Differential Disclosure of Information

Daniel Aranki
Ruzena Bajcsy

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2014-47
http://www.eecs.berkeley.edu/Pubs/TechRpts/2014/EECS-2014-47.html

May 2, 2014

Acknowledgement

This work was supported in part by TRUST, Team for Research in Ubiquitous Secure Technology, which receives support from the National Science Foundation (NSF award number CCF-0424422).
This publication was made possible by Grant Number HHS 90TR0003/01. The views expressed in this paper are solely the responsibility of the authors and do not necessarily represent the official views of the HHS.

Differential Disclosure of Information

Daniel Aranki and Ruzena Bajcsy
Department of Electrical Engineering and Computer Science
University of California, Berkeley
Berkeley, CA 94720
{daranki,bajcsy}@eecs.berkeley.edu

May 2, 2014

Abstract

Information disclosure is the process of transactions for delivering or revealing information from one party to other parties. The transaction happens between two ends, the first being the disclosing party, usually considered the owner of the information (or owner of the rights of the information). The second end of the transaction is the receiving party. In many cases, the receiving party of the transaction may include entities that are untrusted by the disclosing party (adversaries). In many cases, there is asymmetry in the knowledge between the intended recipient of the information and the adversarial entities, where the intended recipient of the information has more certain knowledge than the adversary about the sender of the information. This asymmetry can be exploited by the disclosing party to protect its privacy. In this report we present a framework of information disclosure under the assumption that adversaries exist in the receiving party such that asymmetry in knowledge between the intended recipient and the adversarial entities exist giving advantage to the intended recipient. We propose a way to disclose the information such that it can have as little utility as possible to these adversarial entities in a classification and inference settings.

This work was supported in part by TRUST, Team for Research in Ubiquitous Secure Technology, which receives support from the National Science Foundation (NSF award number CCF-0424422).

This publication was made possible by Grant Number HHS 90TR0003/01. The views expressed in this paper are solely the responsibility of the authors and do not necessarily represent the official views of the HHS.

Chapter 1

The Setting

1.1 Introduction

The act of revealing private information is a sensitive transaction that involves the disclosure of information of value to the disclosing party. The utility of the information is a function of many factors, including the entities that will get access to the information. For example, trade secrets have high utility long as they are secretly kept among the relevant trusted parties. However, if this information gets disclosed to an untrusted party (e.g. competitor), the utility of this information is drastically reduced. Therefore, information owners prefer to keep the information shared only among trusted parties in order to maintain its utility. However, in many settings, the containment of the data sharing (that is, ensuring the information only reaches trustworthy parties) is hard to ensure.

For example, in vehicle-to-vehicle communication of autonomous (or semi-autonomous) cars, cars may want disclose their GPS location by wireless communication to cars in their neighborhood. The fact that this information is being disclosed such that other cars in the neighborhood can understand it makes it accessible to untrusted entities (thereafter: adversaries) as well (by listening on the communication the same way cars would). The disclosing party of this information may consider this information private and of negative utility if grabbed by the adversary. For instance, an adversary could passively listen to all communication coming from all vehicles and track them continuously (by listening in on the communication without physically following the tracked cars). The question then becomes, how can you disclose information in such a setting, under the assumption of broadcasting, such that it only can be interpreted by cars in the same neighborhood (the ones this information was intended to be understood by).

It is important to note, that in this setting, there is asymmetry between the knowledge known to the intended recipients of the information and the unintended recipients described above. Namely, the neighborhood of the disclosing party is known to cars around it even before receiving its GPS location, but not to someone globally listening to the communication.

One can exploit such asymmetry in knowledge to protect the disclosure of information better from a privacy point of view. For example, one option is to set fixed (and globally known) reference points in the world, one reference point per neighborhood, and only disclose the relative position to the reference point of the neighborhood the car is in (instead of the global position). First, the information can still be interpreted by cars in the same vicinity as the disclosing party since they share the same neighborhood and therefore know the reference point. However, someone who is globally listening in on the communication will not as easily retrieve the global position since he/she doesn't know what reference was used to disclose that information. Furthermore, the reference points can be picked in such a way that maximize the ambiguity of the global position (or the neighborhood from which this information was sent) to someone who doesn't know the neighborhood of where the data was sent.

More generally, knowledge asymmetry between the information provider (disclosing party) and information recipient versus the adversary can be exploited to provide a more private transaction of information disclosure. A clear example of such asymmetry in knowledge is shared-key encryption in computer security, where the adversaries don't share the knowledge of the keys with the sending and receiving parties [1].

In this preliminary technical report, we propose a framework for information disclosure that takes advantage of asymmetry in knowledge between the intended recipients of information versus the adversary about the senders of information. This asymmetry is used to encode the data in a way that maximally hides the original information from an adversary. In this chapter, we present the setting of information disclosure in general, and our setting of differential disclosure of information which will include functions that map the original information into what we call differential information. In Chapter 2, we discuss the solution to finding the "best" mapping functions that satisfy this condition, and in Chapter 3 we present examples

that demonstrate the proposed framework for information disclosure.

1.2 Problem Setting

1.2.1 Definitions

Definition 1.2.1. *An* information space $\mathcal{I}$ *is a set, such that for each element* $x \in \mathcal{I}$*, there is a semantic value attached to* x*. Subsequently, we call* $x \in \mathcal{I}$ *a* piece of information *from the information space* $\mathcal{I}$*.*

For example, let the set $\mathcal{I} = \{(x_1, x_2) \in \mathbb{R}^2 | x_i \geq 0\}$ such that x_1 encodes mass (kg) and x_2 encodes height (m). Then $\mathcal{I}$ is an information space representing all possible combinations of masses and heights of people.

Definition 1.2.2. *An* information providers set $\mathcal{S}$ *over an information space* $\mathcal{I}$ *is a set, such that each element* $s \in \mathcal{S}$ *represents one and only one information provider that supplies information from the information space* $\mathcal{I}$*. Subsequently, we call* $s \in \mathcal{S}$ *an* information provider*.*

For example, the set of patients with heart-failure of a hospital who provide the hospital information regarding their mass and height (from the information space $\mathcal{I}$), defines an information providers set over the information space $\mathcal{I}$.

Definition 1.2.3. *A* provider class space *over an information providers set* $\mathcal{S}$ *is a set* Σ *and a function* $C : \mathcal{S} \to \Sigma$ *where for each information provider* $s \in \mathcal{S}$*,* $C(s) \in \Sigma$ *encodes the class membership of the information provider* s*. Subsequently, we will call* C the provider-class membership function*.*

Remark. For a provider class space $(\Sigma, C(\cdot))$ we will use the shorthand notation Σ and we will refer to the provider-class membership function $C(\cdot)$ as $C_\Sigma(\cdot)$ in case of ambiguity.

In some cases, there may be uncertainty regarding the provider-class membership function, for that we extend a probabilistic analogy to Definition 1.2.3.

Let C be a random variable representing the class, and S be a random variable representing the information provider.

Definition 1.2.4. *A* provider class belief space *over an information providers set* $\mathcal{S}$ *is a set* Σ *and a conditional probability mass function* $p(C = \sigma | S = s)$ *that encodes the likelihood (or belief) of the membership of information provider* $s \in \mathcal{S}$ *to class* $\sigma \in \Sigma$*. Subsequently, we will call* P the provider-class belief function*.*

Remark. For a provider class belief space (Σ, p) we will use the shorthand notation Σ and we will refer to the provider-class belief function p as p_Σ in case of ambiguity.

For example, for the information providers set $\mathcal{S}$, let Σ be the set $\Sigma = \{\sigma_1, \sigma_2\}$ where σ_1 = "Patients with body mass index[1] less than 25" and σ_2 = "Patients with body mass index of at least 25". The hospital can define a provider class space over $\mathcal{S}$ if it possesses the mass and height information about the information providers in $\mathcal{S}$ by using the provider-class membership function. However, an adversary who doesn't possess such information can only define a provider class belief space over $\mathcal{S}$.

Remark. Every provider class space (Σ, C) defines a trivial provider class belief space (Σ, p) such that $p(C = \sigma | S = s) = \mathbb{I}(C(s) = \sigma)$ where $\mathbb{I}(\cdot)$ is the indicator function. Therefore, unless there is a need to stress the fact that some Σ is a provider class space, we will always refer to Σ as a provider class belief space regardless of whether it is a provider class space or a provider class belief space.

[1]Body mass index (BMI) is defined as $BMI \triangleq \frac{mass(kg)}{(height(m))^2}$

Remark. Note that if a provider class belief space Σ over information provider set S is trivially defined by a provider class space, then it satisfies $\mathcal{H}_\Sigma(C|S=s)=0$ for all $s\in\mathcal{S}$ where $\mathcal{H}_\Sigma$ denotes the Shannon Entropy of C given $S=s$ using the distribution P_Σ.[2]

Definition 1.2.5. *Let S be an information provider set and let Σ_+ and Σ_- be two provider class belief spaces over S. If for each $s\in\mathcal{S}$ it is the case that $\mathcal{H}_{\Sigma_+}(C|S=s)\leq\mathcal{H}_{\Sigma_-}(C|S=s)$, we will say that Σ_+* dominates *Σ_- in provider classification knowledge. If the inequality is strict, we say that Σ_+* strongly dominates *Σ_- in provider classification knowledge. Shortly, we say Σ_+* (strongly) dominates *Σ_-. Symbolically, we use $\Sigma_+\succeq\Sigma_-$ for weak domination and $\Sigma_+\succ\Sigma_-$ for strong domination.*

1.2.2 Problem Definition

Consider an information providers set $\mathcal{S}$, an information space $\mathcal{I}$ and two provider class belief spaces Σ_{intended} and $\Sigma_{\text{adversary}}$, the provider class belief spaces of the intended recipient of the information and the adversarial party, respectively.

Assumption 1.2.1 (Certain Knowledge)**.** *The provider class belief space of the intended recipient, $\Sigma_{intended}$ is derived from a provider class space. Equivalently, $\mathcal{H}_{\Sigma_+}(C|S=s)\equiv 0$.*

Assumption 1.2.2 (Adversarial Uncertainty)**.** $\Sigma_{intended}\succ\Sigma_{adversary}$.

Assumption 1.2.1 states that the intended recipient's knowledge regarding the provider classification is certain. That is, it possesses the provider-class membership function $C(\cdot)$.

Assumption 1.2.2 states that the adversary has strictly higher level of uncertainty regarding the provider classification than the intended receiver. Specifically, it does *not* possess the provider-class membership function $C(\cdot)$. Moreover, this assumption asserts that there is asymmetry in the knowledge of the intended recipient of the information versus the adversary about the information providers, giving potential advantage to the intended recipient of the information.

Information provider $s\in\mathcal{S}$ wants to send out a piece of information (data) $x\in\mathcal{I}$ under the assumption that it can be intercepted by an adversarial party.

Note that Assumption 1.2.2 doesn't mean that the adversary can't statistically infer the class of an information provider based on the information that is provided by it. Statistical classification can still be performed. For example, using the information space $\mathcal{I}$, the information providers set $\mathcal{S}$ and the classes set Σ used as examples in Section 1.2.1, an adversary can infer the class of an information provider by the information disclosed by simply calculating the BMI from the disclosed information and figure out the class of the information provider.

Therefore, it is desired to devise an information disclosure process $\mathcal{D}$ that would allow information providers to encode the data x into $z\in I$ in such a way that would satisfy the following set of conditions.

Condition (HC)**.** *$\mathcal{D}$ should disclose as little information as possible to an adversary about the class to which the subject s belongs to, $C(s)$.*

Condition (HX)**.** *$\mathcal{D}$ should hide the original information x as much as possible from an adversary.*

Condition (DECODING)**.** *$\mathcal{D}$ should allow the intended recipient to decode the data z back to x.*

[2]The Shannon Entropy of X given Y is defined as $\mathcal{H}(X|Y=y)\triangleq\mathbb{E}\left[-\log p(X|Y=y)\right]$ where $\mathbb{E}[\cdot]$ is the expected value operator.

For that, we introduce the notion of differential disclosure of information in the following definition.

Definition 1.2.6 (DDI)**.** *Let $\mathcal{I}$ be an information space, $\mathcal{S}$ be an information providers set and Σ be a provider class space over $\mathcal{S}$. Let $\mathbf{R} : \Sigma \to \mathcal{I}^{\underline{\mathcal{I}}}$ (where $\mathcal{I}^{\underline{\mathcal{I}}}$ is the set of all injective functions $\mathcal{I} \rightarrowtail \mathcal{I}$) be a* differential information mapping function*. If $\mathcal{D}$ is an information disclosure process such that*

Sending *The transaction of disclosing a piece of information $x \in \mathcal{I}$ by an information provider $s \in \mathcal{S}$ is performed by applying the following transformation $z \leftarrow [\mathbf{R}(C_\Sigma(s))](x)$ and revealing z. Moreover, we will call z the* piece of differential information*.*

Receiving *The transaction of decoding a piece of information $z \in \mathcal{I}$ sent by an information provider $s \in \mathcal{S}$ is performed by applying $x \leftarrow [\mathbf{R}(C_\Sigma(s))]^l(x)$. Where $[\mathbf{R}(C_\Sigma(s))]^l(\cdot)$ is a left inverse of $[\mathbf{R}(C_\Sigma(s))](\cdot)$.*[3]

Then $\mathcal{D}$ is called a process of differential disclosure of information using the differential information mapping $\mathbf{R}$.

Note that the Definition 1.2.6 (DDI) takes care of the Condition (DECODING), which requires the process $\mathcal{D}$ to allow the intended recipient to decode the data. This is because the range of the differential information mapping function $\mathbf{R}$ is injective functions. Therefore, for any $s \in S$, we know that $[\mathbf{R}(s)](\cdot)$ has a left inverse which is used to decode $z \in \mathcal{I}$ by using the knowledge about the provider's class, $C_\Sigma(s)$.

The two other conditions are still to be satisfied. Namely, we need the differential information mapping function $\mathbf{R}$ to hide the original information $x \in \mathcal{I}$ as much as possible from an adversary (Condition (HX)); and to make the adversary inference of $C_\Sigma(s)$ for a provider $s \in \mathcal{S}$ based on $z \leftarrow [\mathbf{R}(C_\Sigma(s))](x)$ as hard as possible (Condition (HC)). These two requirements will be discussed in more details in Chapter 2.

Bibliography

[1] Hans Delfs and Helmut Knebl. Symmetric-key encryption. In *Introduction to Cryptography*, pages 11–31. Springer, 2007.

[3] Let $f : X \rightarrowtail Y$ be an injective function. A function $g : Y \to X$ that satisfies $g(f(x)) = x$ for all $x \in X$ is called *a left inverse* of function f.

Chapter 2

Solution

2.1 Introduction

In Chapter 1, the differential disclosure of information setting was introduced. In this setting, the key component is the differential information mapping function $\mathbf{R}$. Such mapping needs to be defined in a way that satisfies the so-far qualitative Conditions (HC) and (HX). In order to approach this problem, we need to provide quantitative measures for these conditions. In this chapter, we walk through this process in Section 2.2 and provide the solution in Section 2.3.

In this chapter we use the notation x for the pieces of information, z for the pieces of differential information, s for the information provider and σ for the provider's class. Similarly, we use the uppercase letters X for the random variable representing the pieces of information, Z for the random variable representing the pieces of differential information, S for the random variable representing the information provider and C for random variable representing the class. Also, for probability density/mass functions, we use the lowercase letter p (to distinguish from uppercase letter P that is notation for probability distribution functions).

2.2 Approach

In this section, we provide quantitative counterparts for Conditions (HC) and (HX) presented in Chapter 1. This quantification will serve us in finding a differential information mapping function with desirable qualities, better protecting the privacy of information providers.

First, let us examine Condition (HC). The statistical interpretation for the provider-class membership classification is $p(C|S = s)$, which is provided by the provider class belief space $\Sigma_{\text{adversary}}$. Using this, the condition becomes "making $p_{\Sigma_{\text{adversary}}}(C|S, Z)$ as close to $p_{\Sigma_{\text{adversary}}}(C|S)$ as possible", making the disclosure of the differential information $Z = z$ of as little value as possible to the classification problem. To see this, think about the ideal case where $p_{\Sigma_{\text{adversary}}}(C|S, Z) = p_{\Sigma_{\text{adversary}}}(C|S)$, in this case, the disclosure of $Z = z$ would be useless to provide any better classifications of the provider's class for the adversary than his prior belief described in his provider-class belief function $p_{\Sigma_{\text{adversary}}}(C|S)$.

As for Condition (HX), the inference of $X = x$ from $Z = z$ and $S = s$ can be described as $p_{\Sigma_{\text{adversary}}}(X = x|Z = z, S = s)$. Similar to the approach we took to make z as least useful as possible to the classification for the adversary, we can achieve "hiding x by disclosing z" by making $p_{\Sigma_{\text{adversary}}}(X = x|Z = z, S = s)$ as close to $p_{\Sigma_{\text{adversary}}}(X = x|S = s)$ as possible. In the ideal case that $p_{\Sigma_{\text{adversary}}}(X = x|Z = z, S = s)$ is exactly equal to $p_{\Sigma_{\text{adversary}}}(X = x|S = s)$ for all $z \in \mathcal{I}$ and $s \in \mathcal{S}$, Z becomes useless to an adversary in inferring X given S.

Now let us think about the information disclosure setting as presented in Chapter 1 as a whole. We will construct a graphical model to represent the setting of the process of differential disclosure of information. There will be a node for each random variable in our setting. Namely, S, C, X and Z. To figure out the direct dependencies, we will start with an information provider $s \in \mathcal{S}$, the class membership σ of provider s is directly dependent on s, therefore, creating a directed edge from node S to C. Let us now examine the direct dependence of the original information X. The explicit semantic relationship between the class and information provider from the provider class space implicitly defines a direct dependence between C and X. That is, once a class is observed, this relationship defines the statistical distribution of the information X. Moreover, generally each information provider (even inside the same class) has a different distribution of the information X, adding an extra direct dependence between S and X. From that, two directed edges go into the node X, one from C and another one from S. Finally, Z is intuitively directly dependent on the original information X and the class of the provider C (the relationship provided by the differential information mapping function $\mathbf{R}$). That adds two directed edges entering the node Z from X and from C. To summarize, the relationships described here can be described using the graphical model in Figure 2.2.1.

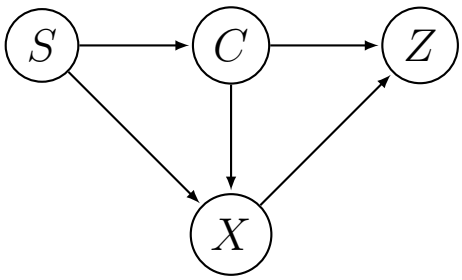


Figure 2.2.1: The graphical model describing the general process of differential disclosure of information.

To reiterate, for the model in Figure 2.2.1, the following conditional probabilities are needed. $p(S)$, the distribution of the information providers. $p(C|S)$, the provider-class belief function. $p(X|S,C)$, the distribution of the original information per provider and class. Lastly, $p(Z|X,C)$, the distribution of the differential information conditioned by the actual data and the class that the data comes from.

$p(Z|X,C)$ is actually trivially defined by the differential information mapping function $\mathbf{R}$ as follows.

$$\forall \sigma \in \Sigma, x \in \mathcal{I}, z \in \mathcal{I} : p(Z = z|X = x, C = \sigma) \triangleq \delta([\mathbf{R}(\sigma)](x) - z) \tag{2.1}$$

where $\delta(\cdot)$ is the Dirac delta function (or the Kronecker delta function for the discrete case of information).

As mentioned above, $p(C|S)$ is provided by the provider class belief space. This is where the asymmetry of knowledge lies between the intended recipient of the information versus the adversary.

Finally, $p(S)$, the distribution of information providers, and $p(X|S,C)$, the model of the information distribution per provider and class can be learned from data, or modeled based on our knowledge about the adversary and his/her model of the world.

2.3 Solution

Using the reasoning introduced in Section 2.2, we will discuss the final steps towards achieving a solution to finding a differential information mapping function $\mathbf{R}$, satisfying Conditions (HC) and (HX). A key notion in our approach is making some distribution "as close as possible" to another distribution. One measure of distance between distributions is the Kullback-Leibler (KL) divergence[1], and minimizing it is what we desire.

Remark. In this section, we will be referring to $p_{\Sigma_{\text{adversary}}}$ when we talk about probability distributions, therefore, we will omit the subscript $\Sigma_{\text{adversary}}$.

It is important to note that we have a collection of distributions that we would like to fit to each others. For instance, in order to satisfy Condition (HC), we need to make $p(C|Z = z, S = s)$ as close to $p(C|S = s)$ as possible for all $z \in \mathcal{I}$ and $s \in \mathcal{S}$ (minimizing multiple KL divergence measures, one for each combination of z and s). In the case of a continuous information space $\mathcal{I}$, there are infinite possible values for z and therefore infinite number of distributions to fit. We also note that, depending on the distribution of the information providers S, the information X and subsequently the differential information Z, some information might be more likely to be disclosed than others. For that, in the cases where an exact match between $p(C|Z = z, S = s)$

[1]Note that the Kullback-Leiber divergence is not a distance metric since it does not satisfy the symmetry condition of distance metrics.

and $p(C|S = s)$ cannot be achieved for all $z \in \mathcal{I}$ and $s \in \mathcal{S}$, we would like to put more importance in making these distributions close for the more likely pieces of information that are disclosed. One natural choice to achieve this importance weighting is to take the expected value of the KL divergence (expected with respect to to Z and S) and minimize that. That is, we would like to minimize $\mathbb{E}\left[D_{KL}\left(p(C|Z,S)||p(C|S)\right)\right]$. The minimization of the expected KL divergence also takes care of the multiple optimization objectives and makes our optimization with a single objective.

Intuitively, since the class σ of the provider s constitutes the key component by which the information x can be decoded from z, making z of the least use to infer σ, also makes it of little use to retrieve x for an adversary. Therefore, we will make it our objective to directly satisfy Condition (HC) and by that "hide x by z" too.

The expectation of the KL diveregence is over the joint probability of Z and S, namely $p(Z = z, S = s)$. We calculate this joint probability based on the model in Figure 2.2.1. For a given $z \in \mathcal{I}$ and $s \in \mathcal{S}$

$$p(Z = z, S = s) = \sum_{\sigma} \int_{x} p(X = x, S = s, C = \sigma, Z = z)dx =$$

$$\sum_{\sigma} \int_{x} p(S = s) \cdot p(C = \sigma|S = s) \cdot p(X = x|C = \sigma, S = s) \cdot p(Z = z|X = x, C = \sigma)dx =$$

$$p(S = s) \cdot \sum_{\sigma} \left[p(C = \sigma|S = s) \cdot \int_{x} p(X = x|C = \sigma, S = s) \cdot p(Z = z|X = x, C = \sigma)dx \right] \quad (\dagger)$$

Also, we want to calculate $p(C|Z, S)$ in terms of the conditional probabilities provided for the model in Figure 2.2.1. For a given $s \in \mathcal{S}$, $z \in \mathcal{I}$ and $\sigma \in \Sigma$

$$p(C = \sigma|Z = z, S = s) = \frac{p(Z = z|C = \sigma, S = s) \cdot p(C = \sigma|S = s)}{p(Z = z|S = s)} \propto$$

$$\int_{x} \left[p(Z = z, X = x|C = \sigma, S = s)\right] dx \cdot p(C = \sigma|S = s) =$$

$$\int_{x} \left[p(Z = z|X = x, C = \sigma, S = s) \cdot p(X = x|C = \sigma, S = s)\right] dx \cdot p(C = \sigma|S = s) =$$

$$\int_{x} \left[p(Z = z|X = x, C = \sigma) \cdot p(X = x|C = \sigma, S = s)\right] dx \cdot p(C = \sigma|S = s)$$

where the last step is due to the fact that the model in Figure 2.2.1 implies that Z is conditionally independent of S given X and C.

We normalize to get the conditional probability mass function

$$p(C = \sigma|Z = z, S = s) =$$
$$\frac{\int_{x} \left[p(Z = z|X = x, C = \sigma) \cdot p(X = x|C = \sigma, S = s)\right] dx \cdot p(C = \sigma|S = s)}{\sum_{\bar{\sigma}} \int_{x} \left[p(Z = z|X = x, C = \bar{\sigma}) \cdot p(X = x|C = \bar{\sigma}, S = s)\right] dx \cdot p(C = \bar{\sigma}|S = s)} \quad (\ddagger)$$

Recall that the optimization search space is, ideally, the whole space of functions $\mathcal{R} = \left(\Sigma \to \mathcal{I}^{\mathcal{I}}\right)$. In addition to that, our optimization problem is non-convex. These challenges make our problem computationally intractable. Therefore, we will choose to only search a subspace of $\mathcal{R}$ that is a parametric family of differential information mapping functions. Let $\mathcal{I}_{\mathcal{D}} \subset \mathcal{I}^{\mathcal{I}}$ be a parametric family of injective information mapping functions, then our new optimization search space becomes $\mathcal{R}_{\mathcal{D}} \triangleq (\Sigma \to \mathcal{I}_{\mathcal{D}})$. This parameterization depends on the problem in hand. For example, one subspace for information spaces of N dimensions (over the field $\mathbb{R}$) can be defined using

$$\mathcal{I}_{\mathcal{D}} \triangleq \left\{ f : \mathcal{I} \rightarrowtail \mathcal{I} | f(x) = A \cdot x - b, A \in \mathbb{R}^{N \times N} : det(A) \neq 0, b \in \mathbb{R}^{N} \right\}$$

which is the set of all injective affine functions in $\mathbb{R}^N$.

Given a parameterized optimization search subspace $\mathcal{R}_{\mathcal{D}}$ (or equivalently $\mathcal{I}_{\mathcal{D}}$), the differential information mapping function $\mathbf{R}$ becomes a parametric function too, so we use the notation $\mathbf{R}(\cdot;\Theta)$ to indicate that the function is parametric with parameter Θ.[2]

Now we formulate our complete optimization problem as follows

Listing 2.3.1: Finding $\mathcal{R}(\cdot;\Theta)$ based on the model in Figure 2.2.1 by satisfying Condition (HC)

Input: $\mathcal{I}$: Information space
Input: $\mathcal{S}$: Information providers set
Input: $\Sigma \triangleq \Sigma_{\text{intendend}}$: Provider class space
Input: $\mathcal{I}_{\mathcal{D}} \subset \mathcal{I}^{\mathcal{I}}$: Parametric search subspace
Input: $p(C|S), p(X|C,S), p(S)$: Model of the adversary
Output: $\mathbf{R} : \Sigma \to \mathcal{I}_{\mathcal{D}}$

$$\begin{aligned}
&\text{minimize} && \mathbb{E}_{p(Z,S)}\left[D_{KL}\left(p(C|Z,S)||p(C|S)\right)\right](\Theta)\\
&\text{w.r.t} && \Theta\\
&\text{s.t.} && \mathbf{R}(\cdot;\Theta) \in (\Sigma \to \mathcal{I}_{\mathcal{D}})\\
& && \forall \sigma \in \Sigma, x \in \mathcal{I}, z \in \mathcal{I} : p(Z=z|X=x, C=\sigma) = \delta([\mathbf{R}(\sigma;\Theta)](x) - z)\\
& && p(C|Z,S) = \frac{\int_x [p(Z=z|X=x,C=\sigma)\cdot p(X=x|C=\sigma,S=s)]dx \cdot p(C=\sigma|S=s)}{\sum_{\bar{\sigma}} \int_x [p(Z=z|X=x,C=\bar{\sigma})\cdot p(X=x|C=\bar{\sigma},S=s)]dx \cdot p(C=\bar{\sigma}|S=s)}\\
& && p(Z=z, S=s) = (\dagger)
\end{aligned}$$

Note that by using Equations (†) and (‡) to calculate $p(Z,S)$ and $p(C|Z,S)$, we impose our modeling structure from the model in Figure 2.2.1.

[2]Note that the optimization problem remains non-convex in general even when using a parametric subspace $\mathcal{R}_{\mathcal{D}}$ for our optimization, but it, at least, becomes tractable for sensible definitions of $\mathcal{R}_{\mathcal{D}}$.

Chapter 3

Examples

3.1 Introduction

In this chapter, we present a series of examples of differential disclosure of information. For simplicity, in all of the examples we use $p(C|S)$ that is uniform, and $p(S)$ that is uniform (no special knowledge about the information providers by the adversary). Also, to simplify the visualization of the examples, we use models $p(X|C,S) = p(X|C)$. In all examples, we first generate the distribution of the information X given the class C (using different distributions in each example). Then we solve the optimization described in Listing 2.3.1 and present the results. In all examples we assume that the adversary has the ground truth model of $p(X|C)$ as the one we used to generate the information.

We arranged the examples in sections according to the type of the distribution $p(X|C)$. In Section 3.2, the information is generated using a uniform distribution in each class. Afterwards, in Section 3.3 we present examples where the information is generated using a normal distribution in each class. In Section 3.4 we present one example where the information is sampled from a combination of partially overlapping uniform distributions in each class, resulting in a differently shaped distribution of information in the different classes. In Sections 3.2-3.4 all information spaces are one dimensional. Finally, in Section 3.5, we present an example with information that is two dimensional, which was generated using a two-dimensional uniform distribution in each class.

3.2 Uniform Information Per Class (One Dimensional)

In this section, we will present two examples of differential disclosure of information using information distributions that are uniform in each class. In both examples of this section, we will use the notation $\mathcal{I}$ for the information space.

First, we present an example where all the variances of the information distributions for all classes are equal. This is depicted in Figure 3.2.1a, which shows $p(X|C)$. The data was sampled, for each class from a uniform distribution with the same variance (same width of support). Clearly, if X was disclosed directly, the classification for anyone (including the adversary) would be very simple by just applying the following classification function

$$C(x) = \begin{cases} 1 & x \leq 1 \\ 2 & \text{otherwise} \end{cases} \tag{3.1}$$

We learn the differential information mapping function $\mathbf{R}$ by solving the optimization problem in Listing 2.3.1 over the space of differential information mapping functions $\{1,2\} \to AF$ where AF is

$$AF \triangleq \{f : \mathcal{I} \rightarrowtail \mathcal{I} | \exists a \neq 0, b \in \mathbb{R} : \forall x \in \mathcal{I}, f(x) = a \cdot x - b\} \tag{3.2}$$

the set of injective affine functions from and to $\mathcal{I}$.

After optimization, $\mathbf{R}$ is found to be

$$\mathbf{R}(c) = \begin{cases} f(x) = x & c = 1 \\ f(x) = 0.98 \cdot x - 0.86 & c = 2 \end{cases} \tag{3.3}$$

and as a result the conditional distribution of differential information Z given the class C is depicted in Figure 3.2.1b. Note that this is what the intended recipient sees when data is being received since it knows the class of the sender.

On the other hand, the adversary doesn't know the class of the sender and it sees only $p(Z)$ which is depicted in Figure 3.2.1c. The classification distribution $p(C|Z)$ from the point of view of the adversary is depicted in Figure 3.2.1d, and looks pretty uniform as expected.

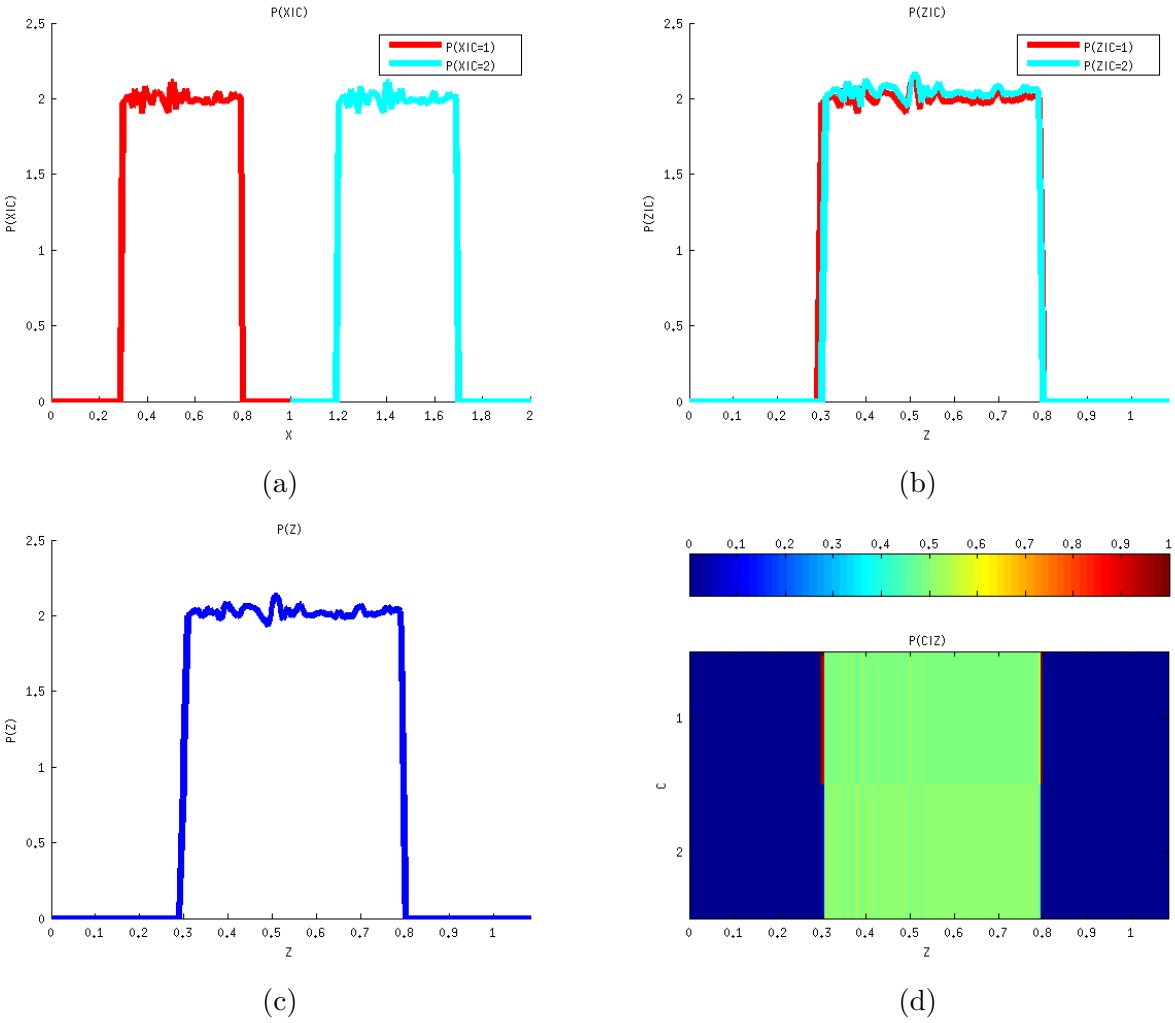


Figure 3.2.1

Next, we present an example with three classes, where the variances of information in the different classes are different this time. This is depicted in Figure 3.2.2a, which shows $p(X|C)$. The data was sampled, for each class, from a uniform distribution with the a different variance (different width of support). Clearly, if X was disclosed directly, the classification for anyone (including the adversary) would be very simple by just applying the following classification function

$$C(x) = \begin{cases} 1 & x \in [0, 1) \\ 2 & x \in [1, 2) \\ 3 & \text{otherwise} \end{cases} \tag{3.4}$$

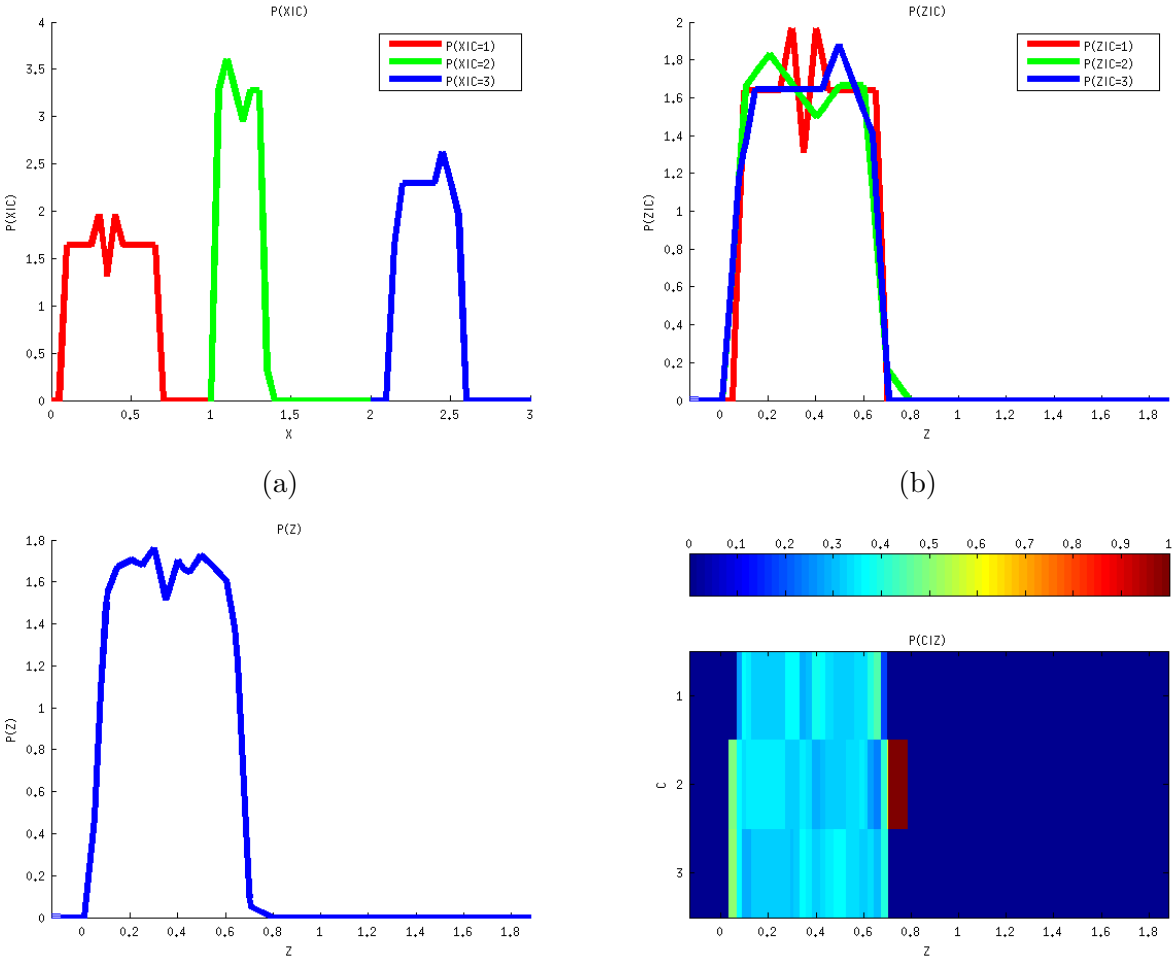
We learn the differential information mapping function $\mathbf{R}$ by solving the optimization problem in Listing 2.3.1 over the space of differential information mapping functions $\{1, 2, 3\} \to AF$ where AF is, again, the set of injective affine functions from and to $\mathcal{I}$.

After optimization, $\mathbf{R}$ is found to be

$$\mathbf{R}(c) = \begin{cases} f(x) = x & c = 1 \\ f(x) = 1.97 \cdot x - 1.96 & c = 2 \\ f(x) = 1.34 \cdot x - 2.92 & c = 3 \end{cases} \tag{3.5}$$

and as a result the conditional distribution of differential information Z given the class C is depicted in Figure 3.2.2b. Like before, this is what the intended recipient sees when data is being received since it knows the class of the sender.

On the other hand, the adversary doesn't know the class of the sender and it sees only $p(Z)$ which is depicted in Figure 3.2.2c. The classification distribution $p(C|Z)$ from the point of view of the adversary is depicted in Figure 3.2.2d, and looks pretty uniform as expected.



(a) (b)

(c) (d)

Figure 3.2.2

3.3 Normally Distributed Information Per Class (One Dimensional)

In this section, we will present two examples of differential disclosure of information using information distributions that are normal in each class. In both examples of this section, we will use the notation $\mathcal{I}$ for the information space.

First, we present an example with two classes, where the variances of the normal distributions in the different classes are equal. This is depicted in Figure 3.3.1a, which shows $p(X|C)$. The data was sampled, for each class from a normal distribution with the same variance. Clearly, if X was disclosed directly, the classification for anyone (including the adversary) would be very simple by just applying the following classification function

$$C(x) = \begin{cases} 1 & x \leq 1 \\ 2 & \text{otherwise} \end{cases} \tag{3.6}$$

We learn the differential information mapping function $\mathbf{R}$ by solving the optimization problem in Listing 2.3.1 over the space of differential information mapping functions $\{1,2\} \to AF$ where AF is, again, the set of injective affine functions from and to $\mathcal{I}$.

After optimization, $\mathbf{R}$ is found to be

$$\mathbf{R}(c) = \begin{cases} f(x) = x & c = 1 \\ f(x) = 0.98 \cdot x - 1.1 & c = 2 \end{cases} \tag{3.7}$$

and as a result the conditional distribution of differential information Z given the class C is depicted in Figure 3.3.1b. Like before, this is what the intended recipient sees when data is being received since it knows the class of the sender.

On the other hand, the adversary doesn't know the class of the sender and it sees only $p(Z)$ which is depicted in Figure 3.3.1c. The classification distribution $p(C|Z)$ from the point of view of the adversary is depicted in Figure 3.3.1d, and looks pretty uniform as expected.

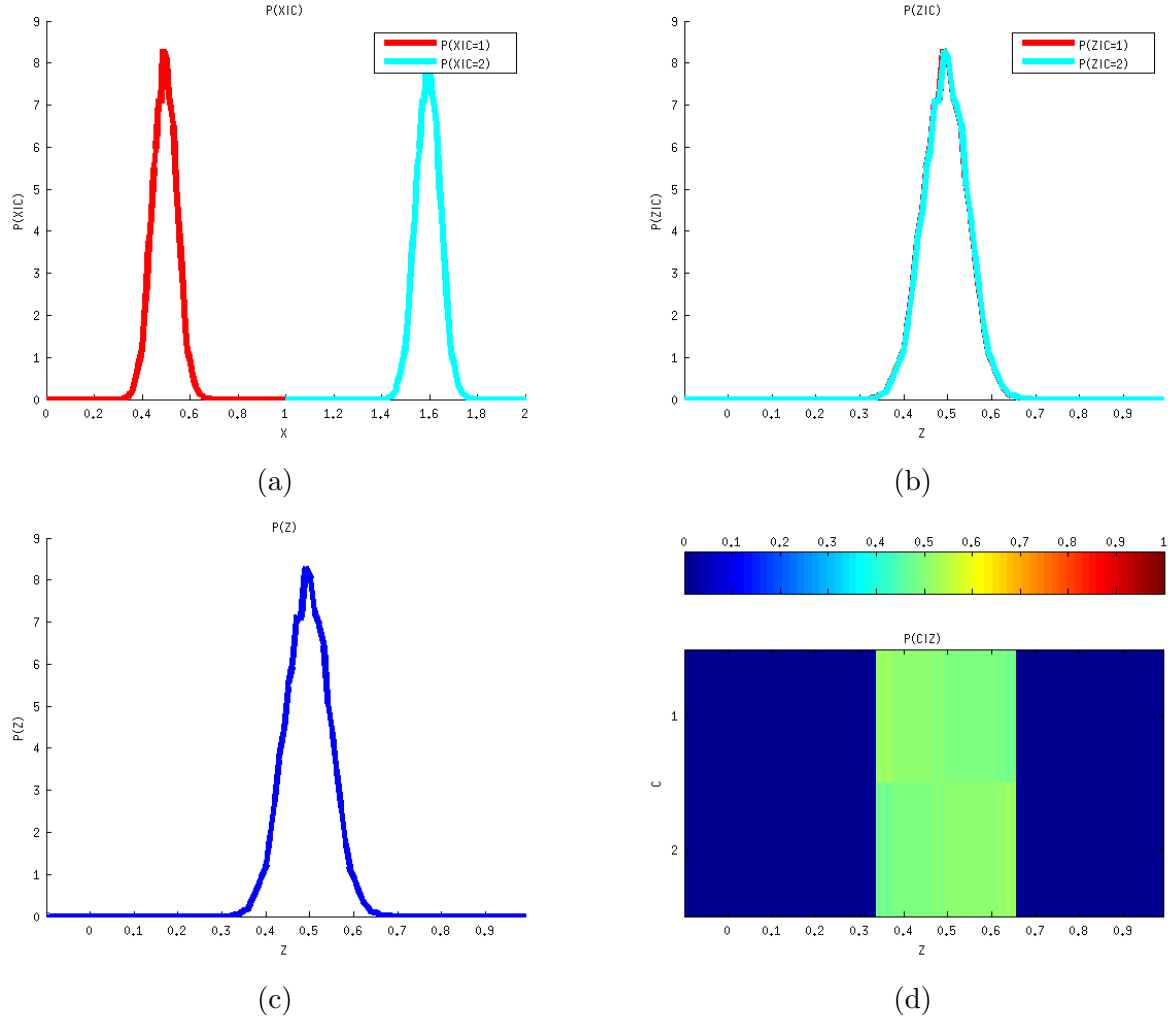


Figure 3.3.1

Next, we present an example with two classes, where the variances of the normal distributions in the different classes are different this time. This is depicted in Figure 3.3.2a, which shows $p(X|C)$. The data was sampled, for each class from a normal distribution with a different variance. Clearly, if X was disclosed directly, the classification for anyone (including the adversary) would be very simple by just applying the following classification function

$$C(x) = \begin{cases} 1 & x \leq 1 \\ 2 & \text{otherwise} \end{cases} \tag{3.8}$$

We learn the differential information mapping function $\mathbf{R}$ by solving the optimization problem in Listing 2.3.1 over the space of differential information mapping functions $\{1, 2\} \to AF$ where AF is, again, the set of injective affine functions from and to $\mathcal{I}$.

After optimization, $\mathbf{R}$ is found to be

$$\mathbf{R}(c) = \begin{cases} f(x) = x & c = 1 \\ f(x) = 2.33 \cdot x - 2.07 & c = 2 \end{cases} \tag{3.9}$$

and as a result the conditional distribution of differential information Z given the class C is depicted in Figure 3.3.2b. Like before, this is what the intended recipient sees when data is being received since it knows the class of the sender.

On the other hand, the adversary doesn't know the class of the sender and it sees only $p(Z)$ which is depicted in Figure 3.3.2c. The classification distribution $p(C|Z)$ from the point of view of the adversary is depicted in Figure 3.3.2d, and looks pretty uniform as expected.

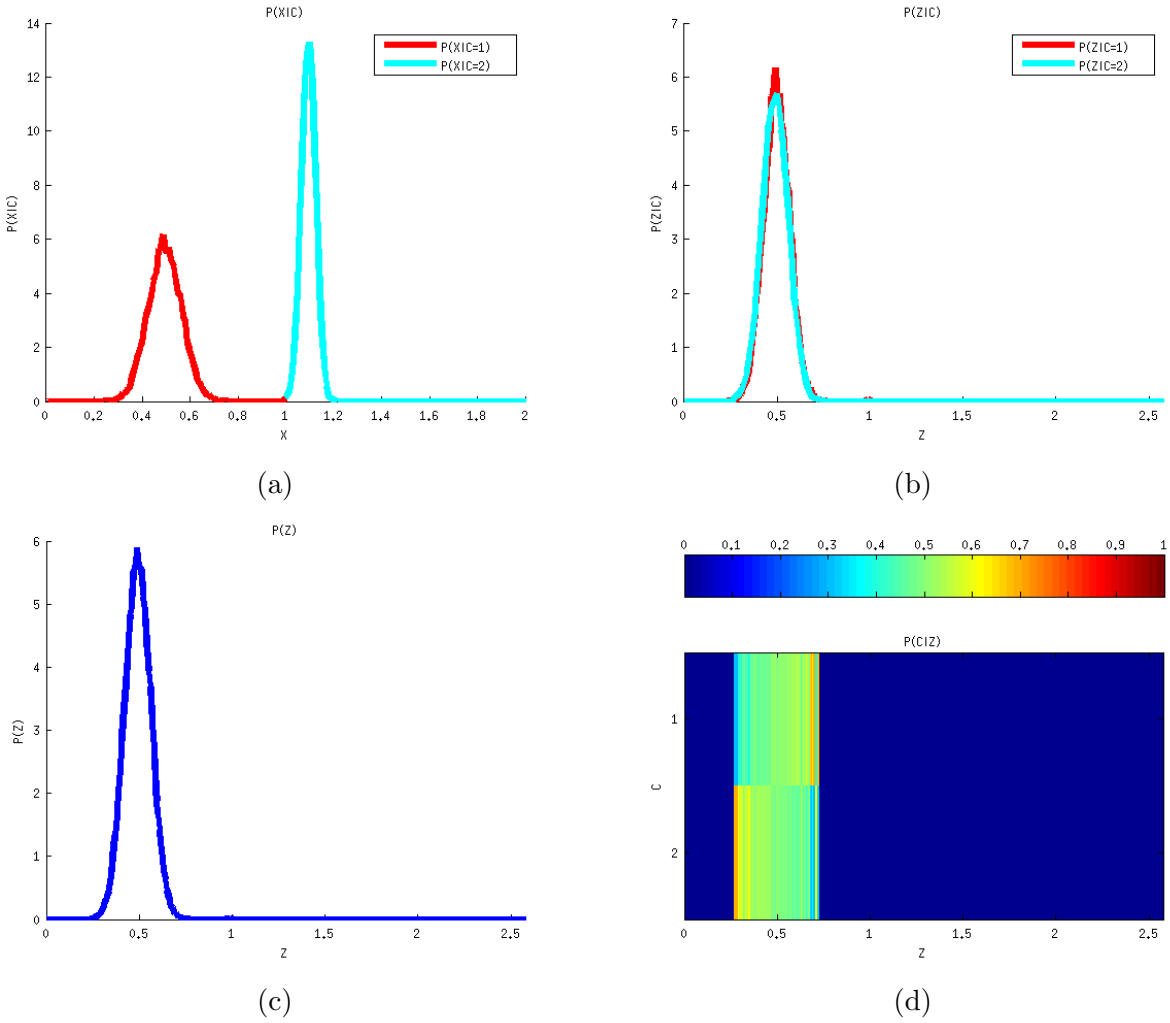


Figure 3.3.2

3.4 More General One Dimensional Information

In this section, we present an example where the information X has a shaped distribution per class. Moreover, the shapes for the distribution of the data are different for different classes. This example includes three classes and the data distribution per class $p(X|C)$ is shown in

Figure 3.4.1a. Note the difference in the shapes of the distributions in the different classes. Still, clearly, if X was disclosed directly, the classification for anyone (including the adversary) would be very simple by just applying the following classification function

$$C(x) = \begin{cases} 1 & x \in [0,1) \\ 2 & x \in [1,2) \\ 3 & x \in [2,3) \end{cases} \tag{3.10}$$

We learn the differential information mapping function $\mathbf{R}$ by solving the optimization problem in Listing 2.3.1 over the space of differential information mapping functions $\{1,2,3\} \to AF$ where AF is, again, the set of injective affine functions from and to $\mathcal{I}$.

After optimization, $\mathbf{R}$ is found to be

$$\mathbf{R}(c) = \begin{cases} f(x) = x & c = 1 \\ f(x) = -1.85 \cdot x + 3.67 & c = 2 \\ f(x) = 1.31 \cdot x - 2.71 & c = 3 \end{cases} \tag{3.11}$$

and as a result the conditional distribution of differential information Z given the class C is depicted in Figure 3.4.1b. Like before, this is what the intended recipient sees when data is being received since it knows the class of the sender.

On the other hand, the adversary doesn't know the class of the sender and it sees only $p(Z)$ which is depicted in Figure 3.4.1c. The classification distribution $p(C|Z)$ from the point of view of the adversary is depicted in Figure 3.4.1d, and looks pretty uniform as expected.

3.5 Uniform Information Per Class (Two Dimensional)

In this last example of this report, we present an example with an information space that is two dimensional and a provider class space that has three classes. To keep the example simple, the information distribution per class is two dimensional uniform with the same support width for all classes and both dimensions. The distribution of information per class is depicted in Figure 3.5.1a as a heatmap. On the horizontal axis is the first dimension of the information space, on the vertical axis is the second dimension of the information space. The colors depict the likelihood for each point in the information space (dark blue = 0, red = 1).

Again it is clear that if x was to be disclosed directly, the classification becomes trivial. The following classification function will classify the data perfectly

$$C(x) = \begin{cases} 1 & x \in [0,1) \times [0,1) \\ 5 & x \in [1,2) \times [1,2) \\ 9 & x \in [2,3) \times [2,3) \end{cases} \tag{3.12}$$

We learn the differential information mapping function $\mathbf{R}$ by solving the optimization problem in Listing 2.3.1 over the space of differential information mapping functions $\{1,2,3\} \to TF$ where TF is

$$TF \triangleq \left\{ f : \mathcal{I} \rightarrowtail \mathcal{I} | \exists b \in \mathbb{R}^2 : \forall x \in \mathcal{I}, f(x) = x - b \right\} \tag{3.13}$$

the set of (injective) functions from and to $\mathcal{I}$ that apply translation only.

After optimization, $\mathbf{R}$ is found to be

$$\mathbf{R}(c) = \begin{cases} f(x) = x & c = 1 \\ f(x) = x - \begin{bmatrix} 0.99 \\ 1 \end{bmatrix} & c = 5 \\ f(x) = x - \begin{bmatrix} 2.09 \\ 2.1 \end{bmatrix} & c = 9 \end{cases} \tag{3.14}$$

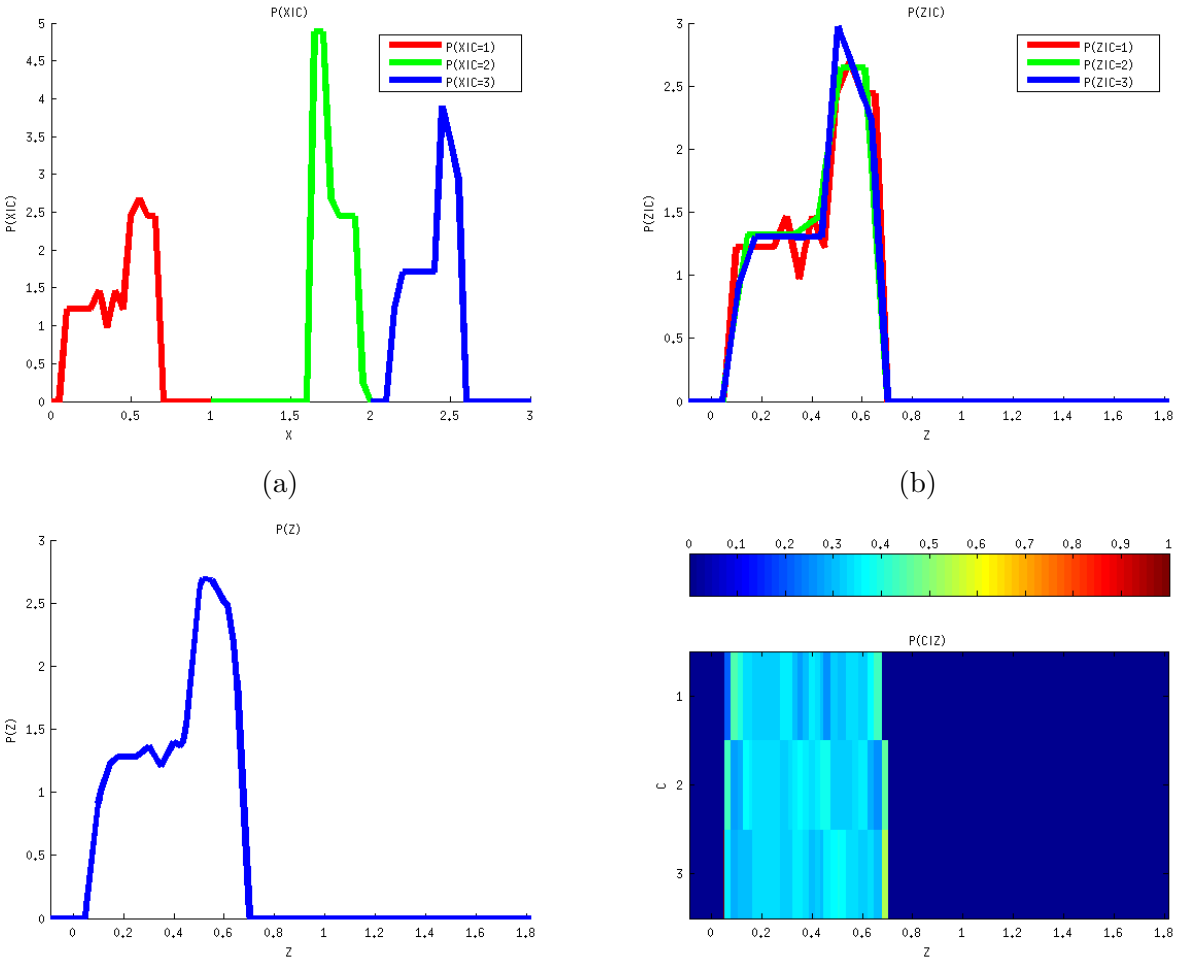
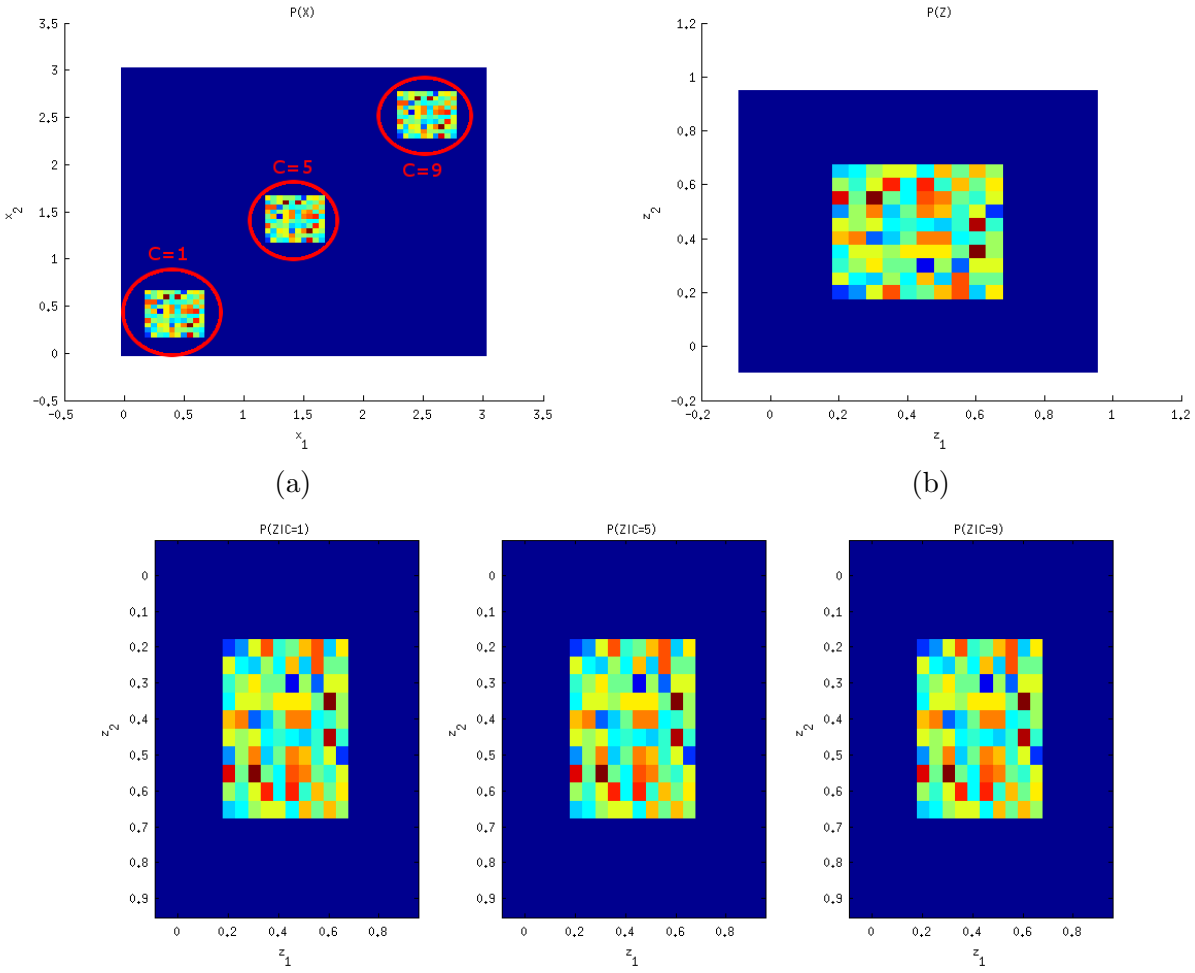


Figure 3.4.1

and as a result the conditional distribution of differential information Z given the class C is depicted in Figure 3.5.1c. Like before, this is what the intended recipient sees when data is received since it knows the class of the sender.

On the other hand, the adversary doesn't know the class of the sender and it sees only $p(Z)$ which is depicted in Figure 3.5.1b. The classification distribution is not viewed here since it is 4-dimensional (2 dimensions for Z, 1 dimension for C, and 1 dimension for the value of the likelihood). But from Figure 3.5.1c, it can be seen that the data overlaps almost completely between the different classes, and therefore the posterior $p(C|Z)$ is pretty uniform as expected.



(a)

(b)

(c)

Figure 3.5.1