BERNOULLI SUMS AND RENYI ENTROPY INEQUALITIES

MOKSHAY MADIMAN, JAMES MELBOURNE AND CYRIL ROBERTO

Abstract. We investigate the Rényi entropy of independent sums of integer valued random variables through Fourier theoretic means, and give sharp comparisons between the variance and the Rényi entropy, for Poisson-Bernoulli variables. As applications we prove that a discrete "min-entropy power" is super additive on independent variables up to a universal constant, and give new bounds on an entropic generalization of the Littlewood-Offord problem that are sharp in the "Poisson regime".

1. Introduction

For a countable set A, $|A|$ will denote its cardinality. The notation $\mathbb{P}$ will be reserved for a probability measure and the probability of an event A will be denoted $\mathbb{P}(A)$. For a discrete random variable X, on a countable set $\mathcal{X}$, we will denote its density function with respect to the counting measure as f_X so that $f_X(x) = \mathbb{P}(X = x)$, $x \in \mathcal{X}$ (when $\mathcal{X} = \mathbb{Z}$, we may use the notation $p_n := f_X(n)$ for simplicity). We will denote for f a function on a countable set $\mathcal{X}$, and $\alpha \in (0, \infty)$

$$\|f\|_\alpha := \left(\sum_{x \in \mathcal{X}} |f|^\alpha(i) \right)^{\frac{1}{\alpha}}.$$

By continuous limits we define $\|f\|_\infty := \sup_{x\in\mathcal{X}} |f(x)|$, and $\|f\|_0 = |\{x \in \mathcal{X} : f(x) \neq 0\}|$. We will be primarily interested in the case that $\mathcal{X} = \mathbb{Z}$, the integers. The subset of the integers $\{a, a+1, \ldots, b-1, b\}$ will be denoted by $[\![a, b]\!]$. When $a = 0$ we will abbreviate $[\![0, b]\!]$ by $[\![b]\!]$.

Definition 1.1 (Rényi Entropy [32]). *For X a random variable taking values $x \in \mathcal{X}$, such that $f_X(x) = \mathbb{P}(X = x)$, define for $\alpha \in (0, 1) \cup (1, \infty)$, the α-Rényi entropy of X,*

$$H_\alpha(X) := (1-\alpha)^{-1} \log \sum_{x\in\mathcal{X}} f_X^\alpha(x).$$

For $\alpha \in \{0, 1, \infty\}$ the Rényi entropy is defined through continuous limits;

$$H_0(X) := \log |\{x \in \mathcal{X} : f_X(x) > 0\}|$$
$$H_1(X) := -\sum_{x\in\mathcal{X}} f_X(x) \log f_X(x)$$
$$H_\infty(X) := -\log \|f_X\|_\infty.$$

Note that $H_1(X)$ agrees with the usual Shannon entropy. As such we will employ the conventional notation $H(X) := H_1(X)$. Note that for $\alpha \in (0, 1) \cup (1, \infty)$ and $\alpha' = \alpha/(\alpha - 1)$ we have the expression $H_\alpha(X) = -\alpha' \log \|f\|_\alpha$. We will also use the notation $H_\alpha(f_X)$ in place

Date: March 2, 2021.

Key words and phrases. Rényi entropy, Littlewood-Offord, Poisson-Bernoulli, entropy power, information theoretic inequalities.

The last author is supported by the Labex MME-DII funded by ANR, reference ANR-11-LBX-0023-01 and ANR-15-CE40-0020-03 - LSD - Large Stochastic Dynamics, and the grant of the Simone and Cino Del Duca Foundation, France.

of $H_\alpha(X)$ when it is more convenient to express the entropy as a function of the densities rather than variables.

The Rényi entropy has a well known analog in the continuous setting, when X is a $\mathbb{R}^d$ random variable with a density function with respect to the usual d-dimensional Lebesgue measure: namely the differential Rényi entropy is defined as

$$h_\alpha(X) \coloneqq (1-\alpha)^{-1} \log \int_{\mathbb{R}^d} f_X^\alpha(x)dx.$$

for $\alpha \in (0,1) \cup (1,\infty)$. For $\alpha \in \{0,1,\infty\}$ the continuous Rényi entropy is defined through continuous limits;

$$\begin{aligned} h_0(X) &\coloneqq \log |\{x \in \mathbb{R}^d : f_X(x) > 0\}| \\ h_1(X) &\coloneqq -\int_{\mathbb{R}^d} f_X(x) \log f_X(x) dx \\ h_\infty(X) &\coloneqq -\log \|f_X\|_\infty. \end{aligned}$$

where $|\cdot|$ denotes the Lebesgue volume, h_1 corresponds to the differential Shannon entropy, and $\|f_X\|_\infty$ denotes the essential suprema of f_X with respect to the Lebesgue measure in this case. Super-additivity properties of the Rényi entropy connect anti-concentration results in Probability [34, 5, 24], the seminal entropy power inequality due to Shannon [35], and the Brunn-Minkowski inequality of convex geometry, see [15] for further background. Such connections can be traced back to [10] where the analogy between the Brunn-Minkowski inequality and the Entropy Power Inequality was first described. In [12] the sharp Young inequality (see [1]) based proofs of the Brunn-Minkowski inequality [9] and entropy [21] were synthesized to prove Rényi entropy inequalities connecting the two results. We direct the reader to [23] for further background. There has been significant recent interest and progress in understanding the behavior of the differential Rényi entropy's behavior on independent summation. In analogy with the Shannon entropy power $N(X) \coloneqq N_1(X) \coloneqq e^{2h_1(X)}$, define for $\alpha \in [0,\infty]$, $N_\alpha(X) = e^{2h_\alpha(X)/d}$.

Theorem 1.2 ([4, 5, 31, 27])**.** *For $\alpha \in [1,\infty]$, there exists $c(\alpha) \geq 1/e$ such that X_i independent $\mathbb{R}^d$ valued random variables implies*

$$N_\alpha(X_1 + \cdots + X_n) \geq c(\alpha) \sum_{i=1}^n N_\alpha(X_i). \tag{1}$$

Further, for $\alpha \in [0,1)$, there exists $c(\alpha) > 0$, such that if X_i are further assumed to be log-concave[1] *then (1) holds.*

Note that $c(\alpha)$ can be given independent of n. When $\alpha = 1$ the Entropy Power Inequality is the fact that one may take $c(\alpha) = 1$, while the Brunn-Minkowski inequality[2] can be written as

$$N_0^{\frac{1}{2}}(X_1 + \cdots + X_n) \geq \sum_{i=1}^n N_0^{\frac{1}{2}}(X_i). \tag{2}$$

For all other α, necessarily $c(\alpha) < 1$. In fact without concavity assumptions on the X_i, (1) fails (see [20]) for any $\alpha \in (0,1)$. Variants of the Rényi entropy power inequalities were studied in [7, 19] and connections with optimal transport theory can be found in [33, 11].

In the discrete setting, it is natural to wonder if a parallel interplay exists, especially in light of the fruitful analogy between additive combinatorics and convex geometry, already

[1]We recall that an $\mathbb{R}^d$ valued random variable is log-concave when it has a density f_i such that $t \in [0,1]$ and $x, y \in \mathbb{R}^d$ implies $f((1-t)x + ty) \geq f^{1-t}(x) f^t(y)$.

[2]Observe that (2) is stronger than the inequality $N_0(X_1 + \cdots + X_n) \geq \sum_{i=1}^n N_0(X_i)$

well known, see [38]. There has been considerable interest in developing discrete versions of the entropy power inequality, see [18, 17, 16, 25]. General super additivity properties of the Rényi entropy on independent summation have proved elusive in the discrete setting, and the mentioned results only succeed in the $\alpha = 1$ case, for special classes of variables. Even in the $\alpha = 2$ case sometimes referred to as the collision entropy in the literature, which has taken a central role in information theoretic learning applications, see [30], little seems to be known.

We introduce the following functional,

Definition 1.3. *For X a discrete random variable on $\mathbb{Z}$, and $\alpha \in (0,1) \cup (1,\infty)$, we set*

$$\Delta_\alpha(X) \coloneqq \left(\sum_{n \in \mathbb{Z}} p_n^\alpha \right)^{\frac{2}{1-\alpha}} - 1.$$

Further (through continuous limits), $\Delta_0(X) \coloneqq |\{n : p_n > 0\}|^2 - 1$, $\Delta_1(X) \coloneqq \prod_{n \in \mathbb{Z}} p_n^{-2p_n} - 1$, and $\Delta_\infty(X) \coloneqq \|p_i\|_\infty^{-2} - 1$.

For example, when H_α is taken in base 2, $\Delta_\alpha(X) = 2^{2H_\alpha(X)} - 1 \in [0,\infty]$ for any $\alpha \in [0,\infty]$. Note, the Cauchy-Davenport theorem (on $\mathbb{Z}$), in loose analogy with (2), can be written as the super additivity of a stronger functional as well, explicitly, $(\Delta_0(X)+1)^{\frac{1}{2}} - 1$.

Recall that a random variable S is a Poisson-Binomial when there exists X_i independent Bernoulli variables, such that $\mathbb{P}(X_i = 1) = p_i = 1 - \mathbb{P}(X_i = 0)$ and $S = X_1 + \cdots + X_n$.

For $\alpha \geq 2$ we will show that for Poisson-Binomials variables and their limits, the functional $\Delta_\alpha(X)$ is, up to absolute constants, proportional to variance. This should be compared to the continuous setting where it is known that the Rényi entropy power is proportional to variance for Gaussian variables, for example, see [2], for a connection to Bourgain's hyperplane conjecture [8].

It will be shown that if S is the weak limit of a sequence of Poisson-Binomial variables, and $\alpha \in [2,\infty]$ then

$$2\alpha' \mathrm{Var}(S) \leq \Delta_\alpha(S).$$

This inequality is sharp for Bernoulli variables with parameter p tending to 0, and for Poisson variables with parameter λ tending to 0 as well.

Further we will prove a "min-Entropy power inequality": for $\alpha = \infty$, we will prove that without qualification, independent X_i satisfy the following Rényi entropy inequality

$$\Delta_\infty(X_1 + \cdots + X_n) \geq c \sum_{i=1}^n \Delta_\infty(X_i),$$

for a universal $c > 0$ independent of n, the number of summands, one can take $c = \frac{1}{20}$ for instance.

As an application we prove bounds on a generalized version of the Littlewood-Offord problem. What is more, we give sharp bounds for a Rényi entropic generalization of the Littlewood-Offord problem $\alpha \geq 2$.

The mathematical underpinning the min-EPI is an identification of extreme points in the space of probability measures with a fixed upper bound on their density functions that was proven in [24], and a rearrangement inequality from [25]. A main technical contribution is an L^p-norm bound on the characteristic function of a Bernoulli variable. Recall that $\hat{f}_X(t) = \mathbb{E}(e^{itX})$, $t \in \mathbb{R}$, denotes the Fourrier transform of the (discrete) random variable X. For $q \geq 1$ we set $\|\hat{f}_X\|_q^q := \frac{1}{2\pi}\int_{-\pi}^{\pi} |\mathbb{E}e^{itX}|^q dt$. We prove that when X is a Bernoulli, with variance σ^2, $\|\hat{f}_X\|_q^q \leq (6\sigma^2 q)^{-1/2} \int_0^{\sqrt{6\sigma^2 q}} e^{-t^2/2} dt$.

Let us outline the contents of the paper. In Section 2 we derive L^p bounds for the characteristic function of a Bernoulli random variable in terms of its variance using a distributional argument due to [29]. Then, we give a general theorem for the extension of such inequalities to independent sums. In Section 3 we demonstrate how the characteristic function bounds of Section 2 can be used to deliver sharp comparisons between the Rényi entropy and the variance for variables with distributions in the closure of the Poisson-Binomial. It is also demonstrated that such bounds cannot be achieved for the case that $\alpha = 1$ by a counter example. In Section 4 we develop the functional analytic tools to reduce the problem of a min-EPI for general random variables to a min-EPI of variables consisting only of Bernoulli and Uniform distributions. In addition, an extension of the inequality to $\mathbb{Z}^n$ is given, as well as reversals and sharpenings of the min-EPI in the case that the X_i are Poisson-Binomial. In Section 5, the Littlewood-Offord problem is introduced, and its reduction to Poisson-Binomial variables is given. The bounds for the min-Entropy of a Poisson-Binomial in terms of its variance are applied, and then it is shown that these results can be extended to deliver Rényi bounds on an entropic Littlewood-Offord problem. Some proofs are suppressed to the appendix.

2. Bernoulli Sums

We derive L^p bounds on the characteristic functions of Bernoulli variables via a comparison with Gaussians. This argument will be distributional. We use V, the capitalization of a non-negative function v to denote its distribution function. Explicitly,

Definition 2.1. *For measurable $v : \mathbb{R} \to [0, \infty)$, its distribution function $V : [0, \infty]$, is defined by*

$$V(t) = |\{x : v(x) > t\}|$$

where $|\cdot|$ denotes the Lebesgue length of a set.

Integrals of functions can be easily computed from integrals of their distribution functions using Fubini-Tonelli, and the following formula

$$\int v = \int_0^\infty |\{x : V(x) > t\}| dt. \tag{3}$$

We will use a technique we learned in [29], which shows that a weak monotonicity condition on the difference of two distribution functions, implies a monotonicity result for the difference in the integral of powers of the original functions. We include the proof from the original paper for the convenience of the reader in the appendix.

Lemma 2.2. [29] *For w and v non-negative functions in L^1, with distribution functions W and V respectively, then if $W - V \leq 0$ on $[0, t_0]$ and $W - V \geq 0$ on $[t_0, \infty)$, the function*

$$\varphi(s) = \frac{\int_{\mathbb{R}} w^s - v^s}{s t_0^s}$$

is increasing on the set of $s > 0$ such that $w^s - v^s$ is in L^1.

In particular if one observes that $\int w^{s_0} \geq \int v^{s_0}$ for some $s_0 > 0$ then

$$\int w^s \geq \int v^s$$

for $s \geq s_0$.

This lemma will be used to derive the following main theorem.

Theorem 2.3. *For X a Bernoulli with variance σ^2, then $q \geq 1$ implies*

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |\mathbb{E}e^{itX}|^q dt \leq \frac{1}{\sqrt{6\sigma^2 q}}\int_0^{\sqrt{6\sigma^2 q}} e^{-t^2/2}dt.$$

Note that for a Bernoulli X with variance σ^2, the norm squared of its characteristic function can be expressed as a convex combination of the cosine function and the constant function one, namely

$$|\mathbb{E}e^{itX}| = \sqrt{(1-\lambda)+\lambda\cos(t)}, \tag{4}$$

where $\lambda = 2\sigma^2 \in [0, 1/2]$. This motivates the following definition for $\lambda \in [0, 1/2]$.

$$v_\lambda(t) = \sqrt{(1-\lambda)+\lambda\cos t}$$
$$w_\lambda(t) = \exp\left[\frac{-3\lambda t^2}{2\pi^2}\right].$$

We first claim that $\int_0^\pi v_\lambda^s(t)dt < \int_0^\pi w_\lambda^s(t)dt$ holds when $s = 1$.

Lemma 2.4. *When $s = 1$, and $\lambda \in (0, 1/2]$, $\int_0^\pi v_\lambda^s(t)dt > \int_0^\pi w_\lambda^s(t)dt$, or*

$$\int_0^\pi \sqrt{(1-\lambda)+\lambda\cos(t)}dt < \int_0^\pi \exp\left\{\frac{-3\lambda t^2}{2\pi^2}\right\}dt.$$

Proof. Using the inequality $\sqrt{1-x} < 1 - \frac{x}{2}$ for $x \in (0, 1]$,

$$\begin{aligned}\int_0^\pi \sqrt{(1-\lambda)+\lambda\cos(t)}dt &< \int_0^\pi 1 - \lambda\frac{1-\cos(t)}{2}dt \\ &= \pi(1-\lambda/2).\end{aligned}$$

Meanwhile the inequality and $e^x \geq 1 + x$, gives

$$\begin{aligned}\int_0^\pi \exp\left[\frac{-3\lambda t^2}{2\pi^2}\right]dt &\geq \int_0^\pi 1 + \frac{-3\lambda t^2}{2\pi^2}dt \\ &= \pi(1-\lambda/2),\end{aligned}$$

so that $\int_0^\pi w_\lambda(t)dt > \int_0^\pi v_\lambda(t)dt$ as claimed. □

Lemma 2.5. *For $\lambda > 0$, the function $t \mapsto w_\lambda(t) - v_\lambda(t)$ has no more than one zero on $(0, \pi]$*

The proof is calculus computations and we leave it to an appendix.

Proof of Theorem 2.3. As functions of t, v_λ and w_λ are both strictly decreasing on $[0, \pi]$. Thus, their respective distribution functions V_λ and W_λ are just their strictly decreasing inverse functions on $[0, 1]$. Since $w_\lambda > v_\lambda$ for small t, $W_\lambda > V_\lambda$ for y close to 1, and since w_λ and v_λ cross at no more than one point (by Lemma 2.5), V_λ and W_λ cross at no more than one point. We consider two cases. If $w_\lambda(\pi) - v_\lambda(\pi) \geq 0$, then $w_\lambda - v_\lambda \geq 0$ on $[0, \pi]$ and the theorem holds immediately. If $w_\lambda(\pi) - v_\lambda(\pi) < 0$, then $w_\lambda - v_\lambda$ has exactly one zero, and hence $W_\lambda - V_\lambda$ has exactly one zero. Therefore W_λ and V_λ satisfy the conditions of Lemma 2.2 for some t_λ, and hence,

$$s \mapsto \frac{\int_0^\pi w_\lambda^s(t) - v_\lambda^s(t)dt}{st_\lambda}$$

is increasing on $(1, \infty)$. Since $\int \frac{w_\lambda - v_\lambda}{t_\lambda} > 0$ by Lemma 2.4, we must have

$$\int_0^\pi w_\lambda^s(t)dt > \int^\pi v_\lambda^s(t)dt$$

for all $s \geq 1$. This leads to the desired conclusion since on the one hand

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} |\mathbb{E}e^{itX}|^q dt = \frac{1}{\pi}\int_0^{\pi} v_\lambda^q,$$

and on the other hand

$$\frac{1}{\pi}\int_0^{\pi} w_\lambda^q = \frac{1}{\pi}\int_0^{\pi} e^{-\frac{3q\lambda t^2}{2\pi^2}} dt = \frac{1}{\sqrt{3q\lambda}}\int_0^{\sqrt{3q\lambda}} e^{-u^2/2} du = \frac{1}{\sqrt{6\sigma^2 q}}\int_0^{\sqrt{6\sigma^2 q}} e^{-u^2/2} du$$

where the second inequality follows from the change of variable $u = \sqrt{3q\lambda}t/\pi$ and the last one from the fact that $\lambda = 2\sigma^2$. □

Our next aim is to extend the previous comparison to a finite or infinite sum of independent Bernoulli variables. We will use the following lemma.

Lemma 2.6. *Fix $\Phi\colon (0,\infty) \to [0,\infty)$. Suppose that X_i are independent random variables such that*

$$\|\hat{f}_{X_i}\|_q^q \leq \Phi(c_i q)$$

holds for all $q \geq 1$ and some $c_i > 0$. Then

$$\|\hat{f}_{\sum_i X_i}\|_q^q \leq \Phi(cq)$$

holds for all $q \geq 1$ as well, with $c = \sum_i c_i$.

Proof. By independence and Hölder's inequality for $\sum_i \frac{1}{q_i} = 1$,

$$\begin{aligned}
\|\hat{f}_{\sum_i X_i}\|_q^q &= \|\prod_i \hat{f}_{X_i}^q\|_1 \\
&\leq \prod_i \|\hat{f}_{X_i}^q\|_{q_i} \\
&= \prod_i \left(\|\hat{f}_{X_i}\|_{qq_i}^{qq_i}\right)^{\frac{1}{q_i}}.
\end{aligned}$$

Applying the hypothesis, and taking $q_i = \frac{c}{c_i}$,

$$\|\hat{f}_{\sum_i X_i}\|_q^q \leq \prod_i \Phi^{\frac{1}{q_i}}(c_i q_i q) = \Phi(cq).$$

□

Recall that a finite sum $\sum_{i=1}^n X_i$ of independent (not necessarily identically distributed) Bernoulli X_i is called a Poisson binomial variable. We refer to [37] for a recent review on this topic.

Definition 2.7 (Bernoulli Sum)**.** *We consider Y to be a Bernoulli Sum, if there exists a sequence of Poisson binomial variables Y_n converging weakly to Y.*

Thanks to the previous Lemma, the bound of Theorem 2.3 transfers to Bernoulli sums.

Theorem 2.8. *Let Y be a Bernoulli sum with variance σ^2. Then*

$$\|\hat{f}_Y\|_q^q \leq \frac{1}{\sqrt{6\sigma^2 q}}\int_0^{\sqrt{6\sigma^2 q}} e^{-t^2/2} dt. \tag{5}$$

Proof. Since Y is a Bernoulli sum, there exist a sequence of Y_n converging weakly to Y. Given a Y_n of the sequence, there exists X_i independent Bernoulli with variance σ_i^2, such that $\sum_{i=1}^{m(n)} X_i$. Taking $c_i = \sigma_i^2$, $\Phi(x) = \frac{1}{\sqrt{6x}} \int_0^{\sqrt{6x}} e^{-t^2/2} dt$, the hypothesis of Lemma 2.6 is satisfied for X_i thanks to Theorem 2.3, and the conclusion of the Lemma is exactly (5).

Since the bound holds for each Y_n, it is enough to observe $\|\hat{f}_{Y_n}\|_q^q \to \|\hat{f}_Y\|_q^q$ and $\sigma_{Y_n}^2 \to \sigma_Y^2 < \infty$. Note, that Y is necessarily log-concave since is the limit of log-concave variables Y_n (see Definition 4.5 for a definition). As a consequence there exists $C, c > 0$ such that $f_{Y_n}(k) \leq Ce^{-c|k|}$ and $f_Y(k) \leq Ce^{-c|k|}$ holds for all n, and all $k \in \mathbb{Z}$. Thus all moments exist and there is no difficulty passing limits, and by Lévy's continuity theorem $\|\hat{f}_{Y_n}\|_q^q \to \|\hat{f}_Y\|_q^q$ to complete the result. $\square$

Let us remark on the nature of the function $z \mapsto \frac{1}{z} \int_0^z e^{-t^2/2} dt$, as it will be useful to have simple upper bounds for our applications.

Lemma 2.9. *For $z \in (0, \infty)$,*

$$\frac{1}{z} \int_0^z e^{-t^2/2} dt \leq \min\left\{ \frac{1}{\sqrt{1 + (z^2/3)}}, \sqrt{\frac{\pi}{2z^2}} \right\}.$$

Remark 2.10. *Note that*

$$\sqrt{\frac{\pi}{2z^2}} \leq \left(1 + \frac{z^2}{3}\right)^{-1/2}$$

exactly when $z \geq \sqrt{\frac{3\pi}{6-\pi}} \approx 1.8158$.

Proof. The second term follows from $\int_0^z e^{-t^2/2} dt \leq \int_0^\infty e^{-t^2/2} dt = \sqrt{\pi/2}$, which implies

$$\int_0^z e^{-t^2/2} dt / z \leq \sqrt{\frac{\pi}{2z^2}}.$$

The first term is more complicated. It is enough to prove that $y \mapsto F(y) := \frac{1}{\left(\int_0^y e^{-t^2/2} dt\right)^2} - \frac{1}{y^2}$ is non-decreasing on $(0, \infty)$. Indeed, this would imply for any $y > 0$ that $F(y) \geq \lim_{y \downarrow 0} F(y) = \frac{1}{3}$ which can be rephrased as the expected bound

$$\frac{\int_0^z e^{-t^2/2} dt}{z} \leq \left(1 + \frac{z^2}{3}\right)^{-1/2}. \tag{6}$$

To prove that F is non-decreasing, we take the derivative and obtain that

$$F'(y) = \frac{2}{y^3 \left(\int_0^y e^{-t^2/2} dt\right)^3} \left(-y^3 e^{-y^2/2} + \left(\int_0^y e^{-t^2/2} dt \right)^3 \right), \qquad y > 0.$$

Set $G(y) := -y^3 e^{-y^2/2} + \left(\int_0^y e^{-t^2/2} dt\right)^3$, $y > 0$, and observe that

$$G'(y) = e^{-y^2/2} \left(-3y^2 + y^4 + 3 \left(\int_0^y e^{-t^2/2} dt \right)^2 \right).$$

Now, since $e^{-t^2/2} \geq 1 - \frac{t^2}{2}$, we have $\int_0^y e^{-t^2/2} dt \geq y - \frac{y^3}{6}$. Therefore, for $y \in (0, \sqrt{6}]$ (so that $y - \frac{y^3}{6} \geq 0$), it holds

$$G'(y) \geq e^{-y^2/2} \left(-3y^2 + y^4 + 3 \left(y - \frac{y^3}{6} \right)^2 \right) = \frac{y^6 e^{-y^2/2}}{12} > 0.$$

On the other hand, for $y \geq \sqrt{6}$, we observe that $-3y^2 + y^4 > 0$ so that $G'(y) > 0$ on $[\sqrt{6}, \infty)$ and therefore on $(0, \infty)$. As a consequence G is increasing on $(0, \infty)$, and since $\lim_{y \downarrow 0} G(y) = 0$, F is non-decreasing on $(0, \infty)$ as expected. The limit as y tends to zero is an easy consequence of the Taylor expansion $\int_0^y e^{-t^2/2} dt = y - y^3/6 + o(y^3)$, while (6) is obtained directly from $F(y) \geq \frac{1}{3}$. □

3. Rényi Entropy Inequalities

In this section we prove Rényi entropy inequalities for Poisson-Binomials and their limits. We will use a less unorthodox formulation of the well-known Hausdorff-Young's inequality to translate $L^{\alpha'}$ bounds on the characteristic functions of Poisson-Binomials into α-Rényi entropy bounds.

Theorem 3.1 (Hausdorff-Young). *For $p \in [2, \infty]$, and a random variable X on $\mathbb{Z}$ with probability mass function f then $\|f\|_p \leq \|\hat{f}\|_q$ where $\frac{1}{p} + \frac{1}{q} = 1$.*

Remark 3.2. *We observe that, in contrast with the continuous setting, the inequality $\|f\|_p \leq \|\hat{f}\|_q$ is sharp for random variables on $\mathbb{Z}$. To see this it is enough to consider a Dirac mass at zero ($f(0) = 1$ and $f(n) = 0$ for all $n \neq 0$ for which $\hat{f} \equiv 1$ so that $\|f\|_p = \|\hat{f}\|_p = 1$ for all p).*

Proof. The inequality follows by the Riesz-Thorin interpolation Theorem since $\|f\|_2 = \|\hat{f}\|_2$ and $\|f\|_\infty \leq \|\hat{f}\|_1$. □

Theorem 3.3. *When Y is a Poisson-Binomial with variance σ^2 and $\alpha \in [2, \infty]$ then*

$$H_\alpha(Y) \geq \log \frac{\sqrt{6\sigma^2\alpha'}}{\int_0^{\sqrt{6\sigma^2\alpha'}} e^{-t^2/2} dt}.$$

In particular,

$$H_\alpha(Y) \geq \frac{1}{2} \max\left\{ \log\left(1 + 2\alpha'\sigma^2\right), \log\left(\frac{12\alpha'\sigma^2}{\pi}\right) \right\}.$$

Remark 3.4. *Notice that for $\alpha = \infty$, the bound $H_\alpha(Y) \geq \frac{1}{2}\log\left(1 + 2\alpha'\sigma^2\right)$ reads*

$$\Delta_\infty(Y) \geq 2 \ \mathrm{Var}(Y). \tag{7}$$

Proof. Recall that $H_\alpha(Y) = -\alpha' \log \|f_Y\|_\alpha = -\log \|f_Y\|_\alpha^{\alpha'}$. Therefore, Hausdorff-Young's inequality $\|f_Y\|_\alpha^{\alpha'} \leq \|\hat{f}_Y\|_\alpha^{\alpha'}$ together with Theorem 2.8 guarantee that

$$\begin{aligned} H_\alpha(Y) &\geq -\log \|\hat{f}_Y\|_{\alpha'}^{\alpha'} \\ &\geq \log \frac{\sqrt{6\sigma^2\alpha'}}{\int_0^{\sqrt{6\sigma^2\alpha'}} e^{-t^2/2} dt}. \end{aligned}$$

The last bound follows from Lemma 2.9 applied with $z = \sqrt{6\sigma^2\alpha'}$. □

Our last ingredient in the proof of Theorem 3.7 is the following lemma.

Lemma 3.5. *For $\alpha \geq 2$, and X Bernoulli implies*

$$\Delta_\alpha(X) \leq 12 \mathrm{Var}(X), \tag{8}$$

with equality when the Bernoulli parameter $\theta = \frac{1}{2}$.

Remark 3.6. *Note that if $\theta = 1/2$, $\Delta_\alpha(X) = 12\mathrm{Var}(X)$ independent of α and cannot be improved for any Rényi parameter $\alpha \geq 2$. Considering the Shannon entropy for small θ shows an analogous result fails for all $\alpha \leq 1$, as $\lim_{\theta \to 0} \frac{\Delta_1(X)}{\mathrm{Var}(X)} = \infty$. It can be shown for $\alpha \in (1,2)$, there exists $C(\alpha)$ such that $\Delta_\alpha(X) \leq C(\alpha)\mathrm{Var}(X)$. However by investigation about θ close to 0, $C(\alpha)$ is necessarily larger than 12 for $\alpha < \frac{6}{5}$.*

Proof of Lemma 3.5. By the monotonicity of Rényi entropy $H_\alpha(X) \leq H_2(X)$, so it suffices to prove $\Delta_2(X) \leq 12\mathrm{Var}(X)$. This is equivalent that for $t \in [0,1]$,

$$(t^2 + (1-t)^2)^{-2} - 1 \leq 12t(1-t).$$

Which is equivalent to proving $P(t) \geq 0$, where P is the polynomial,

$$P(t) = (12t(1-t) + 1)(t^2 + (1-t)^2)^2 - 1 \geq 0$$

on $[0,1]$. However, P can be factored, to

$$-4(t-1)t(2t-1)^2(3t^2 - 3t + 2),$$

from which the non-negativity of P on the interval is obvious. □

The main theorem is the following.

Theorem 3.7. *For $\alpha \geq 2$, X_i Poisson-Binomial then*

$$\Delta_\alpha(X_1 + \cdots + X_n) \geq \frac{\alpha'}{6} \sum_{i=1}^n \Delta_\alpha(X_i), \tag{9}$$

independent of the number of summands.

Remark 3.8. *Note that*

$$c(\alpha) := \frac{\alpha}{6(\alpha - 1)} = \frac{\alpha'}{6} \geq \frac{1}{6}.$$

Also, observe that Inequality (9) fails for the Shannon entropy, as can be seen by considering X_i to be iid with parameter θ. Indeed, the sum $X_1 + \cdots + X_n$ has a Binomial distribution, whose entropy has the well known asymptotic formula,

$$H(X_1 + \cdots + X_n) = \frac{1}{2}\log_2(2\pi e n\theta(1-\theta)) + O(1/n).$$

Therefore

$$\begin{aligned} \frac{\Delta_1(\sum_i X_i)}{\sum_i \Delta_1(X_i)} &= \frac{2^{2(H(X_1+\cdots+X_n)+O(1/n))} - 1}{n\Delta_1(X_1)} \\ &= \frac{2\pi e\theta(1-\theta)2^{O(1/n)} - \frac{1}{n}}{\Delta_1(X_1)} \end{aligned}$$

Note that with $\theta \to 0$, and $H(X_1)$ taken in base e goes to zero, hence, $\Delta_1(X_1) = e^{2H(X_1)} - 1 \approx 2H(X_1)$. Thus

$$\frac{\Delta_1(\sum_i X_i)}{\sum_i \Delta_1(X_i)} \leq O(1)\frac{\mathrm{Var}(X_1)}{H(X_1)} \to 0,$$

with $\theta \to 0$, precluding a summand independent entropy power inequality in the sense of Theorem 3.7 for the Shannon entropy.

Proof of Theorem 3.7. By Theorem 3.3, it holds

$$\begin{aligned}\Delta_\alpha\left(\sum_i X_i\right) &= e^{2H_\alpha(\sum_i X_i)} - 1 \\ &\geq 2\alpha' \mathrm{Var}\left(\sum_i X_i\right).\end{aligned}$$

Using that additivity of the variance of independent variables and Lemma 3.5, we conclude that

$$\begin{aligned}2\alpha' \mathrm{Var}\left(\sum_i X_i\right) &= 2\alpha' \sum_i \mathrm{Var}(X_i) \\ &\geq 2\alpha' \sum_i \frac{\Delta_\alpha(X_i)}{12} \\ &= \frac{\alpha'}{6} \sum_i \Delta_\alpha(X_i).\end{aligned}$$

□

4. Min-Entropy Power

Given independent integer valued random variables $X_1, \cdots, X_n$. We would like to investigate minimizers of the quantity $H_\infty(X_1 + \cdots + X_n)$ on the set $H_\infty(X_j) \geq \log C_j$ where $C_j > 1$. We note that $H_\infty(X) \geq \log C$ corresponds to $\|f\|_\infty \leq \frac{1}{C}$.

4.1. **Extreme points.** Let us denote the set of probability density functions supported on a finite set M, with density f bounded by $\frac{1}{C}$ by

$$\mathcal{P}_C(M) = \left\{ f \colon M \to [0,1] \text{ such that } 0 \leq f \leq \frac{1}{C} \text{ and } \sum_{i \in M} f(i) = 1 \right\}. \tag{10}$$

Note that $\mathcal{P}_C(M)$ is a convex compact subset[3] of $\mathbb{R}^{|M|}$, and hence is the closure of the convex hull of its extreme points. Let us recall the necessary definitions. The extreme points $\mathcal{E}(K)$ associated to a convex set K are defined as

$$\mathcal{E}(K) = \left\{ k \in K : k = \frac{k_1 + k_2}{2} \text{ for } k_i \in K \text{ implies } k_1 = k_2 \right\}.$$

The convex hull of a set K is

$$co(K) = \left\{ x : \exists \lambda_i > 0 \text{ and } k_i \in K \text{ such that } \sum_{i=1}^n \lambda_i = 1, \sum_{i=1}^n \lambda_i k_i = x \right\}.$$

For $x \in \mathbb{R}$, we write $\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}$ for the entire part of x.

Theorem 4.1. *For $\mathcal{P}_C(M)$ defined as in (10) with $C \leq |M|$,*

$$\mathcal{E}(\mathcal{P}_C(M)) = \left\{ f : f = \frac{\mathbb{1}_A}{C} + \left(1 - \frac{\lfloor C \rfloor}{C}\right) \mathbb{1}_{\{x\}}, |A| = \lfloor C \rfloor, x \notin A \right\}.$$

[3] observe that $\mathcal{P}_C(M) = \emptyset$ if $C > |M|$. This is pathological situation that we will exclude.

Note that when C is chosen to be an natural number, $\mathcal{E}(\mathcal{P}_C(M))$ is the uniform distributions on sets of size C contained in M, else the extremal distributions are "nearly uniform" representing an appropriately scaled convex combination of a uniform distribution on a set of size $\lfloor C \rfloor$ and a disjoint point mass. When $1 < C \leq 2$, the extreme points of $\mathcal{P}_C(M)$ are probability mass functions supported on exactly two points. A more general proof of this result is given in [24]. As we will not have use for the generality, we provide a simpler proof to allow this article to be more self-contained.

Proof. We will prove the two inclusions ($\subset$, $\supset$) of the sets.

Given $f \in \mathcal{P}_C(M)$, if there exists $i \neq j$ such that $f(i), f(j) \in (0, \frac{1}{C})$ then $g_1 = f + \varepsilon(\mathbb{1}_{\{i\}} - \mathbb{1}_{\{j\}})$ and $g_2 = f - \varepsilon(\mathbb{1}_{\{i\}} - \mathbb{1}_{\{j\}})$ are distinct elements of $\mathcal{P}_C(M)$ (for ε small enough), and since $\frac{g_1+g_2}{2} = f$, $f \notin \mathcal{E}(\mathcal{P}_C(M))$. This proves that extreme points of $\mathcal{P}_C(M)$ have at most one value in $(0, \frac{1}{C})$ and therefore proves the first inclusion.

Conversely, consider f such that there exists i_o with $f(j) \in \{0, 1/C\}$ for all $j \neq i_o$ and suppose that $f = \frac{g_1+g_2}{2}$ for some $g_1, g_2 \in \mathcal{P}_C(M)$. For $j \neq i_o$, $f(j)$ is an extreme point of the interval $[0, \frac{1}{C}]$ and $\frac{g_1(j)+g_2(j)}{2} = f(j)$. Hence $g_1(j) = g_2(j) = f(j)$. By the constraint that f, g_1, and g_2 are probability mass functions we must have $f(i_o) = g_1(i_o) = g_2(i_o)$ as well and the second inclusion is proved. $\square$

Theorem 4.2. *Let m be a natural number and recall that $[\![m]\!] = \{0, 1, \dots, m\}$. For $\alpha \in [0, \infty]$, a natural number n, constants $C_1, \dots, C_n \leq m+1$, independent random variables X_i with probability mass functions $f_{X_i} \in \mathcal{P}_{C_i}([\![m]\!])$, it holds*

$$H_\alpha(X_1 + \cdots + X_n) \geq \min_{Z \in \mathcal{E}} H_\alpha(Z_1 + \cdots + Z_n), \tag{11}$$

where $\mathcal{E}$ is the collection of all $Z = (Z_1, \dots, Z_n)$ such that Z_i are independent variables with density $f_{Z_i} \in \mathcal{E}(\mathcal{P}_{C_i}([\![m]\!]))$.

Remark 4.3. *We stress that the minimum in the right hand side of (11) is indeed a minimum and therefore is achieved.*

Proof. Note that $X_1 + \cdots + X_n$ is supported on $[\![nm]\!]$, and as a function of densities, the map $f \mapsto H_\alpha(f)$ is continuous and quasi-concave in the sense that densities f and g satisfy $H_\alpha(\frac{f+g}{2}) \geq \min\{H_\alpha(f), H_\alpha(g)\}$. Indeed, continuity is obvious since the probability distributions under consideration have finite support, and quasi-concavity follows from the expression $H_\alpha(f) = \alpha' \log \|f\|_\alpha$, and the convexity of $f \mapsto \|f\|_\alpha$ for $\alpha > 1$ and its concavity for $\alpha < 1$. Thus the continuity of the map from $\mathcal{P}_{C_1}([\![m]\!]) \times \cdots \mathcal{P}_{C_n}([\![m]\!]) \to [0, \infty)$, given by $(f_1, \dots, f_n) \mapsto H_\alpha(f_1 * \cdots * f_n)$ is continuous since the convolution can be expressed as a polynomial of the terms of $f_i(k)$. What is more, the map is coordinate quasi-concave, since convolution is coordinate affine, in the sense that $\frac{f_1+g}{2} * f_2 * \cdots * f_n = \frac{f_1 * f_2 * \cdots * f_n}{2} + \frac{g * f_2 * \cdots * f_n}{2}$. Thus the map $g \mapsto H_\alpha(g * f_{X_2} * \cdots * f_{X_n})$ is continuous and quasi-concave on $\mathcal{P}_{C_1}([\![m]\!])$. Since $\mathcal{P}_{C_1}([\![m]\!])$ is a compact, convex subset of $\mathbb{R}^{m+1}$, and since points of compact convex subsets can be written as convex combinations of their extreme points by Krein-Milman, there exists a minimizer $g_1 \in \mathcal{E}(\mathcal{P}_{C_1}([\![m]\!])$ such that

$$H_\alpha(f_{X_1} * f_{X_2} * \cdots * f_{X_n}) \geq H_\alpha(g_1 * f_{X_2} * \cdots * f_{X_n})$$

Iterating the argument gives the proof. $\square$

We restate the case that $\alpha = \infty$ below.

Corollary 4.4. *For $X_1, \dots, X_n$ independent random variables taking values in a finite set M such that $H_\infty(X_i) \geq \log C_i$, there exists $U_1, \dots, U_n$ independent such that $f_{U_i} \in \mathcal{E}(\mathcal{P}_{C_i}(M))$ and*

$$H_\infty(X_1 + \cdots + X_n) \geq H_\infty(U_1 + \cdots + U_n).$$

4.2. **Rearrangement.** In this section we define the notions of log-concavity and of rearrangement of functions on the integers $\mathbb{Z}$, to be used later on.

Definition 4.5. *A function $f\colon \mathbb{Z} \to [0,\infty)$ is log-concave when*

$$f^2(n) \geq f(n+1)f(n-1).$$

and $f(i)f(j) > 0$ for $i < j$ implies $f(k) > 0$ for $k \in [i,j]$.

Definition 4.6. *For a function $f : \mathbb{Z} \to [0,\infty)$ with finite support,*

$$f = \sum_{i=0}^{n} a_i \mathbb{1}_{\{x_i\}}$$

with $x_1 < x_2 < \cdots < x_n$ denote

$$f^{\#} = \sum_{i=0}^{n} a_i \mathbb{1}_{\{i\}}.$$

When X_i are independent random variables with densities f_i, we denote by $X_i^{\#}$ a collection of random variables such that $X_i^{\#}$ has density $f_i^{\#}$.

4.3. **Integers.** In this section we will make use of a result due to Madiman-Wang-Woo [25] that shows that the Rényi entropy power is somehow decreasing under rearrangeemnt. More precisely, these authors prove that $f_1 * \cdots * f_n$ is majorized by $f_1^{\#} * \cdots * f_n^{\#}$. We refer to [26] for background on majorization.

The next theorem follows from [25], by applying the Schur concavity of Rényi entropy.

Theorem 4.7 (Theorem 1.4 [25])**.** *For $\alpha \in [0,\infty]$, and f_i are such that $f_i^{\#}$ are log-concave then*

$$H_\alpha(f_1 * \cdots * f_n) \geq H_\alpha(f_1^{\#} * \cdots * f_n^{\#}).$$

The significance of the theorem for our pursuits here, is that it will reduce our investigations of a min-entropy power inequality to Bernoulli and Uniform distributions.

Corollary 4.8. *For X_i independent variables with $\|f_{X_i}\|_\infty \leq 1/C_i$ with $C_i \in (1,2] \bigcup \cup_{i=3}^{\infty}\{i\}$ then*

$$H_\infty(X_1 + \cdots + X_n) \geq H_\infty(Z_1 + \cdots + Z_n)$$

where the variables Z_i are independent and $f_{Z_i} \in \mathcal{E}(\mathcal{P}_{C_i}([\![\lfloor C_i \rfloor]\!]))$. Moreover, Z_i is Bernoulli when $C_i \in (1,2]$ and Uniform on $\{0,1,\ldots,C_i - 1\}$ when $C_i \in \cup_{i=3}^{\infty}\{i\}$.

Proof. Let us first assume that the X_i are all supported on a finite set $[\![m]\!]$ for some m. This implies that $C_i \leq |[\![m]\!]| = m+1$. By Theorem 4.2 there exists densities $g_i \in \mathcal{E}(\mathcal{P}_{C_i}([\![m]\!]))$ such that

$$H_\infty(f_{X_1} * \cdots * f_{X_n}) \geq H_\infty(g_1 * \cdots * g_n).$$

However, by the assumption that $C_i \in (1,2] \bigcup \cup_{i=3}^{m+1}\{i\}$ the g_i either takes only two values, in which case $g_i^{\#}$ is a Bernoulli, or g_i is a Uniform distribution on C_i values, in which case $g_i^{\#}$ is a uniform distribution on $\{0,1,\ldots,C_i - 1\}$. In either case $g_i^{\#}$ is log-concave and Theorem 4.7, gives

$$H_\infty(g_1 * \cdots * g_n) \geq H_\infty(g_1^{\#} * \cdots * g_n^{\#}).$$

Combining the two inequalities completes the proof when the X_i have compact support. The general inequality is an approximation argument. Define an auxiliary function

$$f_i^{(m)}(n) = \begin{cases} \min\{f_{X_i}(n), \frac{1}{C_i} - \frac{1}{m}\} & \text{if } n \in [\![-m, m]\!] \\ 0 & \text{else.} \end{cases}.$$

and from $f_i^{(m)}$ define a density

$$f_{\tilde{X}_i}^{(m)}(n) = f_i^{(m)}(n) + \frac{1 - \sum_{k=-m}^{m} f_i^{(m)}(k)}{2m+1} \mathbb{1}_{[\![-m,m]\!]}(n).$$

The probability mass functions $f_{\tilde{X}_i}$ converge pointwise to f_{X_i} with $m \to \infty$, are compactly supported and $\|f_{\tilde{X}_i}^{(m)}(n)\|_\infty \leq \|f_{X_i}\|_\infty \leq 1/C_i$. Further pointwise convergence coincides with weak-convergence on $\mathbb{Z}$ by the Portmanteau theorem since all subsets of $\mathbb{Z}$ are closed and open. Note that $f \mapsto H_\infty(f)$ is upper semi-continuous with respect to weak convergence since for $f_\beta \to f$ pointwise, $f(n) = \lim_\beta f_\beta(n) \leq \liminf_\beta \|f_\beta\|_\infty$. Thus it follows that $\limsup_\beta H_\infty(f_\beta) \to H_\infty(f)$. Taking the max over n gives $\|f\|_\infty \leq \liminf_\beta \|f_\beta\|_\infty$ and hence

$$\limsup_\beta H_\infty(f_\beta) \leq H_\infty(f).$$

Since the map $(f_1, \dots, f_n) \mapsto f_1 * \cdots * f_n$ corresponds to the mapping of a product space of measures $(\mu_1, \dots, \mu)$ to their product measure, a weak continuous operation, and composed with pushing forward $\mu_1 \otimes \cdots \otimes \mu_n$ under the continuous map $T(x) = x_1 + \cdots + x_n$ a weak-continuous mapping, $(f_1, \dots, f_n) \mapsto f_1 * \cdots * f_n$ is a composition of weakly continuous functions and hence weakly continuous as well. Thus, $(f_1, \dots, f_n) \mapsto H_\infty(f_1 * \cdots * f_n)$ is upper semi-continuous and we have

$$H(f_{X_1} * \cdots * f_{X_n}) \geq \limsup_\beta H(f_{\tilde{X}_1^{(m)}} * \cdots * f_{\tilde{X}_n^{(m)}}).$$

Since the $f_{\tilde{X}_1^{(m)}}$ are compactly supported $H(f_{\tilde{X}_1^{(m)}} * \cdots * f_{\tilde{X}_n^{(m)}}) \geq H_\infty(Z_1 + \cdots + Z_n)$. □

4.4. **Min Entropy Inequality.** The aim of this section is to prove the following Min Entropy power inequality which constitutes one of our main theorems.

Theorem 4.9 (Min-EPI)**.** *For independent integer valued random variables X_i, it holds*

$$\Delta_\infty(X_1 + \cdots + X_n) \geq \frac{1}{22} \sum_{i=1}^{n} \Delta_\infty(X_i).$$

In order to prove Theorem 4.9, we will use a comparison between the min-entropy and the variance (in both directions). Such a comparison is essentially known in the literature. The next result hold for all random variables and is sharp for uniform distributions.

Theorem 4.10 (Bobkov-Chistyakov [6, 3])**.** *For a discrete variable X,*

$$\Delta_\infty(X) \leq 12\mathrm{Var}(X). \tag{12}$$

To state the other direction, we need to introduce two definitions. First we say that an integer valued random variable X is log-concave if its probability mass function f_X is log-concave, in the sense of Definition 4.5. In other words, $f_X^2(n) \geq f_X(n-1)f_X(n+1)$ holds for all n, and $f_X(k)f_X(n) > 0$ for $k < m < n$ implies $f_X(m) > 0$. Next we define the notion of symmetry.

Definition 4.11 (Symmetric Random variable)**.** *A real valued random variable X is symmetric when there exists a, such that*

$$\mathbb{P}(X = a + x) = \mathbb{P}(X = a - x).$$

holds for all $x \in \mathbb{R}$.

Note that since we consider only X integer valued, $a \in \frac{1}{2}\mathbb{Z}$. Also recall that both symmetry and log-concavity are preserved under independent summation.

Theorem 4.12 (Bobkov-Marsiglietti-Melbourne [3])**.** *For an integer valued, symmetric, log-concave variable X,*

$$\Delta_\infty(X) \geq 2\mathrm{Var}(X). \tag{13}$$

Strictly speaking, the case described in [3] only covered the case that X was symmetric about an integer point. The argument is to reduce to a "symmetric geometric distribution", that is a distribution on $\mathbb{Z}$ of the form $n \mapsto Cp^{|n|}$, by proving that a symmetric log-concave distribution majorizes the symmetric geometric distribution with the same maximum (see [28]) and the fact proven in [3], that the variance is Schur-concave on the space of symmetric distributions. Thus to check the veracity of Theorem 4.12 is suffices to check that it holds for symmetric geometric distributions, which can be done easily. For the convenience of the reader we cover the (missing) case that X is symmetric about a point in $\frac{1}{2} + \mathbb{Z}$ below.

Proof of Theorem 4.12. Suppose that f is a log-concave density symmetric about a point $n - \frac{1}{2}$. Note that if f is supported on only two points, the inequality is true immediately, since by symmetry f is a translation of Bernoulli with parameter $1/2$. Thus, assume that f is supported on at least 4 points so that $\|f\|_\infty < 1/2$ and observe that $\|f\|_\infty = f(n-1) = f(n)$. Take $p = 1 - 2\|f\|_\infty > 0$ and define a density g by $g(n + k) = \|f\|_\infty p^k$ for $k \geq 0$ and $g(n + k) = \|f\|_\infty p^{-k-1}$ for $k < 0$. Note that g is symmetric log-concave density, satisfying $\|g\|_\infty = \|f\|_\infty$ and $g \prec f$. Translating $g(k) = g(n + k - 1/2)$, and $f(k) = f(n + k - 1/2)$ we obtain densities symmetric about 0 taking values on $\frac{1}{2} + \mathbb{Z}$. It is straight forward to prove from the majorization that $X \sim g$ and $Y \sim f$ that $\mathbb{E}X^2 \geq \mathbb{E}Y^2$ and since both variables are centered, $\mathrm{Var}(X) \geq \mathrm{Var}(Y)$ while by definition $\Delta_\infty(X) = \Delta_\infty(Y)$. Thus it suffices to prove the inequality for X and $p \in (0, 1)$.

In this case, with integer $k \geq 0$, $\mathbb{P}(X = \pm(\frac{1}{2} + k)) = \frac{1-p}{2}p^k$, direct computations gives

$$\begin{aligned}
\Delta_\infty(X) &= \frac{4}{(1-p)^2} - 1, \\
\mathrm{Var}(X) &= (1-p)\sum_{k=0}^{\infty}\left(k + \frac{1}{2}\right)p^k \\
&= \frac{p^2 + 6p + 1}{4(1-p)^2}.
\end{aligned}$$

Thus the inequality $\Delta_\infty(X) \geq 2\mathrm{Var}(X)$ is equivalent to

$$8 - 2(1-p)^2 \geq p^2 + 6p + 1$$

for $p \in (0, 1)$ and since we wish to upper bound a convex function by concave function it suffices to check the end points $p \in \{0, 1\}$ so the inequality follows. $\square$

We are now in position to prove Theorem 4.9.

Proof of Theorem 4.9. Given $X_1, X_2, \ldots, X_n$ independent, we assume without loss of generality that for $i \leq k$, $\|f_{X_i}\|_\infty \geq 1/2$ and $i > k$ implies $\|f_{X_i}\|_\infty < 1/2$. By Corollary 4.4,

$$\Delta_\infty(X_1 + \cdots + X_n) \geq \Delta_\infty(B_1 + \cdots + B_k + Z_{k+1} + \cdots + Z_n)$$

where B_i and Z_j are all independent, the B_i are Bernoulli satisfying $M(B_i) = M(X_i)$ and Z_i is uniformly distributed on $\{1, 2, \ldots, n_i\}$ where n_i is uniquely determined by $M(X_j) \in (\frac{1}{n_i+1}, \frac{1}{n_i}]$. Note that trivially,

$$\Delta_\infty(X_1 + \cdots + X_n) \geq \max\{\Delta_\infty(B_1 + \cdots + B_k), \Delta_\infty(Z_1 + \cdots + Z_n)\}.$$

Using the variance-min entropy comparisons above (Inequalities (12) and (13)) for Bernoulli variables,

$$\begin{aligned}
\Delta_\infty(B_1 + \cdots + B_n) &\geq 2\mathrm{Var}\left(\sum_i B_i\right) \\
&= 2\sum_i \mathrm{Var}(B_i) \\
&\geq \frac{1}{6}\sum_i \Delta_\infty(B_i) \\
&= \frac{1}{6}\sum_i \Delta_\infty(X_i).
\end{aligned}$$

Similarly, since $Z_1 + \cdots + Z_n$ is an independent sum of symmetric (about the point $\frac{n_i-1}{2}$) log-concave variables, using the Variance- min entropy comparisons, this time for symmetric log-concave variables,

$$\begin{aligned}
\Delta_\infty(Z_{k+1} + \cdots + Z_n) &\geq 2\sum_j \mathrm{Var}(Z_j) \\
&\geq \frac{1}{6}\sum_j \Delta_\infty(Z_i).
\end{aligned}$$

Since $n_j \geq 2$

$$\begin{aligned}
\frac{\Delta_\infty(Z_j)}{\Delta_\infty(X_j)} &\geq \frac{n_j^2 - 1}{(n_j+1)^2 - 1} \\
&\geq \frac{2^2 - 1}{(2+1)^2 - 1} \\
&= \frac{3}{8}.
\end{aligned}$$

Thus it follows that

$$\Delta_\infty(Z_{k+1} + \cdots Z_n) \geq \frac{1}{16}\sum_{j=k+1}^{n} \Delta_\infty(Z_j).$$

Finally,

$$\begin{aligned}
\Delta_\infty(X_1 + \cdots + X_n) &\geq \max\left\{\frac{1}{6}\sum_{i=1}^{k}\Delta_\infty(X_i), \frac{1}{16}\sum_{j=k+1}^{n}\Delta_\infty(X_j)\right\} \\
&\geq \frac{1}{22}\sum_{i=1}^{n}\Delta_\infty(X_i)
\end{aligned}$$

where in the last line we used the fact that $\max(\alpha a, \beta b) \geq \frac{\alpha\beta}{\alpha+\beta}(a+b)$, valid for any non-negative α, β, a, b. $\square$

4.5. **Min-Entropy inequalities: Tightenings, and Reversals.** For the min-entropy we may also give a reversal of min-EPI for Poisson-Binomials.

Theorem 4.13. *For X_i independent Poisson-Binomials,*

$$\Delta_\infty(\sum_i X_i) \leq 6 \sum_i \Delta_\infty(X_i).$$

Proof. By Theorems 4.10 and 3.3 (see Inequality (7))

$$\begin{aligned}\Delta_\infty(X_1 + \cdots + X_n) &\leq 12\mathrm{Var}\left(\sum_i X_i\right)\\ &= 12 \sum_i \mathrm{Var}(X_i)\\ &\leq 6 \sum_i \Delta_\infty(X_i).\end{aligned}$$

□

In the case that X_i are concentrated about a point, one can actually tighten the min-EPI beyond the $\Delta_\infty(\sum_i X_i) \geq \frac{1}{6}\sum_i \Delta_\infty(X_i)$ that one would achieve through variance comparisons in the min-EPI reversals for Poisson-Binomials, as we show in what follows. We will need the following Lemma whose proof is suppressed to the appendix.

Lemma 4.14. *For X a Bernoulli random variable, and $t \in [-\pi, \pi]$,*

$$|\mathbb{E}e^{itX}| \leq e^{-\Delta_\infty(X)t^2/24}. \tag{14}$$

Observe that, when X is expanding the inequality at $t = 0$ shows that the constant $1/24$ is optimal in the latter and we will show that this inequality can be used to derive a sharpening of the min-EPI for for Bernoulli sums.

Theorem 4.15. *For X_i independent and integer valued such that $\|f_{X_i}\|_\infty = c_i \geq \frac{1}{2}$,*

$$\Delta_\infty(X_1 + \cdots + X_n) \geq \frac{\pi^2}{36} \sum_{j=1}^n \Delta_\infty(X_i).$$

Proof. By Corollary 4.8 it suffices to prove the result when the X_i are independent Bernoulli. By Lemma 4.14,

$$\begin{aligned}\|\hat{f}_{X_i}\|_q^q &\leq \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\Delta_\infty(X_i)qt^2/24}dt\\ &= \frac{1}{\sqrt{\pi^2\Delta_\infty(X_i)q/12}}\int_0^{\sqrt{\pi^2\Delta_\infty(X_i)q/12}} e^{-t^2/2}dt.\end{aligned}$$

Thus applying Lemma 2.6 with $\Phi(q) = \frac{1}{\sqrt{\pi^2 q/12}}\int_0^{\sqrt{\pi^2 q/12}} e^{-t^2/2}dt$, and $c_i = \Delta_\infty(X_i)$, we have

$$\|\hat{f}_{\sum_i X_i}\|_q^q \leq \Phi\left(q \sum_i \Delta_\infty(X_i)\right).$$

By Hausdorff-Young, this gives

$$\begin{aligned}\|f_{\sum_i X_i}\|_\infty &\leq \|\hat{f}_{\sum_i X_i}\|_1\\ &\leq \frac{1}{\sqrt{\pi^2(\sum_j \Delta_\infty(X_j))/12}}\int_0^{\sqrt{\pi^2(\sum_j \Delta_\infty(X_j))/12}} e^{-t^2/2}dt.\end{aligned}$$

Applying Lemma 2.9 with $z = \sqrt{\pi^2 \sum_j \Delta_\infty(X_j)/12}$ this gives

$$\|f_{\sum_j X_j}\|_\infty \leq \sqrt{\frac{1}{1 + \frac{\pi^2}{36} \sum_j \Delta_\infty(X_j)}}$$

which yields,

$$\Delta_\infty(\sum_j X_j) \geq \frac{\pi^2}{36} \sum_j \Delta_\infty(X_j).$$

□

Let us note that the largest constant c such that $\Delta_\infty(\sum_i X_i) \geq c \sum_i \Delta_\infty(X_i)$ holds for any collection of independent X_i is no larger than $\frac{1}{2}$, as can be seen by taking X_1 and X_2 to be iid Bernoulli with parameter $p = 1/2$ (and $\pi^2/36 \simeq 0.27 > \frac{1}{4}$). Note that one can alternatively apply Theorem 3.3 and 3.5 to obtain a similar result $\Delta_\infty(X_1 + \cdots + X_n) \geq 2 \sum_{i=1}^n \mathrm{Var}(X_i) \geq \frac{1}{6}\Delta_\infty(X_i)$ at the expense of a constant. Also note that applying the Bernoulli tightening to the proof of Theorem 4.9 gives

$$\begin{aligned}\Delta_\infty(X_1 + \cdots + X_n) &\geq \max\left\{\frac{\pi^2}{36}\sum_{i=1}^k \Delta_\infty(X_i), \frac{1}{16}\sum_{j=k+1}^n \Delta_\infty(X_i)\right\} \\ &\geq \frac{1}{16 + 36/\pi^2} \sum_{i=1}^n \Delta_\infty(X_i),\end{aligned}$$

and an improvement to a constant $c = \frac{1}{16+36/\pi^2} > \frac{1}{20}$ in Theorem 4.9.

5. Littlewood-Offord Problem

In this section we take advantage of various results that we developed in the previous sections to deal with the Littlewood-Offord problem. There has been recent effort in understanding a certain generalization of the Littlewood-Offord problem, that is determining upper bounds on $Q(S, 0) := \max_x \mathbb{P}(S = x)$ where $S := \sum_{i=1}^n v_i B_i$ where $v_i \in \mathbb{R} \setminus \{0\}$ and B_i iid Bernoulli of parameter p. We recall the following question of Fox, Kwan, and Sauermann.

Question ([14] Question 6.2)**.** *For $(v_1, \ldots, v_n) \in (\mathbb{R} - \{0\})^n$ and $B_1, \ldots, B_n$ iid Bernoulli with some parameter $0 < p \leq 1/2$ and $S = v_1 B_1 + \cdots + v_n B_n$. What upper bounds (in terms of n and p) can we give on the maximum point probability $Q(S, 0) = \max_{x \in \mathbb{R}} \mathbb{P}(S = x)$?*

Note that when $p = \frac{1}{2}$ bounds this is as a reformulation of the classical problem, determining the number of subsums that fall in a given location [22]. As usual, we denote by the dot sign the usual scalar product in $\mathbb{R}^n$ so that $S_v = v \cdot B$ with $v = (v_1, \ldots, v_n)$ and $B = (B_1, \ldots, B_n)$. The following lemma appears as Theorem 1.2 in [36]. As we are in position to provide a short proof, we do so.

Lemma 5.1. *Let $B = (B_1, \ldots, B_n)$ such that B_i are independent Bernoulli random variables, for $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$ such that $v_i \neq 0$ for all i, it holds*

$$Q(v \cdot B, 0) \leq \max_{a \in \{-1,1\}^n} Q(a \cdot B, 0).$$

Proof. Consider $\mathbb{R}$ as a vector space over $\mathbb{Q}$ and let $T : \mathbb{R} \to \mathbb{Q}$ be a linear map such that $T(v_i) \neq 0$ for all i. Then

$$\begin{aligned}\mathbb{P}(v_1 B_1 + \cdots + v_n B_n = x) &\leq \mathbb{P}(T(v_1 B_1 + \cdots + v_n B_n) = T(x)) \\ &= \mathbb{P}(T(v_1) B_1 + \cdots + T(v_n) B_n = T(x)),\end{aligned}$$

thus it suffices to consider the situation that the $v_i = \frac{p_i}{q_i} \in \mathbb{Q}$. Multiplying by $q = \prod_{i=1}^n q_i$, gives

$$\mathbb{P}(v_1 B_1 + \cdots + v_n B_n = x) = \mathbb{P}(m_1 B_1 + \cdots + m_n B_n = y)$$

for $m_i \in \mathbb{Z}$. Thus it suffices to consider the case of integer coefficients. In this case, by Theorem 4.7 applied with $\alpha = \infty$

$$\begin{aligned}\mathbb{P}(v_1 B_1 + \cdots + v_n B_n = x) &\leq \max_{z \in \mathbb{Z}} \mathbb{P}((v_1 B_1)^{\#} + \cdots (v_n B_1)^{\#} = z) \\ &= \max_{z \in \mathbb{Z}} \mathbb{P}(a_1 B_1 + \cdots + a_n B_n = z),\end{aligned}$$

where $a_i = sign(v_i) \coloneqq \mathbb{1}_{(0,\infty)}(v_i) - \mathbb{1}_{(-\infty,0)}(v_i)$. □

Theorem 2.3 yields the following consequence.

Corollary 5.2. *For $v_i \in \mathbb{R}\backslash\{0\}$, $S_v = v_1 X_1 + \cdots + v_n X_n$ where X_i are independent Bernoulli variables with variance σ_i, and denoting by $\sigma^2 = \sum_j \sigma_j^2$, then*

$$Q(S_v, 0) \leq \frac{1}{\sqrt{6\sigma^2}} \int_0^{\sqrt{6\sigma^2}} e^{-t^2/2} dt. \tag{15}$$

Remark 5.3. *Let us note that by the fact that $X \mapsto Q(X, 0)$ is lower semi-continuous with respect to the weak topology for X taking values on $\mathbb{Z}$, that the inequality $Q(S, 0) \leq \frac{1}{\sqrt{6\sigma^2}} \int_0^{\sqrt{6\sigma^2}} e^{-t^2/2} dt$ holds when S is Poisson of parameter λ, in which case $\sigma^2 = \lambda$ and we have the following Taylor expansions for λ near zero*

$$Q(S, 0) = e^{-\lambda} = 1 - \lambda + o(\lambda)$$

while by Lemma 2.9 (with $\sigma^2 = \lambda$)

$$\frac{1}{\sqrt{6\sigma^2}} \int_0^{\sqrt{6\sigma^2}} e^{-t^2/2} dt \leq \frac{1}{\sqrt{1 + 2\lambda}} = 1 - \lambda + o(\lambda).$$

Thus in the Poisson regime Inequality (15) is tight for small λ.

Proof. By Lemma 5.1, it suffices to consider the case that $v_i = a_i$ take only the values ± 1. If ℓ denotes the number of terms such that $a_i = -1$,

$$\begin{aligned}S_a &= \sum_{i: a_i = 1} B_i - \sum_{i: a_i = -1} B_i \\ &= \sum_{i: a_i = 1} B_i + \sum_{i: a_i = -1} (1 - B_i) - \ell.\end{aligned}$$

Note that if B is Bernoulli of parameter p then $1 - B$ is Bernoulli of parameter $1 - p$. Hence B and $1 - B$ have the same variance. Thus S_a is the translation of a Bernoulli sums, say $\sum_i X_i$, with X_i of variance σ_i. Therefore applying the Haussdorf-Young Inequality and Theorem 2.8

$$\begin{aligned}Q(S_a, 0) &= \|f_{\sum X_i}\|_\infty \\ &\leq \|\hat{f}_{\sum X_i}\|_1 \\ &\leq \frac{1}{\sqrt{6\sigma^2}} \int_0^{\sqrt{6\sigma^2}} e^{-t^2/2} dt.\end{aligned}$$

□

Corollary 5.4. *When $S_v = v \cdot B$ for $B = (B_1, \dots, B_n)$ for B_i iid Bernoulli of parameter p, ,*

$$Q(S_v, 0) \leq \frac{1}{\sqrt{1 + 2np(1-p)}}.$$

Proof. Applying 2.9 to Corollary 5.2 while observing that $\sigma^2 = np(1-p)$ gives the result. □

Corollary 5.2 and Corollary 5.4 are sharp for any n as can be seen by taking small variance Bernoulli, and in this sense they are optimal in the "Poisson regime". In the Gaussian regime, when the local limit theorem applies, for example for a sequence of iid Bernoulli variables $S = X_1 + \cdots + X_n$ gives

$$Q(S, 0) \leq \frac{1}{\sqrt{2\pi\sigma_S^2}} + o(1/\sqrt{n}).$$

Meanwhile by integrating on the whole $[0, \infty)$, $\frac{1}{\sqrt{6\sigma^2}} \int_0^{\sqrt{6\sigma^2}} e^{-t^2/2} dt \leq \frac{1}{\sqrt{6\sigma^2}} \int_0^{\infty} e^{-t^2/2} dt = \sqrt{\frac{\pi}{12\sigma^2}}$, so the bounds cannot be improved by more than a constant factor in the Gaussian regime. In particular when the X_i are iid Bernoulli with parameter $1/2$, then by Erdös's sharp solution (see [13]), to the Littlewood-Offord problem, $Q(S, 0) \leq 2^{-n} \begin{pmatrix} n \\ \lfloor n/2 \rfloor \end{pmatrix} \approx \sqrt{\frac{2}{\pi n}}$, while $\sqrt{\frac{\pi}{12\sigma^2}} = \sqrt{\frac{\pi}{3}} \frac{1}{\sqrt{n}}$ (and we are off by a factor of $\pi/\sqrt{6} \simeq 1.28$).

The Littlewood-Offord problem admits the following generalization in terms of the Rényi entropy.

Theorem 5.5. *For $S_v = v_1 B_1 + \cdots + v_n B_n$, where $v_i \neq 0$, B_i independent Bernoulli of variance σ_i^2 and $\alpha \geq 2$,*

$$\begin{aligned} H_\alpha(S_v) &\geq \log \frac{\sqrt{6\sigma^2\alpha'}}{\int_0^{\sqrt{6\sigma^2\alpha'}} e^{-t^2/2} dt} \\ &\geq \frac{1}{2} \max \left\{ \log(1 + 2\alpha'\sigma^2), \log\left(\frac{12\alpha'\sigma^2}{\pi}\right) \right\}, \end{aligned}$$

where $\sigma^2 = \sum_i \sigma_i^2$.

Proof. Let us reduce to the case that $v_i = \pm 1$. Choose a linear function[4] $T : \mathbb{R} \to \mathbb{Q}$ such that $T(v_i) \neq 0$. Further since $T(v_i) = p_i/q_i$ for $q_i, p_i \in \mathbb{Z}$, writing the function $\tilde{T}(x) = qT(x)$ where $q = \prod_{i=1}^n q_i$, then

$$H_p(S_v) \geq H_p(\tilde{T}(S_v))$$

since (by majorization for example) deterministic functions of a random variable decrease Rényi entropy. Since $\tilde{T}(S_v)$ can be expressed as $k_1 B_1 + \cdots + k_n B_n$ for $k_i \in \mathbb{Z} \setminus \{0\}$. By Theorem 4.7,

$$\begin{aligned} H_p(k_1 B_1 + \cdots + k_n B_n) &\geq H_p((k_1 B_1)^\# + \cdots + (k_n B_n)^\#) \\ &= H_p(a_1 B_1 + \cdots + a_n B_n), \end{aligned}$$

where $a_i = sign(k_i)$. Thus, it suffices to consider the case that $v_i = a_i = \pm 1$.

[4] To find such a linear map, choose a Hamel Basis for $\mathbb{R}$ over $\mathbb{Q}$, then $v_1, \dots, v_n$ can be considered as non-zero elements in a finite dimensional vector space over $\mathbb{Q}$. It is obvious, that when $n = 1$ such a linear map can be found. By induction, choose a linear map such that $T(v_i) \neq 0$ for $i < n$, and in the case that $T(v_n) = 0$ (else we are done!), S such that $S(v_n) \neq 0$. Then choose a map of the form $\lambda T + S$ for $\lambda \in \mathbb{Q}$ but $\lambda \neq S(v_i)/T(v_i)$ for $i < n$.

Since $1-B$ Bernoulli of parameter $1-p$ for B Bernoulli p, $S_a = X_1 + \cdots + X_n - \ell$ where $X_i = B_i$ when $a_i = 1$ and $X_i = 1 - B_i$ when $a_i = -1$ and $\ell = |\{i : a_i = -1\}|$. Thus applying Theorem 3.3,

$$\begin{aligned} H_p(a_1 B_1 + \cdots + a_n B_n) &= H_p(X_1 + \cdots + X_n) \\ &\geq \log \frac{\sqrt{6\sigma^2 q}}{\int_0^{\sqrt{6\sigma^2 q}} e^{-t^2/2} dt}. \end{aligned}$$

The second inequality follows from Lemma 2.9. □

Appendices

In this appendix we will successively prove Lemma 2.2, Lemma 2.5 and Lemma 4.14.

Proof of Lemma 2.2. Recall that for w with distribution function W, when $w^s \in L^1$ it has distribution function $t \mapsto W(t^{\frac{1}{s}})$, so that by equation (3) and a change of variables,

$$\int_{\mathbb{R}} w^s = \int_0^\infty s t^{s-1} W(t) dt.$$

Similarly when $w^s - v^s \in L^1$ one obtains an analogous formula for φ,

$$\varphi(s) = t_0^{-1} \int_0^\infty \left(\frac{t}{t_0} \right)^{s-1} (W(t) - V(t)) dt.$$

Thus for $s > s'$, we have

$$\varphi(s) - \varphi(s') = \varphi(s) = t_0^{-1} \int_0^\infty \left(\left(\frac{t}{t_0} \right)^{s-1} - \left(\frac{t}{t_0} \right)^{s'-1} \right) (W(t) - V(t)) dt.$$

Note that this completes the proof since the integrand above is non-negative. □

Proof of Lemma 2.5. Observe that $w_\lambda(t) - v_\lambda(t)$ has no more than one zero on $(0, \pi)$ if and only if $H(t) := w_\lambda(t)^2 - v_\lambda(t)^2$ has no more than one zero on $(0, \pi)$. Taking a derivative, we see that for $\lambda > 0$ and small enough t

$$H'(t) = \lambda \sin t - \frac{6t\lambda e^{-3\lambda t^2/\pi^2}}{\pi^2} > 0.$$

Further we see that for $\lambda > 0$, $H'(t) = 0$ iff $\lambda = \frac{\pi^2 \log\left(\frac{6t}{\pi^2 \sin t}\right)}{3t^2}$. Now we claim that the function

$$(0, \pi) \ni t \mapsto \lambda(t) = \frac{\pi^2 \log\left(\frac{6t}{\pi^2 \sin t}\right)}{3t^2}$$

is strictly increasing and therefore one to one from $(0, \pi)$ into $(-\infty, \infty)$ (since $\lim_{t \to 0} \lambda(t) = -\infty$ and $\lim_{t \to \pi} \lambda(t) = +\infty$). Hence for fixed λ, there exists exactly one $t = t_\lambda$ such that $H'(t) = 0$. Computing

$$\lambda'(t) = \frac{\pi^2}{3t^2} \left(-t \cot(t) + 2 \log\left(\frac{\pi^2 \sin(t)}{6t} \right) + 1 \right)$$

and since $\lim_{t \to 0} \lambda'(t) = \infty$, it suffices to show that $\lambda'(t)$ has no zeros on $(0, \pi)$. That is that

$$f(t) := 1 - t \cot(t) + 2 \log\left(\frac{\pi^2 \sin(t)}{6t} \right)$$

has no zeros on $(0,\pi)$. Note that $\lim_{t\to 0} f(t) = 2\log\left(\frac{\pi^2}{6}\right) > 0$, so it is enough to show that f is increasing on $(0,\pi)$. We have, for $t\in(0,\pi)$

$$f'(t) = 1+\cot(t)+\frac{t}{\sin^2(t)}-\frac{2}{t} \qquad \text{and} \qquad f''(t) = \frac{2}{t^2}-\frac{2t\cos(t)}{\sin^3(t)}.$$

Now we claim that $f''(t)>0$ on $(0,\pi)$ which is equivalent to saying that

$$\sin^3(t) - t^3\cos(t) > 0, \qquad t\in(0,\pi).$$

For $t\geq \pi/2$ this is immediate. For $t<\pi/2$ we use the Taylor series bounds $\sin(t)\geq t-t^3/6$ and $\cos(t)\leq 1-t^2/2+t^4/24$,

$$\begin{aligned}\sin^3(t)-t^3\cos(t) &\geq (t-t^3/6)^3 - t^3(1-t^2/2+t^4/24)\\ &= t^7(3-t)(t+3)/216\end{aligned}$$

which is clearly positive on $(0,\pi/2)$. The claim is proved and hence f' is increasing. Since $\lim_{t\to 0} f'(t) = 1 > 0$, we infer that f is increasing on $(0,\pi)$ as expected.

Thus H' is positive for small t and has at most one zero, and thus it follows that H has at most one 0 on $(0,\pi)$ since $H(0)=0$, and H is increasing and then decreasing. □

Proof of Lemma 4.14. Since both sides of (14) are invariant in the transformation of the parameter $p\mapsto 1-p$, we may assume $p\in[1/2,1]$ and compute explicitly with (4), it is equivalent to prove

$$(1-p)^2+p^2+2p(1-p)\cos(t) \leq e^{-\left(\frac{1}{p^2}-1\right)t^2/12}.$$

Setting $q=1/p\in[1,2]$, the desired inequality is equivalent to proving that

$$F(q) := q^2 e^{-\frac{q^2-1}{12}t^2} - 2(q-1)\cos(t) - (q-1)^2 - 1 \geq 0.$$

Our first aim is to prove that F is concave on in the interval $[1,2]$ for any given $t\in[0,\pi]$. Observe that

$$F'(q) = \left(2q - \frac{q^3t^2}{6}\right) e^{-\frac{q^2-1}{12}t^2} - 2\cos(t) - 2(q-1),$$

and

$$F''(q) = \left(2 - 2e^{\frac{q^2-1}{12}t^2} + \frac{q^2t^2}{6}\left(-5+\frac{q^2t^2}{6}\right)\right) e^{-\frac{q^2-1}{12}t^2}.$$

Given $t\in[0,\pi]$, set $r=\frac{q^2t^2}{6}\in[\frac{t^2}{6};\frac{2t^2}{3}]$ and $G(r) := 2-2e^{\frac{r}{2}}e^{-\frac{t^2}{12}} + r(-5+r)$ so that F is concave on $[1,2]$ reduces to proving that G is negative on $[\frac{t^2}{6};\frac{2t^2}{3}]$. Observe that

$$G'(r) = -e^{\frac{r}{2}}e^{-\frac{t^2}{12}} + 2r - 5 \qquad \text{and} \qquad G''(r) = -\frac{1}{2}e^{\frac{r}{2}}e^{-\frac{t^2}{12}} + 2.$$

We infer that G'' is decreasing on $[\frac{t^2}{6};\frac{2t^2}{3}]$ and may change sign depending on the value of the parameter t. We need to distinguish between two cases.

(1) Assume first that $t\leq\sqrt{4\log 4}$. Then $G''(2t^2/3) = -\frac{1}{2}e^{\frac{t^2}{4}}+2\geq 0$. In that case we conldude that $G''\geq 0$ on the whole interval and therefore that G' is non-decreasing. Hence $G'(r)\leq G'(2t^2/3) = -e^{\frac{t^2}{4}} + \frac{4}{3}r^2 - 5$. It is easy to see that the mapping $[0,\infty)\ni t\mapsto H(t) := -e^{\frac{t^2}{4}} + \frac{4}{3}r^2 - 5$ is increasing on $[0,\sqrt{4\log\frac{16}{3}}]$ so that, for $t\in[0,\sqrt{4\log 4}]$, $H(t)\leq H(\sqrt{4\log 4}) = -9+\frac{16}{3}\log(4)\simeq -1.6\leq 0$. Therefore $G'(r)\leq 0$ and hence $G(r)\leq G(t^2/6) = \frac{t^2}{6}(-5+\frac{t^2}{6})\leq 0$ since $t\in[0,\sqrt{4\log 4}]$. As an intermediate conclusion we proved that $G\leq 0$ in case (1).

(2) Assume now that $t \geq \sqrt{4 \log 4}$. Then G'' changes sign. Namely, since $G''(t^2/6) = 3/2 \geq 0$ and $G''(2t^2/3) = -\frac{1}{2}e^{\frac{t^2}{4}} + 2 \leq 0$, G'' is non-negative on $[\frac{t^2}{6}, r_o]$ and non-positive on $[r_o, \frac{2t^2}{3}]$, with $r_o := 2(\log 4 + \frac{t^2}{12})$. It follows that, for $t \leq \pi$ and $r \in [\frac{t^2}{6}; \frac{2t^2}{3}]$, $G'(r) \leq G'(r_o) = 4(\log 4 + \frac{t^2}{12}) - 9 \leq 4 \log 4 + \frac{\pi^2}{3} - 9 \simeq -0.16 \leq 0$. We conclude that G is non-increasing and therefore that $G(r) \leq G(t^2/6) = \frac{t^2}{6}\left(-5 + \frac{t^2}{6}\right) \leq 0$ since $t \leq \pi$. As a conclusion we proved that $G \leq 0$ in case (2) and therefore in any case. This shows that F is concave.

Now F being concave, $F \geq 0$ is a consequence of the fact that $F(1) = 0$ and $F(2) = 4e^{-\frac{t^2}{4}} - 2\cos(t) - 2 \geq 0$. To see the latter, one can observe that $2\cos(t) + 2 = 4\cos^2(t/2)$ so that $F(2) \geq 0$ is equivalent to saying that $e^{t^2/4} \geq \cos^2(t/2)$ on $[0, \pi]$ which in turn is equivalent to saying that $e^{u^2/2} \geq \cos(u)$ for any $u = \frac{t}{2} \in [0, \frac{\pi}{2}]$. Taking the logarithm, we end up with proving that $I(u) := -\frac{u^2}{2} - \log\cos(u) \geq 0$ on $[0, \frac{\pi}{2}]$. Since $I'(u) = -u + \tan(u) \geq 0$ the desired conclusion immediately follows. This ends the proof of the Lemma. $\square$

Acknowledgements

The authors thank Arnaud Marsiglietti for stimulating discussion and in particular suggesting the connection to the Littlewood-Offord question of [14].

References

[1] W. Beckner. Inequalities in Fourier analysis. *Ann. of Math. (2)*, 102(1):159–182, 1975.

[2] S. Bobkov and M. Madiman. The entropy per coordinate of a random vector is highly constrained under convexity conditions. *IEEE Trans. Inform. Theory*, 57(8):4940–4954, August 2011.

[3] S. Bobkov, A. Marsiglietti, and J. Melbourne. Concentration functions and entropy bounds for discrete log-concave distributions. Preprint available at arxiv:2007.11030, 2020.

[4] S. G. Bobkov and G. P. Chistyakov. Bounds for the maximum of the density of the sum of independent random variables. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI)*, 408(Veroyatnost i Statistika. 18):62–73, 324, 2012.

[5] S. G. Bobkov and G. P. Chistyakov. Entropy power inequality for the Rényi entropy. *IEEE Trans. Inform. Theory*, 61(2):708–714, February 2015.

[6] S. G. Bobkov and G. P. Chistyakov. On concentration functions of random variables. *J. Theoret. Probab.*, 28:976–988, 2015.

[7] S.G. Bobkov and A. Marsiglietti. Variants of the entropy power inequality. *IEEE Transactions on Information Theory*, 63(12):7747–7752, 2017.

[8] J. Bourgain. On high-dimensional maximal functions associated to convex bodies. *Amer. J. Math.*, 108(6):1467–1476, 1986.

[9] H. J. Brascamp and E. H. Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Functional Analysis*, 22(4):366–389, 1976.

[10] M.H.M. Costa and T.M. Cover. On the similarity of the entropy power inequality and the Brunn-Minkowski inequality. *IEEE Trans. Inform. Theory*, 30(6):837–839, 1984.

[11] T. Courtade, M. Fathi, and A. Pananjady. Quantitative stability of the entropy power inequality. *IEEE Transactions on Information Theory*, 64(8):5691–5703, 2018.

[12] A. Dembo, T.M. Cover, and J.A. Thomas. Information-theoretic inequalities. *IEEE Trans. Inform. Theory*, 37(6):1501–1518, 1991.

[13] P. Erdös. On a lemma of Littlewood and Offord. *Bull. Amer. Math. Soc.*, 51:898–902, 1945.

[14] J. Fox, M. Kwan, and L. Sauermann. Combinatorial anti-concentration inequalities, with applications. *arXiv preprint arXiv:1905.12142*, 2019.

[15] R. J. Gardner. The Brunn-Minkowski inequality. *Bull. Amer. Math. Soc. (N.S.)*, 39(3):355–405 (electronic), 2002.

[16] S. Haghighatshoar, E. Abbe, and E. Telatar. A new entropy power inequality for integer-valued random variables. *IEEE Trans. Inform. Th.*, 60(7):3787–3796, July 2014.

[17] P. Harremoës and C. Vignat. An entropy power inequality for the binomial family. *J. Inequal. Pure Appl. Math.*, 4(5):Article 93, 6 pp. (electronic), 2003.

[18] O. Johnson and Y. Yu. Monotonicity, thinning, and discrete versions of the entropy power inequality. *IEEE Trans. Inform. Theory*, 56(11):5387–5395, 2010.
[19] J. Li. Rényi entropy power inequality and a reverse. *Studia Mathematica*, 242:303–319, 2018.
[20] J. Li, A. Marsiglietti, and J. Melbourne. Further investigations of rényi entropy power inequalities and an entropic characterization of s-concave densities. In *Geometric Aspects of Functional Analysis*, pages 95–123. Springer, 2020.
[21] E. H. Lieb. Proof of an entropy conjecture of Wehrl. *Comm. Math. Phys.*, 62(1):35–41, 1978.
[22] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation. III. *Rec. Math. [Mat. Sbornik] N.S.*, 12(54):277–286, 1943.
[23] M. Madiman, J. Melbourne, and P. Xu. Forward and reverse entropy power inequalities in convex geometry. *Convexity and Concentration*, pages 427–485, 2017.
[24] M. Madiman, J. Melbourne, and P. Xu. Rogozin's convolution inequality for locally compact groups. *arXiv preprint arXiv:1705.00642*, 2017.
[25] M. Madiman, L. Wang, and J. O. Woo. Majorization and Rényi entropy inequalities via Sperner theory. *Discrete Mathematics*, 342(10):2911–2923, 2019.
[26] A. W. Marshall, I. Olkin, and B. C. Arnold. *Inequalities: theory of majorization and its applications*, volume 143. Springer, 1979.
[27] A. Marsiglietti and J. Melbourne. On the entropy power inequality for the Rényi entropy of order [0, 1]. *IEEE Transactions on Information Theory*, 65(3):1387–1396, 2018.
[28] J. Melbourne and T. Tkocz. Reversals of Rényi entropy inequalities under log-concavity. *arXiv preprint arXiv:2005.10930*, 2020.
[29] F. L. Nazarov and A. N. Podkorytov. Ball, Haagerup, and distribution functions. In *Complex analysis, operators, and related topics*, volume 113 of *Oper. Theory Adv. Appl.*, pages 247–267. Birkhäuser, Basel, 2000.
[30] J. C. Principe. *Information theoretic learning: Renyi's entropy and kernel perspectives*. Springer Science & Business Media, 2010.
[31] E. Ram and I. Sason. On Rényi entropy power inequalities. *IEEE Transactions on Information Theory*, 62(12):6800–6815, 2016.
[32] A. Rényi. On measures of entropy and information. In *Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I*, pages 547–561. Univ. California Press, Berkeley, Calif., 1961.
[33] O. Rioul. Rényi entropy power inequalities via normal transport and rotation. *Entropy*, 20(9):641, 2018.
[34] M. Rudelson and R. Vershynin. Small ball probabilities for linear images of high-dimensional distributions. *Int. Math. Res. Not. IMRN*, (19):9594–9617, 2015.
[35] C.E. Shannon. A mathematical theory of communication. *Bell System Tech. J.*, 27:379–423, 623–656, 1948.
[36] M. Singhal. Erdos-Littlewood-Offord problem with arbitrary probabilities. *arXiv preprint arXiv:1912.02886*, 2019.
[37] W. Tang and F. Tang. The Poisson binomial distrubution – old & new. Preprint available at arXiv:1908.10024, 2019.
[38] T. Tao and V. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

University of Delaware, Department of Mathematical Sciences, 501 Ewing Hall, Newark DE 19716, USA.

Centro de Investigación en Matemáticas, Col. Valenciana CP: 36023, Apartado Postal 402, Guanajuato, Gto, Mexico.

Université Paris Nanterre, Modal'X, FP2M, CNRS FR 2036, 200 avenue de la République 92000 Nanterre, France.

Email address: `james.melbourne@cimat.mx, madiman@udel.edu, croberto@math.cnrs.fr`