Discrete Generalised Entropy Power Inequalities for Log-Concave Random Variables

Lampros Gavalakis
LAMA, Université Gustave Eiffel, France
lampros.gavalakis@univ-eiffel.fr

***Abstract*—We prove a discrete analogue of the generalised entropy power inequality for log-concave random variables on the integers. As a special case, we show that a conjecture of Tao (2010) holds true for log-concave random variables on the integers:**

$$H(X_1+\cdots+X_{n+1}) \geq H(X_1+\cdots+X_n)+\frac{1}{2}\log\Big(\frac{n+1}{n}\Big)-o(1),$$

where the $o(1)$-term vanishes as $H(X_1)\to\infty$. Explicit, finite bounds for the error term are provided, which are exponential in $H(X_1)$.

A full version of this paper is available at: arXiv:2210.06624.

I. Introduction

A. Generalised Entropy Power Inequalities for Continuous Random Variables

Let X, Y be two independent random variables with densities in $\mathbb{R}$. We denote by $h(X)$ the differential entropy of X, that is $h(X)=-\int_{\mathbb{R}} f(x)\log f(x)dx$, where f is the density of X and similarly for Y. Throughout this paper '$\log$' denotes the natural logarithm.

As it is well known within information theory, the celebrated Entropy Power Inequality (EPI) goes back to Shannon [1] and was first proven in full generality by Stam [2]. For real-valued random variables, it asserts that

$$N(X+Y)\geq N(X)+N(Y), \tag{1}$$

where $N(X)$ is the *entropy power* of X: $N(X)=\frac{1}{2\pi e}e^{2h(X)}$. If X_1, X_2 are identically distributed, (1) can be rewritten as

$$h(X_1+X_2)\geq h(X_1)+\frac{1}{2}\log 2. \tag{2}$$

The EPI was generalised by Artstein et al. [3]: If $\{X_i\}_{i=1}^{n+1}$ are i.i.d. then

$$h\Big(\frac{1}{\sqrt{n+1}}\sum_{i=1}^{n+1}X_i\Big)\geq h\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i\Big). \tag{3}$$

This is the monotonic increase of entropy along the central limit theorem [4].

Simpler, information-theoretic proofs appeared later in [5], [6]. In particular, the following generalisation was proven in [6]: If $X_1,\ldots,X_n$ are independent continuous random variables and $\mathcal{C}$ is an arbitrary collection of subsets of $\{1,\ldots,n\}$, then

$$e^{2h(X_1+\cdots+X_n)}\geq\frac{1}{r}\sum_{S\in\mathcal{C}}e^{2h(\sum_{j\in S}X_j)}, \tag{4}$$

where r is the maximum number of sets in $\mathcal{C}$ in which any one index appears. Our main result (Theorem 1) is a discrete analogue of (4) for identically distributed, discrete log-concave random variables.

B. Discrete Analogues

There has been interest in formulating discrete analogues of the EPI from various perspectives [7]–[10]. It is not hard to see that the exact statement (2) can not hold for all discrete random variables by considering deterministic (or even close to deterministic) random variables.

Sumset theory for entropy. Suppose G is an additive (abelian) group and X is a random variable supported on a discrete (finite or countable) subset A of G with probability mass function (p.m.f.) p on G. We denote by $H(X)$ the (discrete) entropy of X, that is $H(X)=-\sum_{x\in A}p(x)\log p(x)$. Perhaps most notably, Tao [11] proved that if G is torsion-free and X takes finitely many values then

$$H(X_1+X_2)\geq H(X_1)+\frac{1}{2}\log 2-o(1), \tag{5}$$

where X_1, X_2 are independent copies of X and the $o(1)$-term vanishes as the entropy of X tends to infinity. That work explores the connection between additive combinatorics and entropy, which was identified by Tao and Vu in the unpublished notes [12] and by Ruzsa [13]. Partly, the intuition stems from the fact that random variables in G may be associated with subsets A of G: By the *asymptotic equipartition property* [14], there is a set A_n (the typical set) such that if $X_1,\ldots,X_n$ are i.i.d. copies of X, then X_1^n is approximately uniformly distributed on A_n and $|A_n|=e^{n(H(X)+o(1))}$. Given an inequality involving cardinalities of sumsets, it is natural to guess that a counterpart statement holds true for random variables if the logarithm of the cardinality is replaced by the entropy.

Exploring this connection, Tao [11] proved an inverse theorem for entropy, which characterises random variables for which the addition of an independent copy does not increase

[0]L.G. has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 101034255.

the entropy by much. This is the entropic analogue of the inverse Freiman theorem [15] from additive combinatorics, which characterises sets for which the sumset is not much bigger than the set itself. The discrete EPI (5) is a consequence of the inverse theorem for entropy.

Furthermore, it was conjectured in [11] that for any $n \geq 2$ and $\epsilon > 0$

$$H(X_1+\cdots+X_{n+1}) \geq H(X_1+\cdots+X_n)+\frac{1}{2}\log\Big(\frac{n+1}{n}\Big)-\epsilon, \tag{6}$$

provided that $H(X)$ is large enough depending on n and ϵ, where $\{X_i\}_{i=1}^{n+1}$ are i.i.d. copies of X.

As a special case of our main result (Corollary 1) we show that conjecture (6) holds true for log-concave random variables on the integers with an exact lower bound on $H(X_1)$ in terms of n and ϵ.

Equation (6) may be seen as a discrete analogue of (3). A discrete entropic central limit theorem was recently established in [16]. A discussion relating the above conjecture to the convergence of discrete entropy to its maximum in analogy with (3) may be found there.

The result (5) is asymptotic and no explicit lower bound on $H(X)$ was given in [11]. Besides the theoretical interest of this result, there has been interest in establishing a finite bound. The motivation comes, among others, from *compressed sensing* [7]. In the latter work, the following entopy power inequality over the integers was proved: For any random variable X taking values in $\mathbb{Z}$,

$$H(X_1 + X_2) \geq H(X_1) + g(H(X)), \tag{7}$$

where, again, X_1, X_2 are independent copies of X and $g : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly increasing function such that $\lim_{x\to\infty} g(x) = \frac{1}{8\log 2}$ and $g(x) = 0$ if and only if $x = 0$. The polar coding techniques [17], [18] were extended to arithmetic over $\mathbb{Z}$ rather than arithmetic over $\mathbb{F}_q$. To this end, an appropriate martingale was constructed, whose values represent certain conditional entropies. The convergence of that martingale to zero implies what is called an *absorption phenomenon*, analogously to the polarisation phenomenon of [17], [18]. The fact that $g(x) = 0$ if and only if $x = 0$ was then used to prove convergence to 0. Although the function g was explicitly given and sharp estimates for its behaviour were provided, the constant $\frac{1}{8\log 2} < \frac{1}{2}\log 2$ is not optimal.

In the latter application, it is the regime $H(X_1 + X_2) - H(X_1) \approx 0$, or, as it is shown, equivalently $H(X) \approx 0$ that is important. However, the suboptimality of the limit in the $H(X_1) \to \infty$ regime, naturally raises the question of how much the quantitative bounds can be improved, while obtaining the 'right' constant, even under additional assumptions.

The focus of the present work is in the large entropy regime: For $n = 1$, our main result merely shows that if $H(X_1+X_2)-H(X_1)$ is close to zero, then $H(X_1)$ is less than an absolute constant, whereas (7) already implies that in this case $H(X_1)$ is close to 0 as well. Neverthelss, for $H(X_1) \to \infty$, our proofs yield explicit rates for the $o(1)$-terms, which are exponential in $H(X_1)$.

The class of discrete log-concave distributions has been considered recently by Bobkov et al. [19] in connection with the EPI. In particular, discrete analogues of (1) up to a multiplicative constant were proved for this class and generalised for all Rényi entropy powers. Furthermore, sharp upper and lower bounds on the maximum probability of discrete log-concave random variables in terms of their variance were provided, which we are going to use in the proofs (see Lemma 1 below).

We should mention that log-concavity is a strong assumption: It implies connected support set, meaning that $p(k) = 0$ implies $p(k-1) = 0$ or $p(k+1) = 0$, so we can not have 'gaps'. In addition, log-concave random variables have moments of all orders, e.g. [20]. On the other hand, log-concave distributions form a natural class to consider for applications, as many important distributions are log-concave, e.g. the Bernoulli, Poisson, geometric, negative binomial, uniform and others.

Finally, let us mention that another discrete EPI for the special family of uniform distributions on $\mathbb{Z}$ was proved in [10]. Interestingly, the limiting constant is $\frac{1}{2}$ if we restrict to uniform random variables on the integers.

C. Main results and proof ideas

Our main result is:

Theorem 1. *Let $n \geq 1$ and suppose $X_1, \ldots, X_n$ are i.i.d. log-concave random variables on the integers with common variance σ^2. Let $\mathcal{C}$ be an arbitrary collection of subsets of $\{1, \ldots, n\}$ and r be the maximum number of sets in $\mathcal{C}$ in which any one index appears. Then,*

$$e^{2H(X_1+\cdots+X_n)} \geq \frac{1}{r}\sum_{S\in\mathcal{C}} e^{2H(\sum_{j\in S} X_j)-o(1)}, \tag{8}$$

where the $o(1)$-term vanishes as $\sigma \to \infty$ depending on n.

Remark 1. *In fact, we prove a stronger result, Theorem 2 below. A finite bound for the $o(1)$-term may also be computed from* (13)*, which is of order $O\Big(\frac{2^n}{\sigma\sqrt{n}}\log(2^n\sigma\sqrt{n})\Big)$.*

An important step in the proof of (5) in [11] is reduction to the continuous setting by approximation of the continuous density with a discrete p.m.f.; we briefly outline these key points from that proof below as we are going to take a similar approach.

The first step in the proof method of [11, Theorem 1.9] is to assume that $H(X_1 + X_2) \leq H(X_1) + \frac{1}{2}\log 2 - \epsilon$. Then, because of [11, Theorem 1.8], proving the result for random variables X that can be expressed as a sum $Z + U$, where Z is a random variable with entropy $O(1)$ and U is a uniform on a large arithmetic progression, say P, suffices to get a contradiction. Such random variables satisfy, for every x,

$$\mathbb{P}(X = x) \leq \frac{C}{|P|} \tag{9}$$

for some absolute constant C. Using tools from sumset theory, it is shown that it suffices to consider random variables that

take values in a finite subset of the integers. For such random variables that satisfy (9), the smoothness property

$$\|p_{X_1+X_2} - p_{X_1+X_2+1}\|_{\mathrm{TV}} \to 0 \tag{10}$$

as $H(X) \to \infty$ is established, where $p_{X_1+X_2}, p_{X_1+X_2+1}$ are the p.m.f.s of $X_1 + X_2$ and $X_1 + X_2 + 1$ respectively and $\|\cdot\|_{\mathrm{TV}}$ is the total variation distance defined in (16) below. Using this, it is shown that

$$h(X_1 + X_2 + U_1 + U_2) = H(X_1 + X_2) + o(1), \tag{11}$$

as $H(X) \to \infty$, where U_1, U_2 are independent continuous uniforms on $(0, 1)$. The EPI for continuous random variables is then invoked.

Our proofs lack any additive combinatorial arguments as we already work with random variables, whose support set is an arithmetic progression over the integers (although it may have infinite cardinality). An important technical step in our case is to show that any log-concave random variable X on the integers satisfies

$$\|p_X - p_{X+1}\|_{\mathrm{TV}} \to 0$$

as $H(X) \to \infty$. Using this we show the following generalisation of (11):

Theorem 2. *Let $n \geq 1$ and suppose $X_1, \ldots, X_n$ are i.i.d. log-concave random variables on the integers with common variance σ^2. Let $U_1, \ldots, U_n$ be continuous i.i.d. uniforms on $(0, 1)$. Then*

$$h(X_1+\cdots+X_n+U_1+\cdots+U_n) = H(X_1+\cdots+X_n)+o(1), \tag{12}$$

where the $o(1)$-term vanishes as $\sigma^2 \to \infty$ depending on n. In fact, this term can be bounded absolutely by

$$2^{n+6}e^{-(\sqrt{n}\sigma)^{1/5}}(\sqrt{n}\sigma)^3 + \frac{2^{n+2}}{\sigma\sqrt{n}}\log(2^{n+2}\sigma\sqrt{n}) + \frac{\log n\sigma^2}{8n\sigma^2}, \tag{13}$$

provided that $\sigma > \max\{2^{n+2}/\sqrt{n}, 3^7/\sqrt{n}\}$.

Our main tools are first, to approximate the density of the log-concave sum convolved with the sum of n continuous uniforms with the discrete p.m.f. (Lemma 2) and second, to show a type of concentration for the "information density", $-\log p(S_n)$, using Lemma 3. It is a standard argument to show that log-concave p.m.f.s have exponential tails, since the sum of the probabilities is convergent. Lemma 3 is a slight improvement in that it provides a bound for the ratio depending on the variance.

By an application of the generalised EPI for continuous random variables (4), we obtain our main result.

In particular, choosing $\mathcal{C}$ in Theorem 1 to be the collection of all subsets of cardinality $n-1$, we show that the conjecture (6) is true for log-concave random variables on the integers, with an explicit dependence between $H(X)$ and ϵ, which can be computed from (13):

Corollary 1. *Let $n \geq 1$ and $\epsilon > 0$. Suppose $X_1, \ldots, X_n$ are i.i.d. log-concave random variables on the integers. Then if $H(X_1)$ is sufficiently large depending on n and ϵ,*

$$H(X_1+\cdots+X_{n+1}) \geq H(X_1+\cdots+X_n)+\frac{1}{2}\log\Big(\frac{n+1}{n}\Big)-\epsilon. \tag{14}$$

In fact, for (14) to hold it suffices to take $H(X_1) \geq \log\frac{2}{\epsilon} + \log\log\frac{2}{\epsilon} + n + 27$.

Note that $H(X) \to \infty$ implies $\sigma^2 \to \infty$. For the class of log-concave random variables these two statements are actually equivalent, e.g. by Lemma 1 and Proposition 3.

The proofs of Theorems 1, 2 and Corollary 1 are given in Section III. Before that, in Section II below, we state some preliminary facts about discrete, log-concave random variables. The proofs of those, as well as of the three Lemmas in the beginning of Section III, are omitted and may be found in the full version of this paper.

For $n = 1$, the lower bound for $H(X_1)$ given by Corollary 1 for the case of log-concave random variables on the integers is a significant improvement on the lower bound that can be obtained from the proof given in [11] for discrete random variables in a torsion free group, which is $\Omega\Big(\frac{1}{\epsilon}^{\frac{1}{\epsilon}^{\frac{1}{\epsilon}}}\Big)$.

Finally, let us note that the method of adding a continuous uniform to an integer-valued random variable has also been recently used in [21] and can be traced back to Massey [22]. Here we generalise this method by adding n independent uniforms in order to exploit the "sum of i.i.d." structure and approximate the continuous density with the discrete one, when the variance is large. The log-concavity assumption seems to be essential for this approximation. In other words, although the conclusion of Theorem 2 holds true for all integer-valued random variables when $n = 1$, it seems to be quite strong for $n \geq 2$. Although we suspect that the assumption of log-concavity may be relaxed, we do not expect it to hold in much greater generality; we believe that some structural conditions on the random variables should be necessary.

II. Notation and Preliminaries

For a random variable X with p.m.f. p on the integers denote

$$q := \sum_{k\in\mathbb{Z}} \min\{p(k), p(k+1)\}. \tag{15}$$

The parameter q defined above plays an important role in a technique known as Bernoulli part decomposition, which has been used in [23]–[25] to prove local limit theorems. It was also used in [16] to prove the discrete entropic CLT mentioned in the Introduction.

Here we use $1 - q$ as a measure of smoothness of a p.m.f. on the integers. In what follows we will also write $q(p)$ to emphasise the dependence on the p.m.f..

For two p.m.f.s on the integers p_1 and p_2, we use the notation

$$\|p_1 - p_2\|_1 := \sum_{k\in\mathbb{Z}} |p_1(k) - p_2(k)|$$

for the ℓ_1-distance between p_1 and p_2 and

$$\|p_1 - p_2\|_{\mathrm{TV}} := \frac{1}{2}\|p_1 - p_2\|_1 \tag{16}$$

for the total variation distance.

Proposition 1. *Suppose X has p.m.f. p_X on $\mathbb{Z}$ and let $q = \sum_{k\in\mathbb{Z}} \min\{p_X(k), p_X(k+1)\}$. Then*

$$\|p_X - p_{X+1}\|_{\mathrm{TV}} = (1-q).$$

A consequence of Proposition 1 is that the parameter q, defined in (15), does not decrease on convolution:

Corollary 2. *Let p_1, p_2 be two p.m.f.s on the integers. Then*

$$q(p_1 * p_2) \geq \max\{q(p_1), q(p_2)\}.$$

A p.m.f. p on $\mathbb{Z}$ is called *log-concave*, if for any $k \in \mathbb{Z}$

$$p(k)^2 \geq p(k-1)p(k+1). \tag{17}$$

If a random variable X is distributed according to a log-concave p.m.f. we say that X is log-concave. Throughout we suppose that $X_1, \ldots, X_n$ are i.i.d. random variables having a log-concave p.m.f. on the integers, p, common variance σ^2 and denote their sum with S_n. Also, we denote

$$p_{\max} = p_{\max}(X) := \sup_k \mathbb{P}(X = k)$$

and write

$$N_{\max} = N_{\max}(X) := \max\{k \in \mathbb{Z} : p(k) = p_{\max}\}, \tag{18}$$

i.e. $N_{\max}$ is the last $k \in \mathbb{Z}$ for which the maximum probability is achieved. We will make use of the following bound, which follows immediately from [19, Theorem 1.1]:

Lemma 1. *Suppose X has discrete log-concave distribution with $\sigma^2 = \mathrm{Var}(X) \geq 1$. Then*

$$\frac{1}{4\sigma} \leq p_{\max} \leq \frac{1}{\sigma}. \tag{19}$$

Proposition 2. *Let X be a log-concave random variable on the integers with mean $\mu \in \mathbb{R}$ and variance σ^2, and let $\delta > 0$. Then, if $\sigma > 4^{1/2\delta}$,*

$$\mu - \sigma^{3/2+\delta} - 1 < N_{\max} < \mu + \sigma^{3/2+\delta} + 1 \tag{20}$$

Below we show that for any integer-valued random variable, $q \to 1$ implies $H(X) \to \infty$. It is not hard to see that the converse is not alway true, i.e. $H(X) \to \infty$ does not necessarily imply $q \to 1$: Consider a random variable with a mass of $\frac{1}{2}$ at zero and all other probabilities equal on an increasingly large subset of $\mathbb{Z}$. Nevertheless, using Lemma 1, we show that if X is log-concave this implication is true.

Proposition 3. *Suppose that the random variable X has p.m.f. p_X on the integers and let $q = q(p_X)$ as above. Then*

(i) $e^{-H(X)} \leq 1 - q$.

(ii) If p_X is log-concave, then $1 - q = p_{\max} \leq \frac{1}{\sigma}$.

III. Proofs of Theorems 1, 2 and Corollary 1

Let $U^{(n)} := \sum_{i=1}^n U_i$, where U_i are i.i.d. continuous uniforms on $(0,1)$. Let $f_{S_n+U^{(n)}}$ denote the density of $S_n + U^{(n)}$. We approximate $f_{S_n+U^{(n)}}$ with the p.m.f., say p_{S_n}, of S_n. We recall that the class of discrete log-concave distributions is closed under convolution [26] and hence the following lemma may be applied to S_n.

Lemma 2. *Let S be a log-concave random variable on the integers with variance $\sigma^2 = \mathrm{Var}(S)$ and, for any $n \geq 2$, denote by $f_{S+U^{(n)}}$ the density of $S + U^{(n)}$ on the real line. Then for any $n \geq 2$, each $k \in \mathbb{Z}$ and $x \in [k, k+1)$,*

$$f_{S+U^{(n)}}(x) = p_S(k) + g(k), \tag{21}$$

where

$$\sum_{k'\in\mathbb{Z}} |g(k')| \leq (2^n - 2)\frac{1}{\sigma}. \tag{22}$$

Lemma 3. *Let X be a log-concave random variable on the integers with p.m.f. p, mean zero and variance σ^2, and let $0 < \epsilon < 1/2$. If $\sigma \geq \max\{3^{1/\epsilon}, (12e^3)^{1/(1-2\epsilon)}\}$, there is an $N_0 \in \{N_{\max}, \ldots, N_{\max} + 2\lceil\sigma^2\rceil\}$ such that, for each $k \geq N_0$,*

$$p(k+1) \leq \left(1 - \frac{1}{\sigma^{2-\epsilon}}\right)p(k). \tag{23}$$

Similarly, there is an $N_0^- \in \{N_{\max} - 2\lceil\sigma^2\rceil, \ldots, N_{\max}\}$ such that, for each $k \leq N_0$,

$$p(k-1) \leq \left(1 - \frac{1}{\sigma^{2-\epsilon}}\right)p(k).$$

Finally, we need the following Taylor-type estimate. A similar estimate was used in [11].

Lemma 4. *Let $D, M \geq 1$ and, for $x > 0$, consider $G(x) = F(x) - x\log M$, where $F(x) = -x\log x$. Then, for $0 \leq a, b \leq \frac{D}{M}$ and any $0 < \mu < \frac{1}{e}$, we have the estimate*

$$|G(b) - G(a)| \leq \frac{2\mu}{M}\log\frac{1}{\mu} + |b-a|\log\frac{eD}{\mu}. \tag{24}$$

We are now ready to give the proof of our main technical tool, Theorem 2:

Proof of Theorem 2. Assume without loss of generality that X_1 has zero mean. Let $F(x) = x\log\frac{1}{x}, x > 0$ and note that $F(x)$ is non-decreasing for $x \leq 1/e$. As before denote $S_n = \sum_{i=1}^n X_i$, $U^{(n)} = \sum_{i=1}^n U_i$ and let $f_{S_n+U^{(n)}}$ be the density of $S_n + U^{(n)}$ on the reals. We have

$$\begin{aligned} &h(X_1 + \cdots + X_n + U_1 + \cdots + U_n) \\ &= \sum_{k\in(-5n\sigma^2, 5n\sigma^2)} \int_{[k,k+1)} F(f_{S_n+U^{(n)}}(x))dx \\ &+ \sum_{|k|\geq 5n\sigma^2} \int_{[k,k+1)} F(f_{S_n+U^{(n)}}(x))dx. \end{aligned} \tag{25}$$

First we will show that the "entropy tails", i.e. the second term in (25), vanish as σ^2 grows large.

For $k \geq 5n\sigma^2$, we have $p_{S_n}(k+1) \leq p_{S_n}(k)$, since by Proposition 2 applied to the log-concave random variable S_n,

$N_{\max} \le n\sigma^2+1$ as long as $\sqrt{n}\sigma > 4$. Thus, proceeding as in the proof of Lemma 2, we can show that for $k \ge 5n\sigma^2$ and $x \in [k, k+1)$, $f_{S_n+U^{(n)}}(x) \le 2^n p_{S_n}(k-n+1)$. Hence, for

$$\sigma > \frac{2^n}{\sqrt{n}}e, \tag{26}$$

we have

$$\begin{aligned}
0 &\le \sum_{k\ge 5n\sigma^2} \int_{[k,k+1)} F(f_{S_n+U^{(n)}}(x))dx && (27)\\
&\le \sum_{k\ge 5n\sigma^2} F\big(2^n p_{S_n}(k-n+1)\big) && (28)\\
&= \sum_{k\ge 5n\sigma^2} 2^n p_{S_n}(k-n+1)\log\frac{1}{2^n p_{S_n}(k-n+1)} && (29)\\
&\le 2^n \frac{1}{\sqrt{n}\sigma} \sum_{k\ge 5n\sigma^2} \theta^{k-4\lceil n\sigma^2\rceil}\log\frac{\sqrt{n}\sigma}{2^n\theta^{k-4\lceil n\sigma^2\rceil}} && (30)\\
&= 2^n\frac{\log\frac{1}{\theta}}{\sqrt{n}\sigma}\sum_{m\ge n\sigma^2} m\theta^m + 2^n\frac{\log\frac{\sqrt{n}\sigma}{2^n}}{\sqrt{n}\sigma}\sum_{m\ge n\sigma^2}\theta^m && (31)\\
&\le 2^{n+1}\frac{\log\sqrt{n}\sigma}{\sqrt{n}\sigma}\Big[\frac{\theta^{\lceil n\sigma^2\rceil+1}}{(1-\theta)^2} + \lceil n\sigma^2\rceil\frac{\theta^{\lceil n\sigma^2\rceil}}{1-\theta} + \frac{\theta^{\lceil n\sigma^2\rceil}}{1-\theta}\Big] && (32)\\
&\le 2^{n+1}\frac{\log\sqrt{n}\sigma}{\sqrt{n}\sigma}e^{-(\sqrt{n}\sigma)^\epsilon}\Big[(\sqrt{n}\sigma)^{4-2\epsilon}\\
&\quad + (n\sigma^2+1)(\sqrt{n}\sigma)^{2-\epsilon} + (\sqrt{n}\sigma)^{2-\epsilon}\Big] && (33)\\
&\le 2^{n+3}\frac{\log\sqrt{n}\sigma}{\sqrt{n}\sigma}e^{-(\sqrt{n}\sigma)^\epsilon}(\sqrt{n}\sigma)^{4-\epsilon} && (34)\\
&\le 2^{n+4}e^{-(\sqrt{n}\sigma)^{1/5}}(\sqrt{n}\sigma)^3. && (35)
\end{aligned}$$

Here (30) holds for

$$\sqrt{n}\sigma > 3^7 \tag{36}$$

with $\theta = 1-\frac{1}{(\sqrt{n}\sigma)^{2-\epsilon}} = 1-\frac{1}{(\sqrt{n}\sigma)^{9/5}}$, where we have used Lemma 3 with $\epsilon = 1/5$ (which makes the assumption approximately minimal). In particular, repeated application of (23) yields

$p_{S_n}(k-n+1) \le \theta^{k-4\lceil n\sigma^2\rceil} p_{S_n}\big(4\lceil n\sigma^2\rceil - n+1\big) \le \frac{\theta^{k-4\lceil n\sigma^2\rceil}}{\sqrt{n}\sigma}$.

The left tail can be bounded above by (35) in the same way, using the second part of Lemma 3.

On the other hand, the first term in (25) is equal to

$$\sum_{|k|<5n\sigma^2}\int_{[k,k+1)}\Big[F(f_{S_n+U^{(n)}}(x)) - f_{S_n+U^{(n)}}(x)\log n\sigma^2\Big]dx + \log n\sigma^2\mathbb{P}\big(S_n+U^{(n)}\in(-5n\sigma^2,5n\sigma^2)\big). \tag{37}$$

Now we will apply the estimate of Lemma 4 with $\mu = \frac{1}{\sigma\sqrt{n}}, D = 2^n\sqrt{n}\sigma, M = n\sigma^2, a = f_{S_n+U^{(n)}}(x)$ and $b = p_{S_n}(k)$ to the integrand in (37). Thus, using Lemma 2, we obtain that (37) is equal to

$$\log n\sigma^2\mathbb{P}\big(S_n+U^{(n)}\in(-5n\sigma^2,5n\sigma^2)\big) + \sum_{k\in(-5n\sigma^2,5n\sigma^2)} F(p_{S_n}(k)) - p_{S_n}(k)\log n\sigma^2 \tag{38}$$

plus a term, which can be bounded absolutely by

$$\begin{aligned}
&\frac{2\log(\sigma\sqrt{n})}{\sigma^3 n\sqrt{n}} + \sum_k g(k)\Big(\log(\sigma\sqrt{n}) + \log(e2^n\sigma\sqrt{n})\Big) && (39)\\
&\le \frac{2\log(\sigma\sqrt{n})}{\sigma^3 n\sqrt{n}} + \frac{2^{n+1}}{\sigma\sqrt{n}}\log(e2^n\sigma\sqrt{n}) && (40)\\
&\le \frac{2^{n+2}}{\sigma\sqrt{n}}\log(2^{n+2}\sigma\sqrt{n}), && (41)
\end{aligned}$$

where $g(k)$ is given by Lemma 2 applied to the log-concave random variable S_n and therefore $\sum_k g(k) \le \frac{2^n}{\sigma\sqrt{n}}$. Finally, (38) is equal to

$$H(S_n) + \log n\sigma^2\Big[\mathbb{P}\big(S_n+U^{(n)}\in(-5n\sigma^2,5n\sigma^2)\big) - \mathbb{P}\big(S_n\in(-5n\sigma^2,5n\sigma^2)\big)\Big] - \sum_{|k|>5n\sigma^2} F(p_{S_n}(k)), \tag{42}$$

where the discrete entropy tails, i.e. the last term, can be bounded above in a similar way as the continuous tails (see (29)) by two times the upper bound in (35). Moreover, both probabilities in the brackets can be bounded below crudely by

$$1 - \frac{1}{8n\sigma^2}, \tag{43}$$

since by Chebyshev's inequality

$$\begin{aligned}
&\mathbb{P}\big(S_n+U^{(n)}\notin(-5n\sigma^2,5n\sigma^2)\big)\\
&\le \mathbb{P}\big(|S_n+U^{(n)}-\frac{n}{2}|>4.5n\sigma^2\big) \le \frac{1}{8n\sigma^2},
\end{aligned}$$

where we have used that $\sigma^2 > n$ by assumption and $\mathrm{Var}(U^{(n)}) = \frac{n}{12} \le n$. Putting all these together,

$$\begin{aligned}
\big|h(S_n+U^{(n)}) - H(S_n)\big| &\le 2^{n+6}e^{-(\sqrt{n}\sigma)^{1/5}}(\sqrt{n}\sigma)^3\\
&+ \frac{2^{n+2}}{\sigma\sqrt{n}}\log(2^{n+2}\sigma\sqrt{n}) + \frac{\log n\sigma^2}{8n\sigma^2} \qquad (44)
\end{aligned}$$

as long as $\sigma > \max\{2^{n+2}/\sqrt{n}, 3^7/\sqrt{n}\}$.

□

Proof of Theorem 1. Follows from the generalised EPI (4) applied to $X_1+U_1+\cdots+X_n+U_n$ and Theorem 2. □

Proof of Corollary 1. Let $U_1,\dots,U_n$ be continuous i.i.d. uniforms on $(0,1)$. Then, by the generalised entropy power inequality for continuous random variables (3) and the scaling property of differential entropy [14]

$$\begin{aligned}
&h(X_1+\cdots+X_{n+1}+U_1+\cdots+U_{n+1})\\
&\ge h(X_1+\cdots+X_n+U_1+\cdots+U_n) + \tfrac{1}{2}\log\Big(\tfrac{n+1}{n}\Big).
\end{aligned}$$

Now, since [14]

$$H(X_1) = h(X_1+U_1) \le \frac{1}{2}\log(2\pi e(\sigma^2+1/12)), \tag{45}$$

we have, assuming $\sigma > 0.275$, $e^{H(X_1)} \le 6\sigma$. Thus, if $H(X_1) \ge 26$ then $\sigma > 50^6$ and the error-term in (13) is at most $2^n 144\frac{\log\sqrt{n}\sigma}{\sigma\sqrt{n}}$. Hence, taking $H(X_1) \ge \log\frac{2}{\epsilon} + \log\log\frac{2}{\epsilon} + n + 27$, we have by Theorem 2

$$H(X_1+\cdots+X_{n+1}) \ge H(X_1+\cdots+X_n) + \frac{1}{2}\log\Big(\frac{n+1}{n}\Big) - \epsilon. \tag{46}$$

□

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell system technical journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, no. 2, pp. 101–112, 1959.

[3] S. Artstein, K. Ball, F. Barthe, and A. Naor, "Solution of Shannon's problem on the monotonicity of entropy," *Journal of the American Mathematical Society*, vol. 17, no. 4, pp. 975–982, 2004.

[4] A. R. Barron, "Entropy and the central limit theorem," *The Annals of probability*, pp. 336–342, 1986.

[5] A. M. Tulino and S. Verdú, "Monotonic decrease of the non-gaussianness of the sum of independent random variables: A simple proof," *IEEE Transactions on Information Theory*, vol. 52, no. 9, pp. 4295–4297, 2006.

[6] M. Madiman and A. Barron, "Generalized entropy power inequalities and monotonicity properties of information," *IEEE Transactions on Information Theory*, vol. 53, no. 7, pp. 2317–2329, 2007.

[7] S. Haghighatshoar, E. Abbe, and E. Telatar, "Adaptive sensing using deterministic partial Hadamard matrices," in *2012 IEEE International Symposium on Information Theory Proceedings*. Ieee, 2012, pp. 1842–1846.

[8] S. Haghighatshoar, E. Abbe, and I. E. Telatar, "A new entropy power inequality for integer-valued random variables," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3787–3796, 2014.

[9] P. Harremoés, C. Vignat *et al.*, "An entropy power inequality for the binomial family," *JIPAM. J. Inequal. Pure Appl. Math*, vol. 4, no. 5, p. 93, 2003.

[10] J. O. Woo and M. Madiman, "A discrete entropy power inequality for uniform distributions," in *2015 IEEE International Symposium on Information Theory (ISIT)*. IEEE, 2015, pp. 1625–1629.

[11] T. Tao, "Sumset and inverse sumset theory for Shannon entropy," *Combinatorics, Probability and Computing*, vol. 19, pp. 603–639, 07 2010.

[12] T. Tao and V. H. Vu, "Entropy methods," 2005. [Online]. Available: https://www.math.ucla.edu/~tao/preprints/Expository/

[13] I. Z. Ruzsa, "Sumsets and entropy," *Random Structures & Algorithms*, vol. 34, no. 1, pp. 1–10, 2009.

[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. New York, NY, USA: Wiley-Interscience, 2006.

[15] T. Tao and V. H. Vu, *Additive Combinatorics*, ser. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2006.

[16] L. Gavalakis and I. Kontoyiannis, "Entropy and the discrete central limit theorem," *arXiv preprint arXiv:2106.00514*, 2021.

[17] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, 2009.

[18] ——, "Source polarization," in *2010 IEEE International Symposium on Information Theory*. IEEE, 2010, pp. 899–903.

[19] S. G. Bobkov, A. Marsiglietti, and J. Melbourne, "Concentration functions and entropy bounds for discrete log-concave distributions," *Combinatorics, Probability and Computing*, vol. 31, no. 1, pp. 54–72, 2022.

[20] J. Keilson and H. Gerber, "Some results for discrete unimodality," *Journal of the American Statistical Association*, vol. 66, no. 334, pp. 386–389, 1971.

[21] O. Rioul, "Variations on a theme by Massey," *IEEE Transactions on Information Theory*, vol. 68, no. 5, pp. 2813–2828, 2022.

[22] J. L. Massey, "On the entropy of integer-valued random variables," in *Procs. of Int. Workshop on Information Theory, Beijing, China*, 1988, p. C1.

[23] B. Davis and D. McDonald, "An elementary proof of the local central limit theorem," *Journal of Theoretical Probability*, vol. 8, no. 3, pp. 693–701, 1995.

[24] D. R. McDonald, "On local limit theorem for integer-valued random variables," *Theory of Probability & Its Applications*, vol. 24, no. 3, pp. 613–619, 1980.

[25] J. Mineka, "A criterion for tail events for sums of independent random variables," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 25, no. 3, pp. 163–170, 1973.

[26] S. Hoggar, "Chromatic polynomials and logarithmic concavity," *Journal of Combinatorial Theory, Series B*, vol. 16, no. 3, pp. 248–254, 1974.