delta)$ -PML $\implies I_\infty^\delta(X; Y) \leq \epsilon$	Prop. 5.28
Local information privacy	$\epsilon$ -LIP $\implies \epsilon$ -PML	Def. 3.24
Local differential privacy	$\epsilon$ -LDP $\implies \log \frac{1}{p_{\min} + \epsilon^{-\epsilon}(1 - p_{\min})}$ -PML	Prop. 5.30
Local differential identifiability	$\epsilon$ -LDI $\implies \log \frac{1}{p_{\min}(1 + e^{-\epsilon}(\text{supp}(P_X) - 1))}$ -PML	Prop. 5.32
$f$ -information	$I_f(X; Y) \leq \mathbb{E}_{Y \sim P_Y} \left[ \max \{f(\exp(\ell(X \rightarrow Y))), f(0)\} \right]$	Prop. 5.34
Mutual information	$I(X; Y) \leq \mathbb{E}_{Y \sim P_Y} [\ell(X \rightarrow Y)]$	Prop. 5.35
Total variation privacy	$T(X; Y) \leq \min \left\{ \frac{1}{2} \mathbb{E}_{Y \sim P_Y} \left[ \max \{ \exp(\ell(X \rightarrow Y)) - 1, 1 \} \right], \exp(\mathcal{L}(P_{Y X})) - 1 \right\}$	Prop. 5.34 Prop. 5.36

approximate max-divergence between  $P$  and  $Q$  is defined as

$$D_\infty^\delta(P\|Q) = \log \max_{\mathcal{E} \subseteq \Omega, P(\mathcal{E}) \geq \delta} \frac{P(\mathcal{E}) - \delta}{Q(\mathcal{E})}.$$

Note that if  $\delta = 0$ , then the above definition reduces to the max-divergence between  $P$  and  $Q$ , denoted by  $D_\infty(P\|Q)$ .

**Definition 5.26** ((Approximate) max-information [39]). Suppose  $V$  and  $W$  are random variables on finite sets  $\mathcal{V}$  and  $\mathcal{W}$ , respectively, and let  $P_{VW}$  denote their joint distribution. The max-information between  $V$  and  $W$  is defined as

$$\begin{aligned} I_\infty(V; W) &:= D_\infty(P_{VW}\|P_V \times P_W) \\ &= \log \max_{v \in \mathcal{V}, w \in \mathcal{W}} \frac{P_{VW}(v, w)}{P_V(v)P_W(w)}. \end{aligned}$$

Similarly, the  $\delta$ -approximate max-information between  $V$  and  $W$  is defined as

$$\begin{aligned} I_\infty^\delta(V; W) &:= D_\infty^\delta(P_{VW}\|P_V \times P_W) \\ &= \log \max_{\mathcal{E} \subseteq \mathcal{V} \times \mathcal{W}: P_{VW}(\mathcal{E}) \geq \delta} \frac{P_{VW}(\mathcal{E}) - \delta}{(P_V \times P_W)(\mathcal{E})}, \end{aligned}$$

where  $(P_V \times P_W)(\mathcal{E}) = \sum_{(v,w) \in \mathcal{E}} P_V(v)P_W(w)$ .

It follows from the definition of max-information that, given a fixed prior  $P_X$ , a mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -PML if and only if  $I_\infty(X; Y) \leq \epsilon$ . Therefore, below we examine how  $(\epsilon, \delta)$ -PML privacy compares with guarantees given in terms of the  $\delta$ -approximate max-information. To do this, first, we recall a lemma from [40].

**Lemma 5.27** ([40, Lemma 18]). *Let  $P$  and  $Q$  be probability measures on the finite set  $\Omega$  and suppose  $P \ll Q$ . Let the event  $\mathcal{O} \subset \Omega$  be defined as*

$$\mathcal{O} := \{\omega \in \Omega : \frac{P(\omega)}{Q(\omega)} > e^\epsilon\}.$$

*If  $P(\mathcal{O}) \leq \delta$ , then  $D_\infty^\delta(P\|Q) \leq \epsilon$ .*

Now, we use Lemma 5.27 to relate  $\mathbb{P}_{Y \sim P_Y}[\ell(X \rightarrow Y) \leq \epsilon]$  and the  $\delta$ -approximate max-information. The proposition is proved in Appendix 5.H.

**Proposition 5.28.** *Given an arbitrary but fixed  $P_X$ , if the channel  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -PML, then  $I_\infty^\delta(X; Y) \leq \epsilon$ .*

The previous result shows that  $(\epsilon, \delta)$ -PML is a stronger guarantee compared to  $I_\infty^\delta(X; Y) \leq \epsilon$ . Roughly speaking, this is because under a  $I_\infty^\delta(X; Y) \leq \epsilon$  guarantee, the “good”  $y$ ’s are those that have small information density  $i(x; y)$  with high probability over the  $x$ ’s. However, under an  $(\epsilon, \delta)$ -PML guarantee, the “good”  $y$ ’s need to have small  $i(x; y)$  for all  $x$ ’s in  $\text{supp}(P_X)$ , that is,  $i(x; y)$  must be small with probability one over the  $x$ ’s. In addition, note that  $I_\infty^\delta(X; Y)$  treats random variables  $X$  and  $Y$  symmetrically, and the probability of a good event is calculated according to  $P_{XY}$ . On the other hand, under  $(\epsilon, \delta)$ -PML, the probability of a good event (i.e., low leakage) is calculated according to  $P_Y$  over the  $y$ ’s.

## Local Differential Privacy

Recall from Definition 3.10 that a privacy mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -LDP if

$$\max_{x, x' \in \mathcal{X}, y \in \mathcal{Y}} \frac{P_{Y|X=x}(y)}{P_{Y|X=x'}(y)} \leq e^\epsilon.$$

As discussed in Theorem 3.21, taking the supremum over all  $P_X$  of maximal realizable leakage (Definition 3.19) yields the expression for LDP. More precisely, it holds that

$$\sup_{P_X \in \mathcal{P}_X} D_\infty(P_{Y|X} \times P_X \| P_Y \times P_X) = \log \max_{y \in \mathcal{Y}} \max_{x, x' \in \mathcal{X}} \frac{P_{Y|X=x}(y)}{P_{Y|X=x'}(y)}, \quad (5.10)$$

where  $\mathcal{P}_X$  denotes the set of distributions that have full support on  $\mathcal{X}$ . From this, we immediately get the following result connecting  $\epsilon$ -PML and  $\epsilon$ -LDP.

**Theorem 5.29.** *Given  $\epsilon \geq 0$ , a privacy mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -LDP if and only if it satisfies  $\epsilon$ -PML for all  $P_X$  with full support on  $\mathcal{X}$ .*

Note that in (5.10), the supremum is not attained by any distribution in  $\mathcal{P}_{\mathcal{X}}$ . Instead, there exists a sequence of distributions with decreasing entropy (i.e., converging to a vertex of  $\mathcal{P}_{\mathcal{X}}$ ) for which the quantity  $D_{\infty}(P_{Y|X} \times P_X \| P_Y \times P_X)$  approaches  $\log \max_{y \in \mathcal{Y}} \max_{x, x' \in \mathcal{X}} \frac{P_{Y|X=x}(y)}{P_{Y|X=x'}(y)}$ . Informally, this implies that achieving privacy is more challenging when  $P_X$  has a small entropy. We revisit this point both in the next section where we connect PML and DP and also in Chapter 6.

By Theorem 5.29, if  $P_{Y|X}$  satisfies  $\epsilon$ -LDP, then it satisfies  $\epsilon$ -PML for all  $P_X \in \mathcal{P}_{\mathcal{X}}$ . However, when  $P_X$  is known, we can get a tighter bound on PML. Proposition 5.30 is proved in Appendix 5.I.

**Proposition 5.30.** *Suppose  $X$  is distributed according to  $P_X$  and let  $p_{\min} := \min_{x \in \text{supp}(P_X)} P_X(x)$ . If  $P_{Y|X}$  satisfies  $\epsilon$ -LDP, then it satisfies  $\epsilon'$ -PML with*

$$\epsilon' = -\log \left( p_{\min} + e^{-\epsilon}(1 - p_{\min}) \right). \quad (5.11)$$

Note that as expected,  $\epsilon' \rightarrow \epsilon$  as  $p_{\min} \rightarrow 0$ .

Below, we examine the randomized response (RR) mechanism (Definition 3.11) which is one of the most common implementations of LDP. Notably, the translation between LDP and PML described by (5.11) is precise for the RR mechanism.

**Example 5.31** (Randomized response). Suppose  $\mathcal{X} = \mathcal{Y} = [n]$ . Given  $\epsilon \geq 0$ , let  $P_{Y|X}$  be the RR mechanism satisfying  $\epsilon$ -LDP, that is

$$P_{Y|X=x}(y) = \begin{cases} \frac{e^{\epsilon}}{n-1+e^{\epsilon}} & \text{if } x = y, \\ \frac{1}{n-1+e^{\epsilon}} & \text{if } x \neq y. \end{cases}$$

It is easy to see that the RR mechanism satisfies  $\epsilon'$ -PML with  $\epsilon' = -\log \left( p_{\min} + e^{-\epsilon}(1 - p_{\min}) \right)$ .

### Local Information Privacy

As discussed in Chapter 3, a privacy mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -LIP if

$$e^{-\epsilon} \leq \frac{P_{X|Y=y}(x)}{P_X(x)} \leq e^{\epsilon},$$

for all  $y \in \mathcal{Y}$  and  $x \in \mathcal{X}$ , assuming that  $X$  is distributed according to  $P_X$ .

Clearly,  $\epsilon$ -LIP implies  $\epsilon$ -PML. Observe that LIP differs from PML in the additional lower bound on the information density. That is,  $\epsilon$ -LIP requires that

$$\frac{P_X(x)}{P_{X|Y=y}(x)} \leq e^\epsilon, \quad x \in \mathcal{X}, y \in \mathcal{Y}.$$

The above bound has no clear operational interpretation in our current framework and may be superfluous. To see why, consider the following simple example. Suppose  $X$  is a uniformly distributed binary random variable. Assume  $P_{X|Y=y}(0) = p$  for some  $y \in \mathcal{Y}$ , where  $p > 0$  is a small positive number. Therefore,  $\exp(i_{P_{XY}}(0; y))$  is small. Then,  $P_{X|Y=y}(1) = 1 - p$  with  $1 - p$  close to one, and  $\ell(X \rightarrow y)$  is

$$\ell(X \rightarrow y) = \max_{x \in \{0,1\}} i_{P_{XY}}(x; y) = i_{P_{XY}}(1; y) = \log(2(1 - p)),$$

which is close to  $\epsilon_{\max} = \log 2$ . Hence, the outcome  $y$  has a large information leakage. This example shows that small values of  $\exp(i(x; y))$  can increase the pointwise maximal leakage simply because we must have

$$\sum_x P_{X|Y=y}(x) = \sum_x \exp(i(x; y)) \cdot P_X(x) = 1,$$

for all  $y \in \mathcal{Y}$ . As such, it may not be necessary to impose a lower bound on the information density as a separate constraint; a privacy guarantee defined based on an upper bound on information density will be automatically penalized for small values of  $\exp(i(x; y))$ . See also [56] for more discussions on the relationship between the lower and upper bounds on information density.

## Local Differential Identifiability

Recall from Definition 3.25 that a mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -LDI if

$$\max_{x, x' \in \mathcal{X}, y \in \mathcal{Y}} \frac{P_{X|Y=y}(x)}{P_{X|Y=y}(x')} \leq e^\epsilon,$$

assuming that  $X$  is distributed according to  $P_X$ .

The following result, proved in Appendix 5.J, describes the relationship between  $\epsilon$ -LDI and PML.

**Proposition 5.32.** *Suppose  $X$  is distributed according to  $P_X$  and let  $p_{\min} := \min_{x \in \text{supp}(P_X)} P_X(x)$ . If  $P_{Y|X}$  satisfies  $\epsilon$ -LDI, then it satisfies  $\tilde{\epsilon}$ -PML with*

$$\tilde{\epsilon} = -\log \left( p_{\min} (1 + e^{-\epsilon} (|\text{supp}(P_X)| - 1)) \right).$$

*Remark 5.33.* A common characteristic among the three notions of privacy discussed above, namely LDP, LIP, and LDI, is their strict intolerance of zero-probability assignments in the channel  $P_{Y|X}$ . Put differently, the existence of a single input-output pair  $(x, y) \in \text{supp}(P_{XY})$  with  $P_{Y|X=x}(y) = 0$  immediately implies that the channel does not satisfy any of the above notions of privacy. On the other hand,  $\ell(X \rightarrow y)$  is always bounded by  $\epsilon_{\max}$ , so a guarantee given in terms of pointwise maximal leakage, in general, may not satisfy any of the above notions. Nonetheless, we show in the next chapter (Proposition 6.10) that when  $\epsilon < \log \frac{1}{1-p_{\min}}$ ,  $P_{Y|X=x}(y)$  is forced to be positive for all  $(x, y) \in \text{supp}(P_{XY})$ . In this case,  $\epsilon$ -PML yields guarantees in terms of LDP, LIP, and LDI. See also [55] and [56] for more discussions on this point.

In addition, zero-probability assignments in the channel  $P_{Y|X}$  may not necessarily imply “bad privacy”, at least when  $P_X$  is a high-entropy distribution. Consider the following simple example. Suppose  $X$  and  $Y$  are random variables defined on sets with cardinality  $n$  and assume that  $X$  is uniformly distributed. Consider the following channel:

$$P_{Y|X=x_1}(y_i) = \begin{cases} 0, & \text{if } i = 1, \\ \frac{1}{n-1} & \text{otherwise,} \end{cases}$$

and  $P_{Y|X=x_j}(y_i) = \frac{1}{n}$  with  $i \in \{1, \dots, n\}$  and  $j \in \{2, \dots, n\}$ . Intuitively, if  $n$  is large, then  $P_{Y|X}$  leaks very little information which also becomes apparent by calculating the leakage:

$$\begin{aligned} \ell(X \rightarrow y_1) &= \log \frac{n}{n-1}, \\ \ell(X \rightarrow y_i) &= \log \frac{n^2}{n^2 - n + 1}, \quad i = 2, \dots, n. \end{aligned}$$

However, under LDP/LIP/LDI, no matter how large  $n$  is, the above channel is considered to be equally non-private as a deterministic mapping from  $X$  to  $Y$ . This may be an overly pessimistic assessment.

## ***f*-information**

Let  $f : (0, \infty) \rightarrow \mathbb{R}$  be a convex function satisfying  $f(1) = 0$ . As discussed in Chapter 3,  $f$ -information is defined as the  $f$ -divergence of a joint distribution from the product of the marginals. More precisely,

$$I_f(X; Y) = D_f(P_{XY} \| P_X \times P_Y).$$

Below, we describe how PML upper bounds  $f$ -information. The result is proved in Appendix 5.L.

**Proposition 5.34.** *Let  $f : (0, \infty) \rightarrow \mathbb{R}$  be a convex function satisfying  $f(1) = 0$ , and suppose  $\lim_{t \downarrow 0} f(t) < \infty$ . Then, for all joint distributions  $P_{XY}$  on  $\mathcal{X} \times \mathcal{Y}$  we have*

$$I_f(X; Y) \leq \mathbb{E}_{Y \sim P_Y} \left[ \max \left\{ f \left( \exp \left( \ell(X \rightarrow Y) \right) \right), f(0) \right\} \right],$$

where  $f(0)$  is defined by continuity as  $f(0) := \lim_{t \downarrow 0} f(t)$ .

Proposition 5.34 describes a general bound that holds for all  $f$  satisfying  $\lim_{t \rightarrow 0} f(t) < \infty$ . Naturally, the bound can be tightened for specific  $f$ 's. Below, we consider two examples: mutual information, corresponding to  $f(t) = t \log t$ , and

$$T(X; Y) := \mathbb{E}_{Y \sim P_Y} \left[ \text{TV}(P_{X|Y}(\cdot | Y), P_X) \right],$$

used to define *total variation privacy* [114] which corresponds to  $f(t) = \frac{1}{2}|t - 1|$ .

First, we bound mutual information in terms of PML. The result is proved in Appendix 5.K.

**Proposition 5.35.** *For all joint distributions  $P_{XY}$  on  $\mathcal{X} \times \mathcal{Y}$  we have*

$$I(X; Y) \leq \mathbb{E}_{Y \sim P_Y} [\ell(X \rightarrow Y)], \quad (5.12)$$

with equality if and only if  $P_{Y|X=x}(y) = P_{Y|X=x'}(y)$  for all  $x, x' \in \mathcal{X}$  and  $y \in \mathcal{Y}$  such that  $P_{XY}(x, y) > 0$  and  $P_{XY}(x', y) > 0$ .

Note that the above bound is tighter than both the bound in Proposition 5.34 and also the bound  $I(X; Y) \leq \mathcal{L}(P_{Y|X})$  obtained in [63, Lemma 2].

Now, we bound  $T(X; Y)$  in terms of maximal leakage and also show how the  $(\epsilon, \delta)$ -PML privacy guarantee restricts  $T(X; Y)$ . The result is proved in Appendix 5.M.

**Proposition 5.36.** *For all joint distributions  $P_{XY}$  on  $\mathcal{X} \times \mathcal{Y}$  we have*

$$T(X; Y) \leq \exp(\mathcal{L}(P_{Y|X})) - 1.$$

Furthermore, if the mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -PML, then  $T(X; Y)$  is bounded as follows:

- (i) if  $\epsilon \leq \log \frac{3}{2}$ , then  $T(X; Y) \leq e^\epsilon - 1 + \frac{\delta}{2} (e^{\epsilon_{\max}} - 1)$ ,
- (ii) if  $\log \frac{3}{2} \leq \epsilon \leq \log 2$ , then  $T(X; Y) \leq \frac{1}{2} + \frac{\delta}{2} (e^{\epsilon_{\max}} - 1)$ , and
- (iii) if  $\epsilon \geq \log 2$ , then  $T(X; Y) \leq \frac{1}{2} (e^\epsilon - 1) + \frac{\delta}{2} (e^{\epsilon_{\max}} - 1)$ .

Rassouli and Gündüz [114] derived the following bound between  $T(X; Y)$  and maximal leakage:

$$T(X; Y) \leq (|\mathcal{X}| - 1) \cdot \max_{x \in \mathcal{X}} P_X(x) \cdot (\exp(\mathcal{L}(P_{Y|X})) - 1), \quad (5.13)$$

which is rather loose as it depends on the cardinality of  $\mathcal{X}$  (considering that  $0 \leq T(X; Y) \leq 1$ ). Note that applying Proposition 5.34 with  $f(x) = \frac{1}{2}|x - 1|$  we get

$$T(X; Y) \leq \frac{1}{2} \mathbb{E}_{Y \sim P_Y} \left[ \max \left\{ \exp \left( \ell(X \rightarrow Y) \right) - 1, 1 \right\} \right], \quad (5.14)$$

which is tighter than (5.13).

## 5.6 Relationship to Notions of Database Privacy

In this section, we discuss how PML relates to differential privacy and free-lunch privacy (Definition 3.3). Here, we consider the centralized setting where  $X$  is a database and  $Y$  represents the answer to a question posed about  $X$  returned by the mechanism  $P_{Y|X}$ .

Our setup in this section is slightly more general than the previous sections as we no longer assume that  $Y$  is a finite random variable. This more general setup allows us to examine the counting query and the Laplace mechanism (Definition 3.4) through the lens of PML. Given  $x \in \mathcal{X}$ , we use  $p_{Y|X=x}$  to denote the density of  $P_{Y|X=x}$  with respect to a suitable  $\sigma$ -finite measure on  $\mathcal{Y}$ , for example, the Lebesgue measure when  $\mathcal{Y} = \mathbb{R}$ . Similarly, we use  $p_Y$  to denote the density of  $P_Y$ .

Suppose  $X$  is a random variable representing a database containing  $n$  entries. Given  $i \in [n]$ , let  $D_i$  be the random variable corresponding to the  $i$ -th entry, which takes values in a finite alphabet  $\mathcal{D}$ . Then, each database (realization)  $x = (d_1, \dots, d_n) \in \mathcal{D}^n$  is an  $n$ -tuple and  $X$  is a sequence of  $n$  random variables. Suppose  $P_X = P_{D_1, \dots, D_n}$  denotes the distribution according to which databases are drawn from  $\mathcal{D}^n$ . To obtain the probability distribution describing the  $i$ -th entry we marginalize over the remaining  $n - 1$  entries, that is, for each  $d_i \in \mathcal{D}$  and  $i \in [n]$  we have

$$p_{D_i}(d_i) = \sum_{d_{-i} \in \mathcal{D}^{n-1}} p_{D_i|D_{-i}=d_{-i}}(d_i) p_{D_{-i}}(d_{-i}),$$

where  $d_{-i} := (d_1, \dots, d_{i-1}, d_{i+1}, \dots, d_n) \in \mathcal{D}^{n-1}$  is a tuple describing the database with its  $i$ -th entry removed. Our setup is very general in the sense that the entries can be arbitrarily correlated. Note that for notational consistency, we also use densities, e.g.,  $p_{D_i}$  to represent probability mass functions on  $\mathcal{X}$ .

### PML and Differential Privacy

$\epsilon$ -DP has already been defined in Chapter 3, but here we restate it in our current notation.<sup>3</sup>

<sup>3</sup>This definition corresponds to bounded differential privacy. See Chapter 3 for more details.

**Definition 5.37** ( $\epsilon$ -DP). Given  $\epsilon \geq 0$ , we say that the privacy mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -DP if

$$\sup_{y \in \mathcal{Y}} \max_{d_i, d'_i \in \mathcal{D}: \substack{d_{-i} \in \mathcal{D}^{n-1} \\ i \in [n]}} \max_{d_{-i} \in \mathcal{D}^{n-1}} \log \frac{p_{Y|D_i=d_i, D_{-i}=d_{-i}}(y)}{p_{Y|D_i=d'_i, D_{-i}=d_{-i}}(y)} \leq \epsilon.$$

Let  $\mathcal{P}_{\mathcal{X}}$  denote the set of all distributions with full support on  $\mathcal{X} = \mathcal{D}^n$ . Furthermore, let  $\mathcal{Q}_{\mathcal{X}}$  denote the set of product distributions in  $\mathcal{P}_{\mathcal{X}}$ , that is,  $\mathcal{Q}_{\mathcal{X}} := \{P_X \in \mathcal{P}_{\mathcal{X}} : P_X = \prod_{i=1}^n P_{D_i}\}$ . We now show that differential privacy admits multiple different but equivalent formulations in terms of PML.

**Theorem 5.38** (Differential privacy as a PML constraint). *Given  $\epsilon \geq 0$ , the privacy mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -differential privacy if and only if*

- (i)  $\sup_{y \in \mathcal{Y}} \sup_{P_X \in \mathcal{P}_{\mathcal{X}}} \max_{d_{-i} \in \mathcal{D}^{n-1}: \substack{d_{-i} \in \mathcal{D}^{n-1} \\ i \in [n]}} \ell(D_i \rightarrow y \mid d_{-i}) \leq \epsilon$ , or,
- (ii)  $\sup_{y \in \mathcal{Y}} \sup_{P_X \in \mathcal{Q}_{\mathcal{X}}} \max_{d_{-i} \in \mathcal{D}^{n-1}: \substack{d_{-i} \in \mathcal{D}^{n-1} \\ i \in [n]}} \ell(D_i \rightarrow y \mid d_{-i}) \leq \epsilon$ , or,
- (iii)  $\sup_{y \in \mathcal{Y}} \sup_{P_X \in \mathcal{Q}_{\mathcal{X}}} \max_{i \in [n]} \ell(D_i \rightarrow y) \leq \epsilon$ .

*Proof.* Fix an arbitrary  $i \in [n]$ ,  $y \in \mathcal{Y}$ , and  $P_X \in \mathcal{P}_{\mathcal{X}}$ . We have

$$\begin{aligned} & \max_{d_{-i} \in \mathcal{D}^{n-1}} \exp\left(\ell(D_i \rightarrow y \mid d_{-i})\right) \\ &= \max_{d_{-i} \in \mathcal{D}^{n-1}} \max_{d_i \in \text{supp}(P_{D_i|D_{-i}=d_{-i}})} \frac{p_{Y|D_{-i}=d_{-i}, D_i=d_i}(y)}{p_{Y|D_{-i}=d_{-i}}(y)} \end{aligned} \quad (5.15a)$$

$$= \max_{d_{-i} \in \mathcal{D}^{n-1}} \max_{d_i \in \mathcal{D}} \frac{p_{Y|D_{-i}=d_{-i}, D_i=d_i}(y)}{p_{Y|D_{-i}=d_{-i}}(y)} \quad (5.15b)$$

$$\begin{aligned} &= \max_{d_{-i} \in \mathcal{D}^{n-1}} \max_{d_i \in \mathcal{D}} \frac{p_{Y|D_{-i}=d_{-i}, D_i=d_i}(y)}{\sum_{d'_i \in \mathcal{D}} p_{Y|D_{-i}=d_{-i}, D_i=d'_i}(y) p_{D_i|D_{-i}=d_{-i}}(d'_i)} \\ &\leq \max_{d_{-i} \in \mathcal{D}^{n-1}} \max_{d_i \in \mathcal{D}} \frac{p_{Y|D_{-i}=d_{-i}, D_i=d_i}(y)}{\left(\min_{d'_i \in \mathcal{D}} p_{Y|D_{-i}=d_{-i}, D_i=d'_i}(y)\right) \sum_{d'_i \in \mathcal{D}} p_{D_i|D_{-i}=d_{-i}}(d'_i)} \end{aligned} \quad (5.15c)$$

$$= \max_{d_{-i} \in \mathcal{D}^{n-1}} \max_{d_i \in \mathcal{D}} \frac{p_{Y|D_{-i}=d_{-i}, D_i=d_i}(y)}{\min_{d'_i \in \mathcal{D}} p_{Y|D_{-i}=d_{-i}, D_i=d'_i}(y)}$$

$$= \max_{d_{-i} \in \mathcal{D}^{n-1}} \max_{d_i, d'_i \in \mathcal{D}} \frac{p_{Y|D_{-i}=d_{-i}, D_i=d_i}(y)}{p_{Y|D_{-i}=d_{-i}, D_i=d'_i}(y)},$$

where (5.15a) follows from (5.1), and (5.15b) uses the fact that  $\text{supp}(P_{D_i|D_{-i}=d_{-i}}) = \mathcal{D}$  for each  $P_X \in \mathcal{P}_X$ .

Next, we show that the above inequality holds with equality for a product distribution  $P_X^* \in \mathcal{Q}_X$ . This then proves that (i) and (ii) in the statement of the theorem are equivalent to each other and to differential privacy. Let  $\varepsilon > 0$  be a small constant. Suppose  $P_X^* = \prod_{i=1}^n P_{D_i}^*$ , where

$$p_{D_i}^*(d'_i) := \begin{cases} 1 - \varepsilon, & \text{for some } d'_i \in \arg \min_{\bar{d}_i \in \mathcal{D}} p_{Y|D_{-i}=d_{-i}, D_i=\bar{d}_i}(y), \\ \frac{\varepsilon}{|\mathcal{D}|-1}, & \text{otherwise.} \end{cases}$$

Then,  $\sum_{d'_i \in \mathcal{D}} p_{Y|D_{-i}=d_{-i}, D_i=d'_i}(y) p_{D_i}^*(d'_i) \rightarrow \min_{d'_i \in \mathcal{D}} p_{Y|D_{-i}=d_{-i}, D_i=d'_i}(y)$  as  $\varepsilon \rightarrow 0$ .

Thus, inequality (5.15c) holds with equality for  $P_X^*$ .

Now, we show that (iii) in the statement is also equivalent to differential privacy. Fix an arbitrary  $i \in [n]$  and  $y \in \mathcal{Y}$ . Note that each  $P_X \in \mathcal{Q}_X$  can be written as  $P_X = P_{D_i} \times P_{D_{-i}}$ ; hence, we can optimize over  $P_{D_i}$  and  $P_{D_{-i}}$  separately:

$$\sup_{P_{D_{-i}}} \sup_{P_{D_i}} \exp\left(\ell(D_i \rightarrow y)\right) = \sup_{P_{D_{-i}}} \max_{d_i, d'_i} \frac{p_{Y|D_i=d_i}(y)}{p_{Y|D_i=d'_i}(y)} \quad (5.16a)$$

$$\begin{aligned} &= \max_{d_i, d'_i} \sup_{P_{D_{-i}}} \frac{\sum_{d_{-i}} p_{Y|D_i=d_i, D_{-i}=d_{-i}}(y) p_{D_{-i}}(d_{-i})}{\sum_{d_{-i}} p_{Y|D_i=d'_i, D_{-i}=d_{-i}}(y) p_{D_{-i}}(d_{-i})} \\ &\leq \max_{d_i, d'_i} \max_{d_{-i}} \frac{p_{Y|D_i=d_i, D_{-i}=d_{-i}}(y)}{p_{Y|D_i=d'_i, D_{-i}=d_{-i}}(y)}, \end{aligned} \quad (5.16b)$$

where (5.16a) is due to Theorem 5.29. To show that inequality (5.16b) can be attained, for fixed  $d_i$  and  $d'_i$  let

$$d_{-i}^* = (d_1^*, \dots, d_{i-1}^*, d_{i+1}^*, \dots, d_n^*) \in \arg \max_{\bar{d}_{-i}} \frac{p_{Y|D_i=d_i, D_{-i}=\bar{d}_{-i}}(y)}{p_{Y|D_i=d'_i, D_{-i}=\bar{d}_{-i}}(y)}.$$

Consider the pmf  $q_{D_j}^*$  defined by

$$q_{D_j}^*(d_j) := \begin{cases} 1 - \varepsilon, & d_j = d_j^*, \\ \frac{\varepsilon}{|\mathcal{D}|-1}, & \text{otherwise,} \end{cases} \quad (5.17)$$

for  $j \neq i$ . Let  $q_{D_{-i}}^* = \prod_{j \neq i} q_{D_j}^*$  which satisfies  $q_{D_{-i}}^*(d_{-i}^*) = (1 - \varepsilon)^{n-1}$ , and  $q_{D_{-i}}^*(d_{-i}) \leq \frac{\varepsilon}{|\mathcal{D}|-1} (1 - \varepsilon)^{n-2}$  for all  $d_{-i} \neq d_{-i}^*$ . Then, for fixed  $n$ ,

$$\frac{\sum_{d_{-i}} p_{Y|D_i=d_i, D_{-i}=d_{-i}}(y) q_{D_{-i}}^*(d_{-i})}{\sum_{d_{-i}} p_{Y|D_i=d'_i, D_{-i}=d_{-i}}(y) q_{D_{-i}}^*(d_{-i})} \rightarrow \max_{d_{-i}} \frac{p_{Y|D_i=d_i, D_{-i}=d_{-i}}(y)}{p_{Y|D_i=d'_i, D_{-i}=d_{-i}}(y)},$$

as  $\varepsilon \rightarrow 0$ . Thus, inequality (5.16b) holds with equality for distribution  $Q_{D_{-i}}^*$ , which completes the proof.  $\square$

*Remark 5.39.* It is important to note that in all of the formulations above the supremum is never actually attained by any distribution in  $\mathcal{P}_X$  or  $\mathcal{Q}_X$ . For example, consider statement (i). Fix  $i \in [n]$  and  $d_{-i} \in \mathcal{D}^{n-1}$ , and suppose there exists  $Q_{D_i|D_{-i}=d_{-i}}^*$  such that

$$\sum_{d'_i \in \mathcal{D}} p_{Y|D_{-i}=d_{-i}, D_i=d'_i}(y) q_{D_i|D_{-i}=d_{-i}}^*(d'_i) = \min_{\tilde{d}_i \in \mathcal{D}} p_{Y|D_{-i}=d_{-i}, D_i=\tilde{d}_i}(y). \quad (5.18)$$

This equality holds if and only if  $p_{Y|D_{-i}=d_{-i}, D_i=d'_i}(y) = \min_{\tilde{d}_i \in \mathcal{D}} p_{Y|D_{-i}=d_{-i}, D_i=\tilde{d}_i}(y)$  for all  $d'_i \in \mathcal{D}$ . Since (5.18) must hold for all  $i$  and all  $d_{-i}$ , then  $p_{Y|D_{-i}=d_{-i}, D_i=d_i}(y)$  must be a constant that does not depend on  $d_i$  and  $d_{-i}$  for all  $y$ . However, this implies that  $X$  and  $Y$  are independent.

The first formulation of differential privacy in the above theorem is similar to a result of [36]. Specifically, [36, Claim 3] shows that differential privacy is equivalent to *semantic security* [36, Def. 6], where semantic security is defined by imposing both an upper bound and a lower bound on the posterior-prior ratio of all binary predicates of the data. The above result can then be considered as a generalization of [36, Claim 3] because it only requires an upper bound on the posterior-prior ratio and  $D_i$  is not restricted to be binary.

## PML and Free-lunch Privacy

Below, we define free-lunch privacy in our current notation and show how it can be expressed in terms of PML.

**Definition 5.40** ( $\epsilon$ -free-lunch privacy). Given  $\epsilon \geq 0$ , we say that the privacy mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -free-lunch privacy if

$$\sup_{y \in \mathcal{Y}} \max_{d^n, \tilde{d}^n \in \mathcal{D}^n} \log \frac{p_{Y|X=d^n}(y)}{p_{Y|X=\tilde{d}^n}(y)} \leq \epsilon.$$

**Theorem 5.41** (Free-lunch privacy as a PML constraint). *Given  $\epsilon \geq 0$ , the privacy mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -free-lunch privacy if and only if*

- (i)  $\sup_{y \in \mathcal{Y}} \sup_{P_X \in \mathcal{P}_X} \ell(X \rightarrow y) \leq \epsilon$ , or
- (ii)  $\sup_{y \in \mathcal{Y}} \sup_{P_X \in \mathcal{Q}_X} \ell(X \rightarrow y) \leq \epsilon$ , or
- (iii)  $\sup_{y \in \mathcal{Y}} \sup_{P_X \in \mathcal{P}_X} \max_{i \in [n]} \ell(D_i \rightarrow y) \leq \epsilon$ .

*Proof.* The proof is fairly similar to the proof of Theorem 5.38; thus, some details are removed. First, note that it follows directly from Theorem 5.29 that (i) in

the statement of the theorem is equivalent to  $\epsilon$ -free-lunch privacy. To prove that (ii) is equivalent to (i) we show that  $\sup_{P_X \in \mathcal{Q}_X} \ell(X \rightarrow y) \geq \sup_{P_X \in \mathcal{P}_X} \ell(X \rightarrow y)$  for all  $y \in \mathcal{Y}$  since the reverse inequality holds trivially. Consider the database  $x^* = (d_1^*, \dots, d_n^*) \in \arg \min_x P_{Y|X=x}(y)$ . We can use a construction similar to (5.17) to obtain a product distribution  $Q_X^*$  that satisfies  $q_X^*(x^*) = (1 - \epsilon)^n$  while  $q_X^*(x) \leq \frac{\epsilon}{|\mathcal{D}|-1} (1 - \epsilon)^{n-1}$  for all  $x \neq x^*$ . Then, we get

$$\begin{aligned} \sup_{P_X \in \mathcal{Q}_X} \exp\left(\ell(X \rightarrow y)\right) &\geq \exp\left(\ell_{P_{Y|X} \times Q_X^*}(X \rightarrow y)\right) \\ &= \frac{\max_x p_{Y|X=x}(x)}{\sum_{x'} p_{Y|X=x'}(y) q_X^*(x)} \\ &\geq \frac{\max_x p_{Y|X=x}(x)}{(1 - \epsilon)^n p_{Y|X=x^*}(y) + \frac{\epsilon}{|\mathcal{D}|-1} (1 - \epsilon)^{n-1} \sum_{x' \neq x^*} p_{Y|X=x'}(y)}. \end{aligned}$$

For fixed  $n$ , letting  $\epsilon \rightarrow 0$  yields

$$\begin{aligned} \sup_{P_X \in \mathcal{Q}_X} \exp\left(\ell(X \rightarrow y)\right) &\geq \frac{\max_x p_{Y|X=x}(x)}{\min_{x'} p_{Y|X=x'}(y)} \\ &= \sup_{P_X \in \mathcal{P}_X} \exp\left(\ell(X \rightarrow y)\right), \end{aligned}$$

as desired.

Next, we show that (iii) is equivalent to (i). By the pre-processing inequality for PML in Lemma 5.2, we have  $\ell(D_i \rightarrow y) \leq \ell(X \rightarrow y)$  for all  $i \in [n]$ ,  $y \in \mathcal{Y}$ , and  $P_X \in \mathcal{P}_X$ . So, we show that  $\sup_{P_X \in \mathcal{P}_X} \max_{i \in [n]} \ell(D_i \rightarrow y) \geq \sup_{P_X \in \mathcal{P}_X} \ell(X \rightarrow y)$  for all  $y \in \mathcal{Y}$ . Fix an arbitrary  $i \in [n]$ . We write  $P_X = P_{D_i} \times P_{D_{-i}|D_i}$  and optimize over  $P_{D_i}$  and  $P_{D_{-i}|D_i}$  separately:

$$\begin{aligned} \sup_{P_{D_{-i}|D_i}} \sup_{P_{D_i}} \exp\left(\ell(D_i \rightarrow y)\right) &= \sup_{P_{D_{-i}|D_i}} \max_{d_i} \sup_{P_{D_i}} \frac{p_{Y|D_i=d_i}(y)}{p_Y(y)} \\ &= \sup_{P_{D_{-i}|D_i}} \max_{d_i, d'_i} \frac{p_{Y|D_i=d_i}(y)}{p_{Y|D_i=d'_i}(y)} \tag{5.19a} \\ &= \max_{d_i, d'_i} \sup_{P_{D_{-i}|D_i}} \frac{\sum_{d_{-i}} p_{Y|D_i=d_i, D_{-i}=d_{-i}}(y) p_{D_{-i}|D_i=d_i}(d_{-i})}{\sum_{d'_{-i}} p_{Y|D_i=d'_i, D_{-i}=d'_{-i}}(y) p_{D_{-i}|D_i=d'_i}(d'_{-i})}, \end{aligned}$$

where (5.19a) follows from Theorem 5.29.

Consider the kernel  $P_{D_{-i}|D_i}^*$  described by

$$p_{D_{-i}|D_i=d_i}^*(d_{-i}) := \begin{cases} 1 - \epsilon, & \text{for some } d_{-i} \in \arg \max_{\bar{d}_{-i}} p_{Y|D_{-i}=\bar{d}_{-i}, D_i=d_i}(y), \\ \frac{\epsilon}{|\mathcal{D}|^{n-1}-1}, & \text{otherwise,} \end{cases}$$

and

$$p_{D_{-i}|D_i=d'_i}^*(d_{-i}) := \begin{cases} 1 - \varepsilon, & \text{for some } d_{-i} \in \arg \min_{\tilde{d}_{-i}} p_{Y|D_{-i}=\tilde{d}_{-i}, D_i=d'_i}(y), \\ \frac{\varepsilon}{|\mathcal{D}|^n - 1}, & \text{otherwise.} \end{cases}$$

Then, we get

$$\begin{aligned} \sup_{P_{D_{-i}|D_i}} \sup_{P_{D_i}} \exp(\ell(D_i \rightarrow y)) &\geq \max_{d_i, d'_i} \frac{\sum_{d_{-i}} p_{Y|D_i=d_i, D_{-i}=d_{-i}}(y) p_{D_{-i}|D_i=d_i}^*(d_{-i})}{\sum_{d'_{-i}} p_{Y|D_i=d'_i, D_{-i}=d'_{-i}}(y) p_{D_{-i}|D_i=d'_i}^*(d'_{-i})} \\ &= \max_{d_i, d'_i} \frac{\max_{d_{-i}} p_{Y|D_i=d_i, D_{-i}=d_{-i}}(y)}{\min_{d'_{-i}} p_{Y|D_i=d'_i, D_{-i}=d'_{-i}}(y)} \quad (5.20a) \\ &= \max_{d_i, d'_i} \max_{d_{-i}, d'_{-i}} \frac{p_{Y|D_i=d_i, D_{-i}=d_{-i}}(y)}{p_{Y|D_i=d'_i, D_{-i}=d'_{-i}}(y)}, \end{aligned}$$

where (5.20a) follows by letting  $\varepsilon \rightarrow 0$ .  $\square$

We highlight a few points about the above results. First, note that by the Markov chain  $D_i - X - Y$  and the pre-processing inequality for PML we have  $\ell(D_i \rightarrow y) \leq \ell(X \rightarrow y)$  for all  $i \in [n]$ ,  $y \in \mathcal{Y}$  and  $P_X \in \mathcal{P}_X$ . Theorem 5.41 then implies that under certain distributions, the amount of information leaking about a single entry can be as large as the information leaking about the whole database. Roughly speaking, this happens when the entropy of the whole dataset is concentrated on a single entry. Second, by comparing (iii) in Theorem 5.41 and (i) in Theorem 5.38 we arrive at a similar conclusion to [75] and [149] that the informed adversary assumption may lead to underestimating the information leaking about the entries in the dataset. Nevertheless, this can happen only when the entries in the database are highly correlated. Indeed, if we restrict our attention to product distributions, then by (ii) and (iii) in Theorem 5.38 the conditional and unconditional leakages become equal. Third, in neither of the above results the supremum is ever actually attained by any distribution in  $\mathcal{P}_X$  or  $\mathcal{Q}_X$  (see Remark 5.39). Instead, the proofs construct a sequence of distributions with decreasing (conditional) entropy under which PML converges to the corresponding log-likelihood ratio in the definition of differential privacy or free-lunch privacy. In fact, when the dataset has large entropy, the amount of information leaking through a privacy mechanism can be noticeably smaller than the  $\epsilon$  reported by differential privacy or free-lunch privacy.

## Laplace Mechanism and the Counting Query

So far, we have encountered a recurring pattern: Low-entropy distributions pose greater challenges for privatization, yielding larger leakage. This trend was observed in Theorems 5.29, 5.38, 5.41, and Examples 5.6, 5.10, 5.31. For this reason,

we are motivated to include existing assumptions about the prior distribution in our analysis, aiming to derive more precise bounds on the privacy parameter associated with a given mechanism. It is important to emphasize that certain datasets, such as financial data for fraud detection or health data for studying rare diseases, inherently exhibit features with very small entropy. However, in many everyday applications, we encounter high-entropy datasets with more balanced probabilities. In these cases, we can save on the privacy cost paid, and ultimately, achieve more utility. Below, we illustrate this for the archetypical example of a counting query that is answered by the Laplace mechanism.

We consider the third characterization of differential privacy in Theorem 5.38 and restrict the set of product distributions from which  $X$  may be drawn. Suppose  $X$  is an i.i.d database containing  $n$  entries. Consider a predicate  $f : \mathcal{D} \rightarrow \{0, 1\}$  and suppose we want to answer the counting query “What fraction of the entries in the database satisfy  $f(d_i) = 1$ ?”. Let  $0 \leq c < \frac{1}{2}$  be a constant and assume  $P_X \in \mathcal{P}_c^f$ , where

$$\mathcal{P}_c^f = \left\{ P_X \in \mathcal{Q}_X : P_{D_i}(\{d \in \mathcal{D} : f(d) = 1\}) = p \text{ for all } i \in [n] \text{ and } p \in (c, 1 - c) \right\}.$$

That is, we assume that each entry in the database satisfies the predicate  $f$  with probability  $p \in (c, 1 - c)$ . Let  $\text{Lap}(\mu, b)$  denote the Laplace distribution with mean  $\mu \in \mathbb{R}$  and scale parameter  $b > 0$ . To answer the counting query, the Laplace mechanism returns an outcome according to the distribution  $Y \mid X = (d_1, \dots, d_n) \sim \text{Lap}\left(\frac{f(d_1) + \dots + f(d_n)}{n}, b\right)$ .

**Proposition 5.42.** *Consider the predicate  $f : \mathcal{D} \rightarrow \{0, 1\}$ . Suppose  $X$  is a database of size  $n$  drawn according to a distribution  $P_X \in \mathcal{P}_c^f$ . Let  $P_{Y \mid X}$  denote the Laplace mechanism with scale parameter  $b > 0$  answering the counting query corresponding to  $f$ . Then, the information leaking about each entry in the database is upper bounded by*

$$\sup_{P_X \in \mathcal{P}_c^f} \sup_{y \in \mathbb{R}} \ell(D_i \rightarrow y) \leq \frac{1}{nb} - \log \left( (1 - c) + c \exp \left( \frac{1}{nb} \right) \right),$$

for all  $i \in [n]$ .

When  $nb$  is large we may use  $e^x \geq 1 + x$  and  $\log(1 + x) \geq x - \frac{x^2}{2}$  for  $x \geq 0$  to obtain the simplified bound

$$\sup_{P_X \in \mathcal{P}_c^f} \sup_{y \in \mathbb{R}} \ell(D_i \rightarrow y) \leq \frac{1 - c}{nb} + \frac{c^2}{2n^2b^2},$$

for all  $i \in [n]$ . Observe that  $\frac{1}{nb}$  corresponds to the well-known differential privacy parameter of the Laplace mechanism returning the answer to a query with global

sensitivity  $\frac{1}{n}$  (see Definition 3.4). As expected, the above leakage bound also reduces to  $\frac{1}{nb}$  when  $c = 0$ , describing the situation where  $P_X$  can be any i.i.d distribution in  $\mathcal{Q}_X$  with arbitrarily small entropy. On the other hand, when  $c$  is close to  $\frac{1}{2}$ , then the privacy parameter is reduced by almost a factor of  $\frac{1}{2}$ . We conclude that information leakage analysis using PML allows us to adjust the privacy cost to the entropy of the data, and ultimately, achieve higher utility.



---

# Appendices

---

## 5.A Proof of Lemma 5.2

(i) Upper bound:

$$\begin{aligned}
 \ell(X \rightarrow y) &= \log \max_{x \in \text{supp}(P_X)} \frac{P_{X|Y=y}(x)}{P_X(x)} \\
 &\leq \log \max_{x \in \text{supp}(P_X)} \frac{1}{P_X(x)} \\
 &= \log \frac{1}{\min_{x \in \text{supp}(P_X)} P_X(x)},
 \end{aligned}$$

where the inequality holds with equality if and only if  $P_{X|Y=y}(x^*) = 1$  with  $x^* \in \arg \min_{x \in \text{supp}(P_X)} P_X(x)$ .

Lower bound: Here, the idea is that since both  $P_{X|Y=y}$  and  $P_X$  are probability distributions over  $\text{supp}(P_X)$ , then for any fixed  $y \in \text{supp}(P_Y)$ , there exists at least one  $x \in \text{supp}(P_X)$  such that  $P_{X|Y=y}(x) \geq P_X(x)$ . Suppose to the contrary that for all  $x \in \mathcal{X}$ ,  $P_{X|Y=y}(x) < P_X(x)$ . Then,  $1 = \sum_x P_{X|Y=y}(x) < \sum_x P_X(x) = 1$  which is a contradiction. Therefore,

$$\ell(X \rightarrow y) = \log \max_{x \in \text{supp}(P_X)} \frac{P_{X|Y=y}(x)}{P_X(x)} \geq \log 1 = 0.$$

The above inequality holds with equality if and only if  $\max_x P_{Y|X=x}(y) = P_Y(y) = \sum_x P_{Y|X=x}(y)P_X(x)$  which holds whenever  $P_{Y|X=x}(y) = P_{Y|X=x'}(y)$  for all  $x, x' \in \text{supp}P_X$ .

(ii) Both statements follow directly from the definition.

(iii) Let  $x^* \in \arg \max_x P_{Z|X=x}(z)$ . Then,

$$\begin{aligned}
 \ell(X \rightarrow z) &= \log \max_x \frac{P_{Z|X=x}(z)}{P_Z(z)} \\
 &= \log \frac{\sum_{y \in \text{supp}(P_{Y|X=x^*})} P_{Z|Y=y}(z)P_{Y|X=x^*}(y)}{P_Z(z)} \\
 &\leq \log \max_{y' \in \text{supp}(P_{Y|X=x^*})} \frac{P_{Z|Y=y'}(z)}{P_Z(z)} \sum_y P_{Y|X=x^*}(y)
 \end{aligned}$$

$$\leq \log \max_{y' \in \text{supp}(P_Y)} \frac{P_{Z|Y=y'}(z)}{P_Z(z)} = \ell(Y \rightarrow z),$$

where the first inequality holds with equality if  $P_{Z|Y=y}(z) = P_{Z|Y=y'}(z)$  for all  $y, y' \in \text{supp}(P_{Y|X=x^*})$ , and the second inequality holds with equality if  $\max_{y \in \text{supp}(P_Y)} i(y; z)$  is attained at some  $y \in \text{supp}(P_{Y|X=x^*})$ .

(iv)

$$\begin{aligned} \ell(X \rightarrow z) &= \log \max_x \frac{P_{Z|X=x}(z)}{P_Z(z)} \\ &= \log \max_x \frac{\sum_{y \in \text{supp}(P_Y)} P_{Z|Y=y}(z) P_{Y|X=x}(y)}{\sum_{y \in \text{supp}(P_Y)} P_{Z|Y=y}(z) P_Y(y)} \\ &\leq \log \max_x \max_{y \in \text{supp}(P_Y)} \frac{P_{Y|X=x}(y)}{P_Y(y)} \\ &= \max_{y \in \text{supp}(P_Y)} \ell(X \rightarrow y). \end{aligned}$$

Now, if  $X$  and  $Y$  are independent then  $\ell(X \rightarrow z) = \ell(X \rightarrow y) = 0$  for all  $y, z$ , and the inequality holds with equality. Furthermore, if the distribution  $P_{Y|Z=z}$  is degenerate, then  $z$  is mapped uniquely to some  $y_z \in \text{supp}(P_Y)$ . This implies that  $P_{Z|Y=y}(z) = 0$  for  $y \neq y_z$ , hence, we have

$$\begin{aligned} \ell(X \rightarrow z) &= \log \max_x \frac{P_{Z|X=x}(z)}{P_Z(z)} \\ &= \log \max_x \frac{P_{Y|X=x}(y_z)}{P_Y(y_z)} \\ &= \ell(X \rightarrow y_z) \\ &\leq \max_{y \in \text{supp}(P_Y)} \ell(X \rightarrow y), \end{aligned}$$

with equality if  $\ell(X \rightarrow y_z) = \max_{y \in \text{supp}(P_Y)} \ell(X \rightarrow y)$ .

(v)

$$\begin{aligned} \ell(X \rightarrow y | z) &= \log \max_x \frac{P_{Y|X=x, Z=z}(y)}{P_{Y|Z=z}(y)} \\ &= \log \max_x \frac{P_{Y|X=x}(y) P_Y(y)}{P_Y(y) P_{Y|Z=z}(y)} \\ &= \log \max_x \frac{P_{Y|X=x}(y)}{P_Y(y)} + \log \frac{P_Y(y)}{P_{Y|Z=z}(y)} \\ &= \ell(X \rightarrow y) - i(y; z). \end{aligned}$$

(vi)

$$\begin{aligned}
 \ell(X \rightarrow y, z) &= \log \max_x \frac{P_{YZ|X=x}(y, z)}{P_{YZ}(y, z)} \\
 &= \log \max_x \frac{P_{Y|X=x, Z=z}(y) P_{Z|X=x}(z)}{P_{Y|Z=z}(y) P_Z(z)} \\
 &\leq \log \max_x \frac{P_{Y|X=x, Z=z}(y)}{P_{Y|Z=z}(y)} + \log \max_x \frac{P_{Z|X=x}(z)}{P_Z(z)} \\
 &= \ell(X \rightarrow y | z) + \ell(X \rightarrow z),
 \end{aligned}$$

with equality if there exists  $x^* \in \text{supp}(P_X)$  maximizing both  $i(x; y | z)$  and  $i(x; z)$ .

## 5.B Proof of Lemma 5.5

(i) For all  $P_{Y|X}$  and all  $y \in \text{supp}(P_Y)$  we have

$$\log \max_{x \in \text{supp}(P_X)} \frac{P_{X|Y=y}(x)}{P_X(x)} \leq \log \max_{x \in \text{supp}(P_X)} \frac{1}{P_X(x)} = \log \frac{1}{\min_{x \in \text{supp}(P_X)} P_X(x)}.$$

Note that  $\min_{x \in \text{supp}(P_X)} P_X(x) \leq \frac{1}{2}$  which implies that  $\epsilon_{\max} \geq \log 2$ . The second half of the statement is clear from the above inequality.

(ii) If  $X$  and  $Y$  are independent, then  $P_{X|Y=y}(x) = P_X(x)$  for all  $x, y$ , thus the mechanism  $P_{Y|X}$  satisfies  $\epsilon$ -PML with  $\epsilon = 0$ . Conversely, if  $P_{Y|X}$  satisfies  $\epsilon$ -PML with  $\epsilon = 0$  this implies that  $P_{X|Y=y}(x) = P_X(x)$  for all  $x, y$  which means that  $X$  and  $Y$  are independent.

## 5.C Proof of Proposition 5.9

(i) By the pre-processing property of Lemma 5.2, if  $P_{Z|Y}$  satisfies  $\epsilon$ -PML then for all  $z \in \mathcal{Z}$  we have  $\ell(Y \rightarrow z) \leq \epsilon$ . Hence,

$$\max_{z \in \mathcal{Z}} \ell(X \rightarrow z) \leq \max_{z \in \mathcal{Z}} \ell(Y \rightarrow z) \leq \epsilon,$$

and  $P_{Z|X}$  satisfies  $\epsilon$ -PML.

(ii) By the post-processing property of Lemma 5.2,

$$\max_{z \in \mathcal{Z}} \ell(X \rightarrow z) \leq \max_{y \in \mathcal{Y}} \ell(X \rightarrow y) \leq \epsilon,$$

so  $P_{Z|X}$  satisfies  $\epsilon$ -PML.

(iii) Similarly to the above, the pre-processing property of Lemma 5.2 yields

$$\mathbb{P}_{Z \sim P_Z} [\ell(X \rightarrow Z) > \epsilon] \leq \mathbb{P}_{Z \sim P_Z} [\ell(Y \rightarrow Z) > \epsilon] \leq \delta,$$

hence,  $P_{Z|X}$  satisfies  $(\epsilon, \delta)$ -PML.

### 5.D Proof of Lemma 5.16

Let the function  $f : \mathcal{C}(P_{Y|X}) \times [0, 1] \rightarrow \mathbb{R}^+$  be defined as

$$f(P_{\bar{Y}|X}, \delta) = \sup_{P_{Z|\bar{Y}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \ell_{P_{XZ}}(X \rightarrow z),$$

that is,  $f$  represents the largest PML over all outcomes  $z$  of all post-processing channels with probability at least  $\delta$ . We argue that  $f(\cdot, \delta)$  is constant on  $\mathcal{C}(P_{Y|X})$  for all  $0 \leq \delta \leq 1$ . To see this, fix an arbitrary  $P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})$  and note that the Markov chain  $X - \bar{Y} - Y_r$  holds, where  $Y_r$  denotes the random variable induced by the reduced channel  $P_{Y_r|X}$ . Now, we write

$$\begin{aligned} f(P_{\bar{Y}|X}, \delta) &= \sup_{P_{Z|\bar{Y}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \ell_{P_{XZ}}(X \rightarrow z) \\ &= \sup_{\substack{P_{Z|Y_r}: \\ P_{Z|Y_r} = P_{Z|\bar{Y}} \circ P_{\bar{Y}|Y_r}}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \ell_{P_{XZ}}(X \rightarrow z) \\ &\leq \sup_{P_{Z|Y_r}} \max_{\substack{z \in \text{supp}(P_Z): \\ P_Z(z) \geq \delta}} \ell_{P_{XZ}}(X \rightarrow z) \\ &= f(P_{Y_r|X}, \delta). \end{aligned}$$

By definition,  $I(X; \bar{Y}) = I(X; Y_r)$ , therefore  $Y_r$  is a sufficient statistic of  $\bar{Y}$  for  $X$ , and the Markov chain  $X - Y_r - \bar{Y}$  also holds. Then, reversing the role of  $\bar{Y}$  and  $Y_r$ , it can also be established that  $f(P_{Y_r|X}, \delta) \leq f(P_{\bar{Y}|X}, \delta)$ . Thus, we obtain  $f(P_{Y_r|X}, \delta) = f(P_{\bar{Y}|X}, \delta)$  for all  $P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})$  and  $0 \leq \delta \leq 1$ . Finally, if  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -closedness then  $f(P_{Y|X}, \delta) \leq \epsilon$ , which implies that  $\sup_{P_{\bar{Y}|X} \in \mathcal{C}(P_{Y|X})} f(P_{\bar{Y}|X}, \delta) \leq \epsilon$ .

### 5.E Proof of Proposition 5.21

In both cases, we use Theorem 5.19 and verify the conditions of Definition 5.13. Consider the Markov chain  $X - Y - Z - T$ .

- (i) Fix an arbitrary  $P_{T|Z}$  and  $t \in \text{supp}(P_T)$  satisfying  $P_T(t) \geq \delta$ . Then,

$$\ell(X \rightarrow t) \leq \ell(Y \rightarrow t) \leq \epsilon,$$

where the first inequality is due to Lemma 5.2 and the second inequality follows by the assumption that  $P_{Z|Y}$  satisfies  $(\epsilon, \delta)$ -EML. Thus,  $P_{Z|X}$  satisfies  $(\epsilon, \delta)$ -EML.

- (ii) The result follows directly by noticing that  $T$  is a post-processing of  $Y$  through the channel  $P_{T|Y} = P_{T|Z} \circ P_{Z|Y}$ . Hence,  $P_T(t) \geq \delta$  implies  $\ell(X \rightarrow t) \leq \epsilon$  with  $t \in \text{supp}(P_T)$  and  $P_{Z|X}$  satisfies  $(\epsilon, \delta)$ -EML.

## 5.F Proof of Proposition 5.23

We need to show that  $P_Y(\mathcal{A}) \leq \delta$ . We can write

$$\begin{aligned} \ell(X \rightarrow \mathcal{A}) &= \log \max_{x \in \text{supp}(P_X)} \frac{\sum_{y \in \mathcal{A}} P_{Y|X=x}(y)}{\sum_{y \in \mathcal{A}} P_Y(y)} \\ &= \log \frac{\sum_{y \in \mathcal{A}} \max_{x \in \text{supp}(P_X)} P_{Y|X=x}(y)}{\sum_{y \in \mathcal{A}} P_Y(y)} \end{aligned} \quad (5.21a)$$

$$\begin{aligned} &\geq \log \min_{y \in \mathcal{A}} \max_{x \in \text{supp}(P_X)} \frac{P_{Y|X=x}(y)}{P_Y(y)} \\ &= \log \min_{y \in \mathcal{A}} \ell(X \rightarrow y) \\ &> \epsilon, \end{aligned} \quad (5.21b)$$

where (5.21a) follows from the assumption that all  $P_{Y|X=x}(y)$  are maximized at the same  $x$ , and (5.21b) follows from the definition of the event  $\mathcal{A}$ . Since  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -EML,  $\ell(X \rightarrow \mathcal{A}) > \epsilon$  implies that  $P_Y(\mathcal{A}) < \delta$ .

## 5.G Proof of Theorem 5.24

- (i) This result is an immediate consequence of the composition property given in Lemma 5.2. For all  $(y, z) \in \mathcal{Y} \times \mathcal{Z}$  we have

$$\begin{aligned} \ell(X \rightarrow y, z) &\leq \ell(X \rightarrow y) + \ell(X \rightarrow z | y) \\ &\leq \max_{y \in \mathcal{Y}} \ell(X \rightarrow y) + \max_{(y, z) \in \mathcal{Y} \times \mathcal{Z}} \ell(X \rightarrow z | y) \\ &\leq \epsilon_1 + \epsilon_2. \end{aligned}$$

Therefore,  $P_{YZ|X}$  satisfies  $\epsilon_1 + \epsilon_2$ -PML.

- (ii) Since  $\ell(X \rightarrow y, z) \leq \ell(X \rightarrow y) + \ell(X \rightarrow z | y)$  for all  $(y, z) \in \mathcal{Y} \times \mathcal{Z}$ , we can write

$$\begin{aligned} \mathbb{P}[\ell(X \rightarrow Y, Z) > \epsilon_1 + \epsilon_2] &\leq \mathbb{P}[\ell(X \rightarrow Y) + \ell(X \rightarrow Z | Y) > \epsilon_1 + \epsilon_2] \\ &= 1 - \mathbb{P}[\ell(X \rightarrow Y) + \ell(X \rightarrow Z | Y) \leq \epsilon_1 + \epsilon_2]. \end{aligned}$$

We define the following “good” events:

$$\begin{aligned} \mathcal{G} &:= \{(y, z) \in \mathcal{Y} \times \mathcal{Z}: \ell(X \rightarrow y) \leq \epsilon_1 \text{ and } \ell(X \rightarrow z | y) \leq \epsilon_2\}, \\ \mathcal{G}_Y &:= \{y \in \mathcal{Y}: (y, z) \in \mathcal{G} \text{ for some } z \in \mathcal{Z}\}, \\ \mathcal{G}_Z(y) &:= \{z \in \mathcal{Z}: (y, z) \in \mathcal{G}\}. \end{aligned}$$

Our goal is to lower bound the probability of event  $\mathcal{G}$ . We can write

$$\begin{aligned}
 P_{YZ}(\mathcal{G}) &= \sum_{(y,z) \in \mathcal{G}} P_Y(y) P_{Z|Y=y}(z) \\
 &= \sum_{y \in \mathcal{G}_Y} P_Y(y) P_{Z|Y=y}(\mathcal{G}_Z(y)) \\
 &\geq (1 - \delta_2) \sum_{y \in \mathcal{G}_Y} P_Y(y) \tag{5.22a} \\
 &\geq (1 - \delta_2)(1 - \delta_1), \tag{5.22b}
 \end{aligned}$$

where

- (5.22a) follows from the fact that for all  $y \in \mathcal{Y}$ ,  $P_{Z|X,Y=y}$  satisfies  $(\epsilon_2, \delta_2)$ -PML which implies that

$$\begin{aligned}
 P_{Z|Y=y}(\mathcal{G}_Z(y)) &= \mathbb{P}_{Z \sim P_{Z|Y=y}} \left[ \ell(X \rightarrow Z | y) \leq \epsilon_2 \right] \\
 &\geq 1 - \delta_2,
 \end{aligned}$$

- and (5.22b) follows from the fact that  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta_1)$ -PML, that is,

$$P_Y(\mathcal{G}_Y) = \mathbb{P}_{Y \sim P_Y} \left[ \ell(X \rightarrow Y) \leq \epsilon_1 \right] \geq 1 - \delta_1.$$

It follows that

$$\begin{aligned}
 \mathbb{P}_{(Y,Z) \sim P_{YZ}} \left[ \ell(X \rightarrow y) + \ell(X \rightarrow z | y) \leq \epsilon_1 + \epsilon_2 \right] &\geq P_{YZ}(\mathcal{G}) \\
 &\geq (1 - \delta_2)(1 - \delta_1),
 \end{aligned}$$

which yields

$$\mathbb{P}_{(Y,Z) \sim P_{YZ}} \left[ \ell(X \rightarrow Y, Z) > \epsilon_1 + \epsilon_2 \right] \leq \delta_1 + \delta_2 - \delta_1 \delta_2.$$

(iii) Define the event

$$\mathcal{A}_Y = \{y \in \mathcal{Y} : \ell(X \rightarrow y) \leq \epsilon_1\}.$$

As  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta_1)$ -PML, we have

$$1 - \delta_1 \leq P_Y(\mathcal{A}_Y) = P_{YZ}(\mathcal{A}),$$

where  $\mathcal{A} := \mathcal{A}_Y \times \mathcal{Z}$ . Moreover, we define the event

$$\mathcal{B} := \{(y, z) \in \mathcal{Y} \times \mathcal{Z} : \ell(X \rightarrow z | y) \leq \epsilon_2\}.$$

By assumption,  $P_{YZ}(\mathcal{B}) \geq 1 - \delta_2$ . Therefore,

$$\begin{aligned} & \mathbb{P}_{(Y,Z) \sim P_{YZ}} \left[ \ell(X \rightarrow Y, Z) \leq \epsilon_1 + \epsilon_2 \right] \\ & \geq \mathbb{P}_{(Y,Z) \sim P_{YZ}} \left[ \ell(X \rightarrow Y) + \ell(X \rightarrow Z | Y) \leq \epsilon_1 + \epsilon_2 \right] \\ & \geq P_{YZ}(\mathcal{A} \cap \mathcal{B}) \\ & = 1 - P_{YZ}(\mathcal{A}^c \cup \mathcal{B}^c) \\ & \geq 1 - \delta_1 - \delta_2. \end{aligned}$$

(iv) Let  $\mathcal{E} \subseteq \mathcal{Y} \times \mathcal{Z}$  be an event satisfying  $P_{YZ}(\mathcal{E}) \geq \delta_1$  and

$$0 \leq \delta_2 \leq \min_{y \in \mathcal{E}_Y} P_{Z|Y=y}(\mathcal{E}_Z(y)).$$

Since  $P_{Z|X,Y=y}$  satisfies  $(\epsilon_2, \delta_2)$ -EML for all  $y \in \mathcal{E}_Y$ , we have

$$\max_{x \in \text{supp}(P_X)} \frac{P_{Z|Y=y, X=x}(\mathcal{E}_Z(y))}{P_{Z|Y=y}(\mathcal{E}_Z(y))} \leq \exp(\epsilon_2). \quad (5.23)$$

Now, we write

$$\begin{aligned} \exp(\ell(X \rightarrow \mathcal{E})) &= \max_{x \in \text{supp}(P_X)} \frac{P_{YZ|X=x}(\mathcal{E})}{P_{YZ}(\mathcal{E})} \\ &= \max_{x \in \text{supp}(P_X)} \frac{\sum_{y \in \mathcal{E}_Y} \sum_{z \in \mathcal{E}_Z(y)} P_{YZ|X=x}(y, z)}{\sum_{y \in \mathcal{E}_Y} \sum_{z \in \mathcal{E}_Z(y)} P_{YZ}(y, z)} \\ &= \max_{x \in \text{supp}(P_X)} \frac{\sum_{y \in \mathcal{E}_Y} P_{Y|X=x}(y) \sum_{z \in \mathcal{E}_Z(y)} P_{Z|Y=y, X=x}(z)}{\sum_{y \in \mathcal{E}_Y} P_Y(y) \sum_{z \in \mathcal{E}_Z(y)} P_{Z|Y=y}(z)} \\ &= \max_{x \in \text{supp}(P_X)} \sum_{y \in \mathcal{E}_Y} \frac{P_Y(y) P_{Z|Y=y}(\mathcal{E}_Z(y))}{\sum_{y' \in \mathcal{E}_Y} P_Y(y') P_{Z|Y=y'}(\mathcal{E}_Z(y'))} \cdot \left( \frac{P_{Y|X=x}(y)}{P_Y(y)} \right). \end{aligned}$$

$$\leq e^{\epsilon_2} \cdot \max_{x \in \text{supp}(P_X)} \sum_{y \in \mathcal{E}_Y} \frac{P_Y(y) P_{Z|Y=y}(\mathcal{E}_Z(y))}{\sum_{y' \in \mathcal{E}_Y} P_Y(y') P_{Z|Y=y'}(\mathcal{E}_Z(y'))} \cdot \left( \frac{P_{Z|Y=y, X=x}(\mathcal{E}_Z(y))}{P_{Z|Y=y}(\mathcal{E}_Z(y))} \right) \quad (5.24a)$$

$$\leq e^{\epsilon_2} \cdot \max_{x \in \text{supp}(P_X)} h_x(P_{Y|X=x}, \delta_1) \quad (5.24b)$$

$$\leq e^{\epsilon_2 + \epsilon_1}, \quad (5.24c)$$

where

- (5.24a) follows from inequality (5.23),

- the function  $h_x$  in (5.24b) is defined in (5.5),
  - and (5.24c) follows from the fact that  $P_{Y|X}$  satisfies  $(\epsilon_1, \delta_1)$ -EML.
- (v) Let  $\mathcal{E} \subseteq \mathcal{Y} \times \mathcal{Z}$  be an event satisfying  $P_{YZ}(\mathcal{E}) \geq \delta_1 + \delta_2$ . We define the following “bad” sets

$$\begin{aligned}\mathcal{B}_Y &:= \{y \in \mathcal{E}_Y : P_Z(\mathcal{E}_Z(y)) < \delta_2\}, \\ \mathcal{B} &:= \{(y, z) \in \mathcal{E} : y \in \mathcal{B}_Y\},\end{aligned}$$

and the “good” sets  $\mathcal{G}_Y = \mathcal{E}_Y \setminus \mathcal{B}_Y$  and  $\mathcal{G} = \mathcal{E} \setminus \mathcal{B}$ . Note that

$$P_{YZ}(\mathcal{B}) = \sum_{y \in \mathcal{B}_Y} P_Y(y) P_{Z|Y=y}(\mathcal{E}_Z(y)) < \delta_2,$$

which implies that  $P_{YZ}(\mathcal{G}) = P_{YZ}(\mathcal{E}) - P_{YZ}(\mathcal{B}) > \delta_1$ . Now, similarly to the previous part, we write

$$\begin{aligned}\exp(\ell(X \rightarrow \mathcal{E})) &= \max_{x \in \text{supp}(P_X)} \frac{\sum_{(y,z) \in \mathcal{E}} P_{YZ|X=x}(y, z)}{P_{YZ}(\mathcal{E})} \\ &\leq \max_{x \in \text{supp}(P_X)} \sum_{(y,z) \in \mathcal{B}} \frac{P_{YZ}(y, z)}{P_{YZ}(\mathcal{E})} \cdot \left( \frac{P_{YZ|X=x}(y, z)}{P_{YZ}(y, z)} \right) \\ &\quad + \max_{x \in \text{supp}(P_X)} \frac{\sum_{(y,z) \in \mathcal{G}} P_{YZ|X=x}(y, z)}{P_{YZ}(\mathcal{E})} \\ &\leq \frac{\delta_2}{\delta_1 + \delta_2} e^{\epsilon_{\max}} + \max_{x \in \text{supp}(P_X)} \frac{\sum_{(y,z) \in \mathcal{G}} P_{YZ|X=x}(y, z)}{P_{YZ}(\mathcal{G})} \tag{5.25a}\end{aligned}$$

$$\leq \frac{\delta_2}{\delta_1 + \delta_2} e^{\epsilon_{\max}} + e^{\epsilon_1 + \epsilon_2}, \tag{5.25b}$$

where

- (5.25a) follows from the fact that  $P_{YZ}(\mathcal{B}) < \delta_2$ ,  $P_{YZ}(\mathcal{E}) \geq \delta_1 + \delta_2$ , and for all  $(y, z) \in \text{supp}(P_{YZ})$ ,

$$\max_{x \in \text{supp}(P_X)} \frac{P_{YZ|X=x}(y, z)}{P_{YZ}(y, z)} \leq \exp(\epsilon_{\max}),$$

- and (5.25b) follows from the definition of the set  $\mathcal{G}$  and the previous part.

## 5.H Proof of Proposition 5.28

$$\mathbb{P}_{(X,Y) \sim P_{XY}} \left[ \frac{P_{XY}(X, Y)}{P_X(X)P_Y(Y)} \leq e^\epsilon \right] \geq \mathbb{P}_{Y \sim P_Y} \left[ \max_{x \in \text{supp}(P_X)} \frac{P_{XY}(x, Y)}{P_X(x)P_Y(Y)} \leq e^\epsilon \right]$$

$$\begin{aligned} &= \mathbb{P}_{Y \sim P_Y} [\ell(X \rightarrow Y) \leq \epsilon] \\ &\geq 1 - \delta, \end{aligned}$$

or equivalently,

$$\mathbb{P}_{(X,Y) \sim P_{XY}} \left[ \frac{P_{XY}(X,Y)}{P_X(X)P_Y(Y)} > e^\epsilon \right] \leq \delta.$$

Using the above inequality and Lemma 5.27 we obtain

$$I_\infty^\delta(X; Y) = D_\infty^\delta(P_{XY} \| P_X \times P_Y) \leq \epsilon.$$

## 5.I Proof of Proposition 5.30

Suppose  $P_{Y|X}$  satisfies  $\epsilon$ -LDP. Fix  $y \in \mathcal{Y}$ . We have

$$\begin{aligned} \max_{x \in \text{supp}(P_X)} \frac{P_{Y|X=x}(y)}{P_Y(y)} &= \max_{x \in \text{supp}(P_X)} \frac{P_{Y|X=x}(y)}{\sum_{x' \in \mathcal{X}} P_{Y|X=x'}(y) P_X(x')} \\ &= \max_{x \in \text{supp}(P_X)} \frac{P_{Y|X=x}(y)}{P_{Y|X=x}(y) P_X(x) + \sum_{x' \neq x} P_{Y|X=x'}(y) P_X(x')} \\ &\leq \max_{x \in \text{supp}(P_X)} \frac{P_{Y|X=x}(y)}{P_{Y|X=x}(y) P_X(x) + \sum_{x' \neq x} P_{Y|X=x'}(y) P_X(x') e^{-\epsilon}} \\ &= \max_{x \in \text{supp}(P_X)} \frac{1}{P_X(x) + (1 - P_X(x)) e^{-\epsilon}} \\ &\leq \frac{1}{p_{\min} + (1 - p_{\min}) e^{-\epsilon}}. \end{aligned}$$

## 5.J Proof of Proposition 5.32

Note that  $P_{Y|X}$  satisfies  $\epsilon$ -LDI if

$$\frac{P_{Y|X=x}(y) P_X(x)}{P_{Y|X=x'}(y) P_X(x')} \leq e^\epsilon,$$

for all  $y \in \mathcal{Y}$  and all  $x, x' \in \text{supp}(P_X)$ . Fix  $y \in \mathcal{Y}$ . We write

$$\begin{aligned} \max_{x \in \text{supp}(P_X)} \frac{P_{Y|X=x}(y)}{P_Y(y)} &= \max_{x \in \text{supp}(P_X)} \frac{P_{Y|X=x}(y)}{\sum_{x'} P_{Y|X=x'}(y) P_X(x')} \\ &= \max_{x \in \text{supp}(P_X)} \frac{P_{Y|X=x}(y)}{P_{Y|X=x}(y) P_X(x) + \sum_{x' \neq x} P_{Y|X=x'}(y) P_X(x')} \\ &\leq \max_{x \in \text{supp}(P_X)} \frac{P_{Y|X=x}(y)}{P_{Y|X=x}(y) P_X(x) + \sum_{x' \neq x} P_{Y|X=x'}(y) P_X(x') e^{-\epsilon}} \end{aligned}$$

$$\begin{aligned}
 &= \max_{x \in \text{supp}(P_X)} \frac{1}{P_X(x)(1 + e^{-\epsilon}(|\text{supp}(P_X)| - 1))} \\
 &= \frac{1}{p_{\min}(1 + e^{-\epsilon}(|\text{supp}(P_X)| - 1))}.
 \end{aligned}$$

## 5.K Proof of Proposition 5.35

$$\begin{aligned}
 I(X; Y) &= \mathbb{E}_{y \sim P_Y} \left[ \mathbb{E}_{X \sim P_{X|Y}(\cdot|Y)} \left[ \log \frac{P_{XY}(X, Y)}{P_X(X)P_Y(Y)} \right] \right] \\
 &\leq \mathbb{E}_{Y \sim P_Y} \left[ \max_{x \in \text{supp}(P_X)} \log \frac{P_{XY}(x, Y)}{P_X(x)P_Y(Y)} \right] \\
 &= \mathbb{E}_{Y \sim P_Y} [\ell(X \rightarrow Y)],
 \end{aligned}$$

where the inequality holds with equality if and only if for all  $x, x', y$  such that  $P_{XY}(x, y) > 0$  and  $P_{XY}(x', y) > 0$  we have  $i(x; y) = i(x'; y)$ , or equivalently,  $P_{Y|X=x}(y) = P_{Y|X=x'}(y)$  (the condition for equality has also been noted in [63, Lemma 2]).

## 5.L Proof of Proposition 5.34

$$\begin{aligned}
 I_f(X; Y) &= \mathbb{E}_{Y \sim P_Y} \left[ \mathbb{E}_{X \sim P_X} \left[ f \left( \frac{P_{XY}(X, Y)}{P_X(X)P_Y(Y)} \right) \right] \right] \\
 &\leq \mathbb{E}_{Y \sim P_Y} \left[ \max_{x \in \text{supp}(P_X)} f \left( \frac{P_{XY}(x, Y)}{P_X(x)P_Y(Y)} \right) \right] \\
 &= \mathbb{E} \left[ \max \left\{ f \left( \max_{x \in \text{supp}(P_X)} \frac{P_{XY}(x, Y)}{P_X(x)P_Y(Y)} \right), f \left( \min_{x \in \text{supp}(P_X)} \frac{P_{XY}(x, Y)}{P_X(x)P_Y(Y)} \right) \right\} \right] \tag{5.26a}
 \end{aligned}$$

$$\begin{aligned}
 &\leq \mathbb{E} \left[ \max \left\{ f \left( \max_{x \in \text{supp}(P_X)} \frac{P_{XY}(x, Y)}{P_X(x)P_Y(Y)} \right), f(0) \right\} \right] \tag{5.26b} \\
 &= \mathbb{E} \left[ \max \left\{ f \left( \exp(\ell(X \rightarrow Y)) \right), f(0) \right\} \right],
 \end{aligned}$$

where (5.26a) follows from the fact that the maximum of a convex function is attained at an extreme point, and (5.26b) follows from  $\min_{x \in \text{supp}(P_X)} \frac{P_{X|Y=y}(x)}{P_X(x)} \geq 0$  for all  $y \in \mathcal{Y}$ .

## 5.M Proof of Proposition 5.36

Fix some  $y \in \mathcal{Y}$  and define the set  $\mathcal{A}_y := \{x \in \text{supp}(P_X) : P_{X|Y=y}(x) \geq P_X(x)\}$ . We can write

$$\begin{aligned}
 \text{TV}(P_{X|Y=y}, P_X) &= \frac{1}{2} \sum_{x \in \text{supp}(P_X)} |P_{X|Y=y}(x) - P_X(x)| \\
 &= \sum_{x \in \mathcal{A}_y} P_{X|Y=y}(x) - P_X(x) \\
 &= \sum_{x \in \mathcal{A}_y} \left( \frac{P_{X|Y=y}(x)}{P_X(x)} - 1 \right) P_X(x) \\
 &\leq \max_{x \in \mathcal{A}_y} \left( \frac{P_{X|Y=y}(x)}{P_X(x)} - 1 \right) \sum_{x \in \mathcal{A}_y} P_X(x) \\
 &\leq \exp(\ell(X \rightarrow y)) - 1.
 \end{aligned} \tag{5.27}$$

Taking the expectation of the above expression over  $y$ , we get

$$T(X; Y) \leq \exp(\mathcal{L}(P_{Y|X})) - 1.$$

Now, define the function  $\eta(y) := \exp(\ell(X \rightarrow y)) - 1$  with  $y \in \mathcal{Y}$ . By (5.14) and (5.27), we obtain

$$T(X; Y) \leq \mathbb{E}_{Y \sim P_Y} \left[ \min \left\{ \eta(Y), \max \left\{ \frac{1}{2} \eta(Y), \frac{1}{2} \right\} \right\} \right].$$

Next, suppose the mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \delta)$ -PML. Define  $\eta_{\max} := e^{\epsilon_{\max}} - 1$ . Using the fact that  $\epsilon_{\max} \geq \log 2$ , we conclude that with probability smaller than  $\delta$  over  $Y$ , we have

$$\text{TV}(P_{X|Y=y}, P_X) \leq \frac{1}{2} \eta_{\max} = \frac{1}{2} (e^{\epsilon_{\max}} - 1). \tag{5.28}$$

We need to consider the following three cases for  $\epsilon$ :

- (i)  $\epsilon \leq \log \frac{3}{2}$ : With probability at least  $1 - \delta$  we have  $\text{TV}(P_{X|Y=y}, P_X) \leq \eta(y) \leq e^\epsilon - 1$ , which implies that

$$T(X; Y) \leq e^\epsilon - 1 + \frac{\delta}{2} (e^{\epsilon_{\max}} - 1).$$

- (ii)  $\log \frac{3}{2} \leq \epsilon \leq \log 2$ : With probability at least  $1 - \delta$  we have  $\text{TV}(P_{X|Y=y}, P_X) \leq \frac{1}{2}$ , which implies that

$$T(X; Y) \leq \frac{1}{2} + \frac{\delta}{2} (e^{\epsilon_{\max}} - 1).$$

- (iii)  $\epsilon \geq \log 2$ : With probability at least  $1 - \delta$  we have  $\text{TV}(P_{X|Y=y}, P_X) \leq \frac{1}{2} \eta(y) \leq \frac{1}{2} (e^\epsilon - 1)$ , which implies that

$$T(X; Y) \leq \frac{1}{2} (e^\epsilon - 1) + \frac{\delta}{2} (e^{\epsilon_{\max}} - 1).$$

## 5.N Proof of Proposition 5.42

Given  $i \in [n]$ , let  $B_i = f(D_i)$  be a binary random variable that determines whether or not entry  $D_i$  satisfies the predicate  $f$ . Since the outcome of the Laplace mechanism depends on  $D_i$  only through  $B_i$  the Markov chain  $D_i - B_i - Y$  holds and  $\ell(D_i \rightarrow y) \leq \ell(B_i \rightarrow y)$  for all outcomes  $y \in \mathcal{Y}$ . Thus, we may without loss of generality assume that  $D_i = B_i$ , that is, we assume that the database is binary.

For notational simplicity suppose  $i = 1$ . We write

$$\begin{aligned} \sup_{y \in \mathcal{Y}} \ell(D_1 \rightarrow y) &= \sup_{y \in \mathcal{Y}} \log \frac{\max_{d_1 \in \{0,1\}} p_{Y|D_1=d_1}(y)}{p_Y(y)} \\ &= \sup_{y \in \mathcal{Y}} \log \frac{\max_{d_1 \in \{0,1\}} \mathbb{E}_{D_{-1}} \left[ \exp\left(-\frac{|y - \frac{d_1}{n} - \frac{S_{-1}}{n}|}{b}\right) \right]}{\mathbb{E}_X \left[ \exp\left(-\frac{|y - \frac{S_X}{n}|}{b}\right) \right]}, \end{aligned}$$

where  $S_{-1} := \sum_{i=2}^n D_i$  and  $S_X := \sum_{i=1}^n D_i$ . We argue that it is sufficient to consider  $y > 1$  and  $y < 0$ . This is because in the numerator we have

$$\begin{aligned} &\mathbb{E}_{D_{-1}} \left[ \exp\left(-\frac{|y - \frac{d_1}{n} - \frac{S_{-1}}{n}|}{b}\right) \right] \\ &\leq \min \left\{ \mathbb{E}_{D_{-1}} \left[ \exp\left(-\frac{y - \frac{d_1}{n} - \frac{S_{-1}}{n}}{b}\right) \right], \mathbb{E}_{D_{-1}} \left[ \exp\left(\frac{y - \frac{d_1}{n} - \frac{S_{-1}}{n}}{b}\right) \right] \right\}. \end{aligned}$$

Furthermore, the mapping  $y \mapsto \mathbb{E}_X \left[ \exp\left(-\frac{|y - \frac{S_X}{n}|}{b}\right) \right]$  in the denominator is increasing in  $(-\infty, p]$  and decreasing in  $[p, \infty)$  since  $S_X$  is a Binomial random variable with success probability  $p$ .

Now, if  $y > 1$ , then

$$\ell(D_1 \rightarrow y) = \log \frac{\max_{d_1 \in \{0,1\}} \mathbb{E}_{D_{-1}} \left[ \exp\left(-\frac{y - \frac{d_1}{n} - \frac{S_{-1}}{n}}{b}\right) \right]}{\mathbb{E}_X \left[ \exp\left(-\frac{y - \frac{S_X}{n}}{b}\right) \right]}$$

$$\begin{aligned}
 &= \log \frac{\max_{d_1 \in \{0,1\}} \mathbb{E}_{D_{-1}} \left[ \exp\left(\frac{d_1}{nb} + \frac{S_{-1}}{nb}\right) \right]}{\mathbb{E}_X \left[ \exp\left(\frac{S_X}{nb}\right) \right]} \\
 &= \frac{1}{nb} + \log \frac{\mathbb{E}_{D_{-1}} \left[ \exp\left(\frac{D_2 + \dots + D_n}{nb}\right) \right]}{\mathbb{E}_X \left[ \exp\left(\frac{D_1 + \dots + D_n}{nb}\right) \right]} \\
 &= \frac{1}{nb} + \log \frac{\prod_{j=2}^n \mathbb{E} \left[ \exp\left(\frac{D_j}{nb}\right) \right]}{\prod_{j=1}^n \mathbb{E} \left[ \exp\left(\frac{D_j}{nb}\right) \right]} \\
 &= \frac{1}{nb} - \log \left( (1-p) + p \exp\left(\frac{1}{nb}\right) \right) \\
 &\leq \frac{1}{nb} - \log \left( (1-c) + c \exp\left(\frac{1}{nb}\right) \right),
 \end{aligned}$$

where the inequality is due to the fact that the mapping  $p \mapsto (1-p) + p \exp(\frac{1}{nb})$  is increasing in  $p$ . Similarly, if  $y < 0$ , then

$$\begin{aligned}
 \ell(D_1 \rightarrow y) &= \log \frac{\max_{d_1 \in \{0,1\}} \mathbb{E}_{D_{-1}} \left[ \exp\left(\frac{y - \frac{d_1}{n} - \frac{S_{-1}}{n}}{b}\right) \right]}{\mathbb{E}_X \left[ \exp\left(\frac{y - \frac{S_X}{n}}{b}\right) \right]} \\
 &= \log \frac{\max_{d_1 \in \{0,1\}} \mathbb{E}_{D_{-1}} \left[ \exp\left(-\frac{d_1}{nb} - \frac{S_{-1}}{nb}\right) \right]}{\mathbb{E}_X \left[ \exp\left(-\frac{S_X}{nb}\right) \right]} \\
 &= \log \frac{\mathbb{E}_{D_{-1}} \left[ \exp\left(-\frac{D_2 + \dots + D_n}{nb}\right) \right]}{\mathbb{E}_X \left[ \exp\left(-\frac{D_1 + \dots + D_n}{nb}\right) \right]} \\
 &= \log \frac{\prod_{j=2}^n \mathbb{E} \left[ \exp\left(-\frac{D_j}{nb}\right) \right]}{\prod_{j=1}^n \mathbb{E} \left[ \exp\left(-\frac{D_j}{nb}\right) \right]} \\
 &= \frac{1}{nb} - \log \left( p + (1-p) \exp\left(\frac{1}{nb}\right) \right) \\
 &\leq \frac{1}{nb} - \log \left( (1-c) + c \exp\left(\frac{1}{nb}\right) \right),
 \end{aligned}$$

where the last inequality is due to the fact that the mapping  $p \mapsto p + (1-p) \exp(\frac{1}{nb})$

is decreasing in  $p$ . We conclude that

$$\sup_{P_X \in \mathcal{P}_c^f} \sup_{y \in \mathbb{R}} \ell(D_1 \rightarrow y) = \frac{1}{nb} - \log \left( (1 - c) + c \exp \left( \frac{1}{nb} \right) \right).$$

---

## 6. Disclosure Prevention with PML

---

We discussed in Chapter 3 that the goal of differential privacy is to ensure that an individual partaking in a data processing scheme will not face substantially increased risks due to their participation. This guarantee is achieved by ensuring that the outcome of the data processing is not much affected by whether or not each person participates. On the other hand, differential privacy, by design, does not rule out the possibility of privacy violations by *association*. That is, an adversary can still exploit *correlations* among pieces of data to uncover sensitive information about an individual from the outcome of a differentially private mechanism. To account for potential privacy violations by association, Tschantz et al. [137] argue that differential privacy should be understood as a *causal* property of an algorithm. That is, differential privacy simply ensures that an algorithm produces similar outputs when supplied with inputs that differ in a single parameter. From the causal standpoint, (an unintended) inference about an individual is considered to be a privacy breach only if it is specifically caused by the inclusion of the individual's information in a dataset [78].

On the other hand, the above causal interpretation no longer applies if we adopt a Bayesian perspective and assume that databases are sampled from an underlying probability distribution. In particular, several works argue that from the Bayesian point of view, differential privacy either (implicitly) assumes a product distribution on the database or restricts itself to informed adversaries [75, 76, 58, 92, 149, 86, 156]. These works usually provide examples and attack scenarios involving databases containing highly correlated data points and then argue that differential privacy falls short of providing sufficient protection in these cases. For instance, Kifer and Machanavajjhala [75] give an example about a medical database in which Bob's data is perfectly correlated with the data of a large number of other patients. Then, they argue that the Laplace mechanism does not provide sufficient protection in this case since the effect of Bob's data is amplified by the other data points.

A privacy guarantee that can rule out the possibility of privacy breaches due to association must be *inferential* in nature, that is, it must ensure that an adversary's knowledge about the world after interacting with a mechanism does not change much from her prior knowledge.<sup>1</sup> However, inferential guarantees are generally considered to be impossible to achieve by the negative results of Dwork and

---

<sup>1</sup>We call a privacy notion *inferential* if it is defined by comparing an adversary's posterior and prior distributions. This includes definitions such as maximal leakage, pointwise maximal leakage, and (local) information privacy but excludes frameworks that simply assume an underlying distribution on the data, e.g., Pufferfish privacy [77] or Bayesian differential privacy [149].

Naor [33] and Kifer and Machanavajjhala [75] (see also [78, Sec. 7.1]). Particularly, Dwork and Naor [33] prove that (under certain assumptions), no mechanism providing *non-trivial utility* can prevent disclosures against adversaries who may possess *auxiliary information* about a secret  $X$ . This is because an adversary may exploit auxiliary information to disclose more information than what a privacy mechanism intended to release. As an illustrating example, suppose each person's exact height is a secret, and consider a database containing height measurements of people with different nationalities. Assume that the average heights of women of different nationalities are released. Then, an adversary who observes the released values and has the auxiliary information "Terry Gross is two inches shorter than the average Lithuanian woman" learns Terry Gross' exact height [33]. Here, if we adopt an inferential view of privacy naively we may conclude that Terry Gross' privacy rights are violated.

At a high level, Dwork and Naor [33] demonstrate that to provide utility a privacy mechanism necessarily has to disclose some information. To account for this result, differential privacy was designed to distinguish between data that is part of a dataset  $X$  and data that is not part of  $X$  but correlated with it, where the former is protected but the latter may be disclosed. We call this distinction the *in/out dichotomy*. In this chapter, we argue that the in/out dichotomy is not the only way of distinguishing between information that should be protected through privacy guarantees and information that may be disclosed. In particular, we present an alternative distinction termed the *local/global dichotomy*. The concept of the local/global dichotomy yields a fresh perspective on privacy which is compatible with the Bayesian view rather than the causal one required by differential privacy.

## 6.1 Notes on Notation and Terminology

For the following discussions, we need to distinguish between the true underlying distribution on the data and the belief of an adversary observing the outcome of a privacy mechanism. Suppose  $\mathcal{X}$  is a finite set. We use  $P_X$  to represent the true probability distribution of  $X$  and  $p_X$  to represent the probability mass function (pmf) of  $X$ . Without loss of generality, we assume that  $\mathcal{X} = \text{supp}(P_X)$ . Recall that  $\mathcal{P}_{\mathcal{X}}$  denotes the set of all distributions with full support on  $\mathcal{X}$ . We use  $Q_X \in \mathcal{P}_{\mathcal{X}}$  to represent an adversary's (prior) belief about  $X$ . For convenience, we identify adversaries with their prior beliefs. Note that  $Q_X$  may be different from the true distribution  $P_X$  on  $X$ , but we assume that  $Q_X$  and  $P_X$  are mutually absolutely continuous. We use  $q_X$  to denote the pmf of  $Q_X$ .

Suppose  $Y$  takes values in the set  $\mathcal{Y}$ , which may be finite or infinite. Given  $x \in \mathcal{X}$ , we use  $p_{Y|X=x}$  to denote the density of  $P_{Y|X=x}$  with respect to a suitable  $\sigma$ -finite measure on  $\mathcal{Y}$ . For example, when  $\mathcal{Y}$  is a countable set then we use the counting measure and when  $\mathcal{Y}$  is a Euclidean space then we use the Lebesgue measure. Similarly,  $P_Y$  denotes the distribution of  $Y$  induced by  $P_{Y|X}$  and  $P_X$

and  $p_Y$  denotes the density of  $P_Y$  with respect to a suitable measure on  $\mathcal{Y}$ .

We call a  $U$  satisfying the Markov chain  $U - X - Y$  an *attribute* or *feature* of  $X$ , which is induced by the probability kernel  $P_{U|X}$ . Following the terminology of Dwork and Naor [33, p. 96], we call  $P_{U|X}$  a piece of *auxiliary information* which describes how  $U$  depends on the secret  $X$ . We assume that  $U$  takes values in a finite set  $\mathcal{U}$ .

## Leakage Capacity

In Chapter 5, we defined several privacy guarantees by restricting PML in various ways. Here, we extend the definition of  $\epsilon$ -PML to encompass scenarios where  $P_X$  is not precisely known but is assumed to belong to a subset of  $\mathcal{P}_X$ .

**Definition 6.1** ( $(\epsilon, \mathcal{P})$ -PML). Suppose  $X$  is distributed according to  $P_X \in \mathcal{P} \subseteq \mathcal{P}_X$ . Given  $\epsilon \geq 0$ , we say that the mechanism  $P_{Y|X}$  satisfies  $(\epsilon, \mathcal{P})$ -PML if

$$P_Y \left( \{y \in \mathcal{Y} : \ell_{P_{Y|X} \times P_X}(X \rightarrow y) \leq \epsilon\} \right) = 1,$$

for all  $P_X \in \mathcal{P}$ , or equivalently, if

$$\sup_{P_X \in \mathcal{P}} D_\infty(P_{Y|X} \times P_X \| P_Y \times P_X) \leq \epsilon.$$

For simplicity, we assume that the density  $p_{Y|X=x}(y)$  is continuous on  $\mathcal{Y}$  for all  $x \in \mathcal{X}$ .<sup>2</sup> In this case,  $P_{Y|X}$  satisfies  $(\epsilon, \mathcal{P})$ -PML if

$$\sup_{P_X \in \mathcal{P}} \sup_{y \in \mathcal{Y}} \ell_{P_{Y|X} \times P_X}(X \rightarrow y) \leq \epsilon.$$

Now, we define the notion of the *leakage capacity* of a privacy mechanism motivated by Theorem 5.29. The leakage capacity of a mechanism describes the largest amount of information that can leak through that mechanism.

**Definition 6.2** (Leakage Capacity). The leakage capacity of a privacy mechanism  $P_{Y|X}$  is

$$C(P_{Y|X}) := \log \sup_{y \in \mathcal{Y}} \max_{x, x' \in \mathcal{X}} \frac{p_{Y|X=x}(y)}{p_{Y|X=x'}(y)}.$$

Note that  $C(P_{Y|X})$  is infinite if there exists  $(x, y) \in \mathcal{X} \times \mathcal{Y}$  such that  $p_{Y|X=x}(y) = 0$  but  $p_Y(y) > 0$ . In [49], the quantity  $\exp(C(P_{Y|X}))$  is called *lift capacity* and is used to establish a connection between a privacy measure called *max-case g-leakage* and local differential privacy.

<sup>2</sup>See [119, Remark 3.15] for a discussion on replacing the essential supremum by the actual supremum of a function.

## 6.2 Impossibility of Absolute Disclosure Prevention

Any aspiring inferential privacy framework needs to be reconciled with the results of [33] and [75], and this is the subject we take up in this section. We mainly discuss the results of [33], but we also draw connections to [75].<sup>3</sup>

In [32], Dwork proved a fundamental result that marks the beginning of the developments in the area of differential privacy. This result, dubbed the *impossibility result*, proves that no mechanism providing “non-trivial utility” can prevent disclosures against adversaries who may possess arbitrary *auxiliary information* about a secret  $X$ .<sup>4</sup> Roughly speaking, [33] demonstrates that an adversary can exploit auxiliary information to make unintended inferences about quantities correlated with  $X$ , as illustrated by the example about Terry Gross’ height. Thus, privacy guarantees that ensure neither  $X$  nor any quantity correlated with  $X$  is disclosed can be achieved only at the cost of destroying all utility, because it is not feasible to control for the adversary’s auxiliary information.

The impossibility result states that to provide utility, one necessarily has to disclose some information. This raises the question: What information can we (and should we) protect through privacy guarantees, and what information will we inevitably disclose? The answer differential privacy gives to this question is that privacy guarantees should be limited to information that is directly included in  $X$ . So, when  $X$  is a database, the individuals who have contributed their data to the database should be protected, but no such guarantee is provided to individuals whose data may be correlated with  $X$  in other ways. That is, a distinction is made between information that is directly included in  $X$  and information that is not part of  $X$  but may be correlated with it. We call this distinction the *in/out dichotomy*. As the basis for differential privacy, the in/out dichotomy has proved to be a very useful idea for addressing the impossibility result.

Nevertheless, the in/out dichotomy is not the only way we can distinguish between the information that we protect and the information that we allow to be disclosed. Below, we present an alternative distinction which we call the *local/global dichotomy*. The idea behind the local/global dichotomy is that we protect features of  $X$  that have large entropy (i.e., local features) while we allow disclosing features of  $X$  with small entropy (i.e., global features).<sup>5</sup> This view is motivated by how we define utility and what we consider to be a privacy breach. In particular, we argue that features of the data that capture properties of the population as a whole have small entropy and may be disclosed for the sake

---

<sup>3</sup>We emphasize that [33] and [75] prove conceptually different results. Specifically, [33] proves that absolute disclosure prevention is impossible due to the auxiliary information that may be available to an adversary, even if we assume a single fixed and publicly known prior distribution. On the other hand, [75] proves that guaranteeing privacy under all possible prior distributions severely restricts utility.

<sup>4</sup>The impossibility result is somewhat extended in [33] and we mostly refer to ideas from this later version.

<sup>5</sup>These entropies are calculated using the true prior distribution  $P_X$  on the data.

of utility, whereas instance-dependent features of the data have large entropy and should remain secret. Then, similar to how differential privacy provides guarantees according to the in/out dichotomy, we show that PML’s guarantees are based on the local/global dichotomy. In short, the local/global dichotomy allows us to reconcile the results of [33] with the guarantees of an effective inferential privacy framework. The main advantage of this view is that *those features of  $X$  that may be revealed by the privacy mechanism are exactly those population-level features of the data that we would anyway want to be able to disclose to provide utility*. On top of that, this view is directly applicable to many types of secrets and not just private databases.

In what follows, we assume that  $X$  is any type of data containing sensitive information, for example, a database or a piece of information belonging to a single individual.

### 6.3 What Is Privacy and What Is Utility?

Our results and discussions throughout this chapter depend crucially on definitions of utility and privacy formulated in terms of entropy. As such, in this section, we recall the definitions and assumptions of [33], in particular, the notions of utility and privacy posited there. We then present our own definitions and assumptions and discuss how they differ from that of [33].

Suppose  $X$  is distributed according to  $P_X \in \mathcal{P}_X$ . Dwork and Naor [33] assume that  $P_X$  is publicly known, that is,  $P_X$  also represents the prior belief of an adversary who interacts with a privacy mechanism  $P_{Y|X}$ . To define utility, Dwork and Naor [33] posit a random variable  $U$  satisfying the Markov chain  $U - X - Y$  whose value represents the answer to a question posed about  $X$ . It is assumed that the value of  $U$  cannot be *a priori* predicted from its distribution  $P_U = P_{U|X} \circ P_X$ , that is, the entropy  $H_\infty(P_U)$  is large. However, to provide utility, the mechanism must either disclose the value of  $U$  exactly or allow estimating  $U$  with high accuracy, i.e., it is assumed that there exists  $y \in \mathcal{Y}$  such that the entropy  $H_\infty(P_{U|Y=y})$  is either very small or zero. Furthermore, to define privacy, Dwork and Naor [33] suppose the existence of a random variable  $W$  satisfying the Markov chain  $W - X - Y$  whose value must remain secret. That is, the value of  $W$  must be difficult to guess with or without access to the mechanism, but it is assumed that  $W$  has smaller entropy compared to  $U$ . Formally speaking,  $H_\infty(P_W)$  and  $H_\infty(P_{W|Y=y})$  are both large for all  $y \in \mathcal{Y}$ , but  $H_\infty(P_W) < H_\infty(P_U)$ . It is important to note that the condition  $H_\infty(P_W) < H_\infty(P_U)$ <sup>6</sup> is indispensable in the proof of the impossibility result because [33] assumes that it is possible to extract enough randomness from  $U$  to mask the value of  $W$ .

Our setup differs from [33] in several key aspects. We let distribution  $Q_X \in \mathcal{P}_X$

---

<sup>6</sup>This condition is implied by the lower bound on the entropy of the utility vector in terms of the length of the privacy breach in [33, Assumption 1].

represent the prior belief of an adversary who observes the outcome of the privacy mechanism. This distribution may or may not be equal to  $P_X$ , but  $P_X$  and  $Q_X$  are mutually absolutely continuous. To provide utility, the mechanism  $P_{Y|X}$  releases some *global* information about the secret  $X$ , and releasing this information is *not* considered to be a privacy breach. We define global information as the value of any attribute of  $X$  that can be accurately predicted by an analyst who knows the true distribution  $P_X$  and possibly some auxiliary information but without access to the privacy mechanism  $P_{Y|X}$ . Formally, we posit a Markov chain  $U - X - Y$ , where  $U$  is an attribute of  $X$  and the kernel  $P_{U|X}$  is the analyst's auxiliary information. If  $U$  contains global information about  $X$ , then the entropy  $H_\infty(P_U)$  must be *small* since the value of  $U$  should be predictable using the distribution  $P_U$  alone (where  $P_U = P_{U|X} \circ P_X$ ) and without access to  $P_{Y|X}$ . Heuristically, such attributes describe properties of the population of  $X$  and are largely instance-independent. Hence, they may be disclosed to provide utility. By contrast, to maintain privacy, we wish to protect instance-dependent and *local* properties of  $X$ , which are represented by those attributes of  $X$  that have *large* entropy. Consider an attribute  $W$  of  $X$  satisfying the Markov chain  $W - X - Y$ . If  $H_\infty(P_W)$  is large (where  $P_W = P_{W|X} \circ P_X$ ), then even an analyst who knows the true underlying distribution  $P_X$  and the auxiliary information  $P_{W|X}$  cannot reliably estimate  $W$ ; hence, it is only through the mechanism  $P_{Y|X}$  that the value of  $W$  can be disclosed. Accordingly, we consider it to be a privacy breach if the value of any high-entropy attribute of  $X$  is disclosed.

The above distinction between high-entropy local features of  $X$  and low-entropy global features of it is what was earlier called the local/global dichotomy. This is further illustrated by the examples below, where the second example is inspired by [71].

**Example 6.3.** Suppose the database  $X = (D_1, \dots, D_n)$  is i.i.d, where each entry  $D_i$  is drawn according to a distribution  $P_D$  defined over a finite set of real numbers in the interval  $[a, b]$ . Our goal is to estimate the expectation  $\mu = \mathbb{E}_{P_D}[D_i]$ . We may aim to disclose one of the following two estimates: the quantized sample mean  $\hat{\mu}_1 = q_m \left( \frac{\sum_{i=1}^n D_i}{n} \right)$ , or the first row of the database  $\hat{\mu}_2 = D_1$ . The quantization  $q_m(\cdot)$  can be described as follows: Fix a large integer  $m$ , and values  $c_1, \dots, c_{m-1}$  satisfying  $a = c_0 < c_1 < \dots < c_m = b$ . Let  $\mathcal{C} = \left\{ \frac{c_0 + c_1}{2}, \dots, \frac{c_{m-1} + c_m}{2} \right\}$ . Then,  $q_m : [a, b] \rightarrow \mathcal{C}$  denotes a quantizer that maps real numbers in the interval  $[c_j, c_{j+1})$  to  $\frac{c_j + c_{j+1}}{2}$ .<sup>7</sup>

By the law of large numbers, as  $n \rightarrow \infty$  the sample mean converges in probability to  $\mu$ ; thus,  $H_\infty(\hat{\mu}_1) \rightarrow 0$ . In contrast, the distribution of  $\hat{\mu}_2$  does not depend on  $n$ ; hence,  $\hat{\mu}_2$  has larger entropy compared to  $\hat{\mu}_1$ . Therefore, a privacy mechanism is

<sup>7</sup>By the central limit theorem, the sample mean converges in distribution to a Gaussian random variable as  $n \rightarrow \infty$ . Thus, we use the quantization to ensure that the entropy of our estimator remains well-defined as  $n \rightarrow \infty$ . The quantization introduces some bias, which can be made arbitrarily small by taking  $m$  sufficiently large.

allowed to disclose the value of  $\hat{\mu}_1$  for the sake of utility but  $\hat{\mu}_2$  must be kept secret for the sake of privacy.

The above example also sheds light on Terry Gross' case: If the average height of Lithuanian women is released using a low-entropy accurate estimator with suitable convergence properties (e.g.  $\hat{\mu}_1$ ), then we do not consider the disclosure of her height as a privacy breach. This is because an adversary who knows the distribution of women's height can predict her height even without access to the mechanism.

**Example 6.4.** An insurance company has access to an i.i.d medical database  $X$  of size  $n$  and queries it through  $P_{Y|X}$  to obtain (quantized) relative frequencies  $\hat{p}_s$  and  $\hat{p}_{ns}$  describing the empirical probabilities of developing lung disease for smokers and non-smokers, respectively. Let  $p_s$  and  $p_{ns}$  denote the true probabilities of developing lung disease for smokers and non-smokers, which can be calculated from the prior distribution  $P_X$ . If  $n$  is large, then the estimates  $\hat{p}_s$  and  $\hat{p}_{ns}$  have small entropies and well-approximate the true probabilities.

Now, suppose based on  $\hat{p}_s$  and  $\hat{p}_{ns}$  the company draws some conclusions about Bob's probability of developing lung disease, and adjusts his insurance premium accordingly. Assuming that  $\hat{p}_s$  and  $\hat{p}_{ns}$  well-approximate the true probabilities, we do not consider this to be in violation of Bob's privacy (regardless of his participation in the database). This is because the insurance company could have drawn the same conclusions about Bob from the prior  $P_X$  even without access to the privacy mechanism.

In essence, the differences between our setup and [33] stem from the fundamental principle that if an analyst knows the true distribution  $P_X$  on the data, then they should be granted no further utility. Interestingly, the local/global dichotomy also allows us to distinguish between *adversarial* and *non-adversarial* analysts. The non-adversarial analyst Alice is only interested in the value of low-entropy attributes of  $X$ , which reflect properties of the population as a whole. If Alice knows  $P_X$ , then she gains no further value from interacting with the mechanism  $P_{Y|X}$ . On the other hand, the adversarial analyst Eve even equipped with  $P_X$  is motivated to query  $X$  through  $P_{Y|X}$  to uncover the value of high-entropy, instance-dependent, and local features of  $X$  which she cannot *a priori* predict, even if she possesses arbitrary auxiliary information.

## 6.4 Entropy-based Disclosure Prevention

Equipped with our definitions of privacy and utility, in this section, we state the main results of the chapter: that (a) disclosing a piece of information (in the sense of Definition 6.5) to one adversary in  $\mathcal{P}_X$  is tantamount to disclosing that information to all adversaries in  $\mathcal{P}_X$  (Theorem 6.6), and (b) PML provides privacy guarantees according to the local/global dichotomy (Theorem 6.7). In particular,

we show that if a mechanism  $P_{Y|X}$  satisfies  $(\epsilon, P_X)$ -PML, then it cannot disclose the value of any attribute of  $X$  with entropy greater than  $\epsilon$  to any adversary with prior belief in the set  $\mathcal{P}_X$ . Afterwards, in the spirit of the impossibility result, we prove that when a mechanism discloses the value of an attribute  $U$  of  $X$ , then it also discloses another attribute of  $X$  with smaller prior entropy compared to  $U$ . Finally, towards the end of this section, we discuss *absolute disclosure prevention*, i.e., we examine the condition ensuring that no attribute of  $X$  is disclosed by a privacy mechanism.

We begin by formally defining a notion of *disclosure*. Consider an adversary with prior belief  $Q_X \in \mathcal{P}_X$ , and let  $U$  be an attribute of  $X$ . Then, the adversary's prior belief about  $U$  is  $Q_U = P_{U|X} \circ Q_X$ . We may define disclosure as the event that the adversary's belief about  $U$  changes after observing an outcome of the privacy mechanism.<sup>8</sup> That is, disclosure is the event that  $Q_U \neq Q_{U|Y=y}$  for some  $y \in \mathcal{Y}$ , where  $Q_{U|Y=y} = P_{U|X} \circ Q_{X|Y=y}$  denotes the adversary's posterior belief about  $U$  after observing  $y$ . Thus, disclosure prevention requires that  $Y$  and  $U$  be independent. Clearly, this is a very stringent requirement and may necessitate the independence of  $X$  and  $Y$ ,<sup>9</sup> e.g, if  $U = X$ . Hence, we instead postulate the following weaker but more intuitive definition that also matches the notions of disclosure considered in [33] and [75].

**Definition 6.5** (Disclosure). Let  $U$  be an attribute of  $X$ . We say that the privacy mechanism  $P_{Y|X}$  discloses the value of  $U$  to adversary  $Q_X \in \mathcal{P}_X$  if

$$\inf_{y \in \mathcal{Y}} H_\infty(Q_{U|Y=y}) = 0.$$

Henceforth, we use the terms “disclosure” and “disclose” in the sense of Definition 6.5. The following theorem asserts that, in fact, we do not need to specify to which adversary a piece of information has been disclosed. This is because disclosures are ubiquitous across  $\mathcal{P}_X$ .

**Theorem 6.6** (Ubiquity of Disclosures). *Let  $U$  be an attribute of  $X$ . If the privacy mechanism  $P_{Y|X}$  discloses the value of  $U$  to an adversary  $Q_X \in \mathcal{P}_X$ , then it also discloses the value of  $U$  to all other adversaries in  $\mathcal{P}_X$ .*

*Proof.* Consider the Markov chain  $U - X - Y$  and suppose  $P_{Y|X}$  discloses the value of  $U$  to adversary  $Q_X \in \mathcal{P}_X$ . Fix  $R_X \in \mathcal{P}_X$ . First, we argue that since  $Q_X$  and  $R_X$  are mutually absolutely continuous, then the posterior distributions  $Q_{X|Y=y}$  and  $R_{X|Y=y}$  are also mutually absolutely continuous for all  $y \in \mathcal{Y}$ . Let  $f(x) = \frac{r_X(x)}{q_X(x)}$  denote the Radon-Nikodym derivate of  $R_X$  with respect to  $Q_X$  and observe that

<sup>8</sup>This is often called Dalenius' desideratum in the literature.

<sup>9</sup>[115] show that under certain conditions it is possible to design  $P_{Y|X}$  such that  $Y$  is independent of  $U$  but correlated with  $X$ .

$f(x) > 0$  for all  $x \in \mathcal{X}$ . Fix an arbitrary  $x \in \mathcal{X}$  and  $y \in \mathcal{Y}$ . We have

$$\begin{aligned} r_{X|Y=y}(x) &= \frac{p_{Y|X=x}(y) \cdot r_X(x)}{r_Y(y)} \\ &= \frac{p_{Y|X=x}(y) \cdot f(x) \cdot q_X(x)}{r_Y(y)} \\ &= \frac{q_{X|Y=y}(x) \cdot f(x) \cdot q_Y(y)}{r_Y(y)} \\ &= q_{X|Y=y}(x) \cdot g(x, y), \end{aligned}$$

where  $g(x, y) := \frac{f(x) \cdot q_Y(y)}{r_Y(y)}$  is strictly positive for all  $x \in \mathcal{X}$  and all  $y \in \mathcal{Y}$ . Thus, if  $q_{X|Y=y}(x)$  is positive, then so is  $r_{X|Y=y}(x)$  and vice versa, proving that the posterior distributions  $Q_{X|Y=y}$  and  $R_{X|Y=y}$  are mutually absolutely continuous for all  $y \in \mathcal{Y}$ . Next, we note that  $g(x, y)$  is bounded above because

$$\begin{aligned} \max_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} g(x, y) &= \left( \max_{x \in \mathcal{X}} f(x) \right) \sup_{y \in \mathcal{Y}} \frac{q_Y(y)}{r_Y(y)} \\ &= \left( \max_{x \in \mathcal{X}} f(x) \right) \exp \left( D_\infty(Q_Y \| R_Y) \right) \\ &\leq \left( \max_{x \in \mathcal{X}} f(x) \right) \exp \left( D_\infty(Q_X \| R_X) \right), \end{aligned}$$

where the inequality is due to the data-processing inequality for Rényi divergence [139, Thm. 9]. The Rényi divergence  $D_\infty(Q_X \| R_X)$  is also finite since  $Q_X$  and  $R_X$  are mutually absolutely continuous. Let  $c > 0$  be a constant satisfying  $\max_{x \in \mathcal{X}} \sup_{y \in \mathcal{Y}} g(x, y) < c$ .

Fix a small  $\varepsilon > 0$  and an outcome  $y \in \mathcal{Y}$  with  $H_\infty(Q_{U|Y=y}) < \varepsilon$ . Then, there exists  $u^* \in \mathcal{U}$  such that  $q_{U|Y=y}(u^*) > e^{-\varepsilon}$ , which in turn implies that  $q_{U|Y=y}(u) < 1 - e^{-\varepsilon}$  for all  $u \neq u^*$ . Now, for all  $u \neq u^*$  we can write

$$\begin{aligned} r_{U|Y=y}(u) &= \sum_{x \in \mathcal{X}} p_{U|X=x}(u) \cdot r_{X|Y=y}(x) \\ &= \sum_{x \in \mathcal{X}} p_{U|X=x}(u) \cdot g(x, y) \cdot q_{X|Y=y}(x) \\ &\leq \left( \max_{x \in \mathcal{X}} g(x, y) \right) \sum_{x \in \mathcal{X}} p_{U|X=x}(u) \cdot q_{X|Y=y}(x) \\ &= \left( \max_{x \in \mathcal{X}} g(x, y) \right) q_{U|Y=y}(u) \\ &< \left( \max_{x \in \mathcal{X}} g(x, y) \right) (1 - e^{-\varepsilon}) \\ &< c \cdot (1 - e^{-\varepsilon}). \end{aligned}$$

Thus, we get  $r_{U|Y=y}(u^*) = 1 - \sum_{u \neq u^*} r_{U|Y=y}(u) > 1 - (|\mathcal{U}| - 1) \cdot c \cdot (1 - e^{-\epsilon})$ . Finally, taking  $\epsilon \rightarrow 0$  yields  $r_{U|Y=y}(u^*) \rightarrow 1$  and we conclude that  $\inf_{y \in \mathcal{Y}} H_\infty(R_{U|Y=y}) = 0$ . In other words,  $P_{Y|X}$  discloses the value of  $U$  to adversary  $R_X \in \mathcal{P}_X$ .  $\square$

We now exploit Theorem 6.6 to prove that PML-based privacy guarantees prevent disclosing high-entropy attributes of  $X$  to all adversaries in  $\mathcal{P}_X$ .

**Theorem 6.7** (Disclosure prevention via PML). *Suppose  $X$  is distributed according to  $P_X$ , and let  $U$  be an attribute of  $X$  with entropy  $H_\infty(P_U) > \epsilon$ , where  $\epsilon \geq 0$  and  $P_U = P_{U|X} \circ P_X$ . If the privacy mechanism  $P_{Y|X}$  satisfies  $(\epsilon, P_X)$ -PML, then  $P_{Y|X}$  cannot disclose the value of  $U$  to any adversary  $Q_X \in \mathcal{P}_X$ .*

*Proof.* Fix some  $U$  satisfying the Markov chain  $U - X - Y$  with entropy  $H_\infty(P_U) > \epsilon$ , where  $P_U = P_{U|X} \circ P_X$ . First, consider an adversary with prior belief  $P_X$ . Let  $P_{UY} = (P_{U|X} \times P_{Y|X}) \circ P_X$  denote the joint distribution of  $U$  and  $Y$ . Fix an arbitrary  $y \in \mathcal{Y}$ . We can write

$$\begin{aligned} \ell_{P_{UY}}(U \rightarrow y) &= \log \max_{u \in \text{supp}(P_U)} \frac{p_{U|Y=y}(u)}{p_U(u)} \\ &\geq \log \max_{u \in \text{supp}(P_U)} p_{U|Y=y}(u) + \log \frac{1}{\max_{u \in \text{supp}(P_U)} p_U(u)} \\ &\geq \log \max_{u \in \text{supp}(P_{U|Y=y})} p_{U|Y=y}(u) + H_\infty(P_U) \\ &= H_\infty(P_U) - H_\infty(P_{U|Y=y}), \end{aligned} \tag{6.1a}$$

where (6.1a) is due to the fact that  $\text{supp}(P_{U|Y=y}) \subseteq \text{supp}(P_U)$  for all  $y \in \mathcal{Y}$ . That is, we have

$$\begin{aligned} H_\infty(P_{U|Y=y}) &\geq H_\infty(P_U) - \ell_{P_{UY}}(U \rightarrow y) \\ &\geq H_\infty(P_U) - \ell_{P_{XY}}(X \rightarrow y), \end{aligned} \tag{6.2}$$

where the second inequality follows from the pre-processing inequality for PML. Now, assuming that  $P_{Y|X}$  satisfies  $(\epsilon, P_X)$ -PML, taking the supremum over  $y \in \mathcal{Y}$  yields

$$\inf_{y \in \mathcal{Y}} H_\infty(P_{U|Y=y}) \geq H_\infty(P_U) - \sup_{y \in \mathcal{Y}} \ell_{P_{XY}}(X \rightarrow y) > 0. \tag{6.3}$$

Therefore,  $P_{Y|X}$  cannot disclose the value of  $U$  to adversary  $P_X$ . Finally, by Theorem 6.6,  $P_{Y|X}$  cannot disclose the value of  $U$  to any adversary in  $\mathcal{P}_X$ .  $\square$

The above theorem contains a powerful idea: It states that if we protect the data under its true distribution, then we are simultaneously preventing privacy breaches against all adversaries in  $\mathcal{P}_X$ . Furthermore, Theorem 6.7 demonstrates that the two goals of privacy and utility are not inherently at odds with each other.

This is because while PML imposes lower bounds on the remaining uncertainty in the value of high-entropy local attributes of  $X$ , it does not directly restrict the remaining uncertainty in the value of low-entropy global attributes of  $X$ . Indeed, when the answer to a query describes a feature of  $X$  that has very small entropy, it may even be safe to answer it precisely and without any randomness. We give an example of a query answered deterministically in Section 6.5.

It is worth emphasizing that Theorem 6.7 does *not* mean that mechanism  $P_{Y|X}$  leaks the same amount of information to all adversaries. In fact, an attribute  $U$  of  $X$  that has small entropy under the true distribution  $P_X$  may have very large entropy according to the belief of adversary  $Q_X$ . In this case, a privacy mechanism that discloses the value of  $U$  leaks a large amount of information to adversary  $Q_X$ , and this leakage is captured by  $\ell_{Q_{XY}}(X \rightarrow y)$ , where  $Q_{XY} = P_{Y|X} \times Q_X$ . Nevertheless, Theorem 6.7 asserts that we need not be alarmed by the large value of  $\ell_{Q_{XY}}(X \rightarrow y)$  because despite this large leakage, adversary  $Q_X$  will not be able to infer the value of any local features of  $X$ . Put differently, while we may use PML *subjectively* to calculate the amount of information leaked to each adversary, the parameter  $\epsilon$  of the privacy guarantee should be determined and interpreted *objectively* according to our assumptions about the true underlying distribution on the data.

As a converse to Theorem 6.7, we now show that when the mechanism  $P_{Y|X}$  discloses the value of an attribute  $U$  of  $X$ , then we can no longer guarantee privacy for attributes of  $X$  with entropies smaller than  $H_\infty(P_U)$ . In fact, disclosing  $U$  inevitably leads to disclosing another attribute of  $X$  with a smaller entropy compared to  $U$ . The following result is proved in Appendix 6.A.

**Proposition 6.8.** *Suppose  $X$  is distributed according to  $P_X$ . Assume that the privacy mechanism  $P_{Y|X}$  discloses the value of an attribute of  $X$ , denoted by  $U$ . Then, there exists an attribute of  $X$ , denoted by  $W$ , satisfying  $H_\infty(P_W) < H_\infty(P_U)$  whose value is also disclosed.*

Proposition 6.8 is conceptually similar to the impossibility result (specifically, [33, Thm. 3]); yet, it is interpreted differently in our framework: If  $U$  is disclosed to provide utility, then  $U$  has small entropy and can be estimated accurately using its distribution  $P_U$  alone. Since  $H_\infty(P_W) < H_\infty(P_U)$ , then  $W$  can too be estimated accurately using its distribution  $P_W$ , even without access to the privacy mechanism. Thus, if disclosing  $U$  is not considered as a privacy breach, then disclosing  $W$  should not be considered as a privacy breach either. It is also worth mentioning that the proof of Proposition 6.8 requires no assumptions about the mechanism  $P_{Y|X}$  other than the fact that it discloses  $U$ . Hence, the result holds even if we assume that  $P_{Y|X}$  satisfies  $(\epsilon, P_X)$ -PML with  $\epsilon \geq H_\infty(P_U)$ .

As the final topic in this section, we discuss *absolute disclosure prevention*, i.e., we investigate conditions ensuring that *no* attribute of  $X$  can be disclosed by the mechanism  $P_{Y|X}$ . We show that absolute disclosure prevention can be achieved by mechanisms that have finite leakage capacity (see Definition 6.2). Moreover,

we prove that these mechanisms guarantee a lower bound on the remaining uncertainty in the value of all (non-constant) deterministic attributes of  $X$  for all adversaries in  $\mathcal{P}_X$ .

**Theorem 6.9** (Absolute disclosure prevention). *If  $P_{Y|X}$  satisfies  $C(P_{Y|X}) < \infty$ , then for all  $P_X \in \mathcal{P}_X$  no attribute of  $X$  can be disclosed by  $P_{Y|X}$ . Furthermore, given an arbitrary (non-constant) deterministic function of  $X$ , denoted by  $V$ , the remaining uncertainty in the value of  $V$  for adversary  $Q_X \in \mathcal{P}_X$  is at least*

$$H_\infty(Q_{V|Y=y}) \geq \log \left( 1 + \frac{\min_x q_X(x)}{1 - \min_x q_X(x)} e^{-C(P_{Y|X})} \right),$$

for all  $y \in \mathcal{Y}$ .

*Proof.* Consider an adversary  $Q_X \in \mathcal{P}_X$  and suppose  $C(P_{Y|X}) < \infty$ . We prove the theorem by contradiction. Suppose  $P_{Y|X}$  discloses the value of an attribute of  $X$ , denoted by  $U$ . Then, for each  $\varepsilon > 0$  there exists  $y \in \mathcal{Y}$  such that  $H_\infty(Q_{U|Y=y}) < \varepsilon$ , or equivalently,  $q_{U|Y=y}(u) > e^{-\varepsilon}$  for some  $u \in \mathcal{U}$ . Denote this outcome by  $u_1$ . By Bayes' theorem, we have

$$q_{Y|U=u_1}(y) = \frac{q_{U|Y=y}(u_1)q_Y(y)}{q_U(u_1)} > \frac{q_Y(y)}{q_U(u_1)} \cdot e^{-\varepsilon}.$$

On the other hand, we also have

$$q_{Y|U=u_1}(y) = \sum_{x \in \mathcal{X}} p_{Y|X=x}(y)q_{X|U=u_1}(x) \leq \max_x p_{Y|X=x}(y),$$

hence, we get  $\max_x p_{Y|X=x}(y) > \frac{q_Y(y)}{q_U(u_1)} \cdot e^{-\varepsilon}$ . Furthermore, for all  $u \neq u_1$  we have  $q_{U|Y=y}(u) < 1 - e^{-\varepsilon}$ . Let  $u_2$  be one such outcome. Once again, Bayes' theorem yields

$$\sum_{x \in \mathcal{X}} p_{Y|X=x}(y)q_{X|U=u_2}(x) = q_{Y|U=u_2}(y) = \frac{q_{U|Y=y}(u_2)q_Y(y)}{q_U(u_2)} < \frac{q_Y(y)}{q_U(u_2)}(1 - e^{-\varepsilon})$$

which, in turn, implies that  $p_{Y|X=x}(y)q_{X|U=u_2}(x) < \frac{q_Y(y)}{q_U(u_2)}(1 - e^{-\varepsilon})$  for all  $x \in \mathcal{X}$ . Now, since  $\sum_x q_{X|U=u_2}(x) = 1$ ,  $q_{X|U=u_2}(x)$  must be strictly positive for at least one  $x \in \mathcal{X}$ . Let  $x^* \in \mathcal{X}$  be one such outcome. Hence, we get

$$p_{Y|X=x^*}(y) < \frac{q_Y(y)}{q_U(u_2) \cdot q_{X|U=u_2}(x^*)} (1 - e^{-\varepsilon}).$$

Finally, we get

$$\exp(C(P_{Y|X})) > \frac{\max_x p_{Y|X=x}(y)}{p_{Y|X=x^*}(y)} > \frac{q_U(u_2) \cdot q_{X|U=u_2}(x^*)}{q_U(u_1)} \cdot \frac{e^{-\varepsilon}}{1 - e^{-\varepsilon}} > c \cdot \frac{e^{-\varepsilon}}{1 - e^{-\varepsilon}},$$

where  $c > 0$  is a suitably small constant. Then, by letting  $\varepsilon \rightarrow 0$ , we conclude that the capacity  $C(P_{Y|X})$  is infinite which is a contradiction. This proves the first statement.

To prove the second statement, suppose  $V$  is a deterministic function of  $X$  which is induced by the kernel  $P_{V|X}$  and takes values in the set  $\mathcal{V}$ . Fix an arbitrary  $v \in \mathcal{V}$  and define  $\mathcal{X}_v := \{x \in \mathcal{X} : p_{V|X=x}(v) = 1\}$ . Note that  $p_{V|X=x}(v) = 0$  for all  $x \notin \mathcal{X}_v$ . Fix an arbitrary  $y \in \mathcal{Y}$  and let  $r_{\min} = \min_x p_{Y|X=x}(y)$  and  $r_{\max} = \max_x p_{Y|X=x}(y)$ . Observe that  $\exp(C(P_{Y|X})) \geq \frac{r_{\max}}{r_{\min}}$ . Let  $Q_{VY} = (P_{V|X} \times P_{Y|X}) \circ Q_X$  denote the joint distribution of  $V$  and  $Y$ . We can write

$$\begin{aligned} q_{V|Y=y}(v) &= \frac{q_{VY}(v, y)}{q_Y(y)} = \frac{\sum_{x \in \mathcal{X}} p_{V|X=x}(v) p_{Y|X=x}(y) q_X(x)}{\sum_{x \in \mathcal{X}} p_{Y|X=x}(y) q_X(x)} \\ &= \frac{\sum_{x \in \mathcal{X}_v} p_{Y|X=x}(y) q_X(x)}{\sum_{x \in \mathcal{X}} p_{Y|X=x}(y) q_X(x)} \\ &= \frac{1}{1 + \frac{\sum_{x \notin \mathcal{X}_v} p_{Y|X=x}(y) q_X(x)}{\sum_{x \in \mathcal{X}_v} p_{Y|X=x}(y) q_X(x)}} \\ &\leq \frac{1}{1 + \frac{r_{\min} (1 - Q_X(\mathcal{X}_v))}{r_{\max} Q_X(\mathcal{X}_v)}} \\ &\leq \frac{1}{1 + \frac{\min_x q_X(x)}{\exp(C(P_{Y|X}))(1 - \min_x q_X(x))}}. \end{aligned}$$

Thus, we get

$$\begin{aligned} H_\infty(Q_{V|Y=y}(v)) &= \log \frac{1}{\max_v q_{V|Y=y}(v)} \\ &\geq \log \left( 1 + \frac{\min_x q_X(x)}{1 - \min_x q_X(x)} e^{-C(P_{Y|X})} \right). \end{aligned}$$

□

By Theorem 5.29, a privacy mechanism  $P_{Y|X}$  has finite leakage capacity if and only if it satisfies  $(\epsilon, \mathcal{P}_X)$ -PML with some finite value of  $\epsilon$ . As such, the above result contains a similar idea to the no-free-lunch theorem of [75]. More precisely, [75, Thm. 2.1] states that it is not possible to discriminate between different instances of the secret  $X$  if we guarantee privacy under all possible distributions on the data. That is, utility is essentially destroyed when we make no assumptions about the data-generating distribution. Here, however, we may have a different take on Theorem 6.9 when viewed through the lens of the local/global dichotomy:

Guaranteeing privacy under all possible distributions entails that we no longer can distinguish between local and global features of the data. For example, an attribute  $U$  of  $X$  may have small entropy under distribution  $P_X^{(1)} \in \mathcal{P}_X$  but large entropy under another distribution  $P_X^{(2)} \in \mathcal{P}_X$ . Since no non-trivial attribute of  $X$  can have consistently small entropy under all possible distributions in  $\mathcal{P}_X$ , then no attribute of  $X$  can be considered to capture a property of the whole population. Hence, we inevitably protect all features of  $X$ . In other words, when we make no assumptions about the data-generating distribution, then a privacy mechanism provides no utility because there is no utility to be provided.

## 6.5 How to Pick $\epsilon$ ?

According to Theorem 6.7, if a mechanism  $P_{Y|X}$  satisfies  $(\epsilon, P_X)$ -PML, then it cannot disclose the value of any attribute of  $X$  with entropy larger than  $\epsilon$  to any adversary in  $\mathcal{P}_X$ . Essentially,  $\epsilon$  describes where (in terms of entropy) we draw the line between global and local features of  $X$ , and smaller  $\epsilon$  implies stricter privacy requirements. We may select  $\epsilon$  by asking: Which features of  $X$  do we consider to be sufficiently easy to guess by an analyst who knows  $P_X$  such that they may be disclosed without causing a privacy breach? Conversely, we may ask: Which features of  $X$  do we wish to keep secret even from an analyst who knows  $P_X$  and what is the entropy of those features? In this section, we give a few concrete examples of attributes of  $X$  that are disclosed at different values of  $\epsilon$ . We also argue that  $\epsilon$  should always remain below the entropy of the data  $H_\infty(P_X)$ .

First, we establish the existence of an attribute of  $X$  which can be disclosed at the smallest  $\epsilon$  compared to all other attributes of  $X$ . Let  $p_{\min} := \min_x p_X(x)$  and  $x_{\min} \in \mathcal{X}$  be a realization of  $X$  with probability  $p_{\min}$ . The following result is proved in Appendix 6.B.

**Proposition 6.10.** *Suppose  $X$  is distributed according to  $P_X$ . If the privacy mechanism  $P_{Y|X}$  satisfies  $(\epsilon, P_X)$ -PML with  $\epsilon < \log \frac{1}{1-p_{\min}}$ , then  $C(P_{Y|X}) < \infty$ . Conversely, for each  $\epsilon \geq \log \frac{1}{1-p_{\min}}$  there exists an attribute  $U$  of  $X$  and a privacy mechanism  $P_{Y|X}$  satisfying  $(\epsilon, P_X)$ -PML that discloses the value of  $U$ .*

The second statement in Proposition 6.10 is proved by constructing an attribute of  $X$  which requires the smallest privacy cost (i.e.,  $\epsilon = \log \frac{1}{1-p_{\min}}$ ) to be disclosed. This attribute describes a binary random variable that determines whether or not  $X$  has value  $x \neq x_{\min}$ . Thus, Proposition 6.10 essentially states that giving an affirmative answer (deterministically) to the query “Is  $X \in \mathcal{X} \setminus \{x_{\min}\}$ ?” induces the smallest privacy cost. Note that an analyst who possesses  $P_X$  can correctly predict the answer to this query with probability  $1 - p_{\min}$  even without access to the mechanism. On the other hand, Issa et al. [63, Thm. 1] construct an attribute of  $X$  which takes the largest privacy cost to be disclosed. Roughly speaking, [63, Thm. 1] shows that an affirmative answer

can be given to the query “Is  $X \in \{x_{\min}\}$ ?” when  $\epsilon \geq \log \frac{1}{p_{\min}}$ . Note that the answer to this query is correctly guessed (without access to the mechanism) with the small probability of  $p_{\min}$ , and that at  $\epsilon = \log \frac{1}{p_{\min}}$  a mechanism is allowed to answer all possible queries about  $X$  error-free.

Of course,  $\epsilon$  should be picked such that no realization of  $X$  can be disclosed. That is,  $P_{Y|X}$  should not be able to deterministically give an affirmative answer to any query of the form “Is  $X \in \{x\}$ ?” for any  $x \in \mathcal{X}$ . We call this particularly pernicious type of disclosure *singling out*. When  $\epsilon < H_\infty(P_X)$ ,  $P_{Y|X}$  cannot single out the value of  $X$ .

**Definition 6.11** (Singling out). Suppose  $X$  is distributed according to  $P_X$ . We say that a privacy mechanism  $P_{Y|X}$  singles out the value of  $X$  if  $\inf_{y \in \mathcal{Y}} H_\infty(P_{X|Y=y}) = 0$ .

By noting that  $X$  is an attribute of  $X$ , we obtain the following corollary of Theorem 6.7.

**Corollary 6.12.** *Suppose  $X$  is distributed according to  $P_X$ . If the privacy mechanism  $P_{Y|X}$  satisfies  $(\epsilon, P_X)$ -PML with  $\epsilon < H_\infty(P_X)$ , then it cannot single out the value of  $X$ .*

Thus, when  $P_{Y|X}$  has infinite leakage capacity,  $H_\infty(P_X)$  must be treated as a strict upper bound on  $\epsilon$ . In practice, however,  $H_\infty(P_X)$  will likely be very large and we should opt for much smaller values of  $\epsilon$ . We examine this in the example below about a query that could be answered deterministically under favorable conditions.

**Example 6.13.** Consider a database  $X = (D_1, \dots, D_n)$  containing  $n$  i.i.d entries. Suppose we want to answer the query “Are there more than  $m$  individuals in the database who identify as female?” as accurately as possible but without disclosing the gender of any individual in the database. When  $m \ll n$  or  $n - m \ll n$  it may be safe to answer this query deterministically and with no randomness at all. To see why, suppose the individuals in this population identify as female with probability  $p \in [0.3, 0.7]$ . Let  $S_i$  be a binary random variable that describes whether or not individual  $i \in [n]$  identifies as female, and note that the Markov chain  $S_i - D_i - X - Y$  holds. Let  $y = 1$  denote an affirmative answer to the query and  $y = 0$  denote a negative answer to the query.

First, suppose  $\frac{m}{n} \leq p$ . In this case, answering deterministically with  $y = 1$  causes the information leakage

$$\begin{aligned} \ell_{P_{XY}}(X \rightarrow 1) &= \log \frac{\max_{x \in \mathcal{X}} p_{Y|X=x}(1)}{p_Y(1)} \\ &= \log \frac{1}{1 - P_X(\{x : x \text{ contains less than or equal to } m \text{ females}\})} \end{aligned}$$

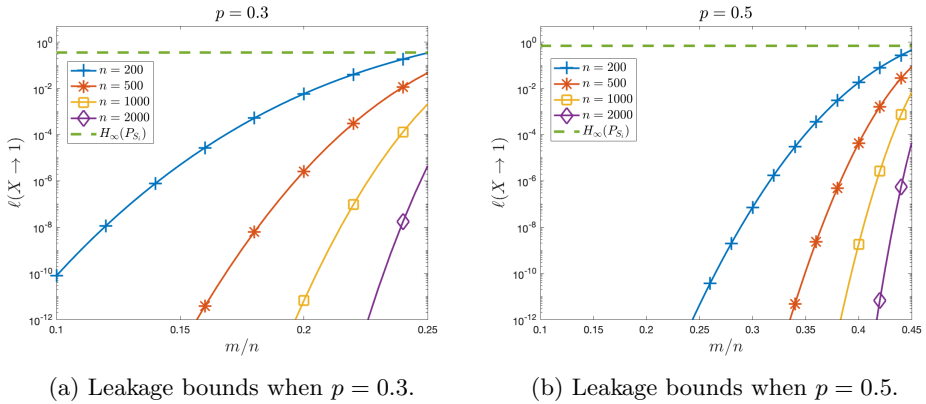


Figure 6.5.1: Upper bounds on  $\ell_{P_{XY}}(X \rightarrow 1)$  in Example 6.13 when  $p \in \{0.3, 0.5\}$  and  $n \in \{200, 500, 1000, 2000\}$ .

$$\begin{aligned}
 &= -\log \left( 1 - \sum_{k=0}^m \binom{n}{k} p^k \cdot (1-p)^{n-k} \right) \\
 &\leq -\log \left( 1 - \exp \left( -nD\left(\frac{m}{n} \parallel p\right) \right) \right),
 \end{aligned}$$

where the last inequality follows from a Chernoff bound on the tail of the Binomial distribution [57], and  $D(q \parallel r) = q \log \frac{q}{r} + (1-q) \log \frac{1-q}{1-r}$  denotes the relative entropy between two Bernoulli distributions with parameters  $q, r \in (0, 1)$ . In Figure 6.5.1, we have plotted the above upper bound on  $\ell_{P_{XY}}(X \rightarrow 1)$  for different values of  $p$  and  $n$ . It can be observed that when  $\frac{m}{n}$  is small, the amount of information leaked by the deterministic query response is several orders of magnitude smaller than  $H_\infty(P_{S_i})$ . Note that by Theorem 6.7, the gender of no individual will be disclosed by the query response as long as  $\ell_{P_{XY}}(X \rightarrow 1) < \min_{p \in [0.3, 0.7]} H_\infty(P_{S_i}) = 0.36$ . Similarly, when

$\frac{m+1}{n} \geq p$ , answering the query deterministically with  $y = 0$  causes the information leakage

$$\ell(X \rightarrow 0) \leq -\log \left( 1 - \exp \left( -nD\left(1 - \frac{m+1}{n} \parallel 1-p\right) \right) \right),$$

which is very small when  $n$  is large and  $m$  is close to  $n$ .

In conclusion,  $\epsilon$  in a PML guarantee is a data-dependent parameter that is easily interpretable in terms of the entropy of the features of  $X$  that we allow to be disclosed. This interpretability is a big advantage over many other privacy

definitions, including differential privacy, where no clear guidelines exist that explain how small the privacy parameter should be in order to maintain meaningful privacy guarantees [41].



---

# Appendices

---

## 6.A Proof of Proposition 6.8

Suppose  $U$  is a random variable taking values in the set  $\mathcal{U} = \{1, \dots, k\}$ . Fix a small  $\varepsilon > 0$  and an outcome  $y \in \mathcal{Y}$  satisfying  $H_\infty(P_{U|Y=y}) < \varepsilon$ . Then, there exists  $u \in \mathcal{U}$  with  $p_{U|Y=y}(u) > e^{-\varepsilon}$ . For simplicity, let this be  $u = 1$ . We now construct an attribute of  $X$  with entropy smaller than  $H_\infty(P_U)$  whose value is also disclosed by  $P_{Y|X}$ . Let  $W$  be a random variable with alphabet  $\mathcal{W} = \mathcal{U}$  defined by the conditional pmf

$$p_{W|U=1}(w) = \begin{cases} 1, & \text{if } w = 1, \\ 0, & \text{if } w \neq 1, \end{cases}$$

and,

$$p_{W|U=i}(w) = \begin{cases} \lambda, & \text{if } w = 1, \\ 1 - \lambda, & \text{if } w = i, \quad \text{for } i = 2, \dots, k, \\ 0, & \text{otherwise,} \end{cases}$$

where  $0 < \lambda < 1$ . Let  $P_W = P_{W|U} \circ P_U$  and  $P_{W|Y=y} = P_{W|U} \circ P_{U|Y=y}$ . Observe that  $p_W(1) = \lambda + (1 - \lambda)p_U(1)$ , and  $p_W(i) = (1 - \lambda)p_U(i)$  for  $i = 2, \dots, k$ . Thus, if

$$\lambda > \frac{\max_{u \in [k]} p_U(u) - p_U(1)}{1 - p_U(1)},$$

then  $p_W(1) > \max_{u \in [k]} p_U(u)$ , which in turn, yields  $H_\infty(P_W) < H_\infty(P_U)$ . Finally, we have

$$\begin{aligned} p_{W|Y=y}(1) &= \sum_{u \in \mathcal{U}} p_{W|U=u}(1) p_{U|Y=y}(u) \\ &\geq p_{W|U=1}(1) p_{U|Y=y}(1) \\ &> e^{-\varepsilon}, \end{aligned} \tag{6.4}$$

which implies that  $H_\infty(P_{W|Y=y}) < \varepsilon$ . Taking  $\varepsilon \rightarrow 0$ , we conclude that  $P_{Y|X}$  discloses the value of  $W$ .

## 6.B Proof of Proposition 6.10

Suppose  $C(P_{Y|X}) = \infty$ . Then, for each  $\varepsilon > 0$ , there exists  $y_\varepsilon \in \mathcal{Y}$  such that

$$\frac{\max_x p_{Y|X=x}(y_\varepsilon)}{\min_x p_{Y|X=x}(y_\varepsilon)} \geq \frac{1}{\varepsilon}.$$

Let  $\bar{x}_\varepsilon \in \arg \max_x p_{Y|X=x}(y_\varepsilon)$  and  $x_\varepsilon \in \arg \min_x p_{Y|X=x}(y_\varepsilon)$ . We have

$$\begin{aligned}
 \sup_{y \in \mathcal{Y}} \ell_{P_{XY}}(X \rightarrow y) &= \sup_{y \in \mathcal{Y}} \log \frac{\max_{x \in \mathcal{X}} p_{Y|X=x}(y)}{p_Y(y)} \\
 &\geq \sup_{\varepsilon > 0} \log \frac{p_{Y|X=\bar{x}_\varepsilon}(y_\varepsilon)}{p_{Y|X=x_\varepsilon}(y_\varepsilon) p_X(x_\varepsilon) + \sum_{x \neq x_\varepsilon} p_{Y|X=x}(y_\varepsilon) p_X(x)} \\
 &\geq \sup_{\varepsilon > 0} \log \frac{p_{Y|X=\bar{x}_\varepsilon}(y_\varepsilon)}{\varepsilon p_{Y|X=\bar{x}_\varepsilon}(y_\varepsilon) p_X(x_\varepsilon) + \sum_{x \neq x_\varepsilon} p_{Y|X=\bar{x}_\varepsilon}(y_\varepsilon) p_X(x)} \\
 &= \sup_{\varepsilon > 0} \log \frac{1}{\varepsilon p_X(x_\varepsilon) + \sum_{x \neq x_\varepsilon} p_X(x)} \\
 &\geq \log \frac{1}{1 - p_{\min}}.
 \end{aligned}$$

Therefore, no mechanism with infinite leakage capacity can satisfy  $(\varepsilon, P_X)$ -PML with  $\varepsilon < \log \frac{1}{1 - p_{\min}}$ .

To prove the second part of the statement, it suffices to construct a mechanism  $P_{Y|X}$  satisfying  $\log \frac{1}{1 - p_{\min}}$ -PML, and an attribute  $U$  of  $X$  which is disclosed by an outcome of  $P_{Y|X}$ . Consider the binary random variable  $U(X) = \mathbf{1}_{\mathcal{X} \setminus \{x_{\min}\}}(X)$  which is a deterministic function of  $X$ .

The posterior distribution  $P_{X|U}$  is given by

$$p_{X|U=0}(x) = \begin{cases} 1, & \text{if } x = x_{\min}, \\ 0, & \text{if } x \neq x_{\min}, \end{cases}$$

and

$$p_{X|U=1}(x) = \begin{cases} 0, & \text{if } x = x_{\min}, \\ \frac{p_X(x)}{1 - p_{\min}}, & \text{if } x \neq x_{\min}. \end{cases}$$

Let  $\alpha > 0$  be a small constant. Suppose  $Y$  be a binary random variable induced by the privacy mechanism  $P_{Y|X}$  defined as

$$p_{Y|X=x}(0) = \begin{cases} 0, & \text{if } x = x_{\min}, \\ \alpha, & \text{if } x \neq x_{\min}, \end{cases}$$

and,

$$p_{Y|X=x}(1) = \begin{cases} 1, & \text{if } x = x_{\min}, \\ 1 - \alpha, & \text{if } x \neq x_{\min}. \end{cases}$$

Then, we have

$$\ell(X \rightarrow 0) = \log \frac{1}{1 - p_{\min}},$$

$$\ell(X \rightarrow 1) = \log \frac{1}{1 - \alpha(1 - p_{\min})}.$$

Note that for small enough  $\alpha$  we have  $\ell(X \rightarrow 0) > \ell(X \rightarrow 1)$ . Hence,  $P_{Y|X}$  satisfies  $\log \frac{1}{1-p_{\min}}$ -PML. We now verify that  $P_{Y|X}$  discloses the value of  $U$ . Let  $P_{Y|U} = P_{Y|X} \circ P_{X|U}$ . We have

$$p_{Y|U=u}(0) = \sum_x p_{Y|X=x}(0)p_{X|U=u}(x) = \begin{cases} 0, & \text{if } u = 0, \\ \alpha, & \text{if } u = 1. \end{cases}$$

That is, if the adversary observes  $y = 0$  she will be certain that  $U$  has value  $u = 1$ . Hence, the privacy mechanism  $P_{Y|X}$  discloses the value of  $U$ , which completes the proof.



---

## 7. Application: Privacy Risk Assessment

---

The purpose of this chapter is to demonstrate how the information leakage measures discussed in the thesis can be used for *privacy risk assessment*. We apply our analysis to a supervised learning framework called *Private Aggregation of Teacher Ensembles* (PATE) [108, 109]. PATE is a general framework for privacy-preserving classification of sensitive data and operates by transferring the knowledge of an ensemble of models (called *teachers*) trained on disjoint partitions of the sensitive data to a *student* classifier. Specifically, the student is trained using a public unlabeled dataset which is labeled by the teachers through an *aggregation mechanism*. The aggregation mechanism is essentially the *report-noisy-max mechanism* [38] and adds noise to the histogram of teachers' predictions.

PATE has several desirable properties as a privacy-preserving supervised learning framework. First, the privacy guarantees result solely from the aggregation mechanism and are agnostic to the specific classification algorithms used by each teacher. Due to the modular structure of PATE, we can invoke the data-processing inequality to uncouple the information leaked through the training and aggregation, and guarantee that the overall leakage is less than both. Second, PATE lends itself well to distributed learning by allowing data owners to separately train their own predictors, hence mitigating the need for centralized storage of sensitive data. Lastly, the aggregation mechanism induces a favorable synergy between privacy and accuracy such that increased agreement among the teachers in labeling a query lowers its associated privacy cost. This synergy is one of our key focus points and will be extensively studied.

Here, our goal is to measure the amount of information leaking about any single data entry in a dataset. To this end, we consider an informed adversary, i.e., an adversary who knows the values of all the entries in the dataset, except for a single data entry of interest. Intuitively, in this setup, observations convey the information contributed by the unknown data entry since all other entries are *a priori* known. To quantify the entrywise information leakage, we use a conditional form of maximal leakage called the *pointwise conditional maximal leakage*, which is a special case of the event-conditional Sibson mutual information introduced by Liao et al. [91]. By allowing the unknown entry to be any of the entries in the dataset, we establish upper bounds on the entrywise information leakage and provide meaningful worst-case guarantees. Moreover, we use majorization theory to prove the privacy-accuracy synergy using analytical arguments. This provides deeper insights into the workings of the framework compared to the previous works [108, 109].

## 7.1 Pointwise Conditional Maximal Leakage

In this section, we introduce *pointwise conditional maximal leakage* and discuss its properties. Recall that in the randomized function view, maximal leakage is defined by assuming that an adversary wishes to guess an arbitrary discrete function of the private input data  $X$  after observing  $Y$ . Here, we assume that the adversary has some *a priori* knowledge about  $X$ . We model this knowledge as the *outcome* of a random variable  $Z$  and, accordingly, define a conditional form of maximal leakage.

Recall that we have also defined a conditional form of maximal leakage given by

$$\mathcal{L}(X \rightarrow Y \mid Z) := \log \sup_{U: U-(X,Z)-Y} \frac{\mathbb{P}(U = \hat{U}(Y, Z))}{\mathbb{P}(U = \tilde{U}(Z))}, \quad (7.1)$$

where  $\hat{U}$  is the optimal estimator of  $U$  given  $(Y, Z)$ , and  $\tilde{U}$  is optimal estimator of  $U$  given  $Z$  (Definition 3.17). It is then shown that  $\mathcal{L}(X \rightarrow Y \mid Z)$  simplifies to

$$\mathcal{L}(X \rightarrow Y \mid Z) = \log \max_{z \in \text{supp}(P_Z)} \sum_{y \in \mathcal{Y}} \max_{x \in \text{supp}(P_{X|Z=z})} P_{Y|X=x, Z=z}(y). \quad (7.2)$$

Our definition of pointwise conditional maximal leakage differs from conditional maximal leakage in that in (7.1), the conditioning is on the random variable  $Z$  itself, which translates into a maximization over the outcomes of  $Z$  in (7.2). We, on the other hand, condition the leakage directly on the outcomes of  $Z$  since we are interested in characterizing the leakage for all possible outcomes of  $Z$ , not just the worst one. As we will see later, the pointwise definition allows us to obtain a data-dependent bound on the information leakage in PATE.

**Definition 7.1** (Pointwise conditional maximal leakage). Let  $P_{XYZ}$  be a distribution on the finite set  $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ . The pointwise conditional maximal leakage from  $X$  to  $Y$  given  $Z = z$  is defined as

$$\mathcal{L}(X \rightarrow Y \mid Z = z) := \log \sup_{U: U-(X,Z)-Y} \frac{\mathbb{P}(U = \hat{U}(Y, Z = z))}{\mathbb{P}(U = \tilde{U}(Z = z))}, \quad (7.3)$$

where  $\hat{U}$  is the optimal estimator of  $U$  given  $Y$  and  $Z = z$ , and  $\tilde{U}$  is optimal estimator of  $U$  given  $Z = z$ .

Pointwise conditional maximal leakage is essentially maximal leakage evaluated with the prior distribution  $P_{X|Z=z}$  and privacy mechanism  $P_{Y|X, Z=z}$ . Thus, to obtain a simple expression for  $\mathcal{L}(X \rightarrow Y \mid Z = z)$  we can condition all the distributions in the proof of Theorem 3.16 on  $Z = z$  and get

$$\mathcal{L}(X \rightarrow Y \mid Z = z) = \log \sum_{y \in \mathcal{Y}} \max_{x \in \text{supp}(P_{X|Z=z})} P_{Y|X=x, Z=z}(y). \quad (7.4)$$

*Remark 7.2.* It is easy to see that

$$\mathcal{L}(X \rightarrow Y | Z) = \max_{z \in \text{supp}(P_Z)} \mathcal{L}(X \rightarrow Y | Z = z).$$

Moreover, if the Markov chain  $Z - X - Y$  holds, then (7.4) becomes

$$\mathcal{L}(X \rightarrow Y | Z = z) = \log \sum_{y \in \mathcal{Y}} \max_{x \in \text{supp}(P_{X|Z=z})} P_{Y|X=x}(y). \quad (7.5)$$

Now, we state two important properties of pointwise conditional maximal leakage: a data-processing inequality and a (non-adaptive) composition inequality. These properties, which follow directly from the properties of maximal leakage discussed in Section 3.4, will be used in the next section to analyze the entrywise information leakage in the PATE framework.

**Lemma 7.3.** *Pointwise conditional maximal leakage satisfies the following properties:*

(i) (Non-adaptive composition) *If the Markov chain  $Y_1 - (X, Z) - Y_2$  holds then,*

$$\mathcal{L}(X \rightarrow Y_1, Y_2 | Z = z) \leq \mathcal{L}(X \rightarrow Y_1 | Z = z) + \mathcal{L}(X \rightarrow Y_2 | Z = z).$$

*More generally, suppose the Markov chain  $Y_i - (X, Z) - Y_j$  holds for all  $i, j \in [k]$  with  $k \geq 2$ . Then,*

$$\mathcal{L}(X \rightarrow Y_1, \dots, Y_k | Z = z) \leq \mathcal{L}(X \rightarrow Y_1 | Z = z) + \dots + \mathcal{L}(X \rightarrow Y_k | Z = z).$$

(ii) (Data-processing inequality) *If the Markov chain  $(X, Z) - Y_1 - Y_2$  holds, then,*

$$\mathcal{L}(X \rightarrow Y_2 | Z = z) \leq \min \left\{ \mathcal{L}(X \rightarrow Y_1 | Z = z), \mathcal{L}(Y_1 \rightarrow Y_2 | Z = z) \right\}.$$

Next, we use pointwise conditional maximal leakage to measure the amount of information leaking about individual data entries in the PATE framework.

## 7.2 System Model

Suppose  $x^n = ((d_1, y_1), \dots, (d_n, y_n)) \in \mathcal{D}^n \times \mathcal{Y}^n$  represents the training data, where  $\mathcal{D}$  is an arbitrary but finite input space and  $\mathcal{Y} = [m]$  is the target space. The pairs  $(d_i, y_i)$  are sampled independently according to some distribution  $P$  over  $\mathcal{D} \times \mathcal{Y}$ , i.e.,  $X^n \sim P^n$ . We use the training data  $d$  to train  $L$  teachers for a classification task with  $m \geq 2$  classes in the PATE framework. Let  $(x^{(1)}, \dots, x^{(L)})$  represent a disjoint partitioning of the training set satisfying  $x^{(i)} \neq \emptyset$  for all  $i \in [L]$ ,  $\bigcup_{i=1}^L x^{(i)} = x^n$  and  $x^{(i)} \cap x^{(j)} = \emptyset$  for all  $i \neq j$ . Each partition  $x^{(i)}$  is used to train a teacher model. This results in a total of  $L$  teacher models classifying queries independently of each other.

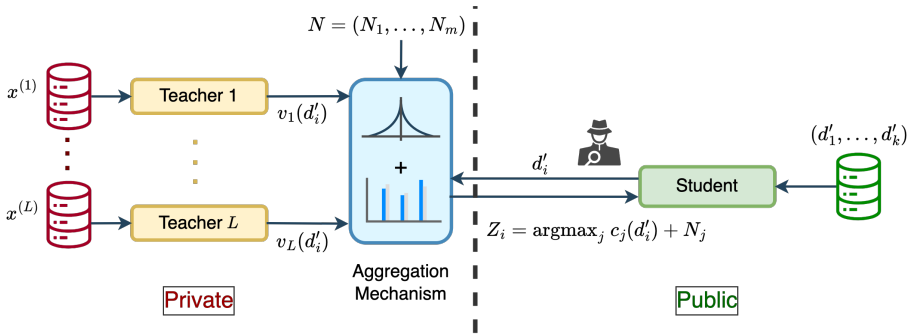


Figure 7.2.1: PATE system model [108]: Each partition of the sensitive training data is used to train a teacher. The student model is trained using a public data set labeled by the noise-perturbed predictions of the teachers. An adversary who knows all the data entries except for  $X_1$  is trying to guess some function of  $X_1$  by observing teachers’ responses to queries made by the student.

The student model is trained using a public and unlabeled dataset, which will be labeled by the teachers’ ensemble in a privacy-preserving manner. Let  $(d'_1, \dots, d'_k) \in \mathcal{D}^k$  be an independently sampled unlabeled dataset and suppose that the student queries the ensemble about the label of  $d'_i$ . Each teacher separately predicts a label for  $d'_i$ , referred to as a *vote*. Let

$$V^L(d'_i) = (V_1(d'_i), \dots, V_L(d'_i)) \in [m]^L,$$

denote the random vector of teachers’ votes and let  $C^m(d'_i) = (C_1(d'_i), \dots, C_m(d'_i))$  denote the histogram of the votes, where  $C_j(d'_i) := |\{l \in [L] : V_l(d'_i) = j\}|$  is the number of teachers who classified  $d'_i$  as belonging to class  $j \in [m]$ . Note that  $\sum_{j=1}^m C_j(d'_i) = L$  for all  $d'_i \in \mathcal{D}$ .

The aggregation mechanism in PATE is essentially the report-noisy-max mechanism [38] which adds i.i.d. noise to the bins of the votes’ histogram and returns the class label with the highest (noisy) value. Let  $\text{Lap}(b)$  denote the zero mean Laplace distribution with scale parameter  $b > 0$ . Suppose  $N^m = (N_1, \dots, N_m)$  is a sequence of i.i.d. Laplace random variables, where  $N_j \sim \text{Lap}(\frac{1}{\gamma})$  for  $j \in [m]$  represents the noise added to the  $j$ th bin. Note that  $\gamma > 0$  determines the dispersion of the noise, and thus, affects the privacy guarantees of the system. Roughly speaking, smaller values of  $\gamma$  correspond to larger noise, and in turn, stronger privacy guarantees. Finally, let  $Z_i = \arg \max_{j \in [m]} C_j(d'_i) + N_j$  be a random variable representing the predicted label for  $d'_i$  returned by the aggregation mechanism. Labeling the entire dataset  $(d'_1, \dots, d'_k)$  produces  $k$  such predictions, each of which entails a privacy cost. The system model is illustrated in Figure 7.2.1.

### 7.3 Privacy-accuracy Synergy

In this section, we assess the information leaking about each entry in the training set using pointwise conditional maximal leakage. Suppose an adversary knows the values of all the entries in the teachers' training set (i.e., the private training set) except for a single entry, say  $X_1 = (D_1, Y_1)$ . The adversary tries to guess the value of  $X_1$  or any arbitrary discrete function of  $X_1$  by observing the queries made by the student and the labels returned by the aggregation mechanism. In this setup, observations leak information only about the unknown entry  $X_1$  since the adversary already knows all the other entries. For notational simplicity and without loss of generality, we assume that  $X_1$  is in the first partition and affects the vote of the first teacher.

Next, we make the following assumptions: (i) The adversary has perfect knowledge of the algorithms used to train the teachers, and (ii) the training is done deterministically. In other words, we posit that all classification algorithms and resulting teacher models operate deterministically. The first assumption allows us to be very conservative about the capabilities of the adversary and derive guarantees that remain valid even against highly knowledgeable adversaries. Furthermore, we use the second assumption to model a scenario where the training leaks a lot of information about  $X_1$ , and the overall privacy guarantee originates only from the aggregation mechanism.

From these assumptions, it naturally follows that the adversary knows all the votes except for the vote of the teacher whose data partition includes  $X_1$ . We emphasize that our analysis considers a highly general setup in which a single data point can arbitrarily affect the vote of its teacher, resulting in observations that may be highly informative for inferring the data entry of interest (as an extreme example, consider a teacher whose vote depends only on  $X_1$ ). Note that if the adversary can already predict the last vote, then there is no information left to be leaked.

Let  $X^- = (X_2, \dots, X_n)$  be a random vector representing the portion of the training set known to the adversary, and given  $i \in [k]$ , let

$$C^-(d'_i) = (C_1^-(d'_i), \dots, C_m^-(d'_i)),$$

be a random vector representing the histogram of the votes that do not depend on  $X_1$ . Note that  $\sum_{j=1}^m C_j^-(d'_i) = L - 1$  for all  $d'_i \in \mathcal{D}$ . Suppose  $Z^k = (Z_1, \dots, Z_k)$  denotes the predicted labels for the queries  $(d'_1, \dots, d'_k)$ . Our goal is to quantify the amount of information leaking about  $X_1$  to  $Z^k$  given that the adversary knows  $X^- = x^-$ , that is,  $\mathcal{L}(X_1 \rightarrow Z^k | x^-)$ .<sup>1</sup> We can write

$$\mathcal{L}(X_1 \rightarrow Z^k | x^-) \leq \sum_{i=1}^k \mathcal{L}(X_1 \rightarrow Z_i | x^-) \tag{7.6a}$$

---

<sup>1</sup>For convenience, we write  $\mathcal{L}(X_1 \rightarrow Z^k | x^-)$  instead of  $\mathcal{L}(X_1 \rightarrow Z^k | X^- = x^-)$

$$\leq \sum_{i=1}^k \min \left\{ \mathcal{L}(X_1 \rightarrow C^m(d'_i) \mid x^-), \mathcal{L}(C^m(d'_i) \rightarrow Z_i \mid x^-) \right\}, \quad (7.6b)$$

where (7.6a) is due to the composition inequality and (7.6b) is due to the data-processing inequality for pointwise conditional maximal leakage (Lemma 7.3), and the Markov chain  $(X_1, X^-) - C^m(d'_i) - Z_i$  for all  $i \in [k]$ .

Fix  $i \in [k]$ . We will now bound the term  $\mathcal{L}(C^m(d'_i) \rightarrow Z_i \mid x^-)$ . For notational convenience, let  $Z := Z_i$ ,  $C^m := C^m(d'_i)$ , and  $C^- := C^-(d'_i)$ . Then, we write

$$\begin{aligned} & \mathcal{L}(C^m \rightarrow Z \mid x^-) \\ &= \sum_{j=1}^m \max_{c^m: P_{C^m \mid X^- = x^-}(c^m) > 0} \mathbb{P}(Z = j \mid C^m = c^m, X^- = x^-) \\ &= \log \sum_{j=1}^m \max_{c^m: P_{C^m \mid X^- = x^-, C^- = c^-(x^-)}(c^m) > 0} \mathbb{P}(Z = j \mid C^m = c^m) \end{aligned} \quad (7.7a)$$

$$\begin{aligned} &\leq \log \sum_{j=1}^m \max_{c^m: P_{C^m \mid C^- = c^-(x^-)}(c^m) > 0} \mathbb{P}(Z = j \mid C^m = c^m) \\ &= \mathcal{L}(C^m \rightarrow Z \mid c^-(x^-)) \\ &= \log \sum_{j=1}^m \max_{\substack{c^m = c^-(x^-) + \delta_{j'} \\ j' \in [m]}} \mathbb{P}(Z = j \mid C^m = c^m) \\ &= \log \sum_{j=1}^m \mathbb{P}(Z = j \mid C^m = c^-(x^-) + \delta_j), \end{aligned} \quad (7.7b)$$

where

- (7.7a) is due to the Markov chain  $X^- - C^m - Z$  and the fact that  $C^-$  is a deterministic function of  $X^-$ ; so when  $X^- = x^-$  we have  $C^- = c^-(x^-)$ ;
- $\delta_j = (0, \dots, 0, 1, 0, \dots, 0)$  denotes a sequence with all elements equal to 0, except for the  $j$ th element which equals 1;
- and (7.7b) follows by noting that the probability of outputting class  $j$  is maximized when the first teacher votes for class  $j$ .

Now, we evaluate  $\mathcal{L}(C^m \rightarrow Z \mid c^-)$  using methods from majorization theory assuming that the noise used in the aggregation mechanism has a log-concave probability density [13, 7].

**Definition 7.4** (Log-concave function). A non-negative function  $f : \mathbb{R}^m \rightarrow \mathbb{R}_+$  is said to be log-concave if it can be written as  $f(x) = \exp \phi(x)$  for some concave function  $\phi : \mathbb{R}^m \rightarrow [-\infty, \infty)$ .

Note that many commonly used probability density functions (and their corresponding CDFs) are log-concave, such as the Laplace and the Gaussian distributions [13].

**Theorem 7.5.** *If noise with log-concave probability density is used in the aggregation mechanism, i.e., the report-noisy-max mechanism, then the mapping  $c^- \mapsto \mathcal{L}(C^m \rightarrow Z \mid c^-)$  is Schur-concave. Hence, the leakage  $\mathcal{L}(C^m \rightarrow Z \mid c^-)$  is maximized by*

$$c_{max}^- := \left( \frac{L-1}{m}, \dots, \frac{L-1}{m} \right), \quad (7.8)$$

and is minimized by

$$c_{min}^- := (0, \dots, 0, L-1, 0, \dots, 0) = (L-1) \delta_j,$$

for some  $j \in [m]$ .

Theorem 7.5 is proved in Appendix 7.A.

The Schur-concavity of  $\mathcal{L}(C^m \rightarrow Z \mid c^-)$  implies that stronger consensus among teachers lowers the amount of information leaked about any individual data entry. This is a useful property of the PATE framework: increased accuracy of the teacher models results in stronger consensus in predicting the label of a query, which in turn, results in stronger privacy guarantees. It is noteworthy that the works by Papernot et al. [108, 109] arrive at similar conclusions regarding the privacy-accuracy synergy, specifically in the context of Laplace and Gaussian noise distributions. In contrast, here we have provided an analytical proof of this property that extends its applicability to the broader class of log-concave distributions.

## 7.4 Leakage Bounds

Now, we apply Theorem 7.5 to (7.7b) to get an upper bound on the leakage of the aggregation mechanism with Laplace noise. The following result is proved in Appendix 7.B.

**Proposition 7.6.** *Suppose noise with Laplace distribution is used in the aggregation mechanism. For all  $c^- \in \{0, \dots, L-1\}^m$  with  $\sum_{i=1}^m c_i^- = L-1$ , the information leaked to the output of a single query is upper bounded as  $\mathcal{L}(C^m \rightarrow Z \mid c^-) \leq \log(B_1)$ , where*

$$B_1 := (1-m)2^{-m}e^{-\gamma} + e^{\gamma} \left( 1 - \left(1 - \frac{1}{2}e^{-\gamma}\right)^m \right) + \frac{m}{2} \left(1 - \frac{1}{2}e^{-\gamma}\right)^{m-1} - \frac{m(m-1)}{4} e^{-\gamma} H(m-2),$$

and

$$H(m) := \begin{cases} \gamma & \text{if } m = 0, \\ \gamma + \sum_{k=1}^m \frac{2^{-k} - (1 - \frac{1}{2}e^{-\gamma})^k}{k} & \text{if } m \geq 1. \end{cases} \quad (7.9)$$

The bound is attained by  $c_{max}^-$  defined in (7.8).

Proposition 7.6 describes a data-independent bound that holds uniformly for all  $c^-$  (and consequently all  $x^-$ ) but depends on  $m$ , the number of classes. It can be easily verified that the bound is non-decreasing in  $m$ . Therefore, by letting  $m$  tend to infinity, we get the following simpler bound which holds for all  $x^-$  and all  $m \geq 2$ .

**Theorem 7.7.** *Suppose noise with Laplace distribution is used in the aggregation mechanism. For all  $x^- \in \mathcal{D}^{n-1} \times [m]^{n-1}$  and all  $m \geq 2$ , the information leaked about  $X_1$  as a result of labeling a single query is upper bounded by*

$$\mathcal{L}(X_1 \rightarrow Z | x^-) \leq \mathcal{L}(C^m \rightarrow Z | c^-(x^-)) \leq \gamma.$$

*Proof.* We show that

$$k(m) := \exp\left(\mathcal{L}(C^m \rightarrow Z | c_{max}^-)\right)$$

is concave in  $m$  and that

$$\lim_{m \rightarrow \infty} \exp\left(\mathcal{L}(C^m \rightarrow Z | c_{max}^-)\right) = e^\gamma. \quad (7.10)$$

Since  $m$  is an integer, we will check the second-order difference of the leakage with respect to  $m$ . The first-order difference is

$$\begin{aligned} \Delta k(m) &= k(m+1) - k(m) \\ &= \left(1 - \frac{1}{2}e^{-\gamma}\right)^m - \frac{1}{2}e^{-\gamma} \left(2^{-(m-1)} + mH(m-1)\right), \end{aligned}$$

and the second-order difference is

$$\begin{aligned} \Delta^2 k(m) &= \Delta k(m+1) - \Delta k(m) \\ &= -\frac{1}{2}e^{-\gamma}H(m) \\ &\leq 0, \end{aligned} \quad (7.11a)$$

where (7.11a) is due to the non-negativity of  $H(m)$ . Thus, the mapping  $m \mapsto \exp\left(\mathcal{L}(C^m \rightarrow Z | c_{max}^-)\right)$  is concave. It is also straightforward to verify that (7.10) holds. Hence, for all  $x^- \in \mathcal{D}^{n-1} \times [m]^{n-1}$  we have

$$\mathcal{L}(X_1 \rightarrow Z | x^-) \leq \mathcal{L}(C^m \rightarrow Z | c^-(x^-)) \leq \mathcal{L}(C^m \rightarrow Z | c_{max}^-) \leq \gamma. \quad (7.12)$$

□

The bounds stated in Proposition 7.6 and Theorem 7.7 give a more accurate characterization of the leakage as consensus among teachers decreases. This is illustrated in the following example where we calculate the leakage in (7.7b) directly using conditional probabilities, and compare it with the bounds.

**Example 7.8.** Suppose the PATE framework has been implemented with  $L = 11$  teachers to classify queries into  $m = 4$  classes. Further, suppose that for a given query  $x'_i$ , the histogram of teachers' votes is (some permutation of)  $c = (5, 3, 2, 1)$  and that Laplace noise with  $\gamma = 0.1$  is used in the aggregation mechanism. Thus, the adversary has  $c^- \in \{(4, 3, 2, 1), (5, 2, 2, 1), (5, 3, 1, 1), (5, 3, 2, 0)\}$ . We can now directly use (7.7b) to calculate the leakage of the aggregation mechanism using the probability density function of the Laplace distribution for each  $c^-$ . Then, we have one of the following four cases:

- $c^- = (4, 3, 2, 1) \implies \mathcal{L}(C^m \rightarrow Z \mid c^-) = 8.50 \times 10^{-2}$ .
- $c^- = (5, 2, 2, 1) \implies \mathcal{L}(C^m \rightarrow Z \mid c^-) = 8.40 \times 10^{-2}$ .
- $c^- = (5, 3, 1, 1) \implies \mathcal{L}(C^m \rightarrow Z \mid c^-) = 8.37 \times 10^{-2}$ .
- $c^- = (5, 3, 2, 0) \implies \mathcal{L}(C^m \rightarrow Z \mid c^-) = 8.35 \times 10^{-2}$ .

Therefore,  $\mathcal{L}(C^m \rightarrow Z \mid c^-) \leq 8.50 \times 10^{-2}$  while Proposition 7.6 gives  $\mathcal{L}(C^m \rightarrow Z \mid c^-) \leq \log(B_1) = 8.61 \times 10^{-2}$  and Theorem 7.7 gives  $\mathcal{L}(C^m \rightarrow Z \mid c^-) \leq 0.1$ . Note that due to the Schur-concavity of  $c^- \mapsto \mathcal{L}(C^m \rightarrow Z \mid c^-)$  it was expected that the information leakage would be largest for  $(4, 3, 2, 1)$ , and it would have sufficed to just consider this case.

Now, suppose  $c = (3, 3, 3, 2)$ . Calculating the leakage using the corresponding conditional probabilities gives  $\mathcal{L}(C^m \rightarrow Z \mid c^-) \leq 8.58 \times 10^{-2}$ , which is close to the value predicted by Proposition 7.6 and Theorem 7.7.

Finally, we obtain the following data-independent bound on the information leaked through multiple queries.

**Corollary 7.9.** *Consider the setting of Theorem 7.7. For all  $x^- \in \mathcal{D}^{n-1} \times [m]^{n-1}$  the information leaked about  $X_1$  as the result of training a student model on  $k$  samples is upper bounded by*

$$\mathcal{L}(X_1 \rightarrow Z^k \mid x^-) \leq k\gamma. \quad (7.13)$$

The bound is a direct consequence of Theorem 7.7 and Lemma 7.3.

Now, we derive a bound that depends on the training data through  $c^-$ . Proposition 7.10 is proved in Appendix 7.C.

**Proposition 7.10.** *Suppose noise with Laplace distribution is used in the aggregation mechanism. Assume  $c^-$  is sorted in non-increasing order and that the first*

$r$  coordinates have equal votes, that is,  $c_1^- = \dots = c_r^- > c_{r+1}^- \geq \dots \geq c_m^-$  for some  $1 \leq r \leq m$ . Then, we have

$$\mathcal{L}(C^m \rightarrow Z \mid c^-) \leq \log(B_2),$$

where

$$B_2 := r \left( 1 - \frac{2 + \gamma(c_1^- + 1 - c_2^-)}{4 \exp(\gamma(c_1^- + 1 - c_2^-))} \right) + \sum_{j=r+1}^m \frac{2 + \gamma(c_1^- - 1 - c_j^-)}{4 \exp(\gamma(c_1^- - 1 - c_j^-))}.$$

In practice, in order to calculate the information leaked through a query response, one has to take the minimum of the data-dependent bound in Proposition 7.10 and the data-independent bound in Proposition 7.6. Roughly speaking, the data-dependent bound is tighter than the data-independent bound when the teachers have strong agreement over the label of a query. This is illustrated in the numerical example below.

**Example 7.11.** Suppose the PATE framework has been implemented for a classification task with  $m = 4$  classes and that Laplace noise with  $\gamma = 0.1$  is used in the aggregation mechanism. First, consider the case where  $L = 11$  and  $c^- = (4, 3, 2, 1)$ . Then,  $\log(B_1) = 8.61 \times 10^{-2}$  while  $\log(B_2) = 6.81 \times 10^{-1}$ , so the data-independent bound is much tighter. Now, suppose  $L = 101$  and  $c^- = (90, 5, 5, 0)$ . Then,  $\log(B_2) = 1.05 \times 10^{-3}$ , while the data-independent bound remains as before. Therefore, the data-dependent bound is tighter when there is a strong consensus among teachers.

---

# Appendices

---

## 7.A Proof of Theorem 7.5

We prove that the entrywise information leakage of the aggregation mechanism is Schur-concave when the injected noise has log-concave probability density. In order to simplify the proof, we will assume that the elements of  $c^-$  (i.e., the histogram of known votes) can take non-negative real values. The results of the proof, however, will be readily applicable to histograms of non-negative integers.

Using (7.7b) we define

$$f_j(c^-) := \mathbb{P}(Z = j \mid C^m = c^- + \delta_j), \quad (7.14)$$

where  $\delta_j = (0, \dots, 0, 1, 0, \dots, 0)$  represents a single vote for class  $j$ . Then,

$$\begin{aligned} \mathcal{L}(C^m \rightarrow Z \mid c^-) &= \log \sum_{j=1}^m f_j(c^-) \\ &= \log f(c^-), \end{aligned} \quad (7.15)$$

where  $f(c^-) := \sum_{j=1}^m f_j(c^-)$ . It is clear from (7.15) that the leakage does not depend on the order of elements in  $c^-$ , thus  $c^- \mapsto \mathcal{L}(C^m \rightarrow Z \mid c^-)$  is symmetric. Moreover, according to [98, 3.B.1], the composition of an increasing function and a Schur-concave function remains Schur-concave. Since  $\log(\cdot)$  is an increasing function, to prove the Schur-concavity of the entrywise leakage we only need to verify Schur's condition for  $f(c^-)$ .

Without loss of generality assume that  $c^- = (c_1^-, \dots, c_m^-)$  is non increasingly ordered, i.e.,  $c_1^- \geq \dots \geq c_m^-$ . Let  $N^m = (N_1, \dots, N_m)$  denote the tuple of noise, where the elements are independent, identically distributed, and have log-concave probability density  $g$ . We write

$$\begin{aligned} f_j(c^-) &= \mathbb{P}(Z = j \mid C^m = c^- + \delta_j) \\ &= \mathbb{P}\left\{c_j^- + N_j + 1 > c_1^- + N_1, \dots, c_j^- + N_j + 1 > c_m^- + N_m\right\} \\ &= \int_{-\infty}^{\infty} \left[ \prod_{\substack{l=1 \\ l \neq j}}^m \mathbb{P}\{N_l < (c_j^- - c_l^- + t + 1)\} \right] g(t) dt \\ &= \int_{-\infty}^{\infty} \left[ \prod_{\substack{l=1 \\ l \neq j}}^m G(c_j^- - c_l^- + t + 1) \right] g(t) dt, \end{aligned} \quad (7.16)$$

where  $G(t) = \int_{-\infty}^t g(t') dt'$  is the cumulative distribution function of  $g$ . According to [7, Proposition 1], if  $g$  is log-concave, then  $G$  is also log-concave. We now check Schur's condition (Theorem 2.17) by writing

$$\frac{\partial f(c^-)}{\partial c_1^-} - \frac{\partial f(c^-)}{\partial c_2^-} = \sum_{j=1}^m \frac{\partial f_j(c^-)}{\partial c_1^-} - \frac{\partial f_j(c^-)}{\partial c_2^-}, \quad (7.17)$$

where we have one of the following three cases:

If  $j = 1$ , then,

$$\begin{aligned} \frac{\partial f_1(c^-)}{\partial c_1^-} - \frac{\partial f_1(c^-)}{\partial c_2^-} &= \int_{-\infty}^{\infty} \left[ \sum_{l=2}^m g(c_1^- - c_l^- + t + 1) \prod_{\substack{k=2 \\ k \neq l}}^m G(c_1^- - c_k^- + t + 1) \right] g(t) dt \\ &\quad - \int_{-\infty}^{\infty} [-g(c_1^- - c_2^- + t + 1)] \left[ \prod_{l=3}^m G(c_1^- - c_l^- + t + 1) \right] g(t) dt, \end{aligned}$$

if  $j = 2$ , then,

$$\begin{aligned} \frac{\partial f_2(c^-)}{\partial c_1^-} - \frac{\partial f_2(c^-)}{\partial c_2^-} &= \int_{-\infty}^{\infty} [-g(c_2^- - c_1^- + t + 1)] \left[ \prod_{l=3}^m G(c_2^- - c_l^- + t + 1) \right] g(t) dt \\ &\quad - \int_{-\infty}^{\infty} \left[ \sum_{\substack{l=1 \\ l \neq 2}}^m g(c_2^- - c_l^- + t + 1) \prod_{\substack{k=1 \\ k \neq 2, l}}^m G(c_2^- - c_k^- + t + 1) \right] g(t) dt, \end{aligned}$$

and if  $j \neq 1, 2$ , then,

$$\begin{aligned} \frac{\partial f_j(c^-)}{\partial c_1^-} - \frac{\partial f_j(c^-)}{\partial c_2^-} &= \int_{-\infty}^{\infty} - \left[ g(c_j^- - c_1^- + t + 1) G(c_j^- - c_2^- + t + 1) \right. \\ &\quad \left. + g(c_j^- - c_2^- + t + 1) G(c_j^- - c_1^- + t + 1) \right] \cdot \left[ \prod_{\substack{l=3 \\ l \neq j}}^m G(c_j^- - c_l^- + t + 1) \right] g(t) dt. \end{aligned}$$

Then,

$$\frac{\partial f(c^-)}{\partial c_1^-} - \frac{\partial f(c^-)}{\partial c_2^-} = A_1 - A_2 + \sum_{j=3}^m B_{(1,j)} - B_{(2,j)} + B_{(3,j)} - B_{(4,j)},$$

where

$$A_1 = 2 \int_{-\infty}^{\infty} g(c_1^- - c_2^- + t + 1) \left[ \prod_{k=3}^m G(c_1^- - c_k^- + t + 1) \right] g(t) dt,$$

$$A_2 = 2 \int_{-\infty}^{\infty} g(c_2^- - c_1^- + t + 1) \left[ \prod_{k=3}^m G(c_2^- - c_k^- + t + 1) \right] g(t) dt,$$

$$B_{(1,j)} = \int_{-\infty}^{\infty} g(c_1^- - c_j^- + t + 1) G(c_1^- - c_2^- + t + 1) \cdot \left[ \prod_{\substack{k=3 \\ k \neq j}}^m G(c_1^- - c_k^- + t + 1) \right] g(t) dt,$$

$$B_{(2,j)} = \int_{-\infty}^{\infty} g(c_j^- - c_1^- + t + 1) G(c_j^- - c_2^- + t + 1) \cdot \left[ \prod_{\substack{k=3 \\ k \neq j}}^m G(c_j^- - c_k^- + t + 1) \right] g(t) dt,$$

$$B_{(3,j)} = \int_{-\infty}^{\infty} g(c_j^- - c_2^- + t + 1) G(c_j^- - c_1^- + t + 1) \cdot \left[ \prod_{\substack{k=3 \\ k \neq j}}^m G(c_j^- - c_k^- + t + 1) \right] g(t) dt,$$

$$B_{(4,j)} = \int_{-\infty}^{\infty} g(c_2^- - c_j^- + t + 1) G(c_2^- - c_1^- + t + 1) \cdot \left[ \prod_{\substack{k=3 \\ k \neq j}}^m G(c_2^- - c_k^- + t + 1) \right] g(t) dt.$$

We now show that both  $A_1 - A_2$  and  $B_{(1,j)} - B_{(2,j)} + B_{(3,j)} - B_{(4,j)}$  are non-positive. Let us recall some properties of log-concave functions.

**Proposition 7.12** ([7, Lemma 1]). *Consider  $g : \mathbb{R} \rightarrow \mathbb{R}_+$  and suppose that  $\{x : g(x) > 0\} = (a, b)$ . Then,  $g(x)$  is log-concave if and only if for all  $a < x_1 \leq x_2 < b$  and all  $\delta \geq 0$  it holds that*

$$g(x_1 + \delta)g(x_2) \geq g(x_1)g(x_2 + \delta). \tag{7.18}$$

**Proposition 7.13** ([13, Remark 2]). *Suppose  $g : \mathbb{R} \rightarrow \mathbb{R}_+$  is a continuously differentiable function and let  $\{x : g(x) > 0\} = (a, b)$ . Then,  $g(x)$  is log-concave if and only if  $\frac{g'(x)}{g(x)}$  is a non-increasing function of  $x$  in  $(a, b)$ .*

We now prove that  $A_1 - A_2 \leq 0$ . By a change of variable in  $A_1$  we let  $c_1^- - c_2^- + t = u$ . Then,

$$A_1 - A_2 = 2 \int_{-\infty}^{\infty} \prod_{k=3}^m G(u + c_2^- - c_k^- + 1) \cdot [g(u + 1)g(u + c_2^- - c_1^-) - g(u)g(u + c_2^- - c_1^- + 1)] du.$$

We apply Proposition 7.12 to the preceding equation by noting that  $u \geq u + c_2^- - c_1^-$  (due to the non-increasing order of the elements in  $c^-$ ) which yields

$$g(u + 1)g(u + c_2^- - c_1^-) - g(u)g(u + c_2^- - c_1^- + 1) \leq 0.$$

Since  $\prod_{k=3}^m G(u + c_2^- - c_k^- + 1) \geq 0$ , we conclude that

$$A_1 - A_2 \leq 0.$$

Next, we show that  $B_{(1,j)} - B_{(2,j)} + B_{(3,j)} - B_{(4,j)} \leq 0$  for all  $j = 3, \dots, m$ . By a change of variable in  $B_{(1,j)}$  we let  $c_1^- - c_j^- + t = u$  which gives

$$B_{(1,j)} = \int_{-\infty}^{\infty} g(u + 1)G(c_j^- - c_2^- + u + 1) \cdot \left[ \prod_{\substack{k=3 \\ k \neq j}}^m G(c_j^- - c_k^- + u + 1) \right] g(c_j^- - c_1^- + u) du.$$

Similarly, by a change of variable in  $B_{(4,j)}$  we let  $c_2^- - c_j^- + t = u$  which gives

$$B_{(4,j)} = \int_{-\infty}^{\infty} g(u + 1)G(c_j^- - c_1^- + u + 1) \cdot \left[ \prod_{\substack{k=3 \\ k \neq j}}^m G(c_j^- - c_k^- + u + 1) \right] g(c_j^- - c_2^- + u) du.$$

Thus, we obtain

$$B_{(1,j)} - B_{(2,j)} + B_{(3,j)} - B_{(4,j)} = \int_{-\infty}^{\infty} \prod_{\substack{k=3 \\ k \neq j}}^m G(c_j^- - c_k^- + u + 1)$$

$$\cdot \left[ G(c_j^- - c_2^- + u + 1) (g(u + 1)g(c_j^- - c_1^- + u) - g(u)g(c_j^- - c_1^- + u + 1)) \right. \\ \left. + G(c_j^- - c_1^- + u + 1) (g(u)g(c_j^- - c_2^- + u + 1) - g(u + 1)g(c_j^- - c_2^- + u)) \right].$$

Furthermore, since  $c_j^- - c_2^- \geq c_j^- - c_1^-$  by Proposition 7.13 we have

$$\frac{g(c_j^- - c_2^- + u + 1)}{G(c_j^- - c_2^- + u + 1)} \leq \frac{g(c_j^- - c_1^- + u + 1)}{G(c_j^- - c_1^- + u + 1)},$$

or alternatively,

$$G(c_j^- - c_1^- + u + 1) \leq \frac{g(c_j^- - c_1^- + u + 1)}{g(c_j^- - c_2^- + u + 1)} G(c_j^- - c_2^- + u + 1).$$

Therefore, we get

$$B_{(1,j)} - B_{(2,j)} + B_{(3,j)} - B_{(4,j)} \leq \int_{-\infty}^{\infty} \left( \prod_{\substack{k=3 \\ k \neq j}}^m G(c_j^- - c_k^- + u + 1) \right) \cdot G(c_j^- - c_2^- + u + 1) \\ \cdot g(u + 1) \cdot \left( g(c_j^- - c_1^- + u) - g(c_j^- - c_2^- + u) \frac{g(c_j^- - c_1^- + u + 1)}{g(c_j^- - c_2^- + u + 1)} \right) \leq 0,$$

where the last equality is due to Proposition 7.12. Therefore, we have verified Schur's condition for  $f(c^-)$  and can conclude that the mapping  $c^- \mapsto \mathcal{L}(C^m \rightarrow Z \mid c^-)$  is Schur-concave. Finally, by Proposition 2.16, the entrywise leakage is maximized by

$$c_{max}^- = \left( \frac{L-1}{m}, \dots, \frac{L-1}{m} \right), \quad (7.19)$$

and is minimized by

$$c_{min}^- = (0, \dots, 0, L-1, 0, \dots, 0) = (L-1) \delta_j, \quad (7.20)$$

for some  $j \in [m]$ .

## 7.B Proof of Proposition 7.6

Let  $N^m = (N_1, \dots, N_m)$  be a sequence of i.i.d. Laplace random variables, where  $N_j \sim \text{Lap}(\frac{1}{\gamma})$  for all  $j \in [m]$ . To find an upper bound on the leakage, we apply Theorem 7.5 and calculate  $\mathcal{L}(C^m \rightarrow Z \mid c^-)$  with  $c_{max}^- = (\frac{L-1}{m}, \dots, \frac{L-1}{m})$ . We have

$$\mathcal{L}(C^m \rightarrow Z \mid c_{max}^-) = \log \sum_{j=1}^m \mathbb{P}(Z = j \mid C^m = c_{max}^- + \delta_j),$$

where

$$\begin{aligned} \mathbb{P}(Z = j \mid C^m = c_{max}^- + \delta_j) &= \mathbb{P}\{N_j + 1 > N_1, \dots, N_j + 1 > N_m\} \\ &= \int_{-\infty}^{\infty} \left[ \prod_{\substack{l=1 \\ l \neq j}}^m \mathbb{P}\{N_l < (t+1)\} \right] \cdot \frac{\gamma}{2} e^{-\gamma|t|} dt, \end{aligned}$$

and

$$\mathbb{P}\{N_l < (t+1)\} = \begin{cases} \frac{1}{2} e^{\gamma(t+1)} & t \leq -1, \\ 1 - \frac{1}{2} e^{-\gamma(t+1)} & t \geq -1. \end{cases}$$

Thus, we have

$$\begin{aligned} \mathbb{P}(Z = j \mid C^m = c_{max}^- + \delta_j) &= \underbrace{\frac{\gamma}{2} \int_{-\infty}^{-1} \left( \frac{1}{2} e^{\gamma(t+1)} \right)^{m-1} \cdot e^{\gamma t} dt}_A \\ &+ \underbrace{\frac{\gamma}{2} \int_{-1}^0 \left( 1 - \frac{1}{2} e^{-\gamma(t+1)} \right)^{m-1} \cdot e^{\gamma t} dt}_B + \underbrace{\frac{\gamma}{2} \int_0^{\infty} \left( 1 - \frac{1}{2} e^{-\gamma(t+1)} \right)^{m-1} \cdot e^{-\gamma t} dt}_C. \end{aligned}$$

It is straightforward to calculate integrals  $A$  and  $C$  as

$$A = \frac{2^{-m}}{m} e^{-\gamma} \quad \text{and} \quad C = \frac{1 - \left(1 - \frac{1}{2} e^{-\gamma}\right)^m}{m} e^{\gamma},$$

and integral  $B$  can be written as

$$B = \frac{1}{2} \left( 1 - \frac{1}{2} e^{-\gamma} \right)^{m-1} - 2^{-m} e^{-\gamma} - \frac{\gamma(m-1)}{4} e^{-\gamma} \int_{-1}^0 \left( 1 - \frac{1}{2} e^{-\gamma(t+1)} \right)^{m-2} dt.$$

We define

$$\begin{aligned} H(m) &:= \gamma \int_{-1}^0 \left( 1 - \frac{1}{2} e^{-\gamma(t+1)} \right)^m dt \\ &= \gamma \sum_{k=0}^m \binom{m}{k} \left( -\frac{1}{2} \right)^k e^{-\gamma k} \int_{-1}^0 e^{-\gamma k t} dt \\ &= \sum_{k=0}^m \binom{m}{k} \left( -\frac{1}{2} \right)^k \frac{1}{k} (1 - e^{-\gamma k}). \end{aligned}$$

Using  $\binom{m}{k} = \binom{m-1}{k} + \binom{m-1}{k-1}$  we get

$$H(m) = \sum_{k=0}^{m-1} \binom{m-1}{k} \left( -\frac{1}{2} \right)^k \frac{1}{k} (1 - e^{-\gamma k}) + \sum_{k=0}^m \binom{m-1}{k-1} \left( -\frac{1}{2} \right)^k \frac{1}{k} (1 - e^{-\gamma k})$$

$$= H(m-1) + \frac{1}{m} \left( 2^{-m} - \left(1 - \frac{1}{2}e^{-\gamma}\right)^m \right),$$

when  $m \geq 1$  and  $H(0) = \gamma$ . Thus,

$$H(m) = \begin{cases} \gamma & \text{if } m = 0, \\ \gamma + \sum_{k=1}^m \frac{2^{-k} - \left(1 - \frac{1}{2}e^{-\gamma}\right)^k}{k} & \text{if } m \geq 1. \end{cases}$$

Note that  $H(m)$  is non-negative and monotonically decreasing in  $m$ . Since  $\sum_{k=1}^{\infty} \frac{t^k}{k} = \log \frac{1}{1-t}$  for  $|t| < 1$ , we have  $\lim_{m \rightarrow \infty} H(m) = 0$ . Hence, integral  $B$  can be written as

$$B = \frac{1}{2} \left( 1 - \frac{1}{2}e^{-\gamma} \right)^{m-1} - 2^{-m}e^{-\gamma} - \frac{m-1}{4}e^{-\gamma}H(m-2).$$

Finally, we have

$$\mathcal{L}(C^m \rightarrow Z \mid c^-) \leq \mathcal{L}(C^m \rightarrow Z \mid c_{max}^-) = \log(B_1),$$

where

$$\begin{aligned} B_1 := & (1-m)2^{-m}e^{-\gamma} + e^{\gamma} \left( 1 - \left(1 - \frac{1}{2}e^{-\gamma}\right)^m \right) \\ & + \frac{m}{2} \left(1 - \frac{1}{2}e^{-\gamma}\right)^{m-1} - \frac{m(m-1)}{4}e^{-\gamma}H(m-2). \end{aligned}$$

## 7.C Proof of Proposition 7.10

Similarly to the proof of Proposition 7.6, we can write

$$\begin{aligned} \mathcal{L}(C^m \rightarrow Z \mid c^-) &= \log \sum_{j=1}^m \mathbb{P}(Z = j \mid C^m = c^- + \delta_j) \\ &= \log \sum_{j=1}^m \mathbb{P}\{N_j + c_j^- + 1 > N_1 + c_1^-, \dots, N_j + c_j^- + 1 > N_m + c_m^-\}. \end{aligned}$$

For  $1 \leq j \leq r$ , we have

$$\begin{aligned} \mathbb{P}(Z = j \mid C^m = c^- + \delta_j) &= \mathbb{P}(Z = 1 \mid C^m = c^- + \delta_1) \\ &\leq \min_{j' \neq 1} \mathbb{P}\{N_1 + c_1^- + 1 > N_{j'} + c_{j'}^-\} \\ &= \mathbb{P}\{N_1 + c_1^- + 1 > N_2 + c_2^-\} \\ &= \mathbb{P}\{N_2 - N_1 < c_1^- + 1 - c_2^-\}, \end{aligned}$$

and for  $r + 1 \leq j \leq m$ , we have

$$\begin{aligned} \mathbb{P}(Z = j \mid C^m = c^- + \delta_j) &\leq \min_{j' \neq j} \mathbb{P}\{N_j + c_j^- + 1 > N_{j'} + c_{j'}^-\} \\ &= \mathbb{P}\{N_j + c_j^- + 1 > N_1 + c_1^-\} \\ &= \mathbb{P}\{N_1 - N_j < c_j^- + 1 - c_1^-\}. \end{aligned}$$

It is straightforward to see that the random variable described as the difference of two  $\text{Lap}(\frac{1}{\gamma})$  random variables has the following CDF:

$$\mathbb{P}\{N_1 - N_2 \leq x\} = \begin{cases} \frac{1}{4} \exp(\gamma x)(2 - \gamma x) & \text{if } x \leq 0, \\ 1 - \frac{1}{4} \exp(-\gamma x)(2 + \gamma x) & \text{if } x > 0. \end{cases}$$

Then, by noting that  $c_1^- + 1 - c_2^- > 0$  and  $c_j^- + 1 - c_1^- \leq 0$  for  $r + 1 \leq j \leq m$ , we get

$$\mathcal{L}(C^m \rightarrow Z \mid c^-) \leq \log(B_2),$$

where

$$B_2 := r \left( 1 - \frac{2 + \gamma(c_1^- + 1 - c_2^-)}{4 \exp(\gamma(c_1^- + 1 - c_2^-))} \right) + \sum_{j=r+1}^m \frac{2 + \gamma(c_1^- - 1 - c_j^-)}{4 \exp(\gamma(c_1^- - 1 - c_j^-))}.$$

---

## 8. Application: Maximal Leakage-Distortion Tradeoff

---

In this chapter, we investigate the problem of privacy-preserving mechanism design. We explore the interplay between privacy and utility, examining the tradeoff between maximal leakage as the measure of privacy and the expected Hamming distortion as the measure of utility.

The privacy-utility tradeoff problem has been previously studied using different notions of privacy and different utility measures. To give a few relevant examples, LDP has been balanced against Hamming distortion [126, 143, 70], Bayes risk [3], minimax risk [31], and a class of convex utility functions [69]. Wu et al. [147] characterize the optimal maximal leakage privacy mechanisms using a class of cost functions called *staircase non decreasing* cost functions. Both maximal leakage and mutual information have been considered as privacy measures in hypothesis testing problems [88, 89]. In [90], maximal  $\alpha$ -leakage is taken as the privacy measure, and the privacy-distortion tradeoff is studied using a hard distortion measure which bounds the distortion with probability one. In [114], total variation privacy is considered and the privacy-utility tradeoff is investigated using mutual information, error probability and mean square error as the utility measures.

Perhaps the most relevant previous work to our current investigation is [70], where the privacy-distortion tradeoff is explored with LDP and Hamming distortion. More precisely, Kalantari et al. [70] consider a setup in which the prior distribution is not exactly known, but belongs to a set of possible distributions. They then divide sets of distributions into three classes. Class I sets are those sets whose convex hull includes the uniform distribution. Class II sets contain distributions that have the same order in the probabilities assigned to the elements of a given alphabet. To illustrate this, let  $\pi = (\pi_1, \pi_2, \pi_3)$  denote a probability distribution on an alphabet with three elements. Then, the sets  $\{\pi : \pi_2 \geq \pi_1 \geq \pi_3\}$  and  $\{\pi : \pi_3 \geq \pi_2 \geq \pi_1\}$  are both examples of Class II sets. Lastly, Class III sets are those sets that belong to neither of the two previous classes. The authors then study the problem of finding the smallest privacy leakage achievable for the worst distribution in a particular set of priors, subject to a bound on the expected distortion. This problem is considered separately for each class of sets of priors.

Our study is similar to [70] in that we are also interested in a robust characterization of the privacy-distortion tradeoff, albeit using a different notion of privacy. More specifically, we have the following setup: Suppose we want to

release some data containing sensitive information subject to a bound on the information leakage. Among all the privacy mechanisms that satisfy the information leakage constraint, we wish to pick the mechanism that incurs the smallest expected distortion. In this setup, we consider three different but related problems:

- Given a fixed prior distribution, what is the smallest expected distortion achievable, subject to an upper bound on the information leakage? What is the optimal privacy mechanism?
- Given a set of priors, what is the smallest expected distortion for the worst-case prior in the set, subject to an upper bound on the information leakage? What is the optimal privacy mechanism?
- Given two privacy mechanisms satisfying the information leakage constraint, and considering the set of all priors, which mechanism can produce larger distortion?

In our study of the second problem, we consider three sub-problems. First, we assume that the set of priors includes the uniform distribution. Then, we relax this assumption and assume that the set of priors includes a distribution which we will call the *least-dominant* distribution. Informally, the least-dominant distribution is the most “uniform-like” distribution in the set (we will formally define this later). Lastly, we will consider arbitrary sets of priors.

Hence, our approach differs from [70] in a few key aspects. First, we will argue that it is not necessary to consider different classes of sets of priors depending on the order of the probabilities assigned to elements in the alphabet. That is, we need not distinguish between Class II and Class III sets. Roughly speaking, this is because maximal leakage does not depend on the labels of the input and output alphabets, and therefore, we can re-label both alphabets without affecting the privacy guarantee of a mechanism. While here we use maximal leakage as the privacy measures, the same argument applies to analysis using LDP since the guarantees of LDP also remain unaffected by the re-labeling of the input/output alphabets.

Another major difference between our approach and previous works is that our objective goes beyond finding the optimal privacy mechanism and characterizing the smallest expected distortion. Here, our goal is to “order” both prior distributions and privacy mechanisms based on the utility they provide. For this, we use methods from majorization theory, which allow us to partially order vectors. In doing so, we show that, roughly speaking, priors that are more uniformly distributed incur larger expected distortion, while privacy mechanisms that distribute the privacy budget more uniformly over the symbols create smaller worst-case distortion.

## 8.1 Problem Setup

Suppose  $|\mathcal{X}| = n \geq 2$ , and  $|\mathcal{Y}| = m$ . Let  $\Delta^{(m-1)}$  denote the  $m - 1$ -dimensional probability simplex. We use  $\mathcal{M}^{n,m}$  to denote the set of all  $n \times m$  row-stochastic matrices, that is, matrices whose rows are elements from  $\Delta^{(m-1)}$ . To simplify the notation, a privacy mechanism  $P_{Y|X}$  will be represented by a row-stochastic matrix  $P = [p_{ij}] \in \mathcal{M}^{n,m}$ , where  $p_{ij} = P_{Y|X}(y_j | x_i)$  for  $i \in [n]$  and  $j \in [m]$ . Using this notation, the maximal leakage of a privacy mechanism  $P$  can be written as

$$\mathcal{L}(P) = \log \sum_{j=1}^m \max_{i \in [n]} p_{ij}, \quad (8.1)$$

that is, maximal leakage is calculated as the logarithm of the sum of the largest elements in each column of  $P$ .

In order to measure the utility of a mechanism  $P_{Y|X}$ , we calculate the expected distortion  $\mathbb{E}[d(X, Y)]$  using Hamming distortion defined as  $d(x, y) = \mathbf{1}(x \neq y)$ . We use a vector  $\pi = (\pi_1, \dots, \pi_n) \in (0, 1)^n$  to denote the distribution of  $X$  (i.e., the prior), where  $\pi_i = P_X(x_i)$  denotes the probability of  $x_i$ . We assume that  $X$  has full support, hence,  $\pi_i > 0$  for all  $i \in [n]$ . The expected Hamming distortion can be written as

$$\begin{aligned} \mathbb{E}_{P,\pi}[d(X, Y)] &= \sum_{j=1}^m \sum_{i=1}^n p_{ij} \pi_i \mathbf{1}(x_i \neq y_j) \\ &= 1 - \sum_{j=1}^m \sum_{i=1}^n p_{ij} \pi_i \mathbf{1}(x_i = y_j). \end{aligned} \quad (8.2)$$

From (8.2), it is easy to see that mechanisms with  $m > n$ , i.e., matrices with more columns than rows, cannot be optimal in terms of minimizing the expected distortion since for  $j > n$ , we have  $\mathbf{1}(x_i = y_j) = 0$  for all  $i \in [n]$ .<sup>1</sup> Hence, in the rest of the chapter, we assume that  $m = n$ . Note that this also includes the case  $m < n$  by having columns in matrix  $P$  consisting only of zeros. Therefore, the expected Hamming distortion can be written as

$$\mathbb{E}_{P,\pi}[d(X, Y)] = 1 - \sum_{j=1}^n p_{jj} \pi_j. \quad (8.3)$$

Let  $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ . We use the following notations to represent a few simple operations on vectors:

- $x_{\downarrow} = (x_{[1]}, \dots, x_{[n]})$  denotes a permutation of  $x$  that orders its components decreasingly, where  $x_{[k]}$  is the  $k$ th largest element in  $x$ . We call  $x_{\downarrow}$  the *decreasing rearrangement* of  $x$ ,

<sup>1</sup>This fact is formally proved in [70, Lemma 3] using LDP, and similar arguments can be made for our case.

- $x_{\uparrow} = (x_{(1)}, \dots, x_{(n)})$  denotes the *increasing rearrangement* of  $x$ , where  $x_{(k)}$  denotes the  $k$ th smallest element of  $x$ ,
- $\tilde{x}_k = \sum_{j=1}^k x_j$  denotes the sum of the first  $k$  elements of  $x$ ,
- $\tilde{x}_{[k]} = \sum_{j=1}^k x_{[j]}$  denotes the sum of the  $k$  largest elements of  $x$ .
- $\tilde{x}_{(k)} = \sum_{j=1}^k x_{(j)}$  denotes the sum of the  $k$  smallest elements of  $x$ ,
- $i_x = (i_x(1), \dots, i_x(n))$  denotes the sequence of indexes corresponding to the decreasing rearrangement of  $x$ , that is,  $x_{\downarrow} = (x_{[1]}, \dots, x_{[n]}) = (x_{i_x(1)}, \dots, x_{i_x(n)})$ , where  $i_x(j) \in [n]$  is the index of the  $j$ th largest element of  $x$ .

## 8.2 Known Prior Distribution

In this section, we study our first problem formulated as follows. Suppose we want to disclose the outcome of a random variable  $X$  such that the information leakage about  $X$  is below a predefined threshold, and assume that the prior distribution  $\pi$  of  $X$  is known. Among all the privacy mechanisms that satisfy the leakage constraint, we want to pick the mechanism that creates the smallest expected Hamming distortion, and therefore, provides the highest utility. Considering this setup, let

$$\mathcal{S}_{\epsilon} = \{P \in \mathcal{M}^{n,n} : \mathcal{L}(P) \leq \epsilon\}$$

be the set of all  $n \times n$  row-stochastic matrices whose maximal leakage is bounded by some  $\epsilon \leq \log n$ , where  $e^{\epsilon}$  represents our overall privacy budget. Since  $\mathcal{L}(P) \leq \log n$ , when  $\epsilon \geq \log n$ , we are allowed to fully disclose all outcomes of  $X$ , in which case  $\mathcal{S}_{\epsilon} = \mathcal{M}^{n,n}$ , i.e., the set  $\mathcal{S}_{\epsilon}$  contains all  $n \times n$  row-stochastic matrices. Thus, in the following, we assume that  $\epsilon \leq \log n$ . Our goal is to find  $D_{\min}(\epsilon, \pi)$  defined as

$$D_{\min}(\epsilon, \pi) := \inf_{P \in \mathcal{S}_{\epsilon}} \mathbb{E}_{P, \pi}[d(X, Y)] = \inf_{P \in \mathcal{S}_{\epsilon}} \left(1 - \sum_j p_{jj} \pi_j\right) = 1 - \sup_{P \in \mathcal{S}_{\epsilon}} \sum_j p_{jj} \pi_j. \quad (8.4)$$

Problem (8.4) describes a constrained convex optimization problem: The objective function is linear, and one can easily verify that the set  $\mathcal{S}_{\epsilon}$  is convex. The following result shows that the optimal mechanism for this problem fully discloses symbols with the largest prior probabilities, and suppresses symbols with the smallest prior probabilities.

**Theorem 8.1.** *Suppose  $k$  is a positive integer such that  $k \leq e^{\epsilon} \leq k + 1$  and  $k \leq n - 1$ . Then, the smallest expected distortion in problem (8.4) is*

$$D_{\min}(\epsilon, \pi) = 1 - \left(\tilde{\pi}_{[k]} + (e^{\epsilon} - k)\pi_{[k+1]}\right). \quad (8.5)$$

In addition, the optimal privacy mechanism  $P^*$  satisfies

$$\max_{i \in [n]} p_{ij}^* = p_{jj}^*, \quad (8.6)$$

for all  $j \in [n]$  (i.e., the largest element in each column is located on the diagonal), and has the following diagonal entries:

$$p_{jj}^* = \begin{cases} 1 & j = i_\pi(1), \dots, i_\pi(k), \\ e^\epsilon - k & j = i_\pi(k+1), \\ 0 & j = i_\pi(k+2), \dots, i_\pi(n). \end{cases} \quad (8.7)$$

*Proof.* Take two vectors  $x, y \in \mathbb{R}_+^n$ . We will define a partial order on  $\mathbb{R}_+^n$  induced by  $i_\pi$  as follows:

$$x \prec_{i_\pi} y \quad \text{if and only if} \quad \sum_{j=1}^l x_{i_\pi(j)} \leq \sum_{j=1}^l y_{i_\pi(j)}, \quad (8.8)$$

for all  $l = 1, \dots, n$ , where  $i_\pi(j)$  is the index of the  $j$ th largest element in  $\pi$ . Note that this partial order is very similar to the weak sub-majorization order, except the elements in the vectors  $x$  and  $y$  are ordered according to  $i_\pi$  instead of decreasingly. Now, let  $p_{\text{diag}} = (p_{11}, \dots, p_{nn})$  denote the vector of diagonal entries for matrix  $P$ . We will use our partial order induced by  $i_\pi$  on vectors with non-negative elements to define a pre-order on the matrices in  $\mathcal{S}_\epsilon$ : for  $P, Q \in \mathcal{S}_\epsilon$ , we have

$$P \prec_{i_\pi} Q \quad \text{if and only if} \quad p_{\text{diag}} \prec_{i_\pi} q_{\text{diag}}. \quad (8.9)$$

Note that  $i_\pi$  only induces a pre-order on matrices since the relation in (8.9) is reflexive and transitive but not anti-symmetric. The result stated in the theorem is then immediate by noting that:

- (i) The function  $f_\pi(P) = \sum_j p_{jj} \pi_j$  is order-preserving (i.e., increasing) with respect to the pre-order  $\prec_{i_\pi}$ , that is,

$$P \prec_{i_\pi} Q \implies f_\pi(P) \leq f_\pi(Q). \quad (8.10)$$

- (ii)  $\sum_{j=1}^n p_{jj} \leq \sum_{j=1}^n \max_i p_{ij} \leq e^\epsilon$ , for all  $P \in \mathcal{S}_\epsilon$ , with equality when  $\max_i p_{ij} = p_{jj}$  for all  $j \in [n]$  and  $\sum_j p_{jj} = e^\epsilon$ .

- (iii) A matrix  $P^*$  described by (8.6) and (8.7) satisfies  $P \prec_{i_\pi} P^*$  for all  $P \in \mathcal{S}_\epsilon$ .

□

*Remark 8.2.* Conditions (8.6) and (8.7) together imply that for  $\epsilon$  such that  $k < e^\epsilon \leq k+1$ , the optimal privacy mechanism for problem (8.4) has  $n - (k+1)$  all-zero columns. Hence, the output alphabet has support of size  $k+1$ .

*Remark 8.3.* The optimal mechanism for problem (8.4) depends on the prior only through  $i_\pi$ . We will frequently use this property in the rest of the chapter.

In Theorem 8.1, if we view  $\max_{i \in [n]} p_{ij}$  as the privacy cost of disclosing the  $j$ th symbol, then the optimal mechanism is highly opportunistic in that the privacy budget is allocated only to the most likely symbols. However, this result must be interpreted carefully. While for a fixed prior an opportunistic mechanism is optimal, we cannot conclude that, in general, privacy mechanisms that allocate the privacy budget uniformly to all symbols will generate larger distortion. In fact, we will see in Section 8.4 that when considering the set of all priors, mechanisms that distribute the privacy budget more uniformly among the symbols generate smaller worst-case distortion.

### 8.3 A Set of Possible Priors

Now, suppose the prior distribution is not known, but belongs to some set  $\Pi$  of probability distributions with support of size  $n$ . Our goal is to find a privacy mechanism in  $\mathcal{S}_\epsilon$  that minimizes the expected distortion for the worst-case prior in  $\Pi$ . Thus, the problem is changed to finding  $D_{\min}(\epsilon, \Pi)$  defined as

$$\begin{aligned} D_{\min}(\epsilon, \Pi) &:= \inf_{P \in \mathcal{S}_\epsilon} \sup_{\pi \in \Pi} \mathbb{E}_{P, \pi}[d(X, Y)] \\ &= \inf_{P \in \mathcal{S}_\epsilon} \sup_{\pi \in \Pi} \left(1 - \sum_j p_{jj} \pi_j\right) \\ &= 1 - \sup_{P \in \mathcal{S}_\epsilon} \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j. \end{aligned} \tag{8.11}$$

We study this problem by considering three sub-problems: First, we assume that the set  $\Pi$  contains the uniform distribution. Then, we relax this condition and assume that  $\Pi$  contains a *least-dominant* distribution. Informally, one can think of the least-dominant distribution as the distribution in  $\Pi$  that is more uniform than any other distribution in  $\Pi$ . Lastly, we consider an arbitrary set  $\Pi$ .

#### Sets Containing the Uniform Distribution

Suppose the set  $\Pi$  contains the uniform distribution. Note that we are not making any other assumptions about  $\Pi$  such as convexity, compactness, etc. In this case, we get the following result characterizing  $D_{\min}(\epsilon, \Pi)$ , which is proved in Appendix 8.A.

**Proposition 8.4.** *Suppose the set  $\Pi$  contains the uniform distribution denoted by  $\pi^u$ . Then, the smallest expected distortion for problem (8.11) is*

$$D_{\min}(\epsilon, \Pi) = 1 - \frac{e^\epsilon}{n}, \tag{8.12}$$

which is achieved by any privacy mechanism  $P \in \mathcal{S}_\epsilon$  satisfying  $\sum_j p_{jj} = e^\epsilon$  with  $\pi^u$  as the prior.

Proposition 8.4 suggests that the worst prior in  $\Pi$  is in fact the uniform distribution. Therefore, in the next section we will look into sets that do not necessarily contain the uniform distribution, but contain a distribution that is more uniform than any other distribution in the set.

### Sets Containing a least-dominant Distribution

Here, we relax the condition on  $\Pi$  in that we no longer require  $\Pi$  to contain the uniform distribution; we only require that  $\Pi$  contains a least-dominant distribution. For the rest of this section, let  $k \leq n - 1$  be a positive integer satisfying  $k \leq e^\epsilon \leq k + 1$ .

**Lemma 8.5.** Consider the function  $h_\epsilon : \mathbb{R}_+^n \rightarrow \mathbb{R}_+$  defined as

$$h_\epsilon(\pi) := \sup_{P \in \mathcal{S}_\epsilon} \sum_j p_{jj} \pi_j.$$

Then,  $h_\epsilon$  depends on  $\pi$  through the function  $f_\epsilon : \mathbb{R}_+^n \rightarrow \mathbb{R}_+^2$  defined as

$$f_\epsilon(\pi) = (\tilde{\pi}_{[k]}, \pi_{[k+1]}).$$

Moreover,  $h_\epsilon$  is increasing and Schur-convex in  $f_\epsilon(\pi)$ ,  $\pi \in \Pi$ . Thus, for all  $\pi, \rho \in \Pi$  such that  $f_\epsilon(\pi) \prec_w f_\epsilon(\rho)$ , we have  $h_\epsilon(\pi) \leq h_\epsilon(\rho)$ .

Lemma 8.5 (proved in Appendix 8.B) suggests that for calculating  $D_{\min}(\epsilon, \Pi)$  one needs to consider the the most uniform prior in  $\Pi$ . We formalize uniformity through the following definition.

**Definition 8.6.** Let  $f_\epsilon$  be the function defined in Lemma 8.5. We say that a distribution  $\pi^* \in \Pi$  is  $k$ -least-dominant if it satisfies  $f_\epsilon(\pi^*) \prec_w f_\epsilon(\pi)$  for all  $\pi \in \Pi$ .

**Proposition 8.7.** Assume that the set  $\Pi$  contains a  $k$ -least-dominant distribution denoted by  $\pi^*$ . Then, the smallest expected distortion for problem (8.11) is

$$D_{\min}(\epsilon, \Pi) = 1 - \left( \tilde{\pi}_{[k]}^* + (e^\epsilon - k)\pi_{[k+1]}^* \right). \quad (8.13)$$

Furthermore,  $D_{\min}(\epsilon, \Pi)$  is achieved by any stochastic matrix satisfying (8.6) and (8.7) for prior  $\pi^*$ .

Proposition 8.7 is proved in Appendix 8.C.

*Remark 8.8.* The uniform distribution over an alphabet of size  $n$  is  $k$ -least informative for all  $k \leq n - 1$ . Therefore, Proposition 8.4 can be considered as a special case of Proposition 8.7.

## General Sets

In the previous section, we considered sets of priors which contain a least-dominant distribution. One should note that, in general, a set  $\Pi$  may not contain a least-dominant distribution since  $\prec_w$  is a partial order, and not all members of  $\Pi$  may be comparable in terms of  $\prec_w$ . Therefore, in this section we present a general approach for finding  $D_{\min}(\epsilon, \Pi)$ .

Let  $\Pi_{\downarrow}$  be the set containing the decreasing rearrangement of the priors in  $\Pi$ , i.e.,  $\Pi_{\downarrow} = \{\pi_{\downarrow} : \pi \in \Pi\}$ .

**Theorem 8.9.** *For all  $\Pi \neq \emptyset$ , the smallest expected distortion  $D_{\min}(\epsilon, \Pi)$  can be obtained as the solution to the following optimization problem:*

$$D_{\min}(\epsilon, \Pi) = 1 - \inf_{\pi \in \Pi_{\downarrow}} \tilde{\pi}_k + (e^{\epsilon} - k)\pi_{k+1}. \quad (8.14)$$

Furthermore, the optimal privacy mechanism  $P \in \mathcal{S}_{\epsilon}$  for problem (8.14) satisfies (8.6) and (8.7) for  $i_{\pi} = (1, \dots, n)$ .

*Proof.* First, we argue that in order to solve problem (8.11), we can consider the set  $\Pi_{\downarrow}$  instead of  $\Pi$ . As stated in (the proof of) Lemma 8.5, the set  $\mathcal{S}_{\epsilon}$  is permutation-invariant. That is, if  $P \in \mathcal{S}_{\epsilon}$ , then  $TP \in \mathcal{S}_{\epsilon}$ , where  $T$  is some  $n \times n$  permutation matrix. From this, we conclude that without loss of generality we can order the elements of each  $\pi \in \Pi$  in some predefined way, for example, decreasingly.

Now, we show that the RHS of (8.14) both lower bounds and upper bounds the LHS.

Lower bound:

$$\begin{aligned} 1 - \sup_{P \in \mathcal{S}_{\epsilon}} \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j &= 1 - \sup_{P \in \mathcal{S}_{\epsilon}} \inf_{\pi \in \Pi_{\downarrow}} \sum_j p_{jj} \pi_j \\ &\geq 1 - \inf_{\pi \in \Pi_{\downarrow}} \sup_{P \in \mathcal{S}_{\epsilon}} \sum_j p_{jj} \pi_j \\ &= 1 - \inf_{\pi \in \Pi_{\downarrow}} \sum_{j=1}^k \pi_j + (e^{\epsilon} - k)\pi_{k+1}, \end{aligned} \quad (8.15)$$

where the last equality follows from Theorem 8.1 since  $i_{\pi} = (1, \dots, n)$  for all  $\pi \in \Pi_{\downarrow}$ .

Upper bound: Let  $P^* \in \mathcal{S}_{\epsilon}$  be some stochastic matrix satisfying (8.6) and (8.7) for  $i_{\pi} = (1, \dots, n)$ . Then, we have

$$\begin{aligned} 1 - \sup_{P \in \mathcal{S}_{\epsilon}} \inf_{\pi \in \Pi_{\downarrow}} \sum_j p_{jj} \pi_j &\leq 1 - \inf_{\pi \in \Pi_{\downarrow}} \sum_j p_{jj}^* \pi_j \\ &= 1 - \inf_{\pi \in \Pi_{\downarrow}} \sum_{j=1}^k \pi_j + (e^{\epsilon} - k)\pi_{k+1}. \end{aligned} \quad (8.16)$$

Finally, the fact that the optimal mechanism for problem (8.14) satisfies (8.6) and (8.7) for  $i_\pi = (1, \dots, n)$  follows from Remark 8.3.  $\square$

Theorem 8.9 states that for an arbitrary set  $\Pi$ , the smallest expected distortion can be calculated as the solution to an optimization problem over two variables. In fact, we can combine the first  $k$  elements of  $\pi$  into  $\tilde{\pi}_k$  for all  $\pi \in \Pi_\downarrow$ , and end up with an optimization problem with two variables over a two-dimensional set. This is of course a very convenient property: Regardless of how large  $n$  is, we only need to optimize over two variables.

In the following two numerical examples, we will illustrate the results of this section.

**Example 8.10.** Suppose we want to solve problem (8.11) with  $\epsilon = \log 2.5$  and for a set  $\Pi^{(1)}$  of distributions over an alphabet with four elements defined as  $\Pi^{(1)} = \{\pi \in \Delta^{(3)} : \pi = (0.4 - 2\delta, 0.3 + \delta, 0.15 + 0.5\delta, 0.15 + 0.5\delta), 0 \leq \delta \leq 0.1\}$ . To solve the problem, first we need to construct the set  $\Pi_\downarrow^{(1)}$ . For this, we note that for  $0 \leq \delta \leq \frac{1}{30}$  we have

$$0.4 - 2\delta \geq 0.3 + \delta > 0.15 + 0.5\delta, \quad (8.17)$$

while for  $\frac{1}{30} \leq \delta \leq 0.1$  we have

$$0.3 + \delta \geq 0.4 - 2\delta \geq 0.15 + 0.5\delta. \quad (8.18)$$

However, since  $k = 2$ , we can sum over the two largest elements of all  $\pi$  to obtain the two dimensional set  $\Pi_\downarrow^{(1)} = \{\pi \in \Delta^{(2)} : \pi = (0.7 - \delta, 0.15 + 0.5\delta, 0.15 + 0.5\delta), 0 \leq \delta \leq 0.1\}$ . Now, since the set  $\Pi_\downarrow^{(1)}$  describes a polytope, we can solve problem (8.14) as a linear program, which gives  $D_{\min}(\epsilon, \Pi^{(1)}) = 0.3$ . Note that the set  $\Pi^{(1)}$  contains a least-dominant distribution for  $k = 2$ , i.e.,  $\pi^* = (0.2, 0.4, 0.2, 0.2)$ , so we could have also used the result of Proposition 8.7 to solve the problem. An example of an optimal mechanism for this problem is:

$$P^* = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0.3 & 0.2 & 0.5 \end{bmatrix}, \quad (8.19)$$

which achieves  $D_{\min}(\epsilon, \Pi^{(1)})$  with  $\pi^*$  as the prior.

**Example 8.11.** Now, suppose  $\epsilon = \log 2.5$  and we wish to find  $D_{\min}(\epsilon, \Pi)$  for  $\Pi = \Pi^{(1)} \cup \Pi^{(2)}$ , where  $\Pi^{(1)}$  was defined in the previous example, and

$$\Pi^{(2)} = \{(0.3, 0.3, 0.1, 0.3), (0.29, 0.28, 0.29, 0.14), (0.05, 0.15, 0.4, 0.4)\}. \quad (8.20)$$

Clearly, the set  $\Pi$  does not contain a least-dominant distribution for  $k = 2$ . However, each of the two sets  $\Pi^{(1)}$  and  $\Pi^{(2)}$  do contain a least-dominant distribution

$((0.29, 0.28, 0.29, 0.14)$  is least-dominant in  $\Pi^{(2)}$ ), so it suffices to compare  $D_{\min}(\epsilon, \pi)$  for  $\pi = (0.2, 0.4, 0.2, 0.2)$  and  $\pi = (0.29, 0.28, 0.29, 0.14)$ . By doing so, we get  $D_{\min}(\epsilon, \Pi) = 0.28$  which is achieved by  $\pi = (0.29, 0.28, 0.29, 0.14)$ . An example of an optimal mechanism for this problem is:

$$P^* = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.1 & 0.5 & 0.4 & 0 \\ 0 & 0 & 1 & 0 \\ 0.3 & 0.4 & 0.3 & 0 \end{bmatrix}. \quad (8.21)$$

We conclude this section by stating an upper bound on  $D_{\min}(\epsilon, \Pi)$  that is valid for all  $\Pi \neq \emptyset$ .

*Remark 8.12.* Consider a mechanism  $Q^* \in \mathcal{S}_\epsilon$  satisfying  $\max_i q_{ij}^* = q_{jj}^* = \frac{e^\epsilon}{n}$  for all  $j \in [n]$ . Then, for all  $\Pi \neq \emptyset$ ,

$$\begin{aligned} D_{\min}(\epsilon, \Pi) &= 1 - \sup_{P \in \mathcal{S}_\epsilon} \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j \\ &\leq 1 - \inf_{\pi \in \Pi} \sum_j \frac{e^\epsilon}{n} \pi_j \\ &= 1 - \frac{e^\epsilon}{n}. \end{aligned} \quad (8.22)$$

As we showed in Proposition 8.4, this upper bound is attained when  $\Pi$  contains the uniform distribution. The intuition behind this upper bound will be made clear in the next section.

## 8.4 Ordering Privacy Mechanisms by Maximum Distortion

In this section, we will consider a slightly different problem. Suppose the set  $\Pi$  contains all prior distributions with support of size  $n$ , that is,  $\Pi$  is the relative interior of  $\Delta^{(n-1)}$ . We are given two privacy mechanisms  $P, Q \in \mathcal{S}_\epsilon$ , and we want to compare the largest distortion generated by them. Thus, we want to find  $D_{\max}(P)$  defined as

$$D_{\max}(P) := \sup_{\pi \in \Pi} (1 - \sum_j p_{jj} \pi_j) = 1 - \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j, \quad (8.23)$$

and compare it with  $D_{\max}(Q)$ .

**Theorem 8.13.** *Let  $p_{\text{diag}} = (p_{11}, \dots, p_{nn})$  denote the vector of diagonal entries for matrix  $P$ .  $D_{\max}(P)$  is Schur-convex and decreasing in the coordinates of  $p_{\text{diag}}$ . Therefore, for  $P, Q \in \mathcal{S}_\epsilon$ ,*

$$q_{\text{diag}} \prec^w p_{\text{diag}} \implies D_{\max}(Q) \leq D_{\max}(P). \quad (8.24)$$

*Proof.* We can write

$$D_{\max}(P) = 1 - \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j = 1 - \min_j p_{jj}. \quad (8.25)$$

Note that the infimum in (8.25) cannot be attained since the prior attaining it is on the boundary of  $\Delta^{(n-1)}$ . Schur-convexity of  $D_{\max}(P)$  follows from the fact that  $1 - \min_j p_{jj}$  is convex and symmetric with respect to permutations of  $p_{\text{diag}}$  [98, 3.C.2]. It is also easy to see that  $D_{\max}(P)$  is decreasing in every component of  $p_{\text{diag}}$ , while keeping the other components constant.  $\square$

Intuitively, we can view  $D_{\max}(P)$  as the capacity of mechanism  $P$  for generating distortion. Therefore, Theorem 8.13 states that mechanisms with larger and more uniform diagonal entries have a lower capacity for generating distortion. Hence, if there is high uncertainty in what the true prior is (for example, when  $\Pi$  is the relative interior of  $\Delta^{(n-1)}$ ), it is better to pick a mechanism with larger and more uniform diagonal entries to avoid large distortion.

**Corollary 8.14.** *Let  $Q^*$  be a mechanism satisfying  $\max_i q_{ij}^* = q_{jj}^* = \frac{\epsilon}{n}$  for all  $j \in [n]$ . By Theorem 8.13, we conclude that  $D_{\max}(Q^*) \leq D_{\max}(P)$  for all  $P \in \mathcal{S}_\epsilon$ . Furthermore, the expected distortion generated by  $Q^*$  does not depend on the prior distribution. Hence, for all  $\pi$ , we have*

$$\mathbb{E}_{P^*, \pi}[d(X, Y)] = 1 - \sum_j p_{jj}^* \pi_j = 1 - \frac{\epsilon}{n} \sum_j \pi_j = 1 - \frac{\epsilon}{n}. \quad (8.26)$$

The previous corollary and Remark 8.12 state that the distortion generated by mechanism  $Q^*$ , which has a uniform privacy cost over the symbols, has no dependency on the prior distribution. We conclude that  $Q^*$  is the most reliable mechanism when either there is high uncertainty in the true value of the prior (i.e., when  $\Pi$  is the set of all distributions), or when the prior is not informative (i.e., when  $\pi$  is the uniform distribution).



---

# Appendices

---

## 8.A Proof of Proposition 8.4

We prove this result by showing that the RHS of (8.12) both lower bounds and upper bounds the LHS.

Lower bound: Let  $\pi^u$  denote the uniform distribution, that is,  $\pi_1^u = \dots = \pi_n^u = \frac{1}{n}$ . Then, for all  $P \in \mathcal{S}_\epsilon$  we have

$$\begin{aligned} \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j &\leq \sum_j p_{jj} \pi_j^u = \frac{1}{n} \sum_j p_{jj} \\ &\leq \frac{1}{n} \sum_j \max_i p_{ij} \leq \frac{e^\epsilon}{n}. \end{aligned}$$

Taking the supremum of both sides we get

$$\sup_{P \in \mathcal{S}_\epsilon} \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j \leq \frac{e^\epsilon}{n},$$

and finally,

$$1 - \sup_{P \in \mathcal{S}_\epsilon} \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j \geq 1 - \frac{e^\epsilon}{n}.$$

Upper bound: Fix some  $Q \in \mathcal{S}_\epsilon$  such that

$$\max_i q_{ij} = q_{jj} = \frac{e^\epsilon}{n}$$

for all  $j \in [n]$ . Then, we can write

$$1 - \sup_{P \in \mathcal{S}_\epsilon} \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j \leq 1 - \inf_{\pi \in \Pi} \sum_j q_{jj} \pi_j = 1 - \frac{e^\epsilon}{n}.$$

Finally, we verify that any matrix  $P \in \mathcal{S}_\epsilon$  satisfying  $\sum_j p_{jj} = e^\epsilon$  achieves  $D_{\min}(\epsilon, \Pi)$  with  $\pi^u$  as the prior:

$$1 - \sum_j p_{jj} \pi_j^u = 1 - \frac{1}{n} \sum_j p_{jj} = 1 - \frac{e^\epsilon}{n}.$$

## 8.B Proof of Lemma 8.5

First, we apply the result of Theorem 8.1 to get

$$h_\epsilon(\pi) = \sup_{P \in \mathcal{S}_\epsilon} \sum_j p_{jj} \pi_j = \tilde{\pi}_{[k]} + (e^\epsilon - k) \pi_{[k+1]},$$

from which we can see that  $h_\epsilon$  depends on  $\pi$  only through  $f_\epsilon(\pi) = (\tilde{\pi}_{[k]}, \pi_{[k+1]})$ . Now, we can prove the Schur-convexity of  $h_\epsilon$  by verifying Schur's condition (Theorem 2.17). Observe that  $h_\epsilon$  is symmetric with respect to permutations of  $\pi$ . This is because maximal leakage  $\mathcal{L}(P)$  does not depend on the order of rows and columns of  $P$ . Therefore, if  $P \in \mathcal{S}_\epsilon$ , then  $TP \in \mathcal{S}_\epsilon$ , where  $T$  is some  $n \times n$  permutation matrix. In addition, since

$$\frac{\partial h_\epsilon(\pi)}{\tilde{\pi}_{[k]}} \geq \frac{\partial h_\epsilon(\pi)}{\pi_{[k+1]}} \geq 0,$$

Schur's condition is satisfied and  $h_\epsilon(\pi)$  is increasing in both  $\tilde{\pi}_{[k]}$  and  $\pi_{[k+1]}$ .

## 8.C Proof of Proposition 8.7

From (the proof of) Lemma 8.5, we know that  $\sup_{P \in \mathcal{S}_\epsilon} \sum_j p_{jj} \pi_j$  is symmetric in  $\pi$ . Therefore, it suffices to consider the decreasing rearrangement of the distributions in  $\Pi$ . That is, we assume  $\pi_1 \geq \dots \geq \pi_n$  for all  $\pi \in \Pi$ . Now, we will prove an upper bound and a lower bound on  $D_{\min}(\epsilon, \Pi)$ .

Lower bound:

$$\begin{aligned} \sup_{P \in \mathcal{S}_\epsilon} \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j &\leq \inf_{\pi \in \Pi} \sup_{P \in \mathcal{S}_\epsilon} \sum_j p_{jj} \pi_j \\ &= \sum_{j=1}^k \pi_j^* + (e^\epsilon - k) \pi_{k+1}^* \\ &= \tilde{\pi}_k^* + (e^\epsilon - k) \pi_{k+1}^*, \end{aligned} \tag{8.27a}$$

where equality (8.27a) follows from Lemma 8.5. Therefore,

$$1 - \sup_{P \in \mathcal{S}_\epsilon} \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j \geq 1 - \left( \sum_{j=1}^k \pi_j^* + (e^\epsilon - k) \pi_{k+1}^* \right).$$

Upper bound: Fix some  $P^* \in \mathcal{S}_\epsilon$  satisfying (8.6), (8.7) for  $i_\pi = (1, \dots, n)$ . Then, we can write

$$1 - \sup_{P \in \mathcal{S}_\epsilon} \inf_{\pi \in \Pi} \sum_j p_{jj} \pi_j \leq 1 - \inf_{\pi \in \Pi} \sum_j p_{jj}^* \pi_j$$

$$\begin{aligned} &= 1 - \inf_{\pi \in \Pi} \tilde{\pi}_k + (e^\epsilon - k)\pi_{k+1} \\ &= 1 - \left( \tilde{\pi}_k^* + (e^\epsilon - k)\pi_{k+1}^* \right). \end{aligned}$$



---

## 9. Conclusions and Future Directions

---

In this thesis, we introduced pointwise maximal leakage, a privacy measure inspired by the strengths of various existing frameworks, notably drawing from the principles of differential privacy and quantitative information flow. PML is applicable across a diverse range of scenarios including centralized and local settings. Furthermore, it has a strong operational meaning even when the private and public data take values in general probability spaces.

In Chapters 4 and 5, we described a theory of privacy that is inherently prior dependent. This characteristic provides the framework with an additional layer of flexibility. If assumptions about the data-generating distributions exist, then those can be exploited to achieve more favorable privacy-utility tradeoffs. In the absence of such assumptions, the compatibility of PML with well-established concepts like differential privacy, local differential privacy, and even the highly stringent definition of free-lunch privacy ensures that our privacy assessments and designs can be as conservative as needed.

In Chapter 6, we made precise the intuitive idea that correlations may lead to disclosing more information than intended. The paradigm described there using definitions of disclosure, privacy breach and utility is likely to cast inferential privacy notions in a more favorable light as well as encourage further research in this direction. Additionally, by dissecting and comparing the impossibility results of Dwork and Naor [33] and Kifer and Machanavajjhala [75] we aimed to dispel existing confusion and misunderstandings surrounding these historically impactful results. Here, an interesting direction for future research would be to investigate the consequences of strengthening the definition of disclosure to the condition where a random variable has small entropy instead of zero entropy. Another compelling avenue is to extend the definitions and results to the case where  $X$  is no longer a discrete random variable. Note that, in general, investigating disclosure prevention is crucial as these results shape our most fundamental expectations and demands from privacy-preserving data release.

In Chapter 7 we demonstrated how PML-based leakage measures can be employed for privacy risk assessment. Our study motivates a line of research where we can quantify the information leaking through various data release systems, even those designed without any formal privacy guarantees. Such an assessment helps clarify the requisites for achieving the privacy goals of each application. It may also serve as the initial step in the privatization process for the sake of compatibility with regulatory frameworks such as the GDPR [45] (see, for example, Article 35). Moreover, privacy risk assessments may help identify high-leakage operations within a system and inspire highly specialized privacy mechanisms.

We emphasize that the privacy risk assessment framework discussed here represents a much more comprehensive approach compared to methods that simply seek to verify the correctness of a differential privacy implementation, namely, *privacy auditing* [64, 136].

Throughout the thesis, most of our definitions, discussions, and results have been largely application-agnostic. Looking ahead, future efforts should be primarily focused on introducing PML into the realm of applications and promoting it as a powerful tool for practitioners. Arguably, the first step for such an undertaking would be to design effective (if not optimal) PML privacy mechanisms that are applicable in various contexts including centralized and local settings. Such mechanisms should be designed with both categorical and numerical data in mind (see [55] for a class of optimal PML mechanisms). Another pivotal direction is the comprehensive analysis of the Gaussian mechanism which is a central component in many differential privacy implementations. Other research paths include developing theoretical and numerical tools for keeping track of information leakage over many iterations of data release (e.g., privacy accounting), exploring the privacy implications of common data manipulations such as sampling and shuffling, and understanding the effects of privacy on fairness in machine learning [12, 47, 96].

Lastly, while originally designed as a privacy measure, PML can be more broadly applied as an information measure. Hence, beyond its role in privacy, PML can be applied to diverse problems such as deriving generalization bounds in adaptive data analysis [117, 59], importance sampling in stochastic gradient descent [73], or importance-aware dataset partitioning in distributed learning [129].

---

# Bibliography

---

- [1] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*, pages 308–318, 2016.
- [2] John M Abowd. The us census bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2867–2867, 2018.
- [3] Mário S Alvim, Miguel E Andrés, Konstantinos Chatzikokolakis, Pierpaolo Degano, and Catuscia Palamidessi. Differential privacy: on the trade-off between utility and information leakage. In *International Workshop on Formal Aspects in Security and Trust*, pages 39–54. Springer, 2011.
- [4] Mário S. Alvim, Kostas Chatzikokolakis, Catuscia Palamidessi, and Geoffrey Smith. Measuring information leakage using generalized gain functions. In *2012 IEEE 25th Computer Security Foundations Symposium*, pages 265–279, 2012. doi: 10.1109/CSF.2012.26.
- [5] Mário S. Alvim, Konstantinos Chatzikokolakis, Annabelle McIver, Carroll Morgan, Catuscia Palamidessi, and Geoffrey Smith. Additive and multiplicative notions of leakage, and their capacities. In *2014 IEEE 27th Computer Security Foundations Symposium*, pages 308–322, 2014. doi: 10.1109/CSF.2014.29.
- [6] Mário S Alvim, Konstantinos Chatzikokolakis, Annabelle McIver, Carroll Morgan, Catuscia Palamidessi, and Geoffrey Smith. *The Science of Quantitative Information Flow*. Springer, 2020.
- [7] Mark Yuying An. Log-concave probability distributions: Theory and statistical testing. *Duke University Dept of Economics Working Paper*, (95-03), 1997.
- [8] Shahab Asoodeh, Fady Alajaji, and Tamás Linder. Notes on information-theoretic privacy. In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1272–1278. IEEE, 2014.
- [9] Shahab Asoodeh, Fady Alajaji, and Tamás Linder. On maximal correlation, mutual information and data privacy. In *2015 IEEE 14th Canadian workshop on information theory (CWIT)*, pages 27–31. IEEE, 2015.

- [10] Shahab Asoodeh, Mario Diaz, Fady Alajaji, and Tamás Linder. Information extraction under privacy constraints. *Information*, 7(1):15, 2016.
- [11] Shahab Asoodeh, Mario Diaz, Fady Alajaji, and Tamás Linder. Estimation efficiency under privacy constraints. *IEEE Transactions on Information Theory*, 65(3):1512–1534, 2018.
- [12] Eugene Bagdasaryan, Omid Poursaeed, and Vitaly Shmatikov. Differential privacy has disparate impact on model accuracy. *Advances in neural information processing systems*, 32, 2019.
- [13] Mark Bagnoli and Ted Bergstrom. Log-concave probability and its applications. *Economic theory*, 26(2):445–469, 2005.
- [14] Borja Balle, Gilles Barthe, and Marco Gaboardi. Privacy amplification by subsampling: Tight analyses via couplings and divergences. *Advances in neural information processing systems*, 31, 2018.
- [15] Borja Balle, Gilles Barthe, and Marco Gaboardi. Privacy profiles and amplification by subsampling. *Journal of Privacy and Confidentiality*, 10(1), 2020.
- [16] Raef Bassily, Adam Groce, Jonathan Katz, and Adam Smith. Coupled-worlds privacy: Exploiting adversarial uncertainty in statistical data privacy. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 439–448. IEEE, 2013.
- [17] Raghav Bhaskar, Abhishek Bhowmick, Vipul Goyal, Srivatsan Laxman, and Abhradeep Thakurta. Noiseless database privacy. In *International Conference on the Theory and Application of Cryptology and Information Security*, pages 215–232. Springer, 2011.
- [18] Matthieu Bloch, Onur Günlü, Aylin Yener, Frédérique Oggier, H Vincent Poor, Lalitha Sankar, and Rafael F Schaefer. An overview of information-theoretic security and privacy: Metrics, limits and applications. *IEEE Journal on Selected Areas in Information Theory*, 2(1):5–22, 2021.
- [19] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. doi: 10.1017/CBO9780511804441.
- [20] Christelle Braun, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Quantitative notions of leakage for one-try attacks. *Electronic Notes in Theoretical Computer Science*, 249:75–91, 2009.
- [21] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds. In *Theory of Cryptography Conference*, pages 635–658. Springer, 2016.

- 
- [22] Flavio P Calmon and Nadia Fawaz. Privacy against statistical inference. In *2012 50th annual Allerton conference on communication, control, and computing (Allerton)*, pages 1401–1408. IEEE, 2012.
- [23] Flavio P Calmon, Ali Makhdoumi, Muriel Médard, Mayank Varia, Mark Christiansen, and Ken R Duffy. Principal inertia components and applications. *IEEE Transactions on Information Theory*, 63(8):5011–5038, 2017.
- [24] Erhan Çinlar. *Probability and stochastics*, volume 261. Springer, 2011.
- [25] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, Ltd, 2005.
- [26] Imre Csiszár. Information-type measures of difference of probability distributions and indirect observation. *studia scientiarum Mathematicarum Hungarica*, 2:229–318, 1967.
- [27] Mario Diaz, Hao Wang, Flavio P Calmon, and Lalitha Sankar. On the robustness of information-theoretic privacy measures and mechanisms. *IEEE Transactions on Information Theory*, 66(4):1949–1978, 2019.
- [28] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. Collecting telemetry data privately. *Advances in Neural Information Processing Systems*, 30, 2017.
- [29] Ni Ding, Mohammad Amin Zarrabian, and Parastoo Sadeghi.  $\alpha$ -information-theoretic privacy watchdog and optimal privatization scheme. In *2021 IEEE International Symposium on Information Theory (ISIT)*, pages 2584–2589. IEEE, 2021.
- [30] Jinshuo Dong, Aaron Roth, and Weijie J Su. Gaussian differential privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84(1):3–37, 2022.
- [31] John C Duchi, Michael I Jordan, and Martin J Wainwright. Local privacy and statistical minimax rates. In *2013 IEEE 54th Annual Symposium on Foundations of Computer Science*, pages 429–438. IEEE, 2013.
- [32] Cynthia Dwork. Differential privacy. In *Automata, Languages and Programming*, pages 1–12. Springer Berlin Heidelberg, 2006. ISBN 978-3-540-35908-1.
- [33] Cynthia Dwork and Moni Naor. On the difficulties of disclosure prevention in statistical databases or the case for differential privacy. *Journal of Privacy and Confidentiality*, 2(1), 2010.
- [34] Cynthia Dwork and Guy N Rothblum. Concentrated differential privacy. *arXiv preprint arXiv:1603.01887*, 2016.

- [35] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual international conference on the theory and applications of cryptographic techniques*, pages 486–503. Springer, 2006.
- [36] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pages 265–284. Springer, 2006.
- [37] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 51–60. IEEE, 2010.
- [38] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3-4):211–407, 2014.
- [39] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toni Pitassi, Omer Reingold, and Aaron Roth. Generalization in adaptive data analysis and holdout reuse. *Advances in Neural Information Processing Systems*, 28, 2015.
- [40] Cynthia Dwork, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Roth. Generalization in adaptive data analysis and holdout reuse. *arXiv preprint arXiv:1506.02629*, 2015.
- [41] Cynthia Dwork, Nitin Kohli, and Deirdre Mulligan. Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*, 9(2), 2019.
- [42] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 1054–1067, 2014.
- [43] Úlfar Erlingsson, Vitaly Feldman, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Abhradeep Thakurta. Amplification by shuffling: From local to central differential privacy via anonymity. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2468–2479. SIAM, 2019.
- [44] Barbara Espinoza and Geoffrey Smith. Min-entropy as a resource. *Information and Computation*, 226:57–75, 2013.
- [45] European Parliament and Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council. URL <https://data.europa.eu/eli/reg/2016/679/oj>.
- [46] Arnold M Faden. The existence of regular conditional probabilities: Necessary and sufficient conditions. *The Annals of Probability*, pages 288–298, 1985.

- 
- [47] Tom Farrand, Fatemehsadat Mireshghallah, Sahib Singh, and Andrew Trask. Neither private nor fair: Impact of data imbalance on utility and fairness in differential privacy. In *Proceedings of the 2020 workshop on privacy-preserving machine learning in practice*, pages 15–19, 2020.
- [48] Serge Fehr and Stefan Berens. On the conditional Rényi entropy. *IEEE Transactions on Information Theory*, 60(11):6801–6810, 2014.
- [49] Natasha Fernandes, Annabelle McIver, and Parastoo Sadeghi. Explaining epsilon in differential privacy through the lens of information theory. *arXiv preprint arXiv:2210.12916*, 2022.
- [50] Carrie Gates and Peter Matthews. Data is the new currency. In *Proceedings of the 2014 New Security Paradigms Workshop*, pages 105–116, 2014.
- [51] Arpita Ghosh and Robert Kleinberg. Inferential privacy guarantees for differentially private mechanisms. *arXiv preprint arXiv:1603.01508*, 2016.
- [52] Arpita Ghosh, Tim Roughgarden, and Mukund Sundararajan. Universally utility-maximizing privacy mechanisms. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 351–360, 2009.
- [53] Atefeh Gilani, Gowtham R Kurri, Oliver Kosut, and Lalitha Sankar. An alphabet of leakage measures. In *2022 IEEE Information Theory Workshop (ITW)*, pages 458–463. IEEE, 2022.
- [54] Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2):270–299, 1984. ISSN 0022-0000.
- [55] Leonhard Grosse, Sara Saeidian, and Tobias J. Oechtering. Extremal mechanisms for pointwise maximal leakage. *Submitted to: IEEE Transactions on Information Forensics and Security*, 2023. URL <https://arxiv.org/pdf/2310.07381.pdf>.
- [56] Leonhard Grosse, Sara Saeidian, Parastoo Sadeghi, Tobias J. Oechtering, and Mikael Skoglund. Quantifying privacy via information density. *To be Submitted to: 2024 IEEE International Symposium on Information Theory (ISIT)*, 2024.
- [57] Torben Hagerup and Christine Rüb. A guided tour of Chernoff bounds. *Information processing letters*, 33(6):305–308, 1990.
- [58] Xi He, Ashwin Machanava, jhala, and Bolin Ding. Blowfish privacy: Tuning privacy-utility trade-offs using policies. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*, pages 1447–1458, 2014.

- [59] Fredrik Hellström, Giuseppe Durisi, Benjamin Guedj, and Maxim Raginsky. Generalization bounds: Perspectives from information theory and PAC-Bayes. *arXiv preprint arXiv:2309.04381*, 2023.
- [60] Siu-Wai Ho and Sergio Verdú. Convexity/concavity of Rényi entropy and  $\alpha$ -mutual information. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 745–749. IEEE, 2015.
- [61] Hsiang Hsu, Shahab Asoodeh, and Flavio P. Calmon. Information-theoretic privacy watchdogs. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 552–556, 2019.
- [62] Hsiang Hsu, Natalia Martinez, Martin Bertran, Guillermo Sapiro, and Flavio P Calmon. A survey on statistical, information, and estimation—theoretic views on privacy. *IEEE BITS the Information Theory Magazine*, 1(1):45–56, 2021.
- [63] Ibrahim Issa, Aaron B Wagner, and Sudeep Kamath. An operational approach to information leakage. *IEEE Transactions on Information Theory*, 66(3):1625–1657, 2019.
- [64] Matthew Jagielski, Jonathan Ullman, and Alina Oprea. Auditing differentially private machine learning: How private is private sgd? *Advances in Neural Information Processing Systems*, 33:22205–22216, 2020.
- [65] Bo Jiang, Ming Li, and Ravi Tandon. Local information privacy and its application to privacy-preserving data aggregation. *IEEE Transactions on Dependable and Secure Computing*, 19(3):1918–1935, 2020.
- [66] Bo Jiang, Mohamed Seif, Ravi Tandon, and Ming Li. Context-aware local information privacy. *IEEE Transactions on Information Forensics and Security*, 2021.
- [67] Eduard Jorswieck and Holger Boche. *Majorization and matrix-monotone functions in wireless communications*, volume 3. Now Publishers Inc, 2007.
- [68] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. The composition theorem for differential privacy. In *International conference on machine learning*, pages 1376–1385. PMLR, 2015.
- [69] Peter Kairouz, Sewoong Oh, and Pramod Viswanath. Extremal mechanisms for local differential privacy. *The Journal of Machine Learning Research*, 17(1):492–542, 2016.
- [70] Kousha Kalantari, Lalitha Sankar, and Anand D Sarwate. Robust privacy-utility tradeoffs under differential privacy and Hamming distortion. *IEEE Transactions on Information Forensics and Security*, 13(11):2816–2830, 2018.

- 
- [71] Shiva P Kasiviswanathan and Adam Smith. On the ‘semantics’ of differential privacy: A Bayesian formulation. *Journal of Privacy and Confidentiality*, 6(1), 2014.
- [72] Shiva Prasad Kasiviswanathan, Homin K Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.
- [73] Angelos Katharopoulos and François Fleuret. Not all samples are created equal: Deep learning with importance sampling. In *International conference on machine learning*, pages 2525–2534. PMLR, 2018.
- [74] Daniel Kifer and Bing-Rong Lin. An axiomatic view of statistical privacy and utility. *Journal of Privacy and Confidentiality*, 4(1), 2012.
- [75] Daniel Kifer and Ashwin Machanavajjhala. No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*, pages 193–204, 2011.
- [76] Daniel Kifer and Ashwin Machanavajjhala. A rigorous and customizable framework for privacy. In *Proceedings of the 31st ACM SIGMOD-SIGACT-SIGAI symposium on Principles of Database Systems*, pages 77–88, 2012.
- [77] Daniel Kifer and Ashwin Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *ACM Transactions on Database Systems*, 39(1): 1–36, January 2014. ISSN 0362-5915, 1557-4644.
- [78] Daniel Kifer, John M Abowd, Robert Ashmead, Ryan Cumings-Menon, Philip Leclerc, Ashwin Machanavajjhala, William Sexton, and Pavel Zhuravlev. Bayesian and frequentist semantics for common variations of differential privacy: Applications to the 2020 census. *arXiv preprint arXiv:2209.03310*, 2022.
- [79] Achim Klenke. *Probability theory: A comprehensive course*. Springer Science & Business Media, 2013.
- [80] Antti Koskela, Joonas Jälkö, and Antti Honkela. Computing tight differential privacy guarantees using fft. In *International Conference on Artificial Intelligence and Statistics*, pages 2560–2569. PMLR, 2020.
- [81] Gowtham R Kurri, Oliver Kosut, and Lalitha Sankar. A variational formula for infinity-Rényi divergence with applications to information leakage. In *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 2493–2498. IEEE, 2022.
- [82] Gowtham R Kurri, Lalitha Sankar, and Oliver Kosut. An operational approach to information leakage via generalized gain functions. *IEEE Transactions on Information Theory*, 2023.

- [83] Jaewoo Lee and Chris Clifton. Differential identifiability. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1041–1049, 2012.
- [84] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd international conference on data engineering*, pages 106–115. IEEE, 2006.
- [85] Ninghui Li, Wahbeh Qardaji, and Dong Su. 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*, pages 32–33, 2012.
- [86] Ninghui Li, Wahbeh Qardaji, Dong Su, Yi Wu, and Weining Yang. Membership privacy: A unifying framework for privacy definitions. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security - CCS '13*, pages 889–900, Berlin, Germany, 2013. ACM Press. ISBN 978-1-4503-2477-9.
- [87] Ninghui Li, Min Lyu, Dong Su, and Weining Yang. *Differential privacy: From theory to practice*. Springer, 2017.
- [88] Jiachun Liao, Lalitha Sankar, Flavio P Calmon, and Vincent YF Tan. Hypothesis testing under maximal leakage privacy constraints. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 779–783. IEEE, 2017.
- [89] Jiachun Liao, Lalitha Sankar, Vincent YF Tan, and Flavio du Pin Calmon. Hypothesis testing under mutual information privacy constraints in the high privacy regime. *IEEE Transactions on Information Forensics and Security*, 13(4):1058–1071, 2017.
- [90] Jiachun Liao, Oliver Kosut, Lalitha Sankar, and Flavio P Calmon. Tunable measures for information leakage and applications to privacy-utility tradeoffs. *IEEE Transactions on Information Theory*, 65(12):8043–8066, 2019.
- [91] Jiachun Liao, Lalitha Sankar, Oliver Kosut, and Flavio P Calmon. Robustness of maximal  $\alpha$ -leakage to side information. In *2019 IEEE International Symposium on Information Theory (ISIT)*, pages 642–646. IEEE, 2019.
- [92] Changchang Liu, Supriyo Chakraborty, and Prateek Mittal. Dependence makes you vulnerable: Differential privacy under dependent tuples. In *NDSS*, volume 16, pages 21–24, 2016.
- [93] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramkrishnan Venkatasubramanian. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3–es, 2007.

- 
- [94] Ashwin Machanavajjhala, Daniel Kifer, John Abowd, Johannes Gehrke, and Lars Vilhuber. Privacy: Theory meets practice on the map. In *2008 IEEE 24th international conference on data engineering*, pages 277–286. IEEE, 2008.
- [95] Ali Makhdoumi, Salman Salamatian, Nadia Fawaz, and Muriel Médard. From the information bottleneck to the privacy funnel. In *2014 IEEE Information Theory Workshop (ITW 2014)*, pages 501–505. IEEE, 2014.
- [96] Karima Makhoulouf, Heber H. Arcolezi, Sami Zhioua, Ghassen Ben Brahim, and Catuscia Palamidessi. On the impact of multi-dimensional local differential privacy on fairness. *arXiv preprint arXiv:2312.04404*, 2023.
- [97] David Malone and Wayne G Sullivan. Guesswork and entropy. *IEEE Transactions on Information Theory*, 50(3):525–526, 2004.
- [98] Albert W. Marshall, Ingram Olkin, and Barry C. Arnold. *Inequalities: Theory of Majorization and Its Applications*. Springer New York, 2011.
- [99] James L Massey. Guessing and entropy. In *Proceedings of 1994 IEEE International Symposium on Information Theory*, page 204. IEEE, 1994.
- [100] Ueli M Maurer. The strong secret key rate of discrete random triples. In *Communications and Cryptography: Two Sides of One Tapestry*, pages 271–285. Springer, 1994.
- [101] Annabelle McIver, Carroll Morgan, Geoffrey Smith, Barbara Espinoza, and Larissa Meinicke. Abstract channels and their robust information-leakage ordering. In *International Conference on Principles of Security and Trust*, pages 83–102. Springer, 2014.
- [102] Frank McSherry. Differential privacy and correlated data, 2016. URL <https://github.com/frankmcsberry/blog/blob/master/posts/2016-08-29.md>.
- [103] Frank McSherry. Lunchtime for data privacy, 2016. URL <https://github.com/frankmcsberry/blog/blob/master/posts/2016-08-16.md>.
- [104] Frank McSherry and Kunal Talwar. Mechanism design via differential privacy. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)*, pages 94–103. IEEE, 2007.
- [105] Sebastian Meiser. Approximate and probabilistic differential privacy definitions. *Cryptology ePrint Archive*, 2018.
- [106] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pages 263–275. IEEE, 2017.

- [107] J. Neyman and E. S. Pearson. On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231:289–337, 1933. ISSN 02643952. URL <http://www.jstor.org/stable/91247>.
- [108] Nicolas Papernot, Martín Abadi, Ulfar Erlingsson, Ian Goodfellow, and Kunal Talwar. Semi-supervised knowledge transfer for deep learning from private training data. *ICLR*, 2017.
- [109] Nicolas Papernot, Shuang Song, Ilya Mironov, Ananth Raghunathan, Kunal Talwar, and Úlfar Erlingsson. Scalable private learning with pate. *ICLR*, 2018.
- [110] Judea Pearl. *Causality*. Cambridge University Press, 2 edition, 2009.
- [111] Balázs Pejó and Damien Desfontaines. *Guide to Differential Privacy Modifications: A Taxonomy of Variants and Extensions*. Springer Nature, 2022.
- [112] Yury Polyanskiy and Sergio Verdú. Arimoto channel coding converse and Rényi divergence. In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1327–1333. IEEE, 2010.
- [113] Yury Polyanskiy and Yihong Wu. Lecture notes on information theory. 2019.
- [114] Borzoo Rassouli and Deniz Gündüz. Optimal utility-privacy trade-off with total variation distance as a privacy measure. *IEEE Transactions on Information Forensics and Security*, 15:594–603, 2019.
- [115] Borzoo Rassouli and Deniz Gündüz. On perfect privacy. *IEEE Journal on Selected Areas in Information Theory*, 2(1):177–191, 2021.
- [116] Alfréd Rényi. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, pages 547–561. University of California Press, 1961.
- [117] Borja Rodríguez Gálvez, Germán Bassi, Ragnar Thobaben, and Mikael Skoglund. Tighter expected generalization error bounds via wasserstein distance. In *Advances in Neural Information Processing Systems*, volume 34, pages 19109–19121, 2021.
- [118] Ryan Rogers, Aaron Roth, Adam Smith, and Om Thakkar. Max-information, differential privacy, and post-selection hypothesis testing. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 487–494. IEEE, 2016.

- 
- [119] Walter Rudin. *Real and Complex Analysis*. McGraw-Hill, May 1986. ISBN 0070542341.
- [120] Parastoo Sadeghi, Ni Ding, and Thierry Rakotoarivelo. On properties and optimization of information-theoretic privacy watchdog. In *2020 IEEE Information Theory Workshop (ITW)*, pages 1–5. IEEE, 2021.
- [121] Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Optimal maximal leakage-distortion tradeoff. In *2021 IEEE Information Theory Workshop (ITW)*, pages 1–6. IEEE, 2021.
- [122] Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Quantifying membership privacy via information leakage. *IEEE Transactions on Information Forensics and Security*, 16:3096–3108, 2021.
- [123] Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Rethinking disclosure prevention with pointwise maximal leakage. *Submitted to: Journal of Privacy and Confidentiality*, 2023. URL <https://people.kth.se/~oech/JPC23.pdf>.
- [124] Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Pointwise maximal leakage. *IEEE Transactions on Information Theory*, 69(12):8054–8080, 2023.
- [125] Sara Saeidian, Giulia Cervia, Tobias J. Oechtering, and Mikael Skoglund. Pointwise maximal leakage on general alphabets. In *2023 IEEE International Symposium on Information Theory (ISIT)*, pages 388–393, 2023.
- [126] Anand D Sarwate and Lalitha Sankar. A rate-distortion perspective on local differential privacy. In *2014 52nd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 903–908, 2014.
- [127] Claude E Shannon. Communication theory of secrecy systems. *The Bell system technical journal*, 28(4):656–715, 1949.
- [128] Claude Elwood Shannon. A mathematical theory of communication. *The Bell system technical journal*, 27(3):379–423, 1948.
- [129] Sina Sheikholeslami, Amir H Payberah, Tianze Wang, Jim Dowling, and Vladimir Vlassov. The impact of importance-aware dataset partitioning on data-parallel training of deep neural networks. In *IFIP International Conference on Distributed Applications and Interoperable Systems*, pages 74–89. Springer, 2023.
- [130] Robin Sibson. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 14(2):149–160, 1969.

- [131] Geoffrey Smith. On the foundations of quantitative information flow. In *International Conference on Foundations of Software Science and Computational Structures*, pages 288–302. Springer, 2009.
- [132] Thomas Steinke. Composition of differential privacy & privacy amplification by subsampling. *arXiv preprint arXiv:2210.00597*, 2022.
- [133] Latanya Sweeney.  $k$ -anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, 10(05):557–570, 2002.
- [134] Andreia Teixeira, Armando Matos, and Luis Antunes. Conditional Rényi entropies. *IEEE Transactions on Information Theory*, 58(7):4273–4277, 2012.
- [135] Abhradeep Guha Thakurta, Andrew H Vyrros, Umesh S Vaishampayan, Gaurav Kapoor, Julien Freudiger, Vivek Rangarajan Sridhar, and Doug Davidson. Learning new words, March 14 2017. US Patent 9,594,741.
- [136] Florian Tramèr, Andreas Terzis, Thomas Steinke, Shuang Song, Matthew Jagielski, and Nicholas Carlini. Debugging differential privacy: A case study for privacy auditing. *arXiv preprint arXiv:2202.12219*, 2022.
- [137] Michael Carl Tschantz, Shayak Sen, and Anupam Datta. Sok: Differential privacy as a causal property. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 354–371, 2020. doi: 10.1109/SP40000.2020.00012.
- [138] US Census Bureau. Disclosure avoidance for the 2020 census: An introduction, 2021.
- [139] Tim van Erven and Peter Harremoës. Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.
- [140] Sergio Verdú.  $\alpha$ -mutual information. In *2015 Information Theory and Applications Workshop (ITA)*, pages 1–6. IEEE, 2015.
- [141] Isabel Wagner and David Eckhoff. Technical privacy metrics: a systematic survey. *ACM Computing Surveys (CSUR)*, 51(3):1–38, 2018.
- [142] Hao Wang, Lisa Vo, Flavio P Calmon, Muriel Médard, Ken R Duffy, and Mayank Varia. Privacy with estimation guarantees. *IEEE Transactions on Information Theory*, 65(12):8025–8042, 2019.
- [143] Weina Wang, Lei Ying, and Junshan Zhang. On the relation between identifiability, differential privacy, and mutual-information privacy. *IEEE Transactions on Information Theory*, 62(9):5018–5029, 2016.

- 
- [144] Ye Wang, Yuksel Ozan Basciftci, and Prakash Ishwar. Privacy-utility tradeoffs under constrained data release mechanisms. *arXiv preprint arXiv:1710.09295*, 2017.
- [145] Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309): 63–69, 1965.
- [146] Larry Wasserman and Shuheng Zhou. A Statistical Framework for Differential Privacy. *Journal of the American Statistical Association*, 105(489):375–389, March 2010. ISSN 0162-1459, 1537-274X.
- [147] Benjamin Wu, Aaron B Wagner, and G Edward Suh. Optimal mechanisms under maximal leakage. In *2020 IEEE Conference on Communications and Network Security (CNS)*, pages 1–6. IEEE, 2020.
- [148] Aaron D Wyner. The wire-tap channel. *Bell system technical journal*, 54(8): 1355–1387, 1975.
- [149] Bin Yang, Issei Sato, and Hiroshi Nakagawa. Bayesian differential privacy on correlated data. In *Proceedings of the 2015 ACM SIGMOD international conference on Management of Data*, pages 747–762, 2015.
- [150] Mengmeng Yang, Lingjuan Lyu, Jun Zhao, Tianqing Zhu, and Kwok-Yan Lam. Local differential privacy and its applications: A comprehensive survey. *arXiv preprint arXiv:2008.03686*, 2020.
- [151] Amirreza Zamani, Tobias J Oechtering, and Mikael Skoglund. Data disclosure with non-zero leakage and non-invertible leakage matrix. *IEEE Transactions on Information Forensics and Security*, 17:165–179, 2021.
- [152] Amirreza Zamani, Tobias J Oechtering, and Mikael Skoglund. A design framework for strongly  $\chi^2$ -private data disclosure. *IEEE Transactions on Information Forensics and Security*, 16:2312–2325, 2021.
- [153] Amirreza Zamani, Tobias J Oechtering, and Mikael Skoglund. On the privacy-utility trade-off with and without direct access to the private data. *IEEE Transactions on Information Theory*, 2023.
- [154] Mohammad Amin Zarrabian, Ni Ding, and Parastoo Sadeghi. Asymmetric local information privacy and the watchdog mechanism. In *2022 IEEE Information Theory Workshop (ITW)*, pages 7–12. IEEE, 2022.
- [155] Mohammad Amin Zarrabian, Ni Ding, and Parastoo Sadeghi. On the lift, related privacy measures, and applications to privacy–utility trade-offs. *Entropy*, 25(4):679, 2023.

- [156] Tianqing Zhu, Ping Xiong, Gang Li, and Wanlei Zhou. Correlated differential privacy: Hiding information in non-iid data set. *IEEE Transactions on Information Forensics and Security*, 10(2):229–242, 2014.
- [157] Yuqing Zhu, Jinshuo Dong, and Yu-Xiang Wang. Optimal accounting of differential privacy via characteristic function. In *International Conference on Artificial Intelligence and Statistics*, pages 4782–4817. PMLR, 2022.