ON EQUIVALENCE OF ADDITIVE-COMBINATORIAL INEQUALITIES FOR SHANNON ENTROPY AND DIFFERENTIAL ENTROPY

BY

ASHOK VARDHAN MAKKUVA

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Electrical and Computer Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Advisers:

Professor Pramod Viswanath
Assistant Professor Yihong Wu, Yale University

ABSTRACT

Entropy inequalities are very important in information theory and they play a crucial role in various communication-theoretic problems, for example, in the study of the degrees of freedom of interference channels. In this thesis, we are concerned with the additive-combinatorial entropy inequalities which are motivated by their combinatorial counterparts: cardinality inequalities for subsets of abelian groups. As opposed to the existing approaches in the literature in the study of the discrete and continuous entropy inequalities, we consider a general framework of balanced linear entropy inequalities. In particular, we show a fundamental equivalence relationship between these classes of discrete and continuous entropy inequalities. In other words, we show that a balanced Shannon entropy inequality holds if and only if the corresponding differential entropy inequality holds. We also investigate these findings in a more general setting of connected abelian Lie groups and in the study of the sharp constants for entropy inequalities.

To my parents and friends, for their love and support.

ACKNOWLEDGMENTS

I would first like to thank my thesis advisors Prof. Yihong Wu and Prof. Pramod Viswanath for their constant support and guidance. I am greatly indebted to them for providing me an opportunity to do first-order research with them.

Working with Yihong has helped me realize and inculcate many crucial aspects of high quality research: intuition, rigor, clarity of thought, and the ability to prove things, no matter how complicated, by oneself. His ability to get to the core of the problem and digest complex ideas through intuition is one thing that I can only aspire for but which I try to imbibe into my research. I greatly benefited from his course on information-theoretic methods for high-dimensional statistics which introduced me to the beautiful theories of statistical estimation, information theory and their connections. I am also thankful to him for providing me an opportunity to travel to ISIT 2016 in Barcelona in my very first year, which helped me meet and interact with some stalwarts of information theory.

Pramod has helped me appreciate the importance of intuition and clarity when trying to understand difficult problems and ideas. His constant guidance and care has helped me to stay motivated and focus on research throughout. His invaluable advice on various aspects of research – creating a problem statement and traversing the unexplored rugged research terrain with a torch called "your own co-ordinate system," identifying the right set of problems to work on, etc. – has greatly shaped my thought process and the perspective to view and understand things.

I would like to thank Prof. Venu Veeravalli, Prof. Pierre Moulin and Prof. Idoia Ochoa for serving as my qualifying exam committee. I would not be able to pursue my Ph.D. without their help. I would also like to thank all the faculty who taught me during the past two years, including Prof. Richard Laugesen, Prof. Maxim Raginsky, Prof. Zhongjin Ruan, Prof. Sewoong Oh,

Prof. Pierre Moulin, Prof. Matus Telgarsky, Prof. Ruoyu Sun, and Prof. Yihong Wu.

I have had a wonderful experience in the last two years, thanks to my amazing friends: Azin Heidarshenas, Aakash Modi, Ishan Deshpande, Harsh Gupta, Konik, Siddartha, Pratik Deogekar, Shantanu Shahane, Surya, Sukanya Patil, Soham Phade, Shreyash, Shailesh, Sai Kiran, Tarek, Vinay Iyer, and Weihao. I would like to thank Anirudh Udupa and Aditi Udupa for the innumerable discussions, after which I always gained something new, about philosophy, life and almost everything which we've had over the years right from my undergrad days. I would also like to thank a close friend, outside the convex hull of UIUC friends, Priya Soundararajan, for putting up with my endless curiosity to talk about anything related to maths, history, and movies.

In the tradition of saving the best for the last, I would like to express deepest gratitude for my parents. I am eternally thankful to them for their self-less love, sacrifice, support and encouragement all these years. To them I dedicate this thesis.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Additive-combinatorial inequalities for cardinalities

In mathematics, the field of additive combinatorics is about the combinatorial estimates associated with arithmetic operations, especially addition and subtraction, on sets that are arbitrary subsets of integers, or more generally, any discrete abelian group. It has invited a lot of exciting mathematical activity in recent years, partially thanks to some landmark results such as Szemerèdi's theorem and Green-Tao's theorem, etc. An important repository of tools in additive combinatorics is the sumset inequalities and the inverse sumset theory. Sumset inequalities are concerned with relating the cardinalities of the sumset and the difference set $A \pm B = \{a \pm b : a \in A, b \in B\}$ to those of A and B, where A and B are arbitrary subsets of an abelian group. A simple example of such an inequality is the following: Given any three sets A, B and C, that are subsets of a discrete abelian group, the *Ruzsa triangle inequality* [1] states that

$$|A - C||B| \leq |A - B||B - C|, \tag{1.1}$$

where $|S|$ denotes the cardinality of a set S. In inverse sumset theory, we seek to characterize and find structure in sets A under some constraints on the cardinalities of the sumset $A + A$ or the difference set $A - A$.

1.2 Extensions to Shannon entropy

While the field of additive-combinatorics by its very nature is host to a wide variety of techniques and tools from number theory to combinatorics, harmonic analysis and ergodic theory, which might be of independent interest

to mathematicians, what is important from an information-theoretic point-of-view is that some interesting analogies can be drawn between these additive-combinatorial inequalities and entropy inequalities.

This interesting connection between additive-combinatorial inequalities about cardinalities of sumsets and the entropy inequalities for sums of random variables was identified by Tao and Vu [2] and Ruzsa [3]. The idea is that one can consider the information-theoretic analogs of additive combinatorial inequalities by replacing the sets by (independent, discrete, group-valued) random variables and, correspondingly, the log-cardinality by the Shannon entropy, defined as

$$H\left(X\right) \triangleq \sum_{x} \mathbb{P}\left[X = x\right] \log \frac{1}{\mathbb{P}\left[X = x\right]}.$$

For example, the inequality

$$\max\{|A|, |B|\} \leq |A + B| \leq |A||B|$$

translates to

$$\max\left\{H\left(X\right), H\left(Y\right)\right\} \leq H\left(X + Y\right) \leq H\left(X\right) + H\left(Y\right), \tag{1.2}$$

which follows from elementary properties of entropy, whereas (1.1) becomes

$$H(X - Z) + H(Y) \leq H(X - Y) + H(Y - Z) \tag{1.3}$$

for independent X, Y, Z which follows from the *submodularity* of Shannon entropy. The motivation for this analogy comes from the well-known asymptotic equipartition property (AEP) [4] for discrete random variables: A random vector $X^n = (X_1, \ldots, X_n)$ consisting of n independent and identical copies of X is concentrated on a set of cardinality $\exp(n(H(X) + o(1))$ as $n \to \infty$. This interpretation has indeed been useful in deducing certain entropy inequalities, e.g., Han's inequality [5], directly from their set counterpart, cf. [3, p. 5]. A similar approach might not be useful in general for inequalities dealing with sums of random variables since the typical set of sums can be exponentially larger than sums of individual typical sets, which in turn necessitates a new set of tools when dealing with discrete entropy inequalities.

It turns out that the functional submodularity property of Shannon en-

tropy serves this role in proving the discrete entropy inequalities. The submodularity property of Shannon entropy states that for any two random variables X_1 and X_2,

$$H(X_0) + H(X_{1,2}) \leq H(X_1) + H(X_2)$$

for $X_0 = f(X_1) = g(X_2)$ and $X_{1,2} = h(X_1, X_2)$ where f, g and h are deterministic functions. Capitalizing on this submodularity of entropy, in the past few years several entropy inequalities for sums and differences have been obtained [2, 6, 7, 8, 9, 10], such as the sum-difference inequality [8, Eq. (2.2)]

$$H(X + Y) \leq 3H(X - Y) - H(X) - H(Y), \tag{1.4}$$

which parallels the following (cf., e.g., [11, Eq. (4)]):

$$|A + B| \leq \frac{|A - B|^3}{|A||B|}.$$

1.3 Extensions to differential entropy

Recall that the *differential entropy* of a real-valued random vector X with probability density function (pdf) f_X is defined as

$$h(X) = \int f_X(x) \log \frac{1}{f_X(x)} dx.$$

Similar to the discrete case, in the sense of AEP, $h(X)$ can be viewed as the log-volume of the effective support of X [4]. In a similar fashion, one can consider additive-combinatorial inequalities for differential entropies on Euclidean spaces. Given this interpretation it is natural to ask: To what extent do Shannon entropy inequalities extend to differential entropy? Kontoyiannis and Madiman [12] and Madiman and Kontoyiannis [9, 13] made important progress in this direction by showing that while the submodularity property, the key ingredient for proving discrete entropy inequalities, fails for differential entropy, several linear inequalities for Shannon entropy nevertheless extend *verbatim* to differential entropy albeit using a different proof strategy: the data processing inequality; for example, the sum-difference in-

equality (1.4) admits an exact continuous analog [12, Theorem 3.7]:

$$h(X+Y) \leq 3h(X-Y) - h(X) - h(Y). \tag{1.5}$$

1.4 A general framework for entropy inequalities

Whereas a significant number of inequalities for Shannon entropy extend verbatim to the continuous case, the proof techniques for the continuous entropy inequalities are not generalizable since the vital ingredient for proving the continuous entropy inequalities, the data processing inequality, has been applied in a case-by-case manner and a principled way of extending the inequalities is missing in the existing literature. Forgoing this approach we consider a more general framework of linear entropy inequalities of the form:

$$\sum_{i=1}^{n} \alpha_i H\left(\sum_{j=1}^{m} a_{ij} Z_j\right) \leq 0, \tag{1.6}$$

and

$$\sum_{i=1}^{n} \alpha_i h\left(\sum_{j=1}^{m} a_{ij} X_j\right) \leq 0, \tag{1.7}$$

where $a_{ij} \in \mathbb{Z}$, $\alpha_i \in \mathbb{R}$, $Z_1, \ldots, Z_m$ are independent discrete group-valued random variables and X_j's are independent $\mathbb{R}^d$-valued continuous random variables. Notice that all of the aforementioned results for Shannon entropy and differential entropy can be cast into this generic form of (1.6) and (1.7) respectively.

Motivated by the absence of a unified and principled framework for the extension of these entropy inequalities from discrete to continuous case (and vice versa) and to understand the fundamental relationship between these two classes of inequalities, we ask the following question, which is the focus point of this thesis:

Question 1. Do all linear inequalities of the form (1.6) for discrete entropy extend to differential entropy inequalities of the form (1.7), and vice versa?

1.5 Main contribution

Towards addressing Question 1, a simple but instructive observation reveals that all linear inequalities for differential entropies are always *balanced*; that is, the sum of all coefficients must be zero. In other words, should (1.7) hold for all independent $\mathbb{R}^d$-valued Z_j's, then we must have $\sum_{i=1}^n \alpha_i = 0$. To see this, recall the fact that $h(aZ) = h(Z) + d \log a$ for any $a > 0$; in contrast, Shannon entropy is scale-invariant. Therefore, whenever the inequality (1.7) is unbalanced, i.e., $\sum_{i=1}^n \alpha_i \neq 0$, scaling all random variables by a and sending a to either zero or infinity leads to a contradiction. For instance, in (1.2), the left inequality (balanced) extends to differential entropy but the right inequality (unbalanced) clearly does not.

With this observation in mind, we establish the equivalence of balanced linear Shannon and differential entropy inequalities in Chapter 2, that is, a balanced linear inequality holds for Shannon entropy if and only if it holds for differential entropy, thereby fully resolving Question 1. This result, in a way, demystifies the striking parallel between discrete and continuous entropy inequalities, thereby bridging the gap between these two notions of entropy inequalities. As a consequence, it follows that the results in [12, 13], which are linear inequalities for mutual information such as $I(X; X+Y) = h(X+Y) - h(Y)$ or Ruzsa distance $\mathrm{dist}_R(X, Y) \triangleq h(X-Y) - \frac{1}{2}h(X) - \frac{1}{2}h(Y)$ [3, 8, 12]) and hence expressible as balanced linear inequalities for differential entropy, can be deduced from their discrete counterparts [8] in a unified manner.

Our result that the differential entropy inequalities follow from their discrete analogs relies on a result due to to Rényi [14] which gives the asymptotic expansion of the Shannon entropy of a finely quantized continuous random variable in terms of its differential entropy, namely,

$$H(\lfloor mX \rfloor) = d \log m + h(X) + o(1), \quad m \to \infty \tag{1.8}$$

for continuous $\mathbb{R}^d$-valued X. Our strategy is to utilize this result in approximating differential entropy inequalities by their discrete counterparts via this quantization, which enables the discrete entropy inequalities to carry over exactly to their continuous analogs, and, even more generally, for connected abelian Lie groups as we prove in Chapter 2. In establishing that all linear discrete entropy inequalities follow from their continuous analogs, the

following are the two key ideas of our approach: First we show that given any finite collection of discrete $\mathbb{R}^d$-valued random variables, we can embed them into a high-dimensional Euclidean space and project them back to $\mathbb{R}^d$ such that the Shannon entropy of any linear combination of the projected random variables is equal to an arbitrarily large multiple of the given random variables. Next we add independent noise, e.g., Gaussian, with arbitrarily small variance to these projected discrete random variables and relate their Shannon entropy to the differential entropy of their noisy versions. Sending the variance to zero and then the dimension to infinity yields the desired inequality for discrete entropy.

Chapter 3 explores the implications of our equivalence result in the context of the sharp constants for additive-combinatorial entropy inequalities. In particular, we establish that the sharp constants for the *sum-difference ratio*, $\frac{h(X-X')-h(X)}{h(X+X')-h(X)}$ for i.i.d. continuous $X, X' \in \mathbb{R}^n$, and $\frac{H(U-U')-H(U)}{H(U+U')-H(U)}$ for i.i.d. discrete $U, U' \in \mathbb{Z}^n$, are equal and, moreover, dimension independent. Whereas the equality of sharp constants for the discrete and continuous versions follows directly from our equivalence result in Chapter 2, the dimension-freeness of the ratio relies on techniques similar to those used for establishing that the continuous entropy inequalities carry over exactly to their discrete versions.

CHAPTER 2

EQUIVALENCE OF SHANNON ENTROPY AND DIFFERENTIAL ENTROPY INEQUALITIES

In this chapter we present the equivalence between the discrete and continuous entropy inequalities. This equivalence completely bridges the gap between these two classes of inequalities in the existing literature. Section 2.3 contains our main equivalence result. We also present the extensions of the result to a more general setting of connected abelian Lie groups. The proofs of the theorems are deferred to Section 2.4, and the key technical lemmas to prove the theorems are proved in Appendix A.1. The material in this chapter has appeared in part in [15].

2.1 Notation

We use the following notations throughout this chapter. For $x \in \mathbb{R}$, let $\lfloor x \rfloor \triangleq \max\{k \in \mathbb{Z} : k \leq x\}$ and $\{x\} = x - \lfloor x \rfloor$ denote its integer and fractional parts, respectively. For any $k \in \mathbb{N}$, define

$$[x]_k \triangleq \frac{\lfloor 2^k x \rfloor}{2^k}, \quad \{x\}_k \triangleq \frac{\{2^k x\}}{2^k}. \tag{2.1}$$

Hence,

$$x = \frac{\lfloor 2^k x \rfloor}{2^k} + \frac{\{2^k x\}}{2^k} = [x]_k + \{x\}_k.$$

For $x \in \mathbb{R}^d$, $[x]_k$ and $\{x\}_k$ are defined similarly by applying the above operations componentwise.

For $N > 0$, denote the hypercube $B_N^{(d)} \triangleq [-N, N]^d$. For a $\mathbb{R}^d$-valued random variable X, let $X^{(N)}$ denote a random variable distributed according to the conditional distribution $P_{X|X \in B_N^{(d)}}$. If X has a pdf f_X, then $X^{(N)}$ has

the following pdf:

$$f_{X^{(N)}}(x) = \frac{f_X(x)\mathbb{1}\{x \in B_N^{(d)}\}}{\mathbb{P}[X \in B_N^{(d)}]}. \tag{2.2}$$

For the sake of concision and excluding trivialities, we assume throughout this thesis that the differential entropies exist and are finite.

2.2 Related work

Before we present the main results of this chapter, we would like to point out that a similar study of establishing the equivalence of linear information inequalities, for *subsets* of random variables, has been done by Chan in [16]. In particular, he established that the class of linear inequalities for Shannon entropy and differential entropy are equivalent provided the inequalities are "balanced" in the following sense. For example, consider the following entropy inequalities for discrete random variables X_1 and X_2:

$$H(X_1) + H(X_2) - H(X_1, X_2) \geq 0, \tag{2.3}$$

$$H(X_1, X_2) - H(X_1) \geq 0. \tag{2.4}$$

The inequality (2.3) is said to be balanced because the sum of the coefficients of the entropy terms in which X_1 appears equal zero and the same is true for X_2 as well. However, the inequality (2.4) is unbalanced because X_2 appears only in the first term. It is worth pointing out that the above notion of balancedness is different from ours, which demands that the sum of the coefficients of all the entropy terms, rather than each individual random variable, be zero. However, the technique employed for extending the discrete entropy inequalities to the continuous case is similar to ours, i.e., through discretization of continuous random variables. In establishing the converse that the discrete inequality follows from its continuous counterpart the method in [16] is to assume, without loss of generality, the discrete random variables are integer-valued and use the fact that $H(A) = h(A + U)$ for any $\mathbb{Z}$-valued A and U independently and uniformly distributed on $[0, 1]$. Clearly this method does not apply to sums of independent random variables.

2.3 Main results

The following are our main results on linear entropy inequalities.

Theorem 1. *Let $(a_{ij}) \in \mathbb{Z}^{n\times m}$ satisfy that $a_{i1}, \ldots, a_{im}$ are relatively prime, for each $i = 1, \ldots, n$. Let $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ be such that $\sum_{i=1}^n \alpha_i = 0$. Suppose for any independent $\mathbb{Z}^d$-valued random variables $U_1, \ldots, U_m$, the following holds:*

$$\sum_{i=1}^{n} \alpha_i H\left(\sum_{j=1}^{m} a_{ij} U_j\right) \leq 0. \tag{2.5}$$

Then for any independent $\mathbb{R}^d$-valued continuous random variables $X_1, \ldots, X_m$, the following holds:

$$\sum_{i=1}^{n} \alpha_i h\left(\sum_{j=1}^{m} a_{ij} X_j\right) \leq 0 \tag{2.6}$$

Remark 1. Without loss of any generality, we can always assume that the coefficients of each linear combination of random variables in (2.5) are relatively prime. This is because for each i we can divide $a_{i1}, \ldots, a_{im}$ by their greatest common divisor so that the resulting entropy inequality remains the same, thanks to the scale invariance of the Shannon entropy.

Theorem 2. *Let $(a_{ij}) \in \mathbb{R}^{n\times m}$ and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$ be such that $\sum_{i=1}^n \alpha_i = 0$. If*

$$\sum_{i=1}^{n} \alpha_i h\left(\sum_{j=1}^{m} a_{ij} X_j\right) \leq 0$$

holds for any $\mathbb{R}^d$-valued independent and continuous random variables $X_1, \ldots, X_m$, then

$$\sum_{i=1}^{n} \alpha_i H\left(\sum_{j=1}^{m} a_{ij} U_j\right) \leq 0$$

holds for any $\mathbb{R}^d$-valued independent and discrete random variables $U_1, \ldots, U_m$.

We now consider the extensions of the above theorems to i.i.d. random variables and to general groups and present their proofs in Section 2.4.

2.3.1 Extensions to i.i.d. random variables

When we further constrain the random variables in additive-combinatorial entropy inequalities, we can obtain strengthened inequalities. For instance, if U and U' are independent and identically distributed (i.i.d.) discrete random variables, then (cf., e.g., [9, Theorems 1.1 and 2.1])

$$\frac{1}{2} \leq \frac{H(U-U') - H(U)}{H(U+U') - H(U)} \leq 2 \tag{2.7}$$

and for i.i.d. continuous X, X',

$$\frac{1}{2} \leq \frac{h(X-X') - h(X)}{h(X+X') - h(X)} \leq 2, \tag{2.8}$$

which are stronger than what would be obtained from (1.4) and (1.5) by substituting $Y = X'$. We refer to these inequalities as *doubling-difference* inequalities which will be the focus of the study in the next chapter.

As evident from the proof, both Theorem 1 and Theorem 2 apply verbatim to entropy inequalities involving independent random variables with arbitrary distributions. Consequently, (2.7) and (2.8) are in fact equivalent. Formally, fix a partition $S_1, \ldots, S_K$ of $[m] \triangleq \{1, \ldots, m\}$. Then (2.5) holds for independent $U_1, \ldots, U_m$ so that $\{U_j\}_{j \in S_k}$ are i.i.d. for $k \in [K]$ if and only if (2.6) holds for independent $X_1, \ldots, X_m$ so that $\{X_j\}_{j \in S_k}$ are i.i.d. for $k \in [K]$. It is worth noting that this result is not a special case of Theorems 1 and 2; nevertheless, the proofs are identical.

2.3.2 Extensions to general groups

We now consider a more general version of Theorem 1. The notion of differential entropy, consequently entropy inequalities, can be extended to a general setting. To this end, let G be a locally compact abelian group equipped with a Haar measure μ. Let X be a G-valued random variable whose distribution is absolutely continuous with respect to μ. Following [13], we define the differential entropy of X as:

$$h(X) = \int f \log \frac{1}{f} d\mu = \mathbb{E}\left[\log \frac{1}{f(X)}\right],$$

where f denotes the pdf of X with respect to μ. This definition subsumes the notion of both the Shannon entropy on $\mathbb{Z}^d$ (with μ being the counting measure) and the differential entropy on $\mathbb{R}^d$ (with μ being the Lebesgue measure).

We now state a generalization of Theorem 1, which holds for connected abelian Lie groups.

Theorem 3. *Under the assumptions of Theorem 1, suppose* (2.5) *holds for any independent random variables $Z_1, \dots, Z_m$ taking values in $\mathbb{Z}^d \times (\mathbb{Z}/2^k\mathbb{Z})^n$ for any $k, d, n \in \mathbb{N}$. Then* (2.6) *holds for any connected abelian Lie group G' and independent G'-valued random variables $X_1, \dots, X_m$.*

Denoting the unit circle in $\mathbb{C}$ by $\mathbb{T}$, we first prove a special case of Theorem 3 when G is a finite cyclic group and G' is the torus $\mathbb{T}^d$. Theorem 3 then follows easily since any connected abelian Lie group is isomorphic to product of torus and Euclidean space. We need the following preliminary fact relating the Haar measures and differential entropies of random variables taking values on isomorphic groups.

Lemma 1. *Let $\phi : G' \to G$ be a group isomorphism between abelian topological groups $(G, +)$ and $(G', +)$ and μ' be a Haar measure on G'. Then the pushforward measure*[1] *$\mu = \phi_* \mu'$ is a Haar measure on G. Furthermore, for any G-valued continuous random variable X,*

$$h(X) = h\left(\phi^{-1}(X)\right).$$

Proof. For any measurable subset A of G and any $g \in G$, then

$$\mu(g + A) = \mu'(\phi^{-1}(g + A)) = \mu'(\phi^{-1}(g) + \phi^{-1}(A)) = \mu'(\phi^{-1}(A)) = \mu(A),$$

which follows the translation invariance of μ'. Similarly, using the fact that ϕ^{-1} is a homeomorphism one can verify that μ is finite on all compacts as well as its inner and outer regularity.

If f is the density function of X with respect to the Haar measure $\phi_* \mu'$ on G, then $f \circ \phi$ is the pdf of $\phi^{-1}(X)$ with respect to the Haar measure μ' on

[1] That is, $(\phi_* \mu')(B) = \mu'(\phi^{-1}(B))$ for any measurable subset B of G.

G'. Hence,

$$\begin{aligned} h(X) &= \int f \log \frac{1}{f} d(\phi_* \mu') \\ &= \int f \circ \phi \log \frac{1}{f \circ \phi} d\mu \\ &= h\left(\phi^{-1}(X)\right). \end{aligned}$$

□

Define the map $\phi : [0,1)^n \to \mathbb{T}^n$ by $\phi(\theta_1, \ldots, \theta_n) = (e^{2\pi i \theta_1}, \ldots, e^{2\pi i \theta_n})$. Let the Haar measure on $\mathbb{T}^n$ be the pushforward of Lebesgue measure under ϕ. For $X \in \mathbb{T}^n$, let $\Theta = \phi^{-1}(X)$. Define the quantization operation of X in terms of the angles

$$[X]_k \triangleq \phi\left(\frac{\lfloor 2^k \Theta \rfloor}{2^k}\right), \quad [\Theta]_k = \frac{\lfloor 2^k \Theta \rfloor}{2^k}. \tag{2.9}$$

Since ϕ is a bijection, $H([X]_k) = H\left(\lfloor 2^k \Theta \rfloor\right)$. We are now ready to prove Theorem 4.

Theorem 4. *Under the assumptions of Theorem 1, suppose* (2.5) *holds for any cyclic group (G)-valued independent random variables $Z_1, \ldots, Z_m$. Then* (2.6) *holds for any $\mathbb{T}^n$-valued independent random variables $X_1, \ldots, X_m$.*

Proof. Let $X_1, \ldots, X_m$ be $\mathbb{T}^n$-valued continuous independent random variables. For each $i \in [m]$, let $\Theta_i = \phi^{-1}(X_i)$. Since $\lfloor 2^k \Theta_i \rfloor$ is $\mathbb{Z}_{2^k}$-valued and $\mathbb{Z}_{2^k}$ is a cyclic group under modulo 2^k addition, to prove Theorem 4, it suffices to prove the following:

$$H([X]_k) = kn \log 2 + h(X) + o_k(1) \tag{2.10}$$

for any $\mathbb{T}^n$-valued continuous random variable X, and

$$H\left(\left[\sum_{i=1}^m a_i X_i\right]_k\right) = H\left(\sum_{i=1}^m a_i [X_i]_k\right) + o_k(1). \tag{2.11}$$

Indeed, (2.10) follows from

$$H([X]_k) = H([\Theta]_k) \overset{(a)}{=} kn \log 2 + h(\Theta) + o_k(1) \overset{(b)}{=} kn \log 2 + h(X) + o_k(1),$$

where (a) is by Lemma 4 since Θ is a continuous $[0,1]$-valued random variable

and (b) is by Lemma 1. To prove (2.11), for each $i \in [m]$, let $\Theta_i = \phi^{-1}(X_i)$. Define

$$A_k \triangleq \left\lfloor 2^k \sum_{i=1}^m a_i \Theta_i \right\rfloor \pmod{2^k}, A'_k = \left\lfloor 2^k \sum_{i=1}^m a_i \Theta_i \right\rfloor,$$
$$B_k \triangleq \sum_{i=1}^m a_i \left\lfloor 2^k \Theta_i \right\rfloor \pmod{2^k}, B'_k = \sum_{i=1}^m a_i \left\lfloor 2^k \Theta_i \right\rfloor,$$
$$Z_k \triangleq \left\lfloor \sum_{i=1}^m a_i \left\{ 2^k \Theta_i \right\} \right\rfloor.$$

Our aim is to prove that $H(A_k) - H(B_k) = o_k(1)$. Since $A'_k = B'_k + Z_k$, $A_k = B_k + Z_k \pmod{2^k}$. Also, $H(A_k) - H(B_k) = I(Z_k; A_k) - I(Z_k; B_k)$. Hence,

$$|H(A_k) - H(B_k)| \leq I(Z_k; A_k) + I(Z_k; B_k) \overset{(a)}{\leq} I(Z_k; A'_k) + I(Z_k; B'_k) \overset{(b)}{\to} 0 \text{ as } k \to \infty,$$

where (a) follows from the data processing inequality and (b) follows from Lemma 10 and Lemma 11. This completes the proof. □

2.4 Proofs

In this section we prove Theorem 1 and Theorem 2. We included the lemmas that are crucial for these theorems before their proofs. The proofs of the lemmas are provided in Appendix A.1.

2.4.1 Proof of Theorem 1

The following lemma lies at the core of the proof of Theorem 1.

Lemma 2. *Let $X_1, \ldots, X_m$ be independent $[0,1]^d$-valued continuous random variables such that both $h(X_j)$ and $H(\lfloor X_j \rfloor)$ are finite for each $j \in [m]$. Then for any $a_1, \ldots, a_m \in \mathbb{Z}$ that are relatively prime,*

$$\lim_{k \to \infty} \left(H\left(\left[\sum_{i=1}^m a_i X_i \right]_k \right) - H\left(\sum_{i=1}^m a_i \left[X_i \right]_k \right) \right) = 0.$$

The next lemma allows us to restrict our focus on bounded random variables.

Lemma 3 (Truncation)**.** *Let $X_1, \ldots, X_m$ be independent $\mathbb{R}^d$-valued random variables and $a_1, \ldots, a_m \in \mathbb{R}$. If each X_j has an absolutely continuous distribution and $h(X_j)$ is finite, then*

$$\lim_{N\to\infty} h\left(\sum_{j=1}^{m} a_j X_j^{(N)}\right) = h\left(\sum_{j=1}^{m} a_j X_j\right).$$

The following lemma is a particularization of [14, Theorem 1] (see (1.8)) to the dyadic subsequence $m = 2^k$:

Lemma 4. *For any $\mathbb{R}^d$-valued random variable X with an absolutely continuous distribution such that both $H(\lfloor X \rfloor)$ and $h(X)$ are finite,*

$$\lim_{k\to\infty} \left(H\left([X]_k\right) - dk\log 2\right) = h(X).$$

We now present the proof of Theorem 1.

Proof. We start by considering the case where $X_j \in [0,1]^d$ for each $j \in [m]$. Since X_j's are independent and $2^k [X_j]_k$ is $\mathbb{Z}^d$-valued for each $j \in [m]$, by assumption,

$$\sum_{i=1}^{n} \alpha_i H\left(\sum_{j=1}^{m} a_{ij} [X_j]_k\right) \leq 0 \tag{2.12}$$

holds where

$$\sum_{i=1}^{n} \alpha_i = 0. \tag{2.13}$$

By Lemma 4, $H\left([X]_k\right) = dk\log 2 + h(X) + o_k(1)$. Thus,

$$\begin{aligned} h\left(\sum_{j=1}^{m} a_{ij} X_j\right) + dk\log 2 + o_k(1) &= H\left(\left[\sum_{j=1}^{m} a_{ij} X_j\right]_k\right) \\ &\overset{(a)}{=} H\left(\sum_{j=1}^{m} a_{ij} [X_j]_k\right) + o_k(1), \end{aligned}$$

where (a) follows from Lemma 2. Multiplying on both sides by α_i and summing over i, and in view of (2.13), we have

$$\sum_{i=1}^{n} \alpha_i h\left(\sum_{j=1}^{m} a_{ij} X_j\right) + o_k(1) = \sum_{i=1}^{n} \alpha_i H\left(\sum_{j=1}^{m} a_{ij} \left[X_j\right]_k\right).$$

By (2.12), sending k to infinity yields the desired result.

For the general case where $X_j \in \mathbb{R}^d$, let $Y_i = \sum_{j=1}^{m} a_{ij} X_j$ for $i \in [n]$. Let $\tilde{X}_j^{(N)} \triangleq \frac{X_j^{(N)}+N}{2N}$, which belongs to $[0,1]^d$. Thus,

$$\begin{aligned}\sum_{i=1}^{n} \alpha_i h\left(\sum_{j=1}^{m} a_{ij} \tilde{X}_j^{(N)}\right) &= \sum_{i=1}^{n} \alpha_i h\left(\sum_{j=1}^{m} a_{ij} X_j^{(N)}\right) + \sum_{i=1}^{n} \alpha_i \cdot \log\left(\frac{1}{2N}\right)^d \\ &= \sum_{i=1}^{n} \alpha_i h\left(\sum_{j=1}^{m} a_{ij} X_j^{(N)}\right), \qquad (2.14)\end{aligned}$$

where (2.14) follows from (2.13). Hence,

$$\begin{aligned}\sum_{i=1}^{n} \alpha_i h\left(Y_i\right) &\overset{(a)}{=} \lim_{N\to\infty} \sum_{i=1}^{n} \alpha_i h\left(\sum_{j=1}^{m} a_{ij} X_j^{(N)}\right) \\ &\overset{(b)}{=} \lim_{N\to\infty} \sum_{i=1}^{n} \alpha_i h\left(\sum_{j=1}^{m} a_{ij} \tilde{X}_j^{(N)}\right) \\ &\overset{(c)}{\leq} 0,\end{aligned}$$

where (a) follows form Lemma 3 and (b) follows from (2.14), and (c) follows from the earlier result for $[0,1]^d$-valued random variables. The proof of Theorem 1 is now complete. □

2.4.2 Proof of Theorem 2

Theorem 2 relies on the following two lemmas. The first result, proved in in Section B.3, is a well-known asymptotic expansion of the differential entropy of a discrete random variable contaminated by weak additive noise. The second result, proved in Section B.4, allows us to blow up the Shannon entropy of linear combinations of discrete random variables arbitrarily.

Lemma 5. *Let U be a discrete $\mathbb{R}^d$-valued random variable such that $H(U) <$*

∞ *and* Z *be a* $\mathbb{R}^d$*-valued continuous random variable with* $h(Z) > -\infty$*. If* U *and* Z *are independent, then*

$$h(U + \varepsilon Z) = h(Z) + \log \varepsilon + H(U) + o_\varepsilon(1).$$

Lemma 6. *Let* $U_1, \ldots, U_m$ *be* $\mathbb{R}^d$*-valued discrete random variables. Let* $k \in \mathbb{N}$*. Then for any* $A = (a_{ij}) \in \mathbb{R}^{n \times m}$*, there exist* $\mathbb{R}^d$*-valued discrete random variables* $U_1^{(k)}, \ldots, U_m^{(k)}$ *such that*

$$H\left(\sum_{j=1}^m a_{ij} U_j^{(k)}\right) = kH\left(\sum_{j=1}^m a_{ij} U_j\right), \forall i \in [n].$$

We now prove Theorem 2.

Proof. Let Z_j be independent $\mathbb{R}^d$-valued Gaussian random variables with zero mean and $U_1, \ldots, U_m$ be independent $\mathbb{R}^d$-valued discrete random variables. Let $U_1^{(k)}, \ldots, U_m^{(k)}$ be independent $\mathbb{R}^d$-valued discrete random variables such that $H\left(\sum_{j=1}^m a_{ij} U_j^{(k)}\right) = kH\left(\sum_{j=1}^m a_{ij} U_j\right)$ for each $i \in [n]$, guaranteed by Lemma 6.

Let $\varepsilon > 0$. For each $j \in [m]$, let $X_j = U_j^{(k)} + \varepsilon Z_j$. Then we have

$$h\left(X_j\right) = H(U_j^{(k)}) + h(Z_j) + \log \varepsilon + o_\varepsilon(1).$$

Hence, for each $i \in [n]$,

$$\begin{aligned} h\left(\sum_{j=1}^m a_{ij} X_j\right) &= h\left(\sum_{j=1}^m a_{ij} U_j^{(k)} + \varepsilon \sum_{j=1}^m a_{ij} Z_j\right) \\ &\overset{(a)}{=} H\left(\sum_{j=1}^m a_{ij} U_j^{(k)}\right) + h\left(\sum_{j=1}^m a_{ij} Z_j\right) + \log \varepsilon + o_\varepsilon(1) \\ &= kH\left(\sum_{j=1}^m a_{ij} U_j\right) + h\left(\sum_{j=1}^m a_{ij} Z_j\right) + \log \varepsilon + o_\varepsilon(1), \end{aligned}$$

where (a) follows from Lemma 5. Since X_j's are independent, by assumption, $\sum_{i=1}^n \alpha_i h\left(\sum_{j=1}^m a_{ij} X_j\right) \leq 0$ where $\sum_{i=1}^n \alpha_i$. Hence,

$$k \sum_{i=1}^n \alpha_i H\left(\sum_{j=1}^m a_{ij} U_j\right) + \sum_{i=1}^n \alpha_i h\left(\sum_{j=1}^m a_{ij} Z_j\right) + o_\varepsilon(1) \leq 0.$$

Thus,

$$\sum_{i=1}^{n} \alpha_i H\left(\sum_{j=1}^{m} a_{ij} U_j\right) + \frac{\sum_{i=1}^{n} \alpha_i h\left(\sum_{j=1}^{m} a_{ij} Z_j\right)}{k} + \frac{o_\varepsilon(1)}{k} \leq 0.$$

The proof is completed by letting $\varepsilon \to 0$ followed by $k \to \infty$. $\square$

CHAPTER 3

ON SHARPNESS OF THE ADDITIVE-COMBINATORIAL LINEAR ENTROPY INEQUALITIES

In the previous chapter we have proved the equivalence of the class of discrete entropy and the class of continuous entropy inequalities. In this chapter we will investigate the implications of this result with regards to the sharp constants in certain additive-combinatorial linear entropy inequalities. In fact, our motivation to study the class of linear entropy inequalities arose from our effort to prove that the constant "2" is the best in the information theoretic analog of the following additive-combinatorial inequality proved by Ruzsa [17]: For any finite $A \subset \mathbb{Z}^n$(or any abelian group)

$$\log \frac{|A-A|}{|A|} \leq 2 \log \frac{|A+A|}{|A|}. \tag{3.1}$$

The information theoretic analogs of the inequality (3.1) are the following: If U and U' are independent and identically distributed (i.i.d.) discrete random variables, then (cf., e.g., [9, Theorems 1.1 and 2.1])

$$\frac{1}{2} \leq \frac{H(U-U') - H(U)}{H(U+U') - H(U)} \leq 2 \tag{3.2}$$

and for i.i.d. continuous X, X',

$$\frac{1}{2} \leq \frac{h(X-X') - h(X)}{h(X+X') - h(X)} \leq 2. \tag{3.3}$$

The constant "2" in (3.1) is known to be sharp (see [18] or [19, p. 107]), i.e., there exists a sequence of sets $\{A_n\}$ such that $\frac{\text{LHS}}{\text{RHS}} \to 2$ as $n \to \infty$. It is not known whether the constants $1/2$ and 2 are the best possible in (3.2) and (3.3). However, as we prove in this chapter, it can be established that the sharp constants for both the discrete and the continuous versions are the same, and dimension-free as a consequence of Theorem 1 and Theorem 2. Since the best constants are unknown theoretically, we performed some nu-

merical simulations for various probability distributions to find out the best possible constants, which we present towards the end of this chapter. So far the best constant we found is around "1.4" achieved for both the Poisson and geometric distributions. From hereafter we refer to the inequalities (3.2) and (3.3) as *doubling-difference* inequalities and the ratio in these inequalities as the *sum-difference ratio*.

3.1 Simplex construction for the sharpness of additive-combinatorial doubling-difference inequality

The idea behind proving that the constant "2" is the best possible in (3.1) is to construct a set whose difference is "much larger" than the sum. Since the precise cardinalities of the sumset and the difference set might be difficult to compute and analyze in general, the idea is to construct sets which contain the lattice points inside a convex body and approximate their cardinality by volume of the body. More formally, for any convex body K in $\mathbb{R}^n$, we consider the set A in (3.1) to be its quantized version $[K]_L \triangleq K \cap (\frac{1}{L}\mathbb{Z}^n)$, where $L \in \mathbb{N}$. The sum and difference sets of $[K]_L$ is related to those of K through $[K \pm K]_L = [K]_L \pm [K]_L$. If we fix the dimension n and let $L \to \infty$, it is well-known that the cardinality of $[K]_L$ is related to the volume of K via $|[K]_L| = \mathrm{vol}(K)L^n(1 + o(1))$. Thus,

$$\frac{|[K]_L \pm [K]_L|}{|[K]_L|} = \frac{\mathrm{vol}(K \pm K)}{\mathrm{vol}(K)}(1 + o(1)).$$

A classical result of Rogers and Shephard [20] states that for any convex $K \in \mathbb{R}^n$, $\mathrm{vol}(K - K) \leq \binom{2n}{n}\mathrm{vol}(K)$ with equality if and only if K is a simplex. Since K is convex, $K + K = 2K$ and thus $\mathrm{vol}(K + K) = 2^n\mathrm{vol}(K)$. Now taking K to be the standard simplex $\Delta_n = \left\{x \in \mathbb{R}^n_+ : \sum_{i=1}^n x_i \leq 1\right\}$, we obtain

$$\frac{\log \frac{|[\Delta_n]_L - [\Delta_n]_L|}{|[\Delta_n]_L|}}{\log \frac{|[\Delta_n]_L + [\Delta_n]_L|}{|[\Delta_n]_L|}} = \frac{\log \frac{\binom{2n}{n}}{n!} - \log \frac{1}{n!} + o_L(1)}{\log \frac{2^n}{n!} - \log \frac{1}{n!} + o_L(1)} = \frac{\log \binom{2n}{n} + o_L(1)}{n \log 2 + o_L(1)},$$

where we used $\mathrm{vol}(\Delta_n) = \frac{1}{n!}$, $\mathrm{vol}(\Delta_n - \Delta_n) = \frac{1}{n!}\binom{2n}{n}$ and $\mathrm{vol}(\Delta_n + \Delta_n) = \frac{2^n}{n!}$. Sending $L \to \infty$ followed by $n \to \infty$ yields the sharpness of (3.1).

3.2 On sharpness of the doubling-difference entropy inequality

Analogous to the tightness of the inequality (3.1), it is natural to ask if the same can be said about the inequalities (3.2) and (3.3). It is unclear if the constants 1/2 and 2 are indeed the best possible. However, one can establish that the sharp constants for the discrete and the continuous versions are the same as a direct consequence of Theorem 1 and Theorem 2. Moreover, these sharp constants are independent of the dimension n. First, we begin with the doubling-difference entropy inequality.

Theorem 5. *([9, Theorems 1.1 and 2.1]) If U and U' are two discrete i.i.d. random variables, and X and X' are two continuous i.i.d. random variables, then:*

$$\frac{1}{2} \leq \frac{H(U - U') - H(U)}{H(U + U') - H(U)} \leq 2,$$
$$\frac{1}{2} \leq \frac{h(X - X') - h(X)}{h(X + X') - h(X)} \leq 2.$$

Proof. See Appendix A.1 □

Remark 2. A careful analysis of the above proof reveals that the equality in the upper bound for the continuous case holds only when X is deterministic in which case it fails to have an absolutely continuous distribution. As a result we can conclude that there exists no continuous distribution for which the equality is attained in (3.3); however, this does not rule out the possibility that we can get arbitrarily close to "2" by a sequence of distributions.

Now we present our main result of this chapter which establishes the uniqueness of the sharp constants for both the continuous and discrete versions.

Theorem 6. *For i.i.d. U and U' and i.i.d. X and X',*

$$\frac{1}{2} \leq \inf_{U \in \mathbb{Z}^n} \frac{H(U-U') - H(U)}{H(U+U') - H(U)} = \inf_{X \in \mathbb{R}^n} \frac{h(X-X') - h(X)}{h(X+X') - h(X)} \tag{3.4}$$

$$\leq \sup_{X \in \mathbb{R}^n} \frac{h(X-X') - h(X)}{h(X+X') - h(X)} = \sup_{U \in \mathbb{Z}^n} \frac{H(U-U') - H(U)}{H(U+U') - H(U)} \leq 2. \tag{3.5}$$

Furthermore, the infimum and the supremum are independent of the dimension n.

Proof. First we will prove that the infima of the ratios for both the discrete and continuous versions are the same. Let $\alpha_n = \inf_{U \in \mathbb{Z}^n} \frac{H(U-U')-H(U)}{H(U+U')-H(U)}$ and $\beta_n = \inf_{X \in \mathbb{R}^n} \frac{h(X-X')-h(X)}{h(X+X')-h(X)}$. For any $U \in \mathbb{Z}^n$, by definition,

$$\frac{H(U-U') - H(U)}{H(U+U') - H(U)} \geq \alpha_n,$$

which in view of Theorem 1 implies that for any $X \in \mathbb{R}^n$

$$\frac{h(X-X') - h(X)}{h(X+X') - h(X)} \geq \alpha_n,$$

and thus

$$\beta_n \geq \alpha_n.$$

In a similar manner using the definition of β_n and Theorem 2 gives

$$\alpha_n \geq \beta_n,$$

which completes the proof that $\alpha_n = \beta_n$. The same argument holds for the result for the supremum. Now we will prove that the infimum and supremum are dimension independent. Clearly $\alpha_n \leq \alpha_1$ by the tensorization property of Shannon entropy. For any $U \in \mathbb{Z}^n$ and its identical copy U', using a similar argument as in the proof of Lemma 6, we can find a linear embedding $f : \mathbb{Z}^n \to \mathbb{Z}$ that preserves the Shannon entropy of $U+U', U-U', U$ and U'. Thus

$$\frac{H(U-U') - H(U)}{H(U+U') - H(U)} = \frac{H(f(U)) - f(U')) - H(f(U))}{H(f(U) + f(U')) - H(f(U))}$$

and $\alpha_1 \leq \alpha_n$. The proof for the supremum is the same after replacing α_n and α_1 with β_n and β_1 respectively. $\square$

It is worth pointing out that the dimension-freeness of the best Shannon entropy ratio follows from standard arguments (tensorization and linear embedding of $\mathbb{Z}^n$ into $\mathbb{Z}$), which have been previously used for proving analogous results for set cardinalities [18]; however, it is unclear how to directly prove that the ratio of differential entropy is dimension-independent without resorting to Theorem 1.

3.3 Numerical plots for the sum-difference ratio

Since the role of the constant "2" being the best upper bound in the doubling-difference inequalities is unclear, we resort to numerically finding the best possible constant for some known distributions. We consider distributions that are not symmetric around 0; otherwise, the sum-difference ratio equals 1. This clearly rules out the possibility of Gaussian and uniform random variables. Since the sum-difference ratio can be rewritten as $\frac{I(X-X';X)}{I(X+X';X)}$, intuitively, to make the ratio arbitrarily close to "2" we need to consider distributions that are highly "asymmetric" where the difference $X - X'$ reveals much more "information" about X than $X + X'$ does about the same.

3.3.1 Bernoulli distribution

Let X be a Bernoulli random variable with parameter $p \in [0, 1]$. Since X and X' are i.i.d., it follows that $X + X'$ takes values in $\{0, 1, 2\}$ with probabilities $(1-p)^2, 2p(1-p)$, and p^2 respectively. Similarly, $X - X'$ takes values in

$\{-1, 0, 1\}$ with probabilities $p(1-p), p^2 + (1-p)^2$, and $p(1-p)$ respectively.

$$\begin{aligned}
\frac{H(X - X') - H(X)}{H(X + X') - H(X)} &= \frac{I(X; X - X')}{I(X; X + X')} \\
&= \frac{H(X) - H(X|X - X')}{H(X) - H(X|X + X')} \\
&= \frac{H(X) - H(X|X - X' = 0)}{H(X) - H(X|X + X' = 1)} \\
&= \frac{h(p) - (p^2 + (1-p)^2) h\left(\frac{p^2}{p^2 + (1-p)^2)}\right)}{h(p) - 2p(1-p) h\left(\frac{1}{2}\right)},
\end{aligned}$$

where $h(p) = p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p}$ denotes the binary entropy function. Figure. 3.1 is the plot of the sum-difference ratio vs. p for Bernoulli random variables. As can be seen, the ratio achieves a maximum value of 1.369 for both $p = 0.06$ and $p = 0.94$.

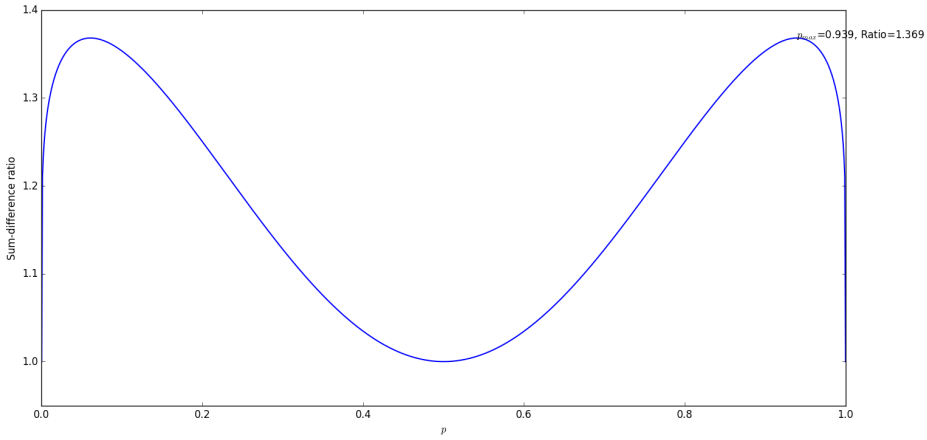

Figure 3.1: Sum-difference ratio vs. p

3.3.2 Poisson distribution

Let $X \sim \text{Poi}(\lambda), \lambda > 0$. Thus $X + X' \sim \text{Poi}(2\lambda)$. The pmf of the difference is given through the modified Bessel functions of first-kind $I_k(\cdot)$:

$$\mathbb{P}[X - X' = k] = e^{-2\lambda} I_{|k|}(2\lambda), \quad k \in \mathbb{Z}.$$

Using the well-known formula for entropy of Poisson random variables, we have

$$H(X) = \lambda(1 - \log \lambda) + e^{-\lambda} \left(\sum_{k=0}^{\infty} \frac{\lambda^k \log k!}{k!} \right),$$
$$H(X + X') = 2\lambda(1 - \log 2\lambda) + e^{-2\lambda} \left(\sum_{k=0}^{\infty} \frac{(2\lambda)^k \log k!}{k!} \right).$$

For the difference, $H(X - X') = p_0 \log \frac{1}{p_0} + 2 \sum_{k=1}^{\infty} p_k \log \frac{1}{p_k}$ where $p_k = \mathbb{P}[X - X' = k]$. Since there are no simplified expressions for these entropies, we truncated the summations to around 10000 terms for good-approximation. We obtained the plot in Figure. 3.2 for the sum-difference ratio as we varied the mean λ.

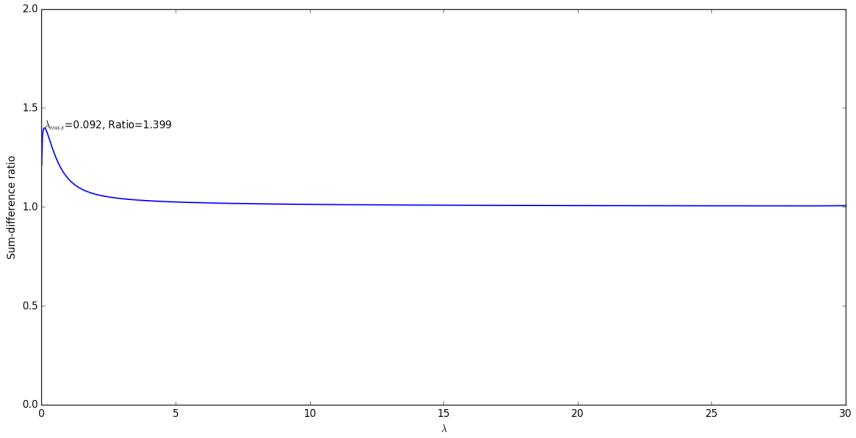

Figure 3.2: Sum-difference ratio vs. λ

Since a Poisson distribution with a large λ resembles a normal distribution with mean and variance λ due to the central limit theorem, it is expected that the ratio is close to 1 for reasonably large values. Figure. 3.2 reveals that it saturates to 1 for λ around 6 after achieving an approximate maximum value of 1.399 for λ close to 0.1.

3.3.3 Geometric distribution

Let $X \sim \text{Geom}(p)$ be a geometric random variable with parameter $p \in [0, 1]$ whose pmf is given by $\mathbb{P}[X = k] = (1-p)p^{k-1}$ for $k \geq 1$. It is straightforward

using the convolutions to derive that the sum and the difference have the following probability mass functions:

$$\begin{aligned}\mathbb{P}\left[X+X'=k\right] &= (k-1)p^2(1-p)^{k-2}, \quad k \geq 1,\\ \mathbb{P}\left[X-X'=k\right] &= \frac{p}{2-p}(1-p)^{|k|}, \quad k \in \mathbb{Z}.\end{aligned}$$

Thus the discrete entropies for $X+X'$ and $X-X'$ are given by

$$\begin{aligned}H(X+X') = -\mathbb{E}[\log P_{X+X'}(X+X')] &= -2\log p - 2\log(1-p)\mathbb{E}[X+X'-2]\\ &\quad - \mathbb{E}[\log(X+X'-1)]\\ &= \frac{2h(p)}{p} - \mathbb{E}[\log(X+X'-1)],\end{aligned}$$

and

$$\begin{aligned}H(X-X') = -\mathbb{E}[\log P_{X-X'}(X-X')] &= \log\frac{2-p}{p} - \log(1-p)\mathbb{E}|X-X'|\\ &= \log\frac{2-p}{p} - \frac{2(1-p)}{p(2-p)}\log(1-p),\end{aligned}$$

where we used the fact that $\min\{X,X'\} \sim \text{Geom}(1-(1-p)^2)$ and $|X-X'| = X+X'-2\min\{X,X'\}$. Plugging the above entropies in the sum-difference ratio together with the fact that $H(X) = \frac{h(p)}{p}$, we obtain

$$\begin{aligned}\frac{H(X-X')-H(X)}{H(X+X')-H(X)} &= \frac{\log\frac{2-p}{p} - \frac{2(1-p)}{p(2-p)}\log(1-p) - \frac{h(p)}{p}}{\frac{2h(p)}{p} - \mathbb{E}[\log(X+X'-1)] - \frac{h(p)}{p}}\\ &= \frac{p\log\frac{2-p}{p} - \frac{2(1-p)}{(2-p)}\log(1-p) - h(p)}{h(p) - p\mathbb{E}[\log(X+X'-1)]}.\end{aligned}$$

After numerically computing $\mathbb{E}[\log(X+X'-1)]$ using the pmf of $X+X'$, we obtain the plot in Figure. 3.3. The maximum value of the ratio in this case is approximately 1.4 attained at $p = 0.89$ as can be seen from the figure.

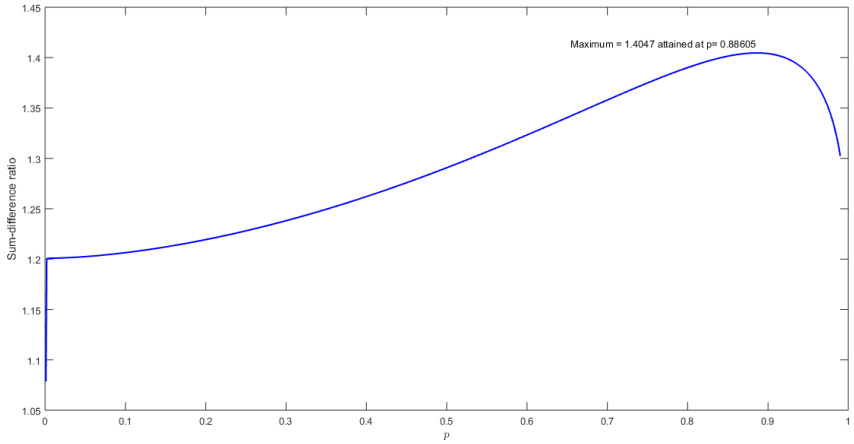

Figure 3.3: Sum-difference ratio vs. p for geometric random variables

3.3.4 Exponential distribution

Let $X \sim \text{Exp}(\lambda), \lambda > 0$. Using the notation $\text{Gamma}(\alpha, \beta)$ to denote a gamma random variable with parameters $\alpha > 0$ and $\beta > 0$ with the pdf

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$$

we know that $X + X' \sim \text{Gamma}(2, \lambda)$. Plugging $\alpha = 2$ and $\beta = \lambda$ in the expression for the differential entropy of a gamma random variable, we obtain

$$\begin{aligned} h(X + X') = \alpha - \log \beta + \log(\Gamma(\alpha)) + (1 - \alpha)\psi(\alpha)|_{\alpha=2,\beta=\lambda} &= 2 - \psi(2) - \log \lambda \\ &= \gamma + 1 - \log \lambda, \end{aligned}$$

where $\gamma \approx 0.5772$ is Euler's constant. Similarly, $h(X) = 1 - \log \lambda$ using the fact that $\text{Exp}(\lambda)$ random variable can be viewed as a $\text{Gamma}(1, \lambda)$ random variable. The difference $X - X'$ follows the Laplace distribution whose pdf and entropies are given by

$$\begin{aligned} f_{X-X'}(z) &= \frac{\lambda}{2} e^{-\lambda|z|}, \quad z \in \mathbb{R}, \\ h(X - X') &= 1 + \log \frac{2}{\lambda}. \end{aligned}$$

Plugging all these values together in the sum-difference ratio, we get

$$\begin{aligned}\frac{h(X-X')-h(X)}{h(X+X')-h(X)} &= \frac{1+\log 2-\log\lambda-1+\log\lambda}{\gamma+1-\log\lambda-1+\log\lambda}\\ &= \frac{\log 2}{\gamma}\\ &\approx 1.2008.\end{aligned}$$

Since the ratio is independent of the mean parameter λ, we obtain the constant plot in Figure. 3.4.

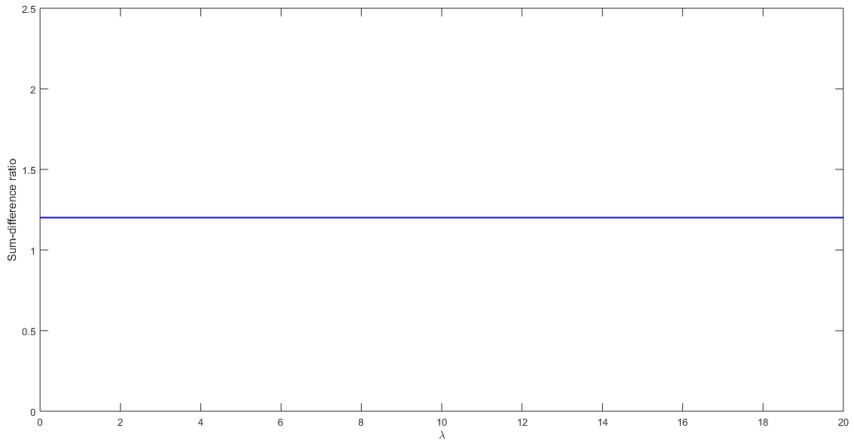

Figure 3.4: Sum-difference ratio vs. λ for exponential random variables

Remark 3. An alternative way to derive that the ratio is independent of λ is through observing that our sum-difference ratio is scale invariant. Since $\lambda X \sim \text{Exp}(1)$, we see that the ratio for any exponential random variable equals that of a Exp(1) random variable which evaluates to $\frac{\log 2}{\gamma}$.

3.3.5 Gamma distribution

Let $X \sim \text{Gamma}(N, 1), N \in \mathbb{N}$. Here we assumed, without loss of generality, that the scale parameter β equals 1 because of the scale invariance of the sum-difference ratio. Thus, $X + X' \sim \text{Gamma}(2N, 1)$. Hence

$$\begin{aligned}h(X) &= N + \log(\Gamma(N)) + (1-N)\psi(N),\\ h(X+X') &= 2N + \log(\Gamma(2N)) + (1-2N)\psi(2N),\end{aligned}$$

where $\psi(x) = \frac{d}{dx}\Gamma(x)$ denotes the digamma function. The probability density function for the difference $X - X'$ is given through the modified Bessel functions of the second-kind $K_\alpha(\cdot)$ as follows:

$$f_{X-X'}(z) = \frac{1}{\sqrt{\pi}\Gamma(N)} \left(\frac{|z|}{2}\right)^{N-\frac{1}{2}} K_{\frac{1}{2}-N}(|z|), \quad z \in \mathbb{R}.$$

Since there are no simplified expressions for the entropy of $X - X'$ in this case, we numerically computed the integral for evaluating $h(X - X') = -\int_{\mathbb{R}} f_{X-X'}(z) \log \frac{1}{f_{X-X'}(z)} dz$. We obtained the plot in Figure. 3.5 for the sum-difference ratio.

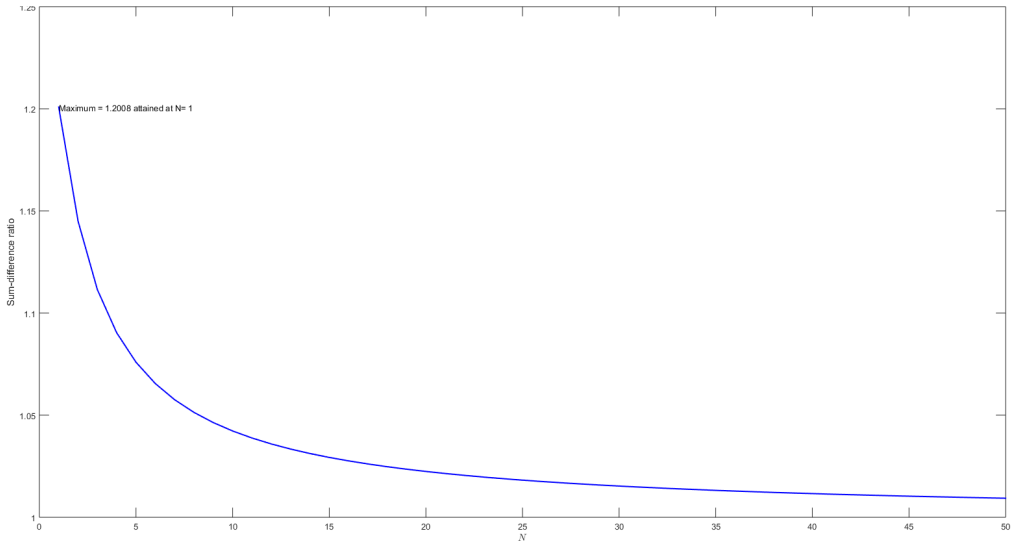

Figure 3.5: Sum-difference ratio vs. N for gamma random variables

Remark 4. We can verify from Figure. 3.5 that the ratio stays close to 1 for large values of N as expected since $h(X \pm X') - h(X) = I(X; X \pm X') \to \frac{1}{2}\log 2$ because of the central limit theorem. The ratio attains a maximum value of around 1.2 for $N = 1$ and monotonically decreases thereafter.

3.3.6 Binomial distribution

Let $X \sim \text{Bin}(n, p)$. Then $X + X' \sim \text{Bin}(2n, p)$. Since the Shannon entropy is translation invariant, we get $H(X - X') = H(X + n - X')$ where $X + n - X' \sim \text{Bin}(n, p) + \text{Bin}(n, 1 - p)$. Though a closed form expression for $H(X)$ can be found ($H(X) = nh(p) - \mathbb{E}\log\binom{n}{X}$), we used the definition that

$H(X) = \sum_{k=0}^{n} p_k \log \frac{1}{p_k}$ and similarly for $H(X + X')$ and $H(X - X')$ to plot the sum-difference ratio for values of n and p. Due to the symmetry of the ratio around $p = 0.5$, we plotted only for values of p in the range $[0, 0.5]$, and Figures $3.6 - 3.10$ are the corresponding plots for sum-difference ratio vs. n for $p \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$.

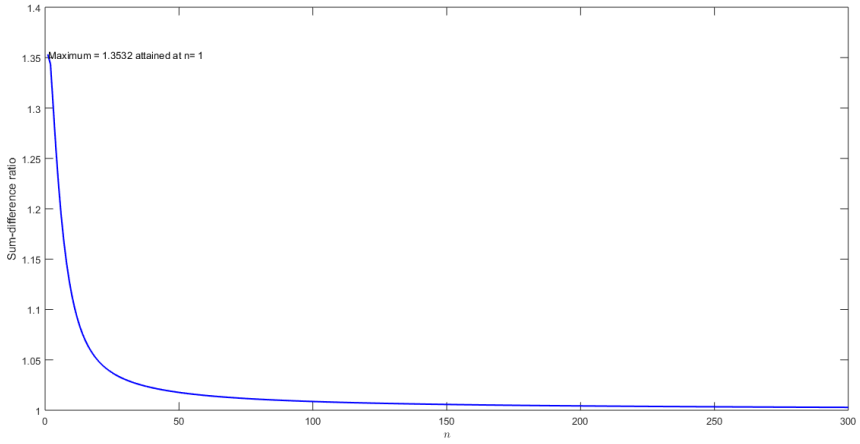

Figure 3.6: Sum-difference ratio vs. n for $p = 0.1$.

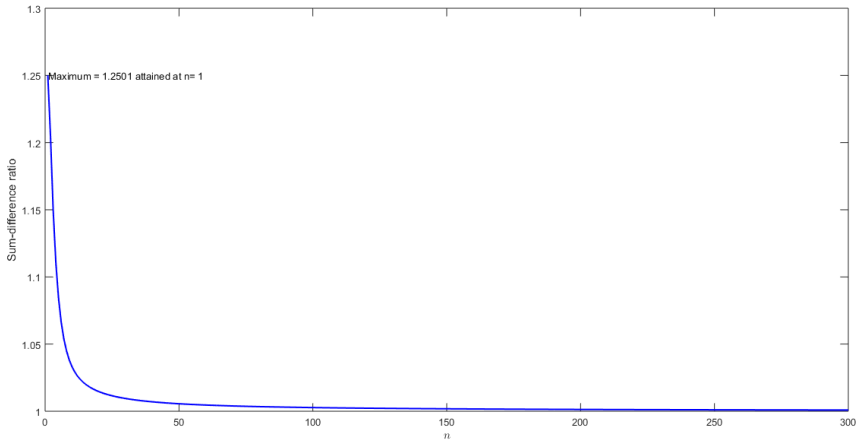

Figure 3.7: Sum-difference ratio vs. n for $p = 0.2$.

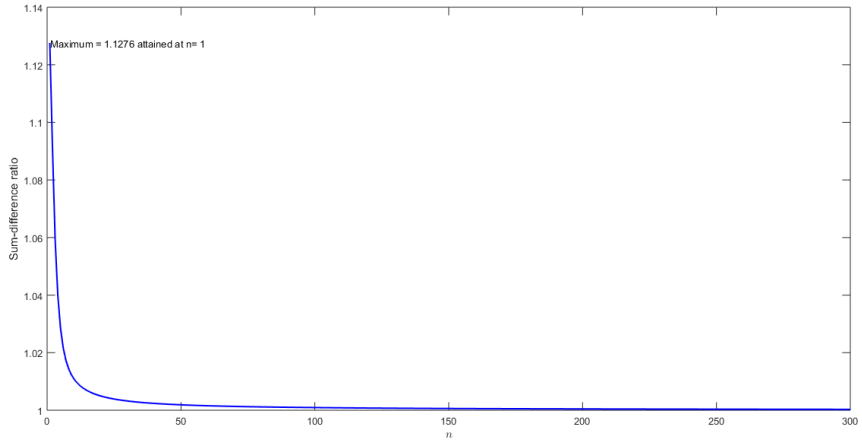

Figure 3.8: Sum-difference ratio vs. n for $p = 0.3$.

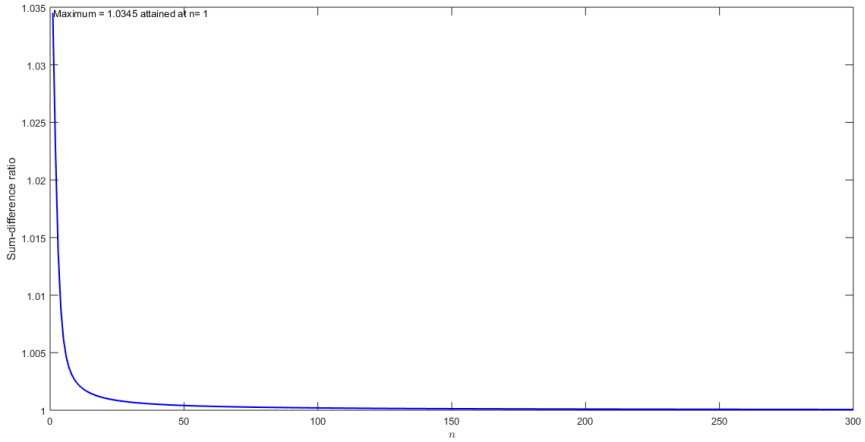

Figure 3.9: Sum-difference ratio vs. n for $p = 0.4$.

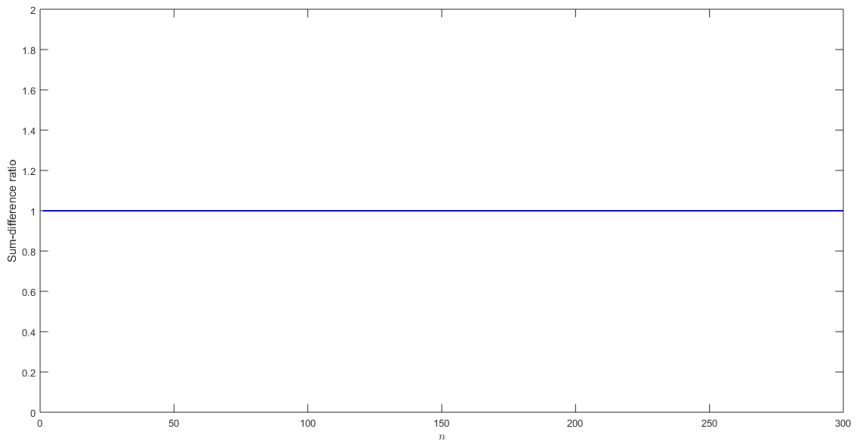

Figure 3.10: Sum-difference ratio vs. n for $p = 0.5$.

Remark 5. Since a binomial random variable is a sum of n i.i.d. Bernoulli random variables, in view of the central limit theorem, we expect that the above ratio converges to one for large values of n for all p. The plots reveal the same behavior with the decay of the ratio to 1 being faster for larger values of p. Since both $X + X'$ and $X + n - X'$ are identical in distribution for $p = 0.5$, we get a constant ratio value of 1 for all values of n.

Remark 6. Notice that the maximum values for the ratio occur at $n = 1$ for all p. Thus it is equivalent to finding the maximum value for the Bernoulli case which we found to be approximately 1.369 for both $p = 0.06$ and $p = 0.94$.

CHAPTER 4

CONCLUSION

In this thesis, we explored the nature of the relationship between the linear entropy inequalities for Shannon entropy and differential entropy. In particular, we established the equivalence of the balanced linear entropy inequalities for the discrete and the continuous versions. We further extended these results for a general group setting. We investigated the implications of our equivalence result in understanding the best constants for certain additive-combinatorial entropy inequalities, in particular, the *doubling-difference* inequality, where we established that the best constants for the continuous and the discrete cases are the same.

To conclude this chapter, we discuss some open problems related to this work.

4.1 Uniform distribution on simplex

In Chapter 3 we proved that the sharp constants for the discrete and continuous versions are the same, and dimension-free. In view of the success of continuous approximation in proving the sharpness of (3.1), proving the sharpness of (3.3) for differential entropies might be more tractable than its discrete counterpart (3.2). Similar to the simplex construction in Section 3.1, one can consider X to be a uniform random variable on simplex Δ_n and let $n \to \infty$ to verify whether this specific construction achieves the upper bound 2 in (3.3). However, the analysis of the entropies of the difference $X - X'$ as well as the sum $X + X'$ seems intractable in this case. It remains an open question whether the uniform distribution on simplex, or any other distribution in general, achieves the constant 2.

APPENDIX A

PROOFS OF CHAPTER 3

A.1 Proof of Theorem 5

We prove the theorem for the continuous version and the same argument holds for the discrete case too. Let X, Y and Z be independent continuous random variables in $\mathbb{R}^n$. By data processing inequality,

$$I(X; X-Y, Y-Z) \geq I(X; X-Z), \tag{A.1}$$

which is equivalent to

$$h(X-Z) + h(Y) \leq h(X-Y, Y-Z).$$

Using the non-negativity of the mutual information we further obtain that

$$h(X-Z) + h(Y) \leq h(X-Y) + h(Y-Z), \tag{A.2}$$

which is also known in the additive-combinatorial literature as Ruzsa triangle inequality for differential entropy [7]. Replacing $-Y$ by Y and taking X, Y and Z to be i.i.d., we obtain

$$h(X-Z) + h(Y) \leq h(X+Y) + h(Y+Z)$$

and further,

$$h(X-Y) - h(X) \leq 2\left(h(X+Y) - h(X)\right).$$

This proves the upper bound. For the lower bound, for any independent X, Y and Z, using the data processing inequality we have

$$I(X+Y+Z;X) \leq I(X+Y,Z;X) = I(X+Y;X).$$

This is equivalent to

$$h(X+Y+Z) + h(Y) \leq h(X+Y) + h(Y+Z).$$

Hence,

$$\begin{aligned} h(X+Z) + h(Y) &\leq h(X+Y+Z) + h(Y), \\ &\leq h(X+Y) + h(Y+Z). \end{aligned}$$

Replacing $-Y$ by Y and taking X, Y and Z to be i.i.d. implies the lower bound:

$$h(X+Y) - h(X) \leq 2\left(h(X-Y) - h(X)\right).$$

Remark 7. The equality in the upper bound holds if and only if we have equalities in (A.1) and (A.2). This happens only when (1) $X \to X - Z \to (X+Y, Y+Z)$ is a Markov chain and (2) $X+Y$ and $Y+Z$ are independent, where X, Y and Z are i.i.d.. The second condition implies that $\mathrm{Cov}(X+Y, Y+Z) = 0$ which gives $\mathrm{Cov}(Y,Y) = 0$. Thus X, Y and Z should have degenerate distributions to achieve the equality in which case their differential entropies are not finite.

APPENDIX B

PROOFS OF CHAPTER 2

B.1 Proof of Lemma 2

Let $a_1, \ldots, a_m \in \mathbb{Z}$ and $X_1, \ldots, X_m$ be $\mathbb{R}^d$-valued random variables. Then

$$\left[\sum_{i=1}^m a_i X_i\right]_k = \frac{\left\lfloor 2^k \sum_{i=1}^m a_i X_i \right\rfloor}{2^k} = \frac{\left\lfloor \sum_{i=1}^m a_i \lfloor 2^k X_i \rfloor + \lfloor \sum_{i=1}^m a_i \{2^k X_i\} \right\rfloor}{2^k}$$
$$= \sum_{i=1}^m a_i [X_i]_k + \frac{\lfloor \sum_{i=1}^m a_i \{2^k X_i\} \rfloor}{2^k}.$$

Define

$$A_k \triangleq 2^k \left[\sum_{i=1}^m a_i X_i\right]_k, \quad B_k \triangleq 2^k \sum_{i=1}^m a_i \left[X_i\right]_k, \quad Z_k \triangleq \left\lfloor \sum_{i=1}^m a_i \{2^k X_i\} \right\rfloor.$$

It is easy to see that $A_k, B_k, Z_k \in \mathbb{Z}^d$ and $A_k = B_k + Z_k$. Since $\{2^k X\} \in [0,1)^d$, each component of Z_k takes integer values in the set $a_1[0,1) + \ldots + a_m[0,1)$ and hence $Z_k \in \mathcal{Z} \triangleq \{a, a+1, \ldots, b-1\}^d$, where $b \triangleq \sum_{i=1}^m a_i \mathbb{1}_{\{a_i > 0\}}$ and $a \triangleq \sum_{i=1}^m a_i \mathbb{1}_{\{a_i < 0\}}$. Hence Z_k takes at most $(b-a)^d$ values, which is bounded for all k.

Next we describe the outline of the proof:

1. The goal is to prove $|H(A_k) - H(B_k)| \to 0$. Since $A_k = B_k + Z_k$, we have

$$H\left(A_k\right) - H\left(B_k\right) = I\left(Z_k; A_k\right) - I\left(Z_k; B_k\right). \tag{B.1}$$

   Hence it suffices to show that both mutual informations vanish as $k \to \infty$.

2. Lemma 10 proves $I\left(Z_k; B_k\right) \to 0$ based on the data processing inequal-

ity and Lemma 7 which asserts that asymptotic independence between the integral part $\lfloor 2^k X \rfloor$ and the fractional part $\{2^k X\}$, in the sense of vanishing mutual information. As will be evident in the proof of Lemma 7, this is a direct consequence of Rényi's result (Lemma 4).

3. Since Z_k takes a bounded number of values, $I(Z_k; A_k) \to 0$ is *equivalent* to the total variation between P_{Z_k,A_k} and $P_{Z_k} \otimes P_{A_k}$ vanishing, known as the T-information [21, 22]. By the triangle inequality and data processing inequality for the total variation, this objective is further reduced to proving the convergence of two pairs of conditional distributions in total variation: one is implied by Pinsker's inequality and Lemma 10, and the other follows from an elementary fact on the total variation between a pdf and a small shift of itself (Lemma 9). Lemma 11 contains the full proof; notably, the argument crucially depends on the assumption that $a_1, \ldots, a_m$ are relatively prime.

We start with the following auxiliary result.

Lemma 7. *Let X be a $[0,1]^d$-valued continuous random variable such that both $h(X)$ and $H(\lfloor X \rfloor)$ are finite. Then*

$$\lim_{k\to\infty} I(\lfloor 2^k X \rfloor; \{2^k X\}) = 0.$$

Proof. Since $X \in [0,1]^d$, we can write X in terms of its binary expansion as:

$$X = \sum_{i\geq 1} X_i 2^{-i}, X_i \in \{0,1\}^d.$$

In other words, $\lfloor 2^k X \rfloor = 2^{k-1} X_1 + \ldots + X_k$. Thus, $\lfloor 2^k X \rfloor$ and $(X_1, \ldots, X_k)$ are in a one-to-one correspondence and so are $\{2^k X\}$ and $(X_{k+1}, \ldots)$. So,

$$I(\lfloor 2^k X \rfloor; \{2^k X\}) = I(X_1^k; X_{k+1}^\infty) \triangleq I(X_1, \ldots, X_k; X_{k+1}, \ldots).$$

Then $I\left(X_1^k; X_{k+1}^\infty\right) = \lim_{m\to\infty} I(X_1^k; X_{k+1}^{k+m})$ cf. [23, Section 3.5]. Let $a_k \triangleq H\left(X_1^k\right) - dk \log 2 - h(X)$. Then Lemma 4 implies $\lim_{k\to\infty} a_k = 0$. Hence

for each $k, m \geq 1$, we have

$$\begin{aligned} I(X_1^k; X_{k+1}^{k+m}) &= H(X_1^k) + H(X_{k+1}^{k+m}) - H(X_1^{k+m}) \\ &= h(X) + dk \log 2 + a_k - (h(X) + d(k+m) \log 2 + a_{k+m}) \\ &\qquad + H(X_{k+1}^{k+m}) \qquad \text{(B.2)} \\ &= a_k - a_{k+m} + H(X_{k+1}^{k+m}) - md \log 2 \\ &\leq a_k - a_{k+m}, \qquad \text{(B.3)} \end{aligned}$$

where (B.3) follows from the fact that X_{k+1}^{k+m} can take only 2^{md} values. Since $I(X_1^k; X_{k+1}^{k+m}) \geq 0$, by (B.3), sending $m \to \infty$ first and then $k \to \infty$ completes the proof. □

Recall that the total variation distance between probability distributions μ and ν is defined as:

$$d_{\mathrm{TV}}(\mu, \nu) \triangleq \sup_F |\mu(F) - \nu(F)|,$$

where the supremum is taken over all measurable sets F.

Lemma 8. *Let X, Y, Z be random variables such that $Z = f(X) = f(Y)$, for some measurable function f. Then for any measurable E such that $\mathbb{P}[Z \in E] > 0$,*

$$d_{\mathrm{TV}}\left(P_{X|Z\in E}, P_{Y|Z\in E}\right) \leq \frac{d_{\mathrm{TV}}(P_X, P_Y)}{\mathbb{P}[Z \in E]}.$$

Proof. For any measurable F,

$$\begin{aligned} \left|P_{X\in F|Z\in E} - P_{Y\in F|Z\in E}\right| &= \frac{|\mathbb{P}[X \in F, f(X) \in E] - \mathbb{P}[Y \in F, f(Y) \in E]|}{\mathbb{P}[Z \in E]} \\ &\leq \frac{d_{\mathrm{TV}}(P_X, P_Y)}{\mathbb{P}[Z \in E]}. \end{aligned}$$

The claim now follows from taking supremum over all F. □

Lemma 9. *If X is a $\mathbb{R}$-valued continuous random variable, then:*

$$d_{\mathrm{TV}}(P_X, P_{X+a}) \to 0 \text{ as } a \to 0.$$

Proof. Let f be the pdf of X. Since continuous functions with compact

support are dense in $\mathcal{L}^1(\mathbb{R})$, for any $\varepsilon > 0$, there exists a continuous and compactly supported function g such that $\|f - g\|_1 < \frac{\varepsilon}{3}$. Because of the uniform continuity of continuous functions on compact sets, there exists a $\delta > 0$ such that, whenever $|a| < \delta$, $\|g(\cdot + a) - g(\cdot)\|_1 < \frac{\varepsilon}{3}$. Hence $\|f(\cdot + a) - f(\cdot)\|_1 < 2\|f(\cdot) - g(\cdot)\|_1 + \|g(\cdot + a) - g(\cdot)\|_1 < \varepsilon$. Hence the claim follows. □

Lemma 10. *If $X_1, \ldots, X_m$ are independent $[0,1]^d$-valued continuous random variables such that both $h(X_j)$ and $H(\lfloor X_j \rfloor)$ are finite for each $j \in [m]$, then*

$$\lim_{k\to\infty} I(Z_k; B_k) = 0.$$

Proof. We have

$$\begin{aligned} I(Z_k; B_k) &= I\Big(\Big\lfloor \sum_{i=1}^m a_i\{2^k X_i\}\Big\rfloor; \sum_{i=1}^m a_i\lfloor 2^k X_i\rfloor\Big) \\ &= I\Big(\Big\lfloor \sum_{i=1}^m a_i\{2^k X_i\}\Big\rfloor; \Big\lfloor \sum_{i=1}^m a_i\lfloor 2^k X_i\rfloor\Big\rfloor\Big) \\ &\overset{(a)}{\le} I\big(a_1\{2^k X_1\}, \ldots, a_m\{2^k X_m\}; a_1\lfloor 2^k X_1\rfloor, \ldots, a_m\lfloor 2^k X_m\rfloor\big) \\ &\overset{(b)}{=} \sum_{i=1}^m I(\{2^k X_i\}; \lfloor 2^k X_i\rfloor), \end{aligned}$$

where (a) follows from the data processing inequality and (b) follows from the fact that $X_1, \ldots, X_m$ are independent. Applying Lemma 7 to each X_i finishes the proof. □

In view of (B.1), Lemma 2 follows from Lemma 10 and the next lemma:

Lemma 11. *Under the assumptions of Lemma 10 and if $a_1, \ldots, a_m \in \mathbb{Z}$ are relatively prime,*

$$\lim_{k\to\infty} I(Z_k; A_k) = 0.$$

Proof. Define the T-information between two random variables X and Y as follows:

$$T(X;Y) \triangleq d_{\mathrm{TV}}(P_{XY}, P_X P_Y).$$

By [22, Proposition 12], if a random variable W takes values in a finite set $\mathcal{W}$, then

$$I(W;Y) \le \log(|\mathcal{W}| - 1)T(W;Y) + h(T(W;Y)), \tag{B.4}$$

where $h(x) = x \log \frac{1}{x} + (1-x) \log \frac{1}{1-x}$ is the binary entropy function.

Since Z_k takes at most $(b-a)^d$ values, by (B.4), it suffices to prove that $\lim_{k\to\infty} T(Z_k; A_k) = 0$. It is well-known that the uniform fine quantization error of a continuous random variable converges to the uniform distribution (see, e.g., [24, Theorem 4.1]). Therefore $\{2^k X_i\} \xrightarrow{\mathcal{L}} \mathrm{Unif}[0,1]^d$ for each $i \in [m]$. Furthermore, since X_i are independent, $Z_k = \lfloor \sum_{i=1}^m a_i \{2^k X_i\} \rfloor \xrightarrow{\mathcal{L}} \lfloor \sum_{i=1}^m a_i U_i \rfloor$ where $U_1, \ldots, U_m$ are i.i.d. $\mathrm{Unif}[0,1]^d$ random variables.

Let $\mathcal{Z}' \triangleq \{z \in \mathcal{Z} : \mathbb{P}[\lfloor \sum_{i=1}^m a_i U_i \rfloor = z] > 0\}$. Since $Z_k \xrightarrow{\mathcal{L}} \lfloor \sum_{i=1}^m a_i U_i \rfloor$, $\lim_{k\to\infty} \mathbb{P}[Z_k = z] > 0$ for any $z \in \mathcal{Z}'$ and $\lim_{k\to\infty} \mathbb{P}[Z_k = z] = 0$ for any $z \in \mathcal{Z} \backslash \mathcal{Z}'$. Since

$$\begin{aligned} T(Z_k; A_k) &= \sum_{z\in\mathcal{Z}} \mathbb{P}[Z_k = z]\, d_{\mathrm{TV}}(P_{A_k}, P_{A_k|Z_k=z}) \\ &\le \sum_{z\in\mathcal{Z}'} d_{\mathrm{TV}}(P_{A_k}, P_{A_k|Z_k=z}) + \sum_{z\in\mathcal{Z}\backslash\mathcal{Z}'} \mathbb{P}[Z_k = z], \end{aligned}$$

it suffices to prove that $d_{\mathrm{TV}}(P_{A_k}, P_{A_k|Z_k=z}) \to 0$ for any $z \in \mathcal{Z}'$.

Using the triangle inequality and the fact that

$$P_{A_k} = \sum_{z'\in\mathcal{Z}} \mathbb{P}[Z_k = z']\, P_{A_k|Z_k=z'},$$

we have

$$\begin{aligned} d_{\mathrm{TV}}(P_{A_k}, P_{A_k|Z_k=z}) &\le \sum_{z'\in\mathcal{Z}} \mathbb{P}[Z_k = z']\, d_{\mathrm{TV}}(P_{A_k|Z_k=z}, P_{A_k|Z_k=z'}) \\ &\le \sum_{z'\in\mathcal{Z}'} d_{\mathrm{TV}}(P_{A_k|Z_k=z}, P_{A_k|Z_k=z'}) + \sum_{z\in\mathcal{Z}\backslash\mathcal{Z}'} \mathbb{P}[Z_k = z]. \end{aligned}$$

Thus it suffices to show that $d_{\mathrm{TV}}(P_{A_k|Z_k=z}, P_{A_k|Z_k=z'}) \to 0$ for any $z, z' \in \mathcal{Z}'$. Since $A_k = B_k + Z_k$, we have

$$\begin{aligned} d_{\mathrm{TV}}(P_{A_k|Z_k=z}, P_{A_k|Z_k=z'}) &= d_{\mathrm{TV}}(P_{B_k+Z_k|Z_k=z}, P_{B_k+Z_k|Z_k=z'}) \\ &= d_{\mathrm{TV}}(P_{B_k+z|Z_k=z}, P_{B_k+z'|Z_k=z'}) \\ &\le d_{\mathrm{TV}}(P_{B_k+z|Z_k=z}, P_{B_k+z|Z_k=z'}) \\ &\qquad + d_{\mathrm{TV}}(P_{B_k+z|Z_k=z'}, P_{B_k+z'|Z_k=z'}) \qquad \text{(B.5)} \\ &= d_{\mathrm{TV}}(P_{B_k|Z_k=z}, P_{B_k|Z_k=z'}) \qquad \text{(B.6)} \\ &\qquad + d_{\mathrm{TV}}(P_{B_k+z|Z_k=z'}, P_{B_k+z'|Z_k=z'}). \qquad \text{(B.7)} \end{aligned}$$

Thus it suffices to prove that each term on the right-hand side of (B.7) vanishes. For the first term, note that

$$d_{\mathrm{TV}}(P_{B_k|Z_k=z}, P_{B_k|Z_k=z'}) \leq d_{\mathrm{TV}}(P_{B_k|Z_k=z}, P_{B_k}) + d_{\mathrm{TV}}(P_{B_k|Z_k=z'}, P_{B_k}),$$

where $d_{\mathrm{TV}}(P_{B_k|Z_k=z}, P_{B_k}) \to 0$ for any $z \in \mathcal{Z}'$ because, from the Pinsker's inequality,

$$\begin{aligned} I(Z_k; B_k) &= \sum_{z \in \mathcal{Z}} \mathbb{P}\left[Z_k = z\right] D(P_{B_k} \| P_{B_k|Z_k=z}) \\ &\geq 2 \sum_{z \in \mathcal{Z}} \mathbb{P}\left[Z_k = z\right] d_{\mathrm{TV}}^2(P_{B_k}, P_{B_k|Z_k=z}) \\ &\geq 2\mathbb{P}\left[Z_k = z\right] d_{\mathrm{TV}}^2(P_{B_k}, P_{B_k|Z_k=z}), \end{aligned}$$

and $I(Z_k; B_k) \to 0$ by Lemma 10 and $\liminf_{k\to\infty} \mathbb{P}\left[Z_k = z\right] > 0$ for any $z \in \mathcal{Z}'$.

Thus it remains to prove the second term on the right-hand of (B.7) vanishes for any $z, z' \in \mathcal{Z}'$. Since $a_1, \ldots, a_m$ are relatively prime, for any $p \in \mathbb{Z}$, there exists $q_1, \ldots, q_m \in \mathbb{Z}$ such that $p = \sum_{i=1}^m a_i q_i$. Hence, for any $z, z' \in \mathbb{Z}^d$, there exists $b_1, \ldots, b_m \in \mathbb{Z}^d$ such that

$$z' - z = \sum_{i=1}^m a_i b_i.$$

Then,

$$B_k + (z' - z) = \sum_{i=1}^m a_i \lfloor 2^k X_i \rfloor + \sum_{i=1}^m a_i b_i = \sum_{i=1}^m a_i \Big\lfloor 2^k (X_i + \frac{b_i}{2^k}) \Big\rfloor.$$

By definition, $Z_k = \lfloor \sum_{i=1}^m a_i \{2^k X_i\} \rfloor = \lfloor \sum_{i=1}^m a_i \{2^k (X_i + \frac{b_i}{2^k})\} \rfloor$. Consider

the second term on the right-hand of (B.7). We have

$$\begin{aligned}
d_{\mathrm{TV}}(P_{B_k+z|Z_k=z'}, P_{B_k+z'|Z_k=z'}) &= d_{\mathrm{TV}}(P_{B_k+(z'-z)|Z_k=z'}, P_{B_k|Z_k=z'}) \\
&= d_{\mathrm{TV}}\big(P_{\sum_{i=1}^m a_i \lfloor 2^k (X_i + \frac{b_i}{2^k})\rfloor | Z_k=z'}, \\
&\qquad\qquad P_{\sum_{i=1}^m a_i \lfloor 2^k X_i \rfloor | Z_k=z'}\big) \\
&\overset{(a)}{\leq} d_{\mathrm{TV}}(P_{X_1+\frac{b_1}{2^k},\ldots,X_m+\frac{b_m}{2^k}|Z_k=z'}, P_{X_1,\ldots,X_m|Z_k=z'}) \\
&\overset{(b)}{\leq} \frac{1}{\mathbb{P}\left[Z_k=z'\right]} d_{\mathrm{TV}}(P_{X_1+\frac{b_1}{2^k},\ldots,X_m+\frac{b_m}{2^k}}, P_{X_1,\ldots,X_m}) \\
&\overset{(c)}{\leq} \frac{1}{\mathbb{P}\left[Z_k=z'\right]} \sum_{i=1}^m d_{\mathrm{TV}}(P_{X_i+\frac{b_i}{2^k}}, P_{X_i}),
\end{aligned}$$

where (a) follows from the data processing inequality for total variation and (b) follows from Lemma 8, and (c) follows from the independence of $X_1, \ldots, X_m$. Letting $k \to \infty$ in view of Lemma 9 finishes the proof. □

B.2 Proof of Lemma 3

Proof. Let $X_1, \ldots, X_m$ be independent and $\mathbb{R}^d$-valued continuous random variables. Without loss of generality, we may assume $a_i \neq 0$. For each $i \in [m]$, $\mathbb{P}\left[X_i \in B_N^{(d)}\right] \xrightarrow{N\to\infty} 1$. Recall the conditional pdf notation (2.2). For $x \in \mathbb{R}^d$, we have

$$f_{a_i X_i^{(N)}}(x) = \frac{1}{|a_i|} f_{X_i^{(N)}}\left(\frac{x}{a_i}\right) = \frac{\frac{1}{|a_i|} f_{X_i}\left(\frac{x}{a_i}\right) \mathbb{1}\left\{\frac{x}{|a_i|} \in B_N^{(d)}\right\}}{\mathbb{P}\left[X_i \in B_N^{(d)}\right]} \tag{B.8}$$

$$= \frac{f_{a_i X_i}(x) \mathbb{1}\left\{\frac{x}{|a_i|} \in B_N^{(d)}\right\}}{\mathbb{P}\left[X_i \in B_N^{(d)}\right]}. \tag{B.9}$$

By the independence of the X_i's, the pdf of $\sum_{i=1}^m a_i X_i$ is given by:

$$\begin{aligned}
g(z) &\triangleq f_{a_1 X_1 + \ldots + a_m X_m}(z) \\
&= \int_{\mathbb{R}^d \times \cdots \times \mathbb{R}^d} f_{a_1 X_1}(x_1) \ldots f_{a_m X_m}(z - x_1 - \ldots - x_{m-1})\, dx_1 \cdots dx_{m-1}.
\end{aligned}$$

Similarly, in view of (B.9), the pdf of $\sum_{i=1}^m a_i X_i^{(N)}$ is given by:

$$\begin{aligned}
g_N(z) &\triangleq f_{a_1X_1^{(N)}+\ldots+a_mX_m^{(N)}}(z) \\
&= \int f_{a_1X_1^{(N)}}(x_1)\ldots f_{a_mX_m^{(N)}}(z-x_1-\ldots-x_{m-1})\,dx_1\ldots dx_{m-1} \\
&= \frac{1}{\prod_{i=1}^m \mathbb{P}\left[X_i \in B_N^{(d)}\right]} \cdot \int f_{a_1X_1}(x_1)\ldots f_{a_mX_m}(z-x_1-\ldots-x_{m-1}) \\
&\qquad \cdot \mathbb{1}\left\{\frac{x}{|a_i|} \in B_N^{(d)},\ldots,\frac{z-x_1-\ldots-x_{m-1}}{|a_m|} \in B_N\right\} dx_1\ldots dx_{m-1}.
\end{aligned}$$

Now taking the limit on both sides, we have $\lim_{N\to\infty} g_N(z) = g(z)$ a.e., which follows the dominated convergence theorem and the fact that $g(z)$ is finite a.e.

Next we prove that the differential entropy also converges. Let $N_0 \in \mathbb{N}$ be so large that

$$\prod_{i=1}^m \mathbb{P}\left[X_i \in B_N^{(d)}\right] \geq \frac{1}{2}$$

for all $N \geq N_0$. Now,

$$\begin{aligned}
\left|h\left(\sum_{j=1}^m a_jX_j\right) - h\left(\sum_{j=1}^m a_jX_j^{(N)}\right)\right| &= \left|\int_{\mathbb{R}^d} g\log\frac{1}{g} - \int_{\mathbb{R}^d} g_N\log\frac{1}{g_N}\right| \\
&\leq \int g_N\log\frac{g_N}{g} + \int\left|(g-g_N)\log\frac{1}{g}\right| \\
&= D\left(P_{\sum_{i=1}^m a_iX_i^{(N)}}\|P_{\sum_{i=1}^m a_iX_i}\right) \\
&\qquad + \int|(g-g_N)\log g| \\
&\overset{(a)}{\leq} \sum_{i=1}^m D\left(P_{X_i^{(N)}}\|P_{X_i}\right) \\
&\qquad + \int|(g-g_N)\log g| \\
&\overset{(b)}{=} \log\frac{1}{\prod_{i=1}^m \mathbb{P}\left[X_i \in B_N^{(d)}\right]} \\
&\qquad + \int|(g-g_N)\log g| \\
&\overset{(c)}{\to} 0 \text{ as } N\to\infty,
\end{aligned}$$

where (a) follows from the data processing inequality and (b) is due to $D\left(P_{X|X\in E}\|P_X\right) = \log\frac{1}{\mathbb{P}[X\in E]}$, and (c) follows from the dominated convergence theorem since $|(g-g_N)\log g| \le 3g\,|\log g|$ for all $N \ge N_0$ and $\int g\,|\log g| < \infty$ by assumption. This completes the proof. □

B.3 Proof of Lemma 5

Proof. In view of the concavity and shift-invariance of the differential entropy, without loss of generality, we may assume that $h(Z) < \infty$. Since U and Z are independent, we have

$$I\left(U; U+\varepsilon Z\right) = h\left(U+\varepsilon Z\right) - h\left(U+\varepsilon Z|U\right) = h\left(U+\varepsilon Z\right) - h(Z) - \log\varepsilon.$$

Hence it suffices to show that $\lim_{\varepsilon\to 0} I(U; U+\varepsilon Z) = H(U)$. Notice that $I(U; U+\varepsilon Z) \le H(U)$ for all ε. On the other hand, $(U, U+\varepsilon Z) \xrightarrow{\mathcal{L}} (U, U)$ and $U+\varepsilon Z \xrightarrow{\mathcal{L}} U$ in distribution, by the continuity of the characteristic function. By the weak lower semicontinuity of the divergence, we have

$$\begin{aligned}\liminf_{\varepsilon\to 0} I(U; U+\varepsilon Z) &= \liminf_{\varepsilon\to 0} D\left(P_{U,U+\varepsilon Z}\|P_U P_{U+\varepsilon Z}\right)\\ &\ge D\left(P_{U,U}\|P_U P_U\right) = H(U),\end{aligned}$$

completing the proof. □

B.4 Proof of Lemma 6

Proof. For any $\mathbb{R}^d$-valued discrete random variable U, let $U_{[k]} \triangleq \left(U_{(1)}, \ldots, U_{(k)}\right)$, where $U_{(i)}$ are i.i.d. copies of U. Thus $H\left(U_{[k]}\right) = kH(U)$ and $\sum_{j=1}^m b_j (U_j)_{[k]} = \left(\sum_{j=1}^m b_j U_j\right)_{[k]}$ for any $b_1, \ldots, b_m \in \mathbb{R}$ and any discrete random variables $U_1, \ldots, U_m \in \mathbb{R}^d$.

Let $U_1, \ldots, U_m$ be $\mathbb{R}^d$-valued discrete random variables and $A = (a_{ij}) \in \mathbb{R}^{n\times m}$. Let $\mathcal{U} \subset \mathbb{R}^d$ be a countable set such that $\sum_{i=1}^m a_{ij} U_j \in \mathcal{U}$ for each $i \in [n]$. Let $f_M : \mathbb{R}^{d\times k} \to \mathbb{R}^d$ be given by $f_M(x_1, \ldots, x_k) = \sum_{i=1}^m x_i M^i$ for $M > 0$. Since for any $x = (x_1, \ldots, x_k)$ and $y = (y_1, \ldots, y_k)$ in $\mathcal{U}^k$, there are at most k values of M such that $f_M(x) = f_M(y)$. Since $\mathcal{U}^k$ is countable,

f_M is injective on $\mathcal{U}^k$ for all but at most countably many values of M. Fix an $M_0 > 0$ such that f_{M_0} is injective on $\mathcal{U}^k$ and abbreviate f_{M_0} by f. Let $U_j^{(k)} = f((U_j)_{[k]})$ for each $j \in [m]$. Thus, for each $i \in [n]$,

$$\begin{aligned} H\left(\sum_{j=1}^m b_j U_j^{(k)}\right) &= H\left(\sum_{j=1}^m a_{ij} f\left((U_j)_{[k]}\right)\right) \overset{(a)}{=} H\left(f\left(\sum_{j=1}^m a_{ij}(U_j)_{[k]}\right)\right) \\ &= H\left(f\left(\left(\sum_{j=1}^m a_{ij} U_j\right)_{[k]}\right)\right) \overset{(b)}{=} H\left(\left(\sum_{j=1}^m a_{ij} U_j\right)_{[k]}\right) \\ &= kH\left(\sum_{j=1}^m a_{ij} U_j\right), \end{aligned}$$

where (a) follows from the linearity of f and (b) follows form the injectivity of f on $\mathcal{U}^k$ and the invariance of Shannon entropy under injective maps. □

B.5 Proof of Theorem 3

Proof of Theorem 3. The proof is almost identical to that of Theorem 4. By the structure theorem for connected abelian Lie groups (cf., e.g., [25, Corollary 1.4.21]), G' is isomorphic to $\mathbb{R}^d \times \mathbb{T}^n$. By Lemma 1 and Lemma 3, we only need to prove the theorem for $[0,1]^d \times \mathbb{T}^n$-valued random variables. Along the lines of the proof of Theorem 4, it suffices to establish the counterparts of (2.10) for any $[0,1]^d \times \mathbb{T}^n$-valued continuous X, and (2.11) for any $[0,1]^d \times \mathbb{T}^n$-valued independent and continuous $X_1, \ldots, X_m$, where the quantization operations are defined componentwise by applying the usual uniform quantization (2.1) to the real-valued components of X and the angular quantization (2.9) to the $\mathbb{T}^n$-component of X. The argument is the same as that of Theorem 4, which we omit for concision. □

REFERENCES

[1] I. Z. Ruzsa, *Sums of Finite Sets*. New York, NY: Springer US, 1996.

[2] T. Tao and V. Vu, "Entropy methods," 2005, unpublished notes, http://www.math.ucla.edu/~tao/preprints/Expository/chapter_entropy.dvi.

[3] I. Z. Ruzsa, "Entropy and sumsets," *Random Structures and Algorithms*, vol. 34, pp. 1–10, Jan. 2009.

[4] T. M. Cover and J. A. Thomas, *Elements of Information Theory, 2nd Ed.* New York, NY, USA: Wiley-Interscience, 2006.

[5] T. S. Han, "Nonnegative entropy measures of multivariate symmetric correlations," *Information and Control*, vol. 36, no. 2, pp. 133 – 156, 1978.

[6] A. Lapidoth and G. Pete, "On the entropy of the sum and of the difference of two independent random variables," *Proc. IEEE 25th Conv. IEEEI*, pp. 623–625, December 2008.

[7] M. Madiman, "On the entropy of sums," in *Proceedings of 2008 IEEE Information Theory Workshop*, Porto, Portugal, 2008, pp. 303–307.

[8] T. Tao, "Sumset and inverse sumset theory for Shannon entropy," *Combinatorics, Probability & Computing*, vol. 19, no. 4, pp. 603–639, 2010.

[9] M. Madiman and I. Kontoyiannis, "The entropies of the sum and the difference of two IID random variables are not too different," in *Proceedings of 2010 IEEE International Symposium on Information Theory*, Austin, TX, June 2010, pp. 1369–1372.

[10] M. Madiman, A. W. Marcus, and P. Tetali, "Entropy and set cardinality inequalities for partition-determined functions," *Random Structures & Algorithms*, vol. 40, no. 4, pp. 399–424, 2012.

[11] K. Gyarmati, F. Hennecart, and I. Z. Ruzsa, "Sums and differences of finite sets," *Funct. Approx. Comment. Math.*, vol. 37, no. 1, pp. 175–186, 2007.

[12] I. Kontoyiannis and M. Madiman, "Sumset and inverse sumset inequalities for differential entropy and mutual information," *Information Theory, IEEE Transactions on*, vol. 60, no. 8, pp. 4503–4514, 2014.

[13] M. Madiman and I. Kontoyiannis, "Entropy bounds on abelian groups and the Ruzsa divergence," *arXiv preprint arXiv:1508.04089*, 2015.

[14] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Mathematica Hungarica*, vol. 10, no. 1 – 2, Mar. 1959.

[15] A. V. Makkuva and Y. Wu, "On additive-combinatorial affine inequalities for shannon entropy and differential entropy," in *Information Theory (ISIT), 2016 IEEE International Symposium on*. IEEE, 2016, pp. 1053–1057.

[16] T. H. Chan, "Balanced information inequalities," *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3261–3267, 2003.

[17] I. Z. Ruzsa, "On the number of sums and differences," *Acta Mathematica Hungarica*, vol. 58, no. 3-4, pp. 439–447, 1991.

[18] F. Hennecart, G. Robert, and A. Yudin, "On the number of sums and differences," *Astérisque*, no. 258, pp. 173–178, 1999.

[19] I. Z. Ruzsa, "Sumsets and structure," in *Combinatorial Number Theory and Additive Group Theory*. Basel, Switzerland: Birkhäuser, 2009.

[20] C. Rogers and G. Shephard, "The difference body of a convex body," *Archiv der Mathematik*, vol. 8, no. 3, pp. 220–233, 1957.

[21] I. Csiszár, "Almost independence and secrecy capacity," vol. 32, no. 1, pp. 48–57, 1996.

[22] Y. Polyanskiy and Y. Wu, "Dissipation of information in channels with input constraints," *IEEE Transaction Information Theory*, vol. 62, no. 1, pp. 35–55, Jan. 2016, also arXiv:1405.3629.

[23] Y. Polyanskiy and Y. Wu, "Lecture Notes on Information Theory," Feb 2015, http://www.ifp.illinois.edu/~yihongwu/teaching/itlectures.pdf.

[24] D. Jimenez, L. Wang, and Y. Wang, "White noise hypothesis for uniform quantization errors," *SIAM Journal on Mathematical Analysis*, vol. 38, no. 6, pp. 2042–2056, 2007.

[25] H. Abbaspour and M. A. Moskowitz, *Basic Lie Theory*. World Scientific, 2007.