Entropy power inequality for a family of discrete random variables

Naresh Sharma **Smarajit Das** **Siddharth Muthukrishnan**
Tata Institute of Fundamental Research
Mumbai 400 005, India
Email: nsharma@tifr.res.in

December 3, 2010

Abstract

It is known that the Entropy Power Inequality (EPI) always holds if the random variables have density. Not much work has been done to identify discrete distributions for which the inequality holds with the differential entropy replaced by the discrete entropy. Harremoës and Vignat showed that it holds for the pair $(B(m,p), B(n,p))$, $m,n \in \mathbb{N}$, (where $B(n,p)$ is a Binomial distribution with n trials each with success probability p) for $p = 0.5$. In this paper, we considerably expand the set of Binomial distributions for which the inequality holds and, in particular, identify $n_0(p)$ such that for all $m,n \geq n_0(p)$, the EPI holds for $(B(m,p), B(n,p))$. We further show that the EPI holds for the discrete random variables that can be expressed as the sum of n independent identical distributed (IID) discrete random variables for large n.

1 Introduction

The Entropy Power Inequality

$$e^{2h(X+Y)} \geq e^{2h(X)} + e^{2h(Y)} \tag{1}$$

holds for independent random variables X and Y with densities, where $h(\cdot)$ is the differential entropy. It was first stated by Shannon in Ref. [1], and the proof was given by Stam and Blachman [2]. See also Refs. [3, 4, 5, 6, 7, 8, 9].

This inequality is, in general, not true for discrete distributions where the differential entropy is replaced by the discrete entropy. For some special cases (binary random variables with modulo 2 addition), results have been provided by Shamai and Wyner in Ref. [10].

More recently, Harremoës and Vignat have shown that this inequality will hold if X and Y are $B(n, 1/2)$ and $B(m, 1/2)$ respectively for all m, n [11]. Significantly, the convolution operation to get the distribution of $X + Y$ is performed over the usual addition over reals and not over finite fields.

Recently, another approach has been expounded by Harremoës et. al. [12] and by Johnson and Yu [13], wherein they interpret Rényi's thinning operation on a discrete random variable as a discrete analog of the scaling operation for continuous random variables. They provide inequalities for the convolutions of thinned discrete random variables that can be interpreted as the discrete analogs of the ones for the continuous case.

In this paper, we take a re-look at the Harremoës and Vignat [11] result for the Binomial family and extend it for all $p \in (0, 1)$. We show that there always exists an $n_0(p)$ that is a function of p, such that for all $m, n \geq n_0(p)$,

$$\mathrm{e}^{2H[B(m+n,p)]} \geq \mathrm{e}^{2H[B(m,p)]} + \mathrm{e}^{2H[B(n,p)]}, \tag{2}$$

where $H(\cdot)$ is the discrete entropy. The result in Ref. [11] is a special case of our result since we obtain $n_0(0.5) = 7$ and it can be checked numerically by using a sufficient condition that the inequality holds for $1 \leq m, n \leq 6$.

We then extend our results for the family of discrete random variables that can be written as the sum of n IID random variables and show that for large n, EPI holds.

We also look at the semi-asymptotic case for the distributions $B(m, p)$ with m small and $B(n, p)$ with n large. We show that even when n is large, there may exist some m such that EPI may not hold.

Lastly, we show that how the EPI for the discrete case can be interpreted as an improvement to the bounds given by Tulino and Verdú for special cases [7].

2 EPI for the Binomial distribution

Our aim is to have an estimate on the threshold $n_0(p)$ such that

$$\mathrm{e}^{2H[B(m+n,p)]} \geq \mathrm{e}^{2H[B(m,p)]} + \mathrm{e}^{2H[B(n,p)]}, \tag{3}$$

holds for all $m, n \geq n_0(p)$.

It is observed that $n_0(p)$ depends on the skewness of the associated Bernoulli distribution. Skewness of a probability distribution is defined as $\kappa_3/\sqrt{\kappa_2^3}$ where κ_2 and κ_3 are respectively the second and third cumulants of the Bernoulli distribution $B(1, p)$, and it turns out to be $(2p-1)/\sqrt{p(1-p)}$. Let

$$\omega(p) = \frac{(2p-1)^2}{p(1-p)}. \tag{4}$$

We find an expression for $n_0(p)$ that depends on $\omega(p)$. The following theorem, known as Taylor's theorem, will be useful for this purpose (see for example p. 110 in Ref. [14]).

Theorem 1 (Taylor)**.** *Suppose f is a real function on $[a,b]$, $n \in \mathbb{N}$, the $(n-1)$th derivative of f denoted by $f^{(n-1)}$ is continuous on $[a,b]$, and $f^{(n)}(t)$ exists for all $t \in (a,b)$. Let α, β be distinct points of $[a,b]$, then there exists a point y between α and β such that*

$$f(\beta) = f(\alpha) + \sum_{k=1}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(\beta-\alpha)^k + \frac{f^{(n)}(y)}{n!}(\beta-\alpha)^n. \tag{5}$$

For $0 \leq p \leq 1$, let $H(p)$ denote the discrete entropy of a Bernoulli distribution with probability of success p, that is, $H(p) \triangleq -p\log(p) - (1-p)\log(1-p)$. We shall use the natural logarithm throughout this paper. Note that we earlier defined $H(\cdot)$ to be the discrete entropy of a discrete random variable. The definition to be used would be amply clear from the context in what follows. Let

$$\hat{H}(x) \triangleq H(p) - H(x), \quad x \in (0,1). \tag{6}$$

Note that $\hat{H}(x)$ satisfies the assumptions in the Theorem 1 in $x \in (0,1)$. Therefore, we can write

$$\hat{H}(x) = \hat{H}(p) + \sum_{k=1}^{n-1} \frac{\hat{H}^{(k)}(p)}{k!}(x-p)^k + \frac{\hat{H}^{(n)}(x_1)}{n!}(x-p)^n, \tag{7}$$

for some $x_1 \in (x,p)$. Note that $\hat{H}(p) = 0$ and

$$F^{(k)}(x) \triangleq \frac{\hat{H}^{(k)}(x)}{k!} = \begin{cases} \log(x) - \log(1-x), & \text{if } k=1, \\ \frac{1}{k(k-1)}\left[(1-x)^{-(k-1)} + (-1)^k x^{-(k-1)}\right], & \text{if } k \geq 2. \end{cases} \tag{8}$$

For even k, $F^{(k)}(x) \geq 0$ for all $x \in (0,1)$, and hence,

$$\hat{H}(x) \geq \sum_{k=1}^{2l+1} F^{(k)}(p)(x-p)^k \tag{9}$$

for all $x \in (0,1)$ and any non-negative integer l. The following useful identity would be employed at times

$$\log(2) - H(p) = \sum_{\nu=1}^{\infty} \frac{2^{2\nu}}{2\nu(2\nu-1)}\left(p - \frac{1}{2}\right)^{2\nu}. \tag{10}$$

Let $P \triangleq \{p_i\}$ and $Q \triangleq \{q_i\}$ be two probability measures over a finite alphabet $\mathcal{A}$. Let $C^{(p)}(P,Q)$ and $\triangle_\nu^{(p)}(P,Q)$ be measures of discrimination defined as

$$C^{(p)}(P,Q) \triangleq pD(P \parallel M) + qD(Q \parallel M), \tag{11}$$

$$\triangle_\nu^{(p)}(P,Q) \triangleq \sum_{i \in A} \frac{|pp_i - qq_i|^{2\nu}}{(pp_i + qq_i)^{2\nu-1}}, \tag{12}$$

where

$$
\begin{aligned}
M &\triangleq pP+qQ,\\
q &\triangleq 1-p.
\end{aligned}
$$

These quantities are generalized capacitory discrimination and triangular discrimination of order ν respectively that were introduced by Topsøe [15].

The following theorem relates $C^{(p)}(P,Q)$ with $\triangle_\nu^{(p)}(P,Q)$ and would be used later to derive an expression for $n_0(p)$. It generalizes Theorem 1 in Ref. [15].

Theorem 2. *Let P and Q be two distributions over the alphabet $\mathcal{A}$ and $0<p<1$. Then*

$$
C^{(p)}(P,Q)=\sum_{\nu=1}^{\infty}\frac{\triangle_\nu^{(p)}(P,Q)}{2\nu(2\nu-1)}-[\log(2)-H(p)]. \tag{13}
$$

Proof. Let

$$
m_i=pp_i+qq_i,\quad \epsilon_i=|2pp_i-m_i|,\quad k_i=\frac{m_i}{\epsilon_i}. \tag{14}
$$

We have

$$
\frac{1}{k_i}=\frac{|pp_i-qq_i|}{pp_i+qq_i} \tag{15}
$$

and $0\le 1/k_i\le 1$. We have

$$
\begin{aligned}
C^{(p)}(P,Q) &= p\sum_{i\in\mathcal{A}}p_i\log\left(\frac{p_i}{m_i}\right)+q\sum_{i\in\mathcal{A}}q_i\log\left(\frac{q_i}{m_i}\right) &(16)\\
&= p\sum_{i\in\mathcal{A}}p_i\log\left(\frac{pp_i}{m_i}\right)+q\sum_{i\in\mathcal{A}}q_i\log\left(\frac{qq_i}{m_i}\right)+H(p) &(17)\\
&\stackrel{a}{=} \sum_{i\in\mathcal{A}}\frac{m_i+\epsilon_i}{2}\log\left(\frac{m_i+\epsilon_i}{m_i}\right)+\sum_{i\in\mathcal{A}}\frac{m_i-\epsilon_i}{2}\log\left(\frac{m_i-\epsilon_i}{m_i}\right)\\
&\qquad -[\log(2)-H(p)] &(18)\\
&= \sum_{i\in\mathcal{A}}\frac{1}{2}\epsilon_i(1+k_i)\log\left(1+\frac{1}{k_i}\right)+\frac{1}{2}\epsilon_i(k_i-1)\log\left(1-\frac{1}{k_i}\right)\\
&\qquad -[\log(2)-H(p)] &(19)\\
&= \sum_{i\in A}\epsilon_ik_i\left[\log(2)-H\left(\frac{1}{2}+\frac{1}{2k_i}\right)\right]-[\log(2)-H(p)] &(20)\\
&\stackrel{b}{=} \sum_{i\in\mathcal{A}}\epsilon_i\sum_{\nu=1}^{\infty}\frac{1}{2\nu(2\nu-1)k_i^{2\nu-1}}-[\log(2)-H(p)] &(21)\\
&= \sum_{\nu=1}^{\infty}\frac{\triangle_\nu^{(p)}(P,Q)}{2\nu(2\nu-1)}-[\log(2)-H(p)], &(22)
\end{aligned}
$$

where a follows by taking two cases $2pp_i > m_i$ and $2pp_i \leq m_i$, and b follows from (10). $\square$

Let $X^{(n)}$ be a discrete random variable that can be written as

$$X^{(n)} = Z_1 + Z_2 + \cdots + Z_n, \tag{23}$$

where Z_i's are IID random variables. We note that when $X^{(n)}$ is defined as above, we have $X^{(n)} + X^{(m)} = X^{(n+m)}$. Let $Y_n \triangleq \mathrm{e}^{2[H(X^{(n)})]}$. We first use a lemma due to Harremoës and Vignat [11].

Lemma 1 (Harremoës and Vignat [11])**.** *If Y_n/n is increasing, then Y_n is super-additive, i.e., $Y_{m+n} \geq Y_m + Y_n$.*

It is not difficult to show that this is a sufficient condition for the EPI to hold [11]. By the above lemma, the inequality

$$H(X^{(n+1)}) - H(X^{(n)}) \geq \frac{1}{2}\log\left(\frac{n+1}{n}\right) \tag{24}$$

is sufficient for EPI to hold.

Let $X^{(n)} = B(n,p)$. We have

$$P_{X^{(n+1)}}(k+1) = pP_{X^{(n)}}(k) + qP_{X^{(n)}}(k+1). \tag{25}$$

Define a random variable $X^{(n)} + 1$ as

$$P_{X^{(n)}+1}(k+1) = P_{X^{(n)}}(k). \tag{26}$$

for all $k \in \{0, 1, \cdots, n\}$. Hence, using $H(X^{(n)} + 1) = H(X^{(n)})$, we have

$$\begin{aligned} P_{X^{(n+1)}} &= pP_{X^{(n)}+1} + qP_{X^{(n)}}, & (27)\\ H(X^{(n+1)}) &= pH(X^{(n)}+1) + qH(X^{(n)}) + pD(P_{X^{(n)}+1} \parallel P_{X^{(n+1)}}) + qD(P_{X^{(n)}} \parallel P_{X^{(n+1)}}) & (28)\\ &= H(X^{(n)}) + pD(P_{X^{(n)}+1} \parallel P_{X^{(n+1)}}) + qD(P_{X^{(n)}} \parallel P_{X^{(n+1)}}). & (29) \end{aligned}$$

Therefore,

$$H(X^{(n+1)}) = H(X^{(n)}) + C^{(p)}(P_{X^{(n)}+1}, P_{X^{(n)}}). \tag{30}$$

We now derive the lower bound for $C^{(p)}(P_{X^{(n)}+1}, P_{X^{(n)}})$.

Lemma 2. *For $l \in \mathbb{N}$,*

$$\begin{aligned} C^{(p)}(P_{X^{(n)}+1}, P_{X^{(n)}}) &= \sum_{i=0}^{n+1} \hat{H}\left(\frac{i}{n+1}\right) P_{X^{(n+1)}}(i), & (31)\\ C^{(p)}(P_{X^{(n)}}, P_{X^{(n)}+1}) &\geq \sum_{k=1}^{2l+1} F^{(k)}(p)(n+1)^{-k}\mu_k^{(n+1)}, & (32) \end{aligned}$$

where $\mu_k^{(n)}$ *is the* k*-th central moment of* $B(n,p)$*, i.e.,*

$$\mu_k^{(n)} = \sum_{i=0}^{n} (i-np)^k P_{X^{(n)}}(i). \tag{33}$$

Proof. Let $P = X^{(n)}+1$ and $Q = X^{(n)}$. We have

$$\begin{aligned}
\frac{|pp_i - qq_i|}{pp_i + qq_i} &= \frac{p\binom{n}{i-1}p^{i-1}q^{n-i+1} - q\binom{n}{i}p^i q^{n-i}}{p\binom{n}{i-1}p^{i-1}q^{n-i+1} + q\binom{n}{i}p^i q^{n-i}} && (34)\\
&= \frac{\left[\binom{n}{i-1} - \binom{n}{i}\right] p^i q^{n-i+1}}{\left[\binom{n}{i-1} + \binom{n}{i}\right] p^i q^{n-i+1}} && (35)\\
&= \frac{2i-n-1}{n+1}, && (36)
\end{aligned}$$

$$\begin{aligned}
\triangle_\nu^{(p)}(P_{X^{(n)}+1}, P_{X^{(n)}}) &= \sum_{i=0}^{n+1} \left(\frac{2i-n-1}{n+1}\right)^{2\nu} P_{X^{(n+1)}}(i) && (37)\\
&= \left(\frac{2}{n+1}\right)^{2\nu} \sum_{i=0}^{n+1} \left(i - \frac{n+1}{2}\right)^{2\nu} P_{X^{(n+1)}}(i). && (38)
\end{aligned}$$

Using Theorem 2, we have

$$\begin{aligned}
C^{(p)}(P_{X^{(n)}+1}, P_{X^{(n)}}) &= \sum_{\nu=1}^{\infty} \left(\frac{2}{n+1}\right)^{2\nu} \frac{1}{2\nu(2\nu-1)} \sum_{i=0}^{n+1} \left(i - \frac{n+1}{2}\right)^{2\nu} P_{X^{(n+1)}}(i)\\
&\quad -[\log(2) - H(p)] && (39)\\
&= \sum_{i=0}^{n+1} \sum_{\nu=1}^{\infty} \frac{2^{2\nu}}{2\nu(2\nu-1)} \left(\frac{i}{n+1} - \frac{1}{2}\right)^{2\nu} P_{X^{(n+1)}}(i)\\
&\quad -[\log(2) - H(p)] && (40)\\
&\stackrel{a}{=} \sum_{i=0}^{n+1} \left[\log(2) - H\left(\frac{i}{n+1}\right)\right] P_{X^{(n+1)}}(i) + H(p) - \log(2) && (41)\\
&= H(p) - \sum_{i=0}^{n+1} H\left(\frac{i}{n+1}\right) P_{X^{(n+1)}}(i) && (42)\\
&\stackrel{b}{=} \sum_{i=0}^{n+1} \hat{H}\left(\frac{i}{n+1}\right) P_{X^{(n+1)}}(i), && (43)
\end{aligned}$$

where 'a' follows by using (10) and 'b' follows by using (6). To prove the lower bound, we have

$$
\begin{aligned}
C^{(p)}(P_{X^{(n)}}, P_{X^{(n)}+1}) &\overset{a}{\geq} \sum_{i=0}^{n+1}\sum_{k=1}^{2l+1} F^{(k)}(p)\left(\frac{i}{n+1}-p\right)^k P_{X^{(n+1)}}(i) && (44)\\
&= \sum_{k=1}^{2l+1} F^{(k)}(p)\sum_{i=0}^{n+1}\left(\frac{i}{n+1}-p\right)^k P_{X^{(n+1)}}(i) && (45)\\
&= \sum_{k=1}^{2l+1} F^{(k)}(p)(n+1)^{-k}\mu_k^{(n+1)}, && (46)
\end{aligned}
$$

where 'a' holds for all nonnegative integers l using (9). □

The following lemma shows that unlike the continuous case, EPI may not always hold.

Lemma 3. *For $p \neq 0.5$, EPI does not hold for all n.*

Proof. It suffices to show that

$$
\mathrm{e}^{2H[B(2,p)]} \leq \mathrm{e}^{2H[B(1,p)]} + \mathrm{e}^{2H[B(1,p)]} \quad \forall \;\; p, \tag{47}
$$

with equality if and only if $p = 0.5$. In other words, we need to show that

$$
H[B(2,p)] - H[B(1,p)] - \frac{1}{2}\log(2) < 0 \quad \forall \;\; p \neq 0.5. \tag{48}
$$

Using Lemma 2 and (10), we have

$$
\begin{aligned}
H[B(2,p)] - H[B(1,p)] &= H(p) - 2p(1-p)\log(2), && (49)\\
H(p) &\leq \log(2) - 2\left(p-0.5\right)^2. && (50)
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
H[B(2,p)] - H[B(1,p)] - \frac{\log(2)}{2} &\leq \frac{\log(2)}{2} - 2\left(p-0.5\right)^2 - 2p(1-p)\log(2) && (51)\\
&= 2\left(p-0.5\right)^2\left[\log(2)-1\right] && (52)\\
&< 0 \text{ if } p \neq 0.5. && (53)
\end{aligned}
$$

In other words, EPI holds for Binomial distributions $B(n,p)$ for all n only if $p = 0.5$. □

For the case $m = 1$ and $n = 2$, Fig. 1 shows the plot of

$$
f(m,n,p) \triangleq \mathrm{e}^{2H[B(m+n,p)]} - \left\{\mathrm{e}^{2H[B(m,p)]} + \mathrm{e}^{2H[B(n,p)]}\right\} \tag{54}
$$

as a function of p. Note that EPI is satisfied for p close to 0.5, while EPI does not hold if p is close to 0 or 1.

This leads us to the question that for a given p, what should m, n be such that the EPI would hold. The main theorem of this section answers this question.

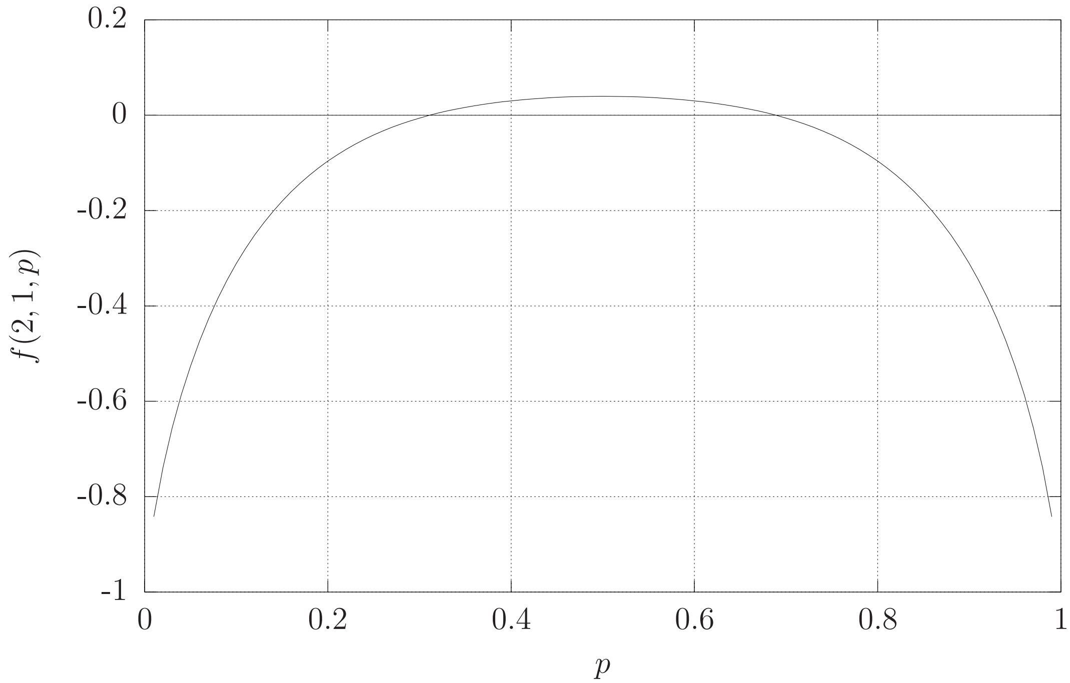

Figure 1: Plot of $\mathrm{e}^{2H[B(3,p)]} - \left\{\mathrm{e}^{2H[B(2,p)]} + \mathrm{e}^{2H[B(1,p)]}\right\}$ versus p.

Theorem 3.

$$H[B(n+1,p)] - H[B(n,p)] \geq \frac{1}{2}\log\left(\frac{n+1}{n}\right) \quad \forall \;\; n \geq n_0(p). \tag{55}$$

Several candidates of $n_0(p)$ are possible such as $n_0(p) = 4.44\ \omega(p) + 7$ and $n_0(p) = \omega(p)^2 + 2.34\ \omega(p) + 7$.

Proof. See Appendix A. □

2.1 Lower bound for the entropy of the Binomial distribution

Unlike the asymptotic expansion of $H[B(n,p)]$ given in Ref. [16], we give non-asymptotic lower bound to it. Let

$$\Gamma_l(j) \triangleq \sum_{k=1}^{2l+1} F^{(k)}(p) j^{-k} \mu_k^{(j)}. \tag{56}$$

We have

$$H(X^{(j)}) - H(X^{(j-1)}) \geq \Gamma_l(j), \tag{57}$$

where $X^{(n)} = B(n,p)$.

Using Faà di Bruno's formula [17], we have

$$\mu_k^{(j)} = \sum \frac{k!}{i_1!(g_1!)^{i_1} \cdots i_s!(g_s!)^{i_s}} \kappa_{g_1}^{i_1} \cdots \kappa_{g_s}^{i_s} j^{i_1+i_2+\cdots+i_s}, \tag{58}$$

where κ_g is the g-th cumulant of the Bernoulli distribution and

$$i_1 g_1 + i_2 g_2 + \cdots + i_s g_s = k. \tag{59}$$

The summation is over all such partitions of k. We have

$$\Gamma_l(j) = \sum_{k=1}^{2l+1} F^{(k)}(p) j^{-k} \mu_k^{(j)} \tag{60}$$

$$= \sum_{w=1}^{2l} j^{-w} c(w), \tag{61}$$

where

$$c(w) = \sum \frac{k!}{i_1!(g_1!)^{i_1} \cdots i_s!(g_s!)^{i_s}} \kappa_{g_1}^{i_1} \cdots \kappa_{g_s}^{i_s} F^{(k)}(p) \tag{62}$$

where the summation is over all such $(i_1, i_2, \cdots, i_s;\ g_1, g_2, \cdots, g_s)$ such that

$$i_1(g_1 - 1) + i_2(g_2 - 1) + \cdots + i_s(g_s - 1) = w. \tag{63}$$

Now

$$H(X^{(n)}) = H(X^{(0)}) + \sum_{j=1}^{n} \left[H(X^{(j)}) - H(X^{(j-1)}) \right] \tag{64}$$

$$\geq \sum_{j=1}^{n} \Gamma_l(j), \qquad \text{since } H(X^{(0)}) = 0 \tag{65}$$

$$= \sum_{j=1}^{n} \sum_{w=1}^{2l} j^{-w} c(w) \tag{66}$$

$$= \sum_{w=1}^{2l} c(w) \sum_{j=1}^{n} j^{-w}. \tag{67}$$

Note that $\sum_{j=1}^{n} j^{-w}$ are the *Generalized Harmonic Numbers* (see for example Ref. [18]).

As an example, we compute the lower bound for $H(X^{(n)})$ for $l = 1$. We have $c(1) = \kappa_2 F^{(2)}(p)$ and $c(2) = 3\kappa_2^2 F^{(4)}(p) + \kappa_3 F^{(3)}(p)$. The first and second cumulants of Bernoulli distribution is given by $\kappa_2 = p(1-p)$ and $\kappa_3 = p(1-p)(1-2p)$. This gives $c(1) = 1/2$ and $c(2) = [1 - p(1-p)]/[12p(1-p)]$ and we get the lower bound as

$$H(X^{(n)}) \geq \frac{1}{2}\left(1 + \frac{1}{2} + \cdots \frac{1}{n}\right) + \frac{1 - p(1-p)}{12p(1-p)}\left(1 + \frac{1}{2^2} + \cdots \frac{1}{n^2}\right). \tag{68}$$

3 EPI for the sum of IID

We showed in the previous section that EPI holds for the pair $(B(n,p), B(m,p))$ for all $m, n \geq n_0(p)$. The question naturally arises whether EPI holds for all such discrete random variables that can be expressed as sum of IID random variables. Let $X^{(n)}$ be a discrete random variable such that

$$X^{(n)} \triangleq X_1 + X_2 + \cdots + X_n, \tag{69}$$

where X_i's are IID random variables and σ^2 is the variance of X_1. We shall use the asymptotic expansion due to Knessl [16].

Lemma 4 (Knessl [16])**.** *For a random variable $X^{(n)}$, as defined above, having finite moments, we have as $n \to \infty$,*

$$g(n) \triangleq H(X^{(n)}) - \frac{1}{2}\log(2\pi e n\sigma^2) \sim -\frac{\kappa_3^2}{12\sigma^6}\frac{1}{n} + \sum_{l=1}^{\infty}\frac{\beta_l}{n^{l+1}}, \tag{70}$$

where κ_j is the jth cumulant of of X_1. If $\kappa_3 = \kappa_4 = \cdots = \kappa_N = 0$ but $\kappa_{N+1} \neq 0$, then

$$g(n) \sim -\frac{\kappa_{N+1}^2}{2(N+1)!\sigma^{2N+2}}n^{1-N} + \sum_{l=N-1}^{\infty}\frac{\beta_l}{n^{l+1}}. \tag{71}$$

Note that the leading term in the asymptotic expansion is always negative. We also note using Lemma 4 that as $n \to \infty$,

$$H(X^{(n)}) < \frac{1}{2}\log(2\pi e n\sigma^2). \tag{72}$$

To see this, we invoke the definition of the asymptotic series to get

$$g(n) = -\frac{\kappa_{N+1}^2}{2(N+1)!\sigma^{2N+2}}n^{1-N} + \frac{\beta_{N-1}}{n^N} + o\left(\frac{1}{n^N}\right). \tag{73}$$

From the definition of the "little-oh" notation, we know that given any $\epsilon > 0$, there exists a $L(\epsilon) > 0$ such that for all $n > L(\epsilon)$,

$$g(n) = -\frac{\kappa_{N+1}^2}{2(N+1)!\sigma^{2N+2}}n^{1-N} + \frac{\beta_{N-1}+\epsilon}{n^N}. \tag{74}$$

Choosing n large enough, we get the desired result.

3.1 Asymptotic case

We first consider the case of the pair $(X^{(n)}, X^{(m)})$ when both m, n are large and have the following result.

Theorem 4. *There exists a $n_0 \in \mathbb{N}$ such that*

$$\mathrm{e}^{2H(X^{(m)}+X^{(n)})} \geq \mathrm{e}^{2H(X^{(m)})} + \mathrm{e}^{2H(X^{(n)})} \tag{75}$$

for all $m, n \geq n_0$.

Proof. We shall prove the sufficient condition for the EPI to hold (as per Lemma 1) and show that

$$H(X^{(n+1)}) - H(X^{(n)}) \geq \frac{1}{2}\log\left(\frac{n+1}{n}\right) \tag{76}$$

for $n \geq n_0$ for some $n_0 \in \mathbb{N}$.

Let us take the first three terms in the above asymptotic series as

$$g(n) \sim -\frac{C_1}{n^{k_1}} + \frac{C_2}{n^{k_2}} + \frac{C_3}{n^{k_3}} \tag{77}$$

where $0 < k_1 < k_2 < k_3$ and C_1 is some non-zero positive constant, and hence,

$$g(n) + \frac{C_1}{n^{k_1}} - \frac{C_2}{n^{k_2}} - \frac{C_3}{n^{k_3}} = o\left(\frac{1}{n^{k_3}}\right). \tag{78}$$

and given any $\epsilon > 0$, there exists a $L(\epsilon) > 0$ such that for all $n > L(\epsilon)$,

$$\left|g(n) + \frac{C_1}{n^{k_1}} - \frac{C_2}{n^{k_2}} - \frac{C_3}{n^{k_3}}\right| \leq \epsilon\left|\frac{1}{n^{k_3}}\right|. \tag{79}$$

Therefore, we have the following inequality

$$\frac{-C_1}{n^{k_1}} + \frac{C_2}{n^{k_2}} + \frac{C_3 - \epsilon}{n^{k_3}} \leq g(n) \leq \frac{-C_1}{n^{k_1}} + \frac{C_2}{n^{k_2}} + \frac{C_3 + \epsilon}{n^{k_3}}. \tag{80}$$

From inequality (80), we can say by using the lower and upper bounds respectively for $g(n+1)$ and $g(n)$ that,

$$\begin{aligned} g(n+1) - g(n) \geq C_1\left[\frac{1}{n^{k_1}} - \frac{1}{(n+1)^{k_1}}\right] + C_2\left[\frac{1}{(n+1)^{k_2}} - \frac{1}{n^{k_2}}\right] \\ + \left[\frac{C_3 - \epsilon}{(n+1)^{k_3}} - \frac{C_3 + \epsilon}{n^{k_3}}\right]. \end{aligned} \tag{81}$$

From the above expression, we can clearly see that the first term is strictly positive and is $O\left(1/n^{k_1+1}\right)$. The second and third terms (their signs are irrelevant) are of the order $O(1/n^{k_2+1})$ and $O(1/n^{k_3})$ respectively. It is clear that there exists some positive integer n_0 such that for all $n \geq n_0$, first (positive) term will dominate and the other two terms will be negligible compared to the first and hence $g(n+1) - g(n) \geq 0$. □

3.2 Semi-asymptotic case

We now consider the pair $(X^{(n)}, X^{(m)})$ where $n \to \infty$ and m is fixed. We already know from the previous result that the EPI holds when both m and n are large.

We start by writing an asymptotic expansion of

$$f(m,n) \triangleq \mathrm{e}^{2H(X^{(m+n)})} - \mathrm{e}^{2H(X^{(n)})} - \mathrm{e}^{2H(X^{(m)})}, \tag{82}$$

by using Knessl's result in Lemma 4 in which an asymptotic expansion for the entropy of $X^{(n)}$ is derived as

$$H(X^{(n)}) \sim \frac{1}{2}\log(2\pi \mathrm{e} n\sigma^2) + \sum_{l=0}^{\infty} \frac{\beta_l}{n^{l+1}}. \tag{83}$$

Let

$$g(n) \triangleq \sum_{l=0}^{\infty} \frac{\beta_l}{n^{l+1}}. \tag{84}$$

We have

$$\begin{aligned} \mathrm{e}^{2H(X^{(n)})} &= (2\pi \mathrm{e} n\sigma^2)\mathrm{e}^{g(n)}, &(85)\\ \mathrm{e}^{2H(X^{(m+n)})} &= [2\pi \mathrm{e}(m+n)\sigma^2]\mathrm{e}^{g(m+n)}, &(86)\end{aligned}$$

and we can rewrite

$$f(m,n) = \left[2\pi \mathrm{e} m\sigma^2 - \mathrm{e}^{2H(X^{(m)})}\right] + 2\pi \mathrm{e}(m+n)\sigma^2\left[\mathrm{e}^{g(m+n)} - 1\right] - 2\pi \mathrm{e} n\sigma^2\left[\mathrm{e}^{g(n)} - 1\right]. \tag{87}$$

The first term in the above equation is a constant since it depends only on m and the second term can be written as

$$2\pi \mathrm{e}(m+n)\sigma^2\left\{\mathrm{e}^{\left[\frac{\beta_0}{m+n} + \frac{\beta_1}{(m+n)^2} + o\left(\frac{1}{(m+n)^2}\right)\right]} - 1\right\}, \tag{88}$$

which can be expanded into

$$2\pi \mathrm{e}(m+n)\sigma^2\left\{\frac{\beta_0}{m+n} + \frac{2\beta_1 + \beta_0^2}{2(m+n)^2} + o\left[\frac{1}{(m+n)^2}\right]\right\}. \tag{89}$$

Similarly, the third term can be written as

$$-2\pi \mathrm{e} n\sigma^2\left[\frac{\beta_0}{n} + \frac{2\beta_1 + \beta_0^2}{2n^2} + o\left(\frac{1}{n^2}\right)\right]. \tag{90}$$

Using the above two expressions, $o\left[1/(m+n)^2\right] = o\left(1/n^2\right)$, and (87), we get

$$f(m,n) = [2\pi \mathrm{e} m\sigma^2 - \mathrm{e}^{2H(X^{(m)})}] + 2\pi \mathrm{e}\sigma^2\left[\frac{2\beta_1 + \beta_0^2}{2}\left(\frac{1}{m+n} - \frac{1}{n}\right) + o\left(\frac{1}{n}\right)\right], \tag{91}$$

where the terms

$$2\pi e\sigma^2\left[o\left(\frac{1}{n}\right)\right] \text{ and } 2\pi e\sigma^2\left(\frac{2\beta_1+\beta_0^2}{2}\right)\left(\frac{1}{m+n}-\frac{1}{n}\right)$$

can be made arbitrarily small as $n \to \infty$. Therefore, for large enough n, we can see that the first term dominates over other terms and moreover, $f(n,m) \geq 0$ if $2\pi e m\sigma^2 - e^{2H(X^{(m)})} > 0$. Therefore,

$$e^{2H(X^{(m)}+X^{(n)})} \geq e^{2H(X^{(m)})} + e^{2H(X^{(n)})} \tag{92}$$

if $n \to \infty$ and

$$H(X^{(m)}) < \frac{1}{2}\log[2\pi e m\sigma^2]. \tag{93}$$

It follows from (72) that the above inequality holds for sufficiently large m. For the Binomial distribution $B(m,p)$, EPI will hold for all such p that satisfy

$$H[B(m,p)] < \frac{1}{2}\log[2\pi e m p(1-p)]. \tag{94}$$

The above relation is not true for all p and m.

4 Discussion and Conclusions

We show that how our results can be used to improve a bound by Tulino and Verdú [7] under special cases.

4.1 Improvement on a bound by Tulino and Verdú

Let X_i, $i = 1, 2, ..., n$ be discrete IID random variables and Z_i, $i = 1, 2, ..., n$, be IID random variables as well with $Z_1 \sim \mathcal{N}(0, \sigma^2)$. Let

$$S^{(n)} = \sum_{i=1}^{n}(X_i + Z_i) \tag{95}$$

$$X^{(n)} = \sum_{i=1}^{n} X_i. \tag{96}$$

Let

$$D(Y) = 0.5\log(2\pi e\sigma_Y^2) - h(Y), \tag{97}$$

where Y is a random variable with density and variance σ_Y^2. Tulino and Verdú [7] interpreted $D(Y)$ as the non-Gaussianess of the random variable X and showed that the non-Gaussianess increases by having more random variables, i.e.,

$$D(S^{(n)}) \leq D(S^{(n-1)}). \tag{98}$$

Expanding it using (97), we get

$$h(S^{(n)}) - h(S^{(n-1)}) \geq \frac{1}{2} \log \left(\frac{n}{n-1} \right). \tag{99}$$

We show that for sufficiently large n, this bound can be made tighter for small σ, i.e.,

$$\lim_{\sigma \to 0} \left[h(S^{(n)}) - h(S^{(n-1)}) \right] \geq \log \left(\frac{n}{n-1} \right). \tag{100}$$

Let

$$I_n \triangleq I(X^{(n)}; S^{(n)}) = h(S^{(n)}) - 0.5 \log(2\pi e n \sigma^2). \tag{101}$$

Note that using Lemma 6 in Ref. [19]

$$H(X^{(n)}) = \lim_{\sigma \to 0} (I_n). \tag{102}$$

We know that for sufficiently large n, the EPI holds and

$$H(X^{(n)}) - H(X^{(n-1)}) \geq \frac{1}{2} \log \left(\frac{n}{n-1} \right). \tag{103}$$

Therefore,

$$\lim_{\sigma \to 0} (I_n - I_{n-1}) \geq \frac{1}{2} \log \left(\frac{n}{n-1} \right) \tag{104}$$

for sufficiently large n. The above limit is due to [19]. On the other hand,

$$I_n - I_{n-1} = h(S^{(n)}) - h(S^{(n-1)}) - \frac{1}{2} \log \left(\frac{n}{n-1} \right). \tag{105}$$

Hence,

$$\lim_{\sigma \to 0} [h(S^{(n)}) - h(S^{(n-1)})] \geq \log \left(\frac{n}{n-1} \right) \tag{106}$$

for sufficiently large n. Comparing (106) with (99), we note that our bound is tighter by a factor of 2.

4.2 Conclusions

In conclusion, we have expanded the set of pairs of Binomial distributions for which the EPI holds. We identified a threshold that is a function of the probability of success beyond which the EPI holds. We further show that EPI would hold for discrete random variables that can be written as sum of IID random variables.

It would be interesting to know if $C^{(p)}(P_{X^{(n)}+1}, P_{X^{(n)}})$ for $X^{(n)} = B(n,p)$ is a concave function in p. It would also be of interest to know that for a given $p \in (0, 0.5)$ if $H[B(n+1,p)] - H[B(n,p)] - 0.5 \log(1 + 1/n)$ would have a single zero crossing as a function of n when n increases from 1 to ∞.

References

[1] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–655, July and Oct. 1948.

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

[3] M. H. M. Costa, “A new entropy power inequality,” *IEEE Trans. Inf. Theory*, vol. 31, pp. 751–760, Nov. 1985.

[4] A. Dembo, T. M. Cover, and J. A. Thomas, “Information theoretic inequalities,” *IEEE Trans. Inf. Theory*, vol. 37, pp. 1501–1518, Nov. 1991.

[5] O. T. Johnson, “Log-concavity and the maximum entropy property of the Poisson distribution,” *Stoch. Proc. Appl.*, vol. 117, pp. 791–802, June 2007.

[6] S. Artstein, K. M. Ball, F. Barthe, and A. Naor, “Solution of Shannon’s problem on the monotonicity of entropy,” *J. Amer. Math. Soc.*, vol. 17, pp. 975–982, May 2004.

[7] A. M. Tulino and S. Verdú, “Monotonic decrease of the non-Gaussianness of the sum of independent random variables: A simple proof,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 4295 –4297, Sep. 2006.

[8] S. Verdú and D. Guo, “A simple proof of the entropy-power inequality,” *IEEE Trans. Inf. Theory*, vol. 52, pp. 2165–2166, May 2006.

[9] M. Madiman and A. Barron, “Generalized entropy power inequalities and monotonicity properties of information,” *IEEE Trans. Inf. Theory*, vol. 53, pp. 2317–2329, July 2007.

[10] S. Shamai (Shitz) and A. D. Wyner, “A binary analog to the entropy-power inequality,” *IEEE Trans. Inf. Theory*, vol. 36, pp. 1428–1430, Nov. 1990.

[11] P. Harremoës and C. Vignat, “An entropy power inequality for the binomial family,” *J. Inequal. Pure Appl. Math.*, vol. 4, no. 5, Oct. 2003.

[12] P. Harremoës, O. Johnshon, and I. Kontoyiannis, “Thinning, entropy, and the law of thin numbers,” *IEEE Trans. Inf. Theory*, vol. 56, pp. 4228–4244, Sep. 2010.

[13] Y. Yu, “Monotonic convergence in an information-theoretic law of small numbers,” *IEEE Trans. Inf. Theory*, vol. 55, pp. 5412–5422, Dec. 2009.

[14] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, 1976.

[15] F. Topsøe, "Some inequalities for information divergence and related measures of discrimination," *IEEE Trans. Inf. Theory*, vol. 46, pp. 1602–1609, July 2000.

[16] C. Knessl, "Integral representations and asymptotic expansions for Shannon and Renyi entropies," *Appl. Math. Lett.*, vol. 11, pp. 69–74, Mar. 1998.

[17] E. Lukacs, "Applications of Faà di Bruno's formula in mathematical statistics," *Am. Math. Monthly*, vol. 62, pp. 340–348, May 1955.

[18] J. H. Conway and R. K. Guy, *The Book of Numbers*. New York: Springer-Verlag, 1996.

[19] D. Guo, S. Shamai (Shitz), and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, pp. 1261–1282, Apr. 2005.

A Proof of Theorem 3

We prove that

$$H[B(n+1,p)] - H[B(n,p)] \geq \frac{1}{2}\log\left(\frac{n+1}{n}\right) \quad \forall\ n \geq n_0(p). \tag{107}$$

Using (46), we have

$$\begin{aligned} H[B(n+1,p)] - H[B(n,p)] &= C^{(p)}(P_{X^{(n)}}, P_{X^{(n)}+1}) &(108)\\ &\geq \sum_{k=1}^{2l+1} F^{(k)}(p)(n+1)^{-k}\mu_k^{(n+1)}. &(109)\end{aligned}$$

Let

$$r \triangleq p - 1/2. \tag{110}$$

We have the first seven central moments of $B(n,p)$ as

$$\begin{aligned}
\mu_2^{(n)} &= \frac{1}{4}n(1-4r^2), &(111)\\
\mu_3^{(n)} &= -\frac{1}{2}nr(1-4r^2), &(112)\\
\mu_4^{(n)} &= \frac{1}{16}n(1-4r^2)[-2+24r^2+3n(1-4r^2)], &(113)\\
\mu_5^{(n)} &= -\frac{1}{4}nr(1-4r^2)[-4+24r^2+5n(1-4r^2)], &(114)\\
\mu_6^{(n)} &= \frac{1}{32}n(1-4r^2)[15n^2(1-4r^2)^2+16(1-30r^2+120r^4)\\
&\quad -10n(3-64r^2+208r^4)], &(115)\\
\mu_7^{(n)} &= -\frac{1}{32}nr(1-4r^2)[105n^2(1-4r^2)^2-14n(17-200r^2+528r^4), &(116)\\
&\quad +8(17-240r^2+720r^4)]. &(117)
\end{aligned}$$

Let $t = \omega(p) = 16r^2/(1-4r^2)$ and hence, $r^2 = t/[4(t+4)]$. Note that $r^2 \in [0, 1/4)$ and $t \in [0, \infty)$. The above seven central moments contain only even powers of r and hence, can be written as a function of t.

We upper bound the right hand side of (107) as

$$\log\left(\frac{n+1}{n}\right) \leq \frac{1}{n} - \frac{1}{2n^2} + \frac{1}{3n^3}. \tag{118}$$

Define

$$f(n,t) \triangleq \sum_{k=1}^{7} F^{(k)}\left[\sqrt{\frac{t}{4(t+4)}} + \frac{1}{2}\right](n+1)^{-k}\mu_k^{(n+1)} - \frac{1}{n} + \frac{1}{2n^2} - \frac{1}{3n^3}. \tag{119}$$

Proving (107) is equivalent to showing that $f(n,t) \geq 0 \; \forall \; n > n_0(p)$. Simplifying

$$\begin{aligned} f(n,t) \;\; = \;\; & \frac{1}{420(n+1)^6 n^3}\Big[35n^7 t + (315t + 35t^2 + 70)n^6 + \\ & (-2989t - 721t^3 - 3339t^2 - 315)n^5 + \\ & (721t - 546t^4 - 826 + 371t^2 - 1568t^3)n^4 + \\ & (-135t^2 - 157t^3 - 10t^5 - 826 - 90t - 66t^4)n^3 - \\ & 630n^2 - 315n - 70\Big]. \end{aligned} \tag{120}$$

Define

$$g(n,t) \triangleq 420(n+1)^6 n^3 f(n,t). \tag{121}$$

A simple but elaborate calculation yields $g(4.44t + 7 + m, t) \approx 35tm^7 + (1122.8t^2 + 2030t + 70)m^6 + (14700.90t^3 + 52210.20t^2 + 48120.80t + 2625)m^5 + (1.01t^4 + 5.32t^3 + 9.57t^2 + 6.06t + 0.40)10^5 m^4 + (3.85t^5 + 26.94t^4 + 72.32t^3 + 88.61t^2 + 43.61t + 3.02)10^5 m^3 + (7.76t^6 + 68.23t^5 + 247.042t^4 + 456.97t^3 + 433.17t^2 + 176.77t + 11.80)10^5 m^2 + (6.47t^7 + 70.91t^6 + 338.88t^5 + 880.98t^4 + 1297.85t^3 + 1030.51t^2 + 361.59t + 20.14)10^5 m + (0.15t^8 + 56.29t^7 + 709.80t^6 + 3485.03t^5 + 8728.40t^4 + 11955.74t^3 + 8613.06t^2 + 2628.77t + 64.15)10^4$.

Note that all the coefficients are positive and hence, $f(4.44t + 7 + m, t) \geq 0$ for all $m \geq 0$ or $f(n,t) \geq 0$ for all $n \geq 4.44t + 7$. A more careful choice would yield all coefficients to be positive for $n \geq 4.438t + 7$. Yet another choice that would yield all coefficients as positive would be $n \geq t^2 + 2.34t + 7$. Note that this choice would be better for $0 < t < 2.1$ and, in particular, for $t = 1$, the first choice yields (after constraining n to be a natural number) $n \geq 12$ while the second one yields $n \geq 11$.

Further refinements are also possible. For example, the expansion of $f(7 + m, t/[4(1+t)])$ yields positive coefficients again. Such a choice constrains $\omega(p) \in (0, 1/4)$ and we get $n_0(p) = 7$.