The Entropies of the Sum and the Difference of Two IID Random Variables are Not Too Different

Mokshay Madiman
Department of Statistics
Yale University
24 Hillhouse Avenue
New Haven, CT 06511, USA.
Email: mokshay.madiman@yale.edu

Ioannis Kontoyiannis
Department of Informatics
Athens University of Economics & Business
Patission 76
Athens 10434, Greece.
Email: yiannis@aueb.gr

***Abstract*—Consider the entropy increase $h(Y+Y')-h(Y)$ of the *sum* of two continuous i.i.d. random variables Y, Y', and the corresponding entropy increase $h(Y-Y')-h(Y)$ of their difference. We show that the ratio between these two quantities always lies between 1/2 and 2. This complements a recent result of Lapidoth and Pete, showing that the difference $h(Y+Y')-h(Y-Y')$ may be arbitrarily large. Corresponding results are discussed for the discrete entropy, and connections are drawn with exciting recent mathematical work in the area of additive combinatorics.**

I. Introduction

Let Y and Y' be independent, identically distributed (i.i.d.) $\mathbb{R}^d$-valued random variables with common density function f (with respect to Lebesgue measure). The differential entropy is denoted by $h(Y) = E[-\log f(Y)]$.[1]

This note is partly motivated by the following question:

Question: How different can $h(Y+Y')$ and $h(Y-Y')$ be?

Lapidoth and Pete [1] recently showed that, given any $M > 0$, there exist i.i.d. random variables Y, Y' of finite differential entropy, such that,

$$h(Y+Y') - h(Y-Y') > M. \tag{1}$$

In other words, the entropies of the sum and difference of two i.i.d. random variables *can differ by an arbitrarily large amount*. They also showed that this result holds in the discrete setting (in fact, they use the discrete version to prove the continuous version).

Our main result, stated next, gives a complementary result:

Theorem 1.1: For any two i.i.d. random variables Y, Y' with finite differential entropy:

$$\frac{1}{2} \le \frac{h(Y+Y') - h(Y)}{h(Y-Y') - h(Y)} \le 2.$$

[1]Throughout, log denotes the natural logarithm. Strictly speaking, we should be referring to Y, Y' are random vectors; in the interest of simplicity, and with a slight abuse of terminology, we still call any such $\mathbb{R}^d$-valued Y a random variable.

In this context, we find that the natural quantities to consider are the differences,

$$\begin{aligned} \Delta_+ &= h(Y+Y') - h(Y) \\ \Delta_- &= h(Y-Y') - h(Y). \end{aligned}$$

Then (1) states that the *difference* $\Delta_+ - \Delta_-$ can be arbitrarily large, while Theorem 1.1 asserts that the *ratio* Δ_+/Δ_- must always lie between $\frac{1}{2}$ and 2.

While the primary motivation for this note was to consider continuous random variables, the main results are also valid in the discrete setting, where they have interesting connections to important problems in additive combinatorics. The field of *additive combinatorics* (see, e.g., [2] for an introduction) is the theory of additive structures in sets equipped with a group structure. The prototypical example is the study of arithmetic progressions in sets of integers, as distinct from the multiplicative structure that underlies prime factorization and much of classical number theory. There have been several major developments and a lot of high-profile mathematical activity in additive combinatorics in recent years, with perhaps the most famous example being the celebrated Green-Tao theorem on the existence of arbitrarily long arithmetic progressions within the primes.

An important collection of tools in the study of additive combinatorics is a variety of *sumset inequalities*. Here, the *sumset* $A+B$ of two discrete subsets A and B of the integers (or any other additive group) is defined as, $A+B = \{a+b : a \in A, b \in B\}$, and a *sumset inequality* is an inequality connecting the cardinality (the number of elements) of $A+B$ with the cardinalities of A and B. Examples of sumset inequalities that have left a deep mark on the field include the Cauchy-Davenport inequality, Freiman's theorem, the Balog-Szemerédi-Gowers theorem, and the Plünnecke-Ruzsa inequalities.

A natural connection between entropy and set cardinalities arises from the fact that the entropy of the uniform distribution on a finite set A is just $\log|A|$, and this is the maximum entropy of any distribution supported on A. The connection between sumset inequalities and analogous inequalities for the entropy of discrete random variables appears to have first been

identified and explored by Imre Ruzsa. In the last couple of years, this connection has also been fleshed out independently and in different directions by Ruzsa [3], Madiman, Marcus and Tetali [4] and Tao [5].

This note is organized as follows. First, in Section II, we describe relevant inequalities from additive combinatorics that may be seen as a set cardinality version of our main result, and discuss inequalities for discrete entropy analogous to them. Then, in Section III, the analogous inequalities for differential entropy are discussed, including why they do not automatically follow from the proof of the discrete case. Section IV contains a brief discussion.

II. Discrete entropy and sumset inequalities

A. Set cardinalities

Throughout this section, A and B are taken to be arbitrary discrete subsets of the integers (or any commutative group).

The following inequality relating sum and difference sets is classical in additive combinatorics,

$$\begin{aligned} \tfrac{1}{2}\big[\log|A-A| - \log|A|\big] &\leq \log|A+A| - \log|A| \\ &\leq 2\big[\log|A-A| - \log|A|\big], \end{aligned}$$

where the difference set $A-B$ is defined in the obvious way as $\{a-b : a \in A, b \in B\}$. The above inequality can be compactly written in terms of the *doubling constant* and *difference constant* of a set. Define the *doubling constant* of A by,

$$\sigma[A] = \frac{|A+A|}{|A|},$$

and the *difference constant* of A by,

$$\delta[A] = \frac{|A-A|}{|A|}.$$

Then the above inequality may be written,

$$\delta[A]^{\frac{1}{2}} \leq \sigma[A] \leq \delta[A]^2. \tag{2}$$

B. Discrete entropy

As mentioned in the Introduction, a significant amount of recent research effort has been devoted to exploring the intriguing idea that any sumset inequality corresponds to an analogous inequality for the entropies of sums of discrete random variables; cf. [3][4][5]. Formally, this translation is performed by replacing discrete sets by independent discrete random variables, and also replacing the log-cardinality of a set by the discrete entropy function.

Let us illustrate this analogy with an example. Let Y, Y' be i.i.d. discrete random variables, and define, the *doubling constant* and the *difference constant* of Y by,

$$\begin{aligned} \sigma[Y] &= \exp\{H(Y+Y') - H(Y)\}, \\ \delta[Y] &= \exp\{H(Y-Y') - H(Y)\}, \end{aligned}$$

respectively, where $H(\cdot)$ denotes the discrete entropy function. Then the entropy analog of the sumset inequality (2) is that,

$$\delta[X]^{\frac{1}{2}} \leq \sigma[X] \leq \delta[X]^2. \tag{3}$$

Note that in this case, and in almost all cases of this correspondence, despite the close analogy between a sumset inequality like the one in (2) and the corresponding entropy version like the one in (3), neither one is a simple consequence of the other. Nevertheless, as emphasized by Tao in [5], there are often deep similarities in their proofs.

The inequality in (3) is equivalently stated in terms of the entropy in Theorem 2.1 below. Although this statement is implicit in recent work (the upper bound was proved in [6] and the lower bound was proved independently by Tao and Vu [7] and by Ruzsa [3]), we find it instructive for our purposes here to outline its proof.

Theorem 2.1: If Y, Y' are i.i.d. discrete random variables, then:

$$\frac{1}{2} \leq \frac{H(Y+Y') - H(Y)}{H(Y-Y') - H(Y)} \leq 2.$$

The proof will be based on the following results.

*Lemma 2.2 (*Submodularity.): [6][5][4]
If $X_0 = F(X_1) = G(X_2)$ and $X_{12} = R(X_1, X_2)$, then:

$$H(X_{12}) + H(X_0) \leq H(X_1) + H(X_2).$$

Proof. By data processing for mutual information and entropy, $H(X_1) + H(X_2) - H(X_{12}) \geq H(X_1) + H(X_2) - H(X_1, X_2) = I(X_1; X_2) \geq I(X_0; X_0) = H(X_0)$. □

C. Ruzsa triangle inequality

*Proposition 2.3 (*Ruzsa triangle inequality.): [5]
If X, Y, Z are independent, then:

$$H(X-Z) \leq H(X-Y) + H(Y-Z) - H(Y).$$

Before giving the proof of the proposition, we recall from [5] that, for any pair X, Y of discrete random variables, the *Ruzsa distance* between X and Y is defined by,

$$\mathrm{dist}_R(X,Y) = H(X'-Y') - \frac{1}{2}H(X') - \frac{1}{2}H(Y'),$$

where $X' \sim X$ and $Y' \sim Y$ are independent. Then, for arbitrary discrete random variables X, Y, Z, the proposition implies that,

$$H(X'-Z') \leq H(X'-Y') + H(Y'-Z') - H(Y'),$$

where (X', Y', Z') are independent, with the same marginals as (X, Y, Z). Rearranging, this yields,

$$\mathrm{dist}_R(X,Z) \leq \mathrm{dist}_R(X,Y) + \mathrm{dist}_R(Y,Z),$$

which explains the name in the proposition.

Proof of Proposition 2.3. By submodularity,

$$H(X,Y,Z) + H(X-Z) \leq H(X-Y, Y-Z) + H(X,Z). \tag{4}$$

Rearranging and using independence,

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

as claimed. □

Replacing Y by $-Y$ and noting that $H(W) = H(-W)$ for any random variable W, this triangle inequality yields:

Corollary 2.4: [7][5] If X, Y, Z are independent, then:

$$H(X - Z) + H(Y) \leq H(X + Y) + H(Y + Z).$$

D. Submodularity of entropy of sums

In a similar vein we also have:

Proposition 2.5: [6][4] If X, Y, Z are independent, then:

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

Proof. Again using independence and data processing for mutual information, the difference, $H(X+Y+Z)-H(Y+Z)$ equals,

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

as required. □

E. Proof of Theorem 2.1

For the upper bound, taking $X, -Y$ and Z i.i.d. in Proposition 2.5, we have,

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

so that,

$$H(X + Z) + H(X) \leq 2H(X - Z),$$

or,

$$H(X + Z) - H(X) \leq 2[H(X - Z) - H(X)],$$

which gives the required upper bound. For the lower bound, Corollary 2.4 with X, Y, Z i.i.d. yields,

$$H(X - Y) + H(X) \leq 2H(X + Y),$$

i.e.,

$$H(X - Y) - H(X) \leq 2[H(X + Y) - H(X)],$$

completing the proof. □

III. Differential entropy

A. The difficulty

Statements or inequalities that hold for discrete entropy frequently fail for differential entropy. Simple examples include the falsity of $h(X) \geq 0$ and of $I(X; X) = h(X)$. Deeper manifestations of such differences are explored in [6]. For instance, while it is trivial that $H(X_1+X_2) \leq H(X_1)+H(X_2)$, the differential entropy of a sum does *not* satisfy this simple subadditivity. Nevertheless, as we will see below, despite this difference between the discrete and continuous settings, a certain version of the submodularity property of entropy of sums continues to hold in the continuous setting.

Many of the above arguments and results leading to the derivation of Theorem 2.1 hold for differential entropy h in place of the discrete entropy H. The main obstacle in generalizing Theorem 2.1 to continuous random variables as in Theorem 1.1 is the use of data processing for entropy "$H(f(X)) \leq H(X)$," in the proof of the generalized submodularity of Lemma 2.2. Therefore, in order to prove the natural analog of the Rusza triangle inequality (Proposition 2.3) for continuous random variables, we need an alternative proof the continuous version of the bound in (4).

B. Continuous analogue of Ruzsa triangle inequality

To prove the analog of (4) we first note that,

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

Therefore, the expression,

$$h(X - Y, Y - Z) + h(X, Z) - h(X, Y, Z),$$

equals,

$$\begin{aligned} & h(X, Z) - h(X|X - Y, Y - Z) \\ = \; & h(X) - h(X|X - Y, Y - Z) + h(Z) \\ = \; & I(X; (X - Y, Y - Z)) + h(Z) \\ \geq \; & I(X; X - Z) + h(Z) \\ = \; & h(X - Z) - h(X - Z|X) + h(Z) \\ = \; & h(X - Z) - h(-Z|X) + h(Z) \\ = \; & h(X - Z), \end{aligned}$$

as required, where the inequality holds by data processing for mutual information. Therefore:

Proposition 3.1: The Ruzsa triangle inequality (Proposition 2.3) and the result of Corollary 2.4 remain valid as stated, with the discrete entropy replaced by differential entropy.

C. Proof of the upper bound in Theorem 2.1

The following lemma on the submodularity of differential entropy is not hard to prove along the lines of the proof of Proposition 2.5 above, although the justification changes (one needs to use translation-invariance of Lebesgue measure instead of invariance of discrete entropy to bijective transformations). To the best of our knowledge, it first appeared in [6].

Lemma 3.2: [6] If X, Y, Y' are independent random variables, then:

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

Since the entropy of a sum is no smaller than entropy of each summand, Lemma 3.2 gives,

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

In the case where $Y \stackrel{(d)}{=} Y'$,

$$h(Y+Y') + h(X) \leq 2h(X+Y),$$
$$\text{or} \quad h(Y+Y') - h(X) \leq 2[h(X+Y) - h(X)].$$

Taking $X \stackrel{(d)}{=} -Y$ now gives,

$$h(Y+Y') - h(Y) \leq 2[h(Y-Y') - h(Y)],$$

which is the desired upper bound.

D. Proof of the lower bound in Theorem 2.1

As stated in Proposition 3.1, the analog of Corollary 2.4 also holds for continuous random variables; that is:

Proposition 3.3: If X, Y, Y' are independent random variables, then:

$$h(Y-Y') + h(X) \leq h(X+Y) + h(X+Y').$$

The result of the proposition can equivalently be stated as:

$$h(Y-Y') + h(Z) \leq h(Y-Z) + h(Y'-Z). \tag{5}$$

This is true for general independent random variables, but in our case becomes,

$$h(Y-Y') - h(Z) \leq 2[h(Y-Z) - h(Z)].$$

Now taking $Z \stackrel{(d)}{=} -X$ gives,

$$h(Y-Y') - h(X) \leq 2[h(Y+X) - h(X)],$$

as required.

IV. Discussion

While Theorem 1.1 was stated for continuous $\mathbb{R}^d$-valued random variables, it holds for continuous random variables taking values in any locally compact topological group $(\mathcal{G}, +)$, provided the random variables are independent, and are absolutely continuous with respect to Haar measure on $\mathcal{G}$. In this case, $Y - Y'$ refers to the group addition of Y and $-Y'$, the latter being the additive inverse of Y. Note that the entropy in this setting would still be defined by $h(Y) = E[-\log f(Y)]$, with f being the density of the probability measure on $\mathcal{G}$ describing the distribution of Y with respect to Haar measure.

There are several directions in which the simple observation described in this note may be extended; as a first step, it would be interesting to develop an analog when Y and Y' are independent but not necessarily identically distributed.

Acknowledgment

MM thanks Amos Lapidoth for communicating the note [1]. MM is also grateful to Prasad Tetali and Adam Marcus for the indirect influence of their joint work [4] on entropy inequalities in additive combinatorics.

References

[1] A. Lapidoth and G. Pete, "On the entropy of the sum and of the difference of two independent random variables," *Proc. IEEEI 2008, Eilat, Israel*, 2008.

[2] T. Tao and V. Vu, *Additive combinatorics*, ser. Cambridge Studies in Advanced Mathematics. Cambridge: Cambridge University Press, 2006, vol. 105.

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

[4] M. Madiman, A. Marcus, and P. Tetali, "Entropy and set cardinality inequalities for partition-determined functions, with applications to sumsets," *Preprint*, 2008. [Online]. Available: http://arxiv.org/abs/0901.0055

[5] T. Tao, "Sumset and inverse sumset theory for Shannon entropy," *Preprint,* arXiv:0906.4387v1, 2009.

[6] M. Madiman, "On the entropy of sums," in *Proc. IEEE Inform. Theory Workshop*, Porto, Portugal, 2008.

[7] T. Tao and V. Vu, "Entropy methods," *Unpublished*. [Online]. Available: http://www.math.ucla.edu/∼tao/