Fifth International Workshop on
"Convex Geometry – Analytic Aspects"
Cortona, Italy, June 12–18, 2011

INFORMATION-THEORETIC EXTENSIONS IN HIGH-DIMENSIONAL CONVEX GEOMETRY

Sergey G. Bobkov

joint work with
Mokshay Madiman

1 Entropy functional

$X = (X_1, \dots, X_n)$ a random vector in $\mathbf{R}^n$ with density $f(x)$

Entropy (differential entropy, Shannon entropy):

$$h(X) = -\int f(x) \log f(x)\, dx.$$

Entropy power (exponentiated entropy, "effective variance"):

$$H(X) = \exp\left[\frac{2}{n}\, h(X)\right].$$

Information content:

$$\tilde{h}(X) = -\log f(X).$$

Connections and similarities between Information Theory and other fields (Convex Geometry, Matrix Analysis, Probability, Sobolev inequalities):

Costa and Cover (1984) "On the similarity of the entropy power inequality and the Brunn-Minkowski inequality"

Dembo, Cover and Thomas (1991) "Information theoretic inequalities"

Johnson (2004) "Information theory and the central limit theorem"

Linnik (1959) "An information-theoretic proof of the central limit theorem with the Lindeberg condition"

2 A few basic properties of the entropy

- For any affine volume preserving map $T : \mathbf{R}^n \to \mathbf{R}^n$,

$$h(TX) = h(X), \qquad H(TX) = H(X).$$

- $H(\lambda X) = \lambda^2 H(X) \ \ (\lambda \neq 0)$.
- $\mathbf{E}\,\tilde{h}(X) = h(X)$.
- For any invertible affine map $T : \mathbf{R}^n \to \mathbf{R}^n$

$$\tilde{h}(TX) - h(TX) = \tilde{h}(X) - h(X).$$

- (Monotonicity) If X and Y and independent, then

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

- (Subadditivity) If $X = (X_1, \ldots, X_n)$, then

$$h(X_1, \ldots, X_n) \leq h(X_1) + \ldots + h(X_n)$$

with equality when X_j are independent.

3 Examples

Uniform distribution: If X is uniformly distributed in a convex body $A \subset \mathbf{R}^n$, then

$$h(X) = \log |A|, \qquad H(X) = |A|^{2/n},$$

where $|A|$ stands for the n-dimensional volume of A.

General property: $X \in A \subset \mathbf{R}^n \ \Rightarrow$

$$h(X) \leq \log |A|, \qquad H(X) \leq |A|^{2/n}.$$

Normal (Gaussian) distribution on R: If $X \sim N(a, \sigma^2)$ $a = \mathbf{E}X$, $\sigma^2 = \mathrm{Var}(X)$. Then

$$h(X) = \log \sqrt{2\pi e\, \sigma^2}, \qquad H(X) = 2\pi e\, \sigma^2.$$

Normal distribution on $\mathbf{R}^n$: If $X \sim N(a, R)$, $R = \mathrm{cov}(X)$,

$$h(X) = \log \sqrt{(2\pi e)^n \det(R)}, \qquad H(X) = 2\pi e \det(R)^{1/n}.$$

Equivalently: If X has density f with $\|f\| = \mathrm{ess\,sup}_x f(x)$, then

$$h(X) \ = \ -\log \|f\| + \frac{n}{2}$$

4 Brunn-Minkowski and entropy power inequalities

Theorem (Brunn-Minkowski):

$$|A+B|^{1/n} \geq |A|^{1/n} + |B|^{1/n},$$

where $|A|$ stands for the n-dimensional volume of A.

Theorem (Shannon 1948, Stam 1959, Lieb 1975). Given independent random vectors X and Y in $\mathbf{R}^n$ with finite entropies,

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

Equality: Iff X and Y are normal, with proportional covariance matrices.

The case of normal distributions:
Minkowski inequality for positive definite matrices. If $X \sim N(a,R)$ and $Y \sim N(b,S)$, then

$$\det(R+S)^{1/n} \geq \det(R)^{1/n} + \det(S)^{1/n}.$$

The case of uniform distribution: If $X \sim U(A)$, $Y \sim U(B)$, then $X+Y$ is not uniform in $A+B$. Nevertheless,

$$\frac{1}{4}\,|A+B|^{2/n} \leq \; H(X+Y) \leq \; |A+B|^{2/n}.$$

5 Reverse Brunn-Minkowski inequality

Brunn-Minkowski inequality:

$$|A+B|^{1/n} \geq |A|^{1/n} + |B|^{1/n}.$$

Question. How sharp is it? For any linear volume preserving maps $u_i : \mathbf{R}^n \to \mathbf{R}^n$, we still have

$$|\widetilde{A}+\widetilde{B}|^{1/n} \geq |A|^{1/n} + |B|^{1/n},$$

where $\widetilde{A} = u_1(A)$, $\widetilde{B} = u_2(B)$.

Theorem (V. D. Milman, mid 80s): Given convex bodies A and B in $\mathbf{R}^n$, for some linear volume preserving maps $u_i : \mathbf{R}^n \to \mathbf{R}^n$,

$$|\widetilde{A}+\widetilde{B}|^{1/n} \leq C \left(|A|^{1/n} + |B|^{1/n}\right),$$

where C is a universal constant.

Equivalently: If $|A| = |B| = 1$, then

$$2 \leq |\widetilde{A}+\widetilde{B}|^{1/n} \leq C.$$

Question. Entropic formulation / extension?

6 Reverse entropy power inequality

Let X and Y be independent random vectors in $\mathbf{R}^n$ with densities $f(x)$ and $g(x)$. We have the entropy power inequality:

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

Theorem 1. If X and Y have log-concave distributions, then for some linear volume preserving maps $u_i : \mathbf{R}^n \to \mathbf{R}^n$,

$$H(\widetilde{X}+\widetilde{Y}) \leq C\,(H(X) + H(Y))$$

where $\widetilde{X} = u_1(X)$, $\widetilde{Y} = u_2(Y)$.

Equivalently: If $\|f\| = \|g\| = 1$, then

$$C_0 \leq H(\widetilde{X}+\widetilde{Y}) \leq C_1,$$

with some absolute constants $C_1 > C_0 > 0$.

Corollary. If $X \sim U(A)$, $Y \sim U(B)$, then $H(X) = |A|^{2/n}$, $H(Y) = |B|^{2/n}$, while

$$H(\widetilde{X}+\widetilde{Y}) \geq \frac{1}{4}\,|\widetilde{A}+\widetilde{B}|^{2/n}.$$

Hence,

$$|\widetilde{A}+\widetilde{B}|^{2/n} \leq C\,\left(|A|^{2/n} + |B|^{2/n}\right).$$

7 Convex measures

Definition (Borell 1974). A Radon probability measure μ on a locally convex space L is κ-concave, where $-\infty \leq \kappa \leq 1$, if

$$\mu(tA + sB) \ \geq \ (t\mu(A)^\kappa + s\mu(B)^\kappa)^{1/\kappa}$$

for all non-empty $A, B \subset L$ and $t, s > 0$, $t + s = 1$.

Characterization on $L = \mathbf{R}^n$ in the absolut. continuous case (necessarily $\kappa \leq 1/n$):

$$\mu \text{ is } \kappa\text{−concave} \iff \mu \text{ has a } \kappa_n\text{−concave density } f,$$

that is, for all x, y from a convex supporting set $\Omega \subset \mathbf{R}^n$, and all $t, s > 0$, $t + s = 1$,

$$f(tx + sy) \geq (tf(x)^{\kappa_n} + sf(y)^{\kappa_n})^{1/\kappa_n}$$

where

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

Particular cases of κ's
1) $\kappa = 1/n$, $\kappa_n = +\infty$: $\mu = U(K)$ (B-M inequality)
2) $\kappa = \kappa_n = 0$: Log-concave measures (Prékopa 1971);

$$\mu(tA + sB) \geq \mu(A)^t \mu(B)^s, \qquad f(tx + sy) \geq f(x)^t f(y)^s.$$

3) $\kappa = -\infty$, $\kappa_n = -1/n$: Convex (or hyperbolic) measures;

$$\mu(tA + sB) \geq \min\{\mu(A), \mu(B)\}.$$

Equivalently: $f = V^{-n}$, for some positive convex V on an open convex set $\Omega \subset \mathbf{R}^n$.

Range $-\infty < \kappa < 0$: densities have the form

$$f(x) = V(x)^{-\beta}, \qquad \beta > n, \ \kappa = -\frac{1}{\beta - n},$$

where $V : \mathbf{R}^n \to (-\infty, +\infty]$ are convex.

Examples

1) Gaussian measures, exponential measures, $\kappa = 0$.
2) Cauchy measures, $\kappa = -1/d$ ($d > 0$ real),

$$f(x) = \frac{1}{Z}\,(1 + |x|^2)^{-(n+d)/2}.$$

3) Pareto distributions on $\mathbf{R}_+^n$, $\kappa = -1/d \quad (d > 0$ real),

$$f(x) = \frac{1}{Z}\,(1 + x_1 + \ldots + x_n)^{-(n+d)}.$$

Theorem 2. Let $\beta_0 > 2$. If X and Y have κ-concave distributions with

$$\beta \geq \max\{\beta_0 n,\, 2n + 1\},$$

then for some linear volume preserving maps $u_i : \mathbf{R}^n \to \mathbf{R}^n$,

$$H(\widetilde{X} + \widetilde{Y}) \leq C_{\beta_0}\,(H(X) + H(Y))$$

where $\widetilde{X} = u_1(X)$, $\widetilde{Y} = u_2(Y)$.

Note. The reverse entropy power inequality is not uniform in the class of all convex measures.

8 Entropy and volume of the support

Let X and Y be independent random vectors in $\mathbf{R}^n$.

Lemma 1. (Borell 1974) If the distributions of X and Y are κ_1- and κ_2-concave, $\kappa_1, \kappa_2 > 0$, then the distribution of $X + Y$ is κ-concave with

$$\frac{1}{\kappa} = \frac{1}{\kappa_1} + \frac{1}{\kappa_2}.$$

Lemma 2. Let X be a random vector in $\mathbf{R}^n$ having an absolutely continuous κ-concave distribution supported on a convex body A, with $0 < \kappa \leq 1/n$. Then

$$\log|A| + n\log(\kappa n) \ \leq\ h(X) \ \leq\ \log|A|.$$

Proof. Apply Berwald's inequality in the form of Borell (Khinchin-type inequality for concave functions).

Example

If X and Y are independent and uniformly distributed in convex bodies A and B in $\mathbf{R}^n$ in which case $\kappa_1 = \kappa_2 = 1/n$, then $X+Y$ has a κ-concave distribution with $\kappa = 1/(2n)$, so

$$\log\left|\frac{A+B}{2}\right| \leq\ h(X+Y) \leq\ \log|A+B|,$$

$$\frac{1}{2}\,|A+B|^{1/n} \leq\ H(X+Y)^{1/2} \leq\ |A+B|^{1/n}.$$

9 Entropy and maximum of density

Assume X has a convex distribution with density

$$f(x) = V(x)^{-\beta}, \quad \beta > n,$$

where V is a positive convex function on $\mathbf{R}^n$. Let

$$\|f\| = \sup_x f(x).$$

Lemma 3. If $\beta \geq n+1$, and if $\|f\|$ is fixed, the entropy $h(X)$ is maximal for the n-dimensional Pareto distribution. Equivalently,

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

Range $\beta \geq \beta_0 n$ with fixed $\beta_0 > 1$:

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

with some constant C_{β_0}, depending on β_0, only.

10 Log-concave case

If X has a log-concave distribution ($\beta = +\infty$),

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

that is,

$$0 \le \frac{1}{n} h(X) + \frac{1}{n} \log \|f\| \le 1.$$

Equivalently, if Z has a normal distribution with maximum of the density the same as for the density of X, then

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

Equality on the left: Uniform distribution in a convex body
Equality on the right: The exponential distribution

Proof. Given $t, s > 0$, $t + s = 1$, write

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

Integrate with respect to x and then maximize over y:

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

Let $t \to 1$.

11 Concentration of the information content

Let X have a log-concave density $f(x)$ on $\mathbf{R}^n$.

Theorem (Klartag-Milman). With some absolute $c_0, c_1 > 0$

$$\mathbf{P}\left\{ f(X) \geq \|f\|\, e^{-c_0 n} \right\} \,\geq\, 1 - e^{-c_1 n}.$$

Thus, with very high probability

$$-\log f(X) + \log \|f\| \leq c_0 n.$$

On the other hand,

$$0 \leq h(X) + \log \|f\| \leq n.$$

Informal conclusion:

$$\frac{1}{n}\, \tilde{h}(X) \equiv -\frac{1}{n}\, \log f(X)$$

is strongly concentrated around its mean $\frac{1}{n}\, h(X)$.

12 Large deviations for the information content

Given $X = (X_1, \dots, X_n)$ with density $f(x)$ on $\mathbf{R}^n$, define

$$\begin{aligned} \Delta(X) &= \tilde{h}(X) - h(X) \\ &= -\log f(X) + \mathbf{E} \log f(X). \end{aligned}$$

Notes
1) $\Delta(TX) = \Delta(X)$, for any invertible affine map T.
2) If X_i are independent,

$$\Delta(X) = \Delta(X_1) + \ldots + \Delta(X_n).$$

3) For X standard normal,

$$\Delta(X) = \sum_{i=1}^{n} \frac{X_i^2 - 1}{2}.$$

Theorem 3. If f is log-concave, then for all $t > 0$, with some absolute $c > 0$

$$\mathbf{P}\left\{ \frac{1}{\sqrt{n}}\, |\Delta(X)| \geq t \right\} \leq 2\, e^{-ct}.$$

Moreover, for $0 < t \leq \sqrt{n}$,

$$\mathbf{P}\left\{ \frac{1}{\sqrt{n}}\, |\Delta(X)| \geq t \right\} \leq 2\, e^{-ct^2}.$$

13 M-ellipsoids

Equivalent form of the reverse Brunn-Minkowski inequality:

Theorem (V.D.Milman). For any symmetric convex body A in $\mathbf{R}^n$ with $|A| = 1$, there is an ellipsoid $\mathcal{E}$, $|\mathcal{E}| = 1$, such that

$$|A \cap \mathcal{E}|^{1/n} \geq c,$$

where $c > 0$ is an absolute constant.

The ellipsoid $\mathcal{E}$ is called an M-ellipsoid.
If $\mathcal{E}$ is the ball, A is said to be in regular or M-position.

Theorem. Given a symmetric convex body A in $\mathbf{R}^n$ with volume $|A| = 1$, let γ_A denote the restricted Gaussian measure, with density

$$\frac{d\gamma_A(x)}{dx} = \frac{1}{Z}\, e^{-|x|^2/2}\, 1_A(x).$$

If γ_A is isotropic, then A is in M-position.

14 Log-concave measures in M-position

Definition. Let μ be a probability measure on $\mathbf{R}^n$ with density $f(x)$ such that $\|f\| = \operatorname{ess\,sup} f(x) = 1$. We say that μ is in M-position (with constant $c > 0$), if

$$\mu(D)^{1/n} \geq c,$$

for some Euclidean ball D of volume one.

Corollary. Any symmetric probability measure μ on $\mathbf{R}^n$ with log-concave density f, $\|f\| = 1$, can be put in M-position.

That is, for some linear volume preserving map $u : \mathbf{R}^n \to \mathbf{R}^n$, the image $u(\mu) = \mu\, u^{-1}$ is in M-position.

Proof. Choose $\lambda = e^{-c_0 n}$ such that the symmetric convex body

$$A = \{x : f(x) > \lambda\}$$

has μ-measure at least 1/2 (by using Klartag-Milman's theorem). Then

$$1 \geq \int_A f \geq \lambda\, |A| \ \Rightarrow\ |A| \leq e^{c_0 n}.$$

Also

$$1 \leq 2\mu(A) \leq 2\,|A| \ \Rightarrow\ |A| \geq 1/2.$$

Then apply Milman's theorem to $A/|A|$.

15 Submodularity of entropy

Theorem (M.Madiman). Given independent random vectors X, Y, Z in $\mathbf{R}^n$ with densities,

$$h(X+Y+Z) + h(Z) \leq h(X+Z) + h(Y+Z).$$

Particular case. If Z is uniformly distributed in the Euclidean ball of volume one, then $h(Z) = 0$ and

$$h(X+Y) \leq h(X+Z) + h(Y+Z).$$

Lemma. Let X have a symmetric log-concave density f on $\mathbf{R}^n$ with $\|f\| = 1$. If the distribution μ of X is in M-position, then

$$h(X+Z) \leq Cn.$$

Proof. Let $g = 1_D$, so that $f * g(x) = \int_D f(x-y)\,dy$ has norm

$$\|f * g\| = f * g(0) = \int_D f(y)\,dy = \mu(D) \geq c^n.$$

Reminder: If X has a log-concave distribution,

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

Applying this to $X+Z$, we have

$$h(X+Z) \leq n - \log \|f * g\|.$$

16 Proof of reverse entropy power inequality

Let X, Y be independent random vectors in $\mathbf{R}^n$ with log-concave densities f, g, such that $\|f\| = \|g\| = 1$.

Symmetric case. By the above theorem and lemma, if X and Y are in M-position then

$$h(X+Y) \le 2Cn,$$

that is,

$$H(X+Y) \le e^{4C}.$$

In general $h(X) \ge n \log \|f\|^{-1/n} = 0 \;\Rightarrow\; H(X) \ge 1,$ so

$$H(X+Y) \;\le\; e^{4C}\,(H(X) + H(Y)).$$

Non-symmetric case. Use symmetrization.

17 Relative entropy

Let X be a random vector in $\mathbf{R}^n$ with $\mathbf{E}\,|X|^2 < +\infty$, with density $f(x)$. Then $h(X)$ is well-defined, $-\infty \leq h(X) < +\infty$.

Important fact: If Z is normal with the same covariance matrix as X, then

$$h(X) \leq h(Z).$$

Definition. The quantity

$$D(X) = h(Z) - h(X)$$

is called the relative entropy of X with respect to Z.
Equivalently,

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

where g is density of Z, provided $\mathbf{E}Z = \mathbf{E}X$, $\mathrm{cov}(Z) = \mathrm{cov}(X)$.

Properties

- $0 \leq D(X) \leq +\infty$
- $D(X) = 0 \Longleftrightarrow X$ is normal
- $D(TX) = D(X)$, for any invertible affine map $T : \mathbf{R}^n \to \mathbf{R}^n$
- Pinsker-type inequality:

$$D(X) \geq \frac{1}{2} \, \|P_X - P_Z\|_{\mathrm{TV}}^2.$$

18 Isotropic constants

Definition. Let X be a random vector in $\mathbf{R}^n$ with log-concave density $f(x)$. The quantity

$$L_X^2 \;=\; L_f^2 \;=\; \|f\|^{2/n} \inf_T \int \frac{|Tx|^2}{n}\, f(x)\,dx$$

is called the isotropic constant of X or f. Equivalently,

$$L_X^2 \;=\; \|f\|^{2/n} \det(R)^{1/n}, \qquad R = \mathrm{cov}(X).$$

Isotropic position: $\mathbf{E}X = 0$ and for all $\theta \in \mathbf{R}^n$ with $|\theta| = 1$,

$$\mathbf{E}\,\langle X, \theta\rangle^2 \;=\; \int \langle x, \theta\rangle^2\, f(x)\,dx \;=\; \|f\|^{-2/n}\, L_f^2.$$

Theorem 4. We have

$$\log\Big[L_X\sqrt{2\pi/e}\,\Big] \le \frac{1}{n}\,D(X) \le \log\Big[L_X\sqrt{2\pi e}\,\Big].$$

Corollary. $L_X \ge \frac{1}{\sqrt{2\pi e}}$.

Corollary. In the class of all log-concave densities on $\mathbf{R}^n$ the following are equivalent:

- $D(X) \le Cn$ with some universal constant C.
- $L_X \le C$ with some universal constant C.

Proof. Let $\mathbf{E}X = 0$, $\text{cov}(X) = R = I$, so that

$$L_X = \|f\|^{1/n}.$$

Reminder: If X has a log-concave distribution,

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

Let $Z \sim N(0, I)$ and put $C = 2\pi e$. Then

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

Thus

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

On the other hand,

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

19 Reverse B-M inequality for isotropic bodies

Theorem (K. Ball 1986) Let A and B be convex bodies in isotropic position. Then with some absolute constant $c > 0$,

$$c\,|A+B|^{1/n} \le L_A\,|A|^{1/n} + L_B\,|B|^{1/n}.$$

Proof (information-theoretic). If $\mathrm{cov}(X) = \mathrm{cov}(Z)$ with Z normal, then

$$h(X) \le h(Z).$$

Equivalently, if $X \sim f$,

$$\frac{1}{2\pi e}\,H(X) \le \int \frac{|x|^2}{n}\,f(x)\,dx.$$

Let X and Y have log-concave isotropic densities f and g. Apply the above to $X+Y$:

$$\frac{1}{2\pi e}\,H(X+Y) \le \int \frac{|x|^2}{n}\,f(x)\,dx + \int \frac{|x|^2}{n}\,g(x)\,dx$$

That is,

$$\frac{1}{2\pi e}\,H(X+Y) \;\le\; L_f^2\,H(X) + L_g^2\,H(Y).$$

If $X \sim U(A)$, $Y \sim U(B)$,

$$\frac{1}{2\pi e}\,H(X+Y) \;\le\; L_A^2\,|A|^{2/n} + L_B^2\,|B|^{2/n}.$$

Reminder: $H(X+Y) \ge \frac{1}{4}\,|A+B|^{2/n}$, so

$$\frac{1}{8\pi e}\,|A+B|^{2/n} \;\le\; L_A^2\,|A|^{2/n} + L_B^2\,|B|^{2/n}.$$