REMARKS ON THE RÉNYI ENTROPY OF A SUM OF IID RANDOM VARIABLES

BENJAMIN JAYE, GALYNA V. LIVSHYTS, GRIGORIS PAOURIS, AND PETER PIVOVAROV

Abstract. In this note we study a conjecture of Madiman and Wang [MW] which predicted that the generalized Gaussian distribution minimizes the Rényi entropy of the sum of independent random variables. Through a variational analysis, we show that the generalized Gaussian fails to be a minimizer for the problem.

1. Introduction

For $p > 1$, the p-Rényi entropy [Re] of a (continuous) random variable X in $\mathbb{R}^d$ distributed with density f is defined by

$$h_p(X) = -\frac{1}{p-1} \log \int_{\mathbb{R}^d} f^p dm_d,$$

where m_d denotes the d-dimensional Lebesgue measure. As $p \to 1^+$, $h_p(X)$ converges to the Shannon entropy $h(X) = -\int_{\mathbb{R}^d} f \log f dm_d$, provided that the density of X is regular enough to justify passage of the limit. See Principe [Pr] for more information about where the Rényi entropy arises; see also Bobkov, Marsiglietti [BM] for a related discussion.

In this note we make some elementary remarks on the following basic mathematical question: *Over all random variables X with $h_p(X)$ some fixed quantity, what are the minimizers of the entropy $h_p(X + X')$, where X' is an independent copy of X?*

We learnt about this question from the papers of Madiman, Melbourne, Xu, and Wang [MW, MMX], who studied unifying entropy power inequalities for the Rényi entropy, which, in the limit $p \to 1^+$ recover the statement that, over all probability distributions with $h(X)$ fixed, $h(X + X')$ is minimized if (and only if) X is a Gaussian, see e.g. [DCT].

Following [LYZ, MW, MMX], for $\beta > 0$, consider the *Generalized Gaussian*

$$G_\beta(x) = \alpha(1 - \beta|x|^2)_+^{1/(p-1)},$$

where α is chosen so that $\int_{\mathbb{R}^d} G_\beta dm_d = 1$. The generalized Gaussian is the distribution with the smallest p-th moment with a given Rényi entropy, see [LYZ]. Madiman and Wang conjectured (Conjecture IV.3 in [MW]) that if X_j, $j = 1, \dots, n$, are independent random variables with densities f_j, and Z_j are independent random variables distributed with respect to G_{β_j} where β_j is chosen so that

$$h_p(X_j) = h_p(Z_j),$$

then $h_p(X_1 + \cdots + X_n) \geq h_p(Z_1 + \cdots + Z_n)$.

In this note we will show that unfortunately this conjecture does not hold in the special case when $d = 1$, $p = 2$, $n = 2$ and X_1 and X_2 are identically distributed, see Section 4. However, we do suspect that a minimizing distribution is a relatively small perturbation of the generalized Gaussian.

Throughout this note we only consider the case where $X_1, \dots, X_n$ are independent copies of a random variable X with density f. The question of finding the minimizer of $h_p(X_1 + \dots X_n)$ with $h_p(X)$ fixed can then be rephrased as a constrained maximization problem, which we introduce in Section 2. Subsequently, in Section 3 we take the first variation of this maximization problem. We have not been able to develop a satisfactory theory of the associated Euler-Lagrange equation (3.1), but we show in Section 4 that the generalized Gaussian is not a solution to (3.1), and so fails to be a maximizer of the extremal problem. We conclude the paper with some elementary remarks and speculation.

Acknowledgement. The first named author is supported by NSF DMS-1830128, DMS-1800015 and NSF CAREER DMS-1847301. The second named author is supported by the NSF CAREER DMS-1753260. The third named author is supported by the NSF DMS-1812240. The fourth named author is supported by the NSF DMS-1612936. The work was partially supported by the National Science Foundation under Grant No. DMS-1440140 while the authors were in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Fall 2017 semester.

2. The constrained maximization problem

Denote by $\mathcal{C}_n(f)$ the $(n-1)$-fold convolution of a given function f with itself, that is, $\mathcal{C}_n(f) = f * f * \cdots * f$, where there are n factors of f (and $n-1$ convolutions). Then $\mathcal{C}_1(f) = f$, and it will be convenient to set $\mathcal{C}_0(f) = \delta_0$, the Dirac delta measure.

Throughout the text, we fix $M > 0$, $n \in \mathbb{N}$ and $p \in (1, \infty)$. We set

$$\mathcal{F} = \left\{ f \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d),\ f \geq 0,\ \|f\|_p^p = M,\ \|f\|_1 = 1 \right\}$$

and consider the extremal problem

$$\begin{cases} \text{Maximize } \mathcal{I}(f) \stackrel{\text{def}}{=} \int_{\mathbb{R}^d} \mathcal{C}_n(f)^p dm_d \\ \text{subject to } f \in \mathcal{F}. \end{cases} \tag{2.1}$$

Put

$$\Lambda = \Lambda(p, M) = \sup\{\mathcal{I}(f) : f \in \mathcal{F}\}. \tag{2.2}$$

We begin with a simple scaling lemma, which we will use often in what follows.

Lemma 2.1. *Suppose that $f \in L^1(\mathbb{R}^d) \cap L^p(\mathbb{R}^d)$ is non-negative, and $\|f\|_1 > 0$. The function*

$$\widetilde{f} = \frac{1}{\lambda^d \|f\|_1} f\Big(\frac{\cdot}{\lambda}\Big), \ \textit{with } \lambda = \Big(\frac{\|f\|_p^p}{M\|f\|_1^p}\Big)^{\frac{1}{d(p-1)}},$$

belongs to $\mathcal{F}$, and

$$\mathcal{I}(\widetilde{f}) = \frac{M}{\|f\|_p^p} \frac{1}{\|f\|_1^{p(n-1)}} \mathcal{I}(f).$$

Proof. Observe that, for any $r \in [1, \infty)$,

$$\|\widetilde{f}\|_r^r = \frac{1}{\lambda^{d(r-1)} \|f\|_1^r} \|f\|_r^r.$$

Plugging in $r = 1$ and $r = p$ (and recalling the definition of λ) we see that $\widetilde{f} \in \mathcal{F}$. Next, observe that

$$\mathcal{C}_n(\widetilde{f})(x) = \frac{1}{\|f\|_1^n \lambda^d} \mathcal{C}_n(f)\big(\frac{x}{\lambda}\big) \text{ for any } x \in \mathbb{R}^d.$$

Whence,

$$\mathcal{I}(\widetilde{f}) = \frac{1}{\lambda^{d(p-1)} \|f\|_1^{pn}} \mathcal{I}(f),$$

and the proof is complete by recalling the definition of λ. □

We next prove that (2.1) has a maximizer. A radial function f on $\mathbb{R}^d$ is called decreasing if $f(y) \leq f(x)$ whenever $|y| \geq |x|$.

Proposition 2.2. *The problem (2.1) has a lower-semicontinuous, radially decreasing, maximizer Q.*

Proof. We begin with two observations.

(1) Repeated application of Young's convolution inequality [LL] yields that, with $p' = p/(p-1)$,
$$\mathcal{I}(f) \le \|f\|_{(np')'}^{np},$$
where $(np')'$ is the Hölder conjugate of np'. Since $n > 1$, we have that $(np')' \in (1, p)$.

(2) By iterating Riesz's rearrangement inequality [LL, Theorem 3.7] we have that $\mathcal{I}(f) \le \mathcal{I}(f^*)$, where f^* is the symmetric rearrangement of f; see [B, Section 3.4] for related multiple convolution rearrangement inequalities and their equality cases.

Now take non-negative functions $f_j \in \mathcal{F}$ such that $\Lambda = \lim_{j\to\infty} \mathcal{I}(f_j)$ (recall Λ from (2.2)). From the second observation we may assume that f_j are radial and decreasing. Passing to a subsequence if necessary, we may in addition assume that $f_j \to f$ weakly in $L^p(\mathbb{R}^d)$. Consequently, f is radial, decreasing, $f \ge 0$, and $\|f\|_p^p \le M$. By modifying f on a set of measure zero if necessary, we may assume that f is lower semicontinuous[1].

Our primary goal will be to show that $f_j \to f$ strongly in $L^{(np')'}$ as $j \to \infty$. From this the first observation above would yield that $\mathcal{I}(f) = \Lambda$.

Claim 2.3. As $j \to \infty$, $f_j \to f$ m_d-almost everywhere.

Proof. For $r > 0$, define $v_j(r) = f_j(x)$ whenever $|x| = r$. Then since f_j converges weakly to f in $L^p(\mathbb{R}^d)$, we have that whenever I is a closed interval of finite measure in $(0, \infty)$,
$$\lim_{j\to\infty} \int_I v_j dm_1 = \int_I v dm_1.$$
Recall that almost every point r of a function $v \in L^1_{\text{loc}}((0,\infty))$ is a *Lebesgue point*, that is,
$$v(r) = \lim_{|I|\to 0, r\in I} \frac{1}{|I|} \int_I v dm_1,$$
where the limit is taken over any sequence of intervals I containing r that shrink to r (not necessarily centered at r). But then if $r > 0$ is a Lebesgue point, and $I_k = [r - 2^{-k}, r]$, then
$$v(r) = \lim_{k\to\infty} \frac{1}{2^{-k}} \int_{I_k} v dm_1 = \lim_{k\to\infty} \lim_{j\to\infty} \frac{1}{2^{-k}} \int_{I_k} v_j dm_1$$

[1] If f is discontinuous at $x \in \mathbb{R}^d$, then define $f(x) = \sup_{|y|>|x|} f(y)$ (i.e. the one-sided radial limit from the right). Then $\{f > \lambda\}$ is open for every $\lambda > 0$.

but since v_j is decreasing we have that $v_j \geq v_j(r)$ on I_k. Thus

$$v(r) \geq \limsup_{j\to\infty} v_j(r).$$

Arguing similarly with intervals whose left end-point is r, we also have that

$$v(r) \leq \liminf_{j\to\infty} v_j(r).$$

Thus $\lim_{j\to\infty} v_j = v$ at every Lebesgue point. From this we readily deduce the claim, since if E is a Lebesgue null set in $(0,\infty)$, then $E \times \mathbb{S}^{d-1}$ is a Lebesgue null set in $\mathbb{R}^d$. □

Notice that, as a consequence of this claim, Fatou's Lemma ensures that $\|f\|_1 \leq 1$. Our next claim is

Claim 2.4. If $1 < q < p$, then $f_j \to f$ strongly in $L^q(\mathbb{R}^d)$ as $j \to \infty$.

The proof of this claim is a variant of the Vitali convergence theorem, but observe that it does not necessarily hold if one was to remove the radially decreasing property of the functions f_j (just consider a sequence of translates of a fixed function).

Proof. Fix $\varepsilon > 0, \delta > 0$. Insofar as the functions f_j and f are radially decreasing,

$$\bigcup_j \{|f_j| \geq \tfrac{\delta}{2}\} \cup \{|f| \geq \tfrac{\delta}{2}\} \subset B,$$

where B is the closed ball centred at 0 of radius $\big(\frac{2}{m_d(B(0,1))\delta}\big)^{1/d}$. (Otherwise we would have $\|f_j\|_1 > 1$ for some j, or $\|f\|_1 > 1$.)

On $\mathbb{R}^d\backslash B$, we have $|f_j| < \delta/2$ for every j, and $|f| < \delta/2$, whence

$$\int_{\mathbb{R}^d\backslash B} |f_j - f|^q dm_d \leq \delta^{q-1}\Big(\|f_j\|_1 + \|f\|_1\Big) \leq 2\delta^{q-1} < \frac{\varepsilon}{3}$$

provided $\delta > 0$ is chosen sufficiently small.

Now fix $\varkappa > 0$. Observe that,

$$\int_{B\cap\{|f_j - f| < \varkappa\}} |f_j - f|^q dm_d \leq m_d(B)\varkappa^q < \frac{\varepsilon}{3}$$

if $\varkappa$ is chosen sufficiently small. On the other hand, inasmuch as B has finite measure, we have that $f_j \to f$ in measure on B as $j \to \infty$. From the inequalities

$$\int_{B\cap\{|f_j - f| \geq \varkappa\}} |f_n - f|^q dm_d \leq m_d(B \cap \{|f_j - f| \geq \varkappa\})^{1-q/p} \|f_j - f\|_p^q$$
$$\leq 2^q M^{q/p} m_d(B \cap \{|f_j - f| \geq \varkappa\})^{1-p/q},$$

we infer that there exists $N \in \mathbb{N}$ such that

$$\int_{B\cap\{|f_j-f|\geq\varkappa\}} |f_j - f|^q dm_d < \frac{\varepsilon}{3} \text{ for all } j \geq N.$$

Bringing these estimates together, it follows that $\|f_j - f\|_q^q < \varepsilon$ for every $j \geq N$. □

In particular, since $(np')' \in (1, p)$, this second claim ensures that $f_j \to f$ in $L^{(np')'}$ as $j \to \infty$. Thus by the remarks preceding Claim 2.3, we have $\mathcal{I}(f) = \Lambda$ (so f is not identically zero). It remains to show that $f \in \mathcal{F}$. To this end, we apply Lemma 2.1: Consider the function

$$\widetilde{f} = \frac{1}{\|f\|_1\lambda^d} f\Big(\frac{\cdot}{\lambda}\Big), \text{ with } \lambda = \Big(\frac{\|f\|_p^p}{M\|f\|_1^p}\Big)^{\frac{1}{d(p-1)}}.$$

Then $\widetilde{f} \in \mathcal{F}$ and $\mathcal{I}(\widetilde{f}) = \frac{M}{\|f\|_p^p}\frac{1}{\|f\|_1^{p(n-1)}}\Lambda$. Consequently, if $\|f\|_p^p < M$ or $\|f\|_1 < 1$, then $\mathcal{I}(\widetilde{f}) > \Lambda$, which is absurd. Thus $f \in \mathcal{F}$ and the proof of the proposition is complete. □

3. The First Variation

With the existence of a maximizer proved, we now wish to analyze it analytically. We shall derive the following criterion.

Proposition 3.1. *A radial non-negative lower-semicontinuous function Q is a maximizer of the problem (2.1) if and only if*

$$\mathcal{C}_{n-1}(Q) * [\mathcal{C}_n(Q)]^{p-1} = \frac{\Lambda}{Mn} Q^{p-1} + \frac{\Lambda(n-1)}{n} \text{ on } \{Q > 0\}. \tag{3.1}$$

Proof. The sufficiency is easy to show. Integrating both sides of (3.1) against Q, and recalling that $Q \in \mathcal{F}$, yields

$$\int_{\mathbb{R}^d} Q \cdot (\mathcal{C}_{n-1}(Q) * [\mathcal{C}_n(Q)]^{p-1}) dm_d = \Lambda.$$

But using Tonelli's theorem, the left hand side equals $\mathcal{I}(Q)$. This is just the fact that, for even functions f, g and h, $\int f(g * h)dm_d = \int (f * g)h \, dm_d$.

Conversely, consider a bounded function φ compactly supported in the open set $\{Q > 0\}$. Since Q is lower-semicontinuous, $\inf_{\mathrm{supp}(\varphi)} Q > 0$. Therefore, (insofar as φ is bounded) there exists a constant $C > 0$ such that

$$|\varphi| \leq CQ \text{ on } \mathbb{R}^d, \tag{3.2}$$

so in particular, there exists $t_0 > 0$ such that for $|t| \le t_0$ it follows that $Q_t \overset{\text{def}}{=} Q + t\varphi$ is non-negative. In the notation of Lemma 2.1 with $f = Q_t$, we consider the function

$$\widetilde{Q}_t = \frac{1}{\lambda^d} \frac{(Q + t\varphi)\left(\frac{\cdot}{\lambda}\right)}{\|Q + t\varphi\|_1},$$

with the corresponding $\lambda > 0$ satisfying $\|\widetilde{Q}_t\|_p^p = \|Q\|_p^p = M$. Of course we also have $\int_{\mathbb{R}} \widetilde{Q}_t \, dm_d = 1$ regardless of λ for $|t| < t_0$. We conclude that $\widetilde{Q}_t$ belongs to $\mathcal{F}$ – and therefore $\mathcal{I}(\widetilde{Q}_t) \le \mathcal{I}(Q) = \Lambda$ – for all $|t| < t_0$. Moreover, as in Lemma 2.1,

$$\mathcal{I}(\widetilde{Q}_t) = \frac{1}{\lambda^{d(p-1)} \|Q + t\varphi\|_1^{np}} \int_{\mathbb{R}^d} \mathcal{C}_n(Q + t\varphi)^p dm_d. \tag{3.3}$$

For $|t| < t_0$, we calculate, using commutativity and associativity of the convolution operator,

$$\frac{d}{dt}\mathcal{C}_n(Q + t\varphi)^p = pn[\varphi * \mathcal{C}_{n-1}(Q + t\varphi)][\mathcal{C}_n(Q + t\varphi)]^{p-1},$$

and

$$\begin{aligned} \frac{d^2}{dt^2}\mathcal{C}_n(Q+t\varphi)^p = {} & pn(n-1)\varphi * \varphi * \mathcal{C}_{n-2}(Q + t\varphi)[\mathcal{C}_n(Q + t\varphi)]^{p-1} \\ & + n^2 p(p-1)[\varphi * \mathcal{C}_{n-1}(Q + t\varphi)]^2[\mathcal{C}_n(Q + t\varphi)]^{p-2}. \end{aligned} \tag{3.4}$$

Crudely employing the bound (3.2) in (3.4), we infer that there is a constant $C > 0$ such that for all $|t| < t_0$,

$$\left| \frac{d^2}{dt^2}\mathcal{C}_n(Q + t\varphi)^p \right| \le C\mathcal{C}_n(Q)^p.$$

Whence, the second order Taylor formula yields that

$$|\mathcal{C}_n(Q+t\varphi)^p - \mathcal{C}_n(Q)^p - npt[\varphi * \mathcal{C}_{n-1}(Q)][\mathcal{C}_n(Q)]^{p-1}| \le Ct^2\mathcal{C}_n(Q)^p, \tag{3.5}$$

for $|t| < t_0$. Integrating the pointwise inequality (3.5) yields

$$\int_{\mathbb{R}^d} \mathcal{C}_n(Q + t\varphi)^p dm_d = \Lambda + npt \int_{\mathbb{R}^d} [\varphi * \mathcal{C}_{n-1}(Q)][\mathcal{C}_n(Q)]^{p-1} dm_d + O(t^2) \tag{3.6}$$

as $t \to 0$.

Now, recalling the definition of λ, we calculate

$$\lambda^{d(p-1)}\|Q+t\varphi\|_1^{np} = \frac{\|Q+t\varphi\|_p^p}{M}\|Q+t\varphi\|_1^{(n-1)p}$$
$$= \Big(1+\frac{pt}{M}\int_{\mathbb{R}^d}\varphi Q^{p-1}dm_d + O(t^2)\Big)\Big(1+t(n-1)p\int \varphi dm_d + O(t^2)\Big), \tag{3.7}$$

where in the expansion of $\|Q+t\varphi\|_p^p$ we have again used the inequality (3.2) to obtain the $O(t^2)$ term.

Plugging the two expansions (3.7) and (3.6) into (3.3) yields that, as $t\to 0$,

$$\mathcal{I}(\widetilde{Q}_t) = \Lambda + pt\Big\{n\int_{\mathbb{R}^d}[\varphi * \mathcal{C}_{n-1}(Q)][\mathcal{C}_n(Q)]^{p-1}dm_d - \frac{\Lambda}{M}\int_{\mathbb{R}^d}\varphi Q^{p-1}dm_d - (n-1)\Lambda\int \varphi\, dm_d\Big\} + O(t^2).$$

Since $\lim_{t\to 0}\frac{\mathcal{I}(\widetilde{Q}_t)-\mathcal{I}(Q)}{t} = 0$, the second term in the above expansion must vanish. Therefore, we get that

$$\int_{\mathbb{R}^d}\varphi\Big\{n[\mathcal{C}_{n-1}(Q)] * [\mathcal{C}_n(Q)]^{p-1} - \frac{\Lambda}{M}Q^{p-1} - (n-1)\Lambda\Big\}dm_d = 0.$$

Since φ was any bounded function compactly supported in $\{Q>0\}$, we conclude that (3.1) holds. $\square$

4. On the Madiman-Wang conjecture

Proposition 4.1. *The generalized Gaussian is not necessarily the extremizer for problem (2.1).*

Proof. Consider the simplest case $d=1$, $p=2$, and $n=2$. We shall show that the function $G(x)=\alpha(1-|x|^2)_+$ does not satisfy the equation

$$\mathcal{C}_3(f) = af+b \text{ on } [-1,1] \text{ with } a,b>0, \tag{4.1}$$

and so no function of the form $\frac{c}{\lambda}G(\frac{\cdot}{\lambda})$, with $c,\lambda>0$, satisfies (3.1), for any value of Λ. In fact, we shall show that $\mathcal{C}_3(G) = G*G*G$ is not a quadratic polynomial near 0.

For this, observe:

$$G'' = 2\alpha(\delta_{-1} - \chi_{[-1,1]} + \delta_1).$$

Thus, $(G*G*G)'''''' = (G''*G''*G'')$ is the threefold convolution of the above measure. The threefold convolution of $-2\chi_{[-1,1]}$ equals $-8(3-|x|^2)_+$, and no other term in the convolution $G''*G''*G''$ is quadratic in $|x|$. Therefore, $G*G*G$ has non-vanishing sixth derivative at 0, but $a+bG$ does have vanishing sixth derivative at 0. □

5. Any radially decreasing solution of (3.1) is compactly supported

In this section, we discuss the following

Proposition 5.1. *Decreasing radial solutions of (3.1) are compactly supported.*

Proof. Suppose that $\int_{\mathbb{R}^d} Q\, dm_d = 1$ and Q is not compactly supported. Since Q is non-negative and radially decreasing, its support is $\mathbb{R}^d$.

Since $\|\mathcal{C}_{n-1}(Q) * (\mathcal{C}_n(Q))^{p-1}\|_1 < \infty$ and Q (along with any multiple convolution of Q) is radially decreasing, we have that $\lim_{|x|\to\infty} |\mathcal{C}_{n-1}(Q) * (\mathcal{C}_n(Q))^{p-1}(x)| = 0$. Since also $Q^{p-1}(x) \to 0$ as $|x| \to \infty$, we have that $\Lambda = 0$ in the equation (3.1). But, on the other hand, $\Lambda > 0$. □

6. Remarks

In this section we make some remarks that suggest that although the generalized Gaussian is not an optimal distribution for the problem (2.1), a reasonably small perturbation of the generalized Gaussian could well be.

Beginning with $f_0(x) = \mathbf{1}_{[-1,1]}$, consider the following iteration for $j \geq 1$

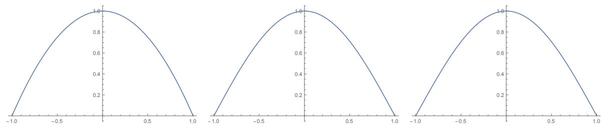
$$f_j(x) = \frac{\mathcal{C}_3(f_{j-1})(x) - \mathcal{C}_3(f_{j-1})(1)}{\mathcal{C}_3(f_{j-1})(0) - \mathcal{C}_3(f_{j-1})(1)}.$$

Numerically, this iteration converges pointwise to a solution of the equation (4.1) for some $a, b > 0$ satisfying the constraints $f(0) = 1$ and $f(1) = 0$ (so the support of f is $[-1,1]$). The resulting function f can then be re-scaled via the transformation $\frac{c}{\lambda} f(\frac{\cdot}{\lambda})$ $(c, \lambda > 0)$ to have any given positive integral and L^2-norm. We do not know if the solution of $\mathcal{C}_3(f) = af + b$ is unique (modulo natural invariants in the problem), so we cannot say that this function f corresponds to a solution of the constrained maximization problem (2.1).

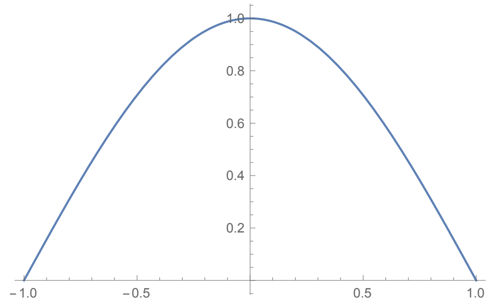
We provide the graphs of f_1, f_2, f_3 and f_4 (see Figure 1 below), and the algebraic expressions for f_1, f_2 and f_3 on $[-1,1]$.

$$f_1(x) = 1 - x^2,\ f_2(x) = 1 - \frac{6x^2}{5} + \frac{x^4}{5}$$

$$f_3(x) = 1 - \frac{62325x^2}{50521} + \frac{12810x^4}{50521} - \frac{1050x^6}{50521} + \frac{45x^8}{50521} - \frac{x^{10}}{50521}.$$



(A) f_1. (B) f_2. (C) f_3.



(D) f_4.

FIGURE 1. The graphs of $f_1, \ldots, f_4$ on $[-1, 1]$.

References

[BM] S. Bobkov, A. Marsiglietti, *Asymptotic behavior of Rényi entropy in the central limit theorem,* submitted, arXiv:1802.10212.

[B] A. Burchard, *Cases of equality in the Riesz rearrangement inequality.* Thesis (Ph.D.)Georgia Institute of Technology. 1994. 94 pp.

[DCT] A. Dembo, T. M. Cover and J. A. Thomas, *Information theoretic inequalities,* IEEE Transactions on Information Theory, vol. 37, no. 6, pp. 1501–1518, Nov. 1991.

[LL] E. Lieb and M. Loss *Analysis, 2nd Edition*, AMS Graduate Studies in Mathematics, 2002.

[LYZ] E. Lutwak, D. Yang and G. Zhang, *Moment-Entropy Inequalities for a Random Vector,* IEEE Transactions on Information Theory, vol. 53, no. 4, pp. 1603–1607, April 2007.

[MW] M. Madiman and L. Wang, *Beyond the entropy power inequality, via rearrangements.* IEEE Transactions on Information Theory, vol. 60, no. 9, pp. 5116-5137, Sept. 2014.

[MMX] M. Madiman, J. Melbourne and P. Xu *Forward and reverse entropy power inequalities in convex geometry.* Convexity and Concentration, E. Carlen, M. Madiman, E. Werner (eds.), The IMA Volumes in Mathematics and its Applications, Vol. 161, pp. 427–485, Springer, 2017.

[Pr] J. C. Principe, *Information theoretic learning,* Springer-Verlag New York 2010.

[Re] A. Rényi. *On Measures of Entropy and Information.* Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics, 547–561, University of California Press, Berkeley, Calif., 1961.

School of Mathematical Sciences, Clemson University
E-mail address: bjaye@clemson.edu

Department of Mathematics, Georgia Tech
E-mail address: glivshyts6@math.gatech.edu

Department of Mathematics, Texas A&M
E-mail address: grigoris@math.tamu.edu

Department of Mathematics, University of Missouri
E-mail address: pivovarovp@missouri.edu