The entropy per coordinate of a random vector is highly constrained under convexity conditions

Sergey Bobkov and Mokshay Madiman, *Member, IEEE*

***Abstract*—The entropy per coordinate in a log-concave random vector of any dimension with given density at the mode is shown to have a range of just 1. Uniform distributions on convex bodies are at the lower end of this range, the distribution with i.i.d. exponentially distributed coordinates is at the upper end, and the normal is exactly in the middle. Thus in terms of the amount of randomness as measured by entropy per coordinate, any log-concave random vector of any dimension contains randomness that differs from that in the normal random variable with the same maximal density value by at most 1/2. As applications, we obtain an information-theoretic formulation of the famous hyperplane conjecture in convex geometry, entropy bounds for certain infinitely divisible distributions, and quantitative estimates for the behavior of the density at the mode on convolution. More generally, one may consider so-called convex or hyperbolic probability measures on Euclidean spaces; we give new constraints on entropy per coordinate for this class of measures, which generalize our results under the log-concavity assumption, expose the extremal role of multivariate Pareto-type distributions, and give some applications.**

***Index Terms*—Maximum entropy; log-concave; slicing problem; inequalities; convex measures.**

I. Introduction

A probability density function (or simply "density") f defined on the linear space $\mathbb{R}^n$ is said to be log-concave if

$$f(\alpha x + (1-\alpha)y) \geq f(x)^\alpha f(y)^{1-\alpha}, \tag{1}$$

for each $x, y \in \mathbb{R}^n$ and each $0 \leq \alpha \leq 1$. If f is log-concave, we will also use the adjective "log-concave" for a random variable X distributed according to f, and for the probability measure induced by it. (For discussion of the justification for such terminology, see the beginning of Section VI.) Given a random vector $X = (X_1, \dots, X_n)$ in $\mathbb{R}^n$ with density $f(x)$, introduce the entropy functional

$$h(f) = -\int_{\mathbb{R}^n} f(x) \log f(x)\, dx,$$

provided that the integral exists in the Lebesgue sense; as usual, we also denote this $h(X)$. Our main contribution in this paper is the observation that when viewed appropriately, every log-concave random vector has approximately the same entropy per coordinate as a related Gaussian vector.

The authors are listed in alphabetical order of last names.

School of Mathematics, University of Minnesota, 127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455 USA. Email: bobkov@math.umn.edu

Department of Statistics, Yale University, 24 Hillhouse Avenue, New Haven, CT 06511, USA. Email: mokshay.madiman@yale.edu

SB was supported by NSF grant DMS-0706866, and MM was supported by a Junior Faculty Fellowship from Yale University.

Log concavity has been deeply studied in probability, statistics, optimization and geometry, and there are a number of results that show that log-concave random vectors resemble Gaussian random vectors. For instance, several functional inequalities that hold for Gaussians also hold for appropriate subclasses of log-concave distributions (see, e.g., [5], [15], [4] for discussion of Poincare and logarithmic Sobolev inequalities for log-concave measures). Observe that this is not at all obvious at first glance– log-concave probability measures include a large variety of distributions including the uniform distribution on any compact, convex set, the (one-sided) exponential distribution, and of course any Gaussian. In this note, we give a strong (quantitative) information-theoretic basis to the intuition that log-concave distributions resemble Gaussian distributions.

To motivate our main results, we first observe that for (one-dimensional) log-concave random variables X,

$$h(X) \approx \log \sigma, \tag{2}$$

where σ is the standard deviation of X. (An exact result to this effect is contained in Proposition II.1 and proved in Section II.) An upper bound for entropy in terms of standard deviation clearly follows from the maximum entropy property of the Gaussian; so it is the lower bound that is not obvious here. Thus the property (2) may be viewed as asserting comparability between the entropy of a one-dimensional log-concave density and that of a Gaussian density with the same standard deviation.

Our main purpose in this note is to describe a way to capture the spirit of the statement (2) in the setting of (multidimensional) random vectors. To describe this extension, recall that the L^∞ norm of a measurable function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as its essential supremum with respect to Lebesgue measure, $\|f\|_\infty = \operatorname{ess\,sup}_x f(x)$. Throughout this paper, we will write $\|f\| = \|f\|_\infty$ for brevity. Any log-concave f is continuous and bounded on the supporting set $\Omega = \{x : f(x) > 0\}$, so we can simply write $\|f\| = \max_{x \in \Omega} f(x)$.

Theorem I.1. *If a random vector X in $\mathbb{R}^n$ has a log-concave density f, let Z in $\mathbb{R}^n$ be any normally distributed random vector with maximum density being the same as that of X. Then*

$$\frac{1}{n} h(Z) - \tfrac{1}{2} \leq \frac{1}{n} h(X) \leq \frac{1}{n} h(Z) + \tfrac{1}{2}.$$

Equality holds in the lower bound if and only if X is uniformly distributed on a convex set with non-empty interior. Equality holds in the upper bound if X has coordinates that are i.i.d. exponentially distributed.

The observation that it is useful to consider Gaussian comparisons by matching $\|f\|$ rather than the first two moments may be considered the key observation of this paper. Theorem I.1 follows easily from the following basic proposition (both are proved in Section IV).

Proposition I.2. *If a random vector X in $\mathbb{R}^n$ has density f, then*

$$\frac{1}{n}\,h(X)\,\geq\,\log\,\|f\|^{-1/n}.$$

If, in addition, f is log-concave, then

$$\frac{1}{n}h(X)\leq\,1+\log\,\|f\|^{-1/n}.$$

Observe that the lower bound here is trivial, since

$$h(X)\geq\int_{\mathbb{R}^n}f(x)\log\frac{1}{\|f\|}\,dx=\log\frac{1}{\|f\|}.$$

On the other hand, let us point out that the upper bound in Proposition I.2 improves upon the naive Gaussian maximum entropy bound (obtained without a log-concavity assumption). Indeed, if the covariance matrix R of X with entries $R_{ij}=\mathrm{cov}(X_i,X_j)$ is fixed, then $h(X)$ is maximized for the normal distribution. This property leads to the upper bound

$$\frac{1}{n}\,h(X)\leq C+\log\sigma, \tag{3}$$

where $\sigma=\det^{1/n}(R)$ and $C=\log\sqrt{2\pi e}$. Now, according to one general comparison principle (stated in Section III), in the class of all probability densities, the quantity $\sigma\,\|f\|^{1/n}$ is minimized for the uniform distribution on ellipsoids. This property yields

$$\sigma\geq c\,\|f\|^{-1/n} \tag{4}$$

for some universal constant $c>0$. Hence, modulo the constant C, (3) would indeed be improved if we replace σ with $\|f\|^{-1/n}$.

While Proposition I.2 is already remarkable in its own right, log-concavity is a relatively strong assumption, and it would be advantageous to loosen it. Inspired by this objective, one wishes to study more general classes of probability distributions, satisfying weaker convexity conditions (in comparison with log-concavity). As a natural generalization, we consider probability densities of the form

$$f(x)=\varphi(x)^{-\beta},\quad x\in\Omega, \tag{5}$$

where φ is a positive convex function on an open convex set Ω in $\mathbb{R}^n$. To see that this is a natural generalization, observe that any log-concave density is of this form for any $\beta>0$ since the exponential function composed with a convex function is convex, and that log-concave distributions have finite moments of all orders, whereas densities of the form (5) can be heavy-tailed. For example, the Cauchy distribution on the real line has density $f(x)=[\pi(1+x^2)]^{-1}=\varphi(x)^{-2}$ with φ being convex, although it is certainly not log-concave.

Another example, which is of significant relevance to our development, is the n-dimensional Pareto distribution. For fixed parameters $\beta>n$ and $a>0$, this has the density

$$f_{\beta,a}(x)=\frac{1}{Z_n(\beta,a)}\,(a+x_1+\cdots+x_n)^{-\beta},\quad x_i>0, \tag{6}$$

where $Z_n(\beta,a)$ is the normalizing factor, i.e.,

$$Z_n(\beta,a)=\int_{\mathbb{R}_+^n}(a+x_1+\cdots+x_n)^{-\beta}\,dx.$$

(As shown in Lemma A.1, Pareto distributions with $\beta\leq n$ do not exist, since $Z_n(\beta,a)$ is finite if and only if $\beta>n$.)

Theorem I.3. *If a random vector X in $\mathbb{R}^n$ has a density f of the form (5) with $\beta\geq n+1$, and if $\|f\|$ is fixed, the entropy $h(X)$ is maximal for the n-dimensional Pareto distribution.*

Since $h(X)+\log\|f\|$ is an affine invariant, one may assume $\|f\|=1$ without loss of generality. Also, put for definiteness $a=1$ and write $Z(\beta)=Z_n(\beta,1)$, and X_β for the random vector with density $f_{\beta,1}$. Then Theorem I.3 may be equivalently written as

$$h(X)+\log\,\|f\|\,\leq\,h(X_\beta)+\log\,\|f_{\beta,1}\|, \tag{7}$$

Moreover, as shown in the Appendix,

$$\frac{1}{Z(\beta)}=(\beta-1)\dots(\beta-n)=(\beta-1)_n,$$

where $(b-1)_n=\Gamma(b)/\Gamma(b-n)$ is the n-th falling factorial of $b-1$, and (7) takes the form

$$h(X)+\log\,\|f\|\,\leq\,\beta\sum_{i=1}^n\frac{1}{\beta-i}. \tag{8}$$

Hence we recover Proposition I.2 in the limit as $\beta\to+\infty$.

It is convenient, for the sake of comparison with Proposition I.2, to write some consequences of Theorem I.3 in the following form.

Corollary I.4. *For the range $\beta\geq\beta_0 n$ with fixed $\beta_0>1$ (and still for $\beta\geq n+1$), we have*

$$\frac{1}{n}\,h(X)\leq\,C_{\beta_0}+\log\,\|f\|^{-1/n},$$

where the constant C_{β_0} depends on β_0 only. In fact, one may take $C_{\beta_0}=\frac{\beta_0}{\beta_0-1}$. However, in the larger range $\beta\geq\beta_0+n$ with fixed $\beta_0\geq1$,

$$\frac{1}{n}\,h(X)\leq\,\log\,\|f\|^{-1/n}+O(\log n),$$

where the $O(\log n)$ term may be explicitly bounded.

For the range $\beta\leq n$, it is not possible to control $h(X)$ in terms of $\|f\|$. In this case $h(X)+\log\,\|f\|$ may be as large, as we wish (which can be seen on the example of the Pareto distribution with $\beta\to n$). One explanation for this observation could be the fact that the measures with densities (5) for $\beta\leq n$ may not be convex (see Remark VI.2), or viewed another way that there do not exist Pareto distributions for $\beta\leq n$ (see Lemma A.1). Thus, we still have a gap $n<\beta<n+1$, when Theorem I.3 is not applicable, and we cannot say whether one may bound $h(X)$ in terms of $\|f\|$.

For ease of navigation, let us outline how this note is organized. In Sections II and III, we expand on the motivation for considering Proposition I.2 by proving the statements (2) and (4).

In Section IV, we prove our main results for log-concave probability measures. In particular, Proposition I.2 and Theorem I.1 emerge as consequences of a more general result that bounds the Rényi entropy of any order $p \geq 1$ using the maximum of the density. As a corollary of this, we also show that any two Rényi entropies become comparable for the class of log-concave densities.

In Section V, we use the preceding development to give a new and easy-to-state entropic formulation of the famous slicing or hyperplane conjecture. Indeed, the hyperplane conjecture can be formulated as a multidimensional analogue of the property (2), different from the multidimensional analogue already represented by Theorem I.1. Specifically, if $D(f)$ is the "entropic distance" of f from Gaussianity (defined precisely later), then the property (2) may be rewritten in the form $0 \leq D(f) \leq c$ for some constant c and every one-dimensional log-concave density f, whereas the hyperplane conjecture is shown to be equivalent to the statement that $D(f) \leq cn$ for some universal constant c and every log-concave density f on $\mathbb{R}^n$. Furthermore existing partial results on the slicing problem are used to deduce a universal bound on $D(f)$ for all log-concave densities f on $\mathbb{R}^n$, although the dominant term in this bound is $\frac{1}{4} n \log n$ rather than linear in n.

Section VI begins the study of a more general class of probability measures, the so-called "convex probability measures". Sections VII and VIII are dedicated to proving Theorem I.3 (and Corollary I.4); the former describes some necessary tools including a result on norms of convex functions, and we complete the proof in the latter.

Section IX develops several applications– to entropy rates of certain discrete-time stochastic processes under convexity conditions, to approximating the entropy of certain infinitely divisible distributions, and to giving a quantitative version of an inequality of Junge concerning the behavior of $\|f\|$ on convolution. We end in Section X with some discussion.

II. One-dimensional log-concave distributions

Proposition II.1. *For a one-dimensional log-concave random variable X with standard deviation σ,*

$$\log(C_0 \sigma) \leq h(X) \leq \log(C_1 \sigma)$$

for some positive constants C_0, C_1. The optimal constant $C_1 = \sqrt{2\pi e}$ is achieved for the normal, and the optimal constant $C_0 > 1/\sqrt{2}$.

Proof: The upper bound holds without the log-concavity assumption, and is obtained simply by using the Gaussian entropy.

Since f is log-concave, it is supported on an interval (a, b) (where a may take the value $-\infty$ and b may take the value ∞), and moreover, it is strictly positive on this support interval (being of the form $e^{-\varphi}$ with φ convex). If F is the cumulative distribution function of f restricted to (a, b), its inverse $F^{-1} : (a, b) \to (0, 1)$ is well defined since the positivity of f implies that F strictly increases on the support interval. Now consider the function

$$I(t) = f(F^{-1}(t)), \;\; 0 < t < 1.$$

In [10, Proposition A1], it was shown that f is log-concave if and only if I is positive and concave on (0,1). Hence for all $t \in (0, 1)$, $\frac{1}{2} I(t) < \frac{1}{2}[I(t) + I(1-t)] \leq I(\frac{1}{2})$, so that

$$I(t) < 2I(\tfrac{1}{2}) = 2f(m).$$

Taking the supremum over all t, one obtains

$$\max_x f(x) < 2f(m). \tag{9}$$

For one-dimensional log-concave densities $f(x)$, it was shown in [15, Proposition 4.1] that

$$\frac{1}{12} \leq \sigma^2 f(m)^2 \leq \tfrac{1}{2}, \tag{10}$$

where m is the median. Combining (9) and (10) gives

$$\frac{1}{12} \leq \sigma^2 \max_x f(x)^2 < 2$$

$$\text{or} \qquad \sigma/\sqrt{2} < \|f\|^{-1} \leq \sqrt{12}\sigma. \tag{11}$$

Applying Proposition I.2,

$$h(X) \geq \log \|f\|^{-1} > \log \sigma - \tfrac{1}{2} \log 2,$$

which is the desired lower bound. ∎

Even in this one-dimensional setting, the best constant C_0 and corresponding extremal situations seem to be unknown; these would be interesting to identify. Note that the inequalities in (10) are sharp and are attained for the uniform and double exponential distributions.

In Section V, we discuss the possible generalization of Proposition II.1 to general dimension n; this is related to the hyperplane conjecture.

III. An extremal property of ellipsoids

Here we recall the comparison property (4), mentioned in Section I, concerning an extremal property of ellipsoids. It goes back to the work of D. Hensley ([28], Lemma 2), who noticed that, if a probability density f on $\mathbb{R}^n$ is maximized at the origin, then the quantity

$$f(0)^{2/n} \int_{\mathbb{R}^n} |x|^2 f(x)\, dx$$

is minimized for the uniform distribution on the Euclidean balls centered at the origin. More precisely, Hensley considered only symmetric quasi-concave probability densities f, and later K. Ball ([6], Lemma 6) simplified the argument and extended this observation to all measurable densities f satisfying $f(x) \leq f(0)$ for all x.

One may further generalize and strengthen this result, by applying affine transformations to the probability measures μ with densities f.

Proposition III.1. *Put*

$$L_{\mu,\rho} = \int_{\mathbb{R}^n} \rho(\|f\|^{1/n} |x|)\, f(x)\, dx,$$

where $\rho = \rho(t)$ is a given non-decreasing function in $t \geq 0$. In the class of all absolutely continuous probability measures μ on $\mathbb{R}^n$, the functional $L_{\mu,\rho}$ is minimized, when μ is a uniform distribution on a Euclidean ball with center at the origin.

Proof: Since $L_{\mu,\rho}$ does not depend on $\|f\|$, we may assume $\|f\| = 1$. Denote by λ the uniform distribution on the Euclidean ball $B(0, r_n)$ with center at the origin and volume one (so that $\omega_n r_n^n = 1$, where ω_n is the volume of the unit ball). We need to show that $L_{\mu,\rho} \geq L_{\lambda,\rho}$. Since both $L_{\mu,\rho}$ and $L_{\lambda,\rho}$ are linear with respect to ρ, it suffices to consider the case $\rho = 1_{(r,\infty)}$, the indicator function of a half-axis. Then the property $L_{\mu,\rho} \geq L_{\lambda,\rho}$ reads as

$$\mu\{|x| \leq r\} \ \leq\ \lambda\{|x| \leq r\} \ =\ \begin{cases} \omega_n r^n, & \text{for } \ 0 \leq r \leq r_n, \\ 1, & \text{for } \ r > r_n. \end{cases}$$

This inequality is automatically fulfilled, when $r > r_n$. In the other case, due to the assumption $f(x) \leq 1$ (almost everywhere), we have

$$\mu\{|x| \leq r\} = \int_{\{|x|\leq r\}} f(x)\,dx \leq \int_{\{|x|\leq r\}} dx = \lambda\{|x| \leq r\},$$

which is the statement. ■

As a corollary, we obtain the following observation.

Corollary III.2. *Let X be a random vector in $\mathbb{R}^n$ with density f and non-singular covariance matrix R. If $\sigma^2 = \det^{1/n}(R)$,*

$$\sigma \geq c\,\|f\|^{-1/n}$$

for some universal constant $c > 0$.

Proof: Let us return to the basic case $\rho(t) = t^2$. Thus, the functional $L_{\mu,\rho} = \|f\|^{2/n} \int |x|^2\, f(x)\,dx$ is minimal for $\mu = \lambda$, the uniform distribution on the Euclidean ball $B(0, r_n)$ with center at the origin and volume one. Hence, the same is true for the functionals

$$\frac{1}{n}\ \|f\|^{2/n} \int |T(x - x_0)|^2\, f(x)\,dx$$

for any point x_0 and any linear map $T : \mathbb{R}^n \to \mathbb{R}^n$ with $|\det T| = 1$. Taking for x_0 the barycenter or mean of μ, this functional may be written as

$$\frac{1}{n}\ \|f\|^{2/n} \mathrm{tr}\,\mathrm{Cov}(TX)$$

where $\mathrm{tr}\,\mathrm{Cov}(TX)$ denotes the trace of the covariance matrix of $T(X)$. Minimizing over all T's, the above integral turns into

$$\|f\|^{2/n}\, (\det R)^{1/n}, \tag{12}$$

where R is the covariance matrix of X. This follows from the classical representation (see, e.g., [9, Proposition II.3.20]) for the determinant of a positive-definite matrix C:

$$(\det C)^{\frac{1}{n}} = \min\left\{\frac{\mathrm{tr}\,(CA)}{n} : A \geq 0, \det A = 1\right\}.$$

The point is that the quantity (12) is invariant both under all shifts and all linear transforms of $\mathbb{R}^n$. In particular, it is constant for the uniform distribution on all ellipsoids, which thus minimize (12). Analytically, for any probability density f,

$$\|f\|^{2/n}\, (\det R)^{1/n} \geq \frac{1}{n} \int_{B(0,r_n)} |x|^2\,dx = \frac{r_n^2}{n+2} = \frac{\omega_n^{-2/n}}{n+2}.$$

Since r_n is of order $\sqrt{n}$ for the growing dimension n, the right side is separated from zero by a universal constant. ■

In fact, this proof allows us to compute the optimal dimension-free constant. Recall that the volume of the unit ball is $\omega_n = \pi^{n/2}/\Gamma(\frac{n}{2}+1)$. Restricting ourselves for simplicity to even dimension n, the optimal dimension-dependent lower bound becomes

$$\frac{\omega_n^{-2/n}}{n+2} = \frac{\pi^{-1}}{(\frac{n}{2})!^{-2/n}} \cdot \frac{1}{n+2},$$

which by Stirling's approximation is multiplicatively well-approximated for large n by

$$\frac{1}{\pi(n+2)}\left[\sqrt{2\pi\left(\frac{n}{2}\right)} \cdot \left(\frac{n}{2e}\right)^{\frac{n}{2}}\right]^{\frac{2}{n}} = \frac{1}{2\pi e} \cdot \frac{n}{n+2} \cdot (\pi n)^{\frac{1}{n}}.$$

As $n \to \infty$ through the subsequence of even numbers, this quantity converges to $c = (2\pi e)^{-1}$, which is therefore the optimal dimension-free constant. Observe that when Corollary III.2 is written with this dimension-free constant, equality is not attained for any finite dimension n but only asymptotically. In Section V, we give a very simple proof of Corollary III.2 using entropy that also naturally yields the exact dimension-free constant.

IV. Rényi entropies of log-concave distributions

Recall the definition of the Rényi entropy of order p: for $p > 1$, and a random vector X in $\mathbb{R}^n$ with density f,

$$h_p(X) = \frac{p}{p-1} \log \frac{1}{\|f\|_p},$$

where

$$\|f\|_p = \left(\int_{\mathbb{R}^n} f^p\,dx\right)^{1/p}$$

is the usual L^p-norm with respect to Lebesgue measure on $\mathbb{R}^n$. By continuity, $h_p(X)$ reduces to the Shannon differential entropy $h(X)$ as $p \to 1$, and to $h_\infty(X) = \log \|f\|^{-1}$ as $p \to \infty$. The definition of $h_p(X)$ continues to make sense for $p \in (0, 1)$ even though $\|f\|_p$ is then not a norm.

Theorem IV.1. *Fix $p \in (1, \infty)$. If a random vector X in $\mathbb{R}^n$ has density f, then*

$$\frac{1}{n}\, h_p(X) \ \geq\ \log\, \|f\|^{-1/n},$$

with equality if and only if X has the uniform distribution on any set of positive finite Lebesgue measure. If, in addition, f is log-concave, then

$$\frac{1}{n} h_p(X) \leq \frac{1}{p-1} \log p + \log\, \|f\|^{-1/n},$$

with equality for the n-dimensional exponential distribution, concentrated on the positive orthant with density $f(x) = e^{-(x_1+\cdots+x_n)}$, $x_i > 0$.

Proof: The lower bound is trivial and holds without any assumption on the density.

Let us derive the upper bound for $p > 1$. By definition of log-concavity, for any $x, y \in \mathbb{R}^n$,

$$f(tx+sy) \geq f(x)^t f(y)^s,\ t,s>0, t+s=1. \quad (13)$$

Taking the t-th root yields

$$f(tx+sy)^{1/t} \geq f(x)\, f(y)^{s/t}.$$

Integrating with respect to x and using the assumption that $\int f = 1$, we get

$$t^{-n}\int f(x)^{1/t}\,dx \geq f(y)^{s/t}.$$

It remains to optimize over y's, so that

$$\int f(x)^{1/t}\,dx \geq t^n\,\|f\|^{s/t}.$$

Taking $p = 1/t$ implies $\int f^p \geq p^{-n}\|f\|^{p-1}$ or

$$\|f\|_p^{-1} \leq p^{n/p}\|f\|^{\frac{1-p}{p}},$$

so that

$$\begin{aligned} h_p(X) &\leq \frac{p}{p-1}\log[p^{n/p}\|f\|^{\frac{1-p}{p}}] \\ &= \frac{n}{p-1}\log p + \log\|f\|^{-1}. \end{aligned}$$

It is easy to check that a product of exponentials is an instance of equality. ■

Thus a maximizer of the Rényi entropy of order p under a log-concavity shape constraint and a supremum norm constraint is the exponential distribution, irrespective of p. This is not the only maximizer– indeed, affine transforms with determinant 1 of an exponentially distributed random vector will also work. Let us remark that if one instead imposes a variance constraint, the maximizers of Rényi entropy are Student's distributions as shown by Costa, Hero and Vignat [22], which specialize to the Gaussian for $p = 1$. (See also Johnson and Vignat [31] and Lutwak, Yang and Zhang [36], [37] for additional related results.)

We may now prove some of the results stated in Section I.

Proof of Proposition I.2: Note that Proposition I.2 is just a limiting version of Theorem IV.1, obtained by letting $p \downarrow 1$. However, it is not automatic, since there exist densities such that $h_p(X) = \infty$ for every $p > 1$ but $h(X) < \infty$. (An example of such a density is

$$f(x) = \frac{c}{x\log^3(1/x)},\ 0 < x < \tfrac{1}{2},$$

where c is a normalizing constant.) Note that by L'Hôpital's rule, what one needs to show is that

$$\lim_{p\downarrow 1} -\left(\int f^p\right)^{-1}\frac{d}{dp}\int f^p$$

exists and equals $h(X)$. This calls for three limit interchanges, each of which can be justified by the Lebesgue dominated convergence theorem if $h_p(X)$ is finite for $p \in (1, 2]$. In our context of log-concave densities, this is always the case because of Theorem IV.1 and the boundedness of log-concave densities. Alternatively, a direct proof of Proposition I.2 can be given similar to that of Theorem IV.1 by integrating (13) with respect to x, maximizing over y, and then comparing derivatives in t at $t = 1$. ■

Proof of Theorem I.1: To see the relationship with the Gaussian, simply observe that the maximum density of the $N(0, \sigma^2 I)$ distribution is $(2\pi\sigma^2)^{-n/2}$. (Here as usual, we use $N(\mu, \Sigma)$ to denote the Gaussian distribution with mean μ and covariance matrix Σ.) Thus matching the maximum density of f and the isotropic normal Z leads to $(2\pi\sigma^2)^{1/2} = \|f\|^{-1/n}$, and

$$\frac{1}{n}h(Z) = \tfrac{1}{2}\log(2\pi e\sigma^2) = \tfrac{1}{2} + \log\|f\|^{-1/n}.$$

This completes the proof of Theorem I.1. ■

Theorem IV.1 also implies that for log-concave random vectors, Rényi entropies of orders p and q are related for any $p, q \geq 1$.

Corollary IV.2. *If X has a log-concave distribution on $\mathbb{R}^n$, and $p, q \in [1, \infty]$, then*

$$\frac{h_p(X)}{n} \leq \frac{\log p}{p-1} + \frac{h_q(X)}{n}.$$

Since Theorem IV.1 is just the special case $q = \infty$ of Corollary IV.2, the two statements are mathematically equivalent.

While the preceding discussion relies heavily on the value of the density at the mode, one can also extract information based on the value at the mean. Let $g : \mathbb{R}^n \to [0, \infty)$ be a log-concave function such that $\int g \in (0, \infty)$. Let x_{mean} be the barycenter or mean of g. Then it was shown by Fradelizi [25] that

$$\sup_{x\in\mathbb{R}^n} g(x) \leq e^n g(x_{\text{mean}}).$$

Combining Proposition I.2 with Fradelizi's lemma immediately yields the following corollary.

Corollary IV.3. *If a random vector X has log-concave density f, with mean x_{mean} and mode x_{mode}, then*

$$h(X) \in [\log f(x_{mean})^{-1} - n, \log f(x_{mode})^{-1} + n].$$

V. An Entropic Formulation of the Slicing Problem

The main observation of this section is a relationship between the entropy distance to Gaussianity $D(f)$ and the isotropic constant L_f for densities of convex measures.

For a random vector X with density f on $\mathbb{R}^n$, the relative entropy from Gaussianity $D(f)$ or $D(X)$ is defined by

$$\int f(x)\log\frac{f(x)}{g(x)}dx,$$

where g is the density of the Gaussian distribution with the same mean and the same covariance matrix as X. If Z has density g, then one may write $D(X) = h(Z) - h(X)$ (see, e.g., Cover and Thomas [24]).

For any probability density function f on $\mathbb{R}^n$ with covariance matrix R, define its isotropic constant L_f by

$$L_f^2 = \|f\|^{2/n}\det{}^{\frac{1}{n}}(R).$$

The isotropic constant has a nice interpretation for uniform distributions on convex sets K. If one rescales K (by a linear transformation) so that the volume of the convex set is 1 and the covariance matrix is a multiple of the identity, then $L_K^2 := L_f^2$ is the value of the multiple.

Observe that both $D(f)$ and L_f are affine invariants. The following result relating them may be viewed as an alternative form of Theorem I.1 relevant to matching first and second moments rather than the supremum norm.

Theorem V.1. *For any density f on $\mathbb{R}^n$,*

$$\frac{1}{n}D(f) \leq \log[\sqrt{2\pi e}L_f],$$

with equality if and only if f is the uniform density on some set of positive, finite Lebesgue measure. If f is a log-concave density on $\mathbb{R}^n$, then

$$\log\left[\sqrt{\frac{2\pi}{e}}L_f\right] \leq \frac{1}{n}D(f),$$

with equality if f is a product of one-dimensional exponential densities.

Proof: Let $X \sim f$ have covariance matrix R. If $Z \sim N(0,R)$,

$$h(Z) = \tfrac{1}{2}\log[(2\pi e)^n \det(R)] = \frac{n}{2}\log(C\sigma^2),$$

where $\sigma^2 = \det(R)^{\frac{1}{n}}$ and $C = 2\pi e$. Thus

$$\begin{aligned}\frac{1}{n}D(X) &= \frac{h(Z)-h(X)}{n}\\ &\leq \tfrac{1}{2}\log(C\sigma^2) - \log\|f\|^{-\frac{1}{n}}\\ &= \tfrac{1}{2}\log[C\sigma^2\|f\|^{2/n}] = \tfrac{1}{2}\log[CL_f^2],\end{aligned}$$

and

$$\begin{aligned}\frac{1}{n}D(X) &= \frac{h(Z)-h(X)}{n}\\ &\geq \tfrac{1}{2}\log(C\sigma^2) - \log\|f\|^{-\frac{1}{n}} - 1\\ &= \tfrac{1}{2}\log\left[\frac{C}{e^2}\sigma^2\|f\|^{2/n}\right] = \tfrac{1}{2}\log\left[\frac{2\pi}{e}L_f^2\right],\end{aligned}$$

where the inequalities come from Proposition I.2. ■

Note that this immediately gives an extremely simple alternate proof of Corollary III.2. Indeed, since $D(f) \geq 0$, we trivially have

$$\sqrt{2\pi e}L_f \geq 1,$$

which is Corollary III.2 with the optimal dimension-free constant.

On the other hand, whether or not the isotropic constant is bounded from above by a universal constant for the class of uniform distributions on convex bodies is an open problem that has attracted a lot of attention in the last 20 years. It was originally raised by J. Bourgain [19] in (a slight variation of) the following form.

Conjecture V.2. *[Slicing Problem or Hyperplane Conjecture] There exists a universal, positive constant c (not depending on n) such that for any convex set K of unit volume in $\mathbb{R}^n$, there exists a hyperplane H such that the $(n-1)$-dimensional volume of the section $K \cap H$ is bounded below by c.*

There are several equivalent formulations of the conjecture, all of a geometric or functional analytic flavor. Whereas Bourgain [19] and Milman and Pajor [42] looked at aspects of the conjecture in the setting of centrally symmetric, convex bodies, a popular formulation developed by Ball [6] is that the isotropic constant of a log-concave measure in any Euclidean space is bounded above by a universal constant independent of dimension. Connections of this question with slices of κ-concave measures are described in [12].

We will now demonstrate that the hyperplane conjecture has a formulation in purely information-theoretic terms. It is useful to start by mentioning the following equivalences.

Corollary V.3. *Let $c(n)$ be any non-decreasing sequence, and $c'(n) = c(n) + \frac{1}{2}\log(2\pi e)$. Then the following statements are equivalent:*

(i) *For any log-concave density f on $\mathbb{R}^n$, $L_f \leq e^{c(n)}$.*
(ii) *For any log-concave density f on $\mathbb{R}^n$, $D(f) \leq nc'(n)$.*
(iii) *$\sup_f \min_g D(f\|g) \leq nc'(n)$, where the minimum is taken over all Gaussian densities on $\mathbb{R}^n$, and the maximum is taken over all log-concave densities on $\mathbb{R}^n$.*

Proof: The equivalence of (i) and (ii) follows from Theorem V.1, and that of (ii) and (iii) follows from the easily verified fact that $D(f) = \min_g D(f\|g)$, where g is allowed to run over all Gaussian distributions. ■

Furthermore, the seminal paper of Hensley [28] (cf. Milman and Pajor [42]) showed that for an isotropic convex body K, and any hyperplane H passing through its barycenter,

$$c_1 \leq L_K \text{Vol}_{n-1}(K \cap H) \leq c_2,$$

where $c_2 > c_1 > 0$ are universal constants. Hence the statements of Corollary V.3, when restricted to uniform distributions on convex sets, are also equivalent to the statement that

$$\text{Vol}_{n-1}(K \cap H) \geq e^{-c(n)}.$$

Thus the slicing problem or the hyperplane conjecture is simply the conjecture that $c(n)$ can be taken to be constant (independent of n), in *any* of the statements of Corollary V.3.

Conjecture V.4 (Entropic Form of Hyperplane Conjecture)**.** *For any log-concave density f on $\mathbb{R}^n$ and some universal constant c,*

$$\frac{D(f)}{n} \leq c.$$

This gives a pleasing formulation of the slicing problem as a statement about the (dimension-free) closeness of an arbitrary log-concave measure to a Gaussian measure.

Let us give another entropic formulation as a statement about the (dimension-free) closeness of an arbitrary log-concave measure to a *product* measure. If f is an arbitrary density on $\mathbb{R}^n$ and f_i denotes the i-th marginal of f, set

$$I(f) = D(f\|f_1 \otimes f_2 \otimes \ldots \otimes f_n);$$

this is the "distance from independence", or the relative entropy of f from the distribution of the random vector that has the same one-dimensional marginals as f but has independent components. (For $n = 2$, this reduces to the mutual information.)

Conjecture V.5 (SECOND ENTROPIC FORM OF HYPERPLANE CONJECTURE)**.** *For any log-concave density f on $\mathbb{R}^n$ and some universal constant c,*

$$\frac{I(f)}{n} \le c.$$

Proof of equivalence of Conjectures V.4 and V.5: The following identity is often used in information theory: if f is an arbitrary density on $\mathbb{R}^n$ and $f^{(0)}$ is the density of some product distribution (i.e., of a random vector with independent components), then

$$D(f\|f_0) = \sum_{i=1}^{n} D(f_i\|f_i^{(0)}) + I(f), \tag{14}$$

where f_i and $f_i^{(0)}$ denote the i-th marginals of f and $f^{(0)}$ respectively.

Now Conjecture V.4 is equivalent to its restriction to those log-concave measures with zero mean and identity covariance (since $D(f)$ is an affine invariant). Applying the identity (14) to such measures,

$$D(f) = \sum_{i=1}^{n} D(f_i) + I(f),$$

since the standard normal is a product measure. The lower bound of Proposition II.1 asserts that $h(X) \ge C + \log\sigma$ for one-dimensional log-concave distributions; thus each $D(f_i)$ is bounded from above by some universal constant. Thus $D(f)$ being uniformly $O(n)$ is equivalent to $I(f)$ being uniformly $O(n)$. ∎

Observe that mimicking Proposition II.1, Conjecture V.4 may be written in the form: for a log-concave random vector X taking values in $\mathbb{R}^n$,

$$\frac{1}{n}\, h(X) \ge C + \log\sigma, \tag{15}$$

or

$$\frac{1}{n}\, h(X) \ge \frac{1}{n}\, h(Z) - C', \tag{16}$$

where C, C' are universal constants, and Z is the normal with the same covariance matrix as X. Owing to (4), the form (15) would strengthen the naive lower bound of Proposition I.2. As for form (16), it looks like the lower bound of Theorem I.1, except that the way in which the matching Gaussian is chosen is to match the covariance matrix rather than the maximum density.

Existing partial results on the slicing problem already give insight into the closeness of log-concave measures to Gaussian measures. For many years, the best known bound in the slicing problem for general bounded convex sets, due to Bourgain [20] in the centrally-symmetric case and generalized by Paouris [44] to the non-symmetric case, was

$$L_K \le cn^{1/4}\log(n+1).$$

Recently Klartag [34] removed the $\log n$ factor and showed that $L_K \le cn^{1/4}$. Using a transference result of Ball [6] from convex bodies to log-concave functions, the same bound is seen to also apply to L_f, for a general log-concave density f. Combining this with Corollary V.3 leads immediately to the following result.

Proposition V.6. *There is a universal constant c such that for any log-concave density f on $\mathbb{R}^n$,*

$$D(f) \le \frac{1}{4} n\log n + cn.$$

Note that the property (2) (quantified by Proposition II.1) for a one-dimensional log-concave density f may be rewritten in the form $0 \le D(f) \le c$ for some constant c. Proposition V.6 is thus a multidimensional version of the statement (2).

VI. CONVEXITY OF MEASURES

Convexity properties of probability distributions may be expressed in terms of inequalities of the Brunn-Minkowski-type. A probability measure μ on $\mathbb{R}^n$ is called κ-concave, where $-\infty \le \kappa \le +\infty$, if it satisfies

$$\mu\big(tA + (1-t)B\big) \ge \big[\,t\mu(A)^\kappa + (1-t)\mu(B)^\kappa\big]^{1/\kappa} \tag{17}$$

for all $t \in (0,1)$ and for all Borel measurable sets $A, B \subset \mathbb{R}^n$ with positive measure. Here $tA + (1-t)B = \{tx + (1-t)y : x \in A, y \in B\}$ stands for the Minkowski sum of the two sets. When $\kappa = 0$, the inequality (17) becomes

$$\mu\big(tA + (1-t)B\big) \ge \mu(A)^t \mu(B)^{1-t},$$

and we arrive at the notion of a log-concave measure, introduced by Prékopa, cf. [46], [47], [35]. In the absolutely continuous case, the log-concavity of a measure is equivalent to the log-concavity of its density, as in (1). When $\kappa = -\infty$, the right-hand side is understood as $\min\{\mu(A), \mu(B)\}$. The inequality (17) is getting stronger as the parameter κ is increasing, so in the case $\kappa = -\infty$ we obtain the largest class, whose members are called convex or hyperbolic probability measures. For general κ's, the family of κ-concave measures was introduced and studied by C. Borell [17], [18].

A remarkable feature of this family is that many important geometric properties of κ-concave measures, like the properties expressed in terms of Khinchin and dilation-type inequalities, may be controlled by the parameter κ, only, and in essence do not depend on the dimension n (although the dimension may appear in the density description of many κ-concave measures).

A full characterization of κ-concave measures was given by C. Borell in [17], [18], cf. also [21]. Namely, any κ-concave probability measure is supported on some (relatively) open convex set $\Omega \subset \mathbb{R}^n$ and is absolutely continuous with respect to Lebesgue measure on Ω. Necessarily, $\kappa \le 1/\dim(\Omega)$, and if Ω has dimension n, we have:

Proposition VI.1. *An absolutely continuous probability measure μ on $\mathbb{R}^n$ is κ-concave, where $-\infty \le \kappa \le 1/n$, if and only if μ is supported on an open convex set $\Omega \subset \mathbb{R}^n$, where it has a positive density f such that, for all $t \in (0,1)$ and $x, y \in \Omega$,*

$$f(tx + (1-t)y) \ge \big[\,tf(x)^{\kappa_n} + (1-t)f(y)^{\kappa_n}\big]^{1/\kappa_n}, \tag{18}$$

where $\kappa_n = \frac{\kappa}{1-n\kappa}$.

Following [3], we call non-negative functions f, satisfying (18), κ_n-concave. Thus, μ is κ-concave if and only if f is κ_n-concave.

If $\kappa < 0$, one may represent the density in the form $f = \varphi^{-\beta}$ with $\beta \geq n$, $\kappa = -1/(\beta - n)$, where φ is an arbitrary positive convex function on Ω, satisfying the normalization condition $\int_\Omega \varphi^{-\beta}\, dx = 1$. Moreover, the condition $\beta \geq n+1$ like in Theorem I.3 corresponds to the range $-1 \leq \kappa < 0$.

Remark VI.2. Note that a density of form $f = \varphi^{-\beta}$, where φ is an arbitrary positive convex function on Ω and $\beta < n$, need not be the density of a convex measure. Indeed, it is not unless φ itself can be written as a convex function raised to a large enough power.

Proposition VI.1 remains to hold without the normalization condition $\mu(\mathbb{R}^n) = 1$. In particular, if f is a positive κ_n-concave function on an open convex set Ω in $\mathbb{R}^n$, then the measure $d\mu(x) = f(x)\, dx$ is κ-concave, that is, it satisfies the Brunn-Minkowski-type inequality (17). For example, the Lebesgue measure on $\mathbb{R}^n$ is $\frac{1}{n}$-concave (in which case $\kappa_n = +\infty$).

This sufficient condition will be used in dimension one as the following:

Corollary VI.3. *Let* $\alpha > 0$. *If* u *is a positive concave function on an interval* $(a,b) \subset \mathbb{R}$, *then the measure on* (a,b) *with density* u^α *is* $\frac{1}{\alpha+1}$*-concave.*

VII. Log-concavity of norms of convex functions

In order to present the proof of our main result for κ-concave probability measures (which we will do in Section VIII), we first need to develop some functional-analytic preliminaries.

Given a measurable function f on a measurable set $\Omega \subset \mathbb{R}^n$ (of positive measure), we write

$$\|f\|_p = \left(\int_\Omega |f|^p\, dx\right)^{1/p}, \quad -\infty < p < +\infty.$$

For the value $p = 0$, the above expression may be understood as the geometric mean $\|f\|_0 = \exp \int \log |f|\, dx$.

It is easy to see that the function $p \to \|f\|_p^p$ is log-convex (which is referred to as Lyapunov's inequality). C. Borell complemented this general property with the following remarkable observation ([16, Theorem 2]).

Proposition VII.1. *If* Ω *is a convex body, and if* f *is positive and concave on* Ω, *then the function*

$$p \longrightarrow C_{n+p}^n \|f\|_p^p = \frac{(p+1)\dots(p+n)}{n!} \int_\Omega f^p\, dx \qquad (19)$$

is log-concave for $p > 0$.

Here we use the standard binomial coefficients

$$C_q^n = \frac{q(q-1)\dots(q-n+1)}{n!}.$$

Borell's theorem, Proposition VII.1, may formally be generalized to the class of κ-concave functions f with $\kappa > 0$, since then $f = \varphi^\kappa$ with concave φ, and one may apply the log-concavity result (19), as well as the inequality (21) to φ. However, for the purpose of proving Theorem I.3, with the aim of going beyond log-concave probability measures, we are mostly interested in the case where $\kappa < 0$, when the function φ is convex.

Thus what we require is a version of Proposition VII.1 for convex functions φ. The following theorem, proved in [14], supplies such a result.

Theorem VII.2. *If* φ *is a positive, convex function on a open convex set* Ω *in* $\mathbb{R}^n$, *then the function*

$$p \longrightarrow C_{p-1}^n \|\varphi\|_{-p}^{-p} = \frac{(p-1)\dots(p-n)}{n!} \int_\Omega \varphi^{-p}\, dx \quad (20)$$

is log-concave on the half-axis $p > n+1$.

It is interesting to note that Borell [16] obtained a different proof of Berwald's inequality [8], which is famous among functional analysts, as a consequence of Proposition VII.1.

Proposition VII.3. *For* $0 < p < q$,

$$\left(C_{n+q}^n |\Omega|^{-1}\right)^{1/q} \|f\|_q \leq \left(C_{n+p}^n |\Omega|^{-1}\right)^{1/p} \|f\|_p. \qquad (21)$$

Equality is achieved when the normalized norms are constant, which corresponds to the linear function $f(x) = x_1 + \dots + x_n$ *on the convex body*

$$\Omega = \{x \in \mathbb{R}^n : x_i > 0,\ x_1 + \dots + x_n < 1\}.$$

Berwald's inequality turns out to have interesting applications to information theory as well as convex geometry (see [14], [13]).

VIII. Entropy of κ-concave distributions

In this section, we explain how to use the remarkable property of convex functions described by Theorem VII.2 in proving Theorem I.3.

Proof: **(of Theorem I.3.)** Let $f = \varphi^{-\beta}$ be a probability density for a random vector X in $\mathbb{R}^n$ with $\beta \geq n+1$, where φ is as in Theorem VII.2. Define f to be zero outside Ω. As is shown in [11], the density admits a bound

$$f(x) \leq \frac{C}{(1+|x|)^\beta}, \quad \text{for all } \ x \in \Omega,$$

with some constant C, so $\varphi(x) \geq c\,(1+|x|)$ with some $c > 0$ (depending on φ). Hence, the function

$$V(p) = \log \int_\Omega \varphi^{-p}\, dx$$

is finite and differentiable for $p > n$ with $V'(p) = -\int_\Omega \varphi^{-p} \log \varphi\, dx / \int_\Omega \varphi^{-p}\, dx$. In particular,

$$h(X) = -\beta\, V'(\beta). \qquad (22)$$

To proceed, assume $\|f\| = 1$, that is, $\inf_\Omega \varphi = 1$. This assumption can be made since the quantity of interest in Theorem I.3, $h(X) + \log \|f\|$ is an affine invariant; so one can scale X to make $\|f\| = 1$. Like in the proof of Theorem IV.1, for $t \in [0,1]$, write

$$f(tx + (1-t)y) \geq \left[t f(x)^{-1/\beta} + (1-t) f(y)^{-1/\beta}\right]^{-\beta},$$

which is valid for any $x, y \in \Omega$. Integrating with respect to x over the whole space, we get

$$t^{-n} \geq \int \left[tf(x)^{-1/\beta} + (1-t) f(y)^{-1/\beta} \right]^{-\beta} dx$$

with equality at $t = 1$. Hence, we may compare derivatives of both sides with respect to t at this point, which gives

$$n \geq \beta \int f(x)^{1+1/\beta} \left[f(x)^{-1/\beta} - f(y)^{-1/\beta} \right] dx.$$

Optimizing over all y's and using $\int f\, dx = 1$, we arrive at the bound

$$n \geq \beta \left(1 - \int_\Omega \varphi^{-\beta-1}\, dx \right). \tag{23}$$

Note that for the Pareto distribution there is equality at every step made before (since in that case φ is affine).

Now, it is time to apply Theorem VII.2. Put $U(p) = \log\left[(p-1)\dots(p-n)\right]$, so that $R(p) = U(p) + V(p)$ is concave on the half-axis $p \geq n+1$. The concavity implies that

$$R'(\beta) \geq R(\beta+1) - R(\beta).$$

Equivalently, since $V(\beta) = 0$ and $U(\beta+1) - U(\beta) = \log \frac{\beta}{\beta-n}$, we have

$$V'(\beta) \geq V(\beta+1) + \log \frac{\beta}{\beta-n} - U'(\beta). \tag{24}$$

But (23) is telling us that $V(\beta+1) \geq \log(1 - \frac{n}{\beta})$, so by (24),

$$V'(\beta) \geq -U'(\beta) = -\sum_{i=1}^n \frac{1}{\beta-i}.$$

With the representation (22) we arrive at the bound (8),

$$h(X) \leq \beta \sum_{i=1}^n \frac{1}{\beta-i}.$$

From Lemma A.2, this is recognized as the entropy of the n-dimensional Pareto density (6), and hence Theorem I.3 describes an extremal property of the Pareto distribution. ∎

In fact, as done for log-concave distributions in Section IV, it is possible obtain analogous bounds for the Rényi entropy of any order. We only state the result here; it is proved in [14].

Theorem VIII.1. *Fix $p \in (1, \infty)$. If a random vector X in $\mathbb{R}^n$ has density f and a κ-concave distribution for $-1 \leq \kappa < 0$, then*

$$\begin{aligned} 0 &\leq \frac{1}{n} h_p(X) - \log \|f\|^{-1/n} \\ &\leq \frac{1}{p-1} \log \left[\frac{(\beta p - 1)\dots(\beta p - n)}{(\beta-1)\dots(\beta-n)} \right], \end{aligned}$$

where $\beta = n + \frac{1}{(-\kappa)}$.

Consequently one has an extension of Corollary IV.2 to the convex measure case: for X as in Theorem VIII.1 and any $1 < p < q \leq \infty$,

$$\begin{aligned} 0 &\leq \frac{h_p(X)}{n} - \frac{h_q(X)}{n} \\ &\leq \frac{1}{p-1} \sum_{i=1}^n \log \left[1 + \left(\frac{\beta}{\beta-i} \right)(p-1) \right]. \end{aligned}$$

To conclude this section, we show how Theorem I.3 implies Corollary I.4.

Proof: **(of Corollary I.4.)** For Corollary I.4, observe that in the regime $\beta \geq \beta_0 n$,

$$\begin{aligned} \frac{\beta}{n} \sum_{i=1}^n \frac{1}{\beta-i} &\leq \frac{1}{n} \sum_{i=1}^n \frac{1}{1 - \frac{i}{\beta_0 n}} = \beta_0 \sum_{i=1}^n \frac{1}{\beta_0 n - i} \\ &\leq \beta_0 \frac{n}{\beta_0 n - n} = \frac{\beta_0}{\beta_0 - 1}. \end{aligned}$$

On the other hand, in the regime $\beta \geq \beta_0 + n$,

$$\begin{aligned} \frac{\beta}{n} \sum_{i=1}^n \frac{1}{\beta-i} &\leq \frac{\beta_0+n}{n} \sum_{i=1}^n \frac{1}{\beta_0+n-i} \\ &\leq \frac{\beta_0+n}{n} \left[\frac{1}{\beta_0+n-1} + \log\left(\frac{\beta_0+n-1}{\beta_0} \right) \right] \\ &\leq \frac{1}{n-1} + \left(1 + \frac{\beta_0}{n} \right) \log \left(1 + \frac{n-1}{\beta_0} \right), \end{aligned}$$

as long as $\beta_0 > 0$. This gives an explicit bound for the $O(\log n)$ term in the second part of Corollary I.4. ∎

IX. Applications

A. Entropy rates

Our first application is to the entropy rate of (strongly) stationary log-concave random processes. We call a discrete-time stochastic process $\mathbf{X} = (X_i)$ *log-concave* if all its finite-dimensional marginals are log-concave distributions. In particular, for the process $\mathbf{X}$ to be log-concave, it is necessary and sufficient for the distribution of $X^n = (X_1, \dots, X_n)$ to be log-concave for each n. Note that an important special case of a log-concave process is a Gaussian process.

An important functional of a discrete-time stochastic process is its *entropy rate*, which is defined by

$$h(\mathbf{X}) = \lim_{n\to\infty} \frac{h(X^n)}{n},$$

when the limit exists.

The only class of processes for which the computation of entropy rate is tractable is the class of stationary Gaussian processes. Indeed, a stationary zero mean Gaussian random process is completely described by its mean correlation function $r_{k,j} = r_{k-j} = E[X_k X_j]$ or, equivalently, by its power spectral density function G, the Fourier transform of the covariance function:

$$G(\lambda) = \sum_{n\in\mathbb{Z}} r_n e^{in\lambda}.$$

For a fixed positive integer n, the probability density function of X^n is the normal density with $n \times n$ covariance matrix R_n, whose entries are $r_{k,j} = r_{k-j}$, and its entropy can be explicitly written. This yields

$$h(\mathbf{X}) = \tfrac{1}{2} \log(2\pi e) + \lim_{n\to\infty} \frac{1}{n} \log\det(R_n).$$

Since R_n is the Toeplitz matrix generated by the power spectral density G (or equivalently by the coefficients $\{r_n\}$), one has from the theory of Toeplitz matrices (see, e.g., (1.11) in Gray [26]) that

$$h(\mathbf{X}) = \tfrac{1}{2}\log(2\pi e) + \frac{1}{2\pi}\int_0^{2\pi} \log G(\lambda) d\lambda.$$

Below we point out that our inequalities give a way of obtaining some information about the entropy rate of a stationary log-concave process.

Corollary IX.1. *For any stationary process* $\mathbf{X}$ *whose finite dimensional marginals are absolutely continuous with respect to Lebesgue measure, let* f_n *be the joint density of* X^n*. If*

$$f_- := \liminf_{n\to\infty} \frac{1}{n} \log \|f_n\|^{-1} > -\infty,$$

then the entropy rate $h(\mathbf{X})$ *exists and* $h(\mathbf{X}) > -\infty$*. If, furthermore,* $\mathbf{X}$ *is a log-concave process and*

$$f_+ := \limsup_{n\to\infty} \frac{1}{n} \log \|f_n\|^{-1} < +\infty,$$

then

$$h(\mathbf{X}) \leq f_+ + 1.$$

Proof: Let $h(X|Y)$ denote conditional entropy. As is well known (see, e.g., Cover and Thomas [24]), for any stationary process $\mathbf{X}$, the sequence $a_n := h(X_n|X^{n-1})$ is a non-increasing sequence, since

$$\begin{aligned} h(X_n|X^{n-1}) &\leq h(X_n|X_2, \ldots, X_{n-1}) \\ &= h(X_{n-1}|X^{n-2}). \end{aligned}$$

(Here one uses the fact that conditioning cannot increase entropy, and the assumed stationarity.) Note also that

$$b_n := \frac{1}{n} h(X^n) = \frac{1}{n}\sum_{i=1}^n a_i.$$

Since

$$\liminf_{n\to\infty} b_n \geq \liminf_{n\to\infty} \frac{1}{n}\log \|f_n\|^{-1} = f_- > -\infty,$$

we must have $\liminf_{n\to\infty} a_n > -\infty$, which combined with the monotonicity of a_n implies that the limit exists and is equal to some $a > -\infty$. Hence the limit of b_n, namely the entropy rate, also exists and is equal to a. The upper bound for the entropy rate follows from Proposition I.2. ∎

One interesting class of processes where this result may be of utility, and where the study of entropy rate has attracted much recent interest, is the class of hidden Markov processes.

Let us also note that reasoning similar to that in Corollary IX.1 can be applied to bound the entropy rate of continuous-time stationary log-concave processes as well (modulo some additional technicalities).

B. The behavior of maximum density on convolution

Our Proposition I.2 can be used to significantly generalize and improve an inequality of Junge [32] for the behavior of the maximum of a density on convolution.

Corollary IX.2. *Let* f *be the density of a* κ*-concave measure on* $\mathbb{R}^n$*, where* $\kappa \in [-1, 0]$*. Then, for any* $m \in \mathbb{N}$*,*

$$\|f^{*m}\| \leq \left(\frac{e^{1-\kappa n}}{\sqrt{m}}\right)^n \|f\|.$$

Proof: We wish to apply Corollary I.4, which requires $\beta \geq \max\{\beta_0 n, n+1\}$ for some $\beta_0 > 1$. Since f is the density of a κ-concave measure, it is of the form $\varphi^{-\beta}$ with φ convex, with $\kappa(\beta - n) = -1$ (see Section VI). Thus the optimal (dimension-dependent!) β_0 that can be chosen in applying Corollary I.4 is given by

$$\kappa = \frac{-1}{\beta - n} \geq \frac{-1}{\max\{1, n(\beta_0 - 1)\}} = -\min\left\{1, \frac{1}{n(\beta_0-1)}\right\};$$

in other words, one may take $\beta_0 = 1 + (-\kappa n)^{-1}$, for which one has $C_{\beta_0} = \frac{\beta_0}{\beta_0 - 1} = 1 - \kappa n$. Now if $X_i \sim f$ are i.i.d., and $S_m = \sum_{i=1}^m X_i$, then

$$\begin{aligned} nC_{\beta_0} + \log \|f^{*m}\|^{-1} &\geq h(S_m) \\ &\geq h(X_1) + \frac{n}{2}\log m \\ &\geq \log \|f\|^{-1} + \frac{n}{2}\log m, \end{aligned}$$

by using Corollary I.4, the Shannon-Stam entropy power inequality [49], and the first part of Proposition I.2. Exponentiating yields the desired result. ∎

In particular, for a log-concave density f on $\mathbb{R}^n$,

$$\|f^{*m}\| \leq \left(\frac{e}{\sqrt{m}}\right)^n \|f\|.$$

While Junge [32] proved that for a symmetric, log-concave density f,

$$\|f^{*m}\| \leq \left(\frac{c}{\sqrt{m}}\right)^n \|f\| \tag{25}$$

for some universal constant c, Corollary IX.2 above generalizes this by removing the symmetry assumption, making the universal constant explicit, and slightly broadening the class of densities allowed; also the proof is far more elementary.

Let us observe that in the three inequalities in the proof of Corollary IX.2, one is tight only for uniforms on convex sets, another only for Gaussians, and the third only for Pareto-type distributions; so Corollary IX.2 is always loose, although it is possible that $c = e$ could be the optimal dimension-free constant in (25).

C. Infinitely divisible distributions

Our third application is to estimating the entropy of certain infinitely divisible distributions. Let X be a random vector in $\mathbb{R}^n$ with density $f(x)$ and characteristic function

$$\varphi(t) = \mathbf{E}\, e^{i\langle X,t\rangle} = \int e^{i\langle x,t\rangle}\, f(x)\, dx.$$

Recall that the distribution of X is *infinitely divisible* if X can be realized as the sum of M independent random vectors, for any natural number M. The Lévy-Khintchine representation theorem [48], [1] asserts that the distribution of X is infinitely divisible if and only if there exist a symmetric nonnegative-definite $n \times n$ matrix Σ, $\gamma \in \mathbb{R}^n$ and a Lévy measure ν such that the characteristic function $\varphi(t)$ of X is given by

$$\varphi(t) = \exp\left\{-\frac{1}{2}\langle \Sigma t, t\rangle + \mathrm{i}\langle \gamma, t\rangle + \int_{\mathbb{R}^n}\left(\mathrm{e}^{\mathrm{i}\langle x,t\rangle} - 1 - \frac{\mathrm{i}\langle x.t\rangle}{1+|x|^2}\right)\nu(\mathrm{d}x)\right\},$$

for each $t \in \mathbb{R}^n$. Here, a measure ν on $\mathbb{R}^n$ is called a *Lévy measure* if it satisfies $\nu(\{0\}) = 0$ and $\int_{\mathbb{R}^n}(1 \wedge |x|^2)\nu(dx) < \infty$. The triplet (Σ, ν, γ) is called the Lévy-Khintchine triplet of X. We write $ID(\Sigma, \nu, \gamma)$ for the distribution of X, and use the abbreviation $ID(\nu) := ID(0, \nu, 0)$. Fixing $\gamma = 0$ is just fixing a location parameter and does not matter for the entropy, whereas fixing $\Sigma = 0$ means that the infinitely divisible measure has no Gaussian part.

We start with a one-dimensional result.

Corollary IX.3. *Let ν be any log-concave Lévy measure supported on $(0, \infty)$. Assume that the density $m(x)$ of ν satisfies $m(0+) \geq 1$. Then if f is the density of $ID(\nu)$,*

$$h(ID(\mu)) \leq 1 - \log \|f\|.$$

Proof: Yamazato [50] (see also Hansen [27] for an alternative proof) showed that for infinitely divisible measures supported on the positive real line, if the Lévy measure has a log-concave density m, then the density of the ID measure is log-concave if and only if $m(0+) \geq 1$. ∎

It is natural to ask how to bound the entropy $h(X)$ in terms of φ, especially when f is not given explicitly but φ is (which is typical in the case of infinitely divisible distributions). We show below that some explicit bounds may be given when we know something about convexity properties of the density. The idea is to utilize the Rényi entropy of order 2, since it is directly connected to the characteristic function by Plancherel's identity.

To start with, assume f is log-concave. Then by applying Corollary IV.2 with $p = 1$ and $q = 2$, one obtains

$$-\log \|f\|_2^2 \leq h(X) \leq n - \log \|f\|_2^2,$$

since by definition, $h_2(X) = -\log \|f\|_2^2$. Note that the lower bound here is universally true (for all random vectors X) as a consequence of Jensen's inequality; indeed, if X has density f,

$$h(X) = E[-\log f(X)] \geq -\log Ef(X) = h_2(X).$$

But Plancherel's formula asserts that $\|f\|_2^2 = (2\pi)^{-n}\|\varphi\|_2^2$. Hence:

Proposition IX.4. *Let X be a log-concave random vector in $\mathbb{R}^n$ with characteristic function $\varphi(t)$. Then*

$$n\log(2\pi) - \log \|\varphi\|_2^2 \leq h(X) \leq n\log(2\pi e) - \log \|\varphi\|_2^2,$$

where

$$\|\varphi\|_2^2 = \int |\varphi(t)|^2\, dt.$$

Equivalently,

$$\log(2\pi) \leq \frac{1}{n}h(X) + \|\varphi\|_2^{2/n} \leq \log(2\pi e).$$

This gives a reasonably strong approximation for the entropy of a log-concave distribution that is only known through its characteristic function: the gap between the upper and lower bounds is just 1.

One would also hope to be able to bound the entropies of the non-normal stable laws (which are not log-concave). As a step in this direction, we have a generalization of Proposition IX.4 to the κ-concave case.

Theorem IX.5. *If X has a κ-concave distribution,*

$$h_2(X) \leq h(X) \leq h_2(X) + \beta\sum_{i=1}^{n}\frac{1}{\beta - i}, \tag{26}$$

provided $\beta = n - \frac{1}{\kappa} \geq n + 1$.

The upper bound is easy to see using Theorem I.3 (more precisely, inequality (8)) and the first part of Theorem IV.1; the lower bound is as for Proposition IX.4.

Observe that $h_2(X)$ can be explicitly computed in many interesting cases via Plancherel's formula. For instance, for *one-dimensional* symmetric α-stable measures with characteristic function

$$\varphi(t) = \exp(-|t|^\alpha),$$

one obtains

$$\|\varphi\|_2^2 = 2\int_0^\infty \exp(-2t^\alpha)dt = \frac{1}{\alpha}\int_0^\infty \left(\frac{z}{2}\right)^{\frac{1-\alpha}{\alpha}} e^{-z}dz = \frac{2^{1-\frac{1}{\alpha}}}{\alpha}\Gamma\left(\frac{1}{\alpha}\right).$$

For the sake of illustration, let us apply our inequalities to approximating the entropy of the Cauchy distribution, which is also explicitly computable. Recall that the standard Cauchy distribution (stable index $\alpha = 1$, skewness parameter $\beta = 0$) has density

$$f(x) = \frac{1}{\pi(1+x^2)}, x \in \mathbb{R},$$

and entropy $\log(4\pi) \approx 1.386 + \log\pi$. It is easy to check that the Cauchy distribution is -1-concave (i.e., one can choose $\beta = 2$), so applying Theorem I.3 gives

$$h(X) \leq \log\pi + 2$$

since $\|f\| = 1/\pi$. On the other hand, $h(X) \geq h_2(X) = -\log[(2\pi)^{-1}\Gamma(1)] = \log(2\pi)$, so that $h(X)$ is trapped in a range of width $2 - \log 2 \approx 1.307$ centered at approximately $1.347 + \log\pi$, which seems fairly good.

For *multivariate* symmetric α-stable probability measures, a representation of the Rényi entropy of order 2 is obtained by Molchanov [43], in terms of the volume of a star body associated with the measure. In particular, [43] uses the identity

$$\int_{\mathbb{R}^n} f(x)^2 dx = 2^{-n/\alpha} f(0)$$

for any symmetric α-stable density f; as pointed out by a reviewer, the left side here is just the value of the self-convolution of f at 0, and for X, X' drawn independently from f, stability implies that $X + X'$ has the same distribution as $(1^\alpha + 1^\alpha)^{n/\alpha} X$. It is well known that symmetric stable random vectors are unimodal with mode at 0 (see, e.g., Kanter [33]); hence one can rewrite this as

$$\frac{1}{n} h_2(X) = \log \|f\|^{-1/n} + \frac{1}{\alpha} \log 2.$$

However, this still does not seem as useful as using Plancherel's formula to connect with the characteristic function.

It seems plausible that large classes of infinitely divisible distributions are κ-concave, although we do not know any existing general results in this direction (other than the log-concavity results mentioned earlier). If one were able to get estimates on κ for an infinitely divisible distribution that is specified through its characteristic function, (26) would immediately yield an upper bound for entropy in terms of $\|\varphi\|_2^2$.

Some negative results on κ-concavity of stable laws actually follow from the preceding discussion. Indeed, if X is symmetric α-stable and κ-concave, one has

$$\begin{aligned}\log \|f\|^{-1/n} + \frac{1}{\alpha}\log 2 = \frac{1}{n} h_2(X) &\leq \frac{1}{n} h(X) \\ &\leq \log \|f\|^{-1/n} + \frac{\beta}{n} \sum_{i=1}^{n} \frac{1}{\beta - i} \\ &\leq \log \|f\|^{-1/n} + (1 - \kappa n),\end{aligned}$$

where the last inequality follows from a similar calculation as in the proof of Corollary IX.2. Thus one obtains

$$\kappa \leq \left(1 - \frac{\log 2}{\alpha}\right) \frac{1}{n}.$$

This is rather loose, since we already know that for $\alpha < 2$, no symmetric α-stable distribution can be log-concave (as it would otherwise have finite moments). However, it does give some negative information for $\alpha < \log 2$. For instance, it shows that for fixed dimension n, symmetric α-stable distributions cease to be κ-concave for any fixed $\kappa \in (-\infty, 0)$ as $\alpha \to 0$. This leads us to the following conjecture.

Conjecture IX.6. *Any strictly stable probability measure on an infinite-dimensional separable Hilbert space is convex. In the finite-dimensional case, one has a threshold phenomenon: For fixed $\kappa \leq 0$, a spherically symmetric stable distribution of index α on $\mathbb{R}^n$ is κ-concave if and only if $\alpha \geq \alpha^*(\kappa, n)$, where $\alpha^*(\kappa, n) \in (0, 2]$ is a constant depending only on κ and n.*

Recall that a random element X in a separable Hilbert space is said to have a strictly stable distribution with index $\alpha \in (0, 2]$ if $X_1 + \ldots + X_m \stackrel{\mathcal{D}}{=} m^{1/\alpha} X$, where the X_i are independent random elements with the same distribution as X.

D. Entropy of Mixtures

Our fourth application is focused on estimating the entropy of scale mixtures of Gaussians (or more generally log-concave distributions). Such distributions are of great interest in Bayesian statistics.

Suppose one starts with a log-concave density $f = e^{-\varphi}$, where φ is convex. A scale mixture using a mixing distribution with density m on the positive real line would have the density

$$\int_0^\infty m(s) \frac{1}{s} \exp\left\{ -\varphi\left(\frac{x}{s}\right) \right\} ds.$$

More generally, one can consider "multivariate scale mixtures" of form

$$f_{\text{mix}}(x) = \int_{P(n)} m(A) f_A(x) \eta(dA),$$

where

$$f_A(x) = \frac{f(A^{-1}x)}{\det(A)}$$

is the density of AX when X is distributed according to f, and η represents the restriction of the Haar measure on the general linear group $GL(n)$ equipped with the multiplicative operation to the subset $P(n)$ (which is both a semigroup with respect to matrix multiplication and a cone, but not a group) of positive-definite matrices.

Note that lower bounds on entropy of mixtures are easy to obtain by using concavity of entropy, but upper bounds are in general difficult. Indeed,

$$\begin{aligned} h(f_{\text{mix}}) &\geq \int_0^\infty m(A) h(f_A) \eta(dA) \\ &= \int_0^\infty m(A)[h(f) + \log\det(A)] \eta(dA) \\ &= h(f) + \int_0^\infty m(A) \log\det(A) \eta(dA). \end{aligned}$$

On the other hand, one has an upper bound under a log-concavity assumption.

Theorem IX.7. *Suppose f_{mix} is a scale mixture of the log-concave density $f = e^{-\varphi}$, using a mixing distribution with density m on the positive-definite cone $P(n)$. Assume f has a mode at 0 (which for instance is the case when it is symmetric), and that f_{mix} is log-concave. Then*

$$h(f_{mix}) \leq n + \varphi(0) - \log \int_{P(n)} \frac{m(A)}{\det(A)} \eta(dA).$$

Proof: The proof is obvious from Proposition I.2 and the fact that the mixture density must also have its mode at 0. ■

The condition on Theorem IX.7 that the mixture be log-concave may not be too onerous to check, at least in the case of mixtures involving a one-dimensional scale, i.e., $A = sI$, with m now a prior on $\mathbb{R}_+$. A sufficient condition for f_{mix} to be log-concave is obtained by requiring that the integrand above is log-concave (thanks to the Prékopa-Leindler inequality), which means

$$\bar{\varphi}(x,s) := \varphi\left(\frac{x}{s}\right) - \log \frac{m(s)}{s}$$

is convex. For given m, this may be checked by verifying positive-definiteness of the $(n+1)\times(n+1)$ Hessian matrix of $\bar{\varphi}$ for $x \in \mathbb{R}^n, s \in (0,\infty)$.

One has an even simpler statement for Gaussian mixtures, which already appears to be new.

Corollary IX.8. *Suppose the mixing distribution with density m on the positive real line is slightly stronger than log-concave, in the sense that $\frac{n}{2}\log v - \log m(v)$ is convex. Writing Z for a standard Gaussian on $\mathbb{R}^n$ with density g, let the random vector $Y = \sqrt{V}Z$, where V is a scalar distributed according to m, have the density g_{mix}. Then*

$$\frac{n}{2}\int_0^\infty m(v)\log v\, dv \le h(g_{mix}) - h(g) \le \frac{n}{2} - \log \int_0^\infty \frac{m(v)}{v} dv.$$

Proof: One can write

$$g_{\text{mix}}(x) = \int_0^\infty \frac{m(v)}{(2\pi v)^{n/2}} \exp\left\{-\frac{\|x\|^2}{2v}\right\} dv,$$

since we are parametrizing using variance (rather than standard deviation). Also, $\|x\|^2/v$ is convex as a function of $(x,v) \in \mathbb{R}^n \times (0,\infty)$; indeed, the quadratic form induced by its Hessian matrix when evaluated at $(a,b) \in \mathbb{R}^n \times (0,\infty)$ is $\|ya - bx\|^2 \ge 0$. Combining this with the assumed log-concavity of $\frac{m(v)}{v^{n/2}}$, one finds that the integrand above is log-concave, and hence so is g_{mix} (by the Prékopa-Leindler inequality). Then the first inequality of Corollary IX.8 follows from concavity of entropy, while the second follows from Proposition I.2. ∎

A limitation (perhaps unavoidable) of this result is that as dimension increases, the shape requirement on the prior m becomes increasingly stringent.

X. Discussion

A central result in our development was the identification of the maximizer of Rényi entropy under log-concavity and supremum norm constraints. We gave a number of probabilistic, information theoretic and convex geometric motivations for considering this entropy maximization problem.

There are some other works in which both log-concavity and entropy appear, although they are only tangentially related to the substance of this paper. Log-concavity plays a role in a few other entropy bounding problems– see, for instance, Cover and Zhang [23] and Yu [52]. Log-concavity (in the discrete sense) also turns out to be relevant to the behavior of discrete entropy; see Johnson [29] and [30] for examples. For instance, Johnson [29] showed that the Poisson is maximum entropy among all ultra-log-concave distributions on the non-negative integers with fixed mean (ultra-log-concavity is a strengthening of discrete log-concavity).

For completeness, let us also mention that a different maximum entropy characterization of one-dimensional generalized Pareto distributions was given by Bercher and Vignat [7]. However, their characterization is rather different: in particular, they use Rényi and Tsallis entropies rather than Shannon entropy, and also it is not clear what the motivation is for the somewhat artificial moment and normalization constraints they impose. While [7] claims a connection to the Balkema-de Haans-Pickands theorem for limiting distribution of excesses over a threshold, log-concavity does not play a role in their development.

Our main goal in this paper was to better understand the behavior of entropy for log-concave (and more generally, hyperbolic) probability measures, particularly as regards phenomena that do not degrade in high dimensions. The information-theoretic perspective on convex geometry suggested in this paper appears to be bearing fruit; for instance, in [13], we use some of the results in this paper as one ingredient (among several) to prove a "reverse entropy power inequality for convex measures" analogous to Milman's reverse Brunn-Minkowski inequality [39], [41], [40], [45] for convex bodies.

We conclude with some open questions. First, the question of characterizing the κ-concavity properties of infinitely divisible laws using only knowledge of the characteristic function is an interesting one, as discussed in Section IX-C. One also hopes that Theorem I.3 can be improved to only require $\beta > n$; this would immediately imply that many of the results in this paper stated for κ-concave measures with $\kappa \ge -1$ (or their densities) would have extended validity to general convex measures. And finally, it would be nice to use the entropic formulation of the hyperplane conjecture to improve the state-of-the-art partial results that exist.

Appendix A
The multivariate Pareto distribution

There does not seem to be a canonical definition of a multivariate version for the Pareto distribution, although various versions appear to have been examined in the actuarial literature (see, e.g., [38], [51], [2]). For our purposes, the distribution with density $f_{\beta,a}$ defined in (6) is the relevant generalization. In this Appendix, we collect some simple observations about this multivariate Pareto family. Recall that

$$f_{\beta,a}(x) = \frac{1}{Z_n(\beta,a)}(a + x_1 + \cdots + x_n)^{-\beta}, \quad x_i > 0.$$

Lemma A.1. *For any $a > 0$, the normalizing factor*

$$Z_n(\beta,a) = \int_{\mathbb{R}_+^n} (a + x_1 + \cdots + x_n)^{-\beta}\, dx$$

is finite if and only if $\beta > n$. Moreover, for $\beta > n$,

$$Z_n(\beta,a) = \frac{1}{(\beta-1)\ldots(\beta-n)} \cdot \frac{1}{a^{\beta-n}}.$$

Proof: We prove the desired statement by induction. First,

$$Z_1(\beta,a) = \int_0^\infty (a+x)^{-\beta}\, dx = \int_a^\infty y^{-\beta}\, dy = \left.\frac{y^{1-\beta}}{1-\beta}\right|_a^\infty = \begin{cases} \infty & \text{if } \beta \le 1 \\ \frac{1}{\beta-1}\cdot\frac{1}{a^{\beta-1}} & \text{if } \beta > 1 \end{cases}$$

Now assume that the statement is true for Z_{n-1}, and observe that

$$
\begin{aligned}
Z_n(\beta,a) &= \int_0^\infty dx_n \int_{\mathbb{R}_+^{n-1}} (a+x_1+\cdots+x_n)^{-\beta}\, dx_1 \dots dx_{n-1} \\
&= \int_0^\infty \frac{1}{(\beta-1)\dots(\beta-n+1)} \cdot \frac{dx_n}{(a+x_n)^{\beta-n+1}} \\
&= \frac{1}{(\beta-1)\dots(\beta-n+1)} Z_1(\beta-n+1,a) \\
&= \frac{1}{(\beta-1)\dots(\beta-n)} \cdot \frac{1}{a^{\beta-n}},
\end{aligned}
$$

which is the required conclusion for Z_n. ∎

In particular, $f_{\beta,a}$ is not a well defined density for $\beta \le n$, and there is no Pareto distribution with such parameters.

Lemma A.2. *For any $a > 0$ and $\beta > n$, the entropy of the multivariate Pareto distribution $f_{\beta,a}$ is given by*

$$
h(f_{\beta,a}) + \log \|f_{\beta,a}\|^{-1} \;=\; \beta \sum_{i=1}^{n} \frac{1}{\beta-i}.
$$

Proof: If $Y \sim f_{\beta,a}$, then

$$
h(Y) = \log Z_n(\beta,a) + \frac{\beta}{Z_n(\beta,a)} L_n(\beta,a),
$$

where

$$
L_n(\beta,a) = \int_{\mathbb{R}_+^n} \frac{\log(a+x_1+\cdots+x_n)}{(a+x_1+\cdots+x_n)^\beta} dx.
$$

With this notation, what we wish to prove is that

$$
\frac{L_n(\beta,a)}{Z_n(\beta,a)} = \log a + \sum_{i=1}^{n} \frac{1}{\beta-i}. \tag{27}
$$

As in the proof of Lemma A.1, one can write the recursion

$$
L_n(\beta,a) = \int_0^\infty L_{n-1}(\beta, a+x_n)\, dx_n,
$$

and it is a simple exercise using integration by parts to see that

$$
L_1(\beta,a) = Z_1(\beta,a) \left[\frac{1}{\beta-1} + \log a \right]. \tag{28}
$$

Our goal is to prove the identity (27) by induction. To this end, we compute using the induction hypothesis for $n-1$:

$$
\begin{aligned}
L_n(\beta,a) &= \int_0^\infty Z_{n-1}(\beta,a+y) \left[\log(a+y) + \sum_{i=1}^{n-1} \frac{1}{\beta-i} \right] dy \\
&= \frac{1}{(\beta-1)\dots(\beta-n+1)} \int_0^\infty \frac{\log(a+y)}{(a+y)^{\beta-n+1}} dy \\
&\quad + Z_{n-1}(\beta,a) \sum_{i=1}^{n-1} \frac{1}{\beta-i}.
\end{aligned}
$$

Recognizing the integral in the last expression as $L_1(\beta-n+1, a)$ and plugging in the evaluation (28), simple manipulations give us (27). Observing that $\|f_{\beta,a}\|^{-1} = Z_n(\beta,a)a^\beta$, the proof of Lemma A.2 is complete. ∎

Acknowledgment

We thank each of the 3 referees for a careful reading and many useful comments.

References

[1] D. Applebaum, *Lévy Processes and Stochastic Calculus*. Cambridge University Press, July 2004.

[2] A. V. Asimit, E. Furman, and R. Vernic, "On a multivariate Pareto distribution," *Insurance Math. Econom.*, vol. 46, no. 2, pp. 308–316, 2010. [Online]. Available: http://dx.doi.org/10.1016/j.insmatheco.2009.11.004

[3] M. Avriel, "r-convex functions," *Math. Programming*, vol. 2, pp. 309–323, 1972.

[4] D. Bakry, F. Barthe, P. Cattiaux, and A. Guillin, "A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case," *Electron. Commun. Probab.*, vol. 13, pp. 60–66, 2008.

[5] D. Bakry and M. Émery, "Diffusions hypercontractives," in *Séminaire de probabilités, XIX, 1983/84*, ser. Lecture Notes in Math. Berlin: Springer, 1985, vol. 1123, pp. 177–206. [Online]. Available: http://dx.doi.org/10.1007/BFb0075847

[6] K. Ball, "Logarithmically concave functions and sections of convex sets in $\mathbf{R}^n$," *Studia Math.*, vol. 88, no. 1, pp. 69–84, 1988.

[7] J. Bercher and C. Vignat, "An entropic view of Pickands' theorem," in *Proc. IEEE Intl. Symp. Inform. Theory*, Toronto, Canada, 2008, pp. 2625–2628.

[8] L. Berwald, "Verallgemeinerung eines Mittelwertsatzes von J. Favard für positive konkave Funktionen," *Acta Math.*, vol. 79, pp. 17–37, 1947.

[9] R. Bhatia, *Matrix analysis*, ser. Graduate Texts in Mathematics. New York: Springer-Verlag, 1997, vol. 169.

[10] S. Bobkov, "Extremal properties of half-spaces for log-concave distributions," *Ann. Probab.*, vol. 24, no. 1, pp. 35–48, 1996. [Online]. Available: http://dx.doi.org/10.1214/aop/1042644706

[11] S. G. Bobkov, "Large deviations and isoperimetry over convex probability measures with heavy tails," *Electron. J. Probab.*, vol. 12, pp. 1072–1100 (electronic), 2007.

[12] ——, "Convex bodies and norms associated to convex measures," *Probab. Theory Related Fields*, vol. 147, no. 1-2, pp. 303–332, 2010. [Online]. Available: http://dx.doi.org/10.1007/s00440-009-0209-7

[13] S. G. Bobkov and M. Madiman, "Reverse Brunn-Minkowski and reverse entropy power inequalities for convex measures," *Preprint*, 2010.

[14] ——, "When can one invert Hölder's inequality? (and why one may want to)," *Preprint*, 2010.

[15] S. Bobkov, "Isoperimetric and analytic inequalities for log-concave probability measures," *Ann. Prob.*, vol. 27, no. 4, pp. 1903–1921, 1999.

[16] C. Borell, "Complements of Lyapunov's inequality," *Math. Ann.*, vol. 205, pp. 323–331, 1973.

[17] ——, "Convex measures on locally convex spaces," *Ark. Mat.*, vol. 12, pp. 239–252, 1974.

[18] ——, "Convex set functions in d-space," *Period. Math. Hungar.*, vol. 6, no. 2, pp. 111–136, 1975.

[19] J. Bourgain, "On high-dimensional maximal functions associated to convex bodies," *Amer. J. Math.*, vol. 108, no. 6, pp. 1467–1476, 1986.

[20] ——, "On the distribution of polynomials on high-dimensional convex sets," in *Geometric aspects of functional analysis (1989–90)*, ser. Lecture Notes in Math. Berlin: Springer, 1991, vol. 1469, pp. 127–137.

[21] H. J. Brascamp and E. H. Lieb, "On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation," *J. Functional Analysis*, vol. 22, no. 4, pp. 366–389, 1976.

[22] J. Costa, A. Hero, and C. Vignat, "On solutions to multivariate maximum alpha-entropy problems," *Lecture Notes in Computer Science*, vol. 2683, no. EMMCVPR 2003, Lisbon, 7-9 July 2003, pp. 211–228, 2003.

[23] T. M. Cover and Z. Zhang, "On the maximum entropy of the sum of two dependent random variables," *IEEE Trans. Inform. Theory*, vol. 40, no. 4, pp. 1244–1246, 1994.

[24] T. Cover and J. Thomas, *Elements of Information Theory*. New York: J. Wiley, 1991.

[25] M. Fradelizi, "Sections of convex bodies through their centroid," *Arch. Math. (Basel)*, vol. 69, no. 6, pp. 515–522, 1997.

[26] R. M. Gray, "Toeplitz and circulant matrices: A review," *Foundations and Trends in Communications and Information Theory*, 2006.

[27] B. G. Hansen, "On log-concave and log-convex infinitely divisible sequences and densities," *Ann. Probab.*, vol. 16, no. 4, pp. 1832–1839, 1988.

[28] D. Hensley, "Slicing convex bodies—bounds for slice area in terms of the body's covariance," *Proc. Amer. Math. Soc.*, vol. 79, no. 4, pp. 619–625, 1980.
[29] O. Johnson, "Log-concavity and the maximum entropy property of the Poisson distribution," *Stochastic Process. Appl.*, vol. 117, no. 6, pp. 791–802, 2007.
[30] O. Johnson, I. Kontoyiannis, and M. Madiman, "On the entropy and log-concavity of compound Poisson measures," *Preprint,* arXiv:0805:4112v1, 2008.
[31] O. Johnson and C. Vignat, "Some results concerning maximum Rényi entropy distributions," *Ann. Inst. H. Poincaré Probab. Statist.*, vol. 43, no. 3, pp. 339–351, 2007.
[32] M. Junge, "Volume estimates for log-concave densities with application to iterated convolutions," *Pacific J. Math.*, vol. 169, no. 1, pp. 107–133, 1995.
[33] M. Kanter, "Unimodality and dominance for symmetric random vectors," *Trans. Amer. Math. Soc.*, vol. 229, pp. 65–85, 1977.
[34] B. Klartag, "On convex perturbations with a bounded isotropic constant," *Geom. Funct. Anal.*, vol. 16, no. 6, pp. 1274–1290, 2006.
[35] L. Leindler, "On a certain converse of Hölder's inequality. II," *Acta Sci. Math. (Szeged)*, vol. 33, no. 3-4, pp. 217–223, 1972.
[36] E. Lutwak, D. Yang, and G. Zhang, "Moment-entropy inequalities," *Ann. Probab.*, vol. 32, no. 1B, pp. 757–774, 2004.
[37] ——, "Moment-entropy inequalities for a random vector," *IEEE Trans. Inform. Theory*, vol. 53, no. 4, pp. 1603–1607, 2007.
[38] K. V. Mardia, "Multivariate Pareto distributions," *Ann. Math. Statist.*, vol. 33, pp. 1008–1015, 1962.
[39] V. D. Milman, "Inégalité de Brunn-Minkowski inverse et applications à la théorie locale des espaces normés," *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 302, no. 1, pp. 25–28, 1986.
[40] ——, "Entropy point of view on some geometric inequalities," *C. R. Acad. Sci. Paris Sér. I Math.*, vol. 306, no. 14, pp. 611–615, 1988.
[41] ——, "Isomorphic symmetrizations and geometric inequalities," in *Geometric aspects of functional analysis (1986/87)*, ser. Lecture Notes in Math. Berlin: Springer, 1988, vol. 1317, pp. 107–131.
[42] V. D. Milman and A. Pajor, "Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space," in *Geometric aspects of functional analysis (1987–88)*, ser. Lecture Notes in Math. Berlin: Springer, 1989, vol. 1376, pp. 64–104.
[43] I. Molchanov, "Convex and star-shaped sets associated with multivariate stable distributions, i : Moments and densities," *J. Multivariate Anal.*, 2009.
[44] G. Paouris, "On the isotropic constant of non-symmetric convex bodies," in *Geometric aspects of functional analysis*, ser. Lecture Notes in Math. Berlin: Springer, 2000, vol. 1745, pp. 239–243.
[45] G. Pisier, *The volume of convex bodies and Banach space geometry*, ser. Cambridge Tracts in Mathematics. Cambridge: Cambridge University Press, 1989, vol. 94.
[46] A. Prékopa, "Logarithmic concave measures with application to stochastic programming," *Acta Sci. Math. (Szeged)*, vol. 32, pp. 301–316, 1971.
[47] ——, "On logarithmic concave measures and functions," *Acta Sci. Math. (Szeged)*, vol. 34, pp. 335–343, 1973.
[48] K.-I. Sato, *Levy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999.
[49] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, pp. 101–112, 1959.
[50] M. Yamazato, "On strongly unimodal infinitely divisible distributions," *Ann. Probab.*, vol. 10, no. 3, pp. 589–601, 1982.
[51] H.-C. Yeh, "Some properties and characterizations for generalized multivariate Pareto distributions," *J. Multivariate Anal.*, vol. 88, no. 1, pp. 47–60, 2004. [Online]. Available: http://dx.doi.org/10.1016/S0047-259X(03)00061-7
[52] Y. Yu, "Letter to the editor: On an inequality of Karlin and Rinott concerning weighted sums of i.i.d. random variables," *Adv. in Appl. Probab.*, vol. 40, no. 4, pp. 1223–1226, 2008.