Communications in Information and Systems
Volume 16, Number 2, 111–115, 2016

Monotonicity of entropy and Fisher information: a quick proof via maximal correlation

Thomas A. Courtade

A simple proof is given for the monotonicity of entropy and Fisher information associated to sums of i.i.d. random variables. The proof relies on a characterization of maximal correlation for partial sums due to Dembo, Kagan and Shepp.

1. Introduction

Assume throughout that X is a random variable with density f and finite variance. The entropy $h(X)$ and, under mild regularity conditions on f, the Fisher information $J(X)$ are defined via

$$h(X) = -\mathbb{E}[\log f(X)], \qquad J(X) = \mathbb{E}\left[\rho_X^2(X)\right],$$

where $\rho_X = f'/f$ denotes the score function associated to X.

Let $X_1, X_2, \ldots$ be i.i.d. copies of X and define $S_n = X_1 + \cdots + X_n$, $n \geq 1$, and its standardized counterpart $U_n = \frac{1}{\sqrt{n}} S_n$. Two celebrated results established by Artstein, Ball, Barthe and Naor [1] are:

i) the entropies $h(U_n)$ are non-decreasing in n; and

ii) the Fisher informations $J(U_n)$ are non-increasing in n.

In other words, the respective central limit theorems for entropy [2] and Fisher information [3] enjoy monotone convergence (the latter holding under mild regularity conditions on f).

The aim of this note is to point out a simple and brief proof of these facts using a characterization of maximal correlation for sums of i.i.d. random variables.

Key words and phrases: Entropy, Fisher information, maximal correlation.
This work supported by the Center for Science of Information (CSoI), a NSF Science and Technology Center, under grant agreement CCF-0939370.

2. Monotonicity of Fisher information and entropy

The maximal correlation associated to a random pair X, Y is defined (in one of its equivalent forms) as

$$r^2(X;Y) = \sup_{\vartheta} \frac{\mathbb{E}[|\mathbb{E}[\vartheta(X)|Y]|^2]}{\mathbb{E}[|\vartheta(X)|^2]}, \tag{1}$$

where the supremum is over all non-constant, real-valued functions ϑ with $\mathbb{E}\vartheta(X) = 0$. An unexpected property enjoyed by r^2, discovered by Dembo, Kagan and Shepp [4], is that $r^2(S_m; S_n) = m/n$ for $1 \le m \le n$. A brief proof of the Dembo-Kagan-Shepp identity has been recently obtained by Kamath and Nair using information-theoretic arguments [5].

As a consequence, if a function $\vartheta : \mathbb{R} \to \mathbb{R}$ satisfies $\mathbb{E}\vartheta(S_m) = 0$, then definition (1) combined with the Dembo-Kagan-Shepp identity yields

$$\mathbb{E}[|\mathbb{E}[\vartheta(S_m)|S_n]|^2] \le \frac{m}{n}\mathbb{E}[|\vartheta(S_m)|^2] \qquad 1 \le m \le n. \tag{2}$$

The contraction (2) is the first ingredient in our proof, and we shall need one more: the behavior of score functions under convolution, observed by Stam [6].

Lemma 1. *Let U, V be independent random variables with smooth densities and put $W = U + V$. If ρ_U and ρ_W denote the score functions of U and W respectively, then*

$$\rho_W(w) = \mathbb{E}[\rho_U(U)|W = w]. \tag{3}$$

Identity (3) is proved by exchanging orders of differentiation and integration, and is justified by smoothness of densities (e.g., [7, Lemma 1.20]).

Theorem 1 (Monotonicity of Fisher Information). *Assume X has smooth density. For $1 \le m \le n$, $J(U_n) \le J(U_m)$.*

Proof. By Lemma 1, we have $\rho_{S_n}(s) = \mathbb{E}[\rho_{S_m}(S_m)|S_n = s]$. Moreover, $\mathbb{E}\rho_{S_m}(S_m) = 0$, so that $\vartheta = \rho_{S_m}$ is a valid choice in (2). Hence, from the definition of Fisher information and (2), we conclude

$$\begin{aligned} J(S_n) = \mathbb{E}[\rho^2_{S_n}(S_n)] &= \mathbb{E}[|\mathbb{E}\left[\rho_{S_m}(S_m)|S_n\right]|^2] \\ &\le \frac{m}{n}\mathbb{E}[\rho^2_{S_m}(S_m)] = \frac{m}{n}J(S_m). \end{aligned} \tag{4}$$

Noting the scaling property $\alpha^2 J\left(\alpha X\right) = J(X)$ finishes the proof. $\square$

Exactly as in [1], the entropy counterpart follows directly from a standard semigroup argument, which derives from Stam's seminal paper [6]. We include it for completeness.

Theorem 2 (Monotonicity of Entropy). *For* $1 \leq m \leq n$, $h(U_m) \leq h(U_n)$.

Proof. For a random variable Z with unit variance, define the Ornstein-Uhlenbeck evolutes $Z_t = e^{-t}Z + (1 - e^{-2t})^{1/2}G$, where G is standard normal independent of Z. Note that Z_t has smooth density for $t > 0$. By de Bruijn's identity (e.g., [6],[7, Appendix C]),

$$h(G) - h(Z) = \int_0^\infty (J(Z_t) - 1)\,\mathrm{d}t. \tag{5}$$

Using these facts, we find that Theorem 2 follows from Theorem 1 by considering the Ornstein-Uhlenbeck evolutes of the X_i's (and consequently U_m and U_n) and integrating along the semigroup. □

3. Historical remarks

Suggested by Shannon's entropy power inequality (EPI), monotonicity of entropy was a long-held conjecture that was eventually verified in 2004 when Artstein, Ball, Barthe and Naor (ABBN) established a '*leave-one-out*' EPI for sums of independent random variables using a variational characterization of Fisher information [1]. Their results imply that the Fisher information and entropy associated to sums of independent – but not necessarily identically distributed – random variables enjoy a monotonicity property that is more general than what we have proved in the present note. Since then, another proof of the ABBN inequality was given by Tulino and Verdú [8] using information-estimation relationships, and Shlyakhtenko has proved a free probability extension in [9]. Madiman and Barron [10, 11] and Madiman and Ghassemi [12] have extended the ABBN results to sums of arbitrary subsets of independent, non-identically distributed random variables.

It is interesting to note that the Dembo-Kagan-Shepp inequality (2) has been known since 2001, but apparently has not been connected to proving monotonicity of entropy until now. In retrospect, however, this connection should not be surprising. Indeed, all of the above referenced proofs (including that of Dembo, Kagan and Shepp [4]) critically hinge on variations of a 'variance drop' inequality due to Hoeffding [13]; once an appropriate variance drop inequality is identified, the respective proofs and that given for Theorem 1 above follow a similar program. The only notable exception in this regard is the proof of (2) by Kamath and Nair [5], which favors an information inequality over a variance drop inequality. In any case, the brief proof of Theorem 1 illustrates that monotonicity of entropy and Fisher information may be viewed as a direct consequence of the contraction $\mathbb{E}[|\mathbb{E}[\vartheta(S_m)|S_n]|^2] \leq \frac{m}{n}\mathbb{E}[|\vartheta(S_m)|^2]$, and may be of interest to those familiar with the Dembo-Kagan-Shepp maximal correlation identity, or the Kamath-Nair strong data processing result.

Finally, we observe that the proof of (2) in [5] goes through verbatim for random vectors, so the argument above extends immediately to the multidimensional setting.

Acknowledgments

The author thanks Mokshay Madiman and an anonymous referee for helpful comments that improved the historical remarks. He also thanks the editor Sergio Verdú for suggestions that improved the presentation.

References

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

[2] A. R. Barron, *Entropy and the central limit theorem*, The Annals of probability (1986), 336–342.

[3] O. Johnson and A. Barron, *Fisher information inequalities and the central limit theorem*, Probability Theory and Related Fields **129** (2004), no. 3, 391–409.

[4] A. Dembo, A. Kagan, and L. A. Shepp, *Remarks on the maximum correlation coefficient*, Bernoulli **7** (2001), no. 2, 343–350.

[5] S. Kamath and C. Nair, *The strong data processing constant for sums of iid random variables*, Proceedings of the 2015 IEEE International Symposium on Information Theory, Hong Kong, June 2015.

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

[7] O. Johnson, *Information theory and the central limit theorem*, Vol. 8. London: Imperial College Press, 2004.

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

[9] D. Shlyakhtenko, *Shannon's monotonicity problem for free and classical entropy*, Proc. Nat. Acad. Sci. **104** (2007), no. 39, 15254–15258.

[10] M. Madiman and A. R. Barron, *The monotonicity of information in the central limit theorem and entropy power inequalities*, Proceedings of the 2006 IEEE International Symposium on Information Theory, Seattle, Washington, July 2006.

[11] M. Madiman and A. R. Barron, *Generalized entropy power inequalities and monotonicity properties of information*, IEEE Transactions on Information Theory **53** (2007), no. 7, 2317–2329.

[12] M. Madiman and F. Ghassemi, *The entropy power of a sum is fractionally superadditive*, Proceedings of the 2009 IEEE International Symposium on Information Theory, Seoul, Korea, July 2009.

[13] W. Hoeffding, *A class of statistics with asymptotically normal distribution*, The annals of mathematical statistics (1948), 293–325.

Department of Electrical Engineering and Computer Sciences
University of California, Berkeley, CA 94720, USA
E-mail address: courtade@berkeley.edu

Received July 29, 2016