On the monotonicity of Shannon entropy

Bogdan Grechuk • 3 Jul 2026

This post continues my series on favorite theorems of the 21st century. For an overview of the categories and earlier selections, see this post.

My choice for 2003 in probability theory is a theorem of Artstein, Ball, Barthe and Naor confirming Lieb's conjecture on the monotonicity of Shannon entropy. The theorem was proved in 2003 and published in 2004 (Artstein et al. 2004).

A central result in probability theory is the *central limit theorem* (CLT). It states that if $X_1, X_2, \dots$ are independent and identically distributed random variables with finite mean $\mu = \mathbb{E}[X_i]$ and finite variance $\sigma^2 = \mathrm{Var}(X_i)$, then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$$

converges in distribution to the normal distribution $N(0, \sigma^2)$. In the special case where the X_i follow a Bernoulli distribution, this result goes back to de Moivre in 1733. A general form was proved by Lyapunov (Lyapunov 1901) in 1901.

Assume now, for simplicity, that $\mu = 0$. Then the CLT says that the normalized sums

$$Y_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$$

converge in distribution to a normal random variable, no matter what the original distribution of the X_i looks like, provided the variance is finite. The question behind Lieb's conjecture is whether this convergence can be made monotone: is Y_{n+1}, in a precise sense, always closer to the normal distribution than Y_n?

The sense of closeness in Lieb's conjecture is given by Shannon entropy (Shannon 1948). For a real-valued random variable X with density f, its continuous Shannon entropy is

$$\mathrm{Ent}(X) = -\int_{\mathbb{R}} f(x)\log f(x)\,\mathrm{d}x,$$

whenever this integral is well defined. Among all real-valued random variables with a fixed variance, the normal distribution has the largest entropy. Thus, if the entropy of Y_n increases with n, then the normalized sums become increasingly Gaussian from the entropic point of view.

This monotonicity was known in a special dyadic form before Lieb's conjecture. Shannon (Shannon and Weaver 1949) outlined the argument that

$$\mathrm{Ent}(Y_2) \geq \mathrm{Ent}(Y_1),$$

and Stam (Stam 1959) later gave a rigorous proof. Applying this inequality repeatedly gives

$$\mathrm{Ent}(Y_1) \leq \mathrm{Ent}(Y_2) \leq \mathrm{Ent}(Y_4) \leq \cdots \leq \mathrm{Ent}(Y_{2^k}) \leq \cdots .$$

Lieb (Lieb 1978) conjectured that the full sequence is monotone:

$$\mathrm{Ent}(Y_n) \leq \mathrm{Ent}(Y_{n+1}) \qquad \text{for all } n \geq 1.$$

Before the work of Artstein, Ball, Barthe and Naor, this conjecture remained open even for the first non-dyadic step, namely the comparison between Y_2 and Y_3. Their theorem settled the conjecture completely.

Theorem 1 *Let $X_1, X_2, \ldots$ be independent and identically distributed real-valued random variables with a density and finite variance. Then*

$$\mathrm{Ent}\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)$$

is a non-decreasing function of n, whenever the entropies are well defined.

The main difficulty in proving Theorem 1 is that there is no obvious way to compare the sum of n independent copies of a random variable with the sum of $n+1$ independent copies. The dyadic inequality avoids this problem because Y_{2n} can be written as the normalized sum of two independent copies of Y_n. No such direct decomposition is available when one passes from n to $n+1$.

Artstein, Ball, Barthe and Naor overcame this obstacle by shifting the problem from entropy to Fisher information. For a random variable X with smooth positive density f, the Fisher information is

$$J(X) = \int_{\mathbb{R}} \frac{(f'(x))^2}{f(x)}\,\mathrm{d}x.$$

Using de Bruijn's identity, the desired entropy inequality can be reduced to the Fisher-information inequality

$$J(Y_n) \geq J(Y_{n+1}). \tag{1}$$

They then proved a new formula for the Fisher information of a normalized sum. This formula is flexible enough to compare sums of different lengths, and it is precisely what makes the passage from n to $n+1$ possible.

The result is beautiful because it strengthens the central limit theorem from a statement about eventual convergence to a statement about steady progress. The CLT tells us that the normalized sums become Gaussian in the limit. The theorem of Artstein, Ball, Barthe and Naor says that, measured by entropy, they move toward the Gaussian distribution at every single step.

References

Artstein, Shiri, Keith Ball, Franck Barthe, and Assaf Naor. 2004. “Solution of Shannon’s Problem on the Monotonicity of Entropy.” *J. Amer. Math. Soc.* 17 (4): 975–82.

Lieb, E. H. 1978. “Proof of Wehrl’s Entropy Conjecture.” *Comm. Math. Phys.* 62: 35–40.

Lyapunov, Aleksandr Mikhailovich. 1901. “A General Proposition of Probability Theory.” *CR Acad. Sci. Paris* 132: 814–15.

Shannon, Claude E. 1948. “A Mathematical Theory of Communication.” *Bell System Technical Journal* 27 (3): 379–423.

Shannon, Claude Elwood, and Warren Weaver. 1949. *The Mathematical Theory of Communication*. University of Illinois Press.

Stam, Aart J. 1959. “Some Inequalities Satisfied by the Quantities of Information of Fisher and Shannon.” *Information and Control* 2 (2): 101–12.