Functional Analysis/Probability Theory

Dimensional behaviour of entropy and information

Comportement dimensionnel de l'entropie et de l'information

Sergey Bobkov\textsuperscript{a}, Mokshay Madiman\textsuperscript{b}

\textsuperscript{a} School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA
\textsuperscript{b} Department of Statistics, Yale University, New Haven, CT 06511, USA

\textbf{Article info}

Article history:
Received 13 October 2010
Accepted after revision 5 January 2011
Available online 22 January 2011
Presented by Gilles Pisier

\section{1. Introduction}

This note announces some of the results obtained in [3--5]. Given a random vector $X$ in $\mathbb{R}^n$ with density $f(x)$, the entropy power is defined by $N(X) = e^{2h(X)/n}$, where, with a common abuse of notation, we write $h(f)$ for the Shannon entropy $h(f) := -\int_{\mathbb{R}^n} f \log f$.

\textbf{Theorem 1.1.} If $X$ and $Y$ are independent random vectors in $\mathbb{R}^n$ with log-concave densities, there exist affine entropy-preserving maps $u_j : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$N(\tilde{X} + \tilde{Y}) \leq C(N(X) + N(Y)),$$

where $\tilde{X} = u_1(X)$, $\tilde{Y} = u_2(Y)$, and where $C$ is a universal constant.

Observe that the Shannon–Stam entropy power inequality [15] implies that $N(\tilde{X} + \tilde{Y}) \geq N(X) + N(Y)$ is always true. Thus Theorem 1.1 may be seen as a reverse entropy power inequality for log-concave measures. The proof of this assertion,
Corollary 2.2. Superlevel sets of log-concave densities has been anticipated before, e.g., by [9], but Theorem 2.1 refines those observations of the uniformity of the distribution of $X$ of $X$. The current best bound known, due to Klartag [8], is equivalent formulations of the conjecture, all of a geometric or functional analytic flavor (even the ones that nominally the theoretic formulation of the hyperplane conjecture. For a random vector $X$ in $\mathbb{R}^n$ with density $f$, let $D(X)$ or $D(f)$ denote

2. Intermediate results

2.1. An equipartition property

Let $X$ be a random vector taking values in $\mathbb{R}^n$, and suppose its distribution has a density $f$ with respect to Lebesgue measure on $\mathbb{R}^n$. The random variable $h(X) = -\log f(X)$ may be thought of as the “information content” of $X$. Note that the entropy is $h(X) = \mathbb{E} h(X)$.

Because of the relevance of the information content in information theory, probability, and statistics, it is intrinsically interesting to understand its behavior. In particular, a natural question arises: Is it true that the information content concentrates around the entropy in high dimension? In general, there is no reason for such a concentration property to hold. However, the following proposition shows that in fact, such a property holds uniformly for the entire class of log-concave densities:

**Theorem 2.1.** If $X$ has a log-concave density $f$ on $\mathbb{R}^n$, then for $0 < \varepsilon \leq 2$,

$$
P\left(\frac{\tilde{h}(X) - h(X)}{n} \geq \varepsilon\right) \leq 4e^{-\varepsilon^2n/16}.
$$

No normalization whatsoever is required for this result, which is proved in [3] using the localization lemma of Lovász–Simonovits, and certain reverse Hölder type inequalities for log-concave measures.

Equivalently, with high probability, $f(x)^{2/n}$ is very close to the entropy power $N(f) = \exp\left(\frac{1}{2} h(X)\right)$, and the distribution of $X$ itself is effectively the uniform distribution on the class of typical observables, or the “typical set” (defined to be the collection of all points $x \in \mathbb{R}^n$ such that $f(x)$ lies between $e^{-(h(X)-\varepsilon n)}$ and $e^{-(h(X)+\varepsilon n)}$, for some small fixed $\varepsilon > 0$). The effective uniformity of the distribution of $X$ on some compact set, entailed by this concentration result, may be seen as an extension of the asymptotic equipartition property (or Shannon–McMillan–Breiman theorem) to non-stationary stochastic processes with log-concave marginals (cf. [3]).

If one is more interested in the effective support rather than an effective uniformity, one can simply consider a superlevel set (necessarily convex and compact) of the density $f$ instead of the annular region above. This effective support on a convex set implied by Theorem 2.1 allows (see [1]) the transference of some results from the setting of convex bodies to that of log-concave measures, in particular, the existence of $M$-ellipsoids [11–14]. (Such a transference technique based on looking at superlevel sets of log-concave densities has been anticipated before, e.g., by [9], but Theorem 2.1 refines those observations and identifies the underlying concentration phenomenon.)

**Corollary 2.2.** Let $\mu$ be a probability measure on $\mathbb{R}^n$ with log-concave density $f$ such that $\|f\|_\infty \geq 1$ (where $\|f\|_\infty$ is the essential supremum and hence the maximum of $f$). Then there exists an ellipsoid $\mathcal{E}$ of volume 1 such that $\mu(\mathcal{E})^{1/n} \geq c_M$ for some universal constant $c_M \in (0, 1)$.

Equivalently, for some linear volume-preserving map $u : \mathbb{R}^n \to \mathbb{R}^n$, $\mu^{-1}(D)^{1/n} \geq c_M$, where $D$ is the Euclidean ball of volume one.

2.2. Entropy and the maximal density value

Trivially $h(X) \geq \log \|f\|_\infty^{-1}$. In fact, one can also bound the entropy from above using the maximal density value under log-concavity (see [4]).

**Theorem 2.3.** If a random vector $X$ in $\mathbb{R}^n$ has log-concave density $f$, then

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

The hyperplane conjecture or slicing problem (cf. Bourgain [6] or Ball [1]) asserts that 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 $H$ passing through its centroid 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 (even the ones that nominally use probability). The current best bound known, due to Klartag [8], is $\Omega(n^{-1/4})$. Theorem 2.3 gives a purely information-theoretic formulation of the hyperplane conjecture. For a random vector $X$ in $\mathbb{R}^n$ with density $f$, let $D(X)$ or $D(f)$ denote
its relative entropy from Gaussianity (which is the relative entropy from the Gaussian \( g \) with the same mean and covariance matrix, and also equals the difference \( h(g) - h(f) \)). The Entropic Form of the Hyperplane Conjecture [4] asserts that for any log-concave density \( f \) on \( \mathbb{R}^n \), \( D(f) \leq cn \) for some universal constant \( c \). It is easy to see then that another equivalent form of the hyperplane conjecture is that the entropic distance from independence (i.e., the relative entropy of any log-concave measure from the product of its marginals) is also bounded by \( cn \) for some universal constant \( c \). As an aside, Klartag’s result combined with our equivalence implies that \( D(f) \leq \frac{1}{4}n \log n + cn \) for any log-concave \( f \). This is already the first quantitative demonstration of the spiritual closeness of log-concave measures to Gaussians, which has been observed in qualitative ways numerous times (e.g., behavior as regards functional inequalities). Let us note en passant that entropy plays a role in Ball’s [2] proof that the KLS conjecture implies the hyperplane conjecture.

3. Proof outline of Theorem 1.1

The following “submodularity” property of the entropy functional with respect to convolutions was obtained in [10]:

\[
\int f + g \leq \int f + \int g
\]

given by universal constants. The following “submodularity” property of the entropy functional with respect to convolutions was obtained in [10]:

\[
\int f + g \leq \int f + \int g
\]

provided that all entropies are well-defined. Let \( Z \sim \text{Unif}(D) \), where \( D \) is the centered Euclidean ball with volume one. Since \( h(Z) = 0 \), the submodularity property implies

\[
\int h(X + Y + Z) + h(Z) \leq \int h(X + Z) + h(Y + Z)
\]

for random vectors \( X \) and \( Y \) in \( \mathbb{R}^n \) independent of each other and of \( Z \).

Let \( X \) and \( Y \) have log-concave densities. Due to homogeneity of Theorem 1.1, assume without loss of generality that \( \| f \|_{\infty} \geq 1 \) and \( \| g \|_{\infty} \geq 1 \). Then, our task reduces to showing that both \( \mathcal{N}(X + Z) \) and \( \mathcal{N}(Y + Z) \) can be bounded from above by universal constants.

By Corollary 2.2, for some affine volume preserving map \( u : \mathbb{R}^n \rightarrow \mathbb{R}^n \), the distribution \( \tilde{\mu} \) of \( \tilde{X} = u(X) \) satisfies \( \tilde{\mu}(D)^{1/n} \geq c_M \) with a universal constant \( c_M > 0 \). Let \( f \) denote the density of \( \tilde{X} = u(X) \). Then the density \( p \) of \( S = \tilde{X} + Z \), given by

\[
p(x) = \int f(x - z) dz = \tilde{\mu}(D - x),
\]

satisfies \( \| p \| \geq \| p(0) \| \geq c_M \). Applying Theorem 2.3 to the random vector \( S \), \( \mathcal{N}(S) \leq C \| p \|^{-2/n} \leq C \cdot c_M^{-2} \), which completes the proof.

Remark 1. Recall C. Borell’s hierarchy of convex measures on \( \mathbb{R}^n \), classified by a parameter \( \kappa \in [-\infty, 1/n] \). In this hierarchy, \( \kappa = 0 \) corresponds to the class of log-concave measures. When \( \kappa > 0 \), a \( \kappa \)-concave probability measure is necessarily compactly supported on some convex set.

For any random vector \( X \) with values in \( A \), there is a general upper bound \( h(X) \leq \log |A| \). Using Berwald’s inequality, we provide a complementary estimate from below depending only on the “strength” of convexity of the density \( f \) of \( X \): Let \( X \) be a random vector in \( \mathbb{R}^n \) having an absolutely continuous \( \kappa \)-concave distribution supported on a convex body \( A \) with \( 0 < \kappa \leq 1/n \). Then \( h(X) \geq \log |A| + n \log(\kappa n) \). Note when \( \kappa = 1/n \), this bound is sharp.

Assume a probability measure \( \mu \) is \( \kappa' \)-concave on \( \mathbb{R}^n \) and a probability measure \( \nu \) is \( \kappa'' \)-concave on \( \mathbb{R}^n \). If \( \kappa', \kappa'' \in [-1, 1] \) satisfy

\[
\kappa' + \kappa'' > 0, \quad \frac{1}{\kappa} = \frac{1}{\kappa'} + \frac{1}{\kappa''},
\]

then their convolution \( \mu * \nu \) is \( \kappa \)-concave. Hence, if random vectors \( X_1 \) and \( X_2 \) are independent and uniformly distributed in convex bodies \( A_1 \) and \( A_2 \) in \( \mathbb{R}^n \), then the sum \( X_1 + X_2 \) has a \( \frac{1}{\kappa} \)-convex distribution supported on the convex body \( A_1 + A_2 \). The preceding entropy bound then implies that \( h(X_1 + X_2) \geq \log |A_1 + A_2| - n \log 2 \). This immediately allows one to deduce Milman’s reverse Brunn–Minkowski inequality from Theorem 1.1.

Remark 2. Theorems 1.1 and 2.3 have been extended to the larger class of convex measures [5,4].

References

[1] K. Ball, Logarithmically concave functions and sections of convex sets in \( \mathbb{R}^n \). Studia Math. 88 (1) (1988) 69–84.
[2] K.M. Ball, Information decrease along semigroups, Talk given at conference on Banach Spaces and Convex Geometric Analysis, Universität Kiel, Germany, April 2003.
[3] S.G. Bobkov, M. Madiman, Concentration of the information in data with log-concave distributions, Ann. Probab., in press, arXiv:1012.5457v1 [math.PR].
[4] S.G. Bobkov, M. Madiman, The entropy per coordinate of a random vector is highly constrained under convexity conditions, preprint, arXiv:1006.2883v2 [cs.IT].
[5] S.G. Bobkov, M. Madiman, Reverse Brunn–Minkowski and reverse entropy power inequalities for convex measures, preprint.
[6] J. Bourgain, On high-dimensional maximal functions associated to convex bodies. Amer. J. Math. 108 (6) (1986) 1467–1476.
[7] A. Dembo, T. Cover, J. Thomas, Information-theoretic inequalities, IEEE Trans. Inform. Theory 37 (6) (1991) 1501–1518.
[8] B. Klartag, On convex perturbations with a bounded isotropic constant, Geom. Funct. Anal. 16 (6) (2006) 1274–1290.
[9] B. Klartag, V.D. Milman, Geometry of log-concave functions and measures, Geom. Dedicata 112 (2005) 169–182.
[10] M. Madiman, On the entropy of sums, in: Proc. IEEE Inform. Theory Workshop, Porto, Portugal, 2008, pp. 303–307.
[11] 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. 302 (1) (1986) 25–28.
[12] V.D. Milman, Isomorphic symmetrizations and geometric inequalities, in: Geometric Aspects of Functional Analysis (1986/87), Lecture Notes in Math., vol. 1317, Springer, Berlin, 1988, pp. 107–131.
[13] V.D. Milman, Entropy point of view on some geometric inequalities, C. R. Acad. Sci. Paris Sér. I Math. 306 (14) (1988) 611–615.
[14] G. Pisier, The Volume of Convex Bodies and Banach Space Geometry, Cambridge Tracts in Mathematics, vol. 94, Cambridge University Press, Cambridge, 1989.
[15] A. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon, Inform. Control 2 (1959) 101–112.