Monotonicity, thinning and discrete versions of the Entropy Power Inequality

Oliver Johnson and Yaming Yu Member, IEEE

Abstract—We consider the entropy of sums of independent discrete random variables, in analogy with Shannon’s Entropy Power Inequality, where equality holds for normals. In our case, infinite divisibility suggests that equality should hold for Poisson variables. We show that some natural analogues of the Entropy Power Inequality do not in fact hold, but propose an alternative formulation which does always hold. The key to many proofs of Shannon’s Entropy Power Inequality is the behaviour of entropy on scaling of continuous random variables. We believe that Rényi’s operation of thinning discrete random variables plays a similar role to scaling, and give a sharp bound on how the entropy of ultra log-concave random variables behaves on thinning.

In the spirit of the monotonicity results established by Artstein, Ball, Barthe and Naor, we prove a stronger version of concavity of the entropy of ultra log-concave random variables behaves on thinning. In particular Shannon [1] stated Theorem 1.1, known as the Entropy Power Inequality. However, results concerning the Central Limit Theorem or in practical applications may come first, with the differential entropy of continuous random variables coming later. Stamp [2] and by Blachman [3] using an argument based on the Power Inequality (EPI), which was later rigorously proved by Rényi [4]. We prove a stronger version of concavity of entropy, which implies a strengthened form of our discrete Entropy Power Inequality.

Keywords: convolution, discrete random variables, entropy, Entropy Power Inequality, monotonicity, Poisson distribution, thinning.

MSC2000 Classification: Primary 94A17 Secondary 62B10; 60E07.

I. REVIEW OF PREVIOUS WORK

It is natural to consider the entropy of the sum of independent random variables, for example in proving theoretical results concerning the Central Limit Theorem or in practical models of information transmission involving addition of noise to the signal.

Pedagogically speaking, the entropy $H$ of discrete random variables usually comes first, with the differential entropy $h$ of continuous random variables coming later. However, results from functional analysis imply properties of the differential entropy which do not yet have discrete counterparts. For example Shannon [1] stated Theorem 1.1, known as the Entropy Power Inequality (EPI), which was later rigorously proved by Stam [2] and by Blachman [3] using an argument based on the heat equation. Write $E(t) = \frac{1}{2} \log(2\pi e t)$ for the entropy of a Gaussian random variable with finite variance $t$, and define $v(X) = E^{-1}(h(X)) = e^{2h(X)/2}\sqrt{2\pi e}$ for the entropy power of random variable $X$ with differential entropy $h(X)$. (We use $log$ to represent the natural logarithm throughout this paper).

Theorem 1.1 (EPI): For independent continuous $X$ and $Y$, the sum $X + Y$ satisfies

$$v(X + Y) \geq v(X) + v(Y),$$

with the only non-trivial case of equality being when $X$ and $Y$ are Gaussian.

A key role is played in many proofs of Theorem 1.1 by the operation of scaling of continuous random variables, using the fact that for any $\alpha$,

$$v(\sqrt{\alpha}X) = \alpha v(X).$$

One major contribution of this paper is Theorem 2.4 below, which shows that a one-sided version of (1) holds for discrete random variables. In this case, the operation of scaling is replaced by the thinning operation introduced by Rényi [4].

As is implicit in the work of Verdú and Guo [5], Theorem 1.1 can be rephrased in terms of scalings, in the form of Corollary 1.2 below. Lieb [6] and Dembo, Cover and Thomas [7] prove the Entropy Power Inequality by working with the Rényi entropy (a generalization of Shannon’s quantity). They use properties of $p$-norms on convolution given by Beckner’s sharp form [8] of the Young inequality. Using a particular parameterization, they show that this Young inequality implies that the differential entropy is concave with respect to normalized linear combinations, that is, for any $0 \leq \alpha \leq 1$:

$$h(\sqrt{\alpha}X + \sqrt{1-\alpha}Y) \geq \alpha h(X) + (1-\alpha)h(Y).$$

The papers [7, 6] show that (3) is equivalent to the Entropy Power Inequality. The form of $\alpha$ used in this proof suggests the following result:

Corollary 1.2: Given independent random variables $X$ and $Y$ with finite and non-zero entropy power, there exist $X^*$ and $Y^*$ such that $X = \sqrt{\alpha}X^*$ and $Y = \sqrt{1-\alpha}Y^*$ for some $0 < \alpha < 1$, and such that $h(X^*) = h(Y^*)$. The Entropy Power Inequality Theorem 1.1 is equivalent to the fact that

$$h(X + Y) \geq h(X^*),$$

with equality if and only if $X$ and $Y$ are Gaussian.

Proof: Applying (1) and taking $\alpha = v(X)/(v(X) + v(Y))$ ensures that $X^* = X/\sqrt{\alpha}$ and $Y^* = Y/\sqrt{1-\alpha}$ have the property that $v(X^*) = v(Y^*) = v(X) + v(Y)$.

Assume (1). Since $h(X + Y) \geq h(X^*)$, applying $E^{-1}$ to both sides we deduce that $v(X + Y) \geq v(X^*)$, which equals $v(X) + v(Y)$, so that the EPI (1) holds.

Assume (1). Since $v(X + Y) \geq v(X) + v(Y) = v(X^*)$, so applying $E$ to both sides, we deduce (1).

It is natural to conjecture that there should be a version of the EPI for discrete entropies $H$. We will show in Theorem 2.5 that an equivalent of this rephrased EPI does hold for discrete variables, whereas in Section V we show that some other apparently natural versions of Theorem 1.1 in fact fail.
In the context of sums of independent continuous random variables, Artstein, Ball, Barthe and Naor [9] proved a stronger type of result, referred to as monotonicity. Alternative proofs were later given by Tulino and Verdú [10] and by Madiman and Barron [11]. For example, Theorem 2 of [9] gives the following:

Theorem 1.3: Given independent continuous random variables \( X_i \) with finite variance, for any positive \( \alpha_i \) such that \( \sum_{i=1}^{n+1} \alpha_i = 1 \), writing \( \alpha^{(j)} = \sum_{i \neq j} \alpha_i = 1 - \alpha_j \), then

\[
nh \left( \sum_{i=1}^{n+1} \sqrt{\alpha_i} X_i \right) \geq \sum_{j=1}^{n+1} \alpha^{(j)} h \left( \sum_{i \neq j} \sqrt{\alpha_i / \alpha^{(j)}} X_i \right).
\]

This is called monotonicity since, choosing \( \alpha_i = 1/(n+1) \), it implies that for independent and identically distributed \( X_i \), the entropy of the normalized sum \( h \left( \sum_{i=1}^{n} X_i / \sqrt{n} \right) \) is monotone increasing in \( n \). Equivalently writing \( d(X) = D(X || \phi_{\kappa X, \sigma^2_{\kappa} X}) \) for the relative entropy from \( X \) to a normal of the same mean and variance, the relative entropy of the normalized sum \( d \left( \sum_{i=1}^{n} X_i / \sqrt{n} \right) \) is monotone decreasing in \( n \).

The other major contribution of this paper is Theorem 3.2, which establishes a discrete analogue of Theorem 1.3. Such monotonicity results as Theorem 1.3 imply strengthened Entropy Power Inequalities. By choosing

\[
\alpha^{(l)} = \frac{nv \left( \sum_{i \neq j} Y_i \right)}{\sum_{j=1}^{n+1} v \left( \sum_{i \neq j} Y_i \right)},
\]

(in the case that all \( \alpha^{(l)} \leq 1 \); if not, the result is automatic) Artstein et al. [9] showed that their Theorem 1.3 implies the following extension of the EPI, Theorem 1.1

Theorem 1.4: Given independent continuous random variables \( Y_i \) with finite variance, the entropy powers satisfy

\[
nv \left( \sum_{i=1}^{n+1} Y_i \right) \geq \sum_{j=1}^{n+1} v \left( \sum_{i \neq j} Y_i \right).
\]

We observe that this strengthened EPI, Theorem 1.4, can be expressed in a similar way to Corollary 1.2. That is, given independent random variables \( Y_i \), if there exist \( \alpha_i \) such that \( \sum_{i=1}^{n+1} \alpha_i = 1 \) and \( Y^*_i = Y_i / \sqrt{\alpha_i} \) have entropies such that \( h \left( \sum_{i \neq j} \sqrt{\alpha_i} Y^*_i \right) / \sqrt{\alpha^{(j)} Y^*_j} \) are constant in \( j \), then

\[
h \left( \sum_{i=1}^{n+1} Y_i \right) \geq h^*.
\]

This again follows by observing that for each \( j \), 2 implies that \( v \left( \sum_{i \neq j} Y_i \right) = v^* \alpha^{(j)} = e^{2h^* \alpha^{(j)} / (2\pi e)} \), so summing over \( j \), the RHS of Theorem 1.4 is equal to \( e^{2h^* n / (2\pi e)} \), and the result follows. Note that in this case, the choice of \( \alpha^{(l)} \) again coincides with that given by 5. In Theorem 3.3 we prove a discrete version of this result.

The structure of the remainder of the paper is as follows. In Section II we introduce the thinning operation, and describe the resulting analogues of the EPI, Theorem 1.1. In Section III we show how these results can extend to provide monotonicity results. In Section IV we discuss two natural versions of the EPI which are not true. In the self-contained Appendices A and B we prove the two main results of the paper, namely the scaling result Theorem 2.3 and the monotonicity result Theorem 3.2. Although these results are related, they are proved independently, the first using a semigroup argument similar to that in 12 and the second using an examination of certain Hessian terms, and previous results from 13.

There has been considerable interest in proving an Entropy Power Inequality for discrete random variables. Some authors [14], [15], [16], [17] have focused on replacing the operation of integer addition + by modulo 2 addition \( \oplus \), and obtained similar results in that case. As in 13, we prefer to retain + as integer addition. Harremoës and Vignat [19] proved that 1 holds when \( X \) and \( Y \) are any independent binomial \( \text{Bin}(nx, 1/2) \) and \( \text{Bin}(ny, 1/2) \) random variables, on re-defining \( v(X) = e^{2H(X) / (2\pi e)} \) (simply replacing differential entropies \( h \) by discrete entropies \( H \)). We prefer to conjecture that the discrete version of the Entropy Power Inequality should be expressed differently, using the entropy of the Poisson distribution.

II. ENTROPY AND THINNING

Recent work of Harremoës, Johnson and Kontoyiannis [20], [21] shows that, in many senses related to Information Theory, the equivalent of scaling continuous random variables by a factor of \( \sqrt{\alpha} \) is the thinning operation \( T_\alpha \) on discrete random variables, as introduced by Rényi [4].

Definition 2.1: The \( \alpha \)-thinned version of random variable \( Y \) is given by the random sum \( T_\alpha Y = \sum_{i=1}^{n} B_i \), where the \( B_1, B_2, \ldots \) are IID Bernoulli Bern(\( \alpha \)), all independent of \( Y \).

We write \( E(t) = H(T_{\alpha t}) \), an increasing concave function, for the entropy of a Poisson random variable \( \Pi_t \) of mean \( t \), and define an analogue of the entropy power as \( V(X) = E^{-1}(H(X)) \). Theorem 2.5 of [12] proves that \( \Pi_{\lambda X} \) maximises entropy within the class of ultra log-concave (ULC) random variables \( X \) (see below) of given mean \( \lambda X \), or that \( V(X) \leq \lambda X \). We investigate the entropy of sums in the context of this restricted ULC class.

Definition 2.2: The ULC random variables are those whose probability mass functions \( P \) satisfy

\[
iP(i)^2 \geq (i+1)P(i+1)P(i-1), \quad \text{for all } i \geq 1.
\]

The ULC class includes the Poisson family and Bernoulli sums. This class was introduced in combinatorics [22], [23], a context in which the Bernoulli random variables are a natural fundamental building block.

The results outlined in [20], [21] suggest an equivalence between scaling by \( \sqrt{\alpha} \) and thinning by \( \alpha \). This idea has developed with the fact that for discrete random variables, a natural equivalent of 5 is given by the following Thinned Entropy Concavity Inequality, proved by Yu and Johnson in [24], extending results in [13], and now a consequence of the more general Theorem 3.2 below.

Theorem 2.3 (TECI): For independent ULC random variables \( X \) and \( Y \), for any \( 0 \leq \alpha \leq 1 \)

\[
H(T_{\alpha}X + T_{1-\alpha}Y) \geq \alpha H(X) + (1-\alpha)H(Y).
\]
For $0 < \alpha < 1$, examination of the proof shows that equality holds if and only if $X$ and $Y$ are Poisson with the same mean.

One major contribution of the present paper is the following theorem, which shows that for ULC random variables a one-sided equivalent of \[2\] holds. This result is proved in Appendix A using a semigroup designed to preserve entropy, a development of techniques in [12]. We refer to this result as the Restricted Thinned Entropy Power Inequality (RTPEI), since it is a special case of the Thinned Entropy Power Inequality [15].

**Theorem 2.4 (RTPEI):** Given any ULC random variable $X$, $V(T_\alpha X) \geq \alpha V(X)$, for any $0 \leq \alpha \leq 1$.

In the continuous case, the equivalents of Theorems 2.3 and 2.4 allowed the full EPI. Theorem 1.1 to be deduced. Despite this, in Section IV we describe how two apparently natural equivalents of the EPI, namely [13] and [15], in fact fail in general. These results are stated as Example 4.1 and 4.2 respectively. In Theorem 4.3 we discuss some conditions under which these results do hold.

However we can prove a discrete analogue of the rephrased Entropy Power Inequality, Corollary 1.2. The key operation is to invert the thinning operation $T_\alpha$ on $X$, to create random variables $X^*$. This additional restriction means that the result holds in less generality than Corollary 1.2.

**Theorem 2.5:** Given independent ULC random variables $X$ and $Y$, suppose there exist $X^*$ and $Y^*$ such that $X = T_\alpha X^*$ and $Y = T_{1-\alpha} Y^*$ for some $0 < \alpha < 1$, and such that $H(X^*) = H(Y^*)$. Then

$$H(X + Y) \geq H(X^*),$$

with equality if and only if $X$ and $Y$ are Poisson.

**Proof:** In analogy with the proof of Corollary 1.2 for any $\alpha$ we define $X^*_{\alpha}$ and $Y^*_{\alpha}$ (if such random variables exist) such that $X = T_\alpha X^*_{\alpha}$ and $Y = T_{1-\alpha} Y^*_{\alpha}$. The Thinned Entropy Concavity Inequality, Theorem 2.3 implies that

$$H(X + Y) = H(T_\alpha X^* + T_{1-\alpha} Y^*) \geq \alpha H(X^*_{\alpha}) + (1-\alpha) H(Y^*_{\alpha}).$$

This bound will hold for any $\alpha$, so choosing $\alpha$ such that $H(X^*_{\alpha}) = H(Y^*_{\alpha})$, we deduce the result.

Unlike the continuous case, in general we cannot prove that this is the right choice of $\alpha$, by optimizing (9). However, we can give a related bound which we optimize, giving an alternative heuristic as to the right value of $\alpha$ to choose. That is, by Theorem 2.4 we deduce that

$$\alpha H(X^*_{\alpha}) + (1-\alpha) H(Y^*_{\alpha}) \leq \alpha E\left(\frac{V(X)}{\alpha}\right) + (1-\alpha) E\left(\frac{V(Y)}{1-\alpha}\right).$$

Because $E(V)$ is concave, the RHS of (10) is maximized by $\alpha = V(X)/(V(X) + V(Y))$.

Note that it is not always possible to find $X^*$ and $Y^*$ as required in Theorem 2.5. For example, for $X \sim \text{Bern}(p)$, there only exists $X^*$ such that $X = T_\alpha X^*$ when $\alpha \geq p$. In general, for any random variable $X$ with support on $\{0,\ldots,L\}$, there does not exist $X^*$ such that $X = T_\alpha X^*$ for $\alpha < E(X)/L$ (since thinning preserves the support, then $L \geq E(X^*) = E(X)/\alpha$). Such an $X^*$ will exist for all $\alpha$ when $X$ lies in certain parametric families, including the geometric and Poisson, since these are preserved by thinning (see [21]).

Some examples illustrate the bounds of Theorem 2.5

**Example 2.6:** Using Theorem 2.5

1) Given $X \sim \text{Poisson}(\lambda)$ and $Y \sim \text{Poisson}(\mu)$, take $X^* \sim Y^* \sim \text{Poisson}(\lambda + \mu)$ and $\alpha = \lambda/((\lambda + \mu)$, to confirm that equality does indeed hold in (8) in this case.

2) Given $X \sim \text{Bin}(n, p)$ and $Y \sim \text{Bin}(n, q)$, if $p + q \leq 1$ then choosing $X^* \sim Y^* \sim \text{Bin}(n, p + q)$ and $\alpha = p/(p + q)$, we deduce that

$$H(\text{Bin}(n, p) + \text{Bin}(n, q)) \geq H(\text{Bin}(n, p + q)).$$

By results in Poisson approximation, we expect that this inequality will be tightest for $n$ large and $p, q$ small. This result also follows from Theorem 1 of Shepp and Olkin [25], which states that if vector $p$ majorizes $q$ then $H(B_p) \leq H(B_q)$, where $B_p$ is the Bernoulli sum $\sum_{i=1}^n \text{Bin}(p_i)$. Vector $(p+q, p+q, \ldots, p+q, 0, 0, \ldots 0)$ majorizes vector $(p, p, \ldots, p, q, q, \ldots q)$.

3) Given any identically distributed ULC random variables $X$ and $Y$, choosing $\alpha = 1/2$, we deduce that if there exists $X^*$ such that $X = T_{1/2} X^*$ then

$$H(X + Y) \geq H(X^*).$$

Note that such an $X^*$ does not exist for the random variables in Example 4.1 which may be relevant to the fact that these provide a counterexample to (13).

### III. Monotonicity results

The other major contribution of this paper is to establish a monotonicity result in Theorem 3.2 which we regard as a discrete analogue of Artstein et al.’s Theorem 1.3.

In [13], corresponding monotonicity results were proved regarding the entropy and relative entropy of sums of thinned random variables, a situation in which the two types of monotonicity are not equivalent. Write $D(X) = D(X|\Pi_{\lambda X})$ for the relative entropy between a random variable $X$ with mean $\lambda X$ and a Poisson with the same mean. Theorems 2 and 3 respectively of [13] showed that for independent and identically distributed $X_i$:

1) the relative entropy $D(\sum_{i=1}^n T_{1/n} X_i)$ is monotone decreasing in $n$.

2) for ULC $X$, the entropy $H(\sum_{i=1}^n T_{1/n} X_i)$ is monotone increasing in $n$.

In the spirit of Theorem 1.3, we will place these results from [13] in a context where they can be deduced from more general results. Lemma 3.1 and Theorem 3.2. As a consequence we give a proof of monotonicity of entropy which uses distinct ideas from the convex ordering techniques used in [13]. The monotonicity of relative entropy is in fact implied by a stronger result which is implicit in [15].

**Lemma 3.1:** Given positive $\alpha_i$ such that $\sum_{i=1}^{n+1} \alpha_i = 1$, and writing $\alpha^{(l)} = \sum_{i \neq l} \alpha_i = 1 - \alpha_l$, then for any independent
Then and the result follows.

**Proof:** Theorem 5 of [13] shows that for independent random variables $Y_i$,
\[
D \left( \sum_{i=1}^{n+1} T_{\alpha_i} X_i \right) \leq \sum_{i=1}^{n+1} D \left( \sum_{i \neq j} T_{\alpha_i/\alpha(j)} X_i \right),
\]
and Lemma 1 of [13] shows that $D(T_{\alpha} X) \leq \alpha D(X)$. Combining these two results we deduce that
\[
nD \left( \sum_{i=1}^{n+1} T_{\alpha_i} X_i \right) \leq \sum_{i=1}^{n+1} D \left( \sum_{i \neq j} T_{\alpha_i/\alpha(j)} X_i \right) = \sum_{i=1}^{n+1} \alpha(i) D \left( \sum_{i \neq j} T_{\alpha_i/\alpha(j)} X_i \right),
\]
and the result follows.

We have to work harder to show that Theorem 3.2 the corresponding result in terms of entropy, holds as well. The proof of this result is given in Appendix [B].

**Theorem 3.2:** Given positive $\alpha_i$ such that $\sum_{i=1}^{n+1} \alpha_i = 1$, and writing $\alpha(i) = \sum_{i \neq j} \alpha_i$, then for any independent ULC $X_i$,
\[
nH \left( \sum_{i=1}^{n+1} T_{\alpha_i} X_i \right) \geq \sum_{i=1}^{n+1} \alpha(i) H \left( \sum_{i \neq j} T_{\alpha_i/\alpha(j)} X_i \right).
\]
This result gives further support to the ‘general conjecture’ of Gnedenko and Korolev [26, Pages 211–2] that ‘the universal principle of non-decrease of uncertainty manifests itself in probability in the form of limit theorems when the limit is taken with respect to infinitely increasing number of “atomic” random variables involved in a model’. In particular Gnedenko and Korolev [26, Page 215] suggest that it is an important problem to ‘give information proofs of limit theorems ... on convergence of random sums’. We believe that the fact that thinning is an operation defined via random summation means that Theorem 3.2 represents progress in the direction proposed by these authors.

Note that Theorem 3.2 is a strengthened form of Theorem 2.3, indeed Theorem 3.2 can be deduced from it by successive deletion of terms.

Just as Theorem 2.3 led to a proof of the rephrased Entropy Power Inequality Theorem 2.4 Theorem 3.3 leads to a strengthened version of Theorem 2.5 analogous to (6).

**Theorem 3.3:** Assume there exist $Y_i^*$ and $\alpha_i$ such that $Y_i = T_{\alpha_i} Y_i^*$ for each $i$, and there exists some constant $H^*$ so that the entropies satisfy $H\left( \sum_{i \neq j} T_{\alpha_i/\alpha(j)} Y_i^* \right) = H^*$ for all $j$. Then
\[
H \left( \sum_{i=1}^{n+1} Y_i \right) \geq H^*.
\]

**Proof:** Theorem 3.2 implies that
\[
nH \left( \sum_{i=1}^{n+1} Y_i \right) = nH \left( \sum_{i=1}^{n+1} T_{\alpha_i} Y_i^* \right) \geq \sum_{i=1}^{n+1} \alpha(i) H \left( \sum_{i \neq j} T_{\alpha_i/\alpha(j)} Y_i^* \right) = nH^*,
\]
giving a discrete version of the rephrased strengthened Entropy Power Inequality, (6).

**IV. TWO NATURAL DISCRETE EPIs**

Since the Poisson distribution shares with the Gaussian the property of infinite divisibility, as in [13] one natural analogue of Theorem 1.1 comes from replacing $v$ by $V$, with equality holding if and only if $X$ and $Y$ are Poisson. However, as a counterexample provided by an anonymous referee previously showed, such a result turns out not to be true.

**Example 4.1:** For independent discrete random variables $X$ and $Y$, it is not always the case that
\[
V(X + Y) \geq V(X) + V(Y).
\]
A counterexample is that $X \sim Y$, $P_X(0) = 1/6$, $P_X(1) = 2/3$, $P_X(2) = 1/6$. Notice that these $X$ and $Y$ are the sum of Bernoulli random variables, and thus restriction of $X$ and $Y$ to the ULC class does not help.

(2) shows that an equivalent form of the EPI Theorem 1.1 is that for any $0 \leq \alpha \leq 1$,
\[
v\sqrt{\alpha X + \sqrt{1-\alpha} Y} \geq \alpha v(X) + (1 - \alpha) v(Y).
\]
(see [7]). In analogy with this, we might make another conjecture, which again turns out not to hold.

**Example 4.2:** A natural conjecture, which we refer to as the Thinned Entropy Power Inequality, is that for independent discrete ULC random variables $X$ and $Y$, for any $0 \leq \alpha \leq 1$,
\[
V(T_X X + T_{1-\alpha} Y) \geq \alpha V(X) + (1 - \alpha) V(Y),
\]
with equality for $0 < \alpha < 1$ if and only if $X$ and $Y$ are Poisson.

However, taking $X \sim \text{Bern}(1/3) + \Pi_1$ and $Y \sim \text{Bern}(0.3) + \Pi_1$ and $\alpha = 0.999$, (15) is false. That is (taking all logs to base 2) $H(X) = 2.08286 \ldots$, and $V(X) = 1.27189 \ldots$. Clearly $V(Y) = 1000$. Hence the RHS is $\alpha V(X) + (1 - \alpha) V(Y) = 2.27062 \ldots$. Then $T_X X + T_{1-\alpha} Y \sim \text{Bern}(\alpha/3) + \Pi_{1-\alpha(1-\alpha)}$ with $H(T_\alpha X + T_{1-\alpha} Y) = 2.55729 \ldots$, and $V(T_\alpha X + T_{1-\alpha} Y) = 2.25374 \ldots$. In this case (15) fails.

Notice that (15) fails even in the restricted case where $Y$ is Poisson, a case where we might hope that even stronger results might hold, in analogy with work of Costa [27]. The same is true of the conjecture [13] – if that result held for $Y$ Poisson, then using Theorem 2.4 would imply that (15) held in the same case.

As previously described, in the continuous case [7] proves (3) is equivalent to the Entropy Power Inequality. The key fact in this proof is the scaling result, (2). Since Theorem 2.4 is a one-sided version of this fact, we combine it with
Theorems 2.3 to obtain the following partial results, which were proved as Proposition 2 and Corollary 2 respectively of [24], conditionally on the then unproved Theorem 2.4 so now hold without qualification.

**Theorem 4.3:** Consider independent ULC random variables \( X \) and \( Y \).

1. For any \( \beta, \gamma \) such that \( \frac{\beta}{\gamma} \leq \frac{V(Y)}{V(X)} \leq \frac{1}{\beta} \) (note that in this case \( \beta + \gamma < 1 \) unless \( V(X) = V(Y) \)), then
   
   \[
   V(T_\beta X + T_\gamma Y) \geq \beta V(X) + \gamma V(Y).
   \]
2. If \( Y \sim \Pi_\mu \), with \( \mu \leq V(X) \), then for all \( 0 \leq \alpha \leq 1 \),
   
   \[
   V(T_\alpha X + T_{1-\alpha} Y) \geq \alpha V(X) + (1 - \alpha) V(Y).
   \]

We conjecture that there exist some \( \alpha_- = \alpha_-(X, Y) \) and \( \alpha_+ = \alpha_+(X, Y) \) (perhaps defined in terms of the means and entropies of \( X \) and \( Y \)) such that for \( \alpha_- \leq \alpha \leq \alpha_+ \), \[13\] holds. However, as Example 4.2 shows, the unrestricted version of this equation fails.

It is worth noticing that the condition on \( \beta \) and \( \gamma \) in Theorem 4.3 can be restated as \( \beta V(X) + (1 - \gamma) V(Y) \leq \min(V(X), V(Y)) \). Hence by assuming a weaker bound, this theorem proves a stronger one.

**APPENDIX A**

**PROOF OF RTEPI THEOREM 2.4**

We prove the Restricted Thinned Entropy Power Inequality, Theorem 2.4 using a quantity \( L(X) \) that plays a role analogous to the Fisher information in the work of Blachman [3] and Stam [2].

**Definition A.1:** For a random variable \( X \) with probability mass function \( P \), define the quantity

\[
L(X) = \sum_{z=0}^{\infty} (z + 1) P(z + 1) \log \left( \frac{P(z)}{P(z + 1)} \right).
\]

We develop the argument in [12], where we adapted random variables by thinning and then adding an independent Poisson random variable:

**Definition A.2:** For a positive function \( f(\alpha) \), define the combined map \( U_{\alpha, f(\alpha)} \) that thins and then adds an independent Poisson random variable:

\[
U_{\alpha, f(\alpha)} X = T_\alpha X + \Pi_{f(\alpha)}.
\]

For most of this section, we assume that the random variable \( X \) has finite support.

**Proposition A.3:** Consider a continuously differentiable function \( f \) with \( f(1) = 0 \). Assume either (a) \( f(t) = 0 \) for all \( t \) or (b) \( f(t) > 0 \) for \( t < 1 \). Given ULC \( X \) with finite support, writing \( X_t = U_{t, f(t)} X \) and \( P_t(z) = P(X_t = z) \), then for any \( 0 < t < 1 \)

\[
\frac{\partial}{\partial t} H(X_t) = \frac{L(X_t)}{t} - r(t) \sum_{z=0}^{\infty} P_t(z) \log \frac{P_t(z)}{P_t(z + 1)}.
\]

where \( r(t) = f(t)/t - f'(t) \). Equivalently, \( f(t) = tf(1) + t \int_1^t r(\beta)/\beta d\beta \).

**Proof:** From Equation (8) of [12], we know that the mass function of \( X_t \) satisfies

\[
\frac{\partial}{\partial t} P_t(z) = \Delta^* \left( \frac{(z + 1) P(t + 1) - r(t) P_t(z)}{t} \right),
\]

where adjoint operators \( \Delta \) and \( \Delta^* \) are defined by \( \Delta^* g(x) = (g(x - 1) - g(x)) \) and \( \Delta g(x) = (g(x + 1) - g(x)) \). Then we simply differentiate the entropy, using [16] to obtain

\[
\frac{\partial}{\partial t} H(P_t) = -\sum_{z=0}^{\infty} \frac{\partial P_t(z)}{\partial t} \log P_t(z) - \sum_{z=0}^{\infty} \frac{\partial P_t(z)}{\partial t} \log P_t(z).
\]

and the result follows, where this final step uses Fubini’s theorem.

The differentiation of the infinite series at \( t \) can be justified in the case (a) since then the sum is simply a finite one. In case (b) it can be justified by a result (see [28]) concerning \( H(s) = \sum_{z=0}^{\infty} u_s(z) \) with \( a \leq s \leq b \). The derivative \( \frac{\partial H}{\partial s} = \sum_{z=0}^{\infty} \frac{\partial u_s(z)}{\partial s} \) for \( a \leq s \leq b \), assuming that \( \frac{\partial u_s(z)}{\partial s} \) exist, and are uniformly bounded as \( |\frac{\partial u_s(z)}{\partial s}| \leq M(z) \), for all \( a \leq s \leq b \), where \( \sum_{z=0}^{\infty} M(z) < \infty \).

Given a particular \( 0 < t < 1 \), we can choose \( a < s < b \) such that this result holds. In this case, writing \( \lambda = EX \), Equation (9) of [21] shows that \( P(T_s X = 0) \geq (1 - s)^\lambda \), so that

\[
P_s(z) \geq P(T_s X = 0) \Pi_{f(s)}(z) = z (1 - s)^\lambda e^{-f(s)} f(s) z^z / z!.
\]

hence for \( a < s < b \), for all \( z \), we can bound

\[
| - \log P_s(z) | \leq -\lambda \log(1 - s) + f(s) + z \log f(s)) + | \log z |.
\]

Since \( f(s) \) is continuous and bounded away from zero on \((a, b)\), Stirling’s formula means that this can be uniformly bounded by \( C_1 + C_2 z^2 \), where \( C_1 \) and \( C_2 \) depend on \( a \) and \( b \).

Similarly, the triangle inequality means that

\[
| \frac{\partial P_s}{\partial s} | \leq \frac{z P_s(z)}{s} + |r(s)| P_s(z - 1) + \frac{(z + 1) P_s(z + 1)}{s} + |r(s)| P_s(z),
\]

so the fact that \( X \), and hence \( X_s \), is ULC means that \( P_s(z) \leq (P_s(1)/P_s(0))^z / z! P_s(0) \). Hence, since \[17\] means that the ratio \( P_s(1)/P_s(0) \) is uniformly bounded on \((a, b)\), the result follows by continuity (and hence boundedness) of \( r(t) \).

Note that although this result is stated for ULC \( X \) with finite support, it should hold for any random variables such that the differentiation step can be justified.

Writing \( J(t) = E'(|t|) = \sum_{z=0}^{\infty} \Pi_t(z) \log((z + 1)/t) \) (a positive function), we state the following isoperimetric inequality, equivalent to the RTEPI Theorem 2.4 a technique suggested by [18]. This result may be of independent interest.
Theorem A.4: For all ULC random variables $X$ with finite support,  
\[ L(X) \leq V(X)J(V(X)). \]

Lemma A.5: For random variables $X$ with finite support, Theorems 2.4 and A.4 are equivalent.

Proof: Write $g(\alpha)$ for $V(T_\alpha X)$. Assume Theorem 2.4 holds, so that $g(\alpha) \geq g(1)$ or, rearranging, that for $0 < \alpha < 1$  
\[ \frac{g(\alpha) - g(1)}{\alpha - 1} \leq g(1), \]
(the change of direction of the inequality comes since $\alpha < 1$).

Letting $\alpha \to 1$, we see that the RTEPI implies that $g'(1) \leq g(1)$.

The key is to observe that using Proposition A.3, the derivative of $H(T_\alpha X)$ with respect to $\alpha$ is $L(T_\alpha X)/\alpha$. This means that by the chain rule the derivative 
\[ g'(\alpha) = (E^{-1})'(H(T_\alpha X)) \frac{L(T_\alpha X)}{\alpha} \]
\[ = \frac{L(T_\alpha X)}{\alpha J(E^{-1}(H(T_\alpha X)))} \]
\[ = \frac{L(T_\alpha X)}{\alpha J(V(T_\alpha X))}, \]
so taking $\alpha = 1$, the result $g'(1) \leq g(1)$ becomes Theorem A.4.

We deduce the reverse implication by using (18), and applying Theorem A.4 to the random variable $T_\alpha X$, to deduce that  
\[ g'(\alpha) = \frac{L(T_\alpha X)}{\alpha J(V(T_\alpha X))} \leq \frac{V(T_\alpha X)}{\alpha} = \frac{g(\alpha)}{\alpha}. \]
This implies that $g(\alpha)/\alpha$ is decreasing in $\alpha$, which means that $g(\alpha)/\alpha \geq g(1)/1$, which is Theorem 2.4.

We prove Theorem A.4 next, and hence deduce that Theorem 2.4 holds by Lemma A.5. Our approach involves the map $U_{\alpha,f(\alpha)}$ which preserves the entropy (as opposed to preserving the mean as in (12)).

Proof of Theorem A.4: Since $L(X) = \frac{dH}{d\alpha}(T_\alpha X)|_{\alpha=1}$, we know that $L(X)$ need not always be positive (consider for example $X \sim Bern(p)$ with $p > 1/2$). However, note that if $L(X) \leq 0$, then automatically $L(X) \leq 0 \leq V(X)J(V(X))$, as required. Hence, we can restrict our interest to the case where $L(X) > 0$.

Now, $H(T_\alpha X)$ is a positive concave function of $\alpha$ which (since by (12) it is upper bounded by the entropy of a $\Pi_{\alpha X}$ random variable) tends to zero as $\alpha$ tends to zero. Hence, $H(T_\alpha X)$ can only be decreasing in $\alpha$ for $\alpha \in (\alpha^*, 1]$, for some $\alpha^* > 0$. Hence, if $L(X) > 0$, then $L(T_\alpha X) \geq 0$ for all $\alpha \in [0, 1]$ and $H(T_\alpha X)$ is a increasing function of $\alpha$ for all $\alpha \in [0, 1]$. Hence, it is possible to perform an interpolation argument – that is, we can find $f(t) \geq 0$ such that $X_t = \Pi_{f(t)} X$ has constant entropy. We write $\lambda_t$ for the mean of $X_t$.

This means that, since the semigroup interpolates between $X_1 \sim X$ and $X_0 \sim \Pi_{\lambda}$, a Poisson random variable with mean $\lambda$, we can deduce that  
\[ H(X) = H(X_1) = H(X_0) = H(\Pi_{\lambda}) = E(\lambda), \]
or that $\lambda' = V(X)$.

Motivated by Proposition A.3 we consider properties of $r(t) = L(X_t)/\left( \sum_{z=0}^{\infty} P_t(z) \log \left( \frac{P_t(z+1)}{P_t(z)} \right) \right)$. Note that by Chebyshev’s rearrangement lemma (see for example Equation (1.7) of [29])
\[ L(X_t) = \sum_{z=0}^{\infty} P_t(z) \left( \frac{(z+1)P_t(z+1)}{P_t(z)} \right) \log \left( \frac{P_t(z)}{P_t(z+1)} \right) \]
is the expectation of the product of an increasing and decreasing function, so $L(X_t) \leq \lambda_t \sum_{z=0}^{\infty} P_t(z) \log \left( \frac{P_t(z)}{P_t(z+1)} \right)$, or $r(t) \leq \lambda_t/t$. We can write $L(X_t)$ as  
\[ -\lambda_t D(P_t^\pi || P_t) + \sum_{z=0}^{\infty} (z+1)P_t(z+1) \log \left( \frac{z+1}{\lambda_t} \right) \]
\[ \leq -D(P_t || \Pi_{\lambda_t}) + \sum_{z=0}^{\infty} (z+1)P_t(z+1) \log \left( \frac{z+1}{\lambda_t} \right) + \sum_{z=0}^{\infty} (z+1)P_t(z+1) \log (z+1), \]
where $P_t^\pi(x) = P(x+1)(x+1)/\lambda_t$ is the size-biased version of $P_t$, and (19) follows by Equation (0.6) of Wu [39].

Theorem A.4 will follow if we can prove that this expression (20), which we shall refer to as $U(X_t)$, is a decreasing function of $t$. That would mean that  
\[ L(X_t) = L(X) \leq U(X_1) \leq U(X_0) \]
\[ = \lambda' J(\lambda') = V(X)J(V(X)). \]

In fact, since $H(X_t)$ is constant, equivalently, we will prove that $U(X_t) - H(X_t)$ is a decreasing function of $t$.

Case A: $r(t) > 0$ for all $t$. We simply differentiate (20), using Equation (16), and express $\frac{dU(X_t)}{dt}$ as  
\[ \sum_{z=0}^{\infty} P_t(z+1) \left( \frac{(z+2)P_t(z+2)}{tP_t(z+1)} - r(t) \right) (z+1) \log \frac{z+2}{z+1} + r(t) = \frac{\lambda_t}{t}. \]

The term-by-term differentiation can be justified as before, since the assumption that $r(t) = -(f(t)/t)' > 0$ implies that $f(t) > 0$ for $t < 1$, so the assumptions of Proposition A.3 hold. Hence the entropy can indeed be differentiated, and the functions $\log z$ and $\log z$ can be controlled using a similar argument. Since $-(z+1) \log \frac{z+2}{z+1} + 1 \geq 0$, Equation (21) is increased on replacing $r(t)$ by the (larger) value $\lambda_t/t$, so we deduce that $\frac{dU(X_t)}{dt}$ is less than or equal to  
\[ \sum_{z=0}^{\infty} P_t(z+1) \left( \frac{(z+2)P_t(z+2)}{tP_t(z+1)} - \lambda_t \right) (z+1) \log \frac{z+2}{z+1}. \]

Observe that (23) is the covariance of decreasing and increasing functions, and hence is negative by the Chebyshev rearrangement lemma. We have shown that if $L(X_t) > 0$ for
all $t$, so that $r(t) > 0$ for all $t$, then $L(X_t)$ is a decreasing function at $t$.

**Case B:** $r(t) \leq 0$ for some $t$. Recall that we need only consider the case where $L(X) = L(X_1) > 0$. Define $t^* = \sup\{t \geq 0 : r(t) \leq 0\}$. Suppose that $t^* > 0$. For all $t > t^*$, $r(t) > 0$, so that for all $t > t^*$, we know that $L(X_t) \geq L(X)$ by considering $t$ arbitrarily close to $t^*$, continuity of $L(X_t)$ implies that $L(X_t) > 0$ for all $t \in (t^* - \epsilon, t^*)$. This contradicts the assumption that $t^* > 0$, so we deduce that $r(t) > 0$ for all $t > 0$, and the result follows.

**Proof of Theorem 2.4.** By Lemma A.5 we deduce from Theorem A.2 that the RTEPI, Theorem 2.4 holds for ULC $X$ with finite support.

For general ULC $X$, let $X^{(k)}$ be the random variable $X$ truncated at $k$, for $k = 1, 2, \ldots$. Then the mass function of $T_0 X^{(k)}$ tends to that of $T_0 X$ pointwise, for all $0 < \alpha \leq 1$. Moreover, the mean of $T_0 X^{(k)}$ tends to that of $T_0 X$. The argument of Part 2) in Theorem 1 of [13] shows that $H(T_0 X^{(k)}) \to H(T_0 X)$ as $k \to \infty$ (the basic argument is to apply Fatou’s lemma twice). Because $E^{−1}(.)$ is continuous, we have $V(T_0 X^{(k)}) \to V(T_0 X)$ as $k \to \infty$. Thus Theorem 2.4 holds by taking a limit on the finite support result.

**APPENDIX B**

**PROOF OF MONOTONICITY THEOREM 3.2**

In this section, we prove monotonicity of entropy by analysing certain direction derivatives of an ‘energy’ functional $\Lambda$. For $X$ with expectation $\lambda_X$, we write $\Lambda(X) = -\mathbb{E} \log P_{\lambda_X}(X) = \lambda_X + \mathbb{E} \log X! - \lambda_X \log \lambda_X$. In this section, we will establish the following proposition:

**Proposition B.1:** Given positive $\alpha_i$ such that $\sum_{i=1}^{n+1} \alpha_i = 1$, and writing $\alpha^{(i)} = \sum_{i \neq j} \alpha_i$, then for any independent ULC $X_i$,

$$n\Lambda \left( \sum_{i=1}^{n+1} T_{\alpha_i} X_i \right) \geq \sum_{i=1}^{n+1} \alpha^{(i)} \Lambda \left( \sum_{i \neq j} T_{\alpha_i/\alpha^{(i)}} X_i \right).$$  \hspace{1cm} (23)

As in [24], Lemma 3.1 can be subtracted from Proposition B.1 to deduce that Theorem 3.2 holds. We will write $\alpha = (\alpha_1, \ldots, \alpha_{n+1})$ and given independent ULC $X_i$ with means $\lambda_i$ we will define the function $\Phi(\alpha) = \Lambda \left( \sum_{i=1}^{n+1} T_{\alpha_i} X_i \right)$. We write $P_{\alpha}(x) = P(T_{\alpha_1} X_1 = x_1, \ldots, T_{\alpha_{n+1}} X_{n+1} = x_{n+1})$ and $Q_{\alpha}(s) = \sum_{x: \sum_i x_i = s} P_{\alpha}(x)$. In order to establish Proposition B.1 we will need to understand the properties of the Hessian matrix $\Phi''$, which we write as the sum of two matrices $\Phi'' = \Phi''_1 + \Phi''_2$. The first matrix,

$$\Phi''_1(\alpha)_{ij} = \frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \sum_{s=0}^{\infty} \mathbb{Q}_{\alpha}(s) \log s!,$$

can be evaluated using Equation (16) – we omit the details for brevity:

**Lemma B.2:** For any $\alpha$, $i$ and $j$ the derivative

$$\Phi''_1(\alpha)_{ij} = \sum_{s=0}^{\infty} \sum_{x: \sum_i x_i = s} P_{\alpha}(x) x_i(x_i - \delta_{ij}) \alpha_i^{-1} \log \left( \frac{s}{s-1} \right).$$  \hspace{1cm} (24)

The second term of the Hessian, $\Phi''_2$, can be explicitly evaluated by writing $\theta(t) = t - t \log t$ and expressing

$$\Phi''_2(\alpha)_{ij} = \frac{\partial^2}{\partial \alpha_i \partial \alpha_j} \theta \left( \sum_{k=1}^{n+1} \alpha_k \lambda_k \right) = - \frac{\lambda_i \lambda_j}{\sum_{k=1}^{n+1} \alpha_k \lambda_k}.$$  \hspace{1cm} (25)

We now examine the Hessian $\Phi''$ in more detail, using techniques that extend the proof of Theorem 2.3 given in [24], first introducing a sufficient condition.

**Condition 1:** We say that vectors $\mu$ and $\beta$ satisfy the positive splitting condition if there exist positive $u_{ij}$ such that

1) For all $i, j$ the terms

$$u_{ij} + u_{ji} = v_{ij}(\beta, \mu) := \left( \frac{\mu_i}{\beta_i} - \frac{\mu_j}{\beta_j} \right)^2 \beta_i \beta_j \lambda_i \lambda_j.$$

2) For all $j$ the terms $\left( \sum_{i \neq j} u_{ij} \right) / (\beta_j \lambda_j)$ take the same value, $S$ say.

Observe that if Condition 1 holds, then multiplying the terms in Part 2. by $\beta_j \lambda_j$ and summing over $j$ we deduce that

$$S = \sum_{i < j} u_{ij}(\beta, \mu) \frac{\beta_j \lambda_j}{\sum_k \beta_k \lambda_k} = \sum_{i < j} \left( \frac{\mu_i}{\beta_i} - \frac{\mu_j}{\beta_j} \right)^2 \beta_i \beta_j \lambda_i \lambda_j \frac{\beta_k \lambda_k}{\sum_k \beta_k \lambda_k} = \sum_{i \neq j} \frac{\beta_i \beta_j \lambda_i \lambda_j}{\sum_k \beta_k \lambda_k} \left( \sum_{j \neq i} \beta_j \lambda_j \right) - S.$$

so that

$$\left( \sum_i \frac{\mu_i^2 \lambda_i \beta_i}{\beta_i} \right) - S = \frac{1}{\sum_k \beta_k \lambda_k} \left( \sum_i \frac{\mu_i^2 \lambda_i \beta_i}{\beta_i} \lambda_i \right) - S = \frac{1}{\sum_k \beta_k \lambda_k} \left( \sum_i \frac{\mu_i^2 \lambda_i \beta_i}{\beta_i} \lambda_i \right) - S = \frac{1}{\sum_k \beta_k \lambda_k}.$$  \hspace{1cm} (26)

This property allows us to deduce the following result:

**Theorem B.3:** If $\mu$ and $\beta$ satisfy the positive splitting condition, Condition 1 then $\mu^T \Phi''(\beta) \mu \leq 0$.

**Proof:** We use Lemma B.2 to deduce that, writing $e_i$ for the $i$th unit vector, $s = \sum_i x_i$ and $x^{(i)} = x - e_i$, then we
can express the product \( \mu^T \Phi_1'(\beta) \mu \) as

\[
\sum_{x} \mathbb{P}_\beta(x) \sum_{i=1}^{n+1} \left[ \frac{\mu_i \Phi_i(x_i - 1)}{\beta_i} + \sum_{j \neq i} \frac{\mu_i \mu_j x_i x_j}{\beta_i \beta_j} \right] \log \left( \frac{s}{s-1} \right) = \sum_{i=1}^{n+1} \mathbb{P}_\beta(x_i) \log \left( \frac{s}{s-1} \right) \times \left[ \frac{\mu_i^2}{\beta_i^2} \left( \sum_{k} x_k - 1 \right) - \sum_{j \neq i} \frac{u_{ij} x_i}{\beta_j \beta_i \lambda_i} \right]
\]

\leq \sum_{i=1}^{n+1} \beta_i \lambda_i \mathbb{P}_\beta(x^{i-1}) \log \left( \frac{s}{s-1} \right) \times \left[ \frac{\mu_i^2}{\beta_i^2} (s-1) - \sum_{j \neq i} \frac{u_{ij} x_i}{\beta_j \beta_i \lambda_i} \right]
\]

\leq \sum_{s=0}^{\infty} \mathbb{Q}(s) s \log \left( \frac{s+1}{s} \right) \left[ \sum_{i=1}^{n+1} \left( \frac{\mu_i^2 \lambda_i}{\beta_i} \right) \right] - S
\]

\leq \frac{\sum_{k} \mu_k \lambda_k}{\sum_{i=1}^{n+1} \beta_i \lambda_i} = - \mu^T \Phi_2' \mu.
\]

Here Equation (27) follows by comparing coefficients of \( x_i x_j \), using Part 1 of Condition [1] Equation (28) follows as in [24], using Chebyshev’s rearrangement lemma, and the fact that \( \log((x_i + w)/(x_i + w - 1)) \) is increasing in \( x_i \) and \( \log((x_i + w)/(x_i + w - 1)) \) is decreasing in \( x_i \) (coupled with the assumption that \( u_{ij} \geq 0 \)). Equation (29) uses Part 2 of Condition [1] Equation (30) follows using (26) since, as in [24], \( s \log((s+1)/s) \leq 1 \). Finally we use the expression for \( \Phi_2' \) given in Equation (25).  

We can use this result to complete the proof of monotonicity of entropy, Theorem [3.3] by proving Proposition [B.1].

**Proof of Proposition B.1.** For each \( i \), we can define a one-parameter map which interpolates between the values of \( \alpha \). That is, for each \( i \), define

\[
A_i(t) = (1 - t) \alpha^{(i)} + t \epsilon_i,
\]

where \( \alpha^{(i)} = (\alpha_1, \ldots, \alpha_{i-1}, 0, \alpha_{i+1}, \ldots, \alpha_n)/\alpha_i^{(i)} \) is the renormalized ‘leave one out’ vector, and \( \epsilon_i \) is the \( i \)th unit vector. We write \( \mu_i = \epsilon_i - \alpha^{(i)} = \frac{d}{dt} A_i(t) \). Observe that \( A_i(0) = \alpha^{(i)} \) and \( A_i(1) = \alpha \), meaning that by Taylor’s theorem, for some \( t'_i \in [0, 1] \), if the relevant Hessian term is negative,

\[
\Phi(\alpha) - \Phi(\alpha) = \alpha_i \mu_i^T T \Phi(\alpha) + \frac{\alpha_i^2}{2} \mu_i^T \Phi''(A_i(t'_i)) \mu_i
\]

\[
\leq \alpha_i \mu_i^T \Phi'(
\alpha).
\]

If this is true for each \( i \), on multiplying by \( \alpha^{(i)} \) and summing over \( i \) we deduce that \( \sum_{i} \alpha^{(i)} \Phi(\alpha^{(i)}) \leq n \Phi(\alpha) \), and the proof is complete. (This uses the property that \( \sum_{i} \alpha^{(i)} \alpha_i = 0 \), which is a consequence of the fact that \( \sum_{i} \alpha^{(i)} \alpha_i = \sum_{i} \alpha_1(\alpha_1, \ldots, \alpha_{i-1}, 0, \alpha_{i+1}, \ldots, \alpha_n) = (\alpha_1(\alpha_1, \ldots, \alpha_{i+1}, \alpha_{i+1}, \ldots, \alpha_n)) = \sum_{j=1}^{n+1} \alpha_j \epsilon_j \), as required.)  

We complete the proof by checking the negativity of the relevant Hessians by testing positive splitting, Condition [1] and applying Theorem [3.3]. There are considerable simplifications in this case, since the majority of the values of \( \nu_{ij}(A_i(t'_i), \mu_i) \) vanish. That is, if \( i, j \neq l \) then for any \( t \) the \( \nu_{ij}(A_i(t), \mu_i) \) becomes

\[
\left( \frac{\alpha_i / \alpha^{(l)}}{\alpha_i(1 - t)/\alpha^{(l)}} - \frac{\alpha_j / \alpha^{(l)}}{\alpha_j(1 - t)/\alpha^{(l)}} \right)^2 \alpha_i(1 - t) \lambda_i \lambda_j = 0.
\]

In the remaining case, when \( i \neq l \) and \( j = l \), the \( \nu_{ij}(A_i(t), \mu_i) \) is

\[
\left( \frac{\alpha_i / \alpha^{(l)}}{\alpha_i(1 - t)/\alpha^{(l)}} + \frac{1}{t} \right)^2 \alpha_i(1 - t) \lambda_i = \frac{\alpha_i \lambda_i A_l}{(1 - t)^4}(1 - t).
\]

We can exhibit a set of positive solutions to the required equations by writing \( \lambda(t) = \sum \alpha_i(t) \lambda_i, \lambda^{(i)}(t) = \sum_{i \neq l} \alpha_i(t) \lambda_i = \lambda(t) - t \lambda_i \) and \( \lambda(t) = \lambda^{(i)}(t)/((t - 1)^2 \lambda(t)) \). Then define \( u_{ij} \) to be zero unless \( i \neq j \) equals \( l \), in which case for \( i \neq l \),

\[
u_{il} = \frac{S \alpha_i \lambda_i(1 - t)}{(1 - t)^4 \alpha^{(l)}} \quad \text{and} \quad u_{il} = \frac{\lambda_i^2 \alpha_i \lambda_i}{(1 - t)^4 \alpha^{(l)}}.
\]

Finally, for \( u \) as defined in (33), writing \( A_{i,j}(t) \) for the \( j \)th component of \( A_i(t) \), the sums \( \sum_{i \neq j} u_{ij} / (A_{i,j}(t) \lambda_j) \) do indeed equal \( S \) for each \( i \). Specifically, for \( j \neq l \) there is only non-zero term in the sum, giving \( u_{ij} / (A_{i,j}(t) \lambda_j) = S \), since \( A_{i,j}(t) = \alpha_j(1 - t)/\alpha^{(l)} \). For \( j = l \), since \( A_{i,j}(t) = t \), the sum becomes

\[
\sum_{i \neq j} u_{ij} / A_{i,j}(t) \lambda_l = \lambda_l \frac{\sum_{i \neq l} \alpha_i / \alpha_j}{(1 - t)^4 \alpha^{(l)}(t) \lambda_j} = S,
\]

as required. Hence Condition [1] holds in this case, so we can apply Theorem [3.3] to deduce that \( \mu^T \Phi''(A_i(t)) \mu_i \leq 0 \) for all \( l \). This means that (31) holds for each \( l \), and the proof of Proposition B.1 is complete.  

**ACKNOWLEDGEMENTS**

The authors would like to thank Ioannis Kontoyiannis and Peter Harremoës for discussions concerning the discrete Entropy Power Inequality, and in particular for some of the notation used in this paper.

**References**

[1] C. E. Shannon and W. W. Weaver, *A Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press, 1949.

[2] A. J. Stam, “Some inequalities satisfied by the quantities of information of Fisher and Shannon,” *Information and Control*, vol. 2, pp. 101–112, 1959.

[3] N. M. Blachman, “The convolution inequality for entropy powers,” *IEEE Trans. Information Theory*, vol. 11, pp. 267–271, 1965.

[4] A. Rényi, “A characterization of Poisson processes,” *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, vol. 1, pp. 519–527, 1956.

[5] S. Verdú and D. Guo, “A simple proof of the entropy-power inequality,” *IEEE Trans. Inform. Theory*, vol. 52, no. 5, pp. 2165–2166, 2006.
[6] E. Lieb, “Proof of an entropy conjecture of Wehrl,” *Comm. Math. Phys.*, vol. 62, pp. 35–41, 1978.

[7] A. Dembo, T. M. Cover, and J. A. Thomas, “Information theoretic inequalities,” *IEEE Trans. Information Theory*, vol. 37, no. 6, pp. 1501–1518, 1991.

[8] W. Beckner, “Inequalities in Fourier analysis,” *Ann. of Math. (2)*, vol. 102, no. 1, pp. 159–182, 1975.

[9] S. Arzstein, K. M. Ball, F. Barthe, and A. Naor, “Solution of Shannon’s problem on the monotonicity of entropy,” *J. Amec Math. Soc.*, vol. 17, no. 4, pp. 975–982 (electronic), 2004.

[10] A. Tulino and S. Verdú, “Monotonic decrease of the non-Gaussianess of the sum of independent random variables: a simple proof,” *IEEE Trans. Inform. Theory*, vol. 52, no. 9, pp. 4295–4297, 2006.

[11] M. Madiman and A. Barron, “Generalized entropy power inequalities and monotonicity properties of information,” *IEEE Trans. Inform. Theory*, vol. 53, no. 7, pp. 2317–2329, 2007.

[12] O. T. Johnson, “Log-concavity and the maximum entropy property of the Poisson distribution,” *Stoch. Proc. Appl.*, vol. 117, no. 6, pp. 791–802, 2007.

[13] Y. Yu, “Monotonic convergence in an information-theoretic law of small numbers,” *IEEE Trans. Inform. Theory*, vol. 55, no. 12, pp. 5412–5422, 2009.

[14] S. Shamai and A. Wyner, “A binary analog to the entropy-power inequality,” *IEEE Trans. Inform. Theory*, vol. 36, no. 6, pp. 1428–1430, Nov 1990.

[15] H. S. Witsenhausen, “Entropy inequalities for discrete channels,” *IEEE Trans. Information Theory*, vol. 20, no. 5, pp. 610–616, 1974.

[16] A. D. Wyner and J. Ziv, “A theorem on the entropy of certain binary sequences and applications. I,” *IEEE Trans. Information Theory*, vol. 19, no. 6, pp. 769–772, 1973.

[17] A. D. Wyner, “A theorem on the entropy of certain binary sequences and applications. II,” *IEEE Trans. Information Theory*, vol. 19, no. 6, pp. 772–777, 1973.

[18] I. Kontoyiannis, “The Entropy Power Inequality and the mutual information game,” 2004, Unpublished notes, private communication.

[19] P. Harremoës and C. Vignat, “An Entropy Power Inequality for the binomial family,” *JIPAM. J. Inequal. Pure Appl. Math.*, vol. 4, 2003, issue 5, Article 93; see also [http://jipam.vu.edu.au/](http://jipam.vu.edu.au/).

[20] P. Harremoës, O. T. Johnson, and I. Kontoyiannis, “Thinning and the Law of Small Numbers,” in *Proceedings of ISIT 2007*, 24th - 29th June 2007, Nice, 2007, pp. 1491–1495.

[21] ——, “Thinning, entropy and the law of thin numbers,” *IEEE Trans. Inform. Theory (to appear)*, 2010, see [arXiv:0908.0690](https://arxiv.org/abs/0908.0690).

[22] T. M. Liggett, “Ultra logconcave sequences and negative dependence,” *J. Combin. Theory Ser. A*, vol. 79, no. 2, pp. 315–325, 1997.

[23] R. Pemantle, “Towards a theory of negative dependence,” *J. Math. Phys.*, vol. 41, no. 3, pp. 1371–1390, 2000.

[24] Y. Yu and O. T. Johnson, “Concavity of entropy under thinning,” in *Proceedings of ISIT 2009*, 28th June - 3rd July 2009, Seoul, 2009, pp. 144–148.

[25] L. A. Shepp and I. Olkin, “Entropy of the sum of independent Bernoulli random variables and of the multinomial distribution,” in *Contributions to probability*. New York: Academic Press, 1981, pp. 201–206.

[26] B. V. Gnedenko and V. Y. Korolev, *Random Summation: Limit Theorems and Applications*. Boca Raton, Florida: CRC Press, 1996.

[27] M. H. M. Costa, “A new entropy power inequality,” *IEEE Trans. Inform. Theory*, vol. 31, no. 6, pp. 751–760, 1985.

[28] M. B. Porter, “On the differentiation of an infinite series term by term,” *The Annals of Mathematics*, vol. 3, no. 1/4, pp. 19–20, 1901.

[29] J. F. C. Kingman, “Uses of exchangeability,” *Ann. Probability*, vol. 6, no. 2, pp. 183–197, 1978.

[30] L. Wu, “A new modified logarithmic Sobolev inequality for Poisson point processes and several applications,” *Probab. Theory Related Fields*, vol. 118, no. 3, pp. 427–438, 2000.

---

Oliver Johnson received the BA degree in 1995, Part III Mathematics in 1996 and a PhD in 2000, all from the University of Cambridge. He was Clayton Research Fellow of Christ’s College and Max Newman Research Fellow of Cambridge University until 2006, during which time he published the book *Information Theory and the Central Limit Theorem* in 2004. Since 2006 he has been Lecturer in Statistics at Bristol University.

Yaming Yu (M'08) received the B.S. degree in mathematics from Beijing University, Beijing, China, in 1999, and the Ph.D. degree in statistics from Harvard University, Cambridge, MA, in 2005. Since 2005 he has been an Assistant Professor in the Department of Statistics at the University of California, Irvine.