Secret Sharing and Algorithmic Information Theory

Tarik Kaced

▶ **To cite this version:**

Tarik Kaced. Secret Sharing and Algorithmic Information Theory. Information Theory [cs.IT]. Université Montpellier II - Sciences et Techniques du Languedoc, 2012. English. NNT : . tel-00763117

HAL Id: tel-00763117
https://theses.hal.science/tel-00763117v1

Submitted on 10 Dec 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ DE MONTPELLIER 2

ÉCOLE DOCTORALE I2S

THÈSE

Partage de secret et théorie algorithmique de l'information

présentée pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER 2

Spécialité : Informatique

par

Tarik KACED

sous la direction d'Andrei ROMASHCHENKO et d'Alexander SHEN

soutenue publiquement le 4 décembre 2012

JURY

Eugène ASARIN	CNRS & Université Paris Diderot	*Rapporteur*
Bruno DURAND	CNRS & Université Montpellier 2	*Examinateur*
Konstantin MAKARYCHEV	Microsoft Research	*Rapporteur*
František MATÚŠ	Institute of Information Theory and Automation	*Rapporteur*
Andrei ROMASHCHENKO	CNRS & Université Montpellier 2	*Co-encadrant*
Alexander SHEN	CNRS & Université Montpellier 2	*Directeur de thèse*

Résumé

Le *partage de secret* a pour but de répartir une donnée secrète entre plusieurs participants. Les participants sont organisés en une *structure d'accès* recensant tous les groupes qualifiés pour accéder au secret. L'objectif est de fournir une information à chaque participant tel que:

1. un groupe qualifié doit pouvoir recouvrer le secret (intégrité),
2. tout autre groupe ne peut obtenir aucune information sur le secret (confidentialité).

Si cet objectif est atteint on parle alors de sécurité inconditionnelle, car la confidentialité est garantie même si l'adversaire est doté d'une puissance de calcul illimitée. Nous utilisons la Théorie de l'Information pour formaliser cette définition et ses propriétés (intégrité et confidentialité). Notre travail se concentre sur les structures d'accès monotones où dès qu'un groupe de participants connaît le secret, tout groupe plus grand le connaît aussi.

Depuis les travaux de Benaloh *et al* en 1988 et Ito *et al* en 1987, on sait que toute structure d'accès monotone possède un schéma de partage parfait. Le problème est que ce schéma fournit de l'information sous forme de *part* dont la taille peut croître exponentiellement avec le nombre de participants. En définissant l'efficacité d'un schéma par le rapport entre la taille maximale d'une part et la taille du secret, la meilleure borne inférieure connue (un résultat de Csirmaz en 1994) est (seulement) sous-linéaire. La méthode habituelle pour obtenir ces bornes inférieures consiste à utiliser des inégalités linéaires pour l'entropie de Shannon, appelées inégalités d'information.

Ma thèse s'articule autour du lien, qui s'avère être très puissant, entre le partage de secret et les inégalités d'information. Dans un premier temps, je définis la notion de partage *quasi-parfait* de secret, pour lequel les propriétés de partage parfait de secret (intégrité et confidentialité) sont légèrement relâchées. Nous avons introduit deux modèles de fuites d'information pour exprimer ce relâchement: le *ratio d'information manquante* comme étant la proportion maximale d'information manquant à un groupe qualifié pour reconstituer le secret, et le *ratio de fuite d'information* comme étant la proportion maximale d'information qu'un groupe interdit peut obtenir sur le secret. Nous travaillons dans un modèle où les fuites tendent vers zéro lorsque que la taille du secret augmente. Pour notre notion de partage quasi-parfait, il s'avère que le respect strict de l'intégrité ne pose pas de problème (Sect. 5.1.2), et que l'on peut supposer que le secret obéit à une distribution uniforme (Sect. 5.1.3). Dans le cas où l'augmentation de la taille du secret n'aurait pas de sens pratique nous avons travaillé sur une taille finie et montré qu'elle peut être réduite à un seul bit tout en gardant un niveau de sécurité acceptable (voir Sect. 5.1.4).

Notre définition de partage quasi-parfait a motivé l'étude une notion analogue dans le cadre de la Théorie Algorithmique de l'Information : le *partage de secret algorithmique*. Ce type de partage de secret est intrinsèquement non-parfait parce que la complexité de Kolmogorov n'est définie qu'à une constante additive près. Mais nous montrons que les notions de partage quasi-parfait sont équivalentes dans le cas probabiliste et algorithmique (Sect. 5.2.2). Ainsi, la question ouverte principale de notre sujet est de déterminer si la notion de partage parfait

de secret coïncide avec celle du partage quasi-parfait. Nous montrons que les inégalités d'informations ne suffisent pas à séparer directement ces deux notions (Sect. 5.1.5), nous conjecturons que certaines inégalités d'information *conditionnelles* pourraient y parvenir.

Motivés par cette problématique, nous avons étudié de plus près les inégalités conditionnelles d'information. Nous introduisons une nouvelle inégalité non-triviale et montrons que celle-ci ainsi que d'autres déjà connues sont *essentiellement conditionnelles*, dans le sens où elles ne découlent pas d'autres inégalités inconditionnelles (Sect. 6.1). Nous montrons aussi que les inégalités conditionnelles se divisent en deux types. Certaines sont valides pour tout les tuples de réels correspondant aux entropies d'une distribution de probabilité (on appelle ces tuples des points entropiques), alors que d'autres sont de surcroît valides pour toutes les limites de points entropiques (dits points presqu'entropiques) (Sect. 6.3). Finalement, nous montrons que pour certaines des inégalités d'information conditionnelles (pour l'entropie de Shannon) il existe des inégalités conditionnelles algorithmiques associées (pour la complexité de Kolmogorov) (Sect. 6.4).

Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier LIRMM, 161 rue Ada, 34095 Montpellier Cedex 5 - France

---

Ce travail est soutenu par les projets NAFIT ANR-08-EMER-008-01 et EMC ANR-09-BLAN-0164-01.

Abstract

Secret sharing deals with splitting a secret data between several participants. The participants are organized within an *access structure* consisting of groups that may gain access to the secret. The goal is to provide each participant with a piece of information such that the following two requirements are met:

1. RECOVERABILITY: a qualified group can recover the secret,

2. PRIVACY: any other group cannot obtain any information on the secret.

Whenever these two requirements are fulfilled, the scheme is said to be unconditionally secure – security is guaranteed even against adversaries with unbounded computational power. We use Information Theory to formalize this definition and its properties (recoverability, privacy). Our work focuses on monotone access structures, for which any group containing a qualified subgroup is also qualified.

Since Benaloh *et al* and Ito *et al*, we know that every access structure admits a perfect secret-sharing scheme. However, a possible issue is that the size of information pieces assigned as shares to participants can grow exponentially with the number of participants. Defining the efficiency of a scheme as the ratio between the maximum size of a share and the size of the secret, the best known lower bound – from a result of Csirmaz in 1994 – is (only) almost linear. The usual method for obtaining these lower bounds consists in applying linear inequalities for Shannon entropy, also called information inequalities.

The research axis of my thesis is based upon the link – which turns out to be quite powerful – between secret sharing and information inequalities. First, I define the notion of *quasi-perfect* secret sharing, where the perfect secret-sharing requirements (recoverability, privacy) are slightly relaxed. To this aim, we have introduced two quantities for leakage: the *missing information ratio* as the maximal proportion of information a qualified group needs to recover the secret completely, and the *information leak ratio* as the maximal proportion of information forbidden groups can learn about the secret. In our model, these leaks are vanishing as the size of the secret grows. For our notion of quasi-perfect secret sharing, it turns out that strict recoverability is never an issue (see Section 5.1.2), and that we can assume the secret uniformly distributed (see Section 5.1.3). If the size of the secret should remain small, we also worked on fixed finite size for the secret and showed that the secret size can be reduced to a single bit, while keeping an acceptable level of privacy in terms of information leak (see Section 5.1.4).

Our definition of quasi-perfect sharing motivated our investigation of a similar notion for the case of Algorithmic Information Theory: *algorithmic secret sharing*. This kind of secret sharing is intrinsically non-perfect since Kolmogorov complexity is only defined up to a constant additive term. However, we show that the notions of algorithmic and probabilistic quasi-perfect secret sharing are equivalent (see Section 5.2.2). An open question remains: whether perfect and quasi-perfect secret sharing schemes can achieve significantly better ratios than perfect ones. We show that the direct use of information inequalities is unable

to separate these two notions (see Section 5.1.5), though we conjecture that a type of *conditional* information inequalities might be useful.

This was our incentive to further investigate *conditional information inequalities*. We introduce a new nontrivial conditional inequality and show that it belongs to the class of *essentially conditional* inequalities together with all nontrivial ones known to-date in the literature. They are essentially conditional in the sense that they are not a direct consequence of an unconditional inequality (see Section 6.1). We also show that the known conditional inequalities are of two types. Some are valid for tuples of reals that correspond to the entropies of a probability distribution (those tuples are called entropic points), while others are also valid for limits of entropic points (called almost entropic points) (see Section 6.3). Furthermore, we extend this result to the framework of Kolmogorov Complexity. We show that for some of these conditional information inequalities (for Shannon Entropy), there exists a conditional algorithmic inequality (for Kolmogorov complexity) (see Section 6.4).

Contents

List of Symbols

$\binom{n}{k}$	The number of ways of choosing k elements from a set of n elements
$\mathbb{E}$	The expectation operator
Γ	An access structure
Γ_n^*	The set of entropic points for n random variables
Γ_n	The set of polymatroids on n elements
$C(\cdot)$	Kolmogorov complexity
$C(\cdot\|\cdot)$	Conditional Kolmogorov complexity
$\lceil x \rceil$	The smaller integer greater or equal to x
$\lfloor x \rfloor$	The greater integer less or equal to x
$[\![n]\!]$	The set of integers from 1 to n
ln	Natural logarithm (base exp(1))
log	The logarithm to base 2
$\mathbb{F}_q$	The finite field with q elements for a prime power q
$\mathbb{N}$	The set of integers
$\mathbb{R}$	The set of real numbers
$\mathbb{R}_+$	The set of non-negative real numbers
$\mathscr{P}(X)$	The power set of X
Pr	A probability distribution function
rk$(\cdot)$	Rank function for vector spaces
$\vec{H}(X_{\mathcal{N}})$	The entropy profile of the tuple $X_{\mathcal{N}}$
$\|A\|$	The cardinality of the set A
$\|v\|$	The absolute value of the v.
$H(\cdot)$	Shannon's entropy function
$H(\cdot\|\cdot)$	Conditional entropy
$I(\cdot:\cdot)$	Mutual information

$I(\cdot : \cdot \mid \cdot)$	Conditional mutual information
shortlex order:	The order obtained by sorting the lengths and ordering lexicographically for each length

List of Figures

Chapter 1

Introduction

Contents

Our work deals with *secret sharing* in the theoretical points of view of Shannon's *Information Theory* and Kolmogorov's *Algorithmic Information Theory*. We are going to see why and how these *three* subjects are naturally deeply intertwined.

1.1 What is Secret Sharing About ?

Secret sharing in its physical implementation has existed for centuries. Imagine a safe box with two locks; the keys for these locks are given to different people. To access the contents of the safe we do need both their consent. It can sometimes be useful (e.g., preventing a crazy person from launching a nuclear assault).

Sometimes we may need a more complicated arrangement. Imagine that we are afraid of the loss of a key. It would be preferable to have a safe with three keys that can only be opened if at least two keys are present. By giving these three keys to three different people, we protect ourselves against the loss of one key, and we require at least two people who agree to open the safe.

Now this sounds like a good idea, but how to make such a safe box? A mechanical device can be used, but a simpler solution should exist in our digital era. For the first example (when two people agree to open the safe) the solution is easy. Use a standard lock with a digital code (say a n-digit number), but first ask the producer to represent the code c as the sum of two numbers a and b, and send these two numbers to different people (say, Alice and Bob). Then we have the same advantages without the price of making a physical lock with two keys. Acting in agreement, Alice and Bob add their numbers and uncover the code. On the other hand, if one of them refuses to participate, the other one cannot open the safe since she does not know anything about the code.

Can a similar technique be used for more complicated arrangements, e.g., for the three people problem described above? Yes, and it was Adi Shamir who invented an elegant scheme to achieve this goal. This scheme is based on simple properties of polynomials over a finite field. It even generalizes our examples: one can invent a scheme with, say, 7 people such

that the consent of at least, say, 5 among them is required to get into the safe (or to launch the rocket)[1].

However, the simple majority (or any threshold) is not the only rule that makes sense. Any person with administrative experience knows that some people are more equal than others, so the number of participants is not the only important parameter, but also the exact composition of the group. This leads us to the general problem of secret sharing: there are n participants, and some groups are considered as *authorized*, they should be able to reconstruct the secret, while all other, *forbidden*, groups should not be able to recover the least bit of information.

This real-life problem can be formalized via Shannon information theory: the secret is a random variable, and the shares (given to participants) are random variables, too. The requirements for authorized and forbidden sets can be reformulated in terms of equalities involving Shannon entropy and mutual information. We therefore obtain the following mathematical problem: implement a given structure efficiently, where the efficiency is measured by the size of the shares (compared to the size of the secret).

Unfortunately, this problem is open: the known schemes (for an arbitrary collection of authorized groups) are very inefficient; on the other hand, no known results prove that efficient schemes could not exist. This problem seems to be very difficult, and one could look more closely at its variations and tools for its analysis.

1.2 The Results

Part of our contribution is making several (small) steps in this direction. First, we introduce the notion of approximate secret sharing. This means that some information leak is allowed: the forbidden groups can get a small amount of information about the secret (which does not allow them to reconstruct it fully or even come close). Such schemes can be of practical use, especially if one can find an approximate scheme that is much more efficient than the known perfect ones. (But this goal has not been achieved yet.)

Secondly, the introduction of approximate schemes allows us to compare the Shannon information-theoretic setting to the algorithmic information theory setting. In reality the key and the shares are bit sequences, not random variables — so it would be natural to formulate the requirements in this language. (Imagine that an auditing company wants to control whether the producer of a lock with a shared secret combination did a good job. Then auditors only have the actual key and shares, and should approve or disapprove them based only on the final values, not the process of their creation.) We show that these two approaches (Shannon and algorithmic) are in a sense equivalent: if we have an efficient scheme in one framework, we can construct an efficient scheme in the other one.

Finally, we take a closer look on the tools used to prove the non-existence of secret-sharing schemes for some values of the parameters. Such a non-existence result can usually be formulated as the dissatisfaction of a conditional inequality for random variables: if some information quantities are zeroes, then the size of the shares is bounded. Information inequalities play a central role in this text. They are the inequalities for Shannon entropy, but they are also in one-to-one correspondence with the inequalities for Kolmogorov complexity. Kolmogorov complexity formalizes the idea of randomness for strings. These two reasons alone justified to consider the notion of secret sharing in the Algorithmic framework (if one can share a random secret one can share anything). Switching to approximate sharing, we naturally want to transform these conditional inequalities to unconditional ones. This is a

[1] This scheme is actually used to physically protect the DNSSEC Root key and "reboot the Internet" in case of a cyberterrorist attack

difficult question: nobody knows how to describe all conditional inequalities (a large number of non-trivial inequalities, called non-Shannon, were discovered in the last decade). We continue this research by proving some new conditional and unconditional inequalities and (which is probably more important) by showing that some conditional inequalities cannot be converted (by Lagrange multipliers) to unconditional ones.

1.3 Two Other Examples

Sometimes, secret sharing arises naturally as a byproduct of a concrete situation where a secret seems to be shared between two participants, even without the intervention of a third party. In fact, this is what could happen in a simplified penalty shootout (penalty kick). In this simple game, two players are facing each other: a goalkeeper trying to stop a soccer ball and a kicker taking the shot who wants to score a goal. Both players first make a choice and select a private strategy which consist of one of two options: `left` or `right` (the possible choices are the same for both players). For the kicker, the `left` option means she will shoot on the left side, symmetrically `right` means she will shoot on the right side. For the goalkeeper, the `left` option means she will move to her left side, symmetrically `right` means she will move to her right side. After these strategies are chosen – and assuming both would play perfectly – the secret outcome is whether a goal will be scored or not. This outcome remains secret until it is revealed when both the players make their moves. This example may show that secret sharing is (implicitly) around us more than we thought. But it will not be much more relevant in the rest of this text.

Secret sharing was originally introduced to protect cryptographic keys. Cryptography, most of the time, relies on secret data that should remain private. This secret data or key is often intended to encode a secret message, and must be used to decrypt the message. To transmit the secret message safely, we publish the encrypted message but we should also transmit the key safely! It looks like we are back to our original problem: we avoided the transmission of the message and now we have to deal with the key. This is where secret sharing comes into play and overcomes this issue. It allows the transmission of the key without revealing it.

The following second example is in fact the well-known method of encryption called the *Vernam Cipher* or *one-time pad*. This cipher is intended to encrypt a secret message securely with a key of the same length in the following way. For a message consisting of n letters from `A` to `Z`, we create a secret and "random" key of the same size on the same alphabet. To encrypt the message, shift circularly each letter of the message by the corresponding letter of the key (e.g. for letters `Y` and `E` we obtain `D`).

message	THEKEYOFTHISMESSAGEISBELOWTHIS
key	THISSECRETKEYISUNDERTHEMESSAGE
encrypted message	NPNDXDRXYBTXLNLNOKJAMJJYTPMIPX

Figure 1.1: An example of one-time pad

Giving the encrypted message to Alice and the key to Bob, they can meet and uncover the message.

1.4 Outline of the Thesis

- Chapter 2 provides elementary notions of discrete probability theory and information theory. Shannon entropy and related information measures are then defined and their

basic properties are presented. The focus is then set on information inequalities: linear inequalities for entropies. The central inequality is Claude Shannon's basic inequality, it generates what we call Shannon-type inequalities. For a fixed number of random variables one can try to find all the possible valid inequalities for their entropies. We explain Zhang and Yeung's general setting, the Entropy Region, to investigate the valid inequalities; we prove non-trivial inequalities and give an overview of the current state of the research.

- Chapter 3 can be used as a short course on perfect secret sharing. First, we describe the setting of perfect secret sharing and explain the fundamental constructions. We discuss the efficiency of perfect secret-sharing schemes and present the ideal case and its relation with matroid theory. We also give a new property of linear secret-sharing schemes. Then, we discuss a connection between conditional information inequalities and the efficiency of such schemes.

- Chapter 4 contains a short description of Kolmogorov complexity and its relation to Shannon entropy. We formulate useful lemmas in these frameworks. We prove two 5-variable non-Shannon type inequalities, one of which is new. The results will also be used as lemmas in some of the proofs of the next chapters.

- Chapter 5 investigates new versions of secret-sharing schemes (possibly non-perfect). The first one deals with schemes that are sequences of individual subschemes. The idea is to study the asymptotics of the parameters of a scheme implementing a fixed access structure. The second is a version of secret sharing in the algorithmic setting (for Kolmogorov complexity). These two notions turn out to be are equivalent. We further investigate the basic properties of our notions.

- Chapter 6 is the result of a joint work with A. Romashchenko. We study of the notion of essentially conditional inequality. Some nontrivial conditional inequalities from the literature have been conjecture to be consequences of (yet unknown) unconditional inequalities. We show that this is not always the case. Some conditional inequality cannot be deduced from any unconditional inequality. We prove there are conditional inequalities that do not hold for almost entropic points. We define what are conditional Algorithmic inequalities in the framework of Kolmogorov complexity and give some examples of such inequalities.

After receiving remarks from the referees, the following major updates have been added to the manuscript: a proof on the relation of matroids and ideal perfect secret-sharing schemes; a proof of equivalence of two techniques for obtaining non-Shannon-type inequalities; a geometric interpretation of the different types of conditional inequalities.

Chapter 2

The Entropy Region & Information Inequalities

Contents

Introduction

In this chapter, we introduce the study of discrete random variables and their interdependencies via the geometric approach of entropy regions. Information Theory finds its foundations in the elements of probability theory. The central notion is that of Shannon entropy, and its related information measures. We present the basic properties of this measure of uncertainty, putting the focus on the study of information inequalities. We prove Shannon's basic inequality, the first linear inequality used to bound the entropies of tuples of random variables. The further investigation of information inequalities lead to the general framework introduced by Zhang and Yeung: the Entropy Region. Its characterization for every size of tuples remains open in general. Finally we present the main known techniques for proving nontrivial information inequalities. For a more thorough study of the concepts described hereafter, the interested reader is referred to [CF11, Yeu08, Gra90, CT91, Mac03].

2.1 Probability Distributions and Random Variables

We start with some definitions from probability theory.

2.1.1 Probability spaces

A finite probability space consists of a non-empty finite set $\mathcal{W}$ together with a function $\Pr : \mathcal{W} \to \mathbb{R}_+$ such that

$$\sum_{w \in \mathcal{W}} \Pr(w) = 1.$$

Such a function Pr is called a *probability mass function* or a *(probability) distribution*. The set $\mathcal{W}$ is called the sample space and its elements are elementary events. More generally, an *event* is a subset of the sample space $\mathcal{W}$, with the convention that $\Pr(\varnothing) = 0$. The probability of an event $A \in \mathscr{P}(\mathcal{W})$ is defined as the sum of all elementary events it contains:

$$\Pr(A) = \sum_{a \in A} \Pr(a).$$

An example of a basic and important distribution is the *uniform distribution*. The uniform distribution on m elements charges each of the m possible outcomes with the same probability mass, so that all elementary events are equiprobable. Formally, if the sample space consists of $m = |\mathcal{W}|$ elementary events, then each of them has probability $\frac{1}{m}$.

The knowledge of some information might change uncertainty in a drastic way. This notion is expressed via the conditional probability distribution. Let $A, B \in \mathscr{P}(\mathcal{W})$ be any two events such that $\Pr(B) > 0$, then the conditional probability of A given B is defined by

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}.$$

Hereafter, we will use the standard notation $\Pr(A, B) = \Pr(A \cap B)$.

2.1.2 Discrete Random Variables

Random variables. A (discrete) finite random variable X on a probability space $(\mathcal{W}, \Pr)$ is any function X on the set $\mathcal{W}$. We call the set of possible outcomes $\mathcal{X} = X(\mathcal{W})$ the domain (or sometimes alphabet) of X. Any random variable comes with its probability distribution denoted p_X defined as the function that associates each event "$X = x$" with its probability, for $x \in \mathcal{X}$.

$$p_X(x) = \Pr(X = x) = \sum_{w \in \mathcal{W} : X(w) = x} \Pr(w).$$

We may write $X \sim p_X$ to say that X follows the probability distribution p_X. The support of X, denoted by $\mathcal{S}_X$, is the subset of the alphabet $\mathcal{X}$ consisting of outcomes with positive probability.

Joint probability. Let X, Y be two random variables defined on the same probability space, one can consider the random variable consisting of the pair (X, Y), it also has some distribution. This distribution, denoted by p_{XY}, is called the *joint* distribution of X and Y and is defined on $\mathcal{X} \times \mathcal{Y}$ by

$$p_{XY}(x, y) = \Pr(X = x, Y = y).$$

Given any joint distribution p_{XY} for X, Y one can recover the *marginal* distribution of X via:

$$p_X(x) = \sum_{y \in \mathcal{S}_Y} p_{XY}(x, y).$$

This formula can be generalized to any finite tuple of random variables. The joint probability distribution p_{XY} can be pictured using the $|\mathcal{X}| \times |\mathcal{Y}|$ matrix $[\Pr[X = i, Y = j]]$.

Example 1. *Suppose you roll a die D_1 and then a die D_2 until the value of D_2 is less or equal to the value of D_1. The joint probability distribution of D_1 and D_2 is displayed in on Figure 2.1.*

D_1/D_2	1	2	3	4	5	6
1	$\frac{1}{6}$	0	0	0	0	0
2	$\frac{1}{12}$	$\frac{1}{12}$	0	0	0	0
3	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	0	0	0
4	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	0	0
5	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	$\frac{1}{30}$	0
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

Figure 2.1: Example of joint probability distribution

Conditional probability. For any two jointly distributed random variables X, Y,

$$\Pr(X = x|Y = y) = \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)}$$

Thus, for each value $y \in \mathcal{S}_Y$ of Y, there is a (conditional) distribution on X given that "$Y = y$", denoted by

$$p_{X|Y}(x|y) = p_{X|Y=y}(x) = \Pr(X = x|Y = y).$$

As customary, it shall be space-efficient to drop the subscripts and use $p(x)$ or $p(x|y)$ in place of $p_X(x)$ or $p_{X|Y}(x|y)$ if no confusion arises from the context.

2.1.3 Dependence Relationships

Mutual Independence. We say that two random variables X and Y are independent, written $X \perp Y$, if

$$p(x, y) = p(x) \cdot p(y) \text{ for all } (x, y) \in \mathcal{X} \times \mathcal{Y},$$

or equivalently if $p(x|y) = p(x)$ for all $(x, y) \in (\mathcal{X}, \mathcal{S}_Y)$.

The second (equivalent) definition makes intuitive sense in terms of information. It says that the knowledge of the outcome of Y does not provide any information on the outcome of X, *i.e.*, the conditional distributions of X given that $Y = y$ are the same for all $y \in \mathcal{S}_Y$ such that $p(y) > 0$.

In terms of the matrix defined earlier, the mutual independence means that the matrix $[p(x, y)]$ has rank one, since any two rows should be proportional. A set of random variables is *pairwise independent* if any two of them are mutually independent, this set is *mutually independent* if moreover each possible disjoint subsets are mutually independent.

Conditional independence. We say that two random variables X and Y are conditionally independent (given Z), written $X \perp Y|Z$, if

$$p(x, y, z) \cdot p(z) = p(x, z) \cdot p(y, z) \text{ for all } (x, y, z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$$

or equivalently if $p(x|z) = p(x|y,z)$ for $p(y,z) > 0$.

Functional dependence. We say that X functionally depends on Y if

$$\forall y \, \exists x \, , p(x|y) = 1,$$

or equivalently if $y \in \mathcal{S}_Y$ then $p(x,y) = p(x)$ for a unique $x \in \mathcal{S}_X$.

The functional dependency expresses the case when a random variable becomes deterministic when another is known. The definition implies that there exists a surjective function $f : \mathcal{S}_Y \to \mathcal{S}_X$ such that $X = f(Y)$.

Examples. Let us consider random points and lines in the finite affine plane.

Example 2. *Pick uniformly two distinct points A, B and let C be the line going through A and B. In this case, C is a function of the random variables A, B.*

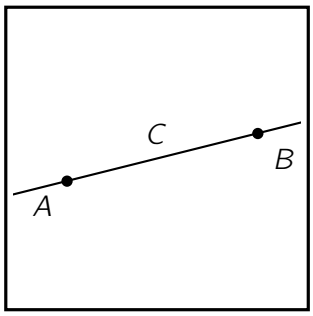


Figure 2.2: Two points and a line

Example 3. *Slightly tilt the previous example. Pick a uniformly random line C, then pick uniformly two points A and B independently amongst all points of C. This time, $A \perp B|C$ but C is no longer a function of A, B.*

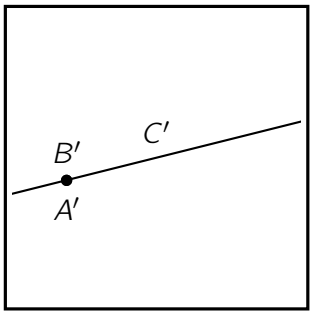


Figure 2.3: Two points and a line, a variant

Example 4. *Let $A, B \in \mathbb{B}^n$ be two random binary strings of length $n \geq 1$. Suppose further that A and B are independent and uniformly distributed. Define $C = A \oplus B$ to be the bitwise* `xor` *of A and B. In this example :*

- *A, B and C are pairwise independent.*
- *A, B, C are not mutually independent.*
- *Each variable is a function of the two others.*

2.2 Information Measures

We introduce Shannon's information measures, also known as entropy, conditional entropy, mutual information and conditional mutual information. We also introduce the Kullback–Leibler divergence.

2.2.1 Shannon's Information Measures

Shannon's Entropy

Definition 1 (Shannon Entropy)**.** *The Shannon entropy of a random variable X is defined by*

$$H(X) = \sum_{x \in \mathcal{S}_X} p(x) \log \frac{1}{p(x)} = -\sum_{x \in \mathcal{S}_X} p(x) \log p(x)$$

The base of the logarithm is unimportant (as long as it is greater than 1), it only accounts for the base unit of the measure of information. Computer scientists usually adopt the logarithm to base 2, and call *bit* the unit of entropy[1]. For a binary random variable X, whose support $\mathcal{S}_X$ consists of two values, say $\mathbb{B} = \{0, 1\}$, let $p = \Pr(X = 0) = 1 - \Pr(X = 1)$. In this case, the entropy of X is called the *binary entropy function* $h(p)$, depicted in Figure 2.4, and rewrites to

$$H(X) = h(p) = -p \log p - (1 - p) \log (1 - p).$$

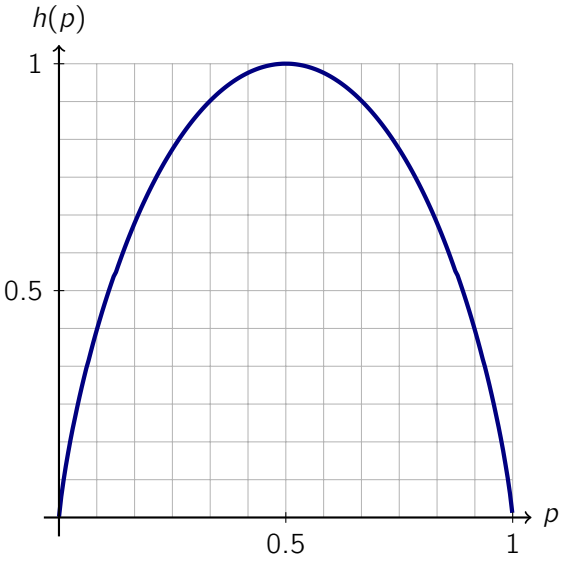


Figure 2.4: The binary entropy function $h(p)$

Expectation. The expectation of a real random variable X is defined by

$$\mathbb{E}[X] = \sum_{x \in \mathcal{S}_X} p(x) \cdot x,$$

which denotes the average value taken by a real random variable according to the "weights" (probabilities) of each possible values. The Shannon entropy can be understood as the expectation of the random variable $-\log p(X)$, *i.e.*, $H(X) = \mathbb{E}[-\log p(X)]$.

Definition 2. *A function $f : (a, b) \to \mathbb{R}$ is said convex if for all $\lambda \in [0, 1]$ and all $x, y \in (a, b)$:*

$$\lambda f(x) + (1 - \lambda) f(y) \geq f(\lambda x + (1 - \lambda) y)$$

We say that f is strictly convex iff the equality holds only when $\lambda \in \{0, 1\}$. Whenever $-f$ is (strictly) convex then f is called (strictly) concave.

[1] Here, the bit unit is not to be confounded with binary digit, which is a value in $\mathbb{B} = \{0, 1\}$.

Shannon's entropy function is continuous in each of its parameters, and is a strictly concave function. A basic result for convex function is Jensen's inequality: for any convex f defined on the interval D and any real-valued random variable X taking values in D,

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

The intuitive explanation is that the barycenter of a set of values of a convex function f is always above the graph of f.

Proposition 1. $0 \leq H(X) \leq \log|\mathcal{S}_X|$, *for all random variables* X.

Equality occurs for a deterministic distribution, for the left part, and for the uniform distribution, for the right part.

Proof. The quantity $-y\log y$ is always non-negative for $y \in (0,1]$ with equality iff $y = 1$. This gives the left-hand inequality for which equality occurs when $|\mathcal{S}_X| = 1$. The other inequality is a direct corollary of Jensen's inequality :

$$\mathbb{E}\left[\log\frac{1}{p_X(x)}\right] \leq \log\left(\mathbb{E}\left[\frac{1}{p_X(x)}\right]\right) = \log|\mathcal{S}_X|.$$

If X is uniformly distributed then

$$H(X) = \sum_{1}^{|\mathcal{S}_X|}\frac{1}{|\mathcal{S}_X|}\log|\mathcal{S}_X| = \log|\mathcal{S}_X|.$$

The uniform distribution hence makes the right-hand inequality an equality. □

Motivation: entropy as the average length of a code. The following interpretation is an incentive for the theoretic study of Shannon entropy.

A monkey types a message on a typewriter, a message made of characters from his favorite alphabet. Our monkey has some preferences amongst tuples of letters, some are more likely than others. We want to transmit this message using bits 0 and 1 in the most efficient way possible. For this we design a code by associating a codeword to each letter ℓ in the alphabet. A codeword consists of a binary string $c_\ell \in \mathbb{B}^*$.

The plan is to encode the message by replacing each letter by its codeword and transmit the corresponding sequence of bits. For this plan to work, the code needs to be uniquely decipherable, meaning that given such an encoded message, there must exist a unique way of breaking it into codewords. An example of such a code is a prefix-free code, which is a set of codewords such that no codeword is a prefix of another codeword. As we want to minimize the length of the encoded message, the average codeword length is the reasonable measure of efficiency for a code. Shannon's noiseless coding theorem asserts the following:

Theorem 1 (C. E. Shannon, [Sha48a, Sha48b])**.** *Let X be a random variable,*

(a) For every uniquely decodable code C for X, the average length a codeword is always greater than $H(X)$.

(b) There exists a uniquely decodable C_X (and even prefix-free code) such that the average length of a codeword is at most $H(X)+1$.

There may exist many prefix-free codes achieving the bound in (b). For example, Huffman's code is an optimal prefix-free code. If we allow the encoding of information by blocks, instead of single letters, the constant 1 can be improved to any constant $\epsilon > 0$ (as small as we want

provided that the block-length is large enough). This confirms the idea that entropy is indeed a good interpretation of the amount of information a random variable contains.

Measures related to entropy

Shannon suggested other useful information measures related to entropy.

Conditional entropy. We denote by $H(X|Y=y)$ the entropy of a random variable whose distribution is $p_{X|Y=y}$. We can define the conditional entropy of X given Y as

$$H(X|Y) = \mathbb{E}[H(X|Y=y)].$$

This is just the average entropy of the distributions $p_{X|Y=y}$ for all y.

One can massage this equation into the well-known identity for conditional entropy:

$$H(X|Y) = H(XY) - H(X).$$

From this one can use induction to recover the *chain rule* for conditional entropy

$$H(X_1, X_2, \ldots, X_m) = \sum_{i=1}^{m} H(X_i|\{X_j : j < i\}).$$

Mutual information. The mutual information $I(X{:}Y)$ is defined as follows:

$$I(X{:}Y) = H(X) + H(Y) - H(XY).$$

Defined as above, it is the excess of information from both parts (independently) over the information of the pair. The mutual information is symmetric in both its variables[2]. An equivalent form using conditional entropy also carries a meaningful interpretation :

$$I(X{:}Y) = H(X) - H(X|Y) = H(Y) - H(Y|X),$$

which defines the mutual information as the amount of information that Y knows about X (the same amount of information that X knows about Y).

Conditional mutual information. For three random variables X, Y, Z, one can look at the conditional distribution of X and Y given a value z of Z, and compute the mutual information between these two distributions. We denote this mutual information by $I(X{:}Y|Z=z)$. The conditional mutual information can be now defined using the expectation as follows:

$$I(X{:}Y|Z) = \mathbb{E}[I(X{:}Y|Z=z)].$$

It is only a matter of rewriting to see that this quantity has the following equivalent form in (plain) entropy terms:

$$I(X{:}Y|Z) = H(XZ) + H(YZ) - H(XYZ) - H(Z).$$

[2]The symmetry explains our preference for the colon (:) rather than the semicolon (;) in $I(\cdot{:}\cdot)$.

From similar previous arguments, the conditional mutual information is also symmetric in X, Y, and can be rewritten into:

$$I(X{:}Y|Z) = H(X|Z) - H(X|YZ) = H(Y|Z) - H(Y|XZ).$$

Using these measures of information, the dependency relationships defined in Section 2.1.3 on page 7 can be reformulated in an information-theoretic way.

- *Mutual independence*: X and Y are mutually independent ($X \perp Y$) iff $H(XY) = H(X) + H(Y)$ or $I(X{:}Y) = 0$ or $H(X|Y) = H(Y)$. Each identity has a perfectly intuitive physical meaning: The first identity means that the amount of uncertainty of the pair XY equals the sum of uncertainties of each component. That is another way of explaining the second identity which says that X and Y have no mutual information. The last identity expresses that knowing Y provides no information on X (so they must be independent).
- *Conditional independence*: $X \perp Y|Z$ iff $I(X{:}Y|Z) = 0$ or $H(XYZ) - H(Z) = H(XZ) + H(YZ)$. This can be seen as a relativized version (conditional to Z) of the previous item.
- *Functional dependence*: X is a function of Y iff $H(X|Y) = 0$, which expresses the intuition that given Y, we need 0 bits of information to recover X.

A general information measure can be deduced from the definition of Shannon's information measures.

Multivariate (conditional) mutual information. There are various ways of generalizing the previous definitions, the main two are :

$$I(X_1{:}X_2{:}\cdots{:}X_n) = \sum_{J \in [\![n]\!]} (-1)^{|J|} H(X_J),$$
$$I(X_1{:}X_2{:}\cdots{:}X_n|Y) = \sum_{J \in [\![n]\!]} (-1)^{|J|} H(X_J|Y),$$

and their equivalent definition using recursion:

$$I(X_1{:}X_2{:}\cdots{:}X_n) = I(X_1{:}X_2{:}\cdots{:}X_{n-1}) - I(X_1{:}X_2{:}\cdots{:}X_{n-1}|X_n),$$
$$I(X_1{:}X_2{:}\cdots{:}X_n|Y) = \mathop{\mathbb{E}}_{y}[I(X_1{:}X_2{:}\cdots{:}X_n|Y = y)].$$

These notations may be useful from time to time, however for most of them a physical meaning is lacking. For instance, Example 4 (page 8) shows that $I(A{:}B{:}C)$ can be negative.

2.2.2 Kullback–Leibler Information divergence

There is a more general measure of information called the Kullback–Leibler divergence. We will see that this information quantity is also non-negative. We further underline how it can be used to prove many inequalities for Shannon entropy.

Definition 3 (KL-divergence, [KL51])**.** *For two distributions p and q on the same sample space $\mathcal{W}$, the Kullback–Leibler divergence from p to q is defined as*

$$D(p\|q) = \sum_{w \in \mathcal{W}} p(w) \log \frac{p(w)}{q(w)} = \mathbb{E}\left[-\log \frac{q(w)}{p(w)}\right]$$

with the convention $p \log \frac{p}{0} = \infty$ *and* $0 \log 0 = 0$ *(justified by taking the limit).*

This information measure is non-symmetric, thus it is not a distance, hence the word "divergence". The intuitive interpretation for this measure is the following. Suppose we have two random variables $X \sim p$ and $Y \sim q$. The divergence from p to q is the average number of extra bits of information needed to encode X if we were using optimal codewords designed for Y instead of those for X.

Theorem 2 (Divergence Inequality (also known as Gibbs' inequality))**.** *The KL-divergence is always non-negative :*

$$\forall p, q, D(p||q) \geq 0$$

with equality iff $p = q$.

Proof. Using Jensen's inequality :

$$D(p\|q) = \mathbb{E}\left[-\log \frac{q(w)}{p(w)}\right] \geq \log\left(\sum_{w \in \mathcal{W}} p(w) \frac{q(w)}{p(w)}\right) \geq \log 1 = 0$$

□

Remark 1. *Notice that the proof only uses the fact that* $\sum_w q(w) \leq 1$. *The inequality is still true if* q *is not a probability distribution but sums to a value less than one. On the other hand, we can also weaken the requirement for* p *to be any function such that* $\sum_w p(w) \geq 1$, *shedding more light on the asymmetry of this information measure.*

2.3 Information Inequalities

A natural kind of inequalities for entropies are linear ones for they trigger to study the associated polytope delimited by linear inequalities, *i.e.*, half-planes. We call *information inequality* any linear inequality for entropies of tuples that holds for all distributions.

2.3.1 Shannon's Basic Inequality

The *basic inequality* is the most fundamental inequality of Information Theory. It was proven by Claude E. Shannon in his seminal papers.

Proposition 2 (Shannon's basic inequality, [Sha48a])**.** *For any jointly distributed random variables* X, Y, Z,

$$I(X{:}Y|Z) \geq 0. \tag{2.1}$$

In words, this inequality says that the mutual information is always non-negative, even conditionally to another random variable. The basic inequality implies, by instantiating variables accordingly, the non-negativity of all other Shannon Information measures.

$$H(A) \geq 0 \tag{2.2}$$
$$H(A|B) \geq 0 \tag{2.3}$$
$$I(A{:}B) \geq 0 \tag{2.4}$$

All of these inequality make intuitive sense and their physical meaning is well-understood.

Corollary 1. $0 \leq H(X|Y) \leq H(X) \leq H(XY) \leq H(X) + H(Y)$ *for all random variables* X, Y.

This chain of inequalities depicts the behavior of information for two random variables. The meaning of each inequality, from left to right, may be alternatively understood via the following statements:

(i) A random variable always has a non-negative amount of information.

(ii) Conditioning can only decrease uncertainty.

(iii) There is at least as much information in the whole as in the part.

(iv) The uncertainty of a tuple is at least the sum of uncertainties of its components.

However simple these statements can be, one needs to be careful in practice. The reader should be warned that conditional entropy should be understood "on average". For instance, it is very possible for some $y \in \mathcal{S}_Y$ that $H(X|Y = y) > H(X)$.

Proving the basic inequality. In order to prove Shannon's basic inequality, we will notice that some of Shannon's information measure are expressible as the KL-divergence of two well-chosen distributions.

Proof of Proposition 2. First, let us prove that the mutual information is non-negative. For this we will in fact prove the following identity:

$$I(X{:}Y) = D(p_{XY} \| p_X \cdot p_Y).$$

The result will immediately follow from the divergence inequality. The sum $p_X(x) \cdot p_Y(y)$ over all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ is 1, certifying that $p_X \cdot p_Y$ is indeed a probability distribution. Let us compute the divergence from p_{XY} to $p_X \cdot p_Y$:

$$\begin{aligned}
D(p_{XY} \| p_X \cdot p_Y) &\geq 0 \\
\mathbb{E}\left[\log \frac{p(x,y)}{p(x) \cdot p(y)}\right] &\geq 0 \\
\mathbb{E}\left[\log p(x,y)\right] - \mathbb{E}\left[\log p(x)\right] - \mathbb{E}\left[\log p(y)\right] &\geq 0 \\
\mathbb{E}\left[\log p(x,y)\right] - \mathbb{E}\left[\log p(x)\right] - \mathbb{E}\left[\log p(y)\right] &\geq 0 \\
-H(XY) + H(X) + H(Y) &\geq 0. \\
I(X{:}Y) &\geq 0
\end{aligned}$$

Now for the basic inequality, we need to see that

$$I(X{:}Y|Z) = D\left(p_{XYZ} \middle\| \frac{p_{XZ} \cdot p_{YZ}}{p_Z}\right).$$

That is, the conditional mutual information between X and Y given Z is the divergence from the joint distribution of (X, Y, Z) to some other distribution. Let us explicit more this second distribution that shall be hereafter denoted by q:

$$q(x,y,z) = \begin{cases} \frac{p(x,z) \cdot p(y,z)}{p(z)} & \text{if } p(z) > 0. \\ 0 & \text{otherwise.} \end{cases}$$

We can check that q is indeed a probability distribution :

$$\sum_{x,y,z} q(x,y,z) = \sum_{x} \sum_{\substack{y,z \\ p(z)>0}} \frac{p(x,z)\cdot p(y,z)}{p(z)} = \sum_{y} \sum_{\substack{z \\ p(z)>0}} \frac{p(z)\cdot p(y,z)}{p(z)} = \sum_{\substack{z \\ p(z)>0}} p(z) = 1.$$

Now, it is only a matter of term rewriting to verify that our identity hold, and thus finish the proof. □

2.3.2 (Un)conditional Information Inequalities

In Pippenger's words [Pip86], information inequalities are the "laws of information theory". Each inequality separates what can be achieved from what is impossible to reach in terms of entropy. Let $\{X_i\}_{i\in\mathcal{N}}$ be a collection of random variables indexed by a set $\mathcal{N}$ of n elements. For $J \subseteq \mathcal{N}$, we denote by X_J the set of random variables $\{X_j : j \in J\}$.

Definitions and Examples

Definition 4 ((Unconditional) Information Inequality)**.** *An unconditional linear information inequality for n-tuples of random variables is a linear form with* $2^n - 1$ *real coefficients* $(c_J)_{\varnothing\neq J\subseteq\mathcal{N}}$ *such that for all jointly distributed random variables* $\{X_i\}_{i\in\mathcal{N}}$

$$\sum_{\varnothing\neq J\subseteq\mathcal{N}} c_J H(X_J) \geq 0.$$

Definition 5 (Shannon-type Inequalities)**.** *We call* Shannon-type *inequalities the set of all positive linear combinations of instances of the basic inequality. That is, a valid inequality that can be put in the form*

$$\sum_{\substack{\varnothing\neq J\subseteq\mathcal{N} \\ \varnothing\neq K\subseteq\mathcal{N} \\ \varnothing\neq L\subseteq\mathcal{N}}} c_{J,K,L} I(X_J{:}X_K|X_L) \geq 0,$$

where all $c_{J,K,L}$ *are non-negative.*

Canonical form and coordinate systems An inequality written only in plain entropy terms $H(\cdot)$ is said to be in canonical form. One can use other coordinate systems by substituting plain entropies $H(X_1), H(X_2), H(X_{12}), \ldots$ by other information measures. (As long as all possible entropy are still expressible in the system.)

A useful coordinate system is the *atomic form* for n variables. By *atom* we mean an instance of a multivariate information measure (see page 12) that involves all single random variables of a tuple of n variables. Atoms can be written as $I(X_{j_1}{:}\cdots{:}X_{j_m}|X_{N\setminus J})$ for a non-empty subset $J \subseteq N$ of size m. One can rewrite plain entropies as a positive sum of atoms:

$$H(X_J) = \sum_{\substack{T=\{t_1,\ldots,t_k\} \\ T\cap J\neq\varnothing}} I(X_{t_1}{:}\cdots{:}X_{t_k}|X_{\mathcal{N}\setminus T}).$$

The latter equation can be seen as a generalized chain rule. Atoms are thus a suitable coordinate system, an inequality using involving only atoms will be called *in atomic form*.

Examples of unconditional information inequalities. Let us give a few more useful examples of Shannon-type inequalities.

Example 5. *For all random variables A, B, C, D,*

$$H(BC) \geq I(A{:}C|B) + I(B{:}D|AC) + H(BC|AD) \tag{2.5}$$

Proof. We check by term expansion that the following is an information equality:

$$H(BC) = I(A{:}C|B) + I(B{:}D|AC) + H(BC|AD) + I(A{:}B) + I(C{:}D|A).$$

All terms being non-negative, this implies inequality (2.5). □

Example 6. *For all random variables A, B, C, D, X,*

$$H(E) \leq H(E|A) + H(E|B) + I(A{:}B) \tag{2.6}$$
$$H(E|C) \leq H(E|A) + H(E|B) + I(A{:}B|C) \tag{2.7}$$
$$H(E) \leq 2H(E|C) + 2H(E|D) + I(C{:}D|A) + I(C{:}D|B) + I(A{:}B) \tag{2.8}$$

Proof. Put inequality (2.6) in canonical form

$$H(E) + H(AB) \leq H(AE) + H(BE),$$

Since $H(AB) \leq H(ABE)$, it is enough to check that

$$H(E) + H(ABE) \leq H(AE) + H(BE),$$

which is an instance of the basic inequality.

Inequality (2.7) is implied by (2.6) and Proposition 4 (proven hereafter). Indeed, by using Proposition 4 on inequality (2.7), we obtain

$$H(E|C) \leq H(E|AC) + H(E|BC) + I(A{:}B|C). \tag{2.9}$$

Inequality (2.9) follows since removing C in the condition increases the entropy.

Inequality (2.8), follows from the two previous inequalities. One can check that it is the sum of the following instances.

$$H(E) \leq H(E|A) + H(E|B) + I(A{:}B)$$
$$H(E|A) \leq H(E|C) + H(E|D) + I(C{:}D|A)$$
$$H(E|B) \leq H(E|C) + H(E|D) + I(C{:}D|B)$$

□

Conditional information inequalities. One can also look at information inequalities that are only valid when the distribution meets a list of constraints. The most natural type of constraints are linear information equalities. A conditional inequality is thus any inequality which is valid in a hyperplane defined by the condition.

Definition 6 (Conditional Information Inequality)**.** *Let $\alpha(X_{\mathcal{N}})$ and $\beta_1(X_{\mathcal{N}}), \ldots, \beta_m(X_{\mathcal{N}})$ be*

linear functions on the entropies of subtuples of $X_{\mathcal{N}}$:

$$\alpha(X_{\mathcal{N}}) = \sum_{\varnothing \neq J \subseteq \mathcal{N}} \alpha_J H(X_J),$$
$$\beta_i(X_{\mathcal{N}}) = \sum_{\varnothing \neq J \subseteq \mathcal{N}} \beta_{i,J} H(X_J), \text{ for } i \in [\![m]\!],$$

such that the implication

$$(\beta_i(X_{\mathcal{N}}) = 0 \text{ for } i \in [\![m]\!]) \Rightarrow \alpha(X_{\mathcal{N}}) \geq 0$$

holds for all distributions $X_{\mathcal{N}}$. We call this implication a conditional linear information inequality.

Example 7. *We give two examples of trivial conditional inequalities.*

(a) If $I(A{:}B) = 0$, then $H(A)+H(B) \leq H(AB)$. This follows immediately from the definition of the mutual information.

(b) If $I(A{:}B) = 0$, then $H(A) + H(B) + H(C) \leq H(AC) + H(BC)$. This follows from an unconditional Shannon-type inequality: for all A, B, C,

$$H(A) + H(B) + H(C) \leq H(AC) + H(BC) + I(A{:}B).$$

Which is the sum of two basic inequalities: $H(C|AB) \geq 0$ and $I(A{:}B|C) \geq 0$.

Example 8 (Common information)**.** *Let A, B, C, D be random variables and assume further that there exists a random variable X which is both a function of C and a function of D, and whose entropy is $H(X) = I(C{:}D)$. Using inequality (2.8) with $E = X$ and the previous conditions we obtain the following conditional inequality:*

$$H(X|C) = H(X|D) = H(X) - I(C{:}D) = 0 \Rightarrow I(C{:}D) \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B)$$

We will show that we can obtain the same inequality if another condition is met. For this we need the following lemma.

Lemma 1 (Romashchenko, [Rom03a])**.** *Let C, D, E be random variables. $I(C{:}E|D) = I(D{:}E|C) = I(C{:}D|E) = 0$ if and only if there exists a random variable X such that $H(X|C) = H(X|D) = H(X|E) = 0$ and $H(CDE|X) = H(C|X) + H(D|X) + H(E|X)$.*

From the statement of this Lemma we can see that X is a random variable satisfying the conditions of Example 8. Thus we get the following conditional inequality:

$$I(C{:}E|D) = I(D{:}E|C) = I(C{:}D|E) = 0 \Rightarrow I(C{:}D) \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B).$$

Proof of Lemma 1. Assume $I(C{:}E|D) = I(D{:}E|C) = I(C{:}D|E) = 0$. Define a random variable X on the same probability space by $X = p_{C|D=d,E=e}$. In words: the value of the random variable X is the conditional distribution function of C given the event that the pair (D, E) takes the value (d, e).

For now X is a function of (D, E). Moreover,

$p_{C\|X=x}$	$=$	$p_{C\|D=d,E=e}$	by definition
$p_{C\|X=x}$	$=$	$p_{C\|E=e}$	since $C \perp E\|D$
$p_{C\|X=x}$	$=$	$p_{C\|D=d}$	since $C \perp D\|E$

By definition of X, the last two equalities imply that X is a function of D and also a function of E. From inequality (2.7), X is also a function of C:

$$H(X|C) \leq H(X|D) + H(X|E) + I(D{:}E|C) = 0.$$

Now, the entropy of X can be computed using

$$H(X) = H(X|C) + I(X{:}C) = I(C{:}D) = I(C{:}D{:}E),$$

and thus

$$H(CDE|X) = H(CDE) - H(X) = H(CDE) - I(C{:}D{:}E) = H(C|X) + H(D|X) + H(E|X).$$

Conversely, suppose there exists a random variable X such that X is a function of each of C, D, E and C, D, E are mutually independent given X. We show that $I(D{:}E|C) = 0$ (the two other equalities being symmetric), by the following computation:

$$\begin{array}{rcl} H(CDE|X) & = & H(C|X) + H(D|C) + H(E|C) - I(D{:}E|C) \\ H(CDE|X) & \leq & H(C|X) + H(D|X) + H(E|X) - I(D{:}E|C) \\ H(CDE|X) & \leq & H(CDE|X) - I(D{:}E|C) \\ I(D{:}E|C) & \leq & 0 \end{array}$$

where first two inequalities hold by the functional dependency of X and the third follows from our assumption, the last inequality finishes the proof □

Motivation and Properties

We are introducing the study of information inequalities, but one should try first to understand what are *information equalities*? In fact, information equalities are all trivial.

Proposition 3. *If $f(X_N) = 0$ is an information equality, then f is identically null.*

Proof. Suppose $f(X_N)$ is a information equality written in atomic form, *i.e.*,

$$f(X_{\mathcal{N}}) = \sum_{\varnothing \neq J \subseteq \mathcal{N} J = \{j_1, \dots j_m\}} \lambda_J I(X_{j_1} : \cdots : X_{j_m} | X_{\mathcal{N} \setminus J}).$$

Let $Zero$ be a deterministic random variable and One be a random variable of entropy 1. For each $\varnothing \neq J \subseteq N$, define the tuple of random variable

$$Z_i^J = \begin{cases} Zero & \text{if } i \notin J. \\ One & \text{if } i \in J. \end{cases}$$

For any Z^J we have

$$f(Z^J) = \lambda_J = 0.$$

Therefore, all coefficients λ_J are equal to zero and thus f is the null function. □

Let us provide one more simple property on information inequalities, we show that if some inequality in canonical form is valid then its "relativized" version, where we add a fresh random variable in the condition, is also valid.

Proposition 4. *The inequality*

$$\sum_{\varnothing \neq J \subseteq N} c_J H(X_J) \geq 0$$

holds for all tuples $X_{\mathcal{N}}$ iff the inequality

$$\sum_{\varnothing \neq J \subseteq N} c_J H(X_J|Y) \geq 0$$

holds for all jointly distributed tuples $X_{\mathcal{N}}$ and variable Y.

Proof. For each value $y \in \mathcal{S}_Y$, there is a conditional distribution on the n-tuple X_N. By assumption our inequality on n-tuples holds for this distribution

$$\sum_{\varnothing \neq J \subseteq N} c_J H(X_J|Y = y) \geq 0.$$

Taking the expectation of this quantity over all $y \in S_Y$, we obtain the inequality.

$$\sum_{\varnothing \neq J \subseteq N} c_J H(X_J|Y) \geq 0.$$

The converse trivially holds by taking a deterministic Y. □

2.4 Venn Information Diagrams

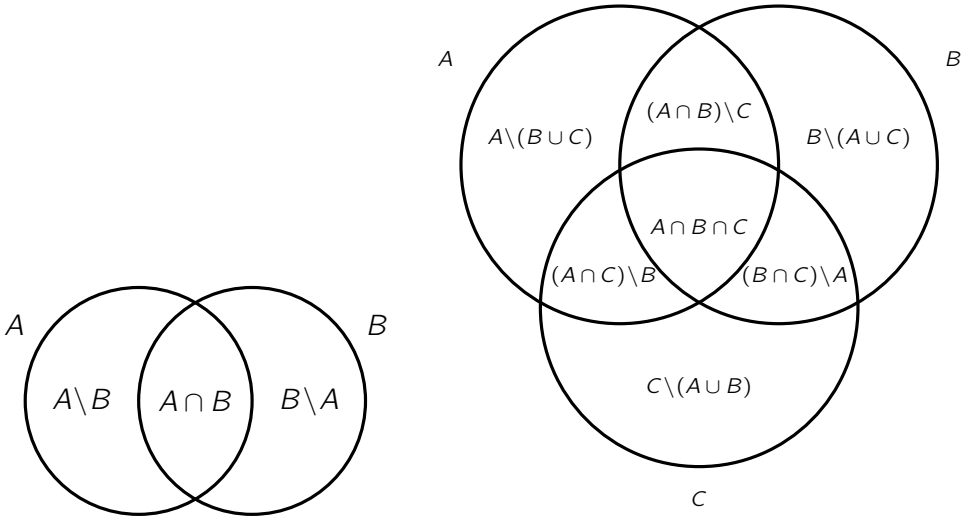
We make a brief parenthesis on Venn diagrams, the interested reader might want to consider the online survey of [RW05].

2.4.1 Information Diagrams

Definition 7 (n-Venn Diagram, [RW05])**.** *Let $C = \{C_1, C_2, \ldots, C_n\}$ be a set of simple closed curves in the plane. Let R_i^{in} be the region enclosed in the interior of C_i and R_i^{out} be the region enclosed in the exterior of C_i. The collection C is called an* independent family *if each of the 2^n regions $\bigcap_{i \in [\![n]\!]} X_i$ is nonempty when X_i is either R_i^{in} or R_i^{out}. If, in addition, each such region is connected and there are only finitely many points of intersection between curves, then C is a* Venn diagram*, or an n-*Venn diagram*.*

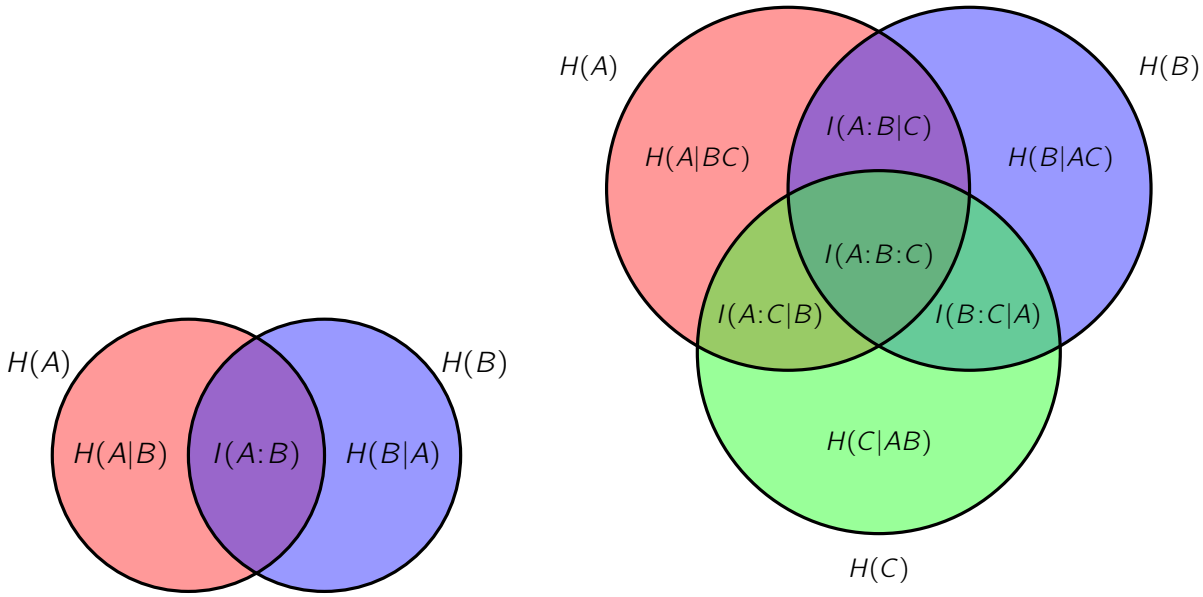
Interpreting each closed curved of a Venn diagram as a set, the diagram should contain every possible intersections. This can be seen, for a few particular cases, in the examples shown in Figure 2.5.

Using Venn-diagrams, one can pictorially represent the information shared by tuples of random variables. These types of diagrams are called (Venn) information diagrams. Each closed curve represents – the Shannon entropy of – a random variable, and each region represents a corresponding information quantity. For instance, the region containing exactly A and B – and not in X and Y, represents the quantity $I(A{:}B|XY)$. In general, information diagrams display the relation between the two coordinate systems defined in Section 2.3.2 (p. 15): the canonical form and the atomic form.



(a) a 2-Venn Diagram (b) a 3-Venn Diagram

Figure 2.5: Examples of Venn Diagrams



(a) For two random variables (b) For three random variables

Figure 2.6: Example of Information Diagrams

For $n = 4$ random variable we will use a diagram inspired of Charles L. Dodgson[3] (see [Dod87]). We describe in Figure 2.7 how the diagram is constructed and how it should be understood and used.

[3] also known as Lewis Carroll.

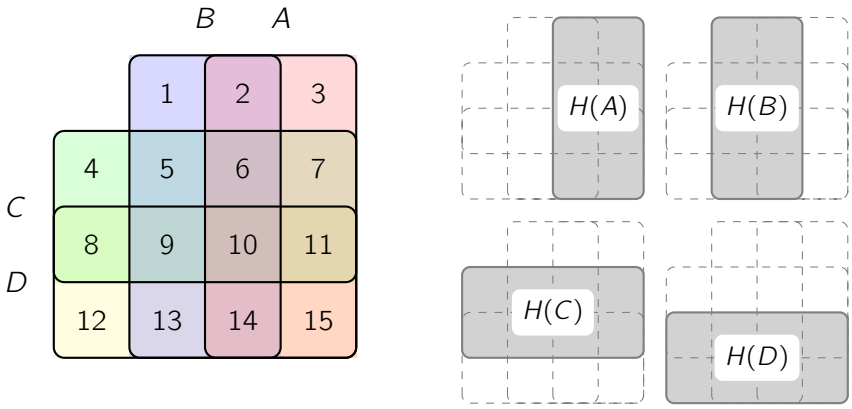


Figure 2.7: An Information Diagram for 4 random variables

The regions numbered 1,3,4,12 correspond to atoms of the type $H(W|XYZ)$ where $W, X, Y, Z \in \{A, B, C, D\}$. Thus they correspond to non-negative conditional entropies. The only numbered regions that may be negative are 6,9,10,11,14.

Remark 2. *We suggest the reader to find all other (non-atomic) regions for which the sum of its components is non-negative due to the non-negativity of the corresponding instance of the basic inequality.*

2.4.2 Proving Inequalities Without Words

How can we certify that an (unconditional) information inequality is Shannon-type ? Recall that a Shannon-type inequality is just a non-negative linear combination of instances of the basic inequality. One could guess the coefficients but there is something better to do: Ask a computer.

Indeed, for n random variables, consider the convex polytope defined by all Shannon-type information inequalities. This polytope is generated by all instances of the basic inequality for n random variables. The very fact that the non-negativity of linear function f is implied by Shannon-type inequalities means that the hyperplane defined by $f \geq 0$ contains this polytope. This can be checked using a reformulation into a Linear Program: we want in fact to minimize f over the convex polytope defined by all Shannon-type inequalities and show that the minimum of f is zero.

Linear Programming has been shown to be in P[4]. Various computer programs, such as `ITIP`, `XITIP`, (see [YY, PPD]) – based upon LP libraries, use these specific LP programs to prove information inequalities.

For n up to 4 or 5, one can try to prove that an inequality is Shannon-type by hand, or with the help of information diagrams. We show in Figure 2.8 a picture proving the 4-variable Shannon-type inequality (2.5):

$$H(BC) \geq I(A{:}C|B) + I(B{:}D|AC) + H(BC|AD).$$

In the picture, the region representing $H(BC)$ is covered by regions representing conditional mutual information quantities, this proves an information **equality**. *The white regions being always non-negative, $H(BC)$ is thus greater than the sum of quantities represented by the light red regions.*

[4] Although it is in P, the size of the LP program for n random variables grows exponentially with n in general.

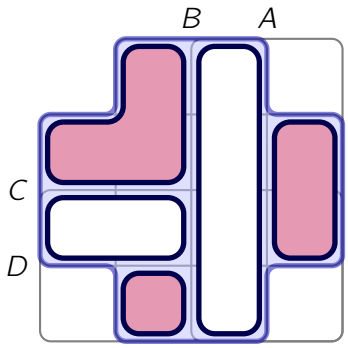


Figure 2.8: Proof of the Shannon-type inequality $H(BC) \geq I(A{:}C|B) + I(B{:}D|AC) + H(BC|AD)$

Remark 3. *Most of the proofs involving information inequalities with few variables can be proven using this visual technique.*

2.5 Linear Random Variables

An important class of random variables is that of *linear random variables*. This class is based on vector spaces and their subspaces - the very basic objects of linear algebra. For simplicity we assume our vector spaces to be finite: the underlying field will be $\mathbb{F}_q$ for some prime power q and the dimension is an integer. A single vector space V of dimension d is then isomorphic to $\mathbb{F}_q^d$. A natural way to define a random variable from V could be to take a uniformly random element from V. The set of elementary events of this probability space consists of the set of vectors in V and the distribution charges each vector with the same probability.

A more elegant (and simpler) way to generalize this construction to tuples is to use vector space duality. We present the construction of a corresponding tuple of random variables from [HRSV00]. More about the relation between random variables and vector spaces can be found in [AKS10, DFZ09].

In a vector space V over the field $\mathbb{F}$, let $(V_i)_{i\in\mathcal{N}}$ be a n-tuple of linear subspaces of V. As usual, we denote by V_J the set $\{V_j\}_{j\in J}$ for $J \subseteq \mathcal{N}$. For each subspace V_i, we construct a corresponding random variable α_i, such that $H(\alpha_J)$ is proportional to $\mathrm{rk}(V_J)$ for all $J \subseteq \mathcal{N}$. We proceed in the following way:

First, we pick uniformly at random a linear function $\alpha : V \to \mathbb{F}$ over all possible such linear functions. The random variable associated to the subspace V_i is defined by $\alpha_i = \alpha_{|V_i}$: the restriction of α to V_i. For $i, j \in \mathcal{N}$, the pair $(\alpha_{|V_i}, \alpha_{|V_j})$ has the same distribution as $\alpha_{|V_i+V_j}$. The same is true for triples, quadruples, quintuples, and so on: generally, for each subset $J \subseteq \mathcal{N}$, the tuple $(\alpha_{|V_i})_{i\in J}$ has the same distribution as $\alpha_{|V_J}$. Moreover, for any linear subspace L of V, the random variable $\alpha_{|L}$ is uniform and takes $|\mathbb{F}|^{\mathrm{rk}(L)}$ different values, *i.e.*, $H(\alpha_{|L}) = \mathrm{rk}(L) \cdot \log|\mathbb{F}|$.

This construction can be generalized to groups (and subgroups) as was studied in [CY02].

2.5.1 Intersection of Vector Spaces

For any two subspaces A, B, the set defined by

$$A + B = \{\lambda a + \mu b : a \in A, b \in B, \lambda, \mu \in \mathbb{F}\}$$

is called the sum of A and B and is also a vector subspace. The sum is called *direct* whenever $A \cap B = \{\varnothing\}$. Vector spaces enjoy another nice structural property : their intersections also makes a vector space, *i.e.*, the set of vectors $A \cap B$ is a vector subspace.

The ranks of the sum and of the intersection of two vector subspaces are related by the following formula.

Proposition 5 (Grassmann's formula, [Gra44])**.** *For every vector subspaces A and B,*

$$\mathrm{rk}(A+B) + \mathrm{rk}(A \cap B) = \mathrm{rk}(A) + \mathrm{rk}(B).$$

Remark 4. *This formula does* **not** *generalize to an inclusion/exclusion principle for more than 2 subspaces!*

$$\begin{gathered}\mathrm{rk}(A \cup B \cup C) \\ \neq \\ \mathrm{rk}(A) + \mathrm{rk}(B) + \mathrm{rk}(C) - \mathrm{rk}(A \cap B) - \mathrm{rk}(A \cap C) - \mathrm{rk}(B \cap C) + \mathrm{rk}(A \cap B \cap C)\end{gathered}$$

Take for instance three distinct lines meeting at the origin of the (affine) plane

2.5.2 Ingleton Inequality

An important inequality for vector spaces – it is also crucial for the study of random variables as we shall see – is Ingleton inequality. It was discovered by A. W. Ingleton, initially motivated by the study of matroid linear representations using vector spaces (see [Ing71]).

Proposition 6 (Ingleton Inequality, [Ing71])**.** *For any vector spaces A, B, C, D, the following rank inequality holds*

$$\begin{aligned}\mathrm{rk}(C) + \mathrm{rk}(D) + \mathrm{rk}(A+B) + \mathrm{rk}(A+C+D) + \mathrm{rk}(B+C+D) \leq \\ \leq \mathrm{rk}(C+D) + \mathrm{rk}(A+C) + \mathrm{rk}(A+D) + \mathrm{rk}(B+C) + \mathrm{rk}(B+D)\end{aligned} \quad (2.10)$$

This inequality is thus valid for linear random variables if we rewrite it using Shannon's information measures.

- The rank of a subspace corresponds to the entropy of its associated random variable.
- $I(A{:}B)$ represents the rank of the intersection of the corresponding subspaces $A \cap B$.
- $I(A{:}B|C)$ represents the rank of the intersection of A/C and B/C (*i.e.*, A and B factorized over C).

Thus, inequality 2.10 rewrites into

$$I(C{:}D) \leq I(C{:}D|B) + I(C{:}D|A) + I(A{:}B) \quad (2.11)$$

Note that this inequality does not hold in general for random variables.

Proof of Proposition 6. We will prove the inequality for linear random variables. In Example 8 we proved the following conditional inequality:

$$H(X|C) = H(X|D) = H(X) - I(C{:}D) = 0 \Rightarrow I(C{:}D) \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B)$$

Now suppose X is the random variable associated to the subspace $C \cap D$. The conditions on the left-hand side hold, thus the inequality in the right hand side is true. Since X is not involved in this inequality, it must hold without conditions. □

What happened in the previous proof? We have just shown that the intersection operator is a very powerful tool: it extracts the common information. For any two linear random variables A, B, the mutual information between A and B can be extracted, or materialized, in its integrality into another linear random variable. Linear random variables are an example where the common information and the mutual information are really the same quantity.

2.6 The Entropy Region Framework

Zhang and Yeung seminal papers from the late 1990's (see [ZY97, ZY98]) leading to the discovery of conditional and unconditional non-Shannon-type inequalities mapped out the start of the search for a complete characterization of the entropy space. Before this turning point, all known inequalities were plain consequences of Shannon's basic inequality. The possibility that Shannon-type inequalities were the only ones was still considered. This possibility is excluded by the very existence of a non-Shannon-type inequality, which, alone, would justify the study of information inequalities in a much general framework. This framework of the Entropy Region was introduced by Zhang and Yeung and further developed in Yeung's textbook [Yeu08] on Information Theory.

2.6.1 Entropic Regions

We are to defined the Entropy Region. An element of this set is a vector of entropies describing a tuple of random variables.

Definition 8 (Entropy profile)**.** *For a tuple $X_{\mathcal{N}}$ of n jointly distributed random variables, its entropy profile*

$$\vec{H}(X_{\mathcal{N}}) = (H(X_1), H(X_2), \dots, H(X_1, X_2), H(X_1, X_3), \dots, H(X_1, X_2, \dots, X_n))$$

is defined as the shortlex-ordered[5] *list of entropies of all subtuples X_J for each non-empty $J \subseteq \mathcal{N}$.*

Of course, a single vector can possibly be the entropy profile of many quite different distributions.

Consider the $(2^n - 1)$-dimensional real Euclidean space $\mathbb{R}^{2^{\mathcal{N}}\setminus\{\varnothing\}}$. Given a point $\vec{h} \in \mathbb{R}^{2^{\mathcal{N}}\setminus\{\varnothing\}}$, we shall denote by h_J its projection onto the J-coordinates for $\varnothing \neq J \subseteq \mathcal{N}$.

- A point $\vec{h}$ is called *entropic* if it is the entropy profile of some tuple of random variables, *i.e.*, such that $\vec{h} = \vec{H}(X_{\mathcal{N}})$ for a tuple $X_{\mathcal{N}}$.
- A point is called *almost entropic* if it is the limit of a sequence of entropy profiles.
- A point is called *linearly entropic* if it is the entropy profile of a tuple of linear random variables.

We are now ready to define subsets of the Euclidean space which are of interest.

Definition 9 (Entropy Regions)**.** *Let n be a positive integer.*

Γ_n^* is the set of all entropic points,

$$\Gamma_n^* = \{h \in \mathbb{R}^{\mathscr{P}(\mathcal{N})\setminus\{\varnothing\}} : h \text{ is entropic } \}$$

[5] First sort by shortest length then lexicographically for each length.

$\overline{\Gamma}_n^*$ *is the set of all almost entropic points*

$$\overline{\Gamma}_n^* = \{h \in \mathbb{R}^{\mathscr{P}(\mathcal{N})\setminus\{\varnothing\}} : h \text{ is almost entropic }\}$$

Γ_n *is the set of all points satisfying all Shannon-type inequalities for n random variables,*

$$\Gamma_n = \{h \in \mathbb{R}^{\mathscr{P}(\mathcal{N})\setminus\{\varnothing\}} : h \text{ satisfies all Shannon-type inequalities }\}$$

L_n *is the set of all linearly entropic points,*

$$L_n = \{h \in \mathbb{R}^{\mathscr{P}(\mathcal{N})\setminus\{\varnothing\}} : h \text{ is a linearly entropic point }\}$$

The following property is immediate:

Proposition 7. $L_n \subseteq \Gamma_n^* \subseteq \overline{\Gamma}_n^* \subseteq \Gamma_n$, *for all positive integer n.*

Proof. By definition, $\overline{\Gamma}_n^*$ is the closure of Γ_n^*, thus $\Gamma_n^* \subseteq \overline{\Gamma}_n^*$. Further, entropic points satisfy all Shannon-type inequalities, so the same must be true for the limit of any sequence of entropic points, *i.e.*, for almost entropic points. □

2.6.2 Geometric Properties and Cones

Some of the previous regions enjoy some nice topological properties.

Proposition 8 (Closure under Sums)**.** *The sum of two entropic points is an entropic point.*

Proof. Let $\vec{H}(x_\mathcal{N})$ and $\vec{H}(y_\mathcal{N})$ are respectively the entropy profiles of the tuples $x_\mathcal{N}$ and $y_\mathcal{N}$, and further assume that these tuples are independent. To see that $\vec{H}(z_\mathcal{N}) = \vec{H}(x_\mathcal{N}) + \vec{H}(y_\mathcal{N})$ is also an entropy profile, consider the variable $Z_\mathcal{N}$ defined by $z_i = (x_i, y_i)$.

Thus, we have $H(z_J) = H(x_J) + H(y_J)$ for any $J \subseteq \mathcal{N}$. So for this $z_\mathcal{N}$, $\vec{H}(z_\mathcal{N}) = \vec{H}(x_\mathcal{N}) + \vec{H}(y_\mathcal{N})$. □

Corollary 2. *The point* $k \cdot h + k' \cdot h'$ *is entropic, for all non-negative integers* k, k' *and entropic points* h, h'.

Definition 10 (N-serialization)**.** *For a random variable x, we can consider the tuple formed by N independent and identically distributed copies of x, denoted by* $X = (x^1, x^2, \ldots, x^N)$. *For a tuple of random variables* $(x, y, z, \ldots)$, *we call N-serialization the tuple* $(X, Y, Z, \ldots)$ *where each component consist of N independent copies of the corresponding component in the original tuple.*

Definition 11 (Convex cone)**.** *Let C be a subset of a vector space,*

- *C is a cone if* $\alpha C \subset C$ *for any non-negative scalar* α,
- *C is convex if* $\lambda C + (1-\lambda)C \subset C$ *for all* $0 \leq \lambda \leq 1$,
- *C is convex cone if* $\alpha C + \beta C \subset C$ *for non-negative scalars* α, β.

For any subset E of a vector space, one can construct $\mathbf{con}(E)$ *the smallest convex cone containing E.*

Definition 12 (Dual Convex Cone)**.** *To any subset E of a vector space V can be associated its dual convex cone:*

$$\mathbf{dual}(E) := \{v \in V^* : \forall w \in E, v(w) \geq 0\},$$

where V^* *is the dual vector space of V.*

The set Γ_n of all points satisfying Shannon-type inequalities is a closed convex cone. This is because the solutions of a system of linear inequalities form a closed convex cone. The dual set **dual**(Γ_n) consists of the set of coefficients of all valid Shannon-type inequalities, it is clearly a closed convex cone. It is sometimes simpler to think in terms of dual cones.

Proposition 9. *$\overline{\Gamma}_n^*$ is a convex cone.*

Proof. Assume first that $\vec{H}(x_{\mathcal{N}})$ and $\vec{H}(y_{\mathcal{N}})$ are two entropy profiles for tuples $x_{\mathcal{N}}$ and (resp.) $y_{\mathcal{N}}$. We must understand why the point $\alpha \cdot \vec{H}(x_{\mathcal{N}}) + \beta \cdot \vec{H}(y_{\mathcal{N}})$ is almost entropic for every $\alpha, \beta \in \mathbb{R}^+$.

For this we construct two variables, a tuple $Z_{\mathcal{N}}$ depending on a "selector" S. Define $X_{\mathcal{N}}$ and $Y_{\mathcal{N}}$ to be n-serializations of $x_{\mathcal{N}}$ and $y_{\mathcal{N}}$ respectively.

$$S = \begin{cases} 0 & \text{with probability } 1-p-q. \\ 1 & \text{with probability } p. \\ 2 & \text{with probability } q. \end{cases} \quad \text{and} \quad Z_i = \begin{cases} 0 & \text{if } S = 0. \\ X_i & \text{if } S = 1. \\ Y_i & \text{if } S = 2. \end{cases}$$

For each subtuple indexed by $\varnothing \neq J \subseteq \mathcal{N}$, we bound the entropy of Z_J using basic inequalities:

$$H(Z_J|S) \leq H(Z) \leq H(S) + H(Z_J|S).$$

We can compute $H(Z_J|S) = np \cdot H(X_J) + nq \cdot H(Y_J)$.

We immediately see that we should take $np = \alpha$ and $nq = \beta$. As the size n-serialization grows, the entropy of S becomes negligible (S tends to a deterministic random variable), while $H(Z_J) \to \alpha \cdot H(X_J) + \beta \cdot H(Y_J)$.

We have proven the theorem for entropic points. The result follows for almost entropic points since they are limits of entropic points. □

2.6.3 Characterization of Entropy Regions

Trivially, we have $\Gamma_1^* = \overline{\Gamma}_1^* = \Gamma_1 = \mathbb{R}$. Indeed, let X be a random variable defined on an alphabet of size m, then the entropy of $H(X)$ belongs to $[0, \log m]$ and can take any value in this interval since it is continuous. By taking m large enough, any point $(a) \in \mathbb{R}$ can be the entropy profile of a single random variable.

Characterization For Two Random Variables

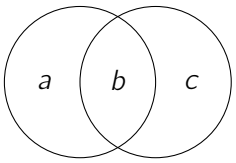


Figure 2.9: Representation of the region for two random variables

Theorem 3. $\Gamma_2^* = \overline{\Gamma}_2^* = \Gamma_2$

Proof. By Proposition 7, we only have to show that $\Gamma_2 \subseteq \Gamma_2^*$. Let a, b, c be any non-negative reals and A, B, C be random mutually independent random variables such that $H(A) = a$, $H(B) = b$ and $H(C) = c$. Let $U = (A, B)$ and $V = (B, C)$, then the entropy profile of (U, V) is the one represented in Figure 2.9. □

For Three Random Variables

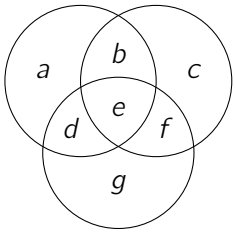


Figure 2.10: Representation of the region for three random variables

Lemma 2. *Let $m \in \mathbb{R}$. The point $\vec{h} = (m, m, m, 2m, 2m, 2m, 2m)$ is entropic iff $m = \log M$ for some positive integer $M \in \mathbb{N}^*$.*

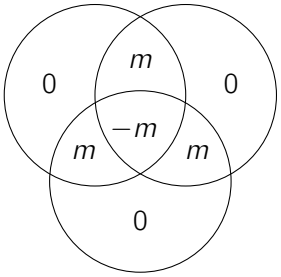


Figure 2.11: The point $\vec{h} = (m, m, m, 2m, 2m, 2m, 2m)$

Proof. Let x, y, z be random variables whose profile is $\vec{h}$.

For any $(x_i, y) \in \mathcal{S}_x \times \mathcal{S}_y$ there exists a unique z such that

$$p(x_i, y) = p(x_i, y, z)$$

Moreover

$$p(x_i)p(z) = p(x_i, z) = p(x_i, y, z) = p(y, z) = p(y)p(z).$$

So $p(x_i) = p(y)$ for any x_i in the support of x. This means that x is uniformly distributed over $\mathcal{S}_x$. By symmetry of the distribution, the same holds for y and z. Summarizing:

$$H(x) = H(y) = H(z) = \log |S_x| = \log |S_y| = \log |S_z| = m.$$

This means that the size of the supports of our three random variables is some integer M such that $m = \log M$.

The point $\vec{h}$ is entropic since it is the entropy profile of the variables in Example 4 (p. 8). □

And now the characterization result:

Theorem 4. $\Gamma_3^* \subset \overline{\Gamma}_3^* = \Gamma_3$

Proof. $\Gamma_3^* \neq \overline{\Gamma}_3^*$ follows immediately from Lemma 2. It remains to see why $\overline{\Gamma}_3^* = \Gamma_3$.

For simplicity, we switch to atomic coordinates. In this coordinate system, each of the

vectors

$$\begin{aligned}\vec{e}_1 &= (1,0,0,0,0,0,0),\\ \vec{e}_2 &= (0,1,0,0,0,0,0),\\ \vec{e}_3 &= (0,0,1,0,0,0,0),\\ \vec{e}_{12} &= (0,0,0,1,0,0,0),\\ \vec{e}_{13} &= (0,0,0,0,1,0,0),\\ \vec{e}_{23} &= (0,0,0,0,0,1,0),\\ \vec{e}_{123} &= (0,0,0,0,0,0,1),\end{aligned}$$

is entropic. For instance $\vec{e}_3$ corresponds to the quantity $H(X_3|X_1X_2)$. Let $J \subseteq \{1,2,3\}$, $Zero$ be a deterministic random variable and One be a uniform bit, set for each $i \in \{1,2,3\}$,

$$X_i = \begin{cases} One & \text{if } i \in J.\\ Zero & \text{if } i \notin J.\end{cases}$$

However, in this coordinate system some quantities can be negative. The only possibly negative quantity is $I(X_1{:}X_2{:}X_3)$, which is associated to the vector $\vec{e}_{123}$. The idea is to use the profile $\vec{g} = (0,0,0,m,m,m,-m)$, entropic by Lemma 2.

Take a vector $\vec{h} \in \Gamma_3$, in our basis

$$\vec{h} = \sum_{\varnothing \neq J \subseteq [\![3]\!]} h_J \cdot \vec{e}_J.$$

Instead, we write $\vec{h}$ as

$$\vec{h} = h_1\vec{e}_1 + h_2\vec{e}_2 + h_3\vec{e}_3 + (h_{12}-g_{12})\vec{e}_{12} + (h_{13}-g_{13})\vec{e}_{13} + (h_{23}-g_{23})\vec{e}_{23} + h_{123}\vec{g}$$

Since all vector in this decomposition are entropic except possibly $\vec{g}$ which is almost entropic: $\vec{h} \in \bar{\Gamma}^*_n$. □

František Matúš further investigated the region Γ^*_3 and confirmed that this region is indeed more complicated, by extending the result of Lemma 2. The lemma implies that Γ^*_3 is not a convex cone. However it could be that the "difference" between this set and its closure is not significant. In fact, this difference *indeed is* significant! In [Mat06], Matúš proved that on some facet F of the convex cone $\bar{\Gamma}^*_n$, the set Γ^*_n intersects F at a region delimited by a piecewise-linear function. This result was recently further developed in [HCG12, YC12].

2.7 Non-Shannon Information Inequalities

The last result hinted what was possibly coming for bigger tuples. For four or more random variables, the situation becomes more delicate.

Theorem 5 (Zhang and Yeung, [ZY98])**.** $\Gamma^*_n \subsetneq \overline{\Gamma}^*_n \subsetneq \Gamma_n$ *for* $n \geq 4$.

This theorem implies the existence of non-Shannon-type inequalities. The following second result by F. Matúš definitively proves the difficulty of the characterization of the convex cone of all almost entropic points.

Theorem 6 (No finite linear characterization, [Mat07b])**.** *$\bar{\Gamma}_n^*$ is not a finitely generated for $n \geq 4$, i.e., there are at least a countable infinite number of independent information inequalities.*

Still, the situation is manageable for the set L_4. Hammer *et al* characterized in [HRSV00] the set of inequalities in 4 variables for linear random variables.

Theorem 7 (Characterization of $\mathbf{con}(L_4)$)**.** $\mathbf{dual}(L_4)$ *consists of all linear combinations of Shannon-type inequalities and the Ingleton inequality.*

2.7.1 Non-Shannon-type Conditional Inequalities

The timeline of the quest for non-Shannon-type inequalities begins with the discovery of Zhang-Yeung's conditional inequality.

Theorem 8 (Zhang-Yeung's conditional information inequality, [ZY97])**.**

$$I(A{:}B|C) = 0 \text{ and } I(A{:}B) = 0 \Leftrightarrow I(C{:}D) \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B).$$

Proof. We prove this inequality using the KL-divergence. The first distribution we use is

$$p'(a,b,c,d) = \begin{cases} \frac{p(a,c,d)\cdot p(b,c,d)}{p(c,d)} & \text{if } p(c,d) > 0. \\ 0 & \text{otherwise.} \end{cases}$$

After summing p' over a, b, c, d in that order, we see that p' is a well-defined distribution. Remember that we can express the assumption $I(A{:}B|C) = I(A{:}B) = 0$ as

$$p(a,b,c) \cdot p(c) = p(a,c) \cdot p(b,c)$$
$$p(a,b) = p(a) \cdot p(b)$$

Consider the function q defined as follows :

$$q(a,b,c,d) = \begin{cases} \frac{p(a,c)\cdot p(b,c)\cdot p(a,d)\cdot p(b,d)}{p(a)\cdot p(b)\cdot p(c)\cdot p(d)} & \text{if } p(a)\cdot p(b)\cdot p(c)\cdot p(d) > 0. \\ 0 & \text{otherwise.} \end{cases}$$

By using both assumptions we can rewrite q into

$$q(a,b,c,d) = \frac{p(a,b,c)\cdot p(a,d)\cdot p(b,d)}{p(a,b)\cdot p(d)}$$

Summing over c, a, b, d in that order, reveals that q is a distribution:

$$\sum_{a,b,c,d} q(a,b,c,d) = \sum_{\substack{a,b,d \\ p(a,b)>0 \\ p(d)>0}} \sum_{c} \frac{p(a,b,c)\cdot p(a,d)\cdot p(b,d)}{p(a,b)\cdot p(d)} = \sum_{\substack{b,d \\ p(d)>0}} \sum_{a} \frac{p(a,d)\cdot p(b,d)}{p(d)} = \sum_{b,d} p(b,d) = 1.$$

Since the support of p' is included in the support of q the divergence inequality is not trivial.

$$D(p'||q) \geq 0$$
$$\sum_{a,b,c,d} \frac{p(a,c,d)\cdot p(b,c,d)}{p(c,d)} \log \frac{p(a,c,d)\cdot p(b,c,d)\cdot p(a)\cdot p(b)\cdot p(c)\cdot p(d)}{p(c,d)\cdot p(a,c)\cdot p(b,c)\cdot p(a,d)\cdot p(b,d)} \geq 0$$

One can verify that it can be expanded into entropies. For instance if we split the logarithm term, the following subterm can be computed as follows:

$$\begin{aligned}\sum_{a,b,c,d} \frac{p(a,c,d)\cdot p(b,c,d)}{p(c,d)} \log \frac{1}{p(a,c)} &= \sum_{a,c,d} \frac{p(a,c,d)\cdot p(c,d)}{p(c,d)} \log \frac{1}{p(a,c)} \\ &= \sum_{a,c,d} p(a,c,d) \log \frac{1}{p(a,c)} \\ &= \sum_{a,c} p(a,c) \log \frac{1}{p(a,c)} \\ &= H(AC)\end{aligned}$$

Computing all subterms gives

$$I(C:D) \leq I(C{:}D|A) + I(C{:}D|B).$$

□

A second conditional inequality, with a similar proof, was discovered by Matúš.

Theorem 9 (Matúš, [Mat99a])**.** *If* $I(A{:}B|C) = 0$ *and* $I(B{:}D|C) = 0$*, then*

$$I(C{:}D) \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B).$$

Proof. Reuse the distribution p' from the previous proof:

$$p'(a,b,c,d) = \begin{cases} \frac{p(a,c,d)\cdot p(b,c,d)}{p(c,d)} & \text{if } p(c,d) > 0. \\ 0 & \text{otherwise.} \end{cases}$$

Notice that the conditions mean:

$$\begin{aligned} p(abc)\cdot p(c) &= p(ac)\cdot p(bc), \\ p(bcd)\cdot p(c) &= p(bc)\cdot p(cd), \end{aligned}$$

and therefore p' has another equivalent form:

$$p'(a,b,c,d) = \begin{cases} \frac{p(a,c,d)\cdot p(a,b,c)}{p(a,c)} & \text{if } p(a,c) > 0. \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

Now take,

$$q(a,b,c,d) = \begin{cases} \frac{p(a,b,c)\cdot p(a,d)\cdot p(b,d)}{p(a,b)\cdot p(d)} & \text{if } p(a,b)\cdot p(d) > 0. \\ \frac{p(c)\cdot p(a,d)\cdot p(b,d)}{p(d)} & \text{if } p(a,b) = 0 \text{ and } p(d) > 0. \\ 0 & \text{otherwise.} \end{cases}$$

This q is a probability distribution. In the KL-divergence from p' to q, the sum when $p(a,b) = 0$ is null (because of equation (2.12)). $D(p'||q)$ thus reduces to:

$$\sum_{a,b,c,d} \frac{p(a,c,d)\cdot p(b,c,d)}{p(c,d)} \log \frac{p(a,c,d)\cdot p(b,c,d)\cdot p(a,b)\cdot p(d)}{p(c,d)\cdot p(a,b,c)\cdot p(a,d)\cdot p(b,d)} \geq 0$$

Splitting the logarithm, every terms rewrites to an entropy quantity. For instance, the trick

to obtain the terms $H(AB)$ and $H(ABC)$ is to use the second form (2.12) of p'. Thus, the KL-divergence from p' to q rewrites to

$$-H(ACD) - H(BCD) - H(AB) - H(D) + H(CD) + H(ABC) + H(AD) + H(BD) \geq 0.$$

Adding $I(A{:}B|C) \geq 0$ to the last inequality finishes the proof. □

For this third inequality we provide two different proofs, one in the style of the two previous ones, and one which does not use the KL-divergence inequality.

Theorem 10 (K., Romashchenko, [KR11])**.** *if* $I(A{:}B|C) = 0$ *and* $H(C|AB) = 0$*, then*

$$I(C{:}D) \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B).$$

First presentation of the proof. Let $p(a, b, c, d)$ denote the distribution of discrete random variables (a, b, c, d) for which $H(c|ab) = I(a{:}b|c) = 0$.

Let us define two new distributions, $p'(a, b, c, d)$ and $q(a, b, c, d)$ as follows:

$$p'(a, b, c, d) = \frac{p(a, c, d) \cdot p(b, c, d)}{p(c, d)}$$

and

$$q(a, b, c, d) = \begin{cases} \frac{p(a,d)\cdot p(b,d)}{p(d)} & \text{if } p(a, b, c) \cdot p(d) > 0. \\ 0 & \text{otherwise.} \end{cases}$$

Since $I(a{:}b|c) = 0$, the condition $p(a, b, c) > 0$ is true if and only if $p(a, c) > 0$ and $p(b, c) > 0$. We notice that q is <u>not</u> a distribution, however $\sum_{a,b,c,d} q(a, b, c, d) \leq 1$. Therefore the KL-divergence inequality still holds (see Remark 1) and thus:

$$0 \leq D(p'||q) = \sum_{\substack{a,b,c,d \\ p(a,b,c)>0}} \frac{p(a, c, d) \cdot p(b, c, d)}{p(c, d)} \cdot \log \frac{p(a, c, d) \cdot p(b, c, d) \cdot p(d)}{p(c, d) \cdot p(a, d) \cdot p(b, d)}.$$

It follows immediately that

$$0 \leq H(AD) + H(BD) + H(CD) - H(ACD) - H(BCD) - H(D).$$

Now we add the values $I(A{:}B|C) = H(AC) + H(BC) - H(ABC) - H(C)$ and $H(C|AB) = H(ABC) - H(AB)$ to the right-hand side of the inequality (both these values are equal to 0 for our distribution) to obtain the desired

$$0 \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B) - I(C{:}D).$$

□

We give a second proof which does not use the KL-divergence.

Second Proof of Theorem 10. The argument consists of two steps: "enforcing conditional independence" and "elimination of conditional entropy". Let us have a joint distribution of random variables a, b, c, d. The first trick of the argument is a suitable transformation of this distribution. We keep the same distribution on the triples (a, c, d) and (b, c, d) but make b independent of a conditional on (c, d). Intuitively it means that we first choose at random (using the old distribution) values of c and d; then given fixed values of c, d we independently choose at random a and b (the conditional distributions of a given (c, d) and b given (c, d)

are the same as in the original distribution). This can be done by modifying b only. More formally, if $p(a, b, c, d)$ is the original distribution, then the new distribution p' is defined as

$$p'(a, b', c, d) = \frac{p(a, c, d) \cdot p(b, c, d)}{p(c, d)}$$

(for all values (a, b', c, d) of the four random variables). With some abuse of notation we denote the new random variables by a, b', c, d. From the construction (a and b' are independent given c, d) it follows that

$$H(AB'CD) = H(CD) + H(A|CD) + H(B'|CD)$$

Since (b, c, d) has exactly the same distribution as the original (b, c, d), we have

$$H(AB'CD) = H(CD) + H(A|CD) + H(B|CD)$$

The same entropy can be bounded in another way:

$$H(AB'CD) \leq H(D) + H(A|D) + H(B'|D) + H(C|AB')$$

Notice that the conditional entropy $H(B'|D)$ is equal to $H(B|D)$ (we again use the fact that b', d has the same distribution as b, d in the original distribution). Thus, we get

$$H(CD) + H(A|CD) + H(B|CD) \leq H(D) + H(A|D) + H(B|D) + H(C|AB')$$

It remains to estimate the value of $H(C|AB')$. We will show that it is zero (and this is the second trick used in the argument).

Here we will use the two conditions of the theorem. We say that some values a, c (b, c or a, b respectively) are *compatible* if in the original distribution these values can appear together, *i.e.*, $p(a, c) > 0$ ($p(b, c) > 0$ or $p(a, b) > 0$ respectively). Since a and b are independent given c, if some values a and b are compatible with the same value c, then these a and b are compatible with each other.

In the new distribution (a, b', c, d), values of a and b' are compatible with each other *only if* they are compatible with some value of c; hence, these values must also be compatible with each other for the original distribution (a, b). Further, since $H(C|AB) = 0$, for each pair of compatible values of a, b there exists only one value of c. Thus, for a random pair of values of (a, b') with probability one there exists only one value of c. In few words: for the new distribution $H(C|AB') = 0$.

Summarizing our arguments, we get

$$H(CD) + H(A|CD) + H(B|CD) \leq H(D) + H(A|D) + H(B|D),$$

which is equivalent to

$$I(C{:}D) \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B).$$

□

Remark 5. *In the argument of the last proof, we constructed a new distribution* (a, b', c, d) *from* (a, b, c, d)*. If we remove the assumption* $H(C|AB) = 0$*, the new distribution satisfies a combinatorial property: the triples* (A, B, C) *and* (A, B', C) *have the same support. Adding the condition* $H(C|AB) = 0$ *indeed automatically makes* $H(C|A'B') = 0$*.*

These three inequalities are non-Shannon-type since they exclude one of the (symmetric)

points

$$\begin{aligned}
\vec{p_1} &= (2,2,2,2,4,3,3,3,3,3,4,4,4,4,4),\\
\vec{p_2} &= (2,2,2,2,3,4,3,3,3,3,4,4,4,4,4),\\
\vec{p_3} &= (2,2,2,2,3,3,4,3,3,3,4,4,4,4,4),\\
\vec{p_4} &= (2,2,2,2,3,3,3,4,3,3,4,4,4,4,4),\\
\vec{p_5} &= (2,2,2,2,3,3,3,3,4,3,4,4,4,4,4),\\
\vec{p_6} &= (2,2,2,2,3,3,3,3,3,4,4,4,4,4,4),
\end{aligned}$$

while Shannon-type inequalities do not.

We give two other inequalities by Matúš, the proof of which is of a rather different nature and will be given in the sequel (see Chapter 6 p. 105).

Theorem 11 (Matúš, [Mat07b])**.**
if $I(A{:}C|D) = 0$ *and* $I(A{:}D|C) = 0$ *then* $I(C{:}D) \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B)$.
if $I(A{:}C|D) = 0$ *and* $I(C{:}D|A) = 0$ *then* $I(C{:}D) \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B)$.

The reader will certainly have noticed that in each of these five conditional inequalities, the conditions imply Ingleton inequality. While Ingleton inequality does not hold in general, these conditional inequalities show it still holds on some facets. These facets defined by the conditions describe a subspace of co-dimension 2 for each conditional inequality except for the one from Theorem 10 where the co-dimension of the facet is 3 since

$$H(C|AB) = H(C|ABD) + I(C{:}D|AB).$$

Conditional inequalities constitute the "most significant" part of the difference between the set of entropic points and the set of almost entropic points. Indeed, a result of Matúš (see [Mat07c, Theorem 1]) explains that the main difference between these sets lies at the border. Each conditional inequalities specifies a bit more what happens on the frontier of entropy regions.

2.7.2 Non-Shannon-type Unconditional Inequalities

A certain number of, sometimes infinite, lists of 4-variable non-Shannon-type inequalities have been discovered in the literature. Some are not independent, some are superseded by others. They were all discovered by proofs relying on a handful of techniques and tricks. We aggregate here an overview.

Theorem 12 (Zhang and Yeung, [ZY98])**.** *The following is a* 4*-variable non-Shannon-type information inequality:*

$$I(C{:}D) \leq I(C{:}D|A) + I(C{:}D|B) + I(A{:}B) + I(C{:}D|A) + I(A{:}C|D) + I(A{:}D|C)$$

Proof. The following is a Shannon-type information inequality

$$\begin{aligned}
I(C{:}D) \leq\ & I(C{:}D|A) + I(C{:}D|B) + I(A{:}B) + \\
& + I(C{:}D|E) + I(E{:}C|D) + I(E{:}D|C) + 3I(E{:}AB|CD) \quad (2.13)
\end{aligned}$$

Let us construct a suitable auxiliary variable E as follows: When the value of (C, D) is given, define E to be a variable with the same distribution as A, such that E is independent of (A, B). The distribution of (E, C, D) is the same as (A, C, D) and by our construction $E \perp AB|CD$.

Plugging this E in inequality 2.13 proves our theorem. □

Using the same trick, Dougherty *et al* proved six independent non-Shannon-type inequalities in [DFZ06]. Matúš proved a few lists of infinitely many information inequalities in [Mat07b]. A slightly more general inequality was proven using a different technique:

Theorem 13 (Makarychev *et al*, Matúš, [MMRV02], [Mat07a])**.** *The following are two* 5*-variable non-Shannon-type information inequality:*

$$I(C:D) \leq I(C:D|A) + I(C:D|B) + I(A:B) + I(C:D|E) + I(E:C|D) + I(E:D|C) \quad (2.14)$$
$$I(C:D) \leq I(C:D|A) + I(C:D|B) + I(A:B) + I(A:C|E) + I(A:E|C) + I(C:E|A) \quad (2.15)$$

We will need the following lemma (more thoroughly presented in Chapter 4 p. 68):

Lemma 3 (Wyner, [Wyn06])**.** *Let (x, y, z) be a jointly distributed triple of random variables. Consider the N-serialization X, Y and Z respectively. Then there exists a random variable W such that*

$$\begin{aligned} H(W|XY) &= 0, \\ H(W) &\leq N \cdot I(xy:z) + o(N), \\ H(X|W) &\leq N \cdot H(x|z) + o(N), \\ H(Y|W) &\leq N \cdot H(y|z) + o(N), \\ H(XY|W) &\leq N \cdot H(xy|z) + o(N). \end{aligned}$$

We denote this W by $AK(Z:XY)$.

Proof of Theorem 13. Let A', B', C', D', E be N-serializations of A, B, C, D and E respectively and $W = AK(E':C'D')$. Note that we have:

$$\begin{aligned} I(C'D':E') &= I(C':D') - I(C':D'|E') + I(E':C'|D') + I(E':D'|C'), \\ H(W|C') &= H(W) + H(C'|W) - H(C') \\ &\leq I(C'D':E') + H(C'|E') - H(C') + o(N) = I(E':D'|C') + o(N), \\ H(W|D') &= H(W) + H(D'|W) - H(D') \\ &\leq I(C'D':E') + H(D'|E') - H(D') + o(N) = I(E':C'|D') + o(N). \end{aligned}$$

Recall (2.8) from page 16 (rewritten for our use)

$$H(W) \leq 2H(W|C') + 2H(W|D') + I(C':D'|A') + I(C':D'|B') + I(A':B')$$

Dividing by N, this inequality rewrites to :

$$I(C:D) \leq I(C:D|A) + I(C:D|B) + I(A:B) + I(C:D|E) + I(E:C|D) + I(E:D|C) + o(1),$$

which implies inequality (2.14).

Let A', B', C', D', E' be N-serializations of A, B, C, D and E respectively and W satisfy the statement of the previous lemma for A, C, E.

Recall inequality 2.9 (see page 16):

$$H(E|C) \leq H(E|AC) + H(E|BC) + I(A:B|C).$$

The following inequalities are true for our variables:

$$\begin{aligned} I(A'C':E') &= I(A':C') - I(A':C'|E') + I(A':E'|C') + I(C':E'|A'), \\ I(A':C') &= I(C':D') + I(A':C'|D') - I(C':D'|A'), \\ H(W) &\leq H(W|A') + H(W|B') + I(A':B'), \\ H(W|B') &\leq H(W|C') + H(W|D') + I(C':D'|B'), \\ H(W|D') &\leq H(W|A') + H(W|C') + I(A':C'|D'). \end{aligned}$$

As before, we have also

$$\begin{aligned} H(W|A') &= I(E':C'|A') + o(N), \\ H(W|C') &= I(E':A'|C') + o(N). \end{aligned}$$

Summing the last 7 inequalities, dividing by N and taking the limit gives inequality (2.15). □

These inequalities were again generalized in two ways to a greater number of variables using induction in [MMRV02].

Comments

For 5 and 6 linear random variables, the problem is being solved. Dougherty *et al* discovered the full list of inequalities generating **dual**(L_5): there are more inequalities than the ones implied by Shannon-type and Ingleton inequalities. They also have found many independent 6-variables inequalities for ranks [DFZ09]. A possible conjecture is that L_n is generated by a finite list of inequalities, for all $n \geq 1$. For general random variables, they have found many families of 4-variable non-Shannon-type inequalities in [DFZ11].

František Matúš and Milan Studený [MS95, Mat95, Mat99a, Stu01] studied conditional independence relations and Matúš solved the problem for 4 variables. He also gave a complete explanation of the Zhang-Yeung unconditional inequality for 4-variables related to polymatroids.

Chapter 3

Perfect Secret Sharing

Contents

Introduction

A secret-sharing scheme is a method by which one can distribute shares of a secret to several participants in such a way that authorized groups of participants can reconstruct the secret but forbidden ones get no information about it. In the standard definition of perfect secret sharing the requirements are strict: every authorized group gets full information about the secret while any other (forbidden) group gets absolutely no information about it.

For instance, assume that a computer scientist wants to share a secret – say, a bit string x of length n – between Alix and Bert in such a way that they can reconstruct x together but neither Alix nor Bert can do this in isolation. The plan is simple: choose a random string r of length n and give $r$to Alix and $r \oplus x$ to Bert ($r \oplus x$ is the bitwise `xor` of x and r). In isolation, both r and $r \oplus x$ are uniformly distributed among all n-bit strings, so they have no information about x.

3.1 Access Structures

The general setting for secret sharing can be described as follows. We consider a finite set $\mathcal{K}$ whose elements are called *secrets*. A important requirement for $\mathcal{K}$ is that it should contain at least two elements – there should be a secret to be shared. We also have a finite set $\mathcal{P}$ of *participants*. An *access structure* is a non-empty set Γ whose elements are groups of participants, *i.e.*, a non-empty subset of $\mathscr{P}(\mathcal{P})$. Elements of Γ are called *authorized* groups of participants (that should be able to reconstruct the secret). Any other subsets of $\mathcal{P}$ are

called *forbidden* groups (that should get no information about the secret). We always assume that Γ is upward-closed (it is natural since a bigger group knows more).[1].

Definition 13 (Minimal authorized groups)**.** *An* access structure Γ *is a monotone subset of* $\mathscr{P}(\mathcal{P})$*,* i.e.*, if* $A \in \Gamma$ *then* $A \cup B \in \Gamma$ *for any* $B \subseteq \mathcal{P}$*. An authorized group* A *is* minimal *if for any* $a \in A$*, the group* $A - \{a\}$ *is forbidden. Any access structure is uniquely defined by the set* Γ^- *of its* minimal *authorized groups.*

Since in our case, forbidden groups are exactly the groups that are not authorized, one can alternatively determine any access structure by the set of all its forbidden groups Δ. This set is sometimes called the *adversarial* or *prohibited* structure. By the definition of an access structure, Δ is downward closed and is uniquely defined by the set Δ^+ consisting of all its *maximal forbidden groups*:

$$\Delta^+ = \{A \in \Delta : \forall p \in \mathcal{P}, A \cup \{p\} \in \Gamma\}$$

There should exist a secret worth sharing. Therefore, one can consider an access structure should not be empty (since otherwise the secret is not shared) and does not consist of all possible groups of participants (since there should be a secret to be shared). If a group consisting of a single participant is authorized, then this participant should get the secret. It is thus safe to assume that a nontrivial access structure should contain no singleton.

3.1.1 Examples of Access Structures

In the introductory example of this chapter, the set of secret keys is $\mathcal{K} = \mathbb{B}^n$ (the set of n-bit strings), the set $\mathcal{P} = \{Alix, Bert\}$ contains two participants, and the access structure Γ consists of the subset $\{Alix, Bert\}$ only. Let us give more examples of basic access structures.

Threshold access structures. For a set of n participants and integer threshold $0 \leq m \leq n$, the authorized groups of an (m, n)-threshold access structure are the sets of participants of size at least m, *i.e.*,

$$\Gamma_{(m,n)} = \{A \subseteq \mathcal{P} : |A| \geq m\}.$$

Graphical access structures. If G is an undirected graph consisting of a set of vertices V and a set of (undirected) edges E, one can define an access structure $\Gamma(G)$ in a natural way from its minimal authorized groups by letting $\Gamma^-(G) = E$. Participants are vertices and minimal authorized groups are edges. For instance, the access structure associated with the complete graph on n vertices is the $(2, n)$-threshold access structure. More generally, we shall denote by $\Gamma(H)$ the access structure associated with a hypergraph H.

3.1.2 On the Number of Access Structures

An *antichain* is a subset of a partially ordered set such that any two elements are incomparable. On the ground set of participants $\mathcal{P}$, there is a bijection between non-empty antichains (for set inclusion) and access structures defined by minimal authorized groups.

Another way to think about access structures is using *monotone boolean functions*. Take an access structure Γ, construct a boolean function f with n bit inputs. Each input bit i correspond to a participant $p_i \in \mathcal{P}$ so that we can think of the whole input as a group of

[1] One can also consider a more general setting where some groups are neither allowed nor forbidden (so there is no restriction on the information they may get about the secret.) We do not consider this more general setting in this chapter.

participants – the i-th input bit correspond to the presence of participant $p_i \in \mathcal{P}$ in the group. The output of f is 1 iff the group A defined by the input belongs to the access structure. This function is monotone because as soon as a group is authorized, flipping input bits from 0 to 1 will only make the group bigger. In fact there is also a bijection between access structures and monotone boolean functions.

So we see that access structures correspond to well-studied objects. However many simple questions on these objects turn out to be open. For instance, computing the exact number of antichains over a n-element set, which is also the number of monotone boolean functions with n inputs, is a difficult task known as the Dedekind problem. The following good approximation has been found in [KM75],

$$2^{\left(1+O\left(\frac{\log_2 n}{n}\right)\right)\binom{n}{\lfloor \frac{n}{2} \rfloor}}.$$

Sperner studied antichains [Spe28] and gave the maximum size of an antichain

$$\binom{n}{\lfloor \frac{n}{2} \rfloor}$$

witnessed by the $\left(\left\lfloor \frac{n}{2} \right\rfloor, n\right)$-threshold and $\left(\left\lceil \frac{n}{2} \right\rceil, n\right)$-threshold access structures.

3.2 Perfect Secret-Sharing Schemes

In general, a *secret-sharing scheme* can be defined as follows. For every participant $p \in \mathcal{P}$ a set $\mathcal{S}_p$ is fixed; its elements are called p's *shares*. For every value of the secret $k \in \mathcal{K}$ we have a tuple of $|\mathcal{P}|$ dependent random variables $\sigma_p \in \mathcal{S}_p$. There are two conditions for this scheme to be *perfect*:

- for every authorized set $A \in \Gamma$ it is possible to reconstruct uniquely the secret k from the shares given to participants in A, i.e., for different secrets k and k' the projections of the corresponding random tuples onto A-coordinates have disjoint ranges;
- for every forbidden set $B \notin \Gamma$ the participants in B altogether get no information about the secret, i.e., for different secrets k and k' the projections of the corresponding random tuples onto B-coordinates are identically distributed.

Various models of combinatorial secret-sharing schemes were introduced, for instance in [BS92] and [BD91]. Note that in this definition we have no probability distribution on the set of secrets. This setting is natural for the case when somebody gives us the secret (i.e., the user chooses her password) and we have to share whatever is given to us.

For instance, if all probabilities of shares are rational numbers, this definition becomes purely combinatorial and schemes can be represented using finite matrices. We present the simplified case where each possible tuple of shares has the same probability. Now a secret-sharing scheme can be represented as a list of distribution rules. A *distribution rule* is an $n+1$-tuple $(k, \sigma_1, \ldots, \sigma_n)$ where k is a possible secret and $(\sigma_1, \cdots, \sigma_n)$ is a possible vector of shares given to the set of participants. A scheme described by a list of distribution rules works as follows: To share the secret k, choose uniformly at random a distribution rule whose first component is k and distribute the shares σ_i to the corresponding participants.

Another way to talk about secret sharing is to say that the secret is also a random variable (see [KGH83] and the further development in [CSGV93]). In this approach, a secret-sharing scheme is a joint distribution of several random variables: one (s) for the secret and one (σ_p) for each participant p. This scheme is called perfect for an access structure Γ if

k	σ_1	σ_2
0	0	0
0	1	1
1	0	1
1	1	0

k	s_1	s_2	s_3
0	0	0	0
0	0	1	1
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	0
1	1	0	0
1	1	1	1

k	s_1	s_2	s_3
0	0	0	0
0	1	1	1
0	2	2	2
1	0	1	2
1	1	2	0
1	2	0	1
2	0	2	1
2	1	0	2
2	2	1	0

Figure 3.1: Distribution rules of some $(2,2)$, $(3,3)$ and $(2,3)$ threshold schemes

- RECOVERABILITY: for every authorized set A the projection $\sigma_A = \{\sigma_p, p \in A\}$ determines s;
- PRIVACY: for every forbidden set B the projection σ_B is independent with s.

These conditions can be rewritten using Shannon information theory: the first condition says that $H(s|\sigma_A) = 0$, and the second says that $I(\sigma_B{:}s) = 0$, *i.e.*,

$$I(s{:}\sigma_A) = \begin{cases} 0 & \text{if } A \in \Gamma. \\ H(s) & \text{if } A \notin \Gamma. \end{cases}$$

To be exact, we should ignore events of probability zero when saying that σ_A determines s. To avoid technicalities, let us also agree that our probability space is finite and all values of the secret have positive probabilities.

The combinatorial and the information-theoretic definitions of a perfect secret-sharing scheme are closely related:

- Assume that a perfect secret-sharing scheme in the sense of the first definition is given. Then for every distribution on secrets (random variable $s \in \mathcal{K}$) we get a scheme in the sense of the second definition as follows. For each secret $k \in \mathcal{K}$ we have a family of dependent random variables σ_p, and we use them as conditional distribution of participants' shares if $s = k$.
- Assume that a perfect secret-sharing scheme in the sense of the second definition is given, and (as we always assume) all secrets have positive probabilities according to s. Then the conditional distributions of σ_p with the condition $s = k$ form a scheme according to the first definition.

This equivalence shows that, in the second version of the definition, the distribution on secrets is irrelevant (as far as all elements in $\mathcal{K}$ have positive probabilities): we can change s keeping the conditional distributions, and still have a perfect secret-sharing scheme (as first observed in [BSV98]). The advantage of the second definition is that we can use standard techniques from Shannon information theory (e.g., information inequalities).

3.2.1 First properties of perfect schemes

Notations. We (ab)use our notations, for the sake of the reader, in order to make our statements less cumbersome. We may identify groups of participants with their tuple of shares, and omit the signs $\{\}$ and $\cup$ for singletons, so that for a participant a and group B we write aB to mean the pair of shares (σ_a, σ_B) as well as the set $\{a\} \cup B$, according to the context.

The first basic property of perfect secret-sharing scheme is that the number of possible shares of a non-redundant participant is at least the number of possible secrets. This is because a non-redundant participant turns a forbidden group into a authorized group, she should therefore be able to tell apart each possible value of the secret (see [KGH83, CSGV93]).

Lemma 4 (Folklore)**.** *For a perfect secret-sharing scheme, if p is not a redundant participant then*

$$H(p) \geq H(s)$$

The proof of this lemma is related to the proof Shannon's secrecy theorem.

Proof. Let $F \in \Delta$ be a forbidden group and $p \in \mathcal{P}$ be a (forbidden) participant which makes the group pF authorized, then

$$H(p) \geq H(p|F) = H(p|Fs) + I(s{:}p|F) \geq H(s). \qquad \square$$

This motivates the particular case of ideal secret sharing for which the bound above is met:

Definition 14 (Ideal Secret Sharing)**.** *We call a perfect secret-sharing scheme* ideal *whenever every share given to a participant has the same size as the secret. An access structure is called ideal if it admits an ideal perfect secret-sharing scheme.*

In general, we have the following basic bounds for any perfect secret-sharing scheme.

Proposition 10 (Van Dijk, [VD95])**.** *For any perfect secret-sharing scheme*

- *if $B \notin \Gamma$ and $AB \in \Gamma$, then*
$$H(A|B) \geq H(s);$$
- *if $C \notin \Gamma$ and $AC, BC \in \Gamma$, then*
$$I(A{:}B|C) \geq H(s);$$
- *if $ABC \in \Gamma$ and $AC, BC \notin \Gamma$, then*
$$I(A{:}B|Cs) \geq H(s).$$

Proof. The first statement is simply a special case of the second. Under the assumptions of perfect secret sharing, the basic inequality

$$I(A{:}B|Cs) = H(ACs) + H(BCs) - H(Cs) - H(ABCs) \geq 0$$

can be rewritten as

$$H(AC) + H(BC) - H(C) - H(s) - H(ABC) \geq 0,$$

which gives our second statement. The third statement is now easily derived from the inequality

$$I(A{:}B|C) \geq 0$$

and perfect secret sharing requirements. □

Special Participants It is sometimes useful to think of the secret as being held by an extra participant, usually called *dealer*.

Remember that redundant participants should be considered useless in the sharing process since they do not turn any forbidden group into an authorized one. Such a participant can be given a "null" share. We can therefore assume, without loss of generality, that Γ contains no redundant participants, *i.e.*,

$$\bigcup_{A\in\Gamma^-} A = \mathcal{P}.$$

A participant is called dictatorial if she belongs to all authorized groups. It is a little less obvious to see why dictatorial participants can be ignored. Suppose Γ contains a dictator $p \in \mathcal{P}$, then to share the value s, choose a random value r uniformly in the same set as s, give $r + s$ to p and share r with the access structure $\{A - v : p \in A \in \Gamma\}$.

Two participants $p, q \in \mathcal{P}$ are called *equivalent* whenever there is no $A \subseteq \mathcal{P}$ such that $Apq \in \Gamma^-$ and for all $A \subseteq \mathcal{P}$, $Ap \in \Gamma \Leftrightarrow Aq \in \Gamma$. We call *reduced* an access structure which does not contain equivalent participants. (Equivalent participants can be given the same information.)

An access structure is said *connected*, if it is not disjoint union of two access structures. Two access structures Γ_1, Γ_2 are said *isomorphic* if there exists a permutation π such that for all group $A \subseteq \mathcal{P}$, $A \in \Gamma_1 \Leftrightarrow \pi(A) \in \Gamma_2$.

3.2.2 Information Ratios as Efficiency Measure

The efficiency or complexity of a scheme can be measured by the amount of information given to participants (as shares) compared to the size of the secret. We will mainly use two related measures. The first one is the *(worst-case) information ratio* ρ which is defined as the maximal size of a participant share compared to the size of the secret:

$$\rho = \max_{p\in\mathcal{P}} \frac{H(p)}{H(s)}.$$

The second is the *average information ratio*, defined as the arithmetical mean of the sizes of all participants shares compared to the size of the secret:

$$\tilde{\rho} = \frac{1}{|\mathcal{P}|} \sum_{p\in\mathcal{P}} \frac{H(p)}{H(s)}$$

The (optimal) information ratios $\rho(\Gamma)$ and $\tilde{\rho}(\Gamma)$ of an access structure is the supremum of ρ_Σ, resp. $\tilde{\rho_\Sigma}$ over all possible perfect secret-sharing schemes Σ for Γ. An *optimal* perfect secret scheme is a scheme satisfying $\tilde{\rho} = \tilde{\rho}(\Gamma)$. Such a scheme does not necessarily exist, as was shown in [Mat99b] and [BLP08].

3.3 Threshold Schemes

Historically, the motivating example for secret sharing was Shamir's scheme (see [Sha79]) which implements threshold access structures (see Paragraph 3.1.1).

3.3.1 A (n, n)-threshold Scheme

First we present the simpler case where the threshold is maximal, this access structure consists of a single authorized group which is the whole participant set. The case of two participants was solved in the Introduction, it is exactly the idea of one-time pads (Vernam cipher) in cryptography. For n participants, a scheme could go as follows: take n random numbers $r_1, \ldots, r_n$ from, say, $\{1 \ldots, 10\}$, chosen independently and with the uniform distribution. The secret is then $r_1 + \ldots + r_n$.

This setting can be generalized to any group. Let $(G, \bullet)$ be a group and $k \in G$ be the secret key. We define a (n, n)-threshold scheme as follows.

(1) Give the first $n-1$ participants an element $g_i \in G$ chosen uniformly at random.

(2) Give the last participant the share $g_n = g_1 \bullet g_2 \bullet \cdots \bullet g_{n-1} \bullet k$.

The set of all participants may then compute

$$k = g_{n-1}^{-1} \bullet \cdots \bullet g_1^{-1} \bullet g_n,$$

and recover the secret k. However, the share vector of any set of $q < n$ participants, takes any value in G^q with the same probability, and thus give no information about k.

3.3.2 Shamir's Threshold Scheme

An elementary explanation of the main idea of this scheme is that two points determine a unique line, three points determine a parabola, and so on. Shamir's pioneer scheme is a method to implement any threshold access structure. In this scheme, secrets are elements of a finite field $\mathbb{F}_q$ of size greater than n. To share a secret value k, we construct a random polynomial

$$P(x) = k + r_1 x + r_2 x^2 + \ldots + r_{t-1} x^{t-1}$$

where the r_i are chosen independently and uniformly in $\mathbb{F}$. The shares are the values $P(x_1), \ldots, P(x_n)$ for distinct nonzero field elements $x_1, \ldots, x_n$ (for each participant a non-zero element of the field is fixed).

The reason why this scheme indeed works can be subsumed in the following basic fact about polynomials: any polynomial of degree d is entirely defined by any $d+1$ of its values on distinct points.

(1) Recoverability: Every group of t participants can reconstruct the polynomial (and therefore k). Indeed, they can for instance recover the polynomial using Lagrange Polynomial interpolation. If the t shares are $(s_1, \ldots, s_t)$ then

$$P_k(x) = \sum_{i=1}^{t} \left(s_i \cdot \prod_{j=1, j \neq i}^{m} \frac{x - s_j}{s_i - s_j} \right)$$

and thus get the secret by computing $P_k(0)$.

(2) Privacy: While for every $t-1$ participants, all combinations of shares are equally probable (for every k), since the number of polynomials going through the $t-1$ points from the shares and any given secret is the same.

Notice that each share has the same size as the secret, which makes this scheme ideal. This solves almost completely the secret-sharing problem for threshold access structures if we

are not concerned about the secret size. Shamir's scheme requires that the size of the field q is greater than the number of participants n. This is because each participant should be assigned a nonzero field element. If we define the secret to be the slope of the tangent at a fixed point, instead of the constant coefficient of the polynomial, then it would be safe to assign $P(0)$ as a share. Thus reducing the requirement to $q \leq n$. When the size of the secret *does* matter, the secret-sharing problem becomes of a much more combinatorial nature.

3.3.3 Mutually Orthogonal Latin Hypercubes

Sometimes, the size of the secret is a relevant parameter, for instance when storage or transmission costs are at stake. It could be crucial that the cardinality of the set of secrets (and shares) is not bound to be a prime power. It could also be a practical issue that the number of participants must be greater than the total number of secrets.

The result below indicates that the efficiency problem of perfect secret-sharing schemes, when the size of the secret matters, is a quite difficult problem. Let us show how it is related to a difficult open problem connected to combinatorics and algebra (as was noticed in [Daw93, Mar91]).

Definition 15 (Mutually Orthogonal Latin Squares)**.** *A Latin square of order m is a $m \times m$ matrix whose rows and columns contain all integers in $[\![m]\!]$. Two latin squares are said* mutually orthogonal *(MOLS) whenever every possible pair from their Cartesian product appears exactly once.*

Theorem 14. *There exist $n-1$ MOLS of order m iff there exists an ideal perfect $(2, n)$-threshold scheme for a set of m secrets.*

Notice this result implies that Shamir's scheme is a way of generating a set of MOLS of any prime power order.

Proof of Theorem 14. Suppose there is an ideal perfect $(2, n)$-threshold scheme for m secrets and assume further that the secret and the shares belong to the set $[\![m]\!]$. We see this scheme as a matrix of unique tuples (with some probability). Let us show that there are exactly m rules for a given secret. Fix a secret s. First, the number of distribution rules for s cannot be less than m. Otherwise, for any participant there exists one value from $[\![m]\!]$ she never receives as her share. This contradicts the fact that she has no information about the secret. Second, each pair of participants should determine the secret, thus for a given s the share of any participant determines a unique rule. Since there is no duplicate rules, the number of distinct rules for a given secret is at most m, which is the number of different shares given to a participant.

Denote by $(i, j, L_1[i, j], \ldots, L_{n-1}[i, j])$ the tuple (or distribution rule) for sharing the secret i that assigns the share j to the first participant. Let us verify that the matrices L_k defined by these rules make a family of $n-1$ MOLS of order m.

(1) Since the first participant has no information on the secret, the pair (i, j) takes all possible values, which ensures each matrix L_k is well-defined.

(2) The other participants also have no information about the secret, and when i is fixed $L_k[i, j]$ takes all possible values in $[\![m]\!]$ which means that the rows of L_m contain each possible value once.

(3) The same is true for columns by considering the authorized group made of the first participant together with another participant k, because when j and k are fixed, $L_k[i, j]$ takes all possible values.

(4) Finally, any two participants have access to the secret and their shares determine a unique distribution rule. Since there are no duplicate rules, this implies that for fixed k_1 and k_2, all pairs $(L_{k_1}[i,j], L_{k_2}[i,j])$ are distinct.

The first three observations show that each L_k is a Latin square. The last observation ensures that any two Latin squares are mutually orthogonal.

Now suppose we have a set of $n-1$ MOLS. We can use the same construction of distribution rules from the argument above and obtain the desired scheme. □

We can extend this theorem to any threshold scheme by considering the following generalization of Latin squares and orthogonality.

Definition 16 (Orthogonal Latin Hypercubes)**.** *A Latin d-hypercube of order m is a m^d array of elements in $[\![m]\!]$ such that the projection over any single coordinate, all other coordinates being fixed, range over $[\![m]\!]$. A family of n Latin d-hypercubes are said* t-orthogonal *if any tuple from the Cartesian product of any t of them appears exactly once.*

Theorem 15. *There exist $n-1$ d-orthogonal d-hypercubes of order m iff there exists an ideal perfect (d,n)-threshold scheme for m secrets.*

Proof. The proof uses essentially the same type of arguments as Theorem 14 □

Orthogonal hypercubes solve definitively the perfect secret-sharing problem for any secret size. Unfortunately, their very existence, even for a given parameter, is a major open problem. For Latin squares of order $n \geq 3$, the existence of $n-1$ MOLS of order n is equivalent to the existence of a projective plane of the same order (see [LM98, Chapter 8]). For the case of order 6, also known as Euler's 36 Officers problem, the non-existence of 5 MOLS of order 6 was only solved in 1900 using exhaustive search by Tarry in [Tar00].

How the Scheme Works. In the case of a general $(2,n)$-threshold scheme, the secret-sharing method can be depicted in a simple manner. We explain the scheme when $n=4$ in Figure 3.2.

Step 1. Public information : each Latin square L_i for $1 \leq i \leq n$ and a column selection matrix $[C]_{ij} = j$.

Step 2. Select a secret row s.

Step 3. Select a secret column c in C and let it be participant p_0's share.

Step 4. The share of participant p_i is the value of the i-th Latin square at position (s,c).

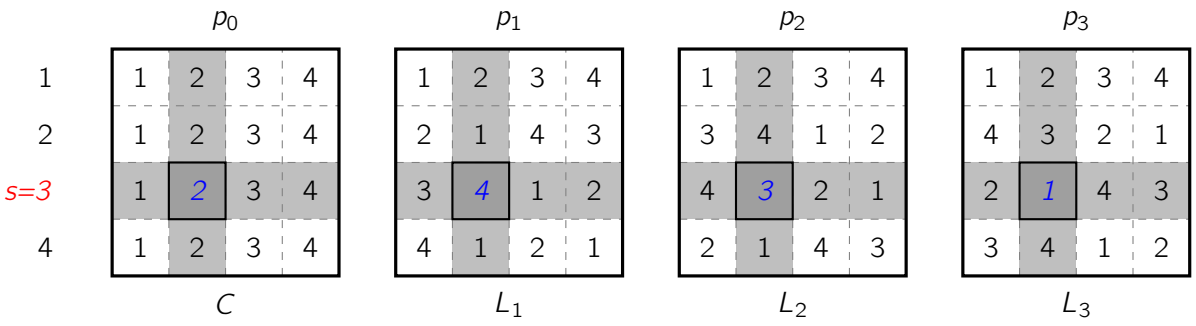


Figure 3.2: A practical $(2,4)$-threshold scheme

3.4 General Access structures and Linear Secret Sharing

In this section, we survey basic and fundamental results in the theory of secret sharing. We first show that any access structure admits a perfect sharing and present a simple general scheme. We introduce the general definition of perfect linear schemes where the random variable are restricted to linear random variables. We show a new property of linear schemes based on a property of vector spaces.

3.4.1 Every Access Structure admits a Perfect Secret-Sharing Scheme

Secret sharing for an access structure is always possible. Various general constructions were proposed for instance by Ito *et al.* in [ISN87] and by Benaloh *et al.* [BL88].

Proposition 11. *Every access structure admits a perfect secret-sharing scheme.*

A general scheme from (m, m)-threshold schemes

The idea of the general scheme we propose below is to use (m, m)-threshold schemes as building blocks. An access structure consists of (possibly) several minimal authorized groups, each of them may be considered as a (sub)access structure. A minimal authorized group $A \in \Gamma^-$ of size m induces a (m, m)-threshold access structure. The proposed scheme goes as follows:

The scheme: For each value of the secret, implement *independently* each minimal group A (using for instance the threshold scheme presented in Section 3.3.1).

Proof. A share assigned by one of the threshold schemes is here referred to as a subshare. By definition, any minimal authorized group can recover the secret. Let F be a forbidden group. We need to see that $H(s|F) = H(s)$.

Call F_i the share given to the group F by the scheme implementing the i-th minimal group. Suppose that F is involved in, and thus receives shares from, M minimal groups. Define

$$F_{[k]} = \{F_j : j \leq k\}.$$

We prove by induction that $H(s|F_k) = H(s)$ for all $1 \leq k \leq M$: Since F is forbidden, the secret s is independent of the share F_1. Suppose now that $H\left(s|F_{[k]}\right) = H(s)$ for some $k < M$, the following information inequality holds

$$H\left(s|F_{[k+1]}\right) = H\left(s|F_{k+1}\right) + H\left(s|F_{[k]}\right) - H(s) + I\left(F_{k+1}:F_{[k]}\right) - I\left(F_{k+1}:F_{[k]}|s\right).$$

By induction, $H\left(s|F_{k+1}\right) = H\left(s|F_{[k]}\right) = H(s)$. By definition of our scheme:

$$I\left(F_{k+1}:F_{[k]}|s\right) = 0,$$

because distinct subshares are independent given the secret. The above inequality then rewrites to

$$H\left(s|F_{[k+1]}\right) \geq H(s) + I\left(F_{k+1}:F_{[k]}\right),$$

and since the left-hand side cannot be greater than $H(s)$ (by Corollary 1 page 13), we get our result. □

Notice that we also proved that all subshares from an independent group are mutually independent (by assumption, they were only independent given the secret).

In this scheme, the number of subshares a participant gets is exactly the number of minimal groups she is involved in. This number can almost grow exponentially with the number of participants (see Section 3.1.2 page 38). Indeed, for a $(m, 2m)$-threshold access structure, this scheme gives a share of size

$$\binom{2m}{m-1} H(s)$$

to any single participant, whereas the shares in Shamir's ideal threshold scheme are of the same size as the secret.

Remark 6. *Notice further that the argument of the previous proof is independent of the subscheme used. This remark will be used later.*

A general scheme from the adversarial structure

We present another proof based the construction from [GM04]. This time the scheme is based on the adversary structure Δ. The following scheme makes sense, for instance when only Δ is given or when $|\Delta^+| \ll |\Gamma^-|$.

General scheme.

1. Let $m = |\Delta^+|$ be the number of maximal forbidden groups, consider a (m, m)-threshold scheme and denote the shares $(s_1, s_2, \ldots, s_m)$.
2. For each maximal forbidden group $F_i \in \Delta^+$, give the share s_i to the each participant in $\mathcal{P} \setminus F_i$.

The nice property about this scheme is that a group of participants B owns the share s_i *if and only if* B is not contained in F_i.

1. RECOVERABILITY: If $A \in \Gamma^+$, from the above mentioned property of this scheme, participants in A have all shares s_i.
2. PRIVACY: For $F_i \in \Delta^+$, participants lack the share s_i and by the privacy property of the threshold scheme, have no information about the secret.

In practice, this scheme is not very efficient for it also provides shares of size $\Theta\left(\binom{2m}{m-1}\right)$ bigger than the secret size for $(2m, m)$-threshold access structures. At this point a question pops up: are there any non-ideal access structures ?

3.4.2 Not Every Access Structure is Ideal

Yes, there are and we present here one of the smallest non-ideal access structure (minimal for the number of participants). This access structure can be represented as a graph: it is the path with four vertices:

$$P_4^- = \{ab, bc, cd\}.$$

This was the first access structure to be shown non-ideal (see [BL88]). The bound on its information ratio was further improved in [KO96, CSGV93].

Proposition 12. *The access structure defined by the minimal authorized groups P_4^- is not ideal and $\rho(P_4) \geq \frac{3}{2}$.*

Proof. Take inequality (2.5) from page 16 (proven without words in Figure 2.8 p. 22):

$$H(bc) \geq I(a{:}c|b) + I(b{:}d|ac) + H(bc|ad).$$

From Proposition 10, each of the terms in the right-hand side is at least $H(s)$. Therefore we proved that $H(bc) \geq 3H(s)$. Now since $H(b) + H(c) \geq H(bc)$, either b or c has a share of size at least $\frac{3}{2}H(s)$. □

Notice that the proof uses only part of the perfect secret-sharing requirements from access structure P_4. In fact, the same proof works also for other access structures. For four participants, these access structures are depicted in Figure 3.3

Corollary 3 (Jackson and Martin, [JM96])**.** *If* $ab, bc, acd \in \Gamma$ *and* $b, ac, ad \in \Delta$ *then* $H(bc) \geq 3H(s)$.

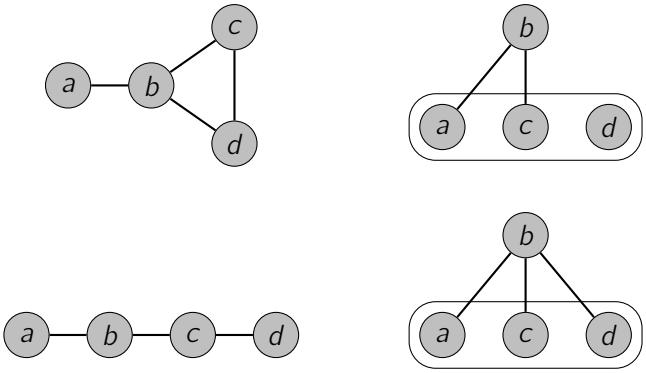


Figure 3.3: Non-ideal access structures with four participants.

As a consequence, the optimal worst-case information rate of P_4 satisfies $\rho^*(P_4) \leq \frac{2}{3}$. We show that the bound is met by providing an explicit scheme. Take two secrets s_1, s_2 and four random elements $r_1, \ldots, r_4$ from the finite field $\mathbb{F}_2$, all chosen independently and uniformly at random. Define the shares of participants a, b, c, d by:

$$\begin{aligned} a &= (r_1, r_2) \\ b &= (r_1 + s_1, r_2 + s_2, r_3) \\ c &= (r_3 + k_1, r_4 + k_2, r_2) \\ d &= (r_3, r_4) \end{aligned}$$

3.4.3 A General Decomposition Construction

In the spirit of building schemes from others, one could try to implement an access structure by covering or decomposing it into substructures (other than plain (m, m)-threshold access structures). This is what Douglas Stinson proposed in a generalized form in [Sti94].

Definition 17 (λ-decomposition of an Access Structure, [Sti94])**.** *A* λ*-decomposition of an access structure is a sequence of access (sub)structures* $(\Gamma_1, \Gamma_2, \ldots, \Gamma_m)$ *such that :*

1. $\Gamma_j \subseteq \Gamma$, *for each* $1 \leq j \leq m$.

2. $|\{j : A \in \Gamma_j\}| \geq \lambda$ *for all* $A \in \Gamma$.

Theorem 16 (Decomposition Technique)**.** *Assume the access structure Γ admits a λ-decomposition $(\Gamma_1, \Gamma_2, \ldots, \Gamma_m)$ such that each access substructure Γ_j can be implemented for the same set of secret $\mathcal{K}'$, where $|\mathcal{K}'|$ is a prime power $q > n$, by a perfect secret-sharing scheme with information ratio ρ_j. Then there exists a perfect secret-sharing schemes for Γ with information ratio*

$$\rho = \max_{p \in \mathcal{P}} \frac{\sum\limits_{j : p \in \bigcup \Gamma_j} \rho_j}{\lambda}$$

The proof of this theorem relies on the following lemma.

Lemma 5. *For any m and d, for all large enough powers of primes q there exists a collection of m vectors in the d-dimensional vector space over $\mathbb{F}_q$ such that any subset d vector is a basis.*

Proof. Take the row vectors $(\alpha_i^k)_{1 \leq k \leq d}$ of the $m \times d$ Vandermonde matrix for distinct elements $\alpha_j \in \mathbb{F}_q$. □

Proof of Theorem 16. Let the cardinality of $\mathcal{K}'$ be a prime power q. Each access substructure has a perfect secret sharing subscheme for sharing a subsecret in $\mathcal{K}'$

Now, define the scheme for Γ as follows: The cardinality of the secret set $\mathcal{K}$ is a power of q to be fixed later. For each subscheme for G_j, share the value s_j independently and assign the subshares to the corresponding participants. By definition of the λ-decomposition, each authorized group can recover at least λ values s_j while every forbidden group knows nothing about all s_j. Now for the secret of our scheme: From the previous lemma, there exists a family of m vectors $(v_1, \ldots, v_m)$ in a λ-dimensional vector space V, over $\mathbb{F}_q$ for a large enough q, such that any λ of them have full rank. Let the secret s be a linear function on V. Fix each s_j to be the restriction of s on the line defined by the vector v_j.

The recoverability property follows since any λ values s_j determine the functional s uniquely. Since the subschemes are implemented independently, the subshares are thus independent given the secret s. The privacy property follows from Remark 6.

By definition of the λ-decomposition, a single participant is given at most

$$\sum_{j : p \in \bigcup \Gamma_j} \rho_j$$

subshares, which gives the value of the information ratio stated by the theorem. □

A further generalized decomposition construction is provided in [vDKST06].

3.4.4 Upper bounds on the Information Ratio

In fact, the previous decomposition technique solves the secret sharing problem for many families of access structures by providing an upper bound matching the lower bound. For instance this technique solved the problem for all access structure up to 4 participants, for most access structures with 5 participants and for most graphical access structures structures for 6 participants (see [FMBPV12, Atı00, JM96, Sti92, VD95]).

Problem solved for some graph families

For individual access structures, upper bounds from the decomposition technique can be obtained using Linear Programming, sometimes leading to a generalization to a whole family of access structures. We compile below some of the results from [CT09, CL09, Sti94].

- A star is a an access structures consisting of a center v and petals w_i such that a minimal group is of the form (v, w_i). Such an access structure is ideal: simply give $r + s$ to v and r to each w_i.

- The path P_n with n vertices satisfies :

$$\rho(\Gamma(P_n)) = \begin{cases} 1 & \text{if } n \leq 3. \\ \frac{3}{2} & \text{if } n \geq 4. \end{cases}$$

- The cycle C_n with n vertices satisfies if $n < 4$ then ideal else

$$\rho(\Gamma(C_n)) = \begin{cases} 1 & \text{if } n \leq 4. \\ \frac{3}{2} & \text{if } n \geq 5. \end{cases}$$

- The d-dimensional hypercube H_n satisfies for $d \geq 2$:

$$\rho(\Gamma(H_d)) = \frac{d}{2}.$$

- Let T be a tree, a core in T is a connected subset of the vertices such that every vertex in the core has a neighbor outside the core. If c is the size of the largest core in T, then

$$\rho(\Gamma(T)) = 2 - \frac{1}{c}.$$

Homogenous Hypergraphs of Bounded Maximum Degree

D. Stinson proved a simple upper bound for graphs with a bounded maximal degree using his decomposition technique:

Theorem 17 (Stinson [Sti94])**.** *Let G be a graph of maximal degree d, then $\rho(\Gamma(G)) \leq \frac{d+1}{2}$.*

Proof. We implement $\Gamma(G)$ using a 2-decomposition into stars. For each vertex v, implement the star with center v whose petals are the neighbors of v. This makes a 2-decomposition: each edge of G is covered twice. each vertex v is covered at most $d + 1$ times: once from the star for which v is the center, and at most d times from the stars associated with its neighbors. By Theorem 16:

$$\rho(\Gamma(G)) \leq \frac{d+1}{2}.$$ □

Note that this upper bound is slightly better than the one obtained from the general scheme presented in Section 3.4.1 p. 46. This result generalizes to a class of hypergraphs. We say that an hypergraph H is k-homogenous if its hyperedges have exactly k vertices.

Theorem 18. *For $k \geq 1$, if H be a k-homogenous hypergraph of maximal degree d, then*

$$\rho(\Gamma(H)) \leq \frac{d^k - 1}{k(d-1)}.$$

Proof. We prove this theorem by induction. For $k = 1$, the access structure consists of singletons, it is therefore ideal. Note that for $k = 2$, we retrieve the bound from the previous theorem.

Let H by a k-homogenous hypergraph of maximal degree d. Let v be a vertex, consider the access structure defined by the set of minimal groups containing v:

$$H_v = \{A \in \Gamma(H) : v \in \Gamma(H)\},$$

In this access structure, v is a dictator so we can implement H_v by assigning $r + s$ to v and sharing r with $H_{-v} = \{A - v : A \in H_v\}$.

The access structure H_{-v} is a $k-1$ homogenous hypergraph of maximal degree d, so it satisfies the induction hypothesis, *i.e.*, H_{-v} has a perfect scheme with information ratio

$$\rho_{k-1} \leq \frac{d^{k-1} - 1}{(k-1)(d-1)}.$$

Implement H using a decomposition into H_v for every vertex v. Each hyperedge e is covered k times, once for each $v \in e$ which means $\lambda = k$. Each participant receives a share from its star, and is at most in d hyperedges containing exactly $k-1$ participants other than her. Thus H can be perfectly implemented with information ratio:

$$\rho_k \leq 1 + \rho_{k-1} \cdot d \cdot (k-1)$$

Therefore

$$\rho(\Gamma(H)) \leq \frac{d^k - 1}{k(d-1)}.$$ □

3.4.5 Linear Secret-Sharing Schemes

Most constructions provide linear secret-sharing schemes, where the random variables are linear. In a linear scheme, the computations of shares and the reconstruction of the secret are very efficient since they only involve linear maps. The following definition encapsulates many others attempts of defining linear secret sharing from the literature.

Let $\mathcal{L}$ be a finite vector space on a field K. A *linear secret-sharing scheme* (on $\mathcal{L}$) assigns a linear subspace L_s to the secret and L_p to participant $p \in P$ such that :

- if A is an authorized group then L_s is contained in the span of the subspaces of participants in A,
- if F is a forbidden group, then the intersection between L_s and the span of all subspaces of participants in F is zero.

This definition is equivalent to the definition of perfect secret-sharing scheme for the case of linear random variables, this class of random variables has been presented in Chapter 2 on page 22. Such a simplification in the definition is natural since the entropies of a secret-sharing scheme based on linear random variables are proportional to the ranks of the corresponding subspaces. In algebraic terms, the information ratios of a linear perfect secret-sharing scheme rewrite to

$$\rho_\Sigma = \max_{p \in \mathcal{P}} \frac{\operatorname{rk} L_p}{\operatorname{rk} L_s} \text{ and } \tilde{\rho} = \frac{1}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \frac{\operatorname{rk} L_p}{\operatorname{rk} L_s}.$$

It is known that linear schemes are strictly less powerful than nonlinear schemes, see [BI01]. One reason for that comes from the existence of information inequalities that hold for linear random variables but not generally, for instance Ingleton Inequality 2.11.

3.4.6 The Size of a Leaf Share

We present a new property of perfect linear schemes. We show that the share of a participant which belongs to exactly one minimal authorized group can always be made of the same size as the secret.

Definition 18 (Leaf)**.** *A* leaf *(participant) in an access structure* Γ *is a participant which belongs to exactly one set in* Γ^-.

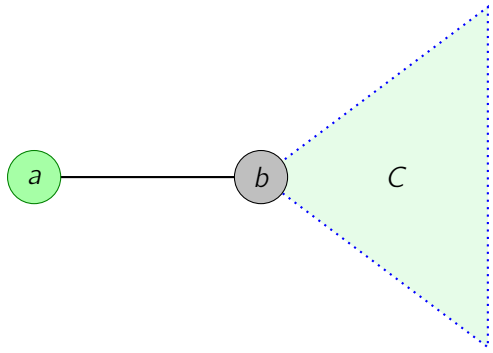


Figure 3.4: Notation for an access structure with a leaf a

Assume, for simplicity, that our access structure contains participants a, b such that $\{a, b\}$ is the only minimal authorized group containing a. We argue why we can make such an assumption without loss of generality. A more general setting is the case where there are disjoint groups of participants A and B such that the only minimal group involving A is $A \cup B \in \Gamma^-$ (Notice this implies that A and B are forbidden groups).

- First, why define leaf participant a and not a leaf group A ? Suppose A is not singleton, the question is whether each participant in A can get a share that has the smallest size. Assume we can solve the problem when $A = \{a\}$ is a singleton, then we could simply distribute a's share among participants of A via an ideal $(|A|, |A|)$-threshold scheme – for which the share size of participants in A is $H(a)$. We have thus reduced the problem to a singleton.
- Why do we assume B is a singleton ? Our proofs will never use the fact that b is a single participant, they remain valid if b is replaced by B.

Therefore we shall assume without loss of generality that our access structures are indeed the ones described by Figure 3.4.

Proposition 13. *Any perfect linear secret-sharing scheme can be converted into another one where a leaf share has the same size as the secret and keeping the same size for other shares.*

The proposition implies that an optimal scheme always gives leaves a share of minimum size. The main ideas of the proof are based on two structural properties of vector spaces, namely intersection and direct sums. The intersection of two vector spaces is also a vector space, which means in terms of the random variables previously defined that the intersection operator extracts the common information, an information quantity introduced in [GK73]. Another property linear random variables enjoy is

Lemma 6. *Let* $\mathcal{L}$ *be a linear vector space and* A *and* B *be two linear subspaces of* $\mathcal{L}$*. There exists a linear subspace* A' *such that:*

(1) $\mathrm{rk}(A + A') = \mathrm{rk}(A)$

(2) $\mathrm{rk}(A') = \mathrm{rk}(A + B) - \mathrm{rk}(B)$

(3) $\mathrm{rk}(A + A' + B) = \mathrm{rk}(A' + B)$

Proof. Let A' be such that $(A \cap B) \oplus A' = A$.

(1) A' is a subspace of A.

(2) By definition of A', we have

$$\mathrm{rk}(A \cap B) + \mathrm{rk}(A') = \mathrm{rk}(A)$$

and using the rank inequality, we obtain

$$\mathrm{rk}(A + B) + \mathrm{rk}(A \cap B) = \mathrm{rk}(B) + \mathrm{rk}(A).$$

(3) A is a subspace of $A' + B$. □

The previous lemma is an exact analogue of Slepian–Wolf coding (see next chapter Section 4.1.2 page 67), exact in the sense that equalities hold without overheads.

Proof of Proposition 13. Let $\Sigma = (L_s, L_a, L_b, L_{c_1}, \ldots, L_{c_{n-2}})$ be a linear perfect secret-sharing scheme for Γ defined by a tuple of subspaces of the finite vector space $\mathcal{L}$.

First, define L''_a as a linear subspace satisfying Lemma 6 applied to L_a and L_b and consider the new scheme Σ'' where the subspace of leaf a is now L''_a. If $C \subseteq P - \{a, b\}$ is a group of participants, then

$$L_C \subseteq L''_a + L_C \subseteq L_a + L_C$$

because L''_a is contained in L_a. Therefore, aC is an authorized group in Σ'' iff aC is authorized in Σ. Also, the span of $L''_a + L_b$ contains L_s because $L''_a + L_b = L_a + L_b$. Thus we have checked that the linear scheme Σ'' for Γ is still perfect. Note that L''_a and L_b intersect in the zero vector, which in terms of shares means that a and b are independent in Σ'' : $I(a{:}b) = 0$.

Next, take $L'_a = L''_a \cap (L_b + L_s)$ and consider the linear scheme

$$\Sigma' = (L_s, L'_a, L_b, L_{c_1}, \ldots, L_{c_{n-2}}).$$

Let us check that Σ' induces a perfect linear scheme for Γ. Since L'_a is contained in L''_a, we only have to check that $\{a, b\}$ is still authorized in Σ', i.e., $L_s \subseteq L'_a + L_b$.

$$\begin{aligned} L_s \subseteq L''_a + L_b &\Leftrightarrow \forall s \in L_s \, \exists a'' \in L''_a, b \in L_b, \; s = a'' + b \\ &\Leftrightarrow \forall s \in L_s \, \exists a' \in L'_a, b \in L_b, \; s = \underbrace{(b - s)}_{a'} + b \\ &\Leftrightarrow L_s \subseteq L'_a + L_b. \end{aligned}$$

Moreover,

$$\begin{aligned} \mathrm{rk}\, L'_a &= \mathrm{rk}(L''_a \cap (L_b + L_s)) \\ &= \mathrm{rk}\, L''_a + \mathrm{rk}(L_b + L_s) - \mathrm{rk}(L''_a + (L_b + L_s)) \\ &= \mathrm{rk}\, L''_a + \mathrm{rk}\, L_b + \mathrm{rk}\, L_s - \mathrm{rk}(L_a + L_b) \\ &= \mathrm{rk}(L_a + L_b) - \mathrm{rk}\, L_b + \mathrm{rk}\, L_b + \mathrm{rk}\, L_s - \mathrm{rk}(L_a + L_b) \\ &= \mathrm{rk}\, L_s \end{aligned}$$

and therefore Σ' is a perfect linear scheme for Γ such that $H(a) = H(s)$. □

The obvious question is whether this result also holds in the general case for non necessarily linear secret-sharing schemes.

Question 1. *Can any perfect secret-sharing scheme be converted into another with a greater or equal average information rate where a leaf share has the same size as the secret ?*

This question remains open for now, the proof above does not generalize because there is no general way of extracting the mutual information. As discussed in Chapter 2, for two random variables the common information is generally less than the mutual information. In the case of linear random variables, these two quantities are equal. However a first step for the resolution of the previous question will be presented in Chapter 5 p. 90.

3.5 Ideal Secret-Sharing Schemes and Matroids

Ideal secret-sharing has been widely studied. Ideal schemes represent what can be done optimally. We present here the relation between ideals schemes and matroids as well as some example of classes of ideal access structures.

First we show an equivalent characterization of ideal schemes in terms of entropy. The following proposition shows that ideal schemes meet strict constraints.

Proposition 14 (Ng and Walker, [NW01])**.** *A perfect secret-sharing scheme is ideal if and only if for all groups $A \in \Delta$ and participants $p \in \mathcal{P}$ such that $Ap \in \Gamma$, it holds that*

- $H(p|A) = H(p)$ *and*
- $H(p|As) = 0$.

Proof. In any ideal perfect secret-sharing scheme, for such A and p we have from the perfect secret-sharing requirements that

$$H(p|A) = H(p|As) + I(s{:}p|A) = H(p|As) + H(s).$$

The property of ideal schemes implies $H(p|As) = 0$ and $H(p|A) = H(p) = H(s)$

Suppose that for any $A \in \Delta$ and $p \in \mathcal{P}$ such that $Ap \in \Gamma$, we have $H(p|A) = H(p)$ and $H(p|As) = 0$. We conclude using the symmetry of the mutual information between s and p conditional to A.

$$\begin{aligned} H(s|A) - H(s|Ap) &= H(p|A) - H(p|As) \\ H(s) &= H(p). \end{aligned}$$

□

The point of showing this equivalence is to understand that ideal schemes have tighter constraints than general schemes. The entropy profile of an ideal perfect secret-sharing scheme should generally satisfy more equalities than the perfect secret sharing requirements. In fact, we have the following corollary which is not true for general schemes:

Corollary 4. *If a perfect secret-sharing scheme where d is the dealer for participants in $\mathcal{P}$ is ideal, then the scheme is also perfectly ideal when $p \in \mathcal{P}$ is the dealer for participants in $\mathcal{P} - \{p\} \cup \{d\}$.*

3.5.1 Ideal Graphical Access Structures

As an example, we characterize the ideal graphical access structures and give a result which further separates the information ratio of non-ideal graphical access structures from ideal ones.

Definition 19 (Complete multipartite graphs)**.** *A complete multipartite graph is a graph whose vertex set can be partitioned into subsets such that the set of edges is exactly the set of all possible edges between two points in distinct parts.*

Proposition 15. *A connected graph G is complete multipartite if and only if G does not contain P_4 nor Q_4 as an induced subgraph.*

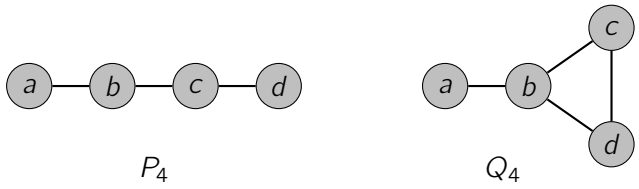


Figure 3.5: P_4 and Q_4: the excluded induced subgraphs for complete multipartite graphs

Proof. Let $G = (V, E)$ be a connected graph. The result holds for graphs with less than 3 vertices, suppose that $|V| \geq 4$.

Suppose G is complete multipartite graph and let $ab \in E$. Let us try to complete this edge into P_4 or Q_4. First case, let us try to add a' such that $ba' \in E$ and $aa' \notin E$. Then for any c such that $ac \in E$ we will also have $a'c \in E$. Second case, let us try to add c such that $ac \in E$ and $bc \in E$. Then for any d such that $cd \in E$, at least one of ad or bd is an edge. In both cases, we have shown that G cannot induce P_4 or Q_4.

Now, suppose G is not complete multipartite, then by a transitivity argument, there should exist $a, b, c \in V$ such that $bc \in E$ but neither $ab \in E$ nor $ac \in E$. Now G being connected, there exists a shortest path connecting a and b or c. Up to renaming, this shortest path is $(a, a_1, a_2, \ldots, a_m, b)$ where $m \geq 1$ and $a_i b \notin E$ for all $1 \leq i < m$. The subgraph induced by vertices a_{m-1}, a_m, b, c is either isomorphic to P_4 or Q_4. □

Using this characterization we can now prove the following theorem.

Theorem 19. *Let G be a connected graph. The following are equivalent*

(a) $\Gamma(G)$ is ideal

(b) $\rho(\Gamma(G)) < \frac{3}{2}$

(c) G is a complete multipartite graph

Proof.

(a) $\Longrightarrow$ (b) is trivial.

(c) $\Longrightarrow$ (a) because the reduced access structure of a complete multipartite graph with n independent sets is isomorphic to the $(2, n)$-threshold access structure.

$\neg$(c) $\Longrightarrow$ $\neg$(b) : If G is not complete multipartite then it induces P_4 or Q_4, therefore $\rho(\Gamma(G))$ is at least $\frac{3}{2} = \min\{\rho(P_4), \rho(Q_4)\}$. □

3.5.2 Access Structures from Matroid Ports

Many results for graphs can be extended to matroids, and the previous one is no counterexample. Ideally-perfect secret sharing is intimately related to Matroid Theory. Matroids were originally introduced to describe and study the concept of linear independence, they pop up in various fields, such as Graph Theory of course, but also linear algebra, greedy algorithms, or Combinatorial Game Theory. For an exhaustive study on Matroid Theory see [Wel76, Oxl92]

We first define what is a matroid in terms of rank functions. Matroids are just restricted polymatroids, and polymatroids are related to entropic points.

Definition 20 (Polymatroids and matroids)**.** *Let $\mathcal{Q}$ be a finite set and $r : \mathscr{P}(\mathcal{Q}) \to \mathbb{R}_+$ be a function. We call $(\mathcal{Q}, r)$ a* polymatroid *whenever r satisfies the following requirements:*

(P1) *$r(A) \geq 0$ for $A \subseteq \mathcal{Q}$ and $r(\varnothing) = 0$*

(P2) *if $\mathcal{Q} \supseteq A \supseteq B$, then $r(A) \geq r(B)$*

(P3) *if $A, B \subseteq \mathcal{Q}$, then $r(A) + r(B) \geq r(A \cup B) + r(A \cap B)$*

The function r is called a rank function. Whenever r is integer-valued and $r(\{q\}) = 1$ for $q \in \mathcal{Q}$ then $(\mathcal{Q}, r)$ is called a matroid*.*

The second property is called subadditivity and the third property is known as submodularity. The set of all polymatroids on a given ground set has already been encountered at this point in Chapter 2. Notice that if the ground set $\mathcal{Q}$ is a set of random variables and r is the entropy function, then we just defined a polymatroid. For instance, submodularity is an analogue of the basic inequality. This remark was first made by Fujishige [Fuj78]. So the set of all polymatroids is related to the set of all the points in the Euclidean space satisfying Shannon-type inequalities.

Matroids are what we call a cryptomorphic object, for they have multiple equivalent axiomatizations. We introduce some of them that will be of interest in this chapter while trying to make the connection with access structure. In the manner of graphs, matroids also relate to access structures. To describe this relationship, it is useful to introduce the extra participant d called dealer and consider the extended participant set $\mathcal{P}^* = \mathcal{P} \cup \{d\}$.

Definition 21 (I-Matroid)**.** *A matroid is a pair $\mathcal{M} = (\mathcal{V}, \mathcal{I})$ where $\mathcal{V}$ is a finite set and $\mathcal{I}$ is a set of subsets of $\mathcal{V}$ such that:*

(I1) *$\varnothing \in \mathcal{I}$.*

(I2) *if $A \in \mathcal{I}$ and $B \subseteq A$, then $B \in \mathcal{I}$*

(I3) *if $A, B \in \mathcal{I}$ and $|A| = |B| + 1$ then there exists $x \in A \setminus B$ such that $B \cup \{x\} \in \mathcal{I}$*

A maximal independent set is called a base*, all bases have the same cardinality, this number is called the* rank *of the matroid.*

In the above axioms a matroid is defined by a collection of *independent* sets. For an access structure Γ, we would like to say that a set of the form $A \cup \{d\}$, where d is the dealer, and $A \in \Gamma$ is dependent, because the share of d can be recovered by participants in A.

Definition 22 (C-Matroid)**.** *A matroid $\mathcal{M} = (\mathcal{V}, \mathcal{C})$ consists of a finite set $\mathcal{V}$ and a collection $\mathcal{C}$ of subsets of $\mathcal{V}$ satisfying the three following properties:*

(C1) *$\varnothing \notin \mathcal{C}$*

(C2) *if $A, B \in \mathcal{C}$ and $A \subseteq B$, then $A = B$*

(C3) *if A, B are distinct members of $\mathcal{C}$ and $x \in A \cap B$ then there exists $C \in \mathcal{C}$ such that $C \subseteq (A \cup B) \setminus \{x\}$*

Proposition 16 (Cryptomorphism)**.** *$\mathfrak{M}$ is a matroid according to definition 20 iff $\mathfrak{M}$ is a I-matroid iff $\mathfrak{M}$ is a C-matroid.*

For instance, given an I-matroid, one can find a C-matroid and whose independent sets are the ones of the I-matroid.

In the previous axiomatization, a circuit $C \in \mathcal{C}$ represents a minimal dependent set which relates to minimal authorized sets in access structures. The notion of circuit is, as we will see, the right one to define access structures. It turns out this has been previously studied by Lehman [Leh64], in a different context, and further developed by Seymour in [Sey76]. Such access structures are called matroid ports:

Definition 23 (Matroid port)**.** *Let $\mathcal{M} = (\mathcal{V}, \mathcal{C})$ be a matroid and $e \in \mathcal{V}$, then the matroid port of $\mathcal{M}$ through e is defined by:*

$$P(\mathcal{M}, e) \triangleq \{C - e : e \in C \in \mathcal{C}\}.$$

An access structure defined by its minimal groups which is isomorphic to the port of a matroid is said matroid-related*.*

Definition 24 (Connectedness)**.** *A matroid is connected if for every pair of distinct elements x and y there is a circuit containing x and y.*

Proposition 17 (Welsh, [Wel76])**.** *Let e be an element of a connected matroid $\mathcal{M}$, then the collection of circuits containing e uniquely determines $\mathcal{M}$.*

Connectedness is only used to obtain uniqueness, and does not change the class of ports. An access structure is thus related to a unique (appropriate) connected matroid. A matroid may still relate to many non-isomorphic access structures.

Example 9. *Let us give a few examples of matroids:*

- Uniform matroids*: On a ground set $\mathcal{V}$ of n elements, let $0 \leq r \leq n$ be a threshold then define $\mathcal{I}$ to be the collection of all subsets of X of size less or equal to r. Then $\mathfrak{U}_{r,n} = (\mathcal{V}, \mathcal{I})$ is a matroid and is said uniform. The matroid $\mathfrak{U}_{r,n}$ is related to the threshold access structure $\Gamma_{(r,n-1)}$.*
- *The Fano and the non-Fano matroids: see Figure 3.6.*

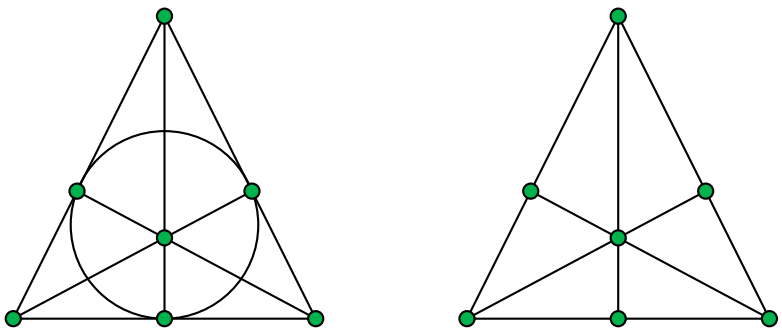


Figure 3.6: The Fano and the non-Fano matroids

A Partial Characterization of Ideal Schemes

Some matroids arise naturally in the context of linear algebra.

Definition 25 (Linearly Representable matroids)**.** *A matroid* $(\mathcal{V}, \mathcal{C})$ *is* representable *over a finite field* $\mathbb{F}$ *if there exists a mapping* $\phi : \mathcal{V} \to \mathbb{F}^d$ *such that*

$$A \subseteq \mathcal{V} \text{ is an independent set } \iff \phi(A) \text{ is linearly independent}$$

Proposition 18. *If* $\mathcal{M}$ *is a representable connected matroid, then there is an ideal secret-sharing scheme realizing an access structure* Γ *isomorphic to the port of* $\mathcal{M}$ *through any element* e*.*

The proof follows directly from the construction of linear random variables from Chapter 2. Less obvious is how any ideal scheme defines a matroid.

Theorem 20 (Brickell and Davenport, [BD91])**.** *If access structure* Γ *admits an ideal secret-sharing scheme then it is the port of a connected matroid* $\mathcal{M}$*.*

This result was slightly generalized in [BK97] and [FP12]. The main ingredient of the proof is the following key lemma:

Lemma 7. *For any ideal perfect secret-sharing scheme for a connected access structure* Γ*,* $H(A)$ *is a multiple of* $H(s)$ *for all* $A \subseteq \mathcal{P}$*.*

Proof. We will prove equivalently that the normalized entropy $h(A) = \frac{H(A)}{H(s)}$ is integer-valued. We will also use $h(B|A) = \frac{H(B|A)}{H(s)}$, for any $A, B \subseteq \mathcal{P}$.

Let A be a set of minimal size such that $h(A)$ is not integer. Since Γ is connected, A contains a non-redundant participant so there exists a minimal $B \subseteq \mathcal{P}$ such that $AB \in \Gamma$ while $B \notin \Gamma$.

First, notice that Proposition 10 gives

$$h(B|A) = \sum_{i=1}^{|B|} h(b_i | A, \{b_j : 1 \leq j \leq i-1\}) = |B|.$$

By minimality of B, there is an non-empty subset A' of A such that $A'B \in \Gamma^-$ is a minimal authorized group.

Thus, for any participant $a \in A'$, we have

$$\begin{aligned} h(a) &= 1, \\ h(a|A'-a, B) &= 1, \\ h(a|A'-a, Bs) &= 0, \\ h(AB) &= h(ABs) = h(A-a, Bs), \\ h(AB) &= h(A) + h(B|A) = h(A) + |B|, \end{aligned}$$

Notice that $|B| = h(B|A) \leq h(B|A-a) \leq |B|$, thus we can write

$$\begin{aligned} h(A-a, Bs) &= h(A-a) + h(B|A-a) + h(s|A-a, B) \\ &= h(A-a) + |B| + h(s|A-a, B). \end{aligned}$$

Summarizing: $h(A) = \underbrace{h(A-a)}_{\text{integer}} + \underbrace{h(s|A-a, B)}_{\text{0 or 1}} \in \mathbb{N}$. □

Theorem 20 provides only a partial characterization of ideal schemes. The question is now to determine whether a matroid port has an ideal secret scheme. We formulate an answer in what follows.

A Characterization of Matroid Ports

Matroids are stable under two basic operations:

Definition 26 (Minors)**.** *Let $\mathfrak{M} = (\mathcal{Q}, f)$ be a matroid defined by its rank function and $Z \subseteq \mathcal{Q}$ be a subset of its ground set, the function defined by*

- $\mathfrak{M}_{\setminus Z} = (\mathcal{Q} \setminus Z, f_{\setminus Z})$ *where $f_{\setminus Z}$ is defined by $f_{\setminus Z}(A) = f(A)$,*
- $\mathfrak{M}_{/Z} = (\mathcal{Q} \setminus Z, f_{/Z})$ *where $f_{/Z}$ is defined by $f_{/Z} = f(A \cup Z) - f(Z)$,*

are also matroids and are called minors *of $\mathfrak{M}$.*

The following result of Seymour characterizes the class of all matroid ports in terms of excluded minors.

Lemma 8 (Seymour, [Sey76])**.** *An access structure is a matroid port if and only if it has no minor isomorphic to P_4, Q_4, Q_4^* or J_s, for $s \geq 3$, where J_s is the access structure defined by $\mathcal{P} = \{p_0, \ldots, p_s\}$ such that $J_s^- = \{p_1, \ldots, p_s\} \cup \{p_0 p_i : 1 \leq i \leq s\}$.*

This result has been used to notice the following surprising separation result for matroid-related access structures.

Theorem 21 (Martí-Farré *et al*, [MFP07])**.** *Let Γ be an access structure. If G is not a matroid port then $\rho(G) \geq \frac{3}{2}$.*

This is an improvement on the result by Brickell and Davenport. The result follows by checking that each excluded minor for matroid ports cannot be implemented with an information ratio less than $\frac{3}{2}$.

Multilinear and Non-representable Matroids

There are matroids which are not linearly representable over any finite fields. For instance, the non-Pappus matroid.

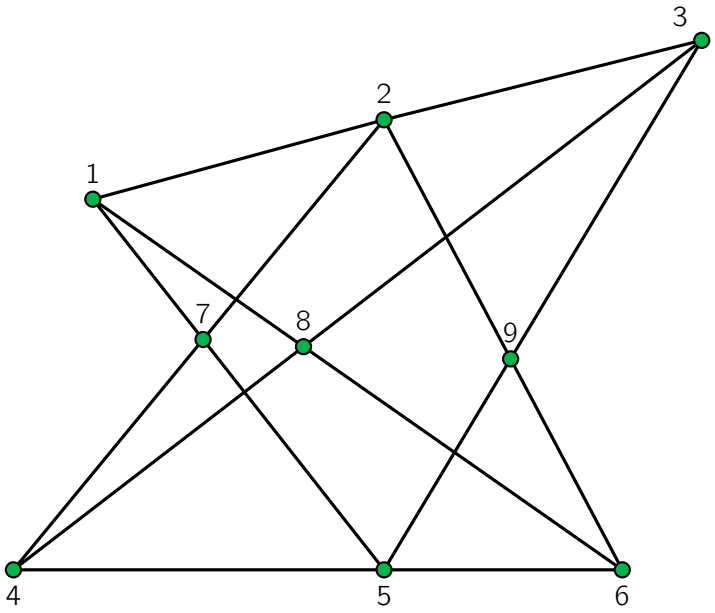


Figure 3.7: A picture of the non-Pappus matroid

Simonis and Ashikmin [SA98] and Matúš [Mat99b] constructed an ideal secret-sharing scheme for access structures related to the *non-Pappus* matroid. Such matroid are said multilinear because the secret can not be represented as subspace of rank 1. For the non-Pappus matroid, the rank of the secret must be at least 2. In terms of entropic points, this result means that the rank function r of the non-Pappus matroid cannot be represented by any linearly entropic points, but $2 \cdot r$ can.

However, some matroid-related access structures do not even admit ideal secret schemes at all. This is the case for the Vámos matroid. The Vámos matroid is defined on the ground set of 8 elements $[\![8]\!]$ and its bases are all subsets of size four except $\{1,2,3,4\}$, $\{1,2,5,6\}$, $\{3,4,5,6\}$, $\{3,4,7,8\}$ and $\{5,6,7,8\}$. The two related non-isomorphic access structures are V_1 and V_6, the matroid ports through the element 1 and 6 respectively. This matroid is not linearly representable because it does not satisfy Ingleton Inequality. Seymour proved moreover in [Sey92], that the access structures related to the Vámos matroid are not ideal. This result was further developed in [BO09, BLP08, MB11, MB09] and the current bound on the information ratio for Vámos matroid is given by the following result.

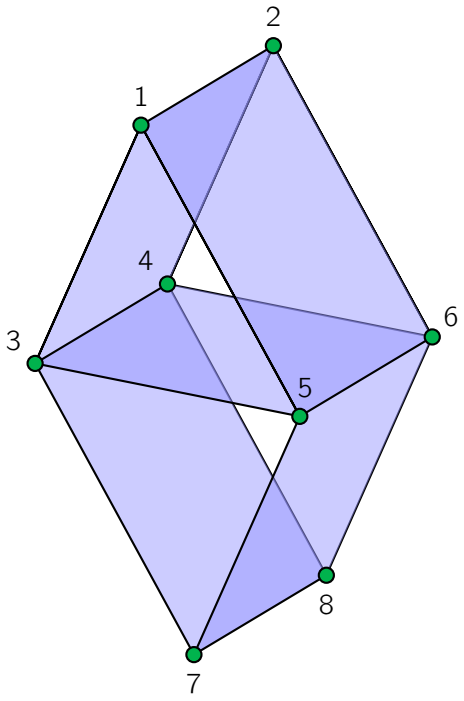


Figure 3.8: A picture of the Vámos matroid

Theorem 22 (Metcalf-Burton, [MB11])**.** *The two non-isomorphic access structures V_1, V_6, related to the Vámos matroid satisfy*

$$\rho(V_1) \geq \frac{9}{8}$$
$$\rho(V_6) \geq \frac{19}{17}$$

3.6 Lower Bounds on the Information Ratio from Information Inequalities

As we have seen in Section 3.4.2, not every access structure is ideal. We explained why no ideal scheme exists for P_4: the access structure with four participants a, b, c, d where the authorized groups are $\{a, b\}$, $\{b, c\}$ and $\{c, d\}$ and all their supersets. We proved that every secret-sharing scheme for this access structure satisfies

$$\log |\mathcal{S}_b| + \log |\mathcal{S}_c| \geq 3 \log |\mathcal{K}|.$$

This follows from Proposition 12 and the fact that the secret can be drawn according to the uniform distribution (remember our discussion in Section 3.2 that the distribution on secret is irrelevant).

It turns out that information inequalities are the only known generic tool to prove bounds for general access structures. Other combinatorial technique have also shown to be punctually useful for this task (see for instance [Sey92, KO96, BL88]). However these bounds can also be achieved, if not improved, by the entropy method. The general technique of the so-called entropy method consists in applying an information inequality for the secret-sharing tuple, taking into account the perfect secret-sharing requirements, and hope that the inequality

rewrites into a nontrivial bound for the entropy of a group of participants. Notice that this is exactly what happened in the proof of Proposition 12 p. 47. In fact, we must underline that the whole argument can be reduced to a single conditional information inequality.

Proofs that use the entropy method usually bound the entropy of the shares of a group. In fact, one cannot prove a nontrivial bound for the share of one single participant.

Lemma 9 (Blundo *et al*, [BSV98])**.** *Let Γ be an access structure, $\mathcal{K}$ be a set of at least two secrets and $p \in \mathcal{P}$ a participant. There exist a perfect secret-sharing scheme for Γ such that*

$$H(p) = \log |\mathcal{K}|.$$

Proof. Notice that for a uniformly distributed secret, this means that $H(p) = H(s)$. The idea is to implement Γ by focusing first on participant p (and assigning him only one share of the size of the secret, the same idea was used for dictators in Paragraph 3.2.1).

Let the secret s be a random element of a field $\mathbb{F}$. We implement Γ as follows:

- Assign the share $r + s$ to p, where r is uniformly distributed on $\mathbb{F}$.
- Share the value r to the substructure $\{A - p : p \in A \in \Gamma^-\}$ on $\mathcal{P} - p$. Minimal groups of Γ^- containing p can now recover the secret.
- Share s to the substructure $\{A : p \notin A \in \Gamma^-\}$ on $\mathcal{P} - p$. Minimal groups of Γ^- that do not contain p can now also recover the secret.

Thus, we implemented Γ and p has a share in $\mathbb{F}$, which concludes our proof. □

3.6.1 The Independent Sequence Technique

In what follows, we present a general technique for proving lower bounds. It first appeared in [BSV94] and was later used in [Csi97, BSSV97, MFP07].

Definition 27 (Independent sequence, [BSV94])**.** *Let Γ be an access structure. For a group $A \subseteq \mathcal{P}$ and groups $B_1 \subseteq \ldots B_m \subseteq \mathcal{P}$, we say that $(B_1, \ldots, B_m|A)$ is an independent sequence in Γ of length m and size s if $|A| = s$, $B_m \notin \Gamma$ and there exist subsets $X_1, \ldots, X_i \subseteq A$ such that*

- *$B_i \cup X_i \in \Gamma$ for all $1 \leq i \leq m$, and*
- *$B_i \cup X_{i+1} \notin \Gamma$ for all $1 \leq i \leq m-1$.*

The independent sequence technique can be subsumed by the following lemma:

Lemma 10 (Independent Sequence Method)**.** *If an access structure Γ admits an independent sequence of length m and size s, then $\rho(\Gamma) \geq \frac{m}{s}$.*

Proof. Let Γ be an access structure and $(B_1, \ldots, B_m|A)$ an independent sequence of length m and size s. For all $1 \leq i \leq m-1$:

$$H(B_{i+1}X_{i+1}) + H(B_iA) \geq H(B_{i+1}A) + H(B_iX_{i+1}) + H(s),$$
$$H(B_{i+1}) + H(B_iX_{i+1}) \geq H(B_{i+1}X_{i+1}) + H(B_i).$$

Adding these two inequalities gives

$$H(B_{i+1}) + H(B_iA) \geq H(B_{i+1}A) + H(B_i) + H(s).$$

Summing up these inequalities for all $1 \leq i \leq m-1$ gives

$$H(B_m) + H(B_1A) \geq H(B_mA) + H(B_1) + (m-1)H(s),$$
$$\text{or equivalently } H(A|B_1) \geq H(A|B_m) + (m-1)H(s)$$

Since $B_m \notin \Gamma$ and $A \cup B_m \in \Gamma$, we get

$$H(A) \geq H(A|B_1) \geq H(A|B_m s) + mH(s),$$

which implies the result. □

Remark 7. *In Proposition 12, we proved that P_4 and all access structure shown in Figure 3.3 admit an independent sequence of length 3 and size 2: $(\varnothing, a, ad|bc)$.*

László Csirmaz used this technique to prove the best known general lower bound on the size of the shares of a perfect secret-sharing scheme.

Theorem 23 (Csirmaz, [Csi97])**.** *There is a family of access structure $(\Gamma)_n$ on n participants such that*

$$\rho(\Gamma) \geq \Omega\left(\frac{n}{\log_2 n}\right).$$

Proof. Let us construct such a family: Take $m \geq 2$ to be an integer. Let $B = \{b_1, \ldots, b_{2^m-2}\}$ be a set of participants and $A = \{a_1, \ldots, a_m\}$ be another set of participants such that $A \cap B = \varnothing$ and $\mathcal{P} = A \cup B$. For $i \leq 1 \leq 2^m - 2$, define

$$X_i = \left\{a_i : \text{the } j\text{-th most significant bit of } \bar{i}_2 \text{ is } 0\right\},$$

assuming the base 2 representation is padded with 0 on the left. Also, denote $B_i = b_1 b_2 \ldots b_i$, now define the minimal authorized groups of Γ to consist of A and all the B_iX_i. By construction, $B_iX_{i+1} \notin \Gamma$, because $B_i \subseteq B_j$ iff $1 \leq i \leq j \leq 2^m - 2$ and $X_i \subseteq X_j$ only if $j \leq i$, so $B_iX_{i+1} \not\subset B_jX_j$ for any j.

Therefore $(\varnothing, B_1, B_2 \ldots, B_{2^m-2}|A)$ is an independent sequence of length $2^m - 1$ and size m. The number of participants in Γ is $n = 2^m - 2 + m$, by Lemma 10,

$$\rho(\Gamma) \geq \frac{2^m - 1}{m} = O\left(\frac{n}{\log n}\right).$$

Thus we have proven Theorem 23. □

For $m = 2$, the access structure defined by our construction is P_4. For $m = 3$, the constructed independent sequence is represented in the figure below.

	$\varnothing$	b_1	b_2	b_3	b_4	b_5	b_6
B_i	000	001	010	011	100	101	110
X_i	abc	ab	ac	a	bc	b	c

Figure 3.9: An independent sequence of length 7 and size 3

3.6.2 The Need for New non-Shannon Inequalities

Csirmaz also investigated the limits of the previous technique and proved that the bound obtained by the independent sequence cannot be improved much. In fact he proved that any general bound obtained using a Shannon-type inequality cannot be greater that $\Omega(n)$.

Theorem 24 (Csirmaz, [Csi97]). *For every access structure Γ, there exists a point $\vec{h}$ satisfying all Shannon-type inequalities and the perfect secret sharing requirement for Γ such that $h_i = n$ for all $i \in [\![n]\!]$.*

So for any access structure, there is a simple point which satisfies all Shannon-type inequalities which could be the entropy profile of a scheme for Γ. This proves a limit of the power of Shannon-type inequalities. *What about non-Shannon type inequalities ?* Amos Beimel and Ilan Orlov extended the result of Csirmaz to all non-Shannon-type inequalities known up to early 2010, and proved they cannot help improve the $\Omega(n)$ lower bound. They also proved it to be the case for all valid rank inequality, all inequalities in 4 or 5 variables.

However non-Shannon type inequalities do help for improving the bound on the information ratio for fixed instances (see [PVY12, MB11, BL08, BLP08, Csi09]). This kind of application of non-Shannon type was first used to prove that the Vámos matroid is not nearly-ideal. The whole argument of the proof of Theorem 22 indeed consists in explaining why some non-Shannon-type inequality of Dougherty *et al* (see [DFZ06]) are indeed enough to prove the bound. This type of argument also works for other matroids. To date, the most general result proves that a finite family related to the Vámos matroid are non-ideal (see [Alb11]).

Discussion & Comments

The use of (n, n)-threshold as building blocks for general access structures can be exported to other models of secret sharing (e.g. by Csirmaz in [Csi12]).

General ideal perfect secret-sharing schemes have been shown to be related to many combinatorial objects such as Orthogonal Arrays [Daw93, Mar91], MDS codes [PZ03], Quasigroup equations [Mat99b]. The notion of a linear secret sharing has been introduced in various forms, Simmons, and Jackson and Martin called them geometric schemes in [Sim90, JM94]. Brickell introduced the Vector Space Construction (see [Bri90]) which was further generalized in [Dij97, Csi09]. Massey showed a relation with linear error correcting codes in [Mas95], Widgerson and Karchmer proved their equivalence with (multi-target) Monotone Span Programs in [KW93].

Unfortunately no textbook on secret sharing exists as of 2012, the curious reader may still be interested in the surveys by Stinson [Sti92], and Beimel [Bei11], by a course by Padro [Pad12] and by the book of Kabatiansky [Kab98].

Chapter 4

An Algorithmic and Information-theoretic Toolbox

Contents

Introduction

This chapter mainly serves as a Toolbox and compiles several key lemmas and results that shall be used in the next chapters. The generality of these results makes them interesting on their own. As an example of use of this toolbox, we show the equivalence of two techniques for creating non-Shannon-type information inequalities. We introduce Algorithmic Information Theory or (exchangeably) Kolmogorov Complexity Theory whose undisputed fathers are Andrei Kolmogorov, Gregory Chaitin and Ray Solomonoff.

We start with a tool from probability theory. Höeffding's inequality, here presented in a slightly less general case, is a bound on the probability that the average value of a sum of independent and identically distributed random variables, deviates from its expectation.

Lemma 11 (Höffding Inequality, [Hoe63])**.** *Let $X_1, \ldots, X_n$ be n i.i.d. real random variables whose range is $[a, \ldots, b]$. If $\overline{X} = \sum_{i=1}^{n} X_i$, then*

$$\Pr(\overline{X} - \mathbb{E}[\overline{X}] \leq t) \geq \exp\left(-\frac{2nt^2}{(b-a)^2}\right)$$

Note that the i.i.d. condition can be changed to the assumption that the X_i are obtained via a sampling without replacement.

4.1 Tools from Shannon's Information Theory

4.1.1 Quasi-Uniform Random Variables

We introduce a new class of random variables studied in [Cha01] and further investigated in [CY02].

Definition 28 (Quasi-uniform random tuple)**.** *A tuple of random variables $x_{\mathcal{N}} = (x_i)_{i\in\mathcal{N}}$ is called* quasi-uniform *if for each subtuple x_J, all of its values that occur with positive probabilities are equiprobable.*

We call quasi-uniform *an entropic point corresponding to the entropy profile of a tuple of quasi-uniform random variables. Denote by Λ_n the set of all quasi-uniform entropic points for n-tuples.*

Terence Chan introduced quasi-uniform random variables in [Cha01] along with their underlying combinatorial representations (called *box assignments*). They are structured random variables: knowing the support of a quasi-uniform random tuple is indeed sufficient to reconstruct its whole distribution. The support can be seen as a (multidimensional) binary array such that for each coordinate, the number of one in every section is the same.

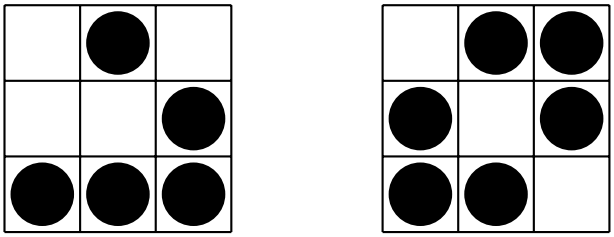


Figure 4.1: Two box assignments: a non-quasi-uniform on the left and a quasi-uniform on the right

The quasi-uniform property for a random tuple $X_{\mathcal{N}}$ can be stated as follows: for any two values $b_i, b_j \in \mathcal{S}_B$ of B we have

$$H(A|B = b_i) = H(A|B = b_j)$$

for any subtuples A, B of $X_{\mathcal{N}}$. Not all distributions are quasi-uniform since the described property of conditional probabilities in general does not hold (see page 13 for a reminder). An example of quasi-uniform random variables are linear random variables (presented in Section 2.5 p. 22).

It turns out that quasi-uniform random variables are enough to characterize the whole entropy region $\bar{\Gamma}^*_n$. Terence H. Chan proved together with Raymond W. Yeung, that set of quasi-uniform entropic point and Γ^*_n are described by the same set of linear inequalities

Theorem 25 (Theorem 4.2, [Cha01])**.** $\overline{\mathbf{con}(\Lambda_n)} = \bar{\Gamma}^*_n$. *where* **con** *is the convex envelope defined on p. 25.*

We are interested in the main lemma used to prove the previous theorem which can be subsumed as follows:

Lemma 12 (Theorem 4.1 [CY02])**.** *For every distribution $x_{\mathcal{N}} = (x_i)_{i\in\mathcal{N}}$ and all $\varepsilon > 0$ there exists a quasi-uniform distribution $y_{\mathcal{N}} = (y_i)_{i\in\mathcal{N}}$ and an integer k such that*

$$\left\| \vec{H}(x_{\mathcal{N}}) - \frac{\vec{H}(y_{\mathcal{N}})}{k} \right\| < \varepsilon.$$

(in some norm; the choice is not important since ε is arbitrarily small, and all norms differ at most by a constant factor).

4.1.2 The Slepian–Wolf Coding Theorem

Say you want to download two copyright-free electronic books X and Y from two different web servers. These two books are from the same author, so the contents of X and Y have some common traits – style, vocabulary, punctuation. One could get the two files separately and download a total amount of information which is equal to the sum of the compressed sizes of X and Y. Assuming that X and Y are values of some random variables, the total traffic is equal to $H(X) + H(Y)$. Can smart-coded web servers exploit the fact that X and Y are correlated and send less information ?

Yes, they can. If they agree in advance, one server can send the whole file X while the second only has to send around $H(Y|X)$ bits of information (without knowing which file X the first server sent). How can this compression happen ?

Wolf and Slepian proved a famous theorem on the characterization of the rate region of this problem – the problem of independent compression of two correlated random variables. Here we only present a special case of the classic Slepian–Wolf theorem. This case actually makes the most important part of the general proof of the standard Slepian–Wolf theorem (see Theorem 2 in the original paper [SW06] and a detailed discussion in Section 5.4.4 of [CT91]).

Lemma 13 (Slepian–Wolf coding)**.** *Let X, Y be N independent copies of random variables x, y,* i.e., *$X = (x_1, \dots, x_N)$, $Y = (y_1, \dots, y_N)$, where pairs (x_i, y_i) are i.i.d. Then there exists X' such that*

- $H(X'|X) = 0$,
- $H(X') = H(X|Y) + o(N)$,
- $H(X|X', Y) = o(N)$.

(This Lemma is also a special case of Theorem 3 in [Mat07c].) Lemma 13 claims that we can construct a hash of a random variable X which is almost independent of Y and has approximately the entropy of X given Y. We will say that X' is the Slepian–Wolf hash of X given Y and write $X' = SW(X|Y)$.

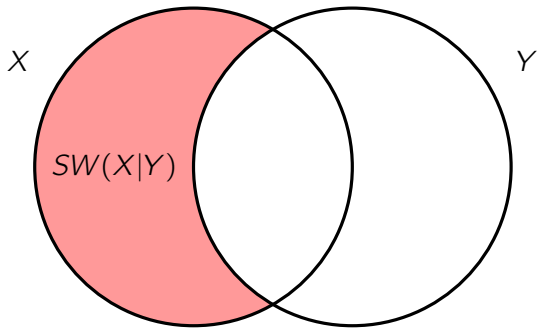


Figure 4.2: A Representation of the Slepian–Wolf Hash $SW(X|Y)$

4.1.3 The Copy Lemma

The idea of the copy lemma was the original trick from [ZY97] to prove the first nontrivial conditional inequality. The lemma was formulated by Randall Dougherty *et al* in [DFZ06]. We present a formalization of this lemma in our notations.

Lemma 14 (Copy lemma, [DFZ06])**.** *Let A, B, C be three jointly distributed random variables. There exists a fourth random variable A' such that (A, B) and (A', B) have the same entropy profile and A' is independent of A, C given B,* i.e.*,*

$$\vec{H}(A, B) = \vec{H}(A', B) \text{ and } I(A' : AC|B) = 0.$$

*Such an A' is called a C-*copy of A given B*.*

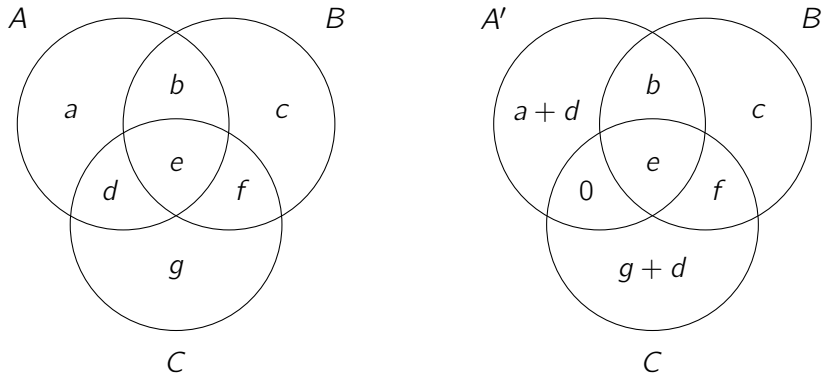


Figure 4.3: Information diagrams for a copy A' of A

The lemma follows from a simple construction for random variables. Suppose you are given A, B, C and want to construct such an A'. We construct a distribution (A', A, B, C) satisfying the properties of Lemma 14:

1. First, sample B using p_B.
2. For each value of C, sample independently (A, C) using $p_{AC|B}$ and A' using $p_{A|B}$.

In this way, we have automatically the required conditional independence property, and by construction $(A, B, C) \sim p_{ABC}$ and $(A', B) \sim p_{AB}$.

In fact, we used such a construction for our proof of the basic inequality using the KL-divergence (see Section 2.3.1 p. 14) and for the proof of some of the conditional and unconditional inequalities (see Section 2.7 p. 28).

4.1.4 The Ahlswede–Körner Lemma

By Ahlswede–Körner lemma, we refer to a result from the works of Ahlswede–Gács–Körner presented in its plain generality in a book by Csiszár and Körner [CK81], whose relevance was underlined by Wyner in [Wyn06]. The result has been presented for the special case of 3 random variables (a generalization of Wyner's result relevant for our purposes) in Section 2.7.2 (p. 34) Lemma 3 can be generalized to any number of variables in the following way.

Lemma 15 (Ahlswede–Körner Lemma, [AGK76, CK81])**.** *Let $x_1, \dots, x_n, y$ be n jointly distributed random variables. Consider their M-serializations $X_1, \dots, X_n, Y$. Then there exists a random variable W such that*

$$H(W|X_1 \dots, X_n) = 0,$$
$$\left\| \vec{H}(X_1, \dots, X_n|W) - m \cdot \vec{H}(x_1, \dots, x_n|y) \right\| = o(M).$$

Denote this W by $AK(Y : X_1, \dots, X_n)$.

For three random variables and in terms of almost entropic points, this lemma states that if there is an entropic point whose profile is shown on the left of Figure 4.4, then there is an almost entropic point whose profile is shown on the right of the figure.

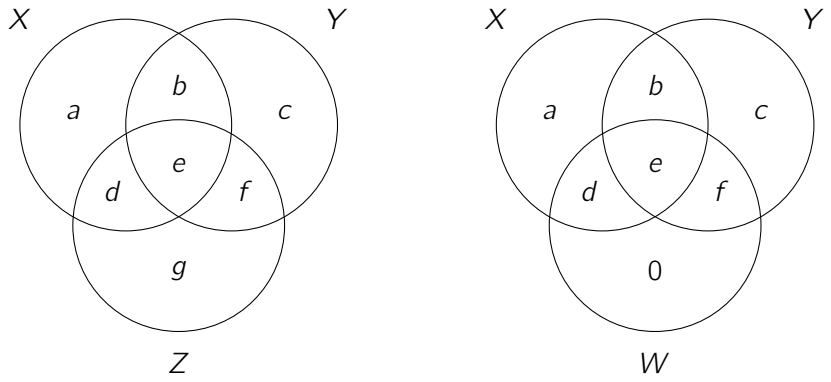


Figure 4.4: Ahlswede–Körner Lemma: from the entropic point on the left, one can construct the almost entropic point on the right.

We proceed to prove two non-Shannon-type information inequalities, one of which (inequality (4.1)) was originally proven using the Copy Lemma in [DFZ06].

Proposition 19. *The following are two* 5*-variable non-Shannon-type inequalities.*

$$I(a{:}b) \le I(ar{:}d)+I(b{:}c|ar)+I(b{:}c|d)+I(a{:}b|cr)+I(a{:}c|br)+I(a{:}b|c)+I(a{:}r|bc) \quad (4.1)$$

$$I(a{:}b) \le I(ar{:}d)+I(b{:}c|ar)+I(b{:}c|d)+I(a{:}b|cr)+I(a{:}c|br)+I(a{:}b|c)+I(d{:}r|bc) \quad (4.2)$$

Proof. Denote by A, B, C, R the N-serialization of a, b, c, r and let $W = AK(a{:}bcr)$. The following (in)equalities hold up to $o(N)$:

$$\begin{aligned}
H(W) &\le H(W|D) + H(W|AR) + I(D{:}AR)\\
H(W|D) &\le H(W|B) + H(W|C) + I(B{:}C|D) - H(W|BCD)\\
H(W|AR) &\le H(W|BR) + H(W|CR) + I(B{:}C|AR)\\
H(W|B) &= I(CR{:}A|B)\\
H(W|C) &= I(A{:}R|BC) + I(A{:}B|C)\\
H(W|BR) &= I(A{:}C|BR)\\
H(W|CR) &= I(A{:}B|CR)\\
I(A{:}B) + I(CR{:}A|B) &= H(W) \quad (= I(BCR{:}A))
\end{aligned}$$

By summing we obtain:

$$\begin{aligned}
I(A{:}B) \le I(AR{:}D) + I(B{:}C|AR) + I(B{:}C|D) + I(A{:}B|CR) +\\
+ I(A{:}C|BR) + I(A{:}R|BC) + I(A{:}B|C) - H(W|BCD),
\end{aligned}$$

which implies (4.1). Now,

$$\begin{aligned}
H(W|BCD) &= H(W|BC) + I(W{:}D|BC)\\
&= I(A{:}R|BC) - I(W{:}D|BC)\\
&\ge I(A{:}R|BC) - [I(R{:}D|BC) + I(W{:}D|BCR)]\\
&\ge I(A{:}R|BC) - I(R{:}D|BC)
\end{aligned}$$

By summing the last two inequalities, we obtain an inequality that implies (4.2).

To see why these inequalities are of non-Shannon type, take r to be deterministic and use the equality:

$$I(a{:}b) = I(b{:}c) + I(a{:}b|c) - I(b{:}c|a).$$

Both inequalities rewrite to an instance of the non-Shannon type inequality (2.14) from Theorem 13 (p. 34) □

4.2 Comparison of Two Proof Techniques for Information Inequalities

Before formalizing the two techniques we have already seen earlier (Chapter 2), we give another theorem on information inequalities.

Theorem 26 (Balanced inequalities, Chan, [Cha03])**.** *Let $(\lambda)_J$ be a list of coefficients, the following are equivalent:*

1. *The inequality*
$$f(X_{\mathcal{N}}) = \sum_{\varnothing \neq J \subseteq \mathcal{N}} \lambda_J H(X_J) \geq 0$$
is a valid information inequality.

2. *The inequality*
$$g(X_{\mathcal{N}}) = \sum_{\varnothing \neq J \subseteq \mathcal{N}} \lambda_J H(X_J) - \sum_{j \in \mathcal{N}} r_j H(X_j|X_{\mathcal{N}-j}) \geq 0,$$
where r_j is the sum of all λ_J involving j, is a valid information inequality.

The latter inequality is said balanced.

For a balanced inequality in atomic form, the coefficients of the terms $H(X_j|X_{\mathcal{N}-j})$ are all zeros. The previous result states that all information inequalities can be put in balanced form by deleting the corresponding terms. These coefficients must obviously be non-negative, therefore the balanced inequality appears to be stronger.

We describe the main two techniques for obtaining the non-Shannon-type information inequalities presented in Chapter 2.

Technique 1: Zhang-Yeung's method. We present this technique as an inference rule:

(A) If we have an information inequality of the form:
$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g(Y_{\mathcal{M}}, Z) + \alpha I(Z{:}X_{\mathcal{N}}|Y_{\mathcal{M}}) \geq 0,$$
for some $\alpha \geq 0$;

(B) then the following stronger inequality is also valid:
$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g(Y_{\mathcal{M}}, Z) \geq 0.$$

Correctness. The correctness of this rule follows from Lemma 14: Take Z' to be a $X_{\mathcal{N}}$-copy of Z given $Y_{\mathcal{M}}$ and apply inequality of step (A) with $Z = Z'$.

Technique 2: Romashchenko's method. The method is subsumed in the following inference rule:

(A) If we have an information inequality of the form:

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g(Y_{\mathcal{M}}, Z) \geq 0;$$

(B) then the following stronger inequality is also valid:

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g(Y_{\mathcal{M}}, Z) - r_Z H(Z|Y_{\mathcal{M}}) \geq 0.$$

where r_Z is the sum of the coefficients of g involving Z.

Correctness. The correctness of this rule follows from Lemma 15 by taking $W = AK(Z{:}Y_{\mathcal{M}})$ (the proof is similar to the one of Theorem 13 on page 34).

We are going to show that these two inference rules are equivalent if we keep in mind Theorem 26 on balanced inequalities. By this we mean the following:

Theorem 27. *For any information inequality $\mathcal{I}_1$ inferred using Technique 1, there is an inequality $\mathcal{I}_2$ inferred using Technique 2 such that $\mathcal{I}_1$ and $\mathcal{I}_2$ have the same balanced form. The converse also holds.*

Proof.
$1 \Rightarrow 2$: Suppose

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g(Y_{\mathcal{M}}, Z) + \alpha I(Z{:}X_{\mathcal{N}}|Y_{\mathcal{M}}) \geq 0, \tag{A_1}$$

for some $\alpha \geq 0$, is a valid information inequality in balanced form. By Technique 1, the stronger

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g(Y_{\mathcal{M}}, Z) \geq 0 \tag{B_1}$$

is also valid. Let us show that inequality (B_1) can also be obtained using Technique 2.

Start from the inequality

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g'(Y_{\mathcal{M}}, Z) \geq 0 \tag{A_2}$$

defined using $g' = g + \alpha H(Z|Y_{\mathcal{M}})$, where r_Z is the sum of coefficients of g involving Z. This inequality is valid since α is non-negative, thus (A_2) follows from $H(Z|Y_{\mathcal{M}}) \geq I(Z{:}X_{\mathcal{N}}|Y_{\mathcal{M}})$ and (A_1).

Using the inference rule of Technique 2, we get

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g'(Y_{\mathcal{M}}, Z) - r'_Z H(Z|Y_{\mathcal{M}}) \geq 0, \tag{B_2}$$

where r'_Z is the sum of coefficients of g' involving Z. By definition of g' we have $r'_Z = \alpha + r_Z$, thus inequality (B_2) rewrites to

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g(Y_{\mathcal{M}}, Z) - r_Z H(Z|Y_{\mathcal{M}}) \geq 0,$$

and since (A_1) is balanced we have $r_Z = 0$, so the last inequality is exactly inequality (B_1).

$2 \Rightarrow 1$: Suppose

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g(Y_{\mathcal{M}}, Z) \geq 0 \tag{A'_2}$$

is a valid information inequality. By Technique 2, the stronger

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g(Y_{\mathcal{M}}, Z) - r_Z H(Z|Y_{\mathcal{M}}) \geq 0 \tag{B'_2}$$

is also valid. Let us show that inequality (B'_2) can also be inferred using Technique 1 and balancing.

Notice first that $H(Z|Y_{\mathcal{M}}) = H(Z|X_{\mathcal{N}}Y_{\mathcal{M}}) + I(Z{:}X_{\mathcal{N}}|Y_{\mathcal{M}})$, therefore ($A'_2$) rewrites to

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + [g(Y_{\mathcal{M}}, Z) - r_Z H(Z|Y_{\mathcal{M}})] + r_Z H(Z|X_{\mathcal{N}}Y_{\mathcal{M}}) + r_Z I(Z{:}X_{\mathcal{N}}|Y_{\mathcal{M}}) \geq 0,$$

where r_Z is the sum of the coefficients of g involving Z. Balancing this inequality gives:

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + [g(Y_{\mathcal{M}}, Z) - r_Z H(Z|Y_{\mathcal{M}})] + r_Z I(Z{:}X_{\mathcal{N}}|Y_{\mathcal{M}}) \geq 0 \qquad (A'_1)$$

Applying the inference rule of Technique 1 to (A'_1) gives

$$f(X_{\mathcal{N}}, Y_{\mathcal{M}}) + g(Y_{\mathcal{M}}, Z) - r_Z H(Z|Y_{\mathcal{M}}) \geq 0, \qquad (B'_1)$$

which is exactly inequality (B'_2). □

4.3 Tools from Kolmogorov Complexity

After his successful axiomatization of probability theory, Andrei N. Kolmogorov reportedly thought that his accomplishment was not satisfactory, in the sense that the theory failed to provide a definition of random objects. Indeed, suppose one sampled the following two sequences of `ASCII` characters of length 52, by selecting each character uniformly at random.

```
Mais que diable allait-il faire dans cette galère ?
T1j-k%<?BveYa3pUh9R4HVPqWy=p;W9hXxTGV09Auhs!1!!one!
```

The first one would clearly appear not random for most of us, whereas the second one would seem to look like an absolutely reasonable "random" sequence. Our intuition tells us this is what most sequences "look like". However both sequences had the same probability 256^{-52} of appearing.

We would like to be able to say that the first one is not random, because we could simply open the right book and find it, or we could compress it because the characters of the French language (or of any spoken language for that matter) show some interdependencies. We would like to be able to say that the second sequence is thus more complex since it does not appears to show any correlation between characters.

Kolmogorov Complexity Theory tries to formalize these ideas by defining the complexity of a string as its compressed size. Using this idea, one can define randomness as incompressibility. We refer the reader to the textbook of Li and Vitanyi (see [LV97]) for an exhaustive survey on Kolmogorov complexity.

4.3.1 Introduction to Algorithmic Information Theory

The Kolmogorov complexity of a finite binary string x is defined as the length of a shortest program that prints x on the empty input. Similarly, the conditional Kolmogorov complexity of a string x given another string y is defined as the length of a shortest program that prints x given y as an input. More formally, for any programming language $\mathcal{L}$, the Kolmogorov complexity $C_{\mathcal{L}}(x|y)$ is defined as

$$C_{\mathcal{L}}(x|y) = \min\{|p| : \text{ program } p \text{ prints } x \text{ on input } y\},$$

and unconditional complexity $C_{\mathcal{L}}(x)$ is defined as complexity of x given the empty y. Our programming language must be powerful enough. The programming model does not much matter; for simplicity we assume that $\mathcal{L}$ is computable (e.g., by a Turing machine, or by some other machine).

The basic fact of Kolmogorov complexity theory is the invariance theorem:

Theorem 28 (Invariance Theorem)**.** *There exists an optimal programming language $\mathcal{U}$ such that for any other language $\mathcal{L}$ we have $C_{\mathcal{U}}(x|y) \leq C_{\mathcal{L}}(x|y) + O(1)$ (the constant depends on $\mathcal{L}$ but not on x and y).*

There exist infinitely many programming languages that satisfy this theorem. We fix one such optimal language $\mathcal{U}$. In what follows we omit the subscript $\mathcal{U}$ and denote the Kolmogorov complexity by $C(x)$, $C(x|y)$.

Let us find out more properties of this complexity notion. It should be natural that any deterministic procedure cannot increase too much the complexity of strings:

Proposition 20 (Computable functions cannot increase the complexity)**.** *$C(f(x)) \leq C(x) + c_f$, for any program f. The constant c_f depends only on f.*

Now we can give a trivial upper bound for the complexity. Any string can be produced by some program. The identity program just reproduces its input as an output: any string is its own description:

Proposition 21 (A string is at most as complex as itself)**.** *$C(x) \leq |x| + c$, for any string x. The constant c is absolute.*

Let us now check that this theory indeed can talk about random strings. By "random" string we mean an "incompressible" string, that is : any description should have approximately the same length as the string itself. For our theory to be interesting, it should indeed be able to prove that there exist random strings. It is indeed the case.

Proposition 22 (Incompressible strings)**.** *For each length n, there is at least one incompressible string.*

In fact, this situation is the typical one. Most strings cannot be compressed by more than a constant additive term:

Proposition 23 (Most strings are incompressible)**.** *The probability that a uniformly random binary string of length n satisfies $C(x) < |x| - k$ is 2^{-k}.*

Suppose now that we have two strings u, v that may share a certain number of bits in some way. Is it true that we can find a program for the pair (u, v) from a program for u and a program for v? The answer is yes:

Theorem 29. $C(u, v) \leq C(u) + C(v) + O(\max\{\log(C(u)), \log(C(v))\})$.

I suggest the reader to try to prove the previous theorem for a first encounter with basic arguments form Algorithmic Information Theory.

Stepping back, one might begin to realize that this complexity notion behaves a lot like the information-theoretic notion of entropy. Our next theorem should settle this thought definitively:

Theorem 30 (Symmetry of Information (Kolmogorov–Levin Theorem), [ZL70])**.** *For all strings u, v:*

$$C(u, v) = C(u) + C(v|u) + O(\log C(u, v)) \tag{KL}$$

This formula is a direct counterpart of the following formula from Shannon's Information Theory:

$$H(UV) = H(U) + H(V|U).$$

This result justifies the definition of the mutual information in this setting, which is an algorithmic version of Shannon's standard definition: the mutual information between binary strings is defined as

$$I(x:y) := C(x) + C(y) - C(x,y),$$

and the conditional mutual information is defined as

$$I(x:y|z) := C(x,z) + C(y,z) - C(x,y,z) - C(z).$$

From the Kolmogorov–Levin theorem it follows that $I(x:y)$ is equal to $C(x) - C(x|y)$, and the conditional mutual information $I(x:y|z)$ is equal to $C(x|z) - C(x|y,z)$ (all these equalities hold only up to logarithmic terms).

A major, yet necessary, drawback of this theory is the uncomputability of the Kolmogorov complexity function itself.

Proposition 24 (Uncomputability)**.** *The function $C(\cdot)$ is not computable,* i.e.*, there is no algorithm c such that $C(x) = c(x)$ for all inputs x.*

We introduce a notation for the complexity profile of a tuple of strings similar to the notation for entropy profile (defined in Section 2.6).

Definition 29 (Complexity Profile)**.** *Let $\mathcal{X} = (x_1, \ldots, x_n)$ be n binary strings, we define the* complexity profile *of* $\mathbf{x}$ *as*

$$\vec{C}(\mathbf{x}) = (C(x_1), C(x_2), \ldots, C(x_1,x_2), C(x_1,x_3), \ldots, C(x_1,x_2,\ldots,x_n))$$

is defined as the shortlex-ordered list of complexities associated with the subtuples x_J for each non-empty $J \subseteq \{1,\ldots,n\}$.

The conditional complexity profile $\vec{C}(\mathbf{x}|y)$ for a string y is defined by

$$\vec{C}(\mathbf{x}|x) = (C(x_1|y), C(x_2|y), \ldots, C(x_1,x_2|y), C(x_1,x_3|y), \ldots, C(x_1,x_2,\ldots,x_n|y))$$

4.3.2 Inequalities Are The Same

A fundamental result of Andrei Romashchenko's validates the link between Information and Algorithmic Information theories. He showed that this parallelism can be pushed to the full extent of inequalities, proving that for every linear inequality for Shannon entropy there exists a counterpart for Kolmogorov complexity.

Theorem 31 (Hammer *et al*, Romashchenko, [HRSV00, Rom00a])**.** *For each family of coefficients $\{\lambda_W\}$ the inequality*

$$\sum_i \lambda_i H(a_i) + \sum_{i<j} \lambda_{i,j} H(a_i,a_j) + \ldots \geq 0$$

is true for every distribution (a_i) if and only if for some constant κ, the inequality

$$\sum_i \lambda_i\, C(x_i) + \sum_{i<j} \lambda_{i,j}\, C(x_i,x_j) + \ldots + \kappa \log N \geq 0$$

is true for all tuples of strings (x_i) (N denotes the sum of the lengths of all strings x_i, constant κ does not depend on strings x_i).

4.3.3 Muchnik's Theorem

Muchnik's theorem is a counterpart of Slepian–Wolf theorem (see Section 4.1.2 p. 67).

Theorem 32 (Muchnik, [Muc02]). *For all strings x, y, there exists a string z such that*

$$\begin{aligned} |z| &= C(x|y), \\ C(z|x) &= O(\log n), \\ C(x|y,z) &= O(\log n), \end{aligned}$$

where $n = C(x,y)$. Such a string is not unique, we denote by $Much(x|y)$ the set of all such strings.

4.3.4 Typization-based Techniques

Typized Sets

Consider the complexity profile $\vec{C}(\mathbf{a})$ of an n-tuple $\mathbf{a} = (a_1, \ldots, a_n)$. Can we find a "clone" tuple $\mathbf{a}'$, which is different but has approximately the same complexity profile (say each components are equal up to a logarithmic term) ? The answer is yes, for instance you could flip one bit of one string a_i, it will only change the complexity by an additive term $O(\log N)$ (where N is the sum of lengths of all strings involved), but this is a bit too naive. In fact many rather different tuples are "clones" of $\mathbf{a}'$. We will see that as soon as a tuple $\mathbf{a}$ has a high enough complexity, many other clones of this $\mathbf{a}$ should also exist. A set of clones of a tuple will generally be referred to as a *typized* set.

Let us define the typized set T of $\mathbf{a}$ as the set of tuples $\mathbf{a}'$ such that every possible conditional complexity of subtuples is upper bounded by the corresponding conditional complexity for $\mathbf{a}$.

Definition 30 (A Typized set T, [Rom03b]).

$$T(\mathbf{a}) = \left\{\mathbf{a}' = (a_1, \ldots, a_n) : \forall U, V \subseteq [\![n]\!] : C(\mathbf{a}'_U|\mathbf{a}'_V) \leq C(\mathbf{a}_U|\mathbf{a}_U)\right\}$$

For now, this set contains also tuples with very small complexity, we take care of this hereafter. Note that our original tuple belongs to its typized set. First, we notice how the size of this set is bounded.

Claim 1. $\log|T(\mathbf{a})| \leq C(\mathbf{a}) + 1$

Proof. The number of possible $\mathbf{a}'$ such that $C(\mathbf{a}') \leq C(\mathbf{a})$ is bounded by the number of programs of size at most $C(\mathbf{a})$ which is $2^{C(\mathrm{a})+1}$. □

Claim 2. $\log|T(\mathbf{a})| \geq C(\mathbf{a}) - O(\log C(\mathbf{a}))$.

Proof. This set $T(\mathbf{a})$ can be enumerated if all possible conditional complexities for subtuples of $(a_1, a_2, \ldots, a_n)$ are given. As we said, T contains the initial tuple $\mathbf{a}$. Therefore, we can describe $\mathbf{a}$ by the parameters needed to enumerate T (*i.e.*, all numbers $C(\mathbf{a}_U|\mathbf{a}_V)$) and the ordinal number of $\mathbf{a}$ in this enumeration. It follows that the complexity of $\mathbf{a}$ is not greater than

$$\log|T| + O(\log C(\mathbf{a})).$$ □

Since this set is large, most of its elements should be of high complexity, *i.e.*, have a complexity close to the one of $\mathbf{a}$. So if we want to keep only the strings with high complexity, we should consider the following set:

$$T_c(\mathbf{a}) = \left\{\mathbf{a}' = (a'_1, \ldots, a'_n) : \left\|\vec{C}(\mathbf{a}) - \vec{C}(\mathbf{a}')\right\| \leq c \log C(\mathbf{a})\right\} \tag{4.3}$$

Claim 3. $\log |T_c(\mathbf{a})| = C(\mathbf{a}) - O(\log C(\mathbf{a}))$ *for large enough constant* c.

Proof. The upper bound is trivial: for each $\mathbf{a}'$ in $T_c(\mathbf{a})$ we have

$$C(\mathbf{a}') \leq C(\mathbf{a}) + c \log N,$$

so $\log |T_c(\mathbf{a})| \leq C(\mathbf{a}) + c \log N + 1$, like in the proof of Claim 1.

Now we need to prove that $T_c(\mathbf{a})$ cannot be too small. The two previous claims showed that

$$\log |T(\mathbf{a})| = C(\mathbf{a}) - O(\log C(\mathbf{a})).$$

Let us look at $T(\mathbf{a}) \cap T_c(\mathbf{a})$. Evidently $|T(\mathbf{a}) \cap T_c(\mathbf{a})| \leq |T|$, let us prove the other bound. If $\mathbf{a}' \in T(\mathbf{a}) \setminus T_c(\mathbf{a})$, then for some U, we have $C(\mathbf{a}'_U) < C(\mathbf{a}_U) - c \log C(\mathbf{a})$, but then

$$\begin{aligned} C(\mathbf{a}') &= C(\mathbf{a}'_U) + C(\mathbf{a}'_{\overline{U}}|\mathbf{a}'_U) + O(\log N) \\ &< C(\mathbf{a}) - c \log C(\mathbf{a}) + O(\log N), \end{aligned}$$

where the constant in the $O(\log N)$ term depends only on the number of strings in the tuples involved, but not on N. This means that if c is large enough (greater than the constant in the $O(\log N)$ term in the inequality above), then the cardinality of the set $T(\mathbf{a}) \setminus T_c(\mathbf{a})$ is only a small fraction of $|T(\mathbf{a})|$:

$$\log |T(\mathbf{a}) \setminus T_c(\mathbf{a})| < \log |T(\mathbf{a})| - c \log C(\mathbf{a}) + O(\log N).$$

Therefore $|T(\mathbf{a}) \cap T_c(\mathbf{a})|$ is large, which proves our claim. □

We fix the value c so that Claim 3 holds (c depends on the size n of the tuple, but not on $C(\mathbf{a})$).

Now we can show that a stronger property holds for clones.

Claim 4. *For every* $\mathbf{a}' \in T_c(\mathbf{a})$

$$C(\mathbf{a}'_U|\mathbf{a}'_V) = C(\mathbf{a}_U|\mathbf{a}_V) + O(\log C(\mathbf{a}))$$

for all $U, V \subseteq [\![n]\!]$.

Proof. By the Symmetry of Information Theorem 30 (p. 73):

$$\begin{aligned} C(\mathbf{a}'_U|\mathbf{a}'_V) &= C(\mathbf{a}'_{U\cup V}) - C(\mathbf{a}'_V) + O(\log C(\mathbf{a}'_{U\cup V})) \\ &= C(\mathbf{a}_{U\cup V}) - C(\mathbf{a}_V) + O(\log C(\mathbf{a}_{U\cup V})) \\ &= C(\mathbf{a}_U|\mathbf{a}_V) + O(\log C(\mathbf{a})) \end{aligned}$$

□

Again, for another string x, one can define a "relativized" version of the previous definition of typized sets by adding x in the conditions.

Definition 31 (Conditional typized sets)**.**

$$T(\mathbf{a}|x) = \left\{\mathbf{a}' = (a'_1, \ldots, a'_n) : \forall U, V \subseteq \{0, \ldots, n\} : C(\mathbf{a}'_U|\mathbf{a}'_V, x) \leq C(\mathbf{a}_U|\mathbf{a}_V, x)\right\} \quad (4.4)$$

$$T_c(\mathbf{a}|x) = \left\{\mathbf{a}' = (a'_1, \ldots, a'_n) : \left\|\vec{C}(\mathbf{a}|x) - \vec{C}(\mathbf{a}'|x)\right\| \leq c \log C(\mathbf{a}|x)\right\} \quad (4.5)$$

All previous claims hold for these relativized sets.

Exchanging a String in a Tuple, Keeping the same Profile

Let us show an application of typized sets. Suppose you are given a tuple of strings $\mathbf{b}$ and another string x whose complexity is close to one of the strings in the tuple, say b_1. We prove that one can find another tuple where the first string is x and whose profile is close to $\vec{C}(\mathbf{b})$.

Lemma 16. *Let $\mathbf{b} = (b_0, b_1, \ldots, b_n)$ be a tuple of strings and b'_0 be another string. Then there exist $b'_1, \ldots, b'_n$ such that tuples $\mathbf{b} = (b_0, \ldots, b_n)$ and $\mathbf{b}' = (b'_0, \ldots, b'_n)$ have complexity profiles that differ (in the corresponding positions) only by $O(\delta + \log N)$ where $\delta = |\, C(b'_0) - C(b_0)|$ and $N = C(\mathbf{b})$.*

Proof. In this argument we use the typized set of "clones" $T(\mathbf{b})$ of the tuple $\mathbf{b}$. The size of $T(\mathbf{b})$ is equal to $2^{C(\mathrm{b}) - O(\log N)}$.

Consider sections of $T(\mathbf{b})$ corresponding to some fixed values of the first coordinate. Denote by $T_0(\hat{b}_0)$ the set of all $(\hat{b}_1, \ldots, \hat{b}_n)$ such that $(\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_n) \in T(\mathbf{b})$.

By definition of $T(\mathbf{b})$, the number of all such sections is at most $2^{C(b_1, \ldots, b_n|b_0)+1}$. The average size of these sections (over all possible $\hat{b}_0$) is equal to

$$\frac{|T(\mathbf{b})|}{\text{the size of projection of } T(\mathbf{b}) \text{ onto the first coordinate}}.$$

We know that $|T(\mathbf{b})| \geq 2^{C(\mathrm{b}) - O(\log N)}$ and the size of the projection of $T(\mathbf{b})$ onto the first coordinate is at most $2^{C(b_0)+1}$. So the average size of such a section of $T(\mathbf{b})$ is at least

$$2^{C(\mathrm{b}) - C(b_0) - O(\log N)} \geq 2^{C(b_1, \ldots, b_n|b_0) - c \log N}$$

for some constant c (*i.e.*, not much less than the maximal size of a section). Thus, the maximal size of a section is $2^{C(b_1, \ldots, b_n|b_0)+1}$, and the average size is at least $2^{C(b_1, \ldots, b_n|b_0) - c \log N}$. From a simple counting it follows that for at least $2^{C(b_0) - O(\log N)}$ strings $\hat{b}_0$ the corresponding sections $T_0(\hat{b}_0)$ contain at least $2^{C(b_1, \ldots, b_n|b_0) - c \log N - 1}$ elements (at least one half of the average size of a section).

So let us fix a threshold

$$\ell = \lfloor C(b_1, \ldots, b_n|b_0) - c \log n - 1 \rfloor$$

and enumerate all possible $\hat{b}_0$ such that $\hat{T}_0(\hat{b}_0)$ contains at least 2^ℓ tuples. Notice that to run this enumeration we only need to know all complexities $C(\mathbf{b}_U|\mathbf{b}_V)$ and the number ℓ, which is $O(\log N)$ bits of information. We want to take from this enumeration the string with the ordinal number b'_0. There may be a technical obstacle at this step: the number of digits in b'_0 can be too large to represent an ordinal number of an element in this list. In this case we just cut off the "redundant" leading bits of b'_0. This string b''_0 obtained from b'_0 by cutting $O(\log N)$ of leading bits satisfies

$$C\left(b''_0|b'_0\right) \leq O(\log N) \text{ and } C\left(b'_0|b''_0\right) \leq \delta + O(\log N),$$

i.e., b'_0 and b''_0 are interchangeable up to $\delta + O(\log N)$ bits.

Now we are allowed to fix $\hat{b}_0$ as the string of index b''_0 (seen as a binary integer) in the previous enumeration. Notice that $\hat{b}_0$ and b''_0 are also interchangeable up to $\delta + O(\log N)$ bits. It remains to find $\hat{b}_1, \ldots, \hat{b}_n$ such that the complexity profile of $(\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_n)$ is close to the profile of $(b_0, b_1, \ldots, b_n)$. We have chosen $\hat{b}_0$ so that $|T_0(\hat{b}_0)| \geq 2^\ell$. Hence, there exists a tuple $(b'_1, \ldots, b'_n)$ in the section $T_0(\hat{b}_0)$ such that

$$C(\hat{b}_1, \ldots, \hat{b}_n | \hat{b_0}) \geq \ell - 1 = C(b_1, \ldots, b_n | b_0) - O(\log N).$$

By the Kolmogorov–Levin theorem we get

$$C(\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_n) = C(b_0, b_1, \ldots, b_n) - \delta + O(\log N).$$

Now we substitute the first element $\hat{b}_0$ by b'_0 in the tuple $(\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_n)$ (recall that they are interchangeable) and let $\mathbf{b'} = (b'_0, \hat{b}_1, \ldots, \hat{b}_n)$. For the latter tuple we have

$$\left\| \vec{C}(\mathbf{b}) - \vec{C}(\mathbf{b'}) \right\| \leq O(\delta + \log N)$$

and the lemma is proven. □

4.3.5 From a Conditional to an Unconditional Profile

We prove a very handy result which is related to relativization of entropy profiles. Given a conditional complexity profile $\vec{C}(Y|x)$, can we find a tuple of strings whose profile is equal to this conditional profile ? It turns out that this relativization is indeed possible in the following sense:

Lemma 17. *For a string x and an m-tuple of strings $\mathcal{Y} = (y_1, \ldots, y_m)$ there exists an m-tuple $\mathcal{Z} = (z_1, \ldots, z_m)$ such that*

$$\vec{C}(\mathcal{Z}) = \vec{C}(\mathcal{Y}|x) + O(\log N),$$

where $N = C(y_1, \ldots, y_m)$.

Proof. We are given a tuple of strings $\mathcal{Y} = (y_1, \ldots, y_m)$. For every subset of indices $W = \{i_1, \ldots, i_k\}$ we denote $\mathcal{Y}_W = (y_{i_1}, \ldots, y_{i_k})$, assuming $1 \leq i_1 < \cdots < i_k \leq m$.

Let S be the set of all tuples $\mathcal{Y}' = (y'_1, \ldots, y'_m)$ such that for all $U, V \subset [\![m]\!]$

$$C(\mathcal{Y}'_U|x) \leq C(\mathcal{Y}_U|x) \text{ and } C(\mathcal{Y}'_U|\mathcal{Y}'_V, x) \leq C(\mathcal{Y}_U|\mathcal{Y}_V, x).$$

In particular, this set contains the initial tuple $\mathcal{Y}$. Further, for each $U \subset [\![m]\!]$ we denote

$$h_U = \left\lceil \log \left[\begin{array}{l} \text{the size of the projection} \\ \text{of } S \text{ onto } U\text{-coordinates} \end{array} \right] \right\rceil$$

and

$$h^{\perp}_U = \left\lceil \log \left[\begin{array}{l} \text{the maximal size of section of } S \\ \text{for some fixed } U\text{-coordinates} \end{array} \right] \right\rceil.$$

E.g., if $U = \{1\}$, then $h_{\{1\}}$ is equal to the number of all strings y'_1 such that for some strings $y'_2, \ldots, y'_m$ the tuple $(y'_1, y'_2, \ldots, y'_m)$ belongs to S. Similarly, $h^{\perp}_{\{1\}}$ is by definition equal to

$$\max_{y'_1} |\{(y'_2, \ldots, y'_m) : (y'_1, y'_2, \ldots, y'_m) \in S\}|.$$

The cardinality of S is less than $2^{C(\mathcal{Y}|x)+1}$ since for each tuple $\mathcal{Y}'$ in S there exists a program of length at most $C(\mathcal{Y}|x)$ which translates x to $\mathcal{Y}'$. Similarly, for each U

$$h_U \le C(\mathcal{Y}_U) + 1 \text{ and } h^{\perp}_U \le C(\mathcal{Y}_{\bar{U}}|\mathcal{Y}_U) + 1,$$

where $\bar{U} = [\![m]\!] \setminus U$.

On the other hand, the cardinality of S cannot be less than $2^{C(\mathcal{Y}|x)-O(\log N)}$, since we can specify $\mathcal{Y}$ given x by the list of numbers h_U and $h^{\perp}_U$ and the ordinal number of $\mathcal{Y}$ in the standard enumeration of S.

We need only $O(\log N)$ bits to specify all numbers h_U and $h^{\perp}_U$ (the constant in this $O(\log N)$ term depends on m but not on N). Given all h_U and $h^{\perp}_U$ we can find some set S' such that the sizes of all projections and maximal sections of S' are equal to the corresponding h_U and $h^{\perp}_U$. Such sets must exist (e.g., there exists a set S, which satisfy all these conditions), so we can find one such set by brute-force search. Note that we do not need to know x to run this search.

For each tuple $\mathcal{Z} = (z_1, \ldots, z_m)$ in S' we have

$$C(\mathcal{Z}) \le h_{[\![m]\!]} = C(\mathcal{Y}|x) + O(\log N)$$

since we can specify this tuple by the list of all h_U and $h^{\perp}_U$ and the index of $\mathcal{Z}$ in the list of elements S'. Similarly, for each set of indices U

$$C(\mathcal{Z}_U) \le h_U = C(\mathcal{Y}_U|x) + O(1) \tag{4.6}$$

and

$$C(\mathcal{Z}_{\bar{U}}|\mathcal{Z}_U) \le h^{\perp}_U = C(\mathcal{Y}_{\bar{U}}|\mathcal{Y}_U, x) + O(1).$$

Let $\mathcal{Z} = (z_1, \ldots, z_m)$ be some tuple in S' with maximal possible complexity. Then $C(\mathcal{Z}) = C(\mathcal{Y}|x) + O(\log N)$ (complexity of $\mathcal{Z}$ cannot be much less since the cardinality of S' is $2^{C(\mathcal{Y}|x)-O(\log N)}$). For this tuple $\mathcal{Z}$, inequality (4.6) becomes an equality

$$C(\mathcal{Z}_U) \le h_U = C(\mathcal{Y}_U|x) + O(\log N).$$

Indeed, if $C(\mathcal{Z}_U)$ is much less than h_U, then

$$C(\mathcal{Z}) = C(\mathcal{Z}_U) + C(\mathcal{Z}_{\bar{U}}|\mathcal{Z}_U) + O(\log N)$$

is much less than

$$C(\mathcal{Y}|x) = C(\mathcal{Y}_U|x) + C(\mathcal{Y}_{\bar{U}}|\mathcal{Y}_U, x) + O(\log N),$$

and we get a contradiction with the choice of $\mathcal{Z}$. Hence, the difference between the corresponding components of complexity profiles $\vec{C}(\mathcal{Z})$ and $\vec{C}(\mathcal{Y}|x)$ is bounded by $O(\log N)$. □

Chapter 5

Quasi-perfect Secret Sharing

Contents

Introduction

In this chapter we relax the original model of perfect secret sharing by including information leaks and incomplete information. If both of these quantities can be made negligible, we have what we call a "quasi-perfect" secret-sharing scheme. We introduce hereafter several definitions formalizing this idea, and study their properties.

We show that missing information is an easier problem and we can get rid of it, possibly at the expense of a bigger information leak. We also show that the distribution on secrets is not important and any quasi-perfect scheme can be converted into another one with the uniform distribution on secrets. This result is a counterpart of a property for perfect schemes.

We establish a relation between the introduced notions and algorithmic information theory. In the algorithmic model, the secret to be shared and the shares are binary strings, and perfect secret-sharing conditions are formulated in terms of conditional complexity and algorithmic information. We use a typization argument to show that for a given access structure and a given information ratio, the existence of a quasi-perfect secret-sharing scheme in the probabilistic model is equivalent to its existence in the algorithmic model. Using similar arguments, we prove in the algorithmic version that it does not matter which secret string we have to share.

The major unsolved problem in secret sharing is to invent new efficient secret-sharing schemes in the cases where they do exist, and prove their nonexistence in the cases where they do not — thus closing the gap between the upper and lower bounds. One can hope that it is easier to construct a quasi-perfect scheme than a perfect one with the same efficiency. It would be also interesting to find a lower bound for perfect schemes that does not apply to quasi-perfect schemes (the known bounds do not have this property). Both problems are still left open for further research; nevertheless, we hope that the definition of an intermediate class of quasi-perfect secret-sharing schemes and investigating its basic properties and relations to algorithmic information theory could be helpful.

We also consider the reduction of the secret size. For perfect secret sharing one can always reduce the size of the secret (restricting the set of secrets) and still have a perfect scheme (though less efficient). If we allow some information leak, the situation becomes

more complicated. Imagine that we have a scheme that distributes, say, 1000-bit secrets with only 10-bit information leak. We cannot use this scheme for 1-bit secrets directly, since the information leak can now exceed the secret size – although the leak was initially small compared to the secret size. We overcome this difficulty and suggest a technique that allows us to reduce both the secret size and the leak at the same time.

5.1 À la Shannon

5.1.1 Definition of Quasi-perfect Secret-sharing Schemes

We begin by formalizing the idea in the probabilistic setting using the tools of Information Theory. First we define the parameters at stake for, non necessarily perfect, secret-sharing schemes.

Definition 32. *Let Γ be an access structure. Consider a* secret-sharing scheme *in the sense of the probabilistic definition: Let s and σ_p, for all participants p, be some random variables defined on the same probability space. The main parameters of these scheme are:*

- entropy $H(s)$ of the secret *(always positive);*
- information ratio: *the maximal entropy of a single share divided by $H(s)$;*
- missing information ratio: *the maximal value of $H(s|\sigma_A)$ over all authorized groups A, divided by $H(s)$;*
- information leak ratio: *the maximal value of $I(s{:}\sigma_B)$ over all forbidden groups B, divided by $H(s)$.*

If the missing information and the information leak are zeros, we obtain a perfect secret-sharing scheme. We interest ourselves in *quasi-perfect* secret-sharing schemes, where the missing information and information leak ratios are asymptotically small.

Definition 33. *An access structure Γ can be* quasi-perfectly implemented with information ratio ρ *if there exists a sequence of secret-sharing schemes such that*

(1) the lim sup *of the information ratio does not exceed ρ;*

(2) the missing information ratio tends to zero;

(3) the information leak ratio tends to zero.

A more general definition could require that the lim sup of the missing information ratio is bounded by some ε, and lim sup of the information leak is bounded by some δ. For these approximate schemes, most of our results below can be extended in a natural way. Notice further that each time we talk about a quasi-perfect scheme, we are in fact handling an infinite sequence of individual secret-sharing schemes. This will permit us to study the asymptotic behavior of secret-sharing schemes.

5.1.2 Secret-Sharing Scheme Without Missing Information

Definition 34 (Scheme without missing information)**.** *We say that an access structure can be quasi-perfectly implemented* without missing information *if it has a quasi-perfect implementation where the missing information (ratio) of each individual scheme is exactly zero.*

Our first observation: we can require the missing information ratio to be exactly zero and still get an equivalent definition.

Theorem 33. *Let Γ be an access structure with n participants. If there exists a secret-sharing scheme for Γ with information ratio ρ, missing information ratio ε and information leak ratio δ, then there exists another scheme with no missing information that has information leak ratio at most $\delta + e$ and information ratio at most $\rho + e$, where $e = \left(2\varepsilon + \frac{3}{H(s)}\right) 2^n$.*

To prove Theorem 33, we proceed by adding the missing information to authorized groups. For that we need to "materialize" this missing information and provide it to each participant, still keeping the information leak small. This plan is achieved using two simple observations from Shannon's Information Theory. The first one:

Lemma 18. *Let α and β be two jointly distributed random variables. There exists a variable γ defined on the same probabilistic space such that $H(\alpha|\beta,\gamma) = 0$ and $H(\gamma) \leq 2H(\alpha|\beta) + 3$.*

Proof. Let β be distributed on a set $\{b_1, \ldots, b_s\}$. For each value b_j of β, we have a conditional distribution on values of α given the condition $\beta = b_j$. We can construct for this conditional distribution a prefix-free binary code $c_{1j}, \ldots, c_{mj}$ with average length at most $H(\alpha|\beta = b_j) + 1$ (e.g., we can take the Huffman code).

Let γ be the corresponding codeword as a random variable defined on the same space where α and β are defined: if $\beta = b_j$ and $\alpha = a_i$ then $\gamma = c_{ij}$ (the i-th codeword from the code constructed for α-distribution with condition $\beta = b_j$).

Given a value b_j of β and the string c_{ij} from j-th code, we can reconstruct the corresponding value of α. Hence, $H(\alpha|\beta,\gamma) = 0$. It remains to estimate the entropy of γ.

This γ ranges over set of all codewords c_{ij} (from all codes constructed for all possible values of β). The average length of bit strings c_{ij} can be easily bounded: for every value b_j of β it is bounded by $H(\alpha|\beta = b_j) + 1$, and taking average over different b_j, we get $H(\alpha|\beta) + 1$.

However, in this way we do *not* get a bound for the entropy of γ, since the set of all c_{ij} is not prefix-free (or uniquely decodable), even if it is prefix-free for any fixed j. But we can replace each string c_{ij} by its encoding $\widehat{c_{ij}}$, where $s \mapsto \hat{s}$ is some prefix-free encoding of all strings. Let us agree that $\hat{s}$ is s with every bit doubled and 01 added; then $|\hat{s}| \leq 2|s| + 2$, and this linear bound remains valid after averaging. So we get a prefix-free code whose average length does not exceed $2H(\alpha|\beta) + 3$, and the entropy of γ does not exceed this average length. □

The second one is a classical (Shannon-type) information inequality:

$$I(\alpha{:}\beta\gamma) \leq I(\alpha{:}\beta) + H(\gamma).$$

It says, for all α, β, γ, that if, in addition to some known β, a participant gets some new information γ, her information about some α increases at most by $H(\gamma)$.

Using these two observations, it is easy to prove Theorem 33. Indeed, let us consider all authorized groups sequentially. For each group we use Lemma 18 to "materialize" the missing information into some random variable (on the same space) and then include this variable into all participants' shares. Then the problem of missing information for this group is solved at the cost of an increased information leak (the increase is bounded because of our second observation) and increased entropy of the shares. Since we need to repeat this at most 2^n times, we get the required bound.

The previous theorem implies the following result in terms of quasi-perfect schemes. It shows that as far as quasi-perfect schemes are concerned, the missing information ratio is unimportant.

Corollary 5. *If an access structure Γ can be quasi-perfectly implemented, then it has a quasi-perfect implementation without missing information for the same information ratio.*

5.1.3 Secrets Drawn According to the Uniform Distribution

The next result shows how the distribution of the secret variable in any quasi-perfect schemes can be made uniform.

Theorem 34. *If some access structure Γ can be quasi-perfectly implemented with information ratio ρ, it can be quasi-perfectly implemented with the same ratio by a sequence of schemes that have uniformly distributed secrets.*

We need to prove that for any given quasi-perfect scheme, we can create another quasi-perfect scheme with the same parameters wherein every individual sharing scheme has a uniform distribution on secrets. To this aim we use a result of T. H. Chan and R. W. Yeung (Lemma 4.1 in [CY02]) on quasi-uniform random variables. This tool is described by Lemma 12 in Section 4.1.1 on page 66 in our Toolbox Chapter.

Since our definition does not restrict strongly the entropy or the size of the shares, as long as the limit ratios are the same, we can use this lemma to construct another secret-sharing scheme and make the tuple quasi-uniform (and therefore to make the secret uniformly distributed). Theorem 34 is proven.

Remark 8. *Notice that we can apply Theorems 34 then 33 to obtain a quasi-perfect implementation with the same parameters where all the schemes have uniform distributions on secrets and missing information is (exactly) zero at the same time.*

5.1.4 Downscaling the Size of the Secret

We prove a result for individual secret-sharing schemes. We consider reducing the size of the secret, and particularly the most drastic reduction going from M possible secrets to only 2 (a 1-bit secret). As we have seen we can assume that the secret is uniformly distributed and that the scheme has no missing information. This time context is more combinatorial, as opposed to the asymptotical study mentioned earlier, for we are interested in a fixed secret size.

Theorem 35. *Suppose there is a secret-sharing scheme for some access structure Γ with: n participants,* uniformly distributed *N-bit secrets, no missing information, shares of entropy at most S and information leak ratio δ.*

If N is large enough compared to n and S, there exists another secret-sharing scheme for the same access structure Γ, the same shares, and 1*-bit secret without missing information and $8\delta^{2/3}$ information leak ratio.*

The exact meaning of "large enough": N should be greater than $5(\log S + \log n)$, and also greater than some fixed constant. So it concerns the vast majority of schemes.

This theorem is about decreasing the secret size. Note that increasing is easy: as for perfect schemes, we may consider N independent copies of a scheme. Then secret and shares contain N times more information, while all ratio parameters remain the same. In the perfect setting, reducing the secret size is also trivial, we can simply reduce the size of the secret support. However, in our more general model, we are now also concerned with leaks, and keeping them small. Using the probabilistic method, we show that there indeed exists such a bit to be shared, that does not leak too much information.

Sketch of the proof of Theorem 35: Construct a new scheme for a 1-bit secret from the initial scheme in the following way. Given a scheme without missing information for Γ sharing a secret of N-bits (uniformly distributed on the set $\mathcal{K} = \{1, \dots, 2^N\}$), take a splitting of $\mathcal{K}$ into two equal parts, say $\mathcal{K}_0$ and $\mathcal{K}_1$. Then define a new scheme as follows: to share the bit i, take a random element of $\mathcal{K}_i$ and share it with the initial scheme. It is easy to see that this new scheme is indeed a scheme without missing information for Γ sharing one uniform bit with some information leak ratio δ' depending on the initial choice of the splitting $\mathcal{K}_0$. We will show that there exists such a splitting which provides a small leak.

We first prove a general lemma about discrete random variables.

Lemma 19. *Let X be a finite discrete random variable on a k-element set A (with k even) such that $H(X) \geq \log_2 k - \delta$ for some positive δ. Let B be a random subset of A of size $k/2$ (B is chosen uniformly, i.e., each $(k/2)$-element subset of A is chosen with probability $1/\binom{k}{k/2}$). Then for every $\gamma \in (0, 1)$, with probability at least*

$$1 - 2e^{-\frac{4\tau^2}{k\gamma^2}}$$

(probability for a random choice of B) we have

$$\left|\Pr[X \in B] - \tfrac{1}{2}\right| \leq 2\tau$$

(probability for the initial distribution X), where $\tau = \frac{1+\delta}{2\log_2 \gamma k}$.

Proof. For each element $x \in A$, denote by ρ_x the non-negative weight (probability) that X assigns to x. Using this notation we have

$$H(X) = \sum_{x \in A} -\rho_x \log_2 \rho_x.$$

A randomly chosen B contains exactly one half of the points x from A. We need to estimate the sum of ρ_x for all $x \in B$. We do it separately for "rather large" ρ_x and for "rather small" ρ_x. To make this idea more precise, fix a threshold $\gamma > 0$ that separates "rather large" and "rather small" values of ρ_x. Denote by p_γ the total measure of all ρ_x that are greater than this threshold. More formally,

$$p_\gamma = \sum_{\rho_x > \gamma} \rho_x.$$

We claim that p_γ is rather small. Indeed, if we need to identify some $x \in A$, we should specify the following information which consists of two parts:

1. We say whether $p_x > \gamma$ or not (one bit of information).

2a. If $p_x > \gamma$, we specify the ordinal number of this "large" point; there are at most $1/\gamma$ points x' such that $\rho_{x'} > \gamma$, so we need at most $\log_2(1/\gamma)$ bits of information;

2b. otherwise, $p_x \leq \gamma$, we simply specify the ordinal number of x in A; here we need at most $\log_2 k$ bits of information.

From the standard coding argument we get

$$H(X) \leq 1 + p_\gamma \log_2(1/\gamma) + (1 - p_\gamma) \log_2 k.$$

Since $H(X) \geq \log_2 k - \delta$, it follows that $p_\gamma \leq \frac{1+\delta}{\log_2(\gamma k)}$.

Thus, we may assume that the total measure of "rather large" values ρ_x is quite small even in the entire set A; hence, "large" points do not contribute much to the measure of a randomly chosen B. It remains to estimate the typical impact of "small" ρ_x on the weight of B.

Technically, it is useful to forget about "large" points x (substitute weights $\rho_x > \gamma$ by 0) and denote

$$\rho'_x = \begin{cases} \rho_x & \text{if } \rho_x \le \gamma. \\ 0 & \text{otherwise.} \end{cases}$$

Now we choose exactly $k/2$ different elements from A and estimate the sum of the corresponding ρ'_x. Note that expectation of this sum is one half of the sum of ρ'_x for all $x \in A$, i.e., $(1 - p_\gamma)/2$. It remains to estimate the deviation of this sum from its expectation. We use the version of Höffding's bound for samplings without replacement, which can be used to estimate deviations for a sampling of $k/2$ points from a k-elements set, (see [Hoe63, Section 6]). The probability of the event that the sum exceeds expected value plus some τ can be bounded as follows:

$$\Pr\left[\sum_{x\in B} \rho'_x \ge (1 - p_\gamma)/2 + \tau\right] \le e^{-\frac{2\tau^2}{|B|\gamma^2}} = e^{-\frac{4\tau^2}{k\gamma^2}}.$$

Together with "large" values ρ_x we have

$$\Pr\left[\sum_{x\in B} \rho_x \ge (1 - p_\gamma)/2 + \tau + p_\gamma\right] \le e^{-\frac{4\tau^2}{k\gamma^2}}$$

Now we fix the parameter τ to be equal to one half of the upper bound for p_γ, i.e.,

$$\tau = \frac{1 + \delta}{2\log_2(\gamma k)}.$$

It follows that,

$$\Pr\left[\sum_{x\in B} \rho_x \ge 1/2 + 2\tau\right] \le e^{-\frac{4\tau^2}{k\gamma^2}}.$$

From this bound, we can deduce the symmetric bound for the sum of ρ_x in $A\setminus B$:

$$\Pr\left[\sum_{x\in A\setminus B} \rho_x \le 1/2 - 2\tau\right] \le e^{-\frac{4\tau^2}{k\gamma^2}}.$$

Since $A\setminus B$ and B share the same distribution (the uniform one), this bound also holds for B. Sum up the two bounds and we are done. □

We are now ready to prove Theorem 35.

Proof of Theorem 35. Let $\mathcal{K}_0$ be a random subset of the set of all secrets $\mathcal{K}$ such that $|\mathcal{K}_0| = 2^{N-1}$. $\mathcal{K}_0$ is chosen uniformly over all possible such fair splittings of $\mathcal{K}$. If s is the random variable for the N-bit secret in the initial scheme, let us define the new secret bit ξ as the bit defined by "$s \in \mathcal{K}_0$" (ξ is indeed a bit since $H(\xi) = 1$). Our goal is to estimate $H(\xi|\sigma_B)$ for each forbidden group $B \notin \Gamma$, and show it is large. Formally, we want to show that $H(\xi|\sigma_B) \ge 1 - \varepsilon'$ where $\varepsilon' = 8\varepsilon^{\frac{2}{3}}$.

First, we notice that for each bit ξ constructed as above, $I(\xi{:}\sigma_B) \le \varepsilon N$ holds for all

$B \notin \Gamma$, so we can assume that $\varepsilon' \leq \varepsilon$, i.e.,

$$\varepsilon' \geq \frac{8^3}{N^2}. \tag{5.1}$$

We know that $H(s|\sigma_B)$ is rather large. More precisely,

$$H(s|\sigma_B) \geq N(1-\varepsilon).$$

We introduce some positive parameter δ (to be fixed later) to separate all values of σ_B into two classes:

$$\text{more typical values } b \text{ such that } H(s|\sigma_B = b) \geq N(1-\delta).$$

and

$$\text{less typical values } b \text{ such that } H(s|\sigma_B = b) < N(1-\delta).$$

Since the entropy $H(s|\sigma_B)$ is large, the total measure of all "less typical" values b is rather small (more precisely, it is not greater than $\frac{\varepsilon}{\delta}$). We do not care about the conditional entropy of ξ when b is non-typical (the total weight of these b is so small that they do not contribute essentially to $H(\xi|\sigma_B)$). We focus on the contribution of $H(\xi|\sigma_B = b)$ for a typical value b. To estimate this quantity we apply Lemma 19 to the distribution k conditional to $\sigma_B = b$, it follows that

$$H(\xi|\sigma_B = b) \geq h(1/2 + 2\tau) \geq 1 - 16\tau^2$$

with probability

$$1 - 2e^{-\frac{4\tau^2}{\gamma^2}2^{-N}}$$

for some new parameter $\gamma > 0$ and

$$\tau = \frac{1+\delta N}{2(\log_2 \gamma + N)}.$$

This inequality is true for all forbidden groups B and typical shares b. Thus if we sum up the bad events, we obtain that the following estimation for $H(\xi|\sigma_B)$:

$$\begin{aligned} H(\xi|\sigma_B) &= \sum_{b \in \mathcal{S}_B} \Pr[\sigma_B = b] H(\xi|\sigma_B = b) \\ &\geq \sum_{\text{typical } b} \Pr[\sigma_B = b] H(\xi|\sigma_B = b) \\ &\geq \left(1 - \frac{\varepsilon}{\delta}\right)(1 - 16\tau^2) \\ &\geq 1 - \frac{\varepsilon}{\delta} - 16\tau^2 \end{aligned}$$

holds with probability at least

$$1 - \left|\overline{\Gamma}\right| |\mathcal{S}_{\mathcal{P}}| 2e^{-\frac{4\tau^2}{\gamma^2}2^{-N}} \tag{5.2}$$

where $\mathcal{S}_{\mathcal{P}}$ is the set of all possible shares given to the group of all participants.

Now, we choose our parameters γ and δ to deduce our result and show that our choice is valid. We take

$$16\tau^2 = \frac{\varepsilon}{\delta} = \frac{1}{2}\varepsilon' = 4\varepsilon^{\frac{2}{3}}.$$

Under these conditions

$$\log_2 \gamma = -N\left[1 - \frac{1}{8}\left(\frac{\varepsilon'}{\varepsilon N} + 2\right)\right]$$

and

$$H(\xi|\sigma_B) \geq 1 - 8\varepsilon^{\frac{2}{3}} = 1 - \varepsilon'.$$

We want to find a simple sufficient condition that guarantees that the probability (5.2) is non-negative. To this end we do some (rather boring) calculations. We take the required inequality and reduce it step by step to a weaker but more suitable form.

$\left\|\overline{\Gamma}\right\| \cdot \|S_{\mathcal{P}}\| 2e^{-\frac{4\tau^2}{\gamma^2}2^{-N}}$	$<$	1	that is what we need, see (5.2)
$\left\|\overline{\Gamma}\right\| \cdot \|S_{\mathcal{P}}\|$	$<$	$2e^{\frac{4\tau^2}{\gamma^2}2^{-N}}$	
$2^n \cdot 2^{nS}$	$<$	$2e^{\frac{4\tau^2}{\gamma^2}2^{-N}}$	trivial upper bounds for $\left\|\overline{\Gamma}\right\|$ and $\|S_{\mathcal{P}}\|$
$2^{n(S+1)}$	$<$	$2^{\frac{4\tau^2}{\gamma^2}2^{-N}}$	since $e > 2$
$n(S+1)$	$<$	$\frac{4\tau^2}{\gamma^2}2^{-N}$	by applying $\log_2$
$2nS$	$<$	$\frac{4\tau^2}{\gamma^2}2^{-N}$	since $S \geq 1$
$2nS$	$<$	$\frac{\varepsilon'}{8}2^{N\left(1-\frac{1}{4}\left(\frac{\varepsilon'}{\varepsilon N}+2\right)\right)}$	from (5.1.4) and (5.1.4)
nS	$<$	$\varepsilon' 2^{\frac{1}{4}N-4}$	since $\varepsilon' \leq \varepsilon N$
2^{cN}	$<$	$\varepsilon' 2^{\frac{1}{4}N-4}$	by assumption
1	$<$	$\varepsilon' 2^{(\frac{1}{4}-c)N-4}$	
0	$<$	$\left(\frac{1}{4}-c\right)N + \log_2 \varepsilon' - 4$	
0	$<$	$\left(\frac{1}{4}-c\right)N - 2\log_2 N + 5$	from (5.1)

The last inequality (which is a sufficient condition for (5.2) to be non-negative) holds when $c < \frac{1}{4}$ and $N > N_0$ for some large enough N_0 depending on c. The statement of the theorem uses $c = 1/5$. □

Sharing exactly one bit in an efficient way seems more difficult than sharing N bits. We do not know whether this bound can be improved. In particular, can we achieve a leak of $O(\varepsilon)$?

5.1.5 Lower Bounds on the Information Ratio of Quasi-Perfect Schemes

Recall the access structure P_4 with participants a, b, c, d and authorized sets $\{a, b\}$, $\{b, c\}$, $\{c, d\}$ and all their supersets. In Chapter 3, we proved a bound on its information ratio for perfect schemes (see Section 3.4 page 47). We now generalize this bound for quasi-perfect schemes.

Proposition 25. *This access structure can be quasi-perfectly implemented with information ratio ρ only if $\rho \geq 3/2$.*

First, we generalize Proposition 10 from page 41 to general secret-sharing schemes where the missing information ratio is ε and information leak ratio is δ.

Proposition 26.

- *if $A \notin \Gamma$ and $AB \in \Gamma$, then*

$$H(A|B) \geq H(A|Bs) + (1 - \varepsilon - \delta) \cdot H(s)$$

- *if* $C \notin \Gamma$ *and* $AC, BC \in \Gamma$ *then*

$$I(A{:}B|C) \geq (1 - 2\varepsilon - \delta) \cdot H(s)$$

- *if* $ABC \in \Gamma$ *and* $AC, BC \notin \Gamma$ *then*

$$I(A{:}B|Cs) \geq (1 - \varepsilon - 2\delta) \cdot H(s)$$

Proof of Theorem 25. The following information inequalities holds:

$$\begin{aligned} H(bc) &\geq I(a{:}c|b) + I(b{:}d|ac) + H(bc|ad) \\ I(a{:}c|b) &\geq (1 - 2\varepsilon - \delta) \cdot H(s) \\ I(b{:}d|ac) &\geq (1 - 2\varepsilon - \delta) \cdot H(s) \\ H(bc|ad) &\geq (1 - \varepsilon - \delta) \cdot H(s) \end{aligned}$$

By summing all of the above we obtain

$$H(bc) \geq (3 - 5\varepsilon - 3\delta) \cdot H(s).$$

Since ε and δ vanish in any quasi-perfect implementation, the desired inequality

$$H(bc) \geq 3H(s)$$

holds, which implies $\rho \geq 3/2$. □

We advertise briefly the topic of the next chapter. One should notice that the inequalities used for perfect schemes are in fact conditional information inequalities, for we apply the inequality to distributions where the perfect secret sharing requirements hold. More particularly, most proofs (known to the author) make use of conditional inequalities which are not "essentially conditional", in the sense of [KR11] as presented in the next chapter. Indeed, such a conditional inequality is a plain consequence of an unconditional one, where the conditions are added to the inequality with some coefficient, quite in the style of Lagrange's multipliers.

Therefore any lower bound for the information ratio of a perfect scheme obtained via a non essentially conditional inequalities, is still a valid lower bound for the information ratio of quasi-perfect schemes for the same access structure.

Theorem 36. *For a fixed access structure, any bound on the complexity of perfect secret-schemes obtained by simply applying an unconditional inequality also holds for quasi-perfect secret-sharing schemes.*

This theorem proves the inability of the current method to obtain better lower bounds for quasi-perfect schemes. A lower bound relying on the use of an essentially conditional inequality could potentially be useful to separate the power of perfect vs. quasi-perfect schemes.

A Weak-Separation Between Perfect and Quasi-perfect Schemes

In spite of the previous negative result, one can still prove that quasi-perfect schemes can sometimes be slightly better than perfect schemes.

Theorem 37. *There exists an access structure that can be quasi-perfectly implemented with information ratio* 1 *but has no ideal perfect scheme.*

The catch here (why we say that this is a weak result) is that this access structure also has perfect secret-sharing schemes with information ratio arbitrarily close to 1.

As we have said in Chapter 3, an access structure Γ is induced by a matroid, defined by its circuits, $M = (\mathcal{Q}, \mathcal{C})$ through $s \in \mathcal{Q}$ if Γ is defined on the set of participants $\mathcal{P} = \mathcal{Q} \setminus \{s\}$ by the upper closure of the collection of subsets $A \subseteq \mathcal{P}$ such that $A \cup \{s\} \in \mathcal{C}$ (here $\mathcal{C}$ is the set of circuits of the matroid $\mathcal{M}$.) Let $\mathcal{F}$ and $\mathcal{F}^-$ be respectively the access structures induced by the Fano and by the non-Fano matroids (through any point). A result of Matúš in [Mat99b] implies that there exist perfect ideal schemes for $\mathcal{F}$, resp. $\mathcal{F}^-$ if and only if $|\mathcal{K}|$ is even, resp. odd.

Consider an access structure Γ consisting of disjoint copies of $\mathcal{F}$ and $\mathcal{F}^-$. From Matúš' argument it follows that Γ cannot be implemented by a perfect scheme with information ratio 1. On the other hand, we show that it can be implemented by a quasi-perfect scheme with limit information ratio 1.

Proof of Theorem 37. First, we construct a scheme Σ consisting of the concatenation of two independent subschemes:

- an ideal perfect scheme for $\mathcal{F}$ for 2^N secrets, and
- a perfect scheme for $\mathcal{F}^-$ for 2^N secrets using $2^N + 1$ shares, a scaled-down version of an ideal perfect scheme where $|\mathcal{K}| = 2^N + 1$

This Σ is a perfect scheme for Γ with information ratio

$$\rho = \frac{\log(2^N + 1)}{N} = 1 + O\left(\frac{1}{N \cdot 2^N}\right).$$

Now we modify the scheme slightly by changing the second subscheme. We substitute the value of the $2^N + 1$-th share with any other possible share value. In the resulting scheme Σ' there are exactly 2^N different shares, but it is not perfect.

Any forbidden set still knows nothing about the secret since it has less information (each new share is a function of the old one). Authorized sets may have lost some information about the secret but only up to $O\left(2^{-N}\right)$. Hence, the new scheme is without information leak, its missing information ratio vanishes, and its information ratio is exactly one. □

5.1.6 A Property of Optimal Quasi-perfect Schemes

Hereafter we pursue the study of leaves, as defined in Section 3.4.6 on page 52. We add a small stone to the resolution of Question 1 by proving the following property of optimal schemes. Recall that in our access structure, a is a leaf participant as witnessed by the only minimal authorized group $\{a, b\}$. The proof of the following result makes use of the Slepian–Wolf coding present in the Toolbox Chapter 4.

Theorem 38. *In an optimal quasi-perfect scheme,*

$$I(a{:}b) \to 0 \text{ as } H(s) \to \infty.$$

Proof of Theorem 38. Assume, for the sake of contradiction, that an optimal quasi-perfect scheme for Γ does not satisfy $I(a{:}b) \to 0$ as $H(s) \to \infty$. Let us construct another scheme with greater or equal average information rate for which $I(a{:}b) = 0$ holds in the limit.

Let $\Sigma = (S, A, B, C_1, \ldots, C_k)$ be an N-serialization of secret-sharing scheme for Γ with parameters an ε and δ. Take the Slepian–Wolf hash $A' = SW(A|B)$ and consider the new

scheme

$$\Sigma' = (S, A', B, C_1, \ldots, C_k).$$

Let us verify that Σ' is a secret-sharing scheme for Γ with the same parameters. Since we only changed the share of the leaf a, we only need to check the secret sharing requirements for authorized and forbidden groups involving a.

(i) $H(S|A') \geq H(S|A) = H(S)$ and

(ii) $H(S|A', C) \geq H(S|A, C)$, for any $C \in P - \{a, b\}$, due to the functional dependency between A' and A.

(iii) $H(S|A', B) = H(S|A, B) + o(N)$ since from A' and B we can almost reconstruct A.

Since A' is a function of A, the average information rate of Σ' cannot be less than in the original scheme Σ. Also, A' and B are almost independent:

$$\begin{aligned} H(A'|B) &= H(A'B) - H(B) = H(AB) + o(N) - H(B) \\ &= H(A|B) + o(N) = H(A') + o(N) \end{aligned}$$

Hence, Σ' is also a secret-sharing scheme for Γ with parameters an ε and δ and whose average information rate is less or equal than the one of the original scheme, and by construction A' and B are almost independent:

$$I(A' : B) = o(N). \qquad \square$$

5.2 À la Kolmogorov

Now let us introduce corresponding notions in the framework of Algorithmic Information Theory (or Kolmogorov Complexity Theory). The advantage of this approach is that we can speak about sharing individual secrets using individual shares, which is not possible in the Shannon framework. On the other hand, our statements become inherently asymptotic, since Kolmogorov complexity is defined only up to a $O(1)$ additive term; another drawback is that Kolmogorov complexity is not computable.

5.2.1 Algorithmic Secret Sharing

We propose the following general definition, a counterpart of Definition 33, for Algorithmic secret-sharing schemes[1].

Definition 35. *An access structure Γ can be* algorithmically implemented with information ratio ρ *if there exists a sequence of secret-sharing schemes (in the algorithmic setting) with secrets s_n such that*

(0) the complexity of s_n tends to infinity;

(1) the lim sup *of the information ratio does not exceed ρ;*

(2) the missing information ratio tends to 0 as $n \to \infty$;

(3) the information leak ratio tends to 0 as $n \to \infty$.

[1] Another version of secret sharing in the Kolmogorov Complexity framework has previously been introduced in [ALPS09]

In this setting the secrets (s_n) and the shares are strings, and the information ratios are defined using Kolmogorov complexity in a natural way. The only new item is requirement (0). It is needed because Kolmogorov complexity is defined only up to constant additive term, so we should consider strings of growing complexity to get an invariant definition.

Note that in the algorithmic setting it is not possible to define an exact counterpart of the notion of *perfect* secret sharing, since the missing information and the information leak are defined only up to a $O(1)$ additive term. Algorithmic secret sharing is thus *intrinsically* "quasi-perfect", and therefore makes a natural counterpart of the probabilistic definition of quasi-perfect secret sharing. Indeed, these two notions turn out to be equivalent.

5.2.2 Equivalence with the Probabilistic Definition

Theorem 39. *Let Γ be an access structure. It can be quasi-perfectly implemented with information ratio ρ (Definition 33) if and only if it can be algorithmically implemented with the same information ratio ρ (Definition 35).*

By Theorem 33, these two notions are equivalent to a third one, namely the notion of quasi-perfect implementation without missing information (Definition 34).

Corollary 6. *The following are equivalent:*

- *Γ can be quasi-perfectly implemented with information ratio ρ.*
- *Γ can be quasi-perfectly implemented with information ratio ρ without missing information.*
- *Γ can be algorithmically implemented with the same information ratio ρ.*

To prove Theorem 39, we convert a sequence of n-tuples of random variables into a sequence of n-tuples of binary strings and vice versa; these conversions will try to preserve complexity/entropy profiles: corresponding tuples of random variables and strings will have similar values in their profiles. The main technical tools are the Kolmogorov–Levin theorem (see Section 4.3.1 p. 73) and the "typicalization" trick for comparing Shannon entropy and Kolmogorov complexity (technique suggested by A. Romashchenko and used in [Rom00b, HRSV00]).

Proof of Theorem 39. In each direction, we construct a new tuple whose profile is close, possibly with a scaling factor, to the profile of the original tuple up to an additive logarithmic term.

[Kolmogorov $\to$ Shannon] Let $\mathbf{a} = (a_1, \ldots, a_n)$ be an n-tuple of binary strings. For a non-negative integer c we consider the typized set: $T_c(\mathbf{a})$ *i.e.*, the set of n-tuples of binary strings whose complexity profile are equal to the one of $\mathbf{x}$ up to a logarithmic term.

Now let $\mathbf{s} = (s_1, \ldots, s_n)$ be a random n-tuple uniformly distributed on $T_c(\mathbf{a})$. Claim 3 guarantees that entropy of $\mathbf{s}$ is close to $C(\mathbf{a})$. Moreover, all components of the entropy profile of $\mathbf{s}$ are close to the corresponding components of the complexity profile of $\mathbf{a}$. We prove this in two steps. First, we obtain an upper bound:

Claim 5. *$H(\mathbf{s}_U) \leq C(\mathbf{a}_U) + 1$ for every $U \subseteq \{1, \ldots, n\}$.*

Proof. The number of possible values for $\mathbf{s}_U$ is the number of different $\mathbf{a}'_U$ for $\mathbf{a}' \in T_c(\mathbf{a})$. By definition, $C(\mathbf{a}'_U) \leq C(\mathbf{a}_U)$, and there is at most $2^{C(\mathbf{a}_U)+1} - 1$ such values for $\mathbf{s}_U$. It remains to note that the entropy of a random variable is bounded by the logarithm of the number of its values. □

It remains to prove the lower bound:

Claim 6. $H(\mathbf{s}_U) \geq C(\mathbf{a}_U) - O(\log C(\mathbf{a}))$ *for every* $U \subseteq \{1, \ldots, n\}$

Proof. Let V be the complement of U. Claim 4 says that

$$C(\mathbf{a}'_V|\mathbf{a}'_U) \leq C(\mathbf{a}_V|\mathbf{a}_U) + O(\log C(\mathbf{a})).$$

So for every value of $\mathbf{a}'_U$ there are at most

$$2^{C(\mathrm{a}_V|\mathrm{a}_U)+O(\log C(\mathrm{a}))}$$

values of $\mathbf{a}'_V$. Therefore for every value of $\mathbf{s}_U$ the conditional distribution of $\mathbf{s}_V$ is concentrated on at most $2^{C(\mathrm{a}_V|\mathrm{a}_U)+O(\log C(\mathrm{a}))}$ elements, and

$$H(\mathbf{s}_V|\mathbf{s}_U) \leq C(\mathbf{a}_V|\mathbf{a}_U) + O(\log C(\mathbf{a})).$$

The upper bound for conditional entropy gives the lower bound for $H(\mathbf{s}_U)$, because

$$H(\mathbf{s}_U) = H(\mathbf{s}) - H(\mathbf{s}_V|\mathbf{s}_U).$$ □

We have shown how an algorithmic secret-sharing scheme can be converted to a probabilistic one such that their profiles are equal up to a logarithmic term. It remains to note that complexity of $C(\mathbf{a})$ increases linearly with the complexity of the secret, for the information ratio is uniformly bounded. Thence logarithmic terms are negligible in the limit and both schemes have the same parameters.

[Shannon → Kolmogorov] Let $\mathbf{s} = (s_1, \ldots, s_n)$ be an n-tuple of random variables. After a suitable approximation we may assume without loss of generality that each value of $\mathbf{s}$ has a rational probability. We fix an integer $M > 0$ (to be specified later) such that all probabilities are rational numbers whose denominators divide M, and construct some $M \times n$ table

$$\mathbf{a} = \begin{bmatrix} a_1^1 & a_2^2 & \ldots & a_1^M \\ a_2^1 & a_2^2 & \ldots & a_2^M \\ \vdots & \vdots & \ddots & \vdots \\ a_n^1 & a_n^2 & \ldots & a_n^M \end{bmatrix}$$

where

(a) The columns of the table (each column is an n-vector) are values of $\mathbf{s}$, and the frequencies of columns correspond to their $\mathbf{s}$-probability. (So, choosing a column at random, we get a random variable distributed as $\mathbf{s}$.)

(b) The table has maximal Kolmogorov complexity among all tables satisfying (a) for fixed M.

It implies, by a rather simple counting argument, that

$$C(\mathbf{a}) \geq M \cdot H(\mathbf{s}) - O(\log M)$$

(Here and below the constant hidden in $O(\log M)$ may depend on the distribution $\mathbf{s}$).

Now we use the n rows of this table as $a_1, \ldots, a_n$. Let us verify that the n-tuple $\mathbf{a} = (a_1, \ldots, a_n)$ has a complexity profile close to the entropy profile of s, multiplied by M.

Claim 7. *$C(\mathbf{a}_U) \le M \cdot H(\mathbf{s}_U) + O(\log M)$ for every $U \subseteq \{1, \dots, n\}$.*

Proof. We extract from the entire table the rows corresponding to U and get a smaller table (of the same width) whose rows form $\mathbf{a}_U$. The frequencies for different columns in this table correspond to the distribution of $\mathbf{s}_U$. By Theorem 5.1 in [ZL70],

$$C(\mathbf{a}_U) \le M \cdot H(\mathbf{s}_U) + O(\log M). \qquad \square$$

Claim 8. *$C(\mathbf{a}_U|\mathbf{a}_V) \le M \cdot H(\mathbf{s}_U|\mathbf{s}_V) + O(\log M)$ for every $U, V \subseteq \{1, \dots, n\}$.*

Proof. The elements of the tuple $\mathbf{a}_V$ are rows of a submatrix. Let us consider its rows: $\mathbf{a}_V = a_V^1 \cdots a_V^M$. We split all positions $i = 1, \dots, M$ into classes corresponding to different values $\alpha_1, \alpha_2, \dots$ of a_V^i. Denote the sizes of these classes by $m_1, m_2, \dots$ By property (a) of the table, each m_j must be proportional to the corresponding probability: the number m_j of positions $i = 1, \dots, M$ such that $a_V^i = \alpha_j$ is equal to

$$\mathbf{Prob}[\mathbf{s}_v = \alpha_j] \cdot M.$$

Given $\mathbf{a}_V$, we describe $\mathbf{a}_U$ by encoding a_U^i separately for different classes of positions corresponding to different values of a_V^i. Similarly to the previous Claim, we get

$$C(\mathbf{a}_U|\mathbf{a}_V) \le \sum_j [m_j H(\mathbf{s}_U|\mathbf{s}_V = \alpha_j) + O(\log m_j)]$$

Therefore,

$$\begin{aligned} C(\mathbf{a}_U|\mathbf{a}_V) &\le M \sum_j \frac{m_j}{M} H(\mathbf{s}_U|\mathbf{s}_V = \alpha_j) + O(\log M) \\ &= M \cdot H(\mathbf{s}_U|\mathbf{s}_V) + O(\log M) \end{aligned} \qquad \square$$

Up to now we had only upper bounds for conditional complexity. To get lower bounds, we recall Kolmogorov–Levin's theorem and note that the complexity of the entire matrix $\mathbf{a}$ is close to $M \cdot H(\mathbf{s})$.

Claim 9. *$C(\mathbf{a}_U) = M \cdot H(\mathbf{s}_U) + O(\log M)$ for every $U \subseteq \{1, \dots, n\}$.*

Proof. The Kolmogorov–Levin theorem says that

$$C(\mathbf{a}) = C(\mathbf{a}_U) + C(\mathbf{a}_V|\mathbf{a}_U) + O(\log C(\mathbf{a})),$$

where V is the complement of U. The error term is $O(\log M)$ since (for fixed $\mathbf{s}$) the complexity of $\mathbf{a}$ is proportional to M. As we already know, both $C(\mathbf{a}_U)$ and $C(\mathbf{a}_V|\mathbf{a}_U)$ are bounded (with $O(\log M)$-precision) by $M \cdot H(\mathbf{s}_U)$ and $M \cdot H(\mathbf{s}_V|\mathbf{s}_U)$, and the left-hand side reaches the sum of these bounds, i.e., $M \cdot H(\mathbf{s})$, by construction. So both terms should also reach their upper bounds. $\square$

The same Kolmogorov–Levin theorem now guarantees that the inequalities of Claims 7 and 8 are equalities with the same $O(\log M)$-precision. Thus, we have constructed an n-tuple of binary strings $\mathbf{a}$ whose complexity profile is close to M times the entropy profile of $\mathbf{s}$, up to some logarithmic term.

Now we increase M enough to make the logarithmic terms insignificant for the information ratio (as well as for missing information and information leaks). This finishes the proof of Theorem 39. $\square$

5.2.3 Sharing Any Secret String

Our definition of Algorithmic secret sharing deals with the sharing of a sequence of strings (s_n). We show using an argument based on the typization method that any sequence of strings is suitable as long as it satisfies requirement (0). So when secret sharing is possible at all, it does not depend of the sequence (s_n) to be shared. This result and the previous one indicate that our definition of quasi-perfect implementation with a given information ratio is quite robust.

Theorem 40. *If some access structure Γ can be algorithmically implemented with information ratio ρ, then for every sequence of strings s_n such that $C(s_n) \to \infty$ there exists a sequence of algorithmic secret-sharing schemes with secrets s_n that implements the same access structure with the same information ratio.*

For this proof, we will use Lemma 16 from The Algorithmic Toolbox on page 77. This result will allow us to construct an algorithm scheme with the same parameters where one secret is replaced with another one of approximately the same complexity.

Proof of Theorem 40. Start with an algorithmic secret-sharing scheme defined by the sequence of tuples $(s_n, \sigma_1, \ldots, \sigma_n)$ and a sequence of secrets (κ_n) that satisfies requirement (0). First of all, we can assume that all secrets κ_n and s_n are incompressible, i.e., $C(k) = |k| + O(\log_2 C(k))$. If this is not the case, simply convert each secret string into one of its shortest programs. The resulting tuple still implements the same access structure with the same information ratio, for the complexity profiles are equal up to a logarithmic term.

Notice that to construct an algorithmic secret-sharing scheme where the sequence of secrets is (κ_n), it is enough to construct a corresponding tuples for large enough κ_ℓ. Fix this large enough index ℓ and let $\jmath$ be an index such that $s_\jmath$ is a longest secret string satisfying $|s_\jmath| \leq \log_2 |\kappa_\ell|$. Since (s_n) satisfies requirement (0) such an index exists for ℓ large enough.

Fix an integer λ such that $\lambda|s_\jmath| = |\kappa_\ell| + O(\log_2 |\kappa_\ell|)$. We begin by constructing a new tuple $\mathbf{Q}$ for a secret of size $\lambda|s_\jmath|$ by concatenating λ "clones" of the tuple $\mathbf{q} = (s_\jmath, \sigma_1, \ldots, \sigma_n)$. Here, by "clone" of $\mathbf{x}$ we mean an element of $T_c(\mathbf{x})$ (in the sense of (4.3)) defined on page 76.

The new tuple $\mathbf{Q}$ is then defined by a

$$\mathbf{Q} = (s_\jmath^1 \cdots s_\jmath^\lambda, \sigma_1^1 \cdots \sigma_1^\lambda, \ldots, \sigma_n^1 \cdots \sigma_n^\lambda)$$

such that $\left\| \vec{C}(\mathbf{Q}) - \lambda \vec{C}(\mathbf{q}) \right\| = O(\lambda \log_2 C(\mathbf{q}))$, which is the case for most tuples $(q^i)_{1 \leq i \leq \lambda}$ of clones in $T_c(\mathbf{q})$.

Now we apply Lemma 16 to obtain a new tuple $\mathbf{Q}'$ for our secret κ_ℓ whose length, and complexity, is approximately the length of the secret $s_\jmath^1 \cdots s_\jmath^\lambda$ in the tuple $\mathbf{Q}$ up to a logarithmic term. The constructed sequence of tuples $\mathbf{Q}'$ is an approximate secret-sharing scheme where the secrets are κ_n for which

$$\left\| \vec{C}(\mathbf{Q}') - \lambda \vec{C}(\mathbf{q}) \right\| = O(\lambda \log_2 C(\mathbf{q})).$$

The logarithmic overhead is tailored to become negligible for ratios, as is showed by the following computation :

$$\begin{aligned} \frac{C(\mathbf{Q}')}{C(\kappa_n)} &= \frac{\lambda\, C(\mathbf{q}) + O(\lambda \log_2 C(\mathbf{q}))}{C(\kappa_n)} \\ &= \frac{C(\mathbf{q})}{C(s_\jmath)} + O\left(\frac{\log_2 C(\mathbf{q})}{C(s_\jmath)}\right) \to \rho \end{aligned}$$

In words, the complexity of the tuple $\mathbf{Q}$ grows linearly in the complexity of the secret κ_n since the information ratio of the original scheme is bounded. The logarithmic overhead is thence not important and the scheme $\mathbf{Q}$ has the same parameters as $\mathbf{q}$, which concludes the proof. □

Conclusion & Comments

This chapter is adapted from the papers [Kac11] and [Kac12].

Chapter 6

Essentially Conditional Information Inequalities

Contents

Introduction

Chapter 2 investigated Entropy Regions and in particular the set $\bar{\Gamma}_n^*$ of almost entropic points. More particularly its dual characterization using unconditional information inequalities. We further pursue this task and shift the focus on conditional inequalities.

We begin by proving all nontrivial conditional inequalities from the literature to be essentially conditional. For two conditional inequalities we even prove a somewhat stronger property. Unlike unconditional inequalities, which hold entropic *and* almost entropic points, it was unknown whether the same was true for all conditional inequalities. Our main result is twofold: we prove that there are two types of essentially conditional inequalities. The ones that are valid for $\bar{\Gamma}_n^*$ and the one that are not. We show that we can extend the notion of conditional information inequalities to the Kolmogorov framework.

To prove that an inequality is essentially conditional we use two elementary techniques: chosen ad-hoc families of binary distributions with suitable limits of entropy values, and a geometric example based on elementary algebra. To prove that some conditional inequality does not hold for almost entropic points, we combine these constructions with Slepian-Wolf coding and the method of quasi-uniform distributions. For the case of algorithmic information theory we use the techniques of typical tuples and Muchnik's theorem on conditional descriptions.

The material in this chapter is adapted from a joint work with Andrei Romashchenko, and extends some results of František Matúš. Most of the results can be found in the conference papers [KR11, KR12b] and submitted journal paper [KR12a].

6.1 Definition and Inequalities

To explain the notion of *essential conditionality*, we start with very basic examples of conditional inequalities. In the three following examples we have statements of the form

> *If some equalities hold for entropies of $x_1, \ldots, x_n$, then some inequality holds for entropies of these variables.*

This is what we call a *conditional information inequality* (also referred to as a *constrained information inequality*).

Example 10. *If $I(a{:}b) = 0$, then*

$$H(a) + H(b) \le H(a, b).$$

This follows immediately from the definition of the mutual information.

Example 11. *If $I(a{:}b) = 0$, then*

$$H(a) + H(b) + H(c) \le H(ac) + H(ac).$$

This follows from an unconditional Shannon-type inequality: for all a, b, c

$$H(a) + H(b) + H(c) \le H(ac) + H(bc) + I(a{:}b).$$

This inequality is the sum of two basic inequalities: $H(c|ab) \ge 0$ and $I(a{:}b|c) \ge 0$.

Example 12. *If $I(e{:}c|d) = I(e{:}d|c) = I(c{:}d|e) = 0$, then*

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b).$$

This is a corollary of the non-Shannon-type inequality from [MMRV02] (see Section 2.7.2 p. 34).

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(e{:}c|d) + I(e{:}d|c) + I(c{:}d|e)$$

which holds for all (a, b, c, d, e) (without any constraints on the distribution).

In examples 10–12 the conditional inequalities are a direct consequence of corresponding unconditional ones. But not all proofs of conditional inequalities are so straightforward. In 1997 Z. Zhang and R.W. Yeung came up with a conditional inequality

$$I(a{:}b) = I(a{:}b|c) = 0 \Rightarrow I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b), \tag{6.1}$$

see [ZY97]. If we wanted to prove (6.1) similarly to examples 10–12 above, then we should first prove an unconditional inequality

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \kappa_1 I(a{:}b) + \kappa_2 I(a{:}b|c) \tag{6.2}$$

with some "Lagrange multipliers" $\kappa_1, \kappa_2 \ge 0$. However, the proof provided in [ZY97] does not follow this scheme. Can we still find an alternative proof of (6.1) that would be based on an instance of (6.2), for some κ_1 and κ_2?

The existence of such an inequality was conjectured in a paper of Makarychev *et al* [MMRV02]. We answer this conjecture in the negative and prove that, whatever the values κ_1, κ_2, the unconditional inequality (6.2) does not hold for Shannon entropy.

We are now ready to define this new class of conditional inequalities:

Definition 36 (Essentially Conditional Information Inequality)**.**
Let $\alpha(X_{\mathcal{N}})$ *and* $\beta_1(X_{\mathcal{N}}), \ldots, \beta_m(X_{\mathcal{N}})$ *be linear functions on the entropies of* $X_{\mathcal{N}} = (X_1, \ldots, X_n)$*:*

$$\alpha(X_{\mathcal{N}}) = \sum_{\varnothing \neq J \subseteq \mathcal{N}} \alpha_J H(X_J),$$
$$\beta_i(X_{\mathcal{N}}) = \sum_{\varnothing \neq J \subseteq \mathcal{N}} \beta_{i,J} H(X_J), \text{ for } i \in [\![m]\!]$$

such that the implication

$$(\beta_i(X_{\mathcal{N}}) = 0 \text{ for all } i \in [\![m]\!]) \Rightarrow \alpha(X_{\mathcal{N}}) \geq 0$$

holds for all distributions $X_{\mathcal{N}}$*. We call this implication* a conditional linear information inequality*. This conditional inequality is said* essentially conditional *if for all* $(\kappa_i)_{1 \leq i \leq m}$ *the inequality*

$$\alpha(X_{\mathcal{N}}) + \sum_{i=1}^{m} \kappa_i \beta_i(X_{\mathcal{N}}) \geq 0$$

does not hold (for some distribution).

Remark 9. *If all functions* β_i *are non-negative, then it only makes sense to consider non-negative values for* κ_i*. In that case, one can equivalently assume that all variables* κ_i *are the same.*

In this chapter, we will concentrate on the following collection of conditional inequalities:

$$\text{if } I(a:b|c) = I(a:b) = 0 \text{ then } I(c:d) \leq I(c:d|a) + I(c:d|b) \quad (\mathcal{I}1)$$
$$\text{if } I(a:b|c) = I(b:d|c) = 0 \text{ then } I(c:d) \leq I(c:d|a) + I(c:d|b) + I(a:b) \quad (\mathcal{I}2)$$
$$\text{if } I(a:b|c) = H(c|ab) = 0 \text{ then } I(c:d) \leq I(c:d|a) + I(c:d|b) + I(a:b) \quad (\mathcal{I}3)$$
$$\text{if } I(a:c|d) = I(a:d|c) = 0 \text{ then } I(c:d) \leq I(c:d|a) + I(c:d|b) + I(a:b) \quad (\mathcal{I}4)$$
$$\text{if } I(a:c|d) = I(c:d|a) = 0 \text{ then } I(c:d) \leq I(c:d|a) + I(c:d|b) + I(a:b) \quad (\mathcal{I}5)$$

Some of these have been presented and discussed in Chapter 2 on page 29.

6.2 Proving the Essential Conditionality

In this section, we show that ($\mathcal{I}1$), ($\mathcal{I}2$), ($\mathcal{I}3$), ($\mathcal{I}4$) and ($\mathcal{I}5$) are essentially conditional. First, we provide ad-hoc (counter)examples for each single inequality. These examples are simple quadruples of binary random variables found with the help of computer search. Then we provide an algebraic proof using an intuitive geometric example which proves that ($\mathcal{I}1$) and ($\mathcal{I}3$) are essentially conditional.

Theorem 41. *Inequalities* ($\mathcal{I}1$), ($\mathcal{I}2$), ($\mathcal{I}3$), ($\mathcal{I}4$) *and* ($\mathcal{I}5$) *are essentially conditional.*

6.2.1 Binary Counterexamples

Claim 10. *Inequality* ($\mathcal{I}1$) *is essentially conditional:*

For any κ the inequality

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + \kappa[I(a{:}b|c) + I(a{:}b)] \tag{6.3}$$

does not hold for some distributions (a, b, c, d).

Proof. For all $\varepsilon \in [0, 1]$, consider the following joint distribution of binary variables (a, b, c, d):

$$\begin{array}{rcl}
\Pr[a=0,\ b=0,\ c=0,\ d=1] &=& (1-\varepsilon)/4,\\
\Pr[a=0,\ b=1,\ c=0,\ d=0] &=& (1-\varepsilon)/4,\\
\Pr[a=1,\ b=0,\ c=0,\ d=1] &=& (1-\varepsilon)/4,\\
\Pr[a=1,\ b=1,\ c=0,\ d=1] &=& (1-\varepsilon)/4,\\
\Pr[a=1,\ b=0,\ c=1,\ d=1] &=& \varepsilon.
\end{array}$$

For each value of a and for each value of b, the value of at least one of variables c, d is uniquely determined: if $a = 0$ then $c = 0$; if $a = 1$ then $d = 1$; if $b = 0$ then $d = 1$; and if $b = 1$ then $c = 0$. Hence, $I(c{:}d|a) = I(c{:}d|b) = 0$. Also it is easy to see that $I(a{:}b|c) = 0$. Thus, if (6.3) is true, then $I(c{:}d) \le \kappa I(a{:}b)$.

Denote the right-hand and left-hand sides of this inequality by $L(\varepsilon) = I(c{:}d)$ and $R(\varepsilon) = \kappa I(a{:}b)$. Both functions $L(\varepsilon)$ and $R(\varepsilon)$ are continuous, and $L(0) = R(0) = 0$ (for $\varepsilon = 0$ both sides of the inequality are equal to 0). However the asymptotics of $L(\varepsilon)$ and $R(\varepsilon)$ as $\varepsilon \to 0$ are different: it is not hard to check that $L(\varepsilon) = \Theta(\varepsilon)$, but $R(\varepsilon) = O\left(\varepsilon^2\right)$. From (6.3) it follows $\Theta(\varepsilon) \le O\left(\varepsilon^2\right)$, which is a contradiction. □

Claim 11. *Inequality ($\mathcal{I}2$) is essentially conditional:*
For any κ the inequality

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \kappa[I(a{:}b|c) + I(b{:}d|c)] \tag{6.4}$$

does not hold for some distributions (a, b, c, d).

Proof. For the sake of contradiction we consider the following joint distribution of binary variables (a, b, c, d) for every value of $\varepsilon \in \left[0, \frac{1}{3}\right]$:

$$\begin{array}{rcl}
\Pr[a=0,\ b=0,\ c=0,\ d=0] &=& 3\varepsilon,\\
\Pr[a=1,\ b=1,\ c=0,\ d=0] &=& 1/3-\varepsilon,\\
\Pr[a=1,\ b=0,\ c=1,\ d=0] &=& 1/3-\varepsilon,\\
\Pr[a=0,\ b=1,\ c=0,\ d=1] &=& 1/3-\varepsilon.
\end{array}$$

We substitute this distribution in (6.4) and obtain

$$I_0 + O(\varepsilon) \le I_0 + 3\varepsilon \log \varepsilon + O(\varepsilon) + O(\kappa\varepsilon),$$

where I_0 is the mutual information between c and d for $\varepsilon = 0$ (which is equal to the mutual information between a and b for $\varepsilon = 0$). We get a contradiction as $\varepsilon \to 0$. □

Claim 12. *Inequality ($\mathcal{I}3$) is essentially conditional:*
For any κ the inequality

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \kappa[I(a{:}b|c) + H(c|ab)] \tag{6.5}$$

does not hold for some distributions (a, b, c, d).

Proof. For every value of $\varepsilon \in \left[0, \frac{1}{2}\right]$ we consider the following joint distribution of binary variables (a, b, c, d):

$$\begin{array}{rcl} \Pr[a=1,\ b=1,\ c=0,\ d=0] &=& 1/2-\varepsilon, \\ \Pr[a=0,\ b=1,\ c=1,\ d=0] &=& \varepsilon, \\ \Pr[a=1,\ b=0,\ c=1,\ d=0] &=& \varepsilon, \\ \Pr[a=0,\ b=0,\ c=1,\ d=1] &=& 1/2-\varepsilon. \end{array}$$

First, it is not hard to check that $I(c{:}d|a) = I(c{:}d|b) = H(c|a,b) = 0$ for every ε. Second,

$$\begin{aligned} I(a{:}b) &= 1 + (2 - 2/\ln 2)\varepsilon + 2\varepsilon\log\varepsilon + O(\varepsilon^2), \\ I(c{:}d) &= 1 + (4 - 2/\ln 2)\varepsilon + 2\varepsilon\log\varepsilon + O(\varepsilon^2), \end{aligned}$$

so $I(a{:}b)$ and $I(c{:}d)$ both tend to 1 as $\varepsilon \to 0$, but their asymptotics are different. Similarly,

$$I(a{:}b|c) = O\left(\varepsilon^2\right).$$

It follows from (6.5) that

$$2\varepsilon + O\left(\varepsilon^2\right) \le O\left(\varepsilon^2\right) + O\left(\kappa\varepsilon^2\right),$$

and with any κ we get a contradiction for small enough ε. □

Claim 13. *Inequality (I4) is essentially conditional:*
For any κ the inequality

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \kappa[I(a{:}c|d) + I(a{:}d|c)] \tag{6.6}$$

does not hold for some distributions.

Proof. For all $\varepsilon \in [0, \frac{1}{4}]$, consider the following joint distribution of binary variables (a, b, c, d):

$$\begin{array}{rcl} \Pr[a=0,\ b=0,\ c=0,\ d=0] &=& \varepsilon, \\ \Pr[a=1,\ b=1,\ c=0,\ d=0] &=& \varepsilon, \\ \Pr[a=0,\ b=1,\ c=1,\ d=0] &=& \frac{1}{4}, \\ \Pr[a=1,\ b=1,\ c=1,\ d=0] &=& \frac{1}{4}-\varepsilon, \\ \Pr[a=0,\ b=0,\ c=0,\ d=1] &=& \frac{1}{4}-\varepsilon, \\ \Pr[a=1,\ b=0,\ c=0,\ d=1] &=& \frac{1}{4}. \end{array}$$

For this distribution, the difference between the right-hand side and the left-hand side of inequality (6.6) rewrites to

$$-I(c{:}d) + I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \kappa[I(a{:}c|d) + I(a{:}d|c)] = -\frac{2}{\ln 2}\varepsilon^2 + O\left(\kappa\varepsilon^3\right),$$

which is negative for ε small enough. □

Claim 14. *Inequality (I5) is essentially conditional:*
For any κ the inequality

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \kappa[I(b{:}c|d) + I(c{:}d|b)] \tag{6.7}$$

does not hold for some distributions.

Proof. For all $\varepsilon \in \left[0, \frac{1}{2}\right]$, consider the following joint distribution for (a, b, c, d):

$$\begin{array}{rcl} \Pr[a=0,\ b=0,\ c=0,\ d=0] & = & \frac{1}{2}-\varepsilon, \\ \Pr[a=0,\ b=1,\ c=0,\ d=1] & = & \frac{1}{2}-\varepsilon, \\ \Pr[a=1,\ b=0,\ c=1,\ d=0] & = & \varepsilon, \\ \Pr[a=1,\ b=1,\ c=0,\ d=0] & = & \varepsilon. \end{array}$$

For this distribution we have $I(c{:}d|a) = I(c{:}d|b) = I(a{:}b) = 0$, and

$$I(c{:}d) = \varepsilon + O\left(\varepsilon^2\right) \text{ and } I(b{:}c|d) = O\left(\varepsilon^2\right).$$

Therefore, the difference between the right-hand side and the left-hand side of inequality (6.7) rewrites to

$$-I(c{:}d) + I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \kappa[I(b{:}c|d) + I(c{:}d|b)] = -\varepsilon + O\left(\kappa\varepsilon^2\right),$$

which is negative for ε small enough. □

6.2.2 An Algebraic Counterexample

Consider a random quadruple $(a, b, c, d)_q$ of geometric objects on the affine plane over the finite field $\mathbb{F}_q$:

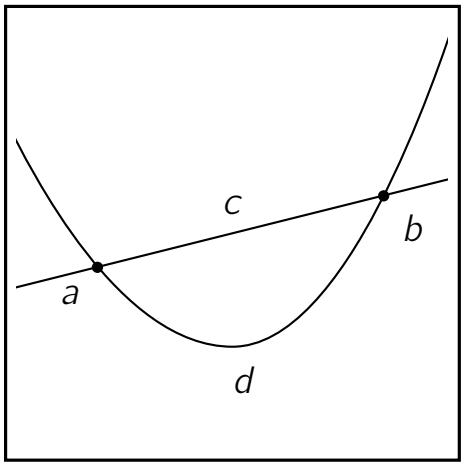


Figure 6.1: A typical configuration of random objects (a, b, c, d).

- First, choose a random non-vertical line c defined by the equation $y = c_0 + c_1 x$ (the coefficients c_0 and c_1 are independent random elements of the field);
- then pick independently and uniformly two points a and b in line c (technically, $a = (a_1, a_2)$ and $b = (b_1, b_2)$, where a_i and b_i are elements of $\mathbb{F}_q$; since the points are chosen independently, they coincide with each other with probability $1/q$);
- pick uniformly at random a parabola d in the set of all non-degenerate parabolas $y = d_0 + d_1 x + d_2 x^2$ (where $d_0, d_1, d_2 \in \mathbb{F}_q$, $d_2 \neq 0$) that intersect c at points a and b (if $a = b$ then c is the tangent line to d at b).

A typical quadruple is shown on Figure 6.1 and elementary events are depicted in Figure 6.2. The probability space defined above has the uniform distribution, *i.e.*, the three configurations shown in figure have the same probability.

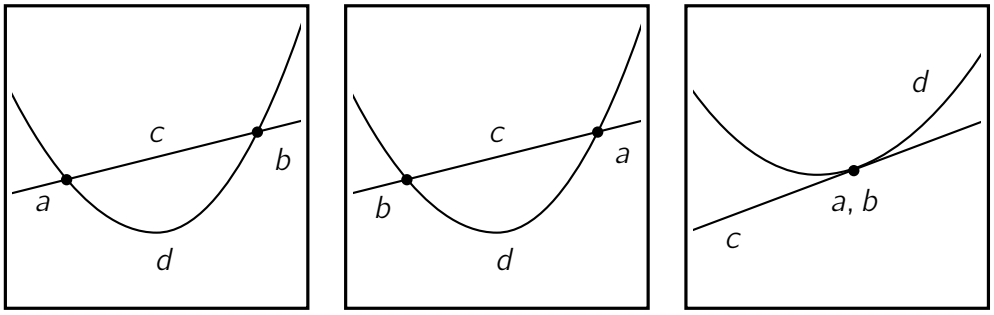


Figure 6.2: Depiction of elementary events of the probability space underlying (a, b, c, d).

Study of the entropy profile of (a, b, c, d)**:** By construction, if the line c is known, then points a and b are independent, *i.e.*,

$$I(a{:}b|c) = 0.$$

Similarly, a and b are independent when d is known, and also c and d are independent given a and given b (when an intersection point is given, the line does not give more information about the parabola), *i.e.*,

$$I(a{:}b|d) = I(c{:}d|a) = I(c{:}d|b) = 0.$$

The mutual information between c and d is approximately 1 bit because randomly chosen line and parabola intersect iff the discriminant of the corresponding equation is a quadratic residue, which happens almost half of the time. A more accurate computation gives $I(c{:}d) = \frac{q-1}{q}$.

When a and b are known and $a \neq b$, then c is uniquely defined (the only line incident with both points). If $a = b$ (which happens with probability $1/q$) we need $\log q$ bits to specify c. Hence,

$$H(c|ab) = \frac{\log q}{q}.$$

To estimate $I(a{:}b)$ we note that a is uniformly distributed on $\mathbb{F}_q^2$. When $b = (b_1, b_2)$ is known, then with probability $1/q$ we have $a = b$, and with probability $1 - 1/q$ the value of a is uniformly chosen among $q(q-1)$ values (a_1, a_2) such that $a_1 \neq b_1$. Hence, $I(a{:}b) = \frac{\log q}{q}$.

By computing both sides of the inequality

$$I(c{:}d) \leq I(c{:}d|a) + I(c{:}d|b) + \kappa_1 I(a{:}b) + \kappa_2 I(a{:}b|c) + \kappa_3 H(c|ab)$$

we obtain

$$1 - \frac{1}{q} \leq \kappa_1 \frac{\log q}{q} + \kappa_3 \frac{\log q}{q}.$$

This leads to a contradiction for large q. It follows that ($\mathcal{I}1$) and ($\mathcal{I}3$) are essentially conditional.

Remark 10. *We can slightly change the probability distribution above. Define* $(a, b, c, d)'_q$ *as follows*

- *Choose a parabola* c *and a line* d *that intersect at* **exactly two points** *uniformly at random over all such possible couples* (c, d).
- *Pick* a *and* b *at random amongst the two intersection points.*

For this quadruple, the third configuration in Figure 6.2 no longer exists, while the first two have the same probability. One can check, rather easily, that the entropy profile of $(a, b, c, d)'_q$ *satisfies:*

$$\begin{aligned} I(c{:}d) &= 1 + [\log_2(q) - \log_2(q-1)], \\ I(a{:}b) &= \log_2(q) - \log_2(q-1), \\ I(a{:}b|c) &= \log_2(q) - \log_2(q-1), \\ I(c{:}d|a) &= \log_2(q) - \log_2(q-1), \\ I(c{:}d|b) &= \log_2(q) - \log_2(q-1), \\ H(c|ab) &= 0. \end{aligned}$$

For instance, given a *there are* q *equiprobable lines* c*. Given* a *and* d*, there are* $q-1$ *equiprobable lines since now the tangent to* d *at* a *is excluded* ($a \neq b$)*. Hence* $I(c{:}d|a) = \log_2(q) - \log_2(q-1)$*. All other computations are similar.*

This distribution (rather a family of distributions parametrized by the size of the field q*) can be also used to prove that* ($\mathcal{I}1$) *and* ($\mathcal{I}3$) *are essentially conditional. Some properties of this particular quadruple (in particular* $H(c|a,b) = 0$*) may appear useful on their own. However, in the sequel we prefer to use the first version of the distribution* $(a, b, c, d)_q$*, which enjoys the property* $I(a; b|c) = 0$*.*

6.2.3 A Stronger Result for Two Conditional Inequalities

The previous construction implies a stronger result than Theorem 41 for inequalities ($\mathcal{I}1$) and ($\mathcal{I}3$). Consider the following information conditional equality which follows from ($\mathcal{I}1$) and ($\mathcal{I}3$).

$$I(a{:}b|c) = I(a{:}b|d) = H(c|ab) = I(c{:}d|a) = I(c{:}d|b) = I(a{:}b) = 0 \Rightarrow I(c{:}d) = 0.$$

We claim that this conditional (in)equality is also *essentially conditional*:

Theorem 42. *There is no constant* κ *such that for all random variables* a, b, c, d

$$I(c{:}d) \leq \kappa[I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(a{:}b|c) + I(a{:}b|d) + H(c|a,b)].$$

Proof. For the quadruple $(a, b, c, d)_q$ from the geometric example defined above, each term in the right-hand side of the inequality vanishes as q tends to infinity, but the left-hand side does not. □

Remark 11. *The conditional (in)equality above is weaker than* ($\mathcal{I}1$) *and* ($\mathcal{I}3$)*; so, Theorem 42 is stronger than what is stated in Theorem 41 for these inequalities.*

6.3 The Case of Almost Entropic Points

Unconditional inequalities for entropic points also hold for almost entropic points. This can be proven directly by taking limits over entropic points. However, for conditional inequalities the limit argument is not valid anymore. Thus, it makes sense to investigate the existence of conditional inequalities that hold for all entropic points but not for all almost entropic points.

6.3.1 Conditional Inequalities for Almost Entropic Points

We recall that some conditional inequalities hold for almost entropic points.

Trivial Inequalities

For instance, any trivial (non-essentially conditional) conditional inequality is also valid for the set of all almost entropic points by definition. This is because such an inequality follows directly from an unconditional one, which still permits the use of the limit argument.

Essentially Conditional Inequalities

Another type of conditional inequalities hold for almost entropic points. For instance, the inequalities proven implicitly by Matúš in [Mat07b]. In his argument, a conditional inequality is obtained as a limit of an infinite family of unconditional inequalities.

Theorem 43 (F. Matúš)**.** *For every distribution* (a, b, c, d, e)

$$I(a{:}c|d) = I(a{:}d|c) = 0 \Rightarrow I(c{:}d) \leq I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(a{:}c|e) + I(a{:}e|c), \tag{$\mathcal{I}4'$}$$

$$I(b{:}c|d) = I(c{:}d|b) = 0 \Rightarrow I(c{:}d) \leq I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(b{:}c|e) + I(c{:}e|b), \tag{$\mathcal{I}5'$}$$

$$I(b{:}c|d) = I(c{:}d|b) = 0 \Rightarrow I(c{:}d) \leq I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(c{:}d|e) + I(c{:}e|d). \tag{$\mathcal{I}5''$}$$

These inequalities hold not only for entropic but also for almost entropic points.

Proof. The following sequences of unconditional inequalities were proven in [Mat07b] for all $k \in \mathbb{N}$:

$$I(c{:}d) \leq I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(a{:}c|e) + I(a{:}e|c) + \frac{1}{k} I(c{:}e|a) + \frac{k-1}{2}[I(a{:}d|c) + I(a{:}c|d)], \tag{6.8}$$

$$I(c{:}d) \leq I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(b{:}c|e) + I(c{:}e|b) + \frac{1}{k} I(b{:}e|c) + \frac{k-1}{2}[I(b{:}c|d) + I(c{:}d|b)], \tag{6.9}$$

$$I(c{:}d) \leq I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(c{:}d|e) + I(c{:}e|d) + \frac{1}{k} I(d{:}e|c) + \frac{k-1}{2}[I(b{:}c|d) + I(c{:}d|b)] \tag{6.10}$$

The constraints in ($\mathcal{I}4'$), ($\mathcal{I}5'$) and ($\mathcal{I}5''$) imply that the terms with the coefficient $\frac{k-1}{2}$ in (6.8),(6.9) and (6.10) (respectively) are equal to zero. The terms with the coefficient $\frac{1}{k}$ vanish as k tends to infinity. So in the limit we obtain from (6.8),(6.9) and (6.10) the required inequalities ($\mathcal{I}4'$), ($\mathcal{I}5'$) and ($\mathcal{I}5''$)

Note that linear inequalities (6.8),(6.9) and (6.10) hold for all points in the cone of almost entropic points. Hence, the limits ($\mathcal{I}4'$), ($\mathcal{I}5'$) and ($\mathcal{I}5''$) are also valid for almost entropic points. □

So we have the following corollary :

Corollary 7. *Inequalities* ($\mathcal{I}4$) *and* ($\mathcal{I}5$) *hold for almost entropic points.*

6.3.2 Conditional Inequalities not Valid for Almost Entropic Points

We have just seen that the conditional inequalities ($\mathcal{I}4$),($\mathcal{I}5$),($\mathcal{I}4'$),($\mathcal{I}5'$) and ($\mathcal{I}5''$) hold for almost entropic points. However, this is not the case for two other essentially conditional inequalities.

Theorem 44. *Inequalities* ($\mathcal{I}1$) *and* ($\mathcal{I}3$) *do not hold for the set of all almost entropic points.*

Proof. The main technical tool used in our proof is Slepian–Wolf coding. We do not need the general version of the classic Slepian–Wolf theorem, we use only its special case. Actually this special case makes the most important part of the general proof of the standard Slepian–Wolf theorem (see Section 4.1.2 in the Toolbox Chapter 4 on page 67).

Construction of an almost entropic counterexample for ($\mathcal{I}1$):

1. Start with distribution $(a, b, c, d)_q$ defined in Section 6.2.2 (page 102). The value of q is specified in what follows. For this distribution, $I(a:b|c) = 0$ holds but $I(a:b) \neq 0$. So far, distribution $(a, b, c, d)_q$ does not satisfy the conditions of ($\mathcal{I}1$).

2. Serialize it: define a new quadruple (A, B, C, D) such that each entropy is N times greater. (A, B, C, D) is obtained by sampling N times independently (a_i, b_i, c_i, d_i) according to the distribution (a, b, c, d) and letting $A = (a_1, \ldots, a_N)$, $B = (b_1, \ldots, b_N)$, $C = (c_1, \ldots, c_N)$, and $D = (d_1, \ldots, d_N)$.

3. Apply Slepian–Wolf coding Lemma 13 and define $A' = SW(A|B)$, then replace A with A' in the quadruple.

   The entropy profile of (A', B, C, D) cannot be far different from the entropy profile for A, B, C, D. Indeed, by construction, $H(A'|A) = 0$ and $H(A|A') \leq I(A:B) + o(N)$. Hence, the difference between entropies involving (A', B, C, D) and (A, B, C, D) is at most
   $$I(A:B) + o(N) = O\left(\frac{\log q}{q} N\right).$$
   Notice that $I(A':B|C) = 0$ since A' functionally depends on A, and $I(a:b|c) = 0$ in the initial distribution.

4. Scale down the entropy profile of (A', B, C, D) by a factor of $1/N$. More precisely, if the entropy profile of (A', B, C, D) is some point $\vec{h} \in \mathbb{R}^{15}$, then for every $\varepsilon > 0$ there exists another distribution (A'', B'', C'', D'') with an entropy profile $\vec{h}'$ such that
   $$\left\| \vec{h}' - \frac{1}{N}\vec{h} \right\| < \varepsilon.$$
   This follows from convexity of the set of almost entropic points (The new distribution can be constructed explicitly, see Proposition 9 on p. 26). We may assume that $\varepsilon = 1/N$.

5. Tend N to infinity. The resulting sequence of entropy profiles have a limit point which is almost entropic. This point does not satisfy ($\mathcal{I}1$) (for q large enough). Indeed, on

one hand, the values of $I(A'':B'')$ and $I(A'':B''|C'')$ converge to zero. On the other hand, inequality $I(C'':D'') \le I(C'':D''|A'') + I(C'':D''|B'')$ results in

$$1 - O\left(\frac{\log_2 q}{q}\right) \le O\left(\frac{\log_2 q}{q}\right),$$

which can not hold for large enough q.

Construction of an almost entropic counterexample for inequality ($\mathcal{I}3$): In this construction we need another lemma based on Slepian–Wolf coding.

Lemma 20. *For every distribution (a, b, c, d) and every integer N there exists a distribution (A', B', C', D') such that the following three conditions hold.*

- $H(C'|A', B') = o(N)$.
- *Denote $\vec{h}$ the entropy profile of (a, b, c, d) and $\vec{h}'$ the entropy profile of (A', B', C', D'); then the components of $\vec{h}'$ differ from the corresponding components of $N \cdot \vec{h}$ by at most*
$$N \cdot H(c|ab) + o(N).$$
- *Moreover, if in the original distribution $I(a:b|c) = 0$, then $I(A':B'|C') = o(N)$.*

Proof of Lemma 20. First we serialize (a, b, c, d), *i.e.*, we take M i.i.d. copies of the initial distribution. The result is a distribution (A, B, C, D) whose entropy profile is exactly the entropy profile of (a, b, c, d) multiplied by M. In particular, we have $I(A:B|C) = 0$. Then, we apply Slepian–Wolf coding (Lemma 13) to get a $Z = SW(C|A, B)$ such that

- $H(Z|C) = 0$,
- $H(Z) = H(C|AB) + o(M)$,
- $H(C|ABZ) = o(M)$.

The entropy profile of the conditional distribution of (A, B, C, D) given Z differs from then entropy profile of (A, B, C, D) by at most $H(Z) = M \cdot H(c|ab) + o(M)$ (*i.e.*, the difference between $H(A)$ and $H(A|Z)$, $H(B)$ and $H(B|Z)$, etc. is not greater than $H(Z)$). Also, if in the original distribution $I(a:b|c) = 0$, then $I(A:B|CZ) = I(A:B|C) = 0$.

We would like to "relativize" (A, B, C, D) conditional on Z and get a new distribution for a quadruple (A', B', C', D') whose unconditional entropies are equal to the corresponding entropies of (A, B, C, D) given Z. This "relativization" procedure is not straightforward since for different values of Z, the corresponding conditional distributions on (A, B, C, D) can be very different. The simplest way to overcome this obstacle is the method of quasi-uniform distributions proposed by Chan and Yeung in [CY02]. The result we need is presented in the Toolbox Chapter 4 on page 65.

For every distribution (A, B, C, D, Z) and every $\delta > 0$ there exists a quasi-uniform distribution $(A'', B'', C'', D'', Z'')$ and an integer k such that

$$\left\| \vec{H}(A, B, C, D, Z) - \frac{1}{k}\vec{H}(A'', B'', C'', D'', Z'') \right\| < \delta.$$

For a quasi-uniform distribution for all values $\mathfrak{z}$ of Z'' the corresponding conditional distributions of (A'', B'', C'', D'') have the same entropies, which are equal to the corresponding conditional entropies. That is, entropies of the distributions of A'', B'', (A'', B''), ... given $Z = \mathfrak{z}$ are

equal to $H(A''|Z)$, $H(B''|Z)$, $H(A''B''|Z)$, ... Thus, for a quasi-uniform distribution we can perform a "relativization" as follows. We fix any value $\mathfrak{z}$ of Z'' and take the conditional distribution on (A'', B'', C'', D'') given $Z'' = \mathfrak{z}$. In this conditional distribution the entropy of C'' given (A'', B'') is not greater than

$$k \cdot [H(C|ABZ) + \delta] = k \cdot [\delta + o(M)].$$

If δ is small enough, then all entropies of (A'', B'', C'', D'') given $Z'' = \mathfrak{z}$ differ from the corresponding components of $kM \cdot \vec{H}(abcd)$ by at most $H(Z'') \leq kM \cdot H(c|ab) + o(kM)$.

Moreover, the mutual information between A'' and B'' given (C'', Z'') is the same as the mutual information between A'' and B'' given only C'', since Z functionally depends on C. If in the original distribution $I(a{:}b|c) = 0$, then the mutual information between (A'', B'') given (C'', Z'') is $o(kM)$.

We choose δ small enough (e.g., $\delta = 1/M$) and let (A', B', C', D') be the constructed above conditional distribution. □

Now we construct an almost entropic counterexample to ($\mathcal{I}3$):

1. Start with the distribution $(a, b, c, d)_q$ from Section 6.2.2 (the value of q is chosen later).

2. Serialize $(a, b, c, d)_q$, *i.e.*, construct (A, B, C, D) by sampling independently N copies of distribution (a, b, c, d).

3. Apply Lemma 20 and get (A', B', C', D') such that $H(C'|A'B') = o(N)$. Lemma 20 guarantees that other entropies of (A', B', C', D') are about N times larger then the corresponding entropies for (a, b, c, d), possibly with an overhead of size

$$O(N \cdot H(c|ab)) = O\left(\frac{\log_2 q}{q} N\right).$$

   From the last bullet of Lemma 20 we also have that $I(A'{:}B'|C') = o(N)$.

4. Scale down the entropy point of (A', B', C', D') by the factor of $1/N$ within precision of $1/N$, similarly to step (4) in the construction above.

5. Tend N to infinity to get an almost entropic point. Conditions of ($\mathcal{I}3$) are satisfied for $I(A'{:}B'|C')$ and $H(C'|A', B')$ both vanish in the limit. Inequality ($\mathcal{I}3$) reduces to

$$1 - O\left(\frac{\log_2 q}{q}\right) \leq O\left(\frac{\log_2 q}{q}\right),$$

   which can not hold if q is large enough. □

The proven result can be rephrased as follows. There exist almost entropic points that satisfy all unconditional linear inequalities for entropies but do not satisfy conditional inequalities ($\mathcal{I}1$) and ($\mathcal{I}3$) respectively.

Remark 12. *Note that one single (large enough) value of q suffices to construct almost entropic counterexamples for ($\mathcal{I}1$) and ($\mathcal{I}3$). However the choice of q in the construction of Theorem 44 provides some freedom: we can control the gap between the left-hand side and the right-hand side of the inequalities. By increasing q we can make the difference between the left-hand side and the right-hand side of inequalities ($\mathcal{I}1$) and ($\mathcal{I}3$) greater than any given term. This property will be used to disprove counterpart of these inequalities on the Kolmogorov Complexity framework in Section 6.4.2.*

In fact, we may combine the two above constructions into one to get a single almost entropic vector to prove the previous result.

Proposition 27. *There exists one almost entropic vector which excludes both* ($\mathcal{I}1$) *and* ($\mathcal{I}3$)*. simultaneously.*

Proof sketch:

1. Generate (A, B, C, D) from $(a, b, c, d)_q$ with entropies N times greater.
2. Construct $A'' = SW(A|B)$ and $C' = SW(C|A, B)$ simultaneously (with the same serialization (A, B, C, D)).
3. Since A'' is a Slepian–Wolf hash of A given B, we have
   - $H(C|A''B) = H(C|AB) + o(N)$ and
   - $H(C|A''BC') = H(C|ABC') + o(N) = o(N)$.

   By inspecting the proof of the Slepian–Wolf theorem we conclude that A'' can be plugged into the argument of Lemma 20 instead of A.
4. The entropy profile of the constructed quadruple (A', B', C', D') is approximately N times the entropy profile of $(a, b, c, d)_q$ with a possible overhead of

$$O(I(A{:}B) + H(C|AB)) + o(N) = O\left(\frac{\log_2 q}{q} N\right),$$

   and further :
   - $I(A'{:}B'|C') = 0$
   - $I(A'{:}B) = o(N)$
   - $H(C'|A'B') = o(N)$
5. Scale the corresponding entropy profile by a factor $1/N$ and tend N to infinity to define the desired almost entropic vector.

6.3.3 The Cone of Almost Entropic Points is not Polyhedral

František Matúš proved the following fundamental result :

Theorem 45 (F. Matúš, [Mat07b])**.** *For $n \geq 4$ the cone $\bar{\Gamma}_n^*$ is not polyhedral. Equivalently, the cone of (unconditional) linear inequalities for the entropies of 4-tuples of random variables is not polyhedral.*

Hereafter, we provide a proof of his result based on essentially conditional inequalities that hold for almost entropic points.

Proof. We prove the theorem for $n = 4$. For the sake of contradiction we assume that the cone of almost entropic points in $\mathbb{R}^{15}$ is polyhedral. That is, the set of almost entropic points is the set of solutions for some finite system of linear inequalities

$$f_1 \geq 0, \ldots, f_s \geq 0,$$

where each f_j is a linear function on $\mathbb{R}^{15}$.

The constraints $I(a:d|c) = I(a:c|d) = 0$ specify a facet (of co-dimension 2) on the boundary of the cone. The corresponding conditional information inequality ($\mathcal{I}4$) specifies a non-degenerate linear function which is non-negative on the corresponding face. Technically, this function is defined by the linear form

$$g = I(c:d|a) + I(c:d|b) + I(a:b) - I(c:d).$$

We will show that this linear function can be extended to the entire space $\mathbb{R}^{15}$ as

$$g' = g + \kappa_1 I(a:d|c) + \kappa_2 I(a:c|d)$$

so that the resulting linear form g' is non-negative on the polyhedral cone. This will imply a contradiction to Theorem 41.

We change the coordinate system. Instead of the standard coordinates $(x_1, \ldots, x_{15})$ corresponding to the entropy values

$$(H(a), H(b), \ldots, H(abcd))$$

we introduce another coordinate systems $(y_1, \ldots, y_{15})$ such that $y_1 = I(a:d|c)$ and $y_2 = I(a:c|d)$. The choice of $y_3, \ldots, y_{15}$ is not important, we only require that the transformation

$$G : (x_1, \ldots, x_{15}) \mapsto (y_1, \ldots, y_{15})$$

is linear and not degenerate.

Inequality ($\mathcal{I}4$) can be reformulated as follows: if $y_1 = y_2 = 0$ then $g \geq 0$ (*i.e.*, linear function g can be represented as a linear form $g = a_3 y_3 + \ldots + a_{15} y_{15}$). On the other hand, in the new coordinate system we can represent each function f_j as

$$f_j = a_{j,1} y_1 + a_{j,2} y_2 + \ldots + a_{j,15} y_{15}$$

Restrictions of each f_j onto the subspace $y_1 = y_2 = 0$ can be specified by 13 real coefficients (instead of 15). Denote

$$f'_j = a_{j,3} y_3 + a_{j,4} y_4 + \ldots + a_{j,15} y_{15}.$$

We know that for all points $\bar{y} = (y_3, \ldots, y_{15})$ such that $f'_j(\bar{y}) \geq 0$ for $j = 1, \ldots, s$, the inequality $g(\bar{y}) \geq 0$ holds. It follows from Farkas' lemma that for some reals $c_j \geq 0$

$$g(\bar{y}) = c_1 f'_1(\bar{y}) + \ldots + c_s f'_s(\bar{y}).$$

From the definition of f'_j we get

$$g(\bar{y}) = c_1(f_1 - a_{1,1} y_1 - a_{1,2} y_2) + \ldots + c_s(f_s - a_{s,1} y_1 - a_{s,2} y_2).$$

This is an identity for linear forms, so their coordinate representations must be equal to each other. Hence, the forms with these coordinate representations are equal to each other on the entire $\mathbb{R}^{15}$. Coming back to the original system of coordinates, we obtain

$$I(c:d|a) + I(c:d|b) + I(a:b) - I(c:d) = \sum c_j f_j - \kappa_1 I(a:d|c) - \kappa_2 I(a:c|d)$$

for some constants κ_1 and κ_2. The sum $\sum c_j f_j$ is non-negative on the entire cone of almost entropic points since all f_j by definition are non-negative on this cone. Thus, we get the

inequality

$$I(c:d) \leq I(c:d|a) + I(c:d|b) + I(a:b) + \kappa_1 I(a:d|c) + \kappa_2 I(a:c|d),$$

which must be true for all distributions (a, b, c, d). This contradicts Theorem 41 (Claim 13), and we are done. □

Remark 13. *The argument above works* mutatis mutandis *for every essentially conditional linear information inequality for the cone of almost entropic points, with constraints that specify some "face" of this cone. In particular, it works for* ($\mathcal{I}4$), ($\mathcal{I}4'$),($\mathcal{I}5$), ($\mathcal{I}5'$),($\mathcal{I}5''$). *The original proof in [Mat07b] corresponds to this argument with inequality* ($\mathcal{I}5$).

6.3.4 On the Geometrical Meaning of Conditional Inequalities

A visual explanation of the different types of 4-variable conditional inequalities may be helpful at this point. However, representing a fifteen dimensional space in this manuscript is not currently possible due to the 3D limitations of our current display technology. We propose to explain the meaning of our conditional inequalities using two-dimensional insights.

The case of trivial conditional inequalities

When a conditional inequality is not essentially conditional, it is a "shade" of one unconditional inequality.

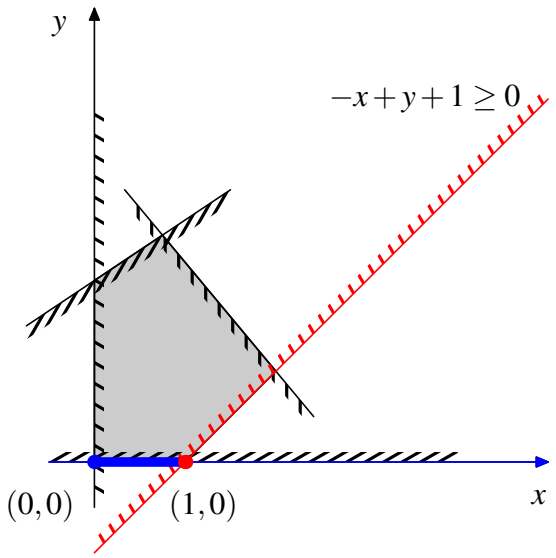


Figure 6.3: Geometric intuition for a trivial conditional inequality

The previous figure shows that a conditional inequality "If $y = 0$ then $x \leq 1$", represented by the thick blue segment, that is valid for the set depicted in gray. The picture shows that in fact this conditional inequality follows from the more general unconditional inequality

$$-x + y + 1 \geq 0.$$

The case of essentially conditional inequalities which are valid for almost entropic points

Now suppose we have a set that still satisfies the conditional inequality: If $y = 0$ then $x \leq 1$, but now this inequality does not follow from an unconditional one. This time, the conditional inequality is implied by an infinite family of tangent half-planes.

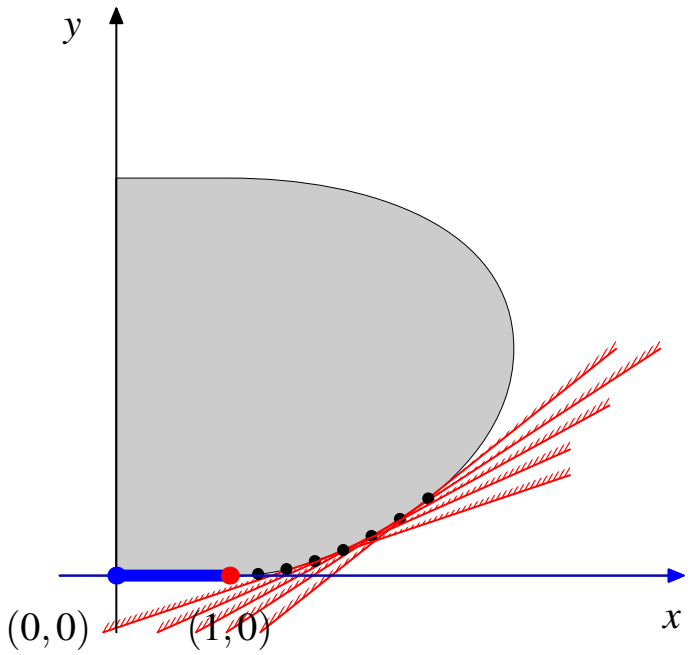


Figure 6.4: Geometric intuition for an essentially conditional inequality for almost entropic points

This geometric interpretation explains the case of the essentially conditional inequalities ($\mathcal{I}4$) and ($\mathcal{I}5$).

The case of essentially conditional inequalities which do not hold for almost entropic points

Sometimes a conditional inequality does not even follow from an infinite family of unconditional ones. The following figure explains what happens for an essentially conditional inequality that does not hold for the closure of the set.

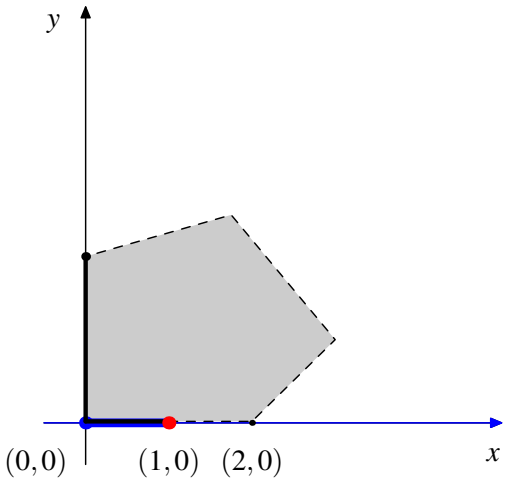


Figure 6.5: Geometric intuition for an essentially conditional inequality which is not valid for almost entropic points

This example explains what happens for an essentially conditional inequality that does not hold for the closure of the set. The conditional inequality If $y = 0$ then $x \leq 1$ is valid for the set in the figure, while for the closure of this set and the same constraint $y = 0$ we have only $x \leq 2$. This example depicts the case of essentially conditional inequalities ($\mathcal{I}1$) and ($\mathcal{I}2$).

6.4 Condition Inequalities for Kolmogorov Complexity

In this section, we suggest a counterpart of conditional inequalities from Information Theory in Algorithmic Information Theory. Recall again that Kolmogorov Complexity is defined up to an additive constant, we thus cannot translate immediately conditional information inequality into conditional algorithmic inequalities. We also know that different flavors of Kolmogorov Complexity are the same up to a logarithmic term in the length of the string, and most statements in Kolmogorov complexity are, indeed, asymptotic. This is why an additive logarithmic term (in the complexity) is considered as a negligible quantity of the most natural kind. Therefore we could naturally translate the constraints of conditional inequalities in that manner. For instance,

$$I(a{:}b|c) = 0 \text{ for Shannon Entropy}$$

is translated into

$$I(a{:}b|c) \leq O(\log n) \text{ for Kolmogorov Complexity.}$$

where n is the length of strings a, b, c. We sometimes choose to keep the utmost generality for expressing the constraints and therefore use a "slack" function $f(n)$ that should satisfy

$$\Omega(\log n) \leq f(n) \leq o(n).$$

Since constraints are softer, we also need to add a possible overhead to the inequality. However this overhead may not be negligible in the same sense as for constraints. Therefore we denote this overhead with another function $g(n)$. For this $g(n)$ we should also assume that

$$\Omega(\log n) \leq g(n) \leq o(n).$$

Such an inequality will be referred to as a *conditional Algorithmic inequality*.

6.4.1 Three Conditional Inequalities for Kolmogorov Complexity

First, we show that inequalities from Theorem 43 can be translated, in some sense, in the language of Kolmogorov complexity.

Theorem 46. *Let $f(n)$ be any function of an integer argument. Then there exists a $\kappa > 0$ such that for every tuple of binary strings (a, b, c, d, e)*

$$\text{if } I(a{:}d|c) \le f(n) \text{ and } I(a{:}c|d) \le f(n), \text{ then}$$
$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(a{:}c|e) + I(a{:}e|c) + \kappa\sqrt{n \cdot f(n)} \quad (C\mathcal{I}4')$$

$$\text{if } I(b{:}c|d) \le f(n) \text{ and } I(c{:}d|b) \le f(n), \text{ then}$$
$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(b{:}c|e) + I(c{:}e|b) + \kappa\sqrt{n \cdot f(n)}, \quad (C\mathcal{I}5')$$

$$\text{if if } I(b{:}c|d) \le f(n) \text{ and } I(c{:}d|b) \le f(n), \text{ then}$$
$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(c{:}d|e) + I(c{:}e|d) + \kappa\sqrt{n \cdot f(n)}, \quad (C\mathcal{I}5'')$$

where n is the sum of lengths of strings a, b, c, d, e.

In this theorem $f(n)$ plays the role of the measure of "precision" of the constraints. Technically the statement of the theorem is true for any $f(n)$, but it is interesting only for $f(n) = o(n)$ (and $f(n) = \Omega(\log n)$, since different definitions of the mutual information in algorithmic information theory are equivalent to each other with only logarithmic precision). For example, assuming $I(a{:}d|c) = O\left(\sqrt{n}\right)$ and $I(a{:}c|d) = O\left(\sqrt{n}\right)$ we get

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(a{:}c|e) + I(a{:}e|c) + O\left(n^{3/4}\right)$$

Proof. By Theorem 31, for every linear inequality for Shannon entropy there exists a counterpart for Kolmogorov complexity that is true for all binary strings up to an additive $O(\log n)$-term. Thus, from inequality (6.8) on page 105 (which holds for the Shannon entropies of any distribution) it follows that a similar inequality holds for Kolmogorov complexity. More precisely, for each integer $k > 0$ there exists a constant D such that for all strings a, b, c, d, e

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \\ + I(a{:}c|e) + I(a{:}e|c) + \frac{1}{k}I(c{:}e|a) + \frac{k-1}{2}[I(a{:}d|c) + I(a{:}c|d)] + D\log n.$$

We choose k that minimizes the sum of $\frac{1}{k}I(c{:}e|a)$ and $\frac{k-1}{2}[I(a{:}d|c) + I(a{:}c|d)]$. The value $\frac{1}{k}I(c{:}e|a)$ is bounded by $O(n/k)$ since all strings are of length at most n; the values $I(a{:}d|c)$ and $I(a{:}c|d)$ are less than $f(n)$. Let $k = \sqrt{n/f(n)}$. Then we get

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + I(a{:}c|e) + I(a{:}e|c) + O\left(\sqrt{n \cdot f(n)}\right)$$

and $(C\mathcal{I}4')$ is proven. Conditional inequalities $(C\mathcal{I}5')$ and $(C\mathcal{I}5'')$ can be proven by a similar argument. □

Theorem 46 involves 5-tuples of strings (a, b, c, d, e) but it implies a nontrivial result for quadruples of strings. By assuming $e = d$ we get from Theorem 46 the following corollary.

Corollary 8. *Let $f(n)$ be a function of an integer argument such that $f(n) \le n$. Then for every tuple of binary strings (a, b, c, d)*

if $I(a{:}d|c) \le f(n)$ and $I(a{:}c|d) \le f(n)$, then

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + O\left(\sqrt{n \cdot f(n)}\right) \quad (C\mathcal{I}4)$$

if $I(b{:}c|d) \le f(n)$ and $I(c{:}d|b) \le f(n)$, then

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + O\left(\sqrt{n \cdot f(n)}\right) \quad (C\mathcal{I}5)$$

where n is the sum of the lengths of all strings involved.

In Theorem 46 and Corollary 8 we deal with two different measures of precision: $f(n)$ in the conditions and $O\left(\sqrt{n \cdot f(n)}\right)$ in the conclusions. These two measures of precision are dramatically different. Assume, for example, that $I(b{:}c|d)$ and $I(c{:}d|b)$ are bounded by $O(\log n)$, which is the most natural conventional assumption of "independence" in algorithmic information theory. Then from Corollary 46 it follows that

$$I(c{:}d) \le I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + O(\sqrt{n \log n}).$$

Can we prove the same inequality with a precision better than $O\left(\sqrt{n \log n}\right)$? The answer is negative: the next proposition shows that this bound is tight.

Proposition 28. *For some $\kappa > 0$, for infinitely many integers n there exists a tuple of strings (a, b, c, d) such that $C(a, b, c, d) = n$, $I(b{:}c|d) = O(\log n)$ and $I(c{:}d|b) = O(\log n)$, and*

$$I(c{:}d) \ge I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + \kappa\sqrt{n \log n}.$$

Proof. Let us take the distribution from Claim 14, p. 101 for some parameter ε, and denote it $(\alpha, \beta, \gamma, \delta)_\varepsilon$. Further, we apply the following simple lemma from [Rom00b].

Lemma 21 (Romashchenko, [Rom00b])**.** *Let $(\alpha, \beta, \gamma, \delta)$ be a distribution on some finite set $\mathcal{M}^4$, and n be an integer. Then there exists a tuple of strings (a, b, c, d) of length n over alphabet $\mathcal{M}$ such that*

$$\vec{C}(a, b, c, d) = n \cdot \vec{H}(\alpha, \beta, \gamma, \delta) + O(|\mathcal{M}| \log n)$$

From this lemma we get a tuple of strings (a, b, c, d) such that the quantities $I(c{:}d|a)$, $I(c{:}d|b)$, $I(a{:}b)$ are bounded by $O(\log n)$, while

$$I(c{:}d) = \Theta(\varepsilon n)$$

and

$$I(b{:}c|d) = O(\varepsilon^2 n).$$

It remains to choose appropriate ε and n. Let $\varepsilon = \sqrt{\frac{\log n}{n}}$. Then $I(b{:}c|d) = O(\varepsilon^2 n) = O(\log n)$ and $I(c{:}d) \ge D\sqrt{n \log n}$ (for some $D > 0$). Hence, we get

$$I(c{:}d) \ge I(c{:}d|a) + I(c{:}d|b) + I(a{:}b) + D\sqrt{n \log n} - O(\log n),$$

and we are done. Keeping in mind details of the construction from Claim 14 we can let here $D = 1$ (though the precise value of D does not matter much for the proof). □

6.4.2 Two Conditional Inequalities not for Kolmogorov Complexity

The following theorem claims that counterparts of ($\mathcal{I}1$) and ($\mathcal{I}3$) do not hold for Kolmogorov complexity.

Theorem 47.
(a) There exists an infinite sequence of tuples of strings $(a, b, c, d)_n$ such that the lengths of all strings a, b, c, d are $\Theta(n)$, $I(a:b) = O(\log n)$, $I(a:b|c) = O(\log n)$, and

$$I(c:d) - I(c:d|a) - I(c:d|b) = \Omega(n)$$

(b) There exists an infinite sequence of tuples of strings $(a, b, c, d)_n$ such that the lengths of all strings a, b, c, d are $\Theta(n)$, $C(c|a, b) = O(\log n)$, $I(a:b|c) = O(\log n)$, and

$$I(c:d) - I(c:d|a) - I(c:d|b) - I(a:b) = \Omega(n).$$

The proof of this theorem is similar to the proof of Theorem 44. Instead of the Slepian–Wolf theorem we use its counterpart in the Kolmogorov framework, which is Muchnik's theorem on conditional description, from our Toolbox p. 75.

Proof of Theorem 47(a). We start with the distribution defined in Section 6.2.2, the value of q is specified in what follows. Let us denote this distribution $(\alpha, \beta, \gamma, \delta)$. Then we apply Lemma 21 and construct strings a, b, c, d such that

$$\vec{C}(a, b, c, d) = n \cdot \vec{H}(\alpha, \beta, \gamma, \delta) + O(\log n).$$

Note that for the constructed a, b, c, d:

$$I(c:d) \gg I(c:d|a) + I(c:d|b) + I(a:b),$$

and $I(a:b|c) = O(\log n)$. But this quadruple of strings does not satisfy the requirements of the theorem since $I(a:b)$ is much greater than $\log n$. Thus, we need to transform a, b, c, d so that

(i) we keep the property $I(c:d) \gg I(c:d|a) + I(c:d|b) + I(a:b)$,

(ii) $I(a:b|c)$ *remains* logarithmic, and

(iii) $I(a:b)$ *becomes* logarithmic.

To this end, we only need to modify the string a.

We apply Theorem 32 for a and b and get $a' \in Much(a|b)$ such that

- $C(a'|a) = O(\log n)$,
- $C(a') = C(a|b) + O(\log n)$,
- $C(a|a', b) = O(\log n)$.

What immediately follows is

- $C(a|a') = I(a:b) + O(\log n)$,
- $I(a':b) = O(\log n)$,
- $I(a':b|c) = O(\log n)$ (since for the original tuple we have $I(a:b|c) = O(\log n)$).

Intuitively a' is the "difference" *a minus b*. In the sequel we investigate the quadruple (a', b, c, d).

The complexity profile of (a', b, c, d) cannot be far different from the complexity profile of (a, b, c, d). Indeed, $C(a'|a) = O(\log n)$ and $C(a|a') = I(a':b) + O(\log n)$. Hence, the difference between corresponding components of complexity profiles $\vec{C}(a', b, c, d)$ and $\vec{C}(a, b, c, d)$ is at most

$$I(a:b) + O(\log n) = O\left(n \cdot \frac{\log q}{q}\right).$$

For the constructed tuple (a', b, c, d) we have

$$\begin{aligned} I(a':b) &= O(\log n) \\ I(a':b|c) &= O(\log n) \end{aligned}$$

On the other hand,

$$\begin{aligned} I(c:d) &= n \cdot \left(1 - \frac{1}{q}\right) + O(\log n), \\ I(c:d|b) &= n \cdot O\left(\frac{\log q}{q}\right) + O(\log n), \end{aligned}$$

and

$$\begin{aligned} I(c:d|a') &\le I(c:d|a) + C(a|a') \\ &= O\left(n \log\left(\frac{\log q}{q}\right)\right) + O(\log n). \end{aligned}$$

Thus, for large enough q we get

$$I(c:d) - I(c:d|a') - I(c:d|b) = \Omega(n). \qquad \square$$

Proof of Theorem 47(b). We again use the distribution from Section 6.2.2 (the value of q is chosen later). Let us denote this distribution $(\alpha, \beta, \gamma, \delta)$ Then we apply to this distribution Lemma 21 and obtain a quadruple of strings (a, b, c, d) such that

$$\vec{C}(a, b, c, d) = n \cdot \vec{H}(\alpha, \beta, \gamma, \delta) + O(\log n).$$

For the constructed a, b, c, d

$$\begin{aligned} I(a:b|c) &= O(\log n) \\ I(c:d) &> I(c:d|a) + I(c:d|b) + I(a:b) - cn \end{aligned}$$

(for some real $c > 0$). However, this quadruple of strings does not satisfy the requirements of the theorem since $H(c|a, b)$ is much greater than $\log n$. It remains to modify a, b, c, d so that

(i) we keep the property $I(c:d) \gg I(c:d|a) + I(c:d|b) + I(a:b)$,

(ii) $I(a:b|c)$ *remains* logarithmic,

(iii) $C(c|a, b)$ *becomes* logarithmic.

We apply Theorem 32 to the constructed strings (a, b, c, d) and get $x \in Much(c|a, b)$ such that

- $C(x|c) = O(\log n)$,
- $C(x) = C(c|a,b) + O(\log n)$,
- $C(x|a,b,x) = O(\log n)$.

Let us consider the conditional complexity profile $\vec{C}(a,b,c,d|x)$. It is not hard to check that the components of $\vec{C}(a,b,c,d|x)$ differ from the corresponding components of $\vec{C}(a,b,c,d)$ by at most

$$C(x) + O(\log n).$$

Moreover, we have $I(a{:}b|c,x) = O(\log n)$ and $C(c|a,b,x) = O(\log n)$. Thus, we would like to "relativize" a, b, c, d given x as an oracle. But this is exactly the goal of Lemma 17 from the Toolbox (p. 78). indeed, from Lemma 17 there exists another tuple of strings (a', b', c', d') such that

$$\vec{C}(a', b', c', d') = \vec{C}(a, b, c, d|x) + O(\log n),$$

where $n = C(a, b, c, d, x)$.

For this tuple: $C(c'|a', b') = O(\log n)$ and $I(a'{:}b'|c) = O(\log n)$. On the other hand,

$$\begin{aligned}
I(c'{:}d') &= n - O\left(n \log\left(\frac{\log q}{q}\right) + \log n\right),\\
I(a'{:}b') &= O\left(n \log\left(\frac{\log q}{q}\right) + \log n\right),\\
I(c'{:}d'|b') &= O\left(n \log\left(\frac{\log q}{q}\right) + \log n\right),\\
I(c'{:}d'|a') &= O\left(n \log\left(\frac{\log q}{q}\right) + \log n\right).
\end{aligned}$$

Hence, for large enough q we get

$$I(c'{:}d') - I(c'{:}d'|a') - I(c'{:}d'|b') - I(a'{:}b') = \Omega(n). \qquad \square$$

References

[AGK76] R. Ahlswede, P. Gács, and J. Körner. Bounds on conditional probabilities with applications in multi-user communication. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 34:157–177, 1976. 68

[AKS10] Andris Ambainis, Julia Kempe, and Or Sattath. A Quantum Lovász Local Lemma. In *Proceedings of the 42nd ACM symposium on Theory of computing*, STOC '10, pages 151–160, New York, NY, USA, 2010. ACM. 22

[Alb11] Boris Albar. Secrets Répartis et Matroïdes. Master's thesis, LIRMM, Université de Montpellier 2, France, 2011. 64

[ALPS09] L. Antunes, S. Laplante, A. Pinto, and L. Salvador. Cryptographic Security of Individual Instances. In *Information Theoretic Security*, volume 4883 of *LNCS*, pages 195–210. Springer Berlin Heidelberg, 2009. 91

[Atı00] Mustafa Atıcı. Information and average information rates of a graphical access structure on six vertices, 2000. 49

[BD91] Ernest F. Brickell and Daniel M. Davenport. On the Classification of Ideal Secret Sharing Schemes. *J. Cryptology*, 4(2):123–134, 1991. 39, 58

[Bei11] Amos Beimel. Secret-sharing schemes: A survey. In *IWCC*, pages 11–46, 2011. 64

[BI01] Amos Beimel and Yuval Ishai. On the Power of Nonlinear Secret-Sharing. In *IEEE Conference on Computational Complexity*, pages 188–202, 2001. 51

[BK97] R. G Blakley and G. A. Kabatiansky. Generalized Ideal Secret-Sharing Schemes and Matroids. *Problems Inform. Transmission*, 33(3):277–284, 1997. 58

[BL88] Josh Cohen Benaloh and Jerry Leichter. Generalized Secret Sharing and Monotone Functions. In *CRYPTO*, pages 27–35, 1988. 46, 47, 61

[BL08] Amos Beimel and Noam Livne. On Matroids and Nonideal Secret Sharing. *IEEE Transactions on Information Theory*, 54(6):2626–2643, 2008. 64

[BLP08] Amos Beimel, Noam Livne, and Carles Padró. Matroids Can Be Far from Ideal Secret Sharing. In *TCC*, pages 194–212, 2008. 42, 60, 64

[BO09] Amos Beimel and Ilan Orlov. Secret Sharing and Non-Shannon Information Inequalities. In *TCC*, pages 539–557, 2009. 60

[Bri90] Ernest F. Brickell. Some Ideal Secret Sharing Schemes. In *EUROCRYPT '89: Proceedings of the workshop on the theory and application of cryptographic techniques on Advances in cryptology*, pages 468–475, New York, NY, USA, 1990. Springer-Verlag New York, Inc. 64

[BS92] E. F. Brickell and D. R. Stinson. Some Improved Bounds on the Information Rate of Perfect Secret Sharing Schemes, 1992. 10.1007/BF02451112. 39

[BSSV97] Carlo Blundo, Alfredo De Santis, Roberto De Simone, and Ugo Vaccaro. Tight Bounds on the Information Rate of Secret Sharing Schemes. *Des. Codes Cryptography*, 11(2):107–122, 1997. 62

[BSV94] Carlo Blundo, Alfredo Santis, and Ugo Vaccaro. Randomness in distribution protocols. In Serge Abiteboul and Eli Shamir, editors, *Automata, Languages and Programming*, volume 820 of *Lecture Notes in Computer Science*, pages 568–579. Springer Berlin Heidelberg, 1994. 62

[BSV98] Carlo Blundo, Alfredo De Santis, and Ugo Vaccaro. On Secret Sharing Schemes. *Inf. Process. Lett.*, 65(1):25–32, 1998. 40, 62

[CF11] Ronald Cramer and Serge Fehr. The Mathematical Theory of information, and Applications (Version 2.0), 2011. 5

[Cha01] Terence Ho Leung Chan. A combinatorial approach to information inequalities. *Communications in Information and Systems*, 1(3), 2001. 65, 66

[Cha03] T. H. Chan. Balanced information inequalities. *IEEE Trans. Inf. Theor.*, 49(12):3261–3267, December 2003. 70

[CK81] Imre Csiszar and Janos Korner. *Information theory : coding theorems for discrete memoryless systems / Imre Csiszar and Janos Korner*. Academic Press ; Akademiai Kiado, New York : Budapest :, 1981. 68

[CL09] László Csirmaz and Péter Ligeti. On an infinite family of graphs with information ratio 2 - $1/k$. *Computing*, 85(1-2):127–136, 2009. 49

[CSGV93] R. M. Capocelli, A. De Santis, L. Gargano, and U. Vaccaro. On the Size of Shares for Secret Sharing Schemes. *J. of Cryptology*, 6:157–168, 1993. 39, 41, 47

[Csi97] László Csirmaz. The Size of a Share Must be Large. *J. Cryptology*, 10(4):223–231, 1997. 62, 63, 64

[Csi09] László Csirmaz. An impossibility result on graph secret sharing. *Des. Codes Cryptography*, 53(3):195–209, 2009. 64

[Csi12] László Csirmaz. Probabilistic infinite secret sharing. *IACR Cryptology ePrint Archive*, 2012:412, 2012. 64

[CT91] Thomas M. Cover and Joy Thomas. *Elements of Information Theory*. Wiley, 1991. 5, 67

[CT09] László Csirmaz and Gábor Tardos. Secret sharing on trees: problem solved. *IACR Cryptology ePrint Archive*, 2009:71, 2009. 49

[CY02] Terence H. Chan and Raymond W. Yeung. On a Relation Between Information Inequalities and Group Theory. *IEEE Trans. on Inform. Theory*, 48:1992–1995, 2002. 22, 65, 66, 84, 107

[Daw93] E. S.; Rahilly Dawson, E.; Mahmoodian. Orthgonal Arrays and Ordered Threshold Schemes. *Australas. J. Combin.*, 8:27–44, 1993. 44, 64

[DFZ06] Randall Dougherty, Christopher Freiling, and Kenneth Zeger. Six new non-shannon information inequalities. In *in Proc. IEEE Int. Symp. Inf. Theory*, pages 233–236, 2006. 34, 64, 67, 68, 69

[DFZ09] Randall Dougherty, Christopher F. Freiling, and Kenneth Zeger. Linear rank inequalities on five or more variables. *CoRR*, abs/0910.0284, 2009. 22, 35

[DFZ11] Randall Dougherty, Christopher F. Freiling, and Kenneth Zeger. Non-shannon information inequalities in four random variables. *CoRR*, abs/1104.3602, 2011. 35

[Dij97] Marten van Dijk. A linear construction of secret sharing schemes. *Designs, Codes and Cryptography*, 12:161–201, 1997. 10.1023/A:1008259214236. 64

[Dod87] Charles L. Dodgson. *The Game of Logic*. London ; New York : Macmillan and Co., 2 edition, 1887. 20

[FMBPV12] Oriol Farràs, Jessica Ruth Metcalf-Burton, Carles Padró, and Leonor Vázquez. On the optimization of bipartite secret sharing schemes. *Des. Codes Cryptography*, 63(2):255–271, 2012. 49

[FP12] Oriol Farras and Carles Padro. Extending brickell-davenport theorem to non-perfect secret sharing schemes. Cryptology ePrint Archive, Report 2012/595, 2012. http://eprint.iacr.org/. 58

[Fuj78] Satoru Fujishige. Polymatroidal dependence structure of a set of random variables. *Information and Control*, 39(1):55 – 72, 1978. 56

[GK73] P. Gács and J. Korner. Common information is far less than mutual information. *Probl. Inform. Control*, 2(2):149–162, 1973. 52

[GM04] Yuan-Bo Guo and Jian-Feng Ma. Practical secret sharing scheme realizing generalized adversary structure. *J. Comput. Sci. Technol.*, 19(4):564–569, July 2004. 47

[Gra44] Hermann Grassmann. *Die lineale Ausdehnungslehre / ein neuer Zweig der Mathematik / dargestellt und durch Anwendungen auf die übrigen Zweige der Mathematik, wie auch auf die Statik, Mechanik, die Lehre vom Magnetismus und die Krystallonomie erläutert*. O. Wigand, Leipzig, 1844. 23

[Gra90] Robert M. Gray. *Entropy and information theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1990. 5

[HCG12] Siu-Wai Ho, Terence Chan, and Alex J. Grant. Non-entropic inequalities from information constraints. In *ISIT*, pages 1256–1260. IEEE, 2012. 28

[Hoe63] Wassily Hoeffding. Probability Inequalities for Sums of Bounded Random Variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. 65, 86

[HRSV00] Daniel Hammer, Andrei Romashchenko, Alexander Shen, and Nikolai Vereshchagin. Inequalities for Shannon Entropy and Kolmogorov Complexity. *J. Comput. System Sci.*, 60(2):442–464, 2000. 22, 29, 74, 92

[Ing71] A. W. Ingleton. Representation of matroids. *Combinatorial Mathematics and its Applications*, 1971. 23

[ISN87] M. Ito, A. Saito, and T. Nishizeki. Secret Sharing Scheme Realizing General Access Structure. In *IEEE Globecom*, pages 99–102, 1987. 46

[JM94] Wen-Ai Jackson and Keith M. Martin. Geometric secret sharing schemes and their duals. *Designs, Codes and Cryptography*, 4:83–95, 1994. 10.1007/BF01388562. 64

[JM96] Wen-Ai Jackson and Keith M. Martin. Perfect secret sharing schemes on five participants. *Designs, Codes and Cryptography*, 9:267–286, 1996. 10.1007/BF00129769. 48, 49

[Kab98] G.A. Kabatianskiy. *Mathematics of Secret Sharing*. New Mathematical Discipline, Moscow, MCCME, 1998. 64

[Kac11] Tarik Kaced. Almost-perfect Secret Sharing. *ISIT 2011, St-Petersburg*, 2011. 96

[Kac12] Tarik Kaced. Quasi-perfect Secret Sharing. *Submitted to Information and Computation*, 2012. 96

[KGH83] Ehud D. Karnin, Jonathan W. Greene, and Martin E. Hellman. On Secret Sharing Systems. *IEEE Trans. on Information Theory*, 29:35–41, 1983. 39, 41

[KL51] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 1951. 12

[KM75] D. Kleitman and G. Markowsky. On Dedekind's problem: The number of isotone Boolean functions. *Trans. AMS*, 214:373–390, 1975. 39

[KO96] Kaoru Kurosawa and Koji Okada. Combinatorial Lower Bounds for Secret Sharing Schemes. *Inf. Process. Lett.*, 60(6):301–304, 1996. 47, 61

[KR11] Tarik Kaced and Andrei Romashchenko. On Essentially Conditional Information Inequalities. *Proceedings IEEE ISIT*, pages 1935–1939, 2011. 31, 89, 97

[KR12a] T. Kaced and A. Romashchenko. Conditional information inequalities for entropic and almost entropic points. *arXiv preprint arXiv:1207.5742*, 2012. 97

[KR12b] T. Kaced and A. Romashchenko. On the non-robustness of essentially conditional information inequalities. *arXiv preprint arXiv:1207.5458*, 2012. 97

[KW93] Mauricio Karchmer and Avi Wigderson. On span programs. In *Structure in Complexity Theory Conference*, pages 102–111, 1993. 64

[Leh64] Alfred Lehman. A Solution of the Shannon Switching Game. *Siam Journal on Applied Mathematics*, 12, 1964. 57

[LM98] Charles F. Laywine and Gary L. Mullen. *Discrete Mathematics Using Latin Squares*. Wiley, 1998. 45

[LV97] M. Li and P. Vitányi. *An Introduction to Kolmogorov complexity and its applications*. Springer-Verlag, second edition, 1997. 72

[Mac03] D. J. C. Mackay. *Information theory, inference, and learning algorithms*. Cambridge University Press, Cambridge, 2003. 5

[Mar91] Keith Murray Martin. *Discrete Structures in the Theory of Secret Sharing*. PhD thesis, University of London, Royal Holloway and Bedford New College, 1991. 44, 64

[Mas95] J. L. Massey. Some applications of coding theory in cryptography. In *Codes and Ciphers: Cryptography and Coding IV*, pages 33–47, 1995. 64

[Mat95] František Matús. Conditional independences among four random variables ii. *Combinatorics, Probability & Computing*, 4:407–417, 1995. 35

[Mat99a] František Matúš. Conditional independences among four random variables iii: Final conclusion. *Combinatorics, Probability and Computing*, 8(03):269–276, 1999. 30, 35

[Mat99b] František Matúš. Matroid Representations by Partitions. *Discrete Mathematics*, 203(1-3):169 – 194, 1999. 42, 60, 64, 90

[Mat06] František Matúš. Piecewise linear conditional information inequality. *IEEE Transactions on Information Theory*, 52(1):236–238, 2006. 28

[Mat07a] Frantisek Matús. Adhesivity of polymatroids. *Discrete Mathematics*, 307(21):2464–2477, 2007. 34

[Mat07b] František Matúš. Infinitely many Information Inequalities. *Proceedings ISIT 2007*, pages 41–44, 2007. 28, 33, 34, 105, 109, 111

[Mat07c] František Matúš. Two constructions on limits of entropy functions. *IEEE Trans. Inf. Theor.*, 53(1):320–330, January 2007. 33, 67

[MB09] Jessica Ruth Metcalf-Burton. *Information Rates for Secret Sharing over Various Access Structures*. PhD thesis, University of Michigan, 2009. 60

[MB11] Jessica Ruth Metcalf-Burton. Improved Upper Bounds for the Information Rates of the Secret Sharing Schemes Induced by the Vamos Matroid. *Discrete Mathematics*, 311(8-9):651 – 662, 2011. 60, 61, 64

[MFP07] Jaume Martí-Farré and Carles Padró. On Secret Sharing Schemes, Matroids and Polymatroids. In *TCC*, pages 273–290, 2007. 59, 62

[MMRV02] K. Makarychev, Y. Makarychev, A. Romashchenko, and N. Vereshchagin. A new class of non-shannon-type inequalities for entropies. *Communications in Information and Systems*, 2(2):147–166, 2002. 34, 35, 98

[MS95] František Matúš and Milan Studený. Conditional independences among four random variables i. *Combinatorics, Probability & Computing*, 4:269–278, 1995. 35

[Muc02] Andrej A. Muchnik. Conditional complexity and codes. *Theor. Comput. Sci.*, 271(1-2):97–109, January 2002. 75

[NW01] Siaw-Lynn Ng and Michael Walker. On the composition of matroids and ideal secret sharing schemes. *Des. Codes Cryptography*, 24(1):49–67, August 2001. 54

[Oxl92] J. G. Oxley. *Matroid Theory*. Oxford University Press, New York, 1992. 56

[Pad12] Carles Padró. Lecture notes in secret sharing. Cryptology ePrint Archive, Report 2012/674, 2012. http://eprint.iacr.org/. 64

[Pip86] Nicholas Pippenger. What are the laws of information theory. *Special Problems on Communication and Computation Conference, Palo Alto, California*, 1986. 15

[PPD] Rethnakaran Pulikkoonattu, Etienne Perron, and Suhas Diggavi. Information theoretic Inequality prover (Xitip). 21

[PVY12] Carles Padró, Leonor Vázquez, and An Yang. Finding lower bounds on the complexity of secret sharing schemes by linear programming. *IACR Cryptology ePrint Archive*, 2012:464, 2012. 64

[PZ03] Josef Pieprzyk and Xian-Mo Zhang. Ideal threshold schemes from mds codes. In *Proceedings of the 5th international conference on Information security and cryptology*, ICISC'02, pages 253–263, Berlin, Heidelberg, 2003. Springer-Verlag. 64

[Rom00a] Andrei Romashchenko. *Inequalities for Kolmogorov complexity and common information*. PhD thesis, Moscow state university, 2000. (In Russian). 74

[Rom00b] Andrei Romashchenko. Pairs of Words with Nonmaterializable Mutual Information. *Problems of Information Transmission*, 36(1):1–18, 2000. 92, 115

[Rom03a] A. Romashchenko. Extracting the mutual information for a triple of binary strings. In *Computational Complexity, 2003. Proceedings. 18th IEEE Annual Conference on*, pages 221–229. IEEE, 2003. 17

[Rom03b] A. Romashchenko. Extracting the mutual information for a triple of binary strings. In *Computational Complexity, 2003. Proceedings. 18th IEEE Annual Conference on*, pages 221–229. IEEE, 2003. 75

[RW05] Frank Ruskey and Mark Weston. A Survey of Venn Diagrams, 2005. http://www.combinatorics.org/files/Surveys/ds5/VennEJC.html. 19

[SA98] Juriaan Simonis and Alexei Ashikhmin. Almost Affine Codes. *Des. Codes Cryptography*, 14(2):179–197, 1998. 60

[Sey76] P. D. Seymour. A forbidden Minor Characterization of Matroid Ports. *The Quarterly Journal of Mathematics*, 27(4):407–413, 1976. 57, 59

[Sey92] P. D. Seymour. On secret-sharing matroids. *J. Comb. Theory Ser. B*, 56(1):69–73, September 1992. 60, 61

[Sha48a] Claude E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, July 1948. 10, 13

[Sha48b] Claude E. Shannon. A mathematical theory of communication (continued). *Bell System Technical Journal*, 27(4):623–656, October 1948. 10

[Sha79] Adi Shamir. How to Share a Secret. *Communications of the ACM*, 22(11):612–613, 1979. 42

[Sim90] G. J. Simmons. How to (really) share a secret. In *Proceedings on Advances in cryptology*, CRYPTO '88, pages 390–448, New York, NY, USA, 1990. Springer-Verlag New York, Inc. 64

[Spe28] Emanuel Sperner. Ein Satz über Untermengen einer Endlichen Menge. *Mathematische Zeitschrift*, 27(1):544–548, December 1928. 39

[Sti92] Douglas R. Stinson. An Explication of Secret Sharing Schemes. *Des. Codes Cryptography*, 2(4):357–390, 1992. 49, 64

[Sti94] Douglas R. Stinson. Decomposition constructions for secret sharing schemes. *IEEE Transactions on Information Theory*, 40(2):118–125, 1994. 48, 49, 50

[Stu01] Milan. Studený. *On mathematical description of probabilistic conditional independence structures*. PhD thesis, Institute of Information Thoery and Automation, Academy of Sciences of the Czech Republic, Prague, 2001. 35

[SW06] D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Trans. Inf. Theor.*, 19(4):471–480, September 2006. 67

[Tar00] G. Tarry. *Le problème des 36 officiers*. Association Française pour l'avancement des sciences fusionnée avec l'association scientifique de France, 1900. 45

[VD95] Marten Van Dijk. On the information rate of perfect secret sharing schemes. *Designs, Codes and Cryptography*, 6:143–169, 1995. 10.1007/BF01398012. 41, 49

[vDKST06] Marten van Dijk, Tom Kevenaar, Geert-Jan Schrijen, and Pim Tuyls. Improved constructions of secret sharing schemes by applying (λ, ω)-decompositions. *Inf. Process. Lett.*, 99(4):154–157, August 2006. 49

[Wel76] D.J.A Welsh. *Matroid Theory*. Academic Press, 1976. 56, 57

[Wyn06] A. Wyner. On source coding with side information at the decoder. *IEEE Trans. Inf. Theor.*, 21(3):294–300, September 2006. 34, 68

[YC12] Raymond W. Yeung and Qi Chen. Characterizing the Entropy Function Region via Extreme Rays. *IEEE Information Theory Workshop*, 2012. 28

[Yeu08] Raymond W. Yeung. *Information Theory and Network Coding*. Springer-Verlag, second edition, 2008. 5, 24

[YY] R. W. Yeung and Y. O. Yan. Information Theoretic Inequality Prover (ITIP). 21

[ZL70] A. K. Zvonkin and L. A. Levin. The Complexity of Finite Objects and the Development of the Concepts of Information and Randomness by means of the Theory of Algorithms. *Russian Math. Surveys*, page 11, 1970. 73, 94

[ZY97] Zhen Zhang and Raymond W. Yeung. A non-Shannon-type Conditional Information Inequality. *IEEE Trans. on Inform. Theory*, 43:1982–1986, 1997. 24, 29, 67, 98

[ZY98] Zhen Zhang and R. W. Yeung. On characterization of entropy function via information inequalities. *IEEE Trans. Inf. Theor.*, 44(4):1440–1452, July 1998. 24, 28, 33