

Chapter 7

Two Remarks on Generalized Entropy Power Inequalities



Mokshay Madiman, Piotr Nayar, and Tomasz Tkocz

Abstract This note contributes to the understanding of generalized entropy power inequalities. Our main goal is to construct a counter-example regarding monotonicity and entropy comparison of weighted sums of independent identically distributed log-concave random variables. We also present a complex analogue of a recent dependent entropy power inequality of Hao and Jog, and give a very simple proof.

Mokshay Madiman was supported in part by the U.S. National Science Foundation through the grant DMS-1409504. Piotr Nayar was partially supported by the National Science Centre Poland grant 2015/18/A/ST1/00553. The research leading to these results is part of a project that has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 637851). This work was also supported by the NSF under Grant No. 1440140, while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, for the “Geometric and Functional Analysis” program during the fall semester of 2017.

M. Madiman (✉)
University of Delaware, Newark, DE, USA
e-mail: madiman@udel.edu

P. Nayar
University of Warsaw, Warsaw, Poland
e-mail: nayar@mimuw.edu.pl

T. Tkocz
Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: ttkocz@math.cmu.edu

7.1 Introduction

The differential entropy of a random vector X with density f (with respect to Lebesgue measure on \mathbb{R}^d) is defined as

$$h(X) = - \int_{\mathbb{R}^d} f \log f,$$

provided that this integral exists. When the variance of a real-valued random variable X is kept fixed, it is a long known fact [11] that the differential entropy is maximized by taking X to be Gaussian. A related functional is the *entropy power* of X , defined by $N(X) = e^{\frac{2h(X)}{d}}$. As is usual, we abuse notation and write $h(X)$ and $N(X)$, even though these are functionals depending only on the density of X and not on its random realization.

The entropy power inequality is a fundamental inequality in both Information Theory and Probability, stated first by Shannon [34] and proved by Stam [36]. It states that for any two independent random vectors X and Y in \mathbb{R}^d such that the entropies of X , Y and $X + Y$ exist,

$$N(X + Y) \geq N(X) + N(Y).$$

In fact, it holds without even assuming the existence of entropies as long as we set an entropy power to 0 whenever the corresponding entropy does not exist, as noted by Bobkov and Chistyakov [6]. One reason for the importance of this inequality in Probability Theory comes from its close connection to the Central Limit Theorem (see, e.g., [21, 25]). It is also closely related to the Brunn–Minkowski inequality, and thereby to results in Convex Geometry and Geometric Functional Analysis (see, e.g., [7, 31]).

An immediate consequence of the above formulation of the entropy power inequality is its extension to n summands: if X_1, \dots, X_n are independent random vectors, then $N(X_1 + \dots + X_n) \geq \sum_{i=1}^n N(X_i)$. Suppose the random vectors X_i are not merely independent but also identically distributed, and that $S_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i$; these are the normalized partial sums that appear in the vanilla version of the Central Limit Theorem. Then one concludes from the entropy power inequality together with the scaling property $N(aX) = a^2 N(X)$ that $N(S_n) \geq N(S_1)$, or equivalently that

$$h(S_n) \geq h(S_1). \tag{7.1}$$

There are several refinements or generalizations of the inequality (7.1) that one may consider. In 2004, Artstein et al. [2] proved (see [13, 26, 35, 38] for simpler proofs and [27, 28] for extensions) that in fact, one has monotonicity of entropy along the Central Limit Theorem, i.e., $h(S_n)$ is a monotonically increasing sequence. If $N(0, 1)$ is the standard normal distribution, Barron [4] had proved

much earlier that $\bar{h}(S_n) \rightarrow h(N(0, 1))$ as long as X_1 has mean 0, variance 1, and $h(X_1) > -\infty$. Thus one has the monotone convergence of $h(S_n)$ to the Gaussian entropy, which is the maximum entropy possible under the moment constraints. By standard arguments, the convergence of entropies is equivalent to the relative entropy between the distribution of S_n and the standard Gaussian distribution converging to 0, and this in turn implies not just convergence in distribution but also convergence in total variation. This is the way in which entropy illuminates the Central Limit Theorem.

A different variant of the inequality (7.1) was recently given by Hao and Jog [20], whose paper may be consulted for motivation and proper discussion. A random vector $X = (X_1, \dots, X_n)$ in \mathbb{R}^n is called *unconditional* if for every choice of signs $\eta_1, \dots, \eta_n \in \{-1, +1\}$, the vector $(\eta_1 X_1, \dots, \eta_n X_n)$ has the same distribution as X . Hao and Jog [20] proved that if X is an unconditional random vector in \mathbb{R}^n , then $\frac{1}{n}h(X) \leq h\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right)$. If X has independent and identically distributed components instead of being unconditional, this is precisely $h(S_n) \geq h(S_1)$ for real-valued random variables X_i (i.e., in dimension $d = 1$).

The goal of this note is to shed further light on both of these generalized entropy power inequalities. We now explain precisely how we do so.

To motivate our first result, we first recall the notion of Schur-concavity. One vector $a = (a_1, \dots, a_n)$ in $[0, \infty)^n$ is *majorised* by another one $b = (b_1, \dots, b_n)$, usually denoted $a < b$, if the nonincreasing rearrangements $a_1^* \geq \dots \geq a_n^*$ and $b_1^* \geq \dots \geq b_n^*$ of a and b satisfy the inequalities $\sum_{j=1}^k a_j^* \leq \sum_{j=1}^k b_j^*$ for each $1 \leq k \leq n - 1$ and $\sum_{j=1}^n a_j = \sum_{j=1}^n b_j$. For instance, any vector a with nonnegative coordinates adding up to 1 is majorised by the vector $(1, 0, \dots, 0)$ and majorises the vector $(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n})$. Let $\Phi: \Delta_n \rightarrow \mathbb{R}$, where $\Delta_n = \{a \in [0, 1]^n : a_1 + \dots + a_n = 1\}$ is the standard simplex. We say that Φ is *Schur-concave* if $\Phi(a) \geq \Phi(b)$ when $a < b$. Clearly, if Φ is Schur-concave, then one has $\Phi(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) \geq \Phi(a) \geq \Phi(1, 0, \dots, 0)$ for any $a \in \Delta_n$.

Suppose X_1, \dots, X_n are i.i.d. copies of a random variable X with finite entropy, and we define

$$\Phi(a) = h\left(\sum \sqrt{a_i} X_i\right) \tag{7.2}$$

for $a \in \Delta_n$. Then the inequality (7.1) simply says that $\Phi(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) \geq \Phi(1, 0, \dots, 0)$, while the monotonicity of entropy in the Central Limit Theorem says that $\Phi(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) \geq \Phi(\frac{1}{n-1}, \dots, \frac{1}{n-1}, 0)$. Both these properties would be implied by (but in themselves are strictly weaker than) Schur-concavity. Thus one is led to the natural question: *Is the function Φ defined in (7.2) a Schur-concave function?* For $n = 2$, this would imply in particular that $h(\sqrt{\lambda} X_1 + \sqrt{1 - \lambda} X_2)$ is maximized over $\lambda \in [0, 1]$ when $\lambda = \frac{1}{2}$. The question on the Schur-concavity of Φ had been floating around for at least a decade, until [3] constructed a counterexample showing that Φ cannot be Schur-concave even for $n = 2$. It was conjectured in [3], however, that for $n = 2$, the Schur-concavity should hold

if the random variable X has a log-concave distribution, i.e., if X_1 and X_2 are independent, identically distributed, log-concave random variables, the function $\lambda \mapsto h\left(\sqrt{\lambda}X_1 + \sqrt{1-\lambda}X_2\right)$ should be nondecreasing on $[0, \frac{1}{2}]$. More generally, one may ask: if X_1, \dots, X_n are n i.i.d. copies of a log-concave random variable X , is it true that $h\left(\sum a_i X_i\right) \geq h\left(\sum b_i X_i\right)$ when $(a_1^2, \dots, a_n^2) \prec (b_1^2, \dots, b_n^2)$? Equivalently, is Φ Schur-concave when X is log-concave?

Our first result implies that the answer to this question is negative. The way we show this is the following: since $(1, \frac{1}{n}, \dots, \frac{1}{n}, \frac{1}{n}) \prec (1, \frac{1}{n-1}, \dots, \frac{1}{n-1}, 0)$, if Schur-concavity held, then the sequence $h\left(X_1 + \frac{X_2 + \dots + X_{n+1}}{\sqrt{n}}\right)$ would be nondecreasing. If we moreover establish convergence of this sequence to $h(X_1 + G)$, where G is an independent Gaussian random variable with the same variance as X_1 , we would have in particular that $h\left(X_1 + \frac{X_2 + \dots + X_{n+1}}{\sqrt{n}}\right) \leq h(X_1 + G)$. We construct examples where the opposite holds.

Theorem 7.1 *There exists a symmetric log-concave random variable X with variance 1 such that if X_0, X_1, \dots are its independent copies and n is large enough, we have*

$$h\left(X_0 + \frac{X_1 + \dots + X_n}{\sqrt{n}}\right) > h(X_0 + Z),$$

where Z is a standard Gaussian random variable, independent of the X_i . Moreover, the left hand side of the above inequality converges to $h(X_0 + Z)$ as n tends to infinity. Consequently, even if X is drawn from a symmetric, log-concave distribution, the function Φ defined in (7.2) is not always Schur-concave.

Here by a symmetric distribution, we mean one whose density f satisfies $f(-x) = f(x)$ for each $x \in \mathbb{R}$.

In contrast to Theorem 7.1, Φ does turn out to be Schur-concave if the distribution of X is a symmetric Gaussian mixture, as recently shown in [15]. We suspect that Schur-concavity also holds for uniform distributions on intervals (cf. [1]).

Theorem 7.1 can be compared with the afore-mentioned monotonicity of entropy property of the Central Limit Theorem. It also provides an example of two independent symmetric log-concave random variables X and Y with the same variance such that $h(X + Y) > h(X + Z)$, where Z is a Gaussian random variable with the same variance as X and Y , independent of them, which is again in contrast to symmetric Gaussian mixtures (see [15]). The interesting question posed in [15] of whether, for two i.i.d. summands, swapping one for a Gaussian with the same variance increases entropy, remains open.

Our proof of Theorem 7.1 is based on sophisticated and remarkable Edgeworth type expansions recently developed by Bobkov et al. [9] en route to obtaining precise rates of convergence in the entropic central limit theorem, and is detailed in Sect. 7.2.

The **second contribution** of this note is an exploration of a technique to prove inequalities akin to the entropy power inequality by using symmetries and invariance properties of entropy. It is folklore that when X_1 and X_2 are i.i.d. from a symmetric distribution, one can deduce the inequality $h(S_2) \geq h(S_1)$ in an extremely simple fashion (in contrast to any full proof of the entropy power inequality, which tends to require relatively sophisticated machinery— either going through Fisher information or optimal transport or rearrangement theory or functional inequalities). In Sect. 7.3, we will recall this simple proof, and also deduce some variants of the inequality $h(S_2) \geq h(S_1)$ by playing with this basic idea of using invariance, including a complex analogue of a recent entropy power inequality for dependent random variables obtained by Hao and Jog [20].

Theorem 7.2 *Let $X = (X_1, \dots, X_n)$ be a random vector in \mathbb{C}^n which is complex-unconditional, that is for every complex numbers z_1, \dots, z_n such that $|z_j| = 1$ for every j , the vector $(z_1 X_1, \dots, z_n X_n)$ has the same distribution as X . Then*

$$\frac{1}{n}h(X) \leq h\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right).$$

Our proof of Theorem 7.2, which is essentially trivial thanks to the existence of complex Hadamard matrices, is in contrast to the proof given by Hao and Jog [20] for the real case that proves a Fisher information inequality as an intermediary step.

We make some remarks on complementary results in the literature. Firstly, in contrast to the failure of Schur-concavity of Φ implied by Theorem 7.1, the function $\Xi: \Delta_n \rightarrow \mathbb{R}$ defined by $\Xi(a) = h(\sum a_i X_i)$ for i.i.d. copies X_i of a random variable X , is actually Schur-convex when X is log-concave [41]. This is an instance of a reverse entropy power inequality, many more of which are discussed in [31]. Note that the weighted sums that appear in the definition of Φ are relevant to the Central Limit Theorem because they have fixed variance, unlike the weighted sums that appear in the definition of Ξ .

Secondly, motivated by the analogies with Convex Geometry mentioned earlier, one may ask if the function $\Psi: \Delta_n \rightarrow \mathbb{R}$ defined by $\Psi(a) = \text{vol}_d(\sum_{i=1}^n a_i B)$, is Schur-concave for any Borel set $B \subset \mathbb{R}^d$, where vol_d denotes the Lebesgue measure on \mathbb{R}^d and the notation for summation is overloaded as usual to also denote Minkowski summation of sets. (Note that unless B is convex, $(a_1 + a_2)B$ is a subset of, but generally not equal to, $a_1 B + a_2 B$.) The Brunn–Minkowski inequality implies that $\Psi(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) \geq \Psi(1, 0, \dots, 0)$. The inequality $\Psi(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}) \geq \Psi(\frac{1}{n-1}, \dots, \frac{1}{n-1}, 0)$, which is the geometric analogue of the monotonicity of entropy in the Central Limit Theorem, was conjectured to hold in [8]. However, it was shown in [16] (cf. [17]) that this inequality fails to hold, and therefore Ψ cannot be Schur-concave, for arbitrary Borel sets B . Note that if B is convex, Ψ is trivially Schur-concave, since it is a constant function equal to $\text{vol}_d(B)$.

Finally, it has recently been observed in [32, 33, 40] that majorization ideas are very useful in understanding entropy power inequalities in discrete settings, such as on the integers or on cyclic groups of prime order.

7.2 Failure of Schur-Concavity

Recall that a probability density f on \mathbb{R} is said to be *log-concave* if it is of the form $f = e^{-V}$ for a convex function $V: \mathbb{R} \rightarrow \mathbb{R} \cup \{\infty\}$. Log-concave distributions emerge naturally from the interplay between information theory and convex geometry, and have recently been a very fruitful and active topic of research (see the recent survey [31]).

This section is devoted to a proof of Theorem 7.1, which in particular falsifies the Schur-concavity of Φ defined by (7.2) even when the distribution under consideration is log-concave.

Let us denote

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

and let p_n be the density of Z_n and let φ be the density of Z . Since X_0 is assumed to be log-concave, it satisfies $\mathbb{E}|X_0|^s < \infty$ for all $s > 0$. According to the Edgeworth-type expansion described in [9, (Theorem 3.2 in Chapter 3)], we have (with any $m \leq s < m + 1$)

$$(1 + |x|^m)(p_n(x) - \varphi_m(x)) = o(n^{-\frac{s-2}{2}}) \quad \text{uniformly in } x,$$

where

$$\varphi_m(x) = \varphi(x) + \sum_{k=1}^{m-2} q_k(x)n^{-k/2}.$$

Here the functions q_k are given by

$$q_k(x) = \varphi(x) \sum H_{k+2j}(x) \frac{1}{r_1! \cdots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \cdots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k},$$

where H_n are Hermite polynomials,

$$H_n(x) = (-1)^n e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2},$$

and the summation runs over all nonnegative integer solutions (r_1, \dots, r_k) to the equation $r_1 + 2r_2 + \cdots + kr_k = k$, and one uses the notation $j = r_1 + \cdots + r_k$. The numbers γ_k are the cumulants of X_0 , namely

$$\gamma_k = i^{-k} \frac{d^k}{dt^k} \log \mathbb{E} e^{itX_0} \Big|_{t=0}.$$

Let us calculate φ_4 . Under our assumption (symmetry of X_0 and $\mathbb{E}X_0^2 = 1$), we have $\gamma_3 = 0$ and $\gamma_4 = \mathbb{E}X_0^4 - 3$. Therefore $q_1 = 0$ and

$$q_2 = \frac{1}{4!}\gamma_4\varphi H_4 = \frac{1}{4!}\gamma_4\varphi^{(4)}, \quad \varphi_4 = \varphi + \frac{1}{n} \cdot \frac{1}{4!}(\mathbb{E}X_0^4 - 3)\varphi^{(4)}. \quad (7.3)$$

We get that for any $\varepsilon \in (0, 1)$

$$(1 + x^4)(p_n(x) - \varphi_4(x)) = o(n^{-\frac{3-\varepsilon}{2}}), \quad \text{uniformly in } x. \quad (7.4)$$

Let f be the density of X_0 . Let us assume that it is of the form $f = \varphi + \delta$, where δ is even, smooth and compactly supported (say, supported in $[-2, -1] \cup [1, 2]$) with bounded derivatives. Moreover, we assume that $\frac{1}{2}\varphi \leq f \leq 2\varphi$, in particular $|\delta| \leq 1/4$. Multiplying δ by a very small constant we can ensure that f is log-concave.

We are going to use Theorem 1.3 from [10]. To check the assumptions of this theorem, we first observe that for any $\alpha > 1$ we have

$$D_\alpha(Z_1||Z) = \frac{1}{\alpha - 1} \log \left(\int \left(\frac{\varphi + \delta}{\varphi} \right)^\alpha \varphi \right) < \infty,$$

since δ has bounded support. We have to show that for sufficiently big $\alpha^* = \frac{\alpha}{\alpha-1}$ there is

$$\mathbb{E}e^{tX_0} < e^{\alpha^*t^2/2}, \quad t \neq 0.$$

Since X_0 is symmetric, we can assume that $t > 0$. Then

$$\begin{aligned} \mathbb{E}e^{tX_0} &= e^{t^2/2} + \sum_{k=1}^{\infty} \frac{t^{2k}}{(2k)!} \int x^{2k} \delta(x) dx \leq e^{t^2/2} + \sum_{k=1}^{\infty} \frac{t^{2k}}{(2k)!} 2^{2k} \int_{-2}^2 |\delta(x)| dx \\ &< e^{t^2/2} + \sum_{k=1}^{\infty} \frac{(2t)^{2k}}{(2k)!} = 1 + \sum_{k=1}^{\infty} \left(\frac{t^{2k}}{2^k k!} + \frac{(2t)^{2k}}{(2k)!} \right) \\ &\leq 1 + \sum_{k=1}^{\infty} \left(\frac{t^{2k}}{k!} + \frac{(2t)^{2k}}{k!} \right) \leq \sum_{k=0}^{\infty} \frac{t^{2k} 4^{2k}}{k!} = e^{16t^2}, \end{aligned}$$

where we have used the fact that $\int \delta(x) dx = 0$, δ has a bounded support contained in $[-2, 2]$ and $|\delta| \leq 1/4$. We conclude that

$$|p_n(x) - \varphi(x)| \leq \frac{C_0}{n} e^{-x^2/64} \quad (7.5)$$

for some constant C_0 independent of n . (In this proof, C_0, C_1, \dots denote sufficiently large constants that may depend on the distribution of X_0 .) Thus

$$p_n(x) \leq \varphi(x) + \frac{C_0}{n} e^{-x^2/64} \leq C_1 e^{-x^2/64}. \tag{7.6}$$

Another consequence of (7.5) is the inequality

$$p_n(x) \geq \frac{1}{10} \quad \text{for } |x| \leq 1 \tag{7.7}$$

and large enough n .

We now prove the convergence part of the theorem. From (7.5) we get that $p_n \rightarrow \varphi$ pointwise. Moreover, from (7.6) and from the inequality $f \leq 2\varphi$ we get, by using Lebesgue’s dominated convergence theorem, that $f * p_n \rightarrow f * \varphi$. In order to show that $\int f * p_n \log f * p_n \rightarrow \int f * \varphi \log f * \varphi$ it is enough to bound $f * p_n | \log f * p_n |$ by some integrable function m_0 independent of n and use Lebesgue’s dominated convergence theorem. To this end we observe that by (7.6) we have

$$(f * p_n)(x) \leq 2(\varphi * p_n)(x) \leq \frac{2C_1}{\sqrt{2\pi}} \int e^{-t^2/2} e^{-(x-t)^2/64} dt \leq 2C_1 e^{-x^2/66}. \tag{7.8}$$

Moreover, by (7.7)

$$(f * p_n)(x) \geq \frac{1}{2}(\varphi * p_n)(x) \geq \frac{1}{20} \int_{-1}^1 \varphi(x - t) dt \geq \frac{1}{10} \varphi(|x| + 1). \tag{7.9}$$

Combining (7.8) with (7.9) we get

$$| \log(f * p_n)(x) | \leq \max \left\{ | \log 2C_1 |, \frac{1}{10} | \log \varphi(|x| + 1) | \right\} \leq C_2(1 + x^2). \tag{7.10}$$

From (7.10) and (7.8) we see that the function $m_0(x) = 2C_1 C_2 e^{-x^2/66} (1 + x^2)$ is the required majorant.

Let us define $h_n = p_n - \varphi_4$. Note that by (7.3) we have $\varphi_4 = \varphi + \frac{c_1}{n} \varphi^{(4)}$, where $c_1 = \frac{1}{4!} (\mathbb{E}X_0^4 - 3)$. We have

$$\begin{aligned} \int f * p_n \log f * p_n &= \int \left(f * \varphi + \frac{c_1}{n} f * \varphi^{(4)} + f * h_n \right) \log f * p_n \\ &= \int f * \varphi \log f * p_n + \frac{c_1}{n} \int f * \varphi^{(4)} \log f * p_n \end{aligned}$$

$$\begin{aligned}
 &+ \int f * h_n \log f * p_n \\
 &= I_1 + I_2 + I_3.
 \end{aligned}$$

We first bound I_3 . Note that using (7.4) with $\varepsilon = 1/2$ we get

$$|(f * h_n)(x)| \leq 2(\varphi * |h_n|)(x) \leq C_3 n^{-5/4} \int e^{-y^2/2} \frac{1}{1 + (x - y)^4} dy \tag{7.11}$$

for sufficiently large n . Assuming without loss of generality that $x > 0$, we have

$$\begin{aligned}
 \int e^{-y^2/2} \frac{1}{1 + (x - y)^4} dy &\leq \int_{y \in [\frac{1}{2}x, 2x]} e^{-y^2/2} \frac{1}{1 + (x - y)^4} dy \\
 &+ \int_{y \notin [\frac{1}{2}x, 2x]} e^{-y^2/2} \frac{1}{1 + (x - y)^4} dy \\
 &\leq \int_{y \in [\frac{1}{2}x, 2x]} e^{-x^2/8} dy \\
 &+ \frac{1}{1 + \frac{1}{16}x^4} \int_{y \notin [\frac{1}{2}x, 2x]} e^{-y^2/2} dy \\
 &\leq \frac{3}{2} x e^{-x^2/8} + \frac{\sqrt{2\pi}}{1 + \frac{1}{16}x^4} \leq \frac{C_4}{1 + x^4}.
 \end{aligned}$$

Combining this with (7.11) one gets for large n

$$|(f * h_n)(x)| \leq C_3 C_4 n^{-5/4} \frac{1}{1 + x^4}. \tag{7.12}$$

Inequalities (7.12) and (7.10) give for large n ,

$$|I_3| \leq C_3 C_4 C_2 n^{-5/4} \int \frac{1 + x^2}{1 + x^4} dx \leq 5 C_3 C_4 C_2 n^{-5/4}. \tag{7.13}$$

We now take care of I_2 by showing that

$$I_2 = \frac{c_1}{n} \int f * \varphi^{(4)} \log f * p_n = \frac{c_1}{n} \int f * \varphi^{(4)} \log f * \varphi + o(n^{-1}). \tag{7.14}$$

To this end it suffices to show that $\int f * \varphi^{(4)} \log f * p_n \rightarrow \int f * \varphi^{(4)} \log f * \varphi$. As we already observed $f * p_n \rightarrow f * \varphi$ pointwise. Taking into account the bound (7.10), to find a majorant m_1 of $f * \varphi^{(4)} \log f * p_n$, it suffices to observe that

$|\varphi^{(4)}(t)| \leq C_5 e^{-t^2/4}$ and thus

$$|f * \varphi^{(4)}|(x) \leq 2(\varphi * |\varphi^{(4)}|)(x) \leq 2C_5 \int e^{-(x-t)^2/2} e^{-t^2/4} dt \leq 8C_5 e^{-x^2/6}.$$

One can then take $m_1(x) = 8C_5 C_2 e^{-x^2/6} (1 + x^2)$.

By Jensen's inequality,

$$I_1 = \int f * \varphi \log f * p_n \leq \int f * \varphi \log f * \varphi = -h(X_0 + Z). \tag{7.15}$$

Putting (7.15), (7.14) and (7.13) together we get

$$\int f * p_n \log f * p_n \leq \int f * \varphi \log f * \varphi + \frac{c_1}{n} \int (f * \varphi)^{(4)} \log(f * \varphi) + o(n^{-1}).$$

This is

$$h(X_0 + Z) \leq h(X_0 + Z_n) + \frac{1}{n} \cdot \frac{1}{4!} (\mathbb{E}X_0^4 - 3) \int (f * \varphi)^{(4)} \log(f * \varphi) + o(n^{-1}).$$

It is therefore enough to construct X_0 (satisfying all previous conditions) such that

$$(\mathbb{E}X_0^4 - 3) \int (f * \varphi)^{(4)} \log(f * \varphi) < 0.$$

It actually suffices to construct a smooth compactly supported even function g such that $\int g = \int g x^2 = \int g x^4 = 0$ and the function $f = \varphi + \varepsilon g$ satisfies

$$\int (f * \varphi)^{(4)} \log(f * \varphi) > 0$$

for some fixed small ε . We then perturb g a bit to get $\mathbb{E}X_0^4 < 3$ instead of $\mathbb{E}X_0^4 = 3$. This can be done without affecting log-concavity.

Let $\varphi_2(x) = (\varphi * \varphi)(x) = \frac{1}{2\sqrt{\pi}} e^{-x^2/4}$. Note that $\varphi_2^{(4)}(x) = \varphi_2(x) (\frac{3}{4} - \frac{3}{4}x^2 + \frac{1}{16}x^4)$. We have

$$\begin{aligned} \int (f * \varphi)^{(4)} \log(f * \varphi) &= \int (\varphi_2 + \varepsilon \varphi * g)^{(4)} \log(\varphi_2 + \varepsilon \varphi * g) \\ &= \int (\varphi_2 + \varepsilon \varphi * g)^{(4)} \left(\log(\varphi_2) + \varepsilon \frac{\varphi * g}{\varphi_2} \right. \\ &\quad \left. - \frac{1}{2} \varepsilon^2 \left(\frac{\varphi * g}{\varphi_2} \right)^2 + r_\varepsilon(x) \right) dx. \end{aligned}$$

We shall show that $\int |(\varphi_2 + \varepsilon\varphi * g)^{(4)}| |r_\varepsilon| \leq C_8 |\varepsilon|^3$. To justify this we first observe that by Taylor's formula with the Lagrange remainder, we have

$$|\log(1+a) - a + a^2/2| \leq \frac{1}{3} \frac{|a|^3}{(1-|a|)^3} \quad |a| < 1. \quad (7.16)$$

Due to the fact that g is bounded and compactly supported, we have

$$|\varphi * g|(x) \leq C_6 \int_{-C_6}^{C_6} \varphi(x-t) dt \leq 2C_6^2 \varphi((|x| - C_6)_+) \leq 2C_6^2 e^{-(|x| - C_6)_+^2/2}.$$

Thus

$$\frac{|\varphi * g|(x)}{\varphi_2(x)} \leq 4\sqrt{\pi} C_6^2 e^{x^2/4} e^{-(|x| - C_6)_+^2/2} \leq C_7.$$

Using (7.16) with $a = \varepsilon \frac{\varphi * g}{\varphi_2}$ and $|\varepsilon| < \frac{1}{2C_7}$ (in which case $|a| \leq 1/2$) we get

$$|r_\varepsilon(x)| = \left| \log \left(1 + \varepsilon \frac{\varphi * g}{\varphi_2} \right) - \varepsilon \frac{\varphi * g}{\varphi_2} + \frac{1}{2} \varepsilon^2 \left(\frac{\varphi * g}{\varphi_2} \right)^2 \right| \leq \frac{|\varepsilon|^3}{3} C_7^3 \frac{1}{(1 - \frac{1}{2})^3}.$$

Thus

$$\int |(\varphi_2 + \varepsilon\varphi * g)^{(4)}| |r_\varepsilon| \leq \frac{8}{3} C_7^3 |\varepsilon|^3 \int \left(|\varphi_2^{(4)}| + \frac{1}{2C_7} \varphi * |g^{(4)}| \right) \leq C_8 |\varepsilon|^3.$$

Therefore

$$\begin{aligned} \int (f * \varphi)^{(4)} \log(f * \varphi) &= \int (\varphi_2 + \varepsilon\varphi * g)^{(4)} \left(\log(\varphi_2) + \varepsilon \frac{\varphi * g}{\varphi_2} \right. \\ &\quad \left. - \frac{1}{2} \varepsilon^2 \left(\frac{\varphi * g}{\varphi_2} \right)^2 \right) + o(\varepsilon^2). \end{aligned}$$

Integrating by parts we see that the leading term in the above equation is

$$\begin{aligned} \int \varphi_2^{(4)} \log \varphi_2 &= \int \varphi_2^{(4)}(x) \log \left(\frac{1}{2\sqrt{\pi}} e^{-x^2/4} \right) dx \\ &= - \int \varphi_2^{(4)}(x) \left(\log(2\sqrt{\pi}) + \frac{1}{4} x^2 \right) dx \\ &= - \int \varphi_2(x) \left(\log(2\sqrt{\pi}) + \frac{1}{4} x^2 \right)^{(4)} dx = 0. \end{aligned}$$

The term in front of ε vanishes. Indeed, $\int \varphi_2^{(4)} \frac{\varphi * g}{\varphi_2} = \int (\frac{3}{4} - \frac{3}{4}x^2 + \frac{1}{16}x^4)(\varphi * g)$ which can be seen to vanish after using Fubini's theorem thanks to g being orthogonal to $1, x, \dots, x^4$. Moreover, $\int (\varphi * g)^{(4)} \log(\varphi_2) = \int (\varphi * g)(\log \frac{1}{2\sqrt{\pi}} - \frac{x^2}{4})^{(4)} = 0$. The term in front of ε^2 is equal to

$$J = \int \frac{(\varphi * g)^{(4)}(\varphi * g)}{\varphi_2} - \frac{1}{2} \int \frac{\varphi_2^{(4)}(\varphi * g)^2}{\varphi_2^2} = J_1 - J_2.$$

The first integral is equal to

$$J_1 = \int \int \int 2\sqrt{\pi} e^{x^2/4} g^{(4)}(s) g(t) \frac{1}{2\pi} e^{-(x-s)^2/2} e^{-(x-t)^2/2} dx ds dt.$$

Now,

$$\int 2\sqrt{\pi} e^{x^2/4} \frac{1}{2\pi} e^{-(x-s)^2/2} e^{-(x-t)^2/2} dx = \frac{2e^{\frac{1}{6}(-s^2+4st-t^2)}}{\sqrt{3}}.$$

Therefore,

$$J_1 = \frac{2}{\sqrt{3}} \int \int e^{\frac{1}{6}(-s^2+4st-t^2)} g^{(4)}(s) g(t) ds dt.$$

If we integrate the first integral four times by parts we get

$$J_1 = \frac{2}{81\sqrt{3}} \int \int e^{\frac{1}{6}(-s^2+4st-t^2)} \left[27 + s^4 - 8s^3t - 72t^2 + 16t^4 - 8st(-9 + 4t^2) + 6s^2(-3 + 4t^2) \right] g(s)g(t) ds dt$$

Since $\varphi_2^{(4)}/\varphi_2^2 = \frac{\sqrt{\pi}}{8}(12 - 12x^2 + x^4)e^{x^2/4}$, we get

$$J_2 = \int \int \int \frac{\sqrt{\pi}}{16} (12 - 12x^2 + x^4) e^{x^2/4} g(s)g(t) \frac{1}{2\pi} e^{-(x-s)^2/2} e^{-(x-t)^2/2} dx ds dt.$$

Since

$$\begin{aligned} & \int \frac{\sqrt{\pi}}{16} (12 - 12x^2 + x^4) e^{x^2/4} \frac{1}{2\pi} e^{-(x-s)^2/2} e^{-(x-t)^2/2} dx \\ &= \frac{1}{81\sqrt{3}} e^{\frac{1}{6}(-s^2+4st-t^2)} \left[27 + (s+t)^2(-18 + (s+t)^2) \right], \end{aligned}$$

we arrive at

$$J_2 = \int \int \frac{1}{81\sqrt{3}} e^{\frac{1}{6}(-s^2+4st-t^2)} \left[27 + (s+t)^2(-18 + (s+t)^2) \right] g(s)g(t) ds dt.$$

Thus $J = J_1 - J_2$ becomes

$$J = J(g) = \frac{1}{81\sqrt{3}} \int \int e^{\frac{1}{6}(-s^2+4st-t^2)} \left[27 + s^4 - 20s^3t - 126t^2 + 31t^4 + 6s^2(-3 + 7t^2) + s(180t - 68t^3) \right] g(s)g(t) ds dt.$$

The function

$$g(s) = \left(\frac{7280}{69} |s|^3 - \frac{11025}{23} s^2 + \frac{49000}{69} |s| - \frac{7875}{23} \right) \mathbf{1}_{[1,2]}(|s|)$$

is compactly supported and it satisfies $\int g = \int gx^2 = \int gx^4 = 0$. Numerical computations show that for this g we have $J(g) > 0.003$. However, this function is not smooth. To make it smooth it is enough to consider $g_\varepsilon = g * \frac{1}{\varepsilon} \psi(\cdot/\varepsilon)$ where ψ is smooth, compactly supported and integrates to 1. Then for any $\varepsilon > 0$ the function g_ε is smooth, compactly supported and satisfies $\int g_\varepsilon = \int g_\varepsilon x^2 = \int g_\varepsilon x^4 = 0$. To see this denote for simplicity $h = \frac{1}{\varepsilon} \psi(\cdot/\varepsilon)$ and observe that, e.g.,

$$\begin{aligned} \int g_\varepsilon(x) x^4 dx &= \int g(t) h(s) (s+t)^4 ds dt \\ &= \int g(t) h(s) (s^4 + 4s^3t + 6s^2t^2 + 4st^3 + t^4) dt ds = 0, \end{aligned}$$

since the integral with respect to t vanishes because of the properties of g . Taking $\varepsilon \rightarrow 0^+$, the corresponding functional $J(g_\varepsilon)$ converges to $J(g)$ due to the convergence of g_ε to g is L_1 and uniform boundedness of g_ε . As a consequence, for small $\varepsilon > 0$ we have $J(g_\varepsilon) > 0.001$. It suffices to pick one particular ε with this property. □

7.3 Entropy Power Inequalities Under Symmetries

The heart of the folklore proof of $h(S_2) \geq h(S_1)$ for symmetric distributions (see, e.g., [39]) is that for possibly dependent random variables X_1 and X_2 , the $SL(n, \mathbb{R})$ -

invariance of differential entropy combined with subadditivity imply that

$$\begin{aligned} h(X_1, X_2) &= h\left(\frac{X_1 + X_2}{\sqrt{2}}, \frac{X_1 - X_2}{\sqrt{2}}\right) \\ &\leq h\left(\frac{X_1 + X_2}{\sqrt{2}}\right) + h\left(\frac{X_1 - X_2}{\sqrt{2}}\right). \end{aligned}$$

If the distribution of (X_1, X_2) is the same as that of $(X_1, -X_2)$, we deduce that

$$h\left(\frac{X_1 + X_2}{\sqrt{2}}\right) \geq \frac{h(X_1, X_2)}{2}. \tag{7.17}$$

If, furthermore, X_1 and X_2 are i.i.d., then $h(X_1, X_2) = 2h(X_1)$, yielding $h(S_2) \geq h(S_1)$. Note that under the i.i.d. assumption, the requirement that the distributions of (X_1, X_2) and $(X_1, -X_2)$ coincide is equivalent to the requirement that X_1 (or X_2) has a symmetric distribution.

Without assuming symmetry but assuming independence, we can use the fact from [23] that $h(X - Y) \leq 3h(X + Y) - h(X) - h(Y)$ for independent random variables X, Y to deduce $\frac{1}{2}[h(X_1) + h(X_2)] \leq h\left(\frac{X_1 + X_2}{\sqrt{2}}\right) + \frac{1}{4} \log 2$. In the i.i.d. case, the improved bound $h(X - Y) \leq 2h(X + Y) - h(X)$ holds [29], which implies $h(X_1) \leq h\left(\frac{X_1 + X_2}{\sqrt{2}}\right) + \frac{1}{6} \log 2$. These bounds are, however, not particularly interesting since they are weaker than the classical entropy power inequality; if they had recovered it, these ideas would have represented by far its most elementary proof.

Hao and Jog [20] generalized the inequality (7.17) to the case where one has n random variables, under a natural n -variable extension of the distributional requirement, namely unconditionality. However, they used a proof that goes through Fisher information inequalities, similar to the original Stam proof of the full entropy power inequality. The main observation of this section is simply that under certain circumstances, one can give a direct and simple proof of the Hao–Jog inequality, as well as others like it, akin to the 2-line proof of the inequality (7.17) given above. The “certain circumstances” have to do with the existence of appropriate linear transformations that respect certain symmetries— specifically Hadamard matrices.

Let us first outline how this works in the real case. Suppose n is a dimension for which there exists a *Hadamard matrix*— namely, a $n \times n$ matrix with all its entries being 1 or -1 , and its rows forming an orthogonal set of vectors. Dividing each row by its length \sqrt{n} results in an orthogonal matrix O , all of whose entries are $\pm \frac{1}{\sqrt{n}}$. By unconditionality, each coordinate of the vector OX has the same distribution as $\frac{X_1 + \dots + X_n}{\sqrt{n}}$. Hence

$$h(X) = h(OX) \leq \sum_{j=1}^n h((OX)_j) = nh\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right),$$

where the inequality follows from subadditivity of entropy. This is exactly the Hao-Jog inequality for those dimensions where a Hadamard matrix exists. It would be interesting to find a way around the dimensional restriction, but we do not currently have a way of doing so.

As is well known, other than the dimensions 1 and 2, Hadamard matrices may only exist for dimensions that are multiples of 4. As of this date, Hadamard matrices are known to exist for all multiples of 4 up to 664 [22], and it is a major open problem whether they in fact exist for all multiples of 4. (Incidentally, we note that the question of existence of Hadamard matrices can actually be formulated in the entropy language. Indeed, Hadamard matrices are precisely those that saturate the obvious bound for the entropy of an orthogonal matrix [19].)

In contrast, complex Hadamard matrices exist in every dimension. A *complex Hadamard matrix* of order n is a $n \times n$ matrix with complex entries all of which have modulus 1, and whose rows form an orthogonal set of vectors in \mathbb{C}^n . To see that complex Hadamard matrices always exist, we merely exhibit the Fourier matrices, which are a well known example of them: these are defined by the entries $H_{j,k} = \exp\{\frac{2\pi i(j-1)(k-1)}{n}\}$, for $j, k = 1, \dots, n$, and are related to the discrete Fourier transform (DFT) matrices. Complex Hadamard matrices play an important role in quantum information theory [37]. They also yield Theorem 7.2.

Proof of Theorem 7.2 Take any $n \times n$ unitary matrix U which all entries are complex numbers of the same modulus $\frac{1}{\sqrt{n}}$; such matrices are easily constructed by multiplying a complex Hadamard matrix by $n^{-1/2}$. (For instance, one could take $U = \frac{1}{\sqrt{n}}[e^{2\pi ikl/n}]_{k,l}$.) By complex-unconditionality, each coordinate of the vector UX has the same distribution, the same as $\frac{X_1 + \dots + X_n}{\sqrt{n}}$. Therefore, by subadditivity,

$$h(X) = h(UX) \leq \sum_{j=1}^n h((UX)_j) = nh\left(\frac{X_1 + \dots + X_n}{\sqrt{n}}\right),$$

which finishes the proof. □

Let us mention that the invariance idea above also very simply yields the inequality

$$D(X) \leq \frac{1}{2}|h(X_1 + X_2) - h(X_1 - X_2)|,$$

where $D(X)$ denotes the relative entropy of the distribution of X from the closest Gaussian (which is the one with matching mean and covariance matrix), and X_1, X_2 are independent copies of a random vector X in \mathbb{R}^n . First observed in [30, Theorem 10], this fact quantifies the distance from Gaussianity of a random vector in terms of how different the entropies of the sum and difference of i.i.d. copies of it are.

Finally, we mention that the idea of considering two i.i.d. copies and using invariance (sometimes called the “doubling trick”) has been used in sophisticated

ways as a key tool to study both functional inequalities [5, 12, 24] and problems in network information theory (see, e.g., [14, 18]).

Acknowledgements We learned about the Edgeworth expansion used in our proof of Theorem 7.1 from S. Bobkov during the AIM workshop *Entropy power inequalities*. We are immensely grateful to him as well as the AIM and the organisers of the workshop which was a valuable and unique research experience.

We would like to thank the anonymous referee for his or her careful reading of the manuscript and suggesting several clarifications and improvements.

References

1. Aimpl: entropy power inequalities (2017). <http://aimpl.org/entropypower/>
2. S. Artstein, K.M. Ball, F. Barthe, A. Naor, Solution of Shannon’s problem on the monotonicity of entropy. *J. Am. Math. Soc.* **17**(4), 975–982 (2004)
3. K. Ball, P. Nayar, A reverse entropy power inequality for log-concave random vectors. *Stud. Math.* **235**(1), 17–30 (2016)
4. A.R. Barron, Entropy and the central limit theorem. *Ann. Probab.* **14**, 336–342 (1986)
5. F. Barthe, Optimal Young’s inequality and its converse: a simple proof. *Geom. Funct. Anal.* **8**(2), 234–242 (1998)
6. S.G. Bobkov, G.P. Chistyakov, Entropy power inequality for the Rényi entropy. *IEEE Trans. Inform. Theory* **61**(2), 708–714 (2015)
7. S. Bobkov, M. Madiman, Dimensional behaviour of entropy and information. *C. R. Acad. Sci. Paris Sér. I Math.* **349**, 201–204 (2011)
8. S.G. Bobkov, M. Madiman, L. Wang, Fractional generalizations of Young and Brunn-Minkowski inequalities, in *Concentration, functional inequalities and isoperimetry*, ed. by C. Houdré, M. Ledoux, E. Milman, M. Milman. Contemporary Mathematics, vol. 545 (American Mathematical Society, Providence, 2011), pp. 35–53
9. S.G. Bobkov, G.P. Chistyakov, F. Götze, Rate of convergence and Edgeworth-type expansion in the entropic central limit theorem. *Ann. Probab.* **41**(4), 2479–2512 (2013)
10. S. Bobkov, G.P. Chistyakov, F. Götze, Rényi divergence and the central limit theorem. *Ann. Probab.* **47**(1), 270–323 (2019). arXiv:1608.01805
11. L. Boltzmann, *Lectures on gas theory* (Dover Publications, Mineola, 1995). Reprint of the 1896–1898 edition (Translated by S. G. Brush)
12. E.A. Carlen, Superadditivity of Fisher’s information and logarithmic Sobolev inequalities. *J. Funct. Anal.* **101**(1), 194–211 (1991)
13. T.A. Courtade, Bounds on the Poincaré constant for convolution measures (preprint, 2018). arXiv:1807.00027
14. T.A. Courtade, A strong entropy power inequality. *IEEE Trans. Inform. Theory* **64**(4, part 1), 2173–2192 (2018)
15. A. Eskenazis, P. Nayar, T. Tkocz, Gaussian mixtures: entropy and geometric inequalities. *Ann. Probab.* **46**(5), 2908–2945 (2018)
16. M. Fradelizi, M. Madiman, A. Marsiglietti, A. Zvavitch, Do Minkowski averages get progressively more convex? *C. R. Acad. Sci. Paris Sér. I Math.* **354**(2), 185–189 (2016)
17. M. Fradelizi, M. Madiman, A. Marsiglietti, A. Zvavitch, The convexifying effect of Minkowski summation. *EMS Surv. Math. Sci.* **5**(1/2), 1–64 (2019). arXiv:1704.05486
18. Y. Geng, C. Nair, The capacity region of the two-receiver Gaussian vector broadcast channel with private and common messages. *IEEE Trans. Inform. Theory* **60**(4), 2087–2104 (2014)
19. H. Gopalkrishna Gadiyar, K.M. Sangeeta Maini, R. Padma, H.S. Sharatchandra, Entropy and Hadamard matrices. *J. Phys. A* **36**(7), L109–L112 (2003)

20. J. Hao, V. Jog, An entropy inequality for symmetric random variables, in Proceedings of IEEE International Symposium on Information Theory, IEEE (2018), pp. 1600–1604. arXiv:1801.03868.
21. O. Johnson, *Information theory and the central limit theorem* (Imperial College Press, London, 2004)
22. H. Kharaghani, B. Tayfeh-Rezaie, A Hadamard matrix of order 428. *J. Combin. Des.* **13**(6), 435–440 (2005)
23. I. Kontoyiannis, M. Madiman, Sumset and inverse sumset inequalities for differential entropy and mutual information. *IEEE Trans. Inform. Theory* **60**(8), 4503–4514 (2014)
24. E.H. Lieb, Gaussian kernels have only Gaussian maximizers. *Invent. Math.* **102**(1), 179–208 (1990)
25. M. Madiman, A primer on entropic limit theorems (preprint, 2017)
26. M. Madiman, A.R. Barron, The monotonicity of information in the central limit theorem and entropy power inequalities, in *Proceedings of IEEE International Symposium on Information Theory*, Seattle (2006), pp. 1021–1025
27. M. Madiman, A.R. Barron, Generalized entropy power inequalities and monotonicity properties of information. *IEEE Trans. Inform. Theory* **53**(7), 2317–2329 (2007)
28. M. Madiman, F. Ghassemi, Combinatorial entropy power inequalities: a preliminary study of the Stam region. *IEEE Trans. Inform. Theory* **65**(3), 1375–1386 (2019). arXiv:1704.01177
29. M. Madiman, I. Kontoyiannis, The entropies of the sum and the difference of two IID random variables are not too different, in *Proceedings of IEEE International Symposium on Information Theory*, Austin (2010)
30. M. Madiman, I. Kontoyiannis, Entropy bounds on Abelian groups and the Ruzsa divergence. *IEEE Trans. Inform. Theory* **64**(1), 77–92 (2018). arXiv:1508.04089
31. M. Madiman, J. Melbourne, P. Xu, Forward and reverse entropy power inequalities in convex geometry, in *Convexity and concentration*, ed. by E. Carlen, M. Madiman, E.M. Werner. IMA Volumes in Mathematics and its Applications, vol. 161 (Springer, Berlin, 2017), pp. 427–485. arXiv:1604.04225
32. M. Madiman, L. Wang, J.O. Woo, Majorization and Rényi entropy inequalities via Sperner theory *Discrete Math.* **342**(10), 2911–2923, (2019). arXiv:1712.00913
33. M. Madiman, L. Wang, J.O. Woo, Rényi entropy inequalities for sums in prime cyclic groups. (preprint, 2017). arXiv:1710.00812
34. C.E. Shannon, A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423, 623–656 (1948)
35. D. Shlyakhtenko, A free analogue of Shannon’s problem on monotonicity of entropy. *Adv. Math.* **208**(2), 824–833 (2007)
36. A.J. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Inform. Control* **2**, 101–112 (1959)
37. W. Tadej, K. Życzkowski, A concise guide to complex Hadamard matrices. *Open Syst. Inf. Dyn.* **13**(2), 133–177 (2006)
38. A.M. Tulino, S. Verdú, Monotonic decrease of the non-Gaussianness of the sum of independent random variables: a simple proof. *IEEE Trans. Inform. Theory* **52**(9), 4295–4297 (2006)
39. L. Wang, M. Madiman, Beyond the entropy power inequality, via rearrangements. *IEEE Trans. Inform. Theory* **60**(9), 5116–5137 (2014)
40. L. Wang, J.O. Woo, M. Madiman, A lower bound on the Rényi entropy of convolutions in the integers, in *IEEE International Symposium on Information Theory*, Honolulu (2014), pp. 2829–2833
41. Y. Yu, Letter to the editor: on an inequality of Karlin and Rinott concerning weighted sums of i.i.d. random variables. *Adv. Appl. Probab.* **40**(4), 1223–1226 (2008)