Journal of Inequalities in Pure and Applied Mathematics

AN ENTROPY POWER INEQUALITY FOR THE BINOMIAL FAMILY

PETER HARREMOËS AND CHRISTOPHE VIGNAT

Department of Mathematics,
University of Copenhagen,
Universitetsparken 5,
2100 Copenhagen, Denmark.
EMail: moes@math.ku.dk

University of Copenhagen and Université de Marne la Vallée,
77454 Marne la Vallée
Cedex 2, France.
EMail: vignat@univ-mlv.fr

ISSN (electronic): 1443-5756
043-03

volume 4, issue 5, article 93, 2003.

Received 03 April, 2003; accepted 21 October, 2003.

Communicated by: S.S. Dragomir

Abstract

In this paper, we prove that the classical Entropy Power Inequality, as derived in the continuous case, can be extended to the discrete family of binomial random variables with parameter $1/2$.

2000 Mathematics Subject Classification: 94A17
Key words: Entropy Power Inequality, Discrete random variable

The first author is supported by a post-doc fellowship from the Villum Kann Rasmussen Foundation and INTAS (project 00-738) and Danish Natural Science Council.
This work was done during a visit of the second author at Dept. of Math., University of Copenhagen in March 2003.

Contents

1. Introduction

The continuous Entropy Power Inequality

$$e^{2h(X)} + e^{2h(Y)} \leq e^{2h(X+Y)} \tag{1.1}$$

was first stated by Shannon [1] and later proved by Stam [2] and Blachman [3]. Later, several related inequalities for continuous variables were proved in [4], [5] and [6]. There have been several attempts to provide discrete versions of the Entropy Power Inequality: in the case of Bernoulli sources with addition modulo 2, results have been obtained in a series of papers [7], [8], [9] and [11].

In general, inequality (1.1) does not hold when X and Y are discrete random variables and the differential entropy is replaced by the discrete entropy: a simple counterexample is provided when X and Y are deterministic.

In what follows, $X_n \sim B\left(n, \frac{1}{2}\right)$ denotes a binomial random variable with parameters n and $\frac{1}{2}$, and we prove our main theorem:

Theorem 1.1. *The sequence X_n satisfies the following Entropy Power Inequality*

$$\forall m, n \geq 1, \quad e^{2H(X_n)} + e^{2H(X_m)} \leq e^{2H(X_n+X_m)}.$$

With this aim in mind, we use a characterization of the superadditivity of a function, together with an entropic inequality.

2. Superadditivity

Definition 2.1. *A function* $n \curvearrowright Y_n$ *is superadditive if*

$$\forall m,n \quad Y_{m+n} \geq Y_m + Y_n.$$

A sufficient condition for superadditivity is given by the following result.

Proposition 2.1. *If* $\frac{Y_n}{n}$ *is increasing, then* Y_n *is superadditive.*

Proof. Take m and n and suppose $m \geq n$. Then by assumption

$$\frac{Y_{m+n}}{m+n} \geq \frac{Y_m}{m}$$

or

$$Y_{m+n} \geq Y_m + \frac{n}{m} Y_m.$$

However, by the hypothesis $m \geq n$

$$\frac{Y_m}{m} \geq \frac{Y_n}{n}$$

so that

$$Y_{m+n} \geq Y_m + Y_n.$$

□

In order to prove that the function

$$Y_n = e^{2H(X_n)} \tag{2.1}$$

is superadditive, it suffices then to show that function $n \curvearrowright \frac{Y_n}{n}$ is increasing.

3. An Information Theoretic Inequality

Denote as $B \sim Ber\,(1/2)$ a Bernoulli random variable so that

$$X_{n+1} = X_n + B \tag{3.1}$$

and

$$P_{X_{n+1}} = P_{X_n} * P_B = \frac{1}{2}\left(P_{X_n} + P_{X_n+1}\right), \tag{3.2}$$

where $P_{X_n} = \{p_k^n\}$ denotes the probability law of X_n with

$$p_k^n = 2^{-n}\binom{n}{k}. \tag{3.3}$$

A direct application of an equality by Topsøe [12] yields

$$\begin{aligned} H\left(P_{X_{n+1}}\right) &= \frac{1}{2}H\left(P_{X_n+1}\right) + \frac{1}{2}H\left(P_{X_n}\right) \\ &\quad + \frac{1}{2}D\left(P_{X_n+1}||P_{X_{n+1}}\right) + \frac{1}{2}D\left(P_{X_n}||P_{X_{n+1}}\right). \end{aligned} \tag{3.4}$$

Introduce the Jensen-Shannon divergence

$$\mathcal{JSD}\left(P,Q\right) = \frac{1}{2}D\left(P\left\|\frac{P+Q}{2}\right.\right) + \frac{1}{2}D\left(Q\left\|\frac{P+Q}{2}\right.\right) \tag{3.5}$$

and remark that

$$H\left(P_{X_n}\right) = H\left(P_{X_n+1}\right), \tag{3.6}$$

since each distribution is a shifted version of the other. We conclude thus that

$$H\left(P_{X_{n+1}}\right) = H\left(P_{X_n}\right) + \mathcal{JSD}\left(P_{X_n+1}, P_{X_n}\right), \tag{3.7}$$

showing that the entropy of a binomial law is an increasing function of n. Now we need the stronger result that $\frac{Y_n}{n}$ is an increasing sequence, or equivalently that

$$\log \frac{Y_{n+1}}{n+1} \geq \log \frac{Y_n}{n} \tag{3.8}$$

or

$$\mathcal{JSD}\left(P_{X_n+1}, P_{X_n}\right) \geq \frac{1}{2} \log \frac{n+1}{n}. \tag{3.9}$$

We use the following expansion of the Jensen-Shannon divergence, due to B.Y. Ryabko and reported in [13].

Lemma 3.1. *The Jensen-Shannon divergence can be expanded as follows*

$$\mathcal{JSD}(P, Q) = \frac{1}{2} \sum_{\nu=1}^{\infty} \frac{1}{2\nu(2\nu-1)} \Delta_\nu(P, Q)$$

with

$$\Delta_\nu(P, Q) = \sum_{i=1}^{n} \frac{|p_i - q_i|^{2\nu}}{(p_i + q_i)^{2\nu-1}}.$$

This lemma, applied in the particular case where $P = P_{X_n}$ and $Q = P_{X_n+1}$ yields the following result.

Lemma 3.2. *The Jensen-Shannon divergence between P_{X_n+1} and P_{X_n} can be expressed as*

$$\mathcal{JSD}\left(P_{X_n+1}, P_{X_n}\right) = \sum_{\nu=1}^{\infty} \frac{1}{\nu\left(2\nu-1\right)} \cdot \frac{2^{2\nu-1}}{\left(n+1\right)^{2\nu}} m_{2\nu}\left(B\left(n+1, \frac{1}{2}\right)\right),$$

where $m_{2\nu}\left(B\left(n+1,\frac{1}{2}\right)\right)$ denotes the order 2ν central moment of a binomial random variable $B\left(n+1,\frac{1}{2}\right)$.

Proof. Denote $P = p_i, Q = p_i^+$ and $\bar{p}_i = (p_i + p_i^+)/2$. For the term $\Delta_\nu\left(P_{X_n+1}, P_{X_n}\right)$ we have

$$\Delta_\nu\left(P_{X_n+1}, P_{X_n}\right) = \sum_{i=1}^{n} \frac{\left|p_i^+ - p_i\right|^{2\nu}}{\left(p_i^+ + p_i\right)^{2\nu-1}} = 2\sum_{i=1}^{n}\left(\frac{p_i^+ - p_i}{p_i^+ + p_i}\right)^{2\nu} \bar{p}_i$$

and

$$\frac{p_i^+ - p_i}{p_i^+ + p_i} = \frac{2^{-n}\binom{n}{i-1} - 2^{-n}\binom{n}{i}}{2^{-n}\binom{n}{i-1} + 2^{-n}\binom{n}{i}} = \frac{2i-n-1}{n+1}$$

so that

$$\begin{aligned}
\Delta_\nu\left(P_{X_n+1}, P_{X_n}\right) &= 2\sum_{i=1}^{n}\left(\frac{2i-n-1}{n+1}\right)^{2\nu}\bar{p}_i \\
&= 2\left(\frac{2}{n+1}\right)^{2\nu}\sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)^{2\nu}\bar{p}_i \\
&= \frac{2^{2\nu+1}}{\left(n+1\right)^{2\nu}} m_{2\nu}\left(B\left(n+1,\frac{1}{2}\right)\right).
\end{aligned}$$

Finally, the Jensen-Shannon divergence becomes

$$\begin{aligned}\mathcal{JSD}\left(P_{X_{n+1}}, P_{X_n}\right) &= \frac{1}{4}\sum_{\nu=1}^{+\infty}\frac{1}{\nu\left(2\nu-1\right)}\Delta_\nu\left(P_{X_{n+1}}, P_{X_n}\right) \\ &= \sum_{\nu=1}^{+\infty}\frac{1}{\nu\left(2\nu-1\right)}\cdot\frac{2^{2\nu-1}}{\left(n+1\right)^{2\nu}}m_{2\nu}\left(B\left(n+1,\frac{1}{2}\right)\right).\end{aligned}$$

□

4. Proof of the Main Theorem

We are now in a position to show that the function $n \curvearrowright \frac{Y_n}{n}$ is increasing, or equivalently that inequality (3.9) holds.

Proof. We remark that it suffices to prove the following inequality

$$\sum_{\nu=1}^{3} \frac{1}{\nu(2\nu-1)} \cdot \frac{2^{2\nu-1}}{(n+1)^{2\nu}} m_{2\nu}\left(B\left(n+1, \frac{1}{2}\right)\right) \geq \frac{1}{2}\log\left(1+\frac{1}{n}\right) \tag{4.1}$$

since the terms $\nu > 3$ in the expansion of the Jensen-Shannon divergence are all non-negative. Now an explicit computation of the three first even central moments of a binomial random variable with parameters $n+1$ and $\frac{1}{2}$ yields

$$m_2 = \frac{n+1}{4}, \quad m_4 = \frac{(n+1)(3n+1)}{16} \quad \text{and} \quad m_6 = \frac{(n+1)(15n^2+1)}{64},$$

so that inequality (4.1) becomes

$$\frac{1}{60}\frac{30n^4+135n^3+245n^2+145n+37}{(n+1)^5} \geq \frac{1}{2}\log\left(1+\frac{1}{n}\right).$$

Let us now upper-bound the right hand side as follows

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

so that it suffices to prove that

$$\frac{1}{60} \cdot \frac{30n^4+135n^3+245n^2+145n+37}{(n+1)^5} - \frac{1}{2}\left(\frac{1}{n} - \frac{1}{2n^2} + \frac{1}{3n^3}\right) \geq 0.$$

Rearranging the terms yields the equivalent inequality

$$\frac{1}{60}\cdot\frac{10n^5-55n^4-63n^3-55n^2-35n-10}{(n+1)^5\, n^3}\geq 0$$

which is equivalent to the positivity of polynomial

$$P(n)=10n^5-55n^4-63n^3-55n^2-35n-10.$$

Assuming first that $n\geq 7$, we remark that

$$P(n)\geq 10n^5-n^4\left(55+\frac{63}{6}+\frac{55}{6^2}+\frac{35}{6^3}+\frac{10}{6^4}\right)=\left(10n-\frac{5443}{81}\right)n^4$$

whose positivity is ensured as soon as $n\geq 7$.

This result can be extended to the values $1\leq n\leq 6$ by a direct inspection at the values of function $n\curvearrowright\frac{Y_n}{n}$ as given in the following table.

n	1	2	3	4	5	6
$\frac{e^{2H(X_n)}}{n}$	4	4	4.105	4.173	4.212	4.233

Table 1: Values of the function $n\curvearrowright\frac{Y_n}{n}$ for $1\leq n\leq 6$.

$\square$

5. Acknowledgements

The authors want to thank Rudolf Ahlswede for useful discussions and pointing our attention to earlier work on the continuous and the discrete Entropy Power Inequalities.

References

[1] C.E. SHANNON, A mathematical theory of communication, *Bell Syst. Tech. J.*, **27** (1948), pp. 379–423 and 623–656.

[2] A.J. STAM, Some inequalities satisfied by the quantities of information of Fisher and Shannon, *Inform. Contr.*, **2** (1959), 101–112.

[3] N. M. BLACHMAN, The convolution inequality for entropy powers, *IEEE Trans. Inform. Theory*, **IT-11** (1965), 267–271.

[4] M.H.M. COSTA, A new entropy power inequality, *IEEE Trans. Inform. Theory*, **31** (1985), 751–760.

[5] A. DEMBO, Simple proof of the concavity of the entropy power with respect to added Gaussian noise, *IEEE Trans. Inform. Theory*, **35** (1989), 887–888.

[6] O. JOHNSON, A conditional entropy power inequality for dependent variables, *Statistical Laboratory Research Reports*, **20** (2000), Cambridge University.

[7] A. WYNER AND J. ZIV, A theorem on the entropy of certain binary sequences and applications: Part I, *IEEE Trans. Inform. Theory*, **IT-19** (1973), 769–772.

[8] A. WYNER, A theorem on the entropy of certain binary sequences and applications: Part II, *IEEE Trans. Inform. Theory*, **IT-19** (1973), 772–777.

[9] H.S. WITSENHAUSEN, Entropy inequalities for discrete channels, *IEEE Trans. Inform. Theory*, **IT-20** (1974), 610–616.

[10] R. AHLSWEDE AND J. KÖRNER, On the connection between the entropies of input and output distributions of discrete memoryless channels, *Proceedings of the Fifth Conference on Probability Theory, Brasov*, Sept. 1974, 13–22, Editura Academiei Republicii Socialiste Romania, Bucuresti 1977.

[11] S. SHAMAI AND A. WYNER, A binary analog to the entropy-power inequality, *IEEE Trans. Inform. Theory*, **IT-36** (1990), 1428–1430.

[12] F. TOPSØE, Information theoretical optimization techniques, *Kybernetika*, **15**(1) (1979), 8–27.

[13] F. TOPSØE, Some inequalities for information divergence and related measures of discrimination, *IEEE Tr. Inform. Theory*, **IT-46**(4) (2000), 1602–1609.