Entropy jumps in the presence of a spectral gap

Keith Ball[*], Franck Barthe[†] and Assaf Naor[‡]

March 6, 2006

Abstract

It is shown that if X is a random variable whose density satisfies a Poincaré inequality, and Y is an independent copy of X, then the entropy of $(X+Y)/\sqrt{2}$ is greater than that of X by a fixed fraction of the entropy gap between X and the Gaussian of the same variance. The argument uses a new formula for the Fisher information of a marginal, which can be viewed as a reverse form of the Brunn-Minkowski inequality (in its functional form due to Prékopa and Leindler).

1 Introduction

The information-theoretic entropy of a real random variable X with density $f : \mathbb{R} \to [0, \infty)$ is defined as

$$\mathrm{Ent}(X) = -\int_{\mathbb{R}} f \log f$$

provided that the integral makes sense. For many random systems, entropy plays a fundamental role in the analysis of how they evolve towards an equilibrium. In this article we are interested in the convergence of the normalised sums

$$\frac{1}{\sqrt{n}} \sum_{1}^{n} X_i$$

of independent copies of X to the Gaussian limit: the convergence in the central limit theorem for IID copies of X.

[*]Supported in part by NSF grant DMS-9796221

[†]Supported in part by EPSRC Grant GR/R37210

[‡]Supported in part by the BSF, Clore Foundation and EU Grant HPMT-CT-2000-00037

Among random variables with a given variance, the Gaussian has the largest entropy. For any random variable, X, the gap between its entropy and that of a Gaussian G with the same variance, is a strong measure of how close X is to being Gaussian. In particular, if X has mean 0, variance 1 and density f, and G is the standard Gaussian with density g, then the Pinsker-Csiszar-Kullback inequality ([13], [7], [10]) implies that

$$\frac{1}{2}\left(\int_{\mathbb{R}} |f-g|\right)^2 \leq \mathrm{Ent}(G) - \mathrm{Ent}(X).$$

It is a consequence of the Shannon-Stam inequality ([15], [17]) that if X and Y are IID, then the normalised sum $(X+Y)/\sqrt{2}$ has entropy at least as large as that of X and Y. Thus, along the sequence of powers of 2, the normalised sums

$$\frac{1}{\sqrt{2^k}}\sum_{1}^{2^k} X_i$$

have increasing entropy. Under fairly weak assumptions it can be shown that this increasing entropy converges to that of the Gaussian limit. The idea of tracking the central limit theorem using entropy goes back to Linnik [12] who used it to give a particularly attractive proof of the central limit theorem. Barron [2] was the first to prove a central limit theorem with convergence in the entropy sense. A stronger result in this direction was obtained by Carlen and Soffer [6] (see also the works of Brown and Shimuzu [5] [16]). The crucial point is that if X is not itself a Gaussian random variable, and X and Y are IID, then the entropy of $(X+Y)/\sqrt{2}$ is strictly larger than that of X.

Carlen and Soffer obtained a uniform result of this type, showing that for any decreasing function K converging to zero at infinity and any real number $\varepsilon > 0$, there is a constant $\delta > 0$ so that if X is a random variable with variance 1 whose entropy is at least ε away from that of the standard Gaussian G, and if X satisfies the tail estimate

$$\int_{|x|>t} x^2 f(x)\, dx \leq K(t), \quad t > 0$$

then the entropy of the sum $(X+Y)/\sqrt{2}$ of IID copies of X is greater than that of X by at least δ. Thus, the entropy experiences a definite jump, as long as it starts off some way below the Gaussian value. Carlen and Soffer used this result to give a very clear entropy proof of the central limit theorem and to study various quantum systems. Their approach uses a compactness argument at one point, and for this reason does not give any quantitative estimates of δ in terms of ε and K. The main purpose of this article is to obtain such quantitative estimates, at least for certain special classes of random variables.

The starting point for the article is a method already used to good effect by (almost all) the authors mentioned earlier. Instead of considering the entropy directly, they study the Fisher information

$$4\int_{\mathbb{R}}(\nabla\sqrt{f})^2 = \int_{\mathbb{R}} f(\nabla \log f)^2.$$

For random variables with a given variance, the Gaussian has the *smallest* Fisher information, and according to the Blachman-Stam inequality ([4], [17]), the Fisher information of a general density, decreases with repeated convolution. It was discovered by Bakry and Emery [1] and again by Barron [2] that there is a remarkable connection between entropy and information provided by the adjoint Ornstein-Uhlenbeck semigroup. This is the semigroup $\{P_t\}_{t\geq 0}$ of convolution operators on L_1 which act on densities in the following way. If f is the density of X then $P_t f$ is the density of the random variable $X_t = \sqrt{e^{-2t}}X + \sqrt{1-e^{-2t}}G$ where G is a standard Gaussian, independent of X. The time derivative of the entropy gap $-\mathrm{Ent}(X_t) + \mathrm{Ent}(G)$ is precisely the gap between the Fisher information of X_t and that of G. Since the semigroup commutes with self-convolution, any linear inequality for the information gap can be integrated up to give the "same" inequality for the entropy gap. This argument is explained in more detail in Section 2.

Quantitative estimates for information jumps, in a very similar spirit to ours, have recently been obtained also by Barron and Johnson, [3]. Their result will be explained more precisely at the end of this introduction.

Our original plan was to try to get estimates for so-called log-concave random variables: those for which the density f has a concave logarithm. Since the Ornstein-Uhlenbeck evolutes of such a random variable are also log-concave, the Fisher information approach makes sense for these random variables. Now the Fisher information can be written as

$$\int_{\mathbb{R}} \frac{(f')^2}{f}.$$

Motivated by the desire to understand log-concave random variables one may observe that if the derivative of f decays at infinity, the integral is equal to

$$\int_{\mathbb{R}} \frac{(f')^2}{f} - f'' = \int_{\mathbb{R}} f(-\log f)''$$

The log-concave random variables are precisely those for which the latter integrand is pointwise non-negative.

Now, if X and Y are IID random variables with density f, the normalised sum $(X+Y)/\sqrt{2}$ has density

$$u \mapsto \int_{\mathbb{R}} f\left(\frac{u+v}{\sqrt{2}}\right) f\left(\frac{u-v}{\sqrt{2}}\right) dv$$

which is a marginal of the joint density on $\mathbb{R}^2$ of the pair (X, Y). It is a consequence of the Brunn-Minkowski inequality (in its functional form due to Prékopa and Leindler, see e.g. [14]) that log-concave random variables have log-concave marginals and hence that that if X and Y are log-concave, then so is $X+Y$. So, in principle, the Prékopa-Leindler inequality tells us something about the positivity of

$$(-\log(f * f))''$$

in terms of the positivity of

$$(-\log f)''.$$

The crucial first idea in this article is to rewrite a proof of the Brunn-Minkowski inequality so that it provides an explicit relation between these two functions which can be used to estimate the Fisher information of the convolution in terms of the information of the original f.

Most proofs of the Brunn-Minkowski inequality are not very well adapted for this purpose, since they involve combining values of the functions at points which are far apart: but since we want to end up just with derivatives, we need a "local" proof. Such a proof is the so-called "transportation" proof explained in Section 3 below. This is a variation of an argument by Henstock and Macbeath [9]. In Section 4, we differentiate the transportation proof to obtain the following identity which relates $(-\log h)''$ to the Hessian of $-\log w$ where $w : \mathbb{R}^2 \to (0, \infty)$ is a positive function on the plane and h is its marginal given by

$$h(x) = \int w(x, y)\, dy.$$

The basic formula *Let $w : \mathbb{R}^2 \to (0, \infty)$ be integrable and let h be the marginal given by,*

$$h(x) = \int w(x, y)\, dy.$$

Then, under appropriate regularity conditions, for each x

$$\begin{aligned} \frac{h'(x)^2}{h(x)} - h''(x) &= \int w.(\partial_y p)^2\, dy + \\ &+ \int w.\,(1, p)\,\mathrm{Hess}(-\log w)\,(1, p)^*\; dy \end{aligned} \tag{1}$$

where p is given by

$$p(x,y) = \frac{1}{w(x,y)} \left(\frac{h'(x)}{h(x)} \int_{-\infty}^{y} w(x,v)\, dv - \int_{-\infty}^{y} \partial_x w(x,v)\, dv \right).$$

The particular function p in the preceding formula arises as the derivative of a transportation map. The right-hand side of the above formula, is a quadratic form in the function p and it makes sense to ask whether the particular choice of p that yields the information of the marginal is the choice that minimises the quadratic form. It is not hard to check that this particular p does indeed satisfy the Euler-Lagrange equation for the minimisation problem. So under appropriate conditions we arrive at a second version of the formula.

Variational inequality for the marginal
Let $w : \mathbb{R}^2 \to (0, \infty)$ be integrable and let h be the marginal given by,

$$h(x) = \int w(x,y)\, dy.$$

Then under appropriate regularity conditions, for each x,

$$\begin{aligned} \frac{h'(x)^2}{h(x)} - h''(x) \quad &\leq \quad \int w.(\partial_y p)^2\, dy + \\ &+ \quad \int w.\, (1,p)\, \mathrm{Hess}(-\log w)\, (1,p)^*\ dy. \end{aligned}$$

By integrating over all x we obtain the estimate for information that we really want.

Theorem 1 (Variational inequality for the information)
Let $w : \mathbb{R}^2 \to (0, \infty)$ be a density with

$$\int \frac{|\nabla w|^2}{w}\ ,\ \int |\partial_{xx} w|\ < \infty.$$

Let h be the marginal density given by,

$$h(x) = \int w(x,y)\, dy$$

and let $J(h)$ be its Fisher information. Then

$$\begin{aligned} J(h) \quad &\leq \quad \int w.(\partial_y p)^2\, dxdy + \\ &+ \quad \int w.\, (1,p)\, \mathrm{Hess}(-\log w)\, (1,p)^*\ dxdy \end{aligned}$$

where $p : \mathbb{R}^2 \to \mathbb{R}$ is any differentiable function for which the terms $\frac{p\partial_x w \partial_y w}{w}$, $p\partial_{xy} w$ and $p^2 \partial_{yy} w$ are integrable.

As one might expect, once this theorem is written down, it can easily be proved directly without any reference to transportation or the calculus of variations. This is done in Section 5. The aim will be to choose a function p which is *not* optimal, but which is simple enough to calculate with, so as to obtain good estimates of the left side. Section 6 of the article explains how a suitable function p may be chosen in the case in which the marginal comes from a convolution. The problem of estimating an entropy jump is quite subtle and, presumably because of this, we were unable to write down an explicit p in any simple way. Our solution to this problem was to make the best possible choice of p from within a restricted family of functions of a certain type for which the formulae are simpler. So, this particular p is chosen to be the solution to the Euler-Lagrange equation for a restricted minimisation problem. Using standard Sturm-Liouville techniques, we then show that this choice of p guarantees a significant information jump, provided that the density f of the random variable X satisfies a spectral gap (or Poincaré) inequality:

$$c \int_{\mathbb{R}} f s^2 \leq \int_{\mathbb{R}} f (s')^2$$

for any function s satisfying

$$\int_{\mathbb{R}} f s = 0.$$

The estimate is given by the following theorem.

Theorem 2 *Let X be a random variable with variance 1 and finite Fisher information J, whose density f satisfies the Poincaré inequality above, and Y an independent copy of X. If the Fisher information of $(X+Y)/\sqrt{2}$ is J_2 then*

$$J - J_2 \geq \frac{c}{2+2c}(J-1).$$

The result states that for a density with a spectral gap inequality, the information decreases by a fixed proportion of its distance to the Gaussian, each time the random variable is added to itself. Using the Ornstein-Uhlenbeck semigroup, this inequality can be integrated up to give the "same" result for entropy.

Theorem 3 *Let X be a random variable with variance 1 and finite entropy, whose density f satisfies the Poincaré inequality above, and Y an independent copy of X. Then*

$$\mathrm{Ent}\left((X+Y)/\sqrt{2}\right) - \mathrm{Ent}(X) \geq \frac{c}{2+2c}(\mathrm{Ent}(G) - \mathrm{Ent}(X)).$$

As mentioned earlier, in a recent article, [3], Barron and Johnson have obtained an inequality very much in the same spirit as that of Theorem 2. Their method is completely different (at least in detail) but they also prove an information jump in the presence of a spectral gap. The main difference is that they have an extra factor $1/J$ on the right side. This non-linearity means that an entropy jump inequality cannot immediately be recovered but the inequality yields good large-time estimates for the rate of information decrease.

As we mentioned earlier, our original aim had been to obtain jump estimates for log-concave random variables. Although the really relevant property seems to a spectral gap, log-concave random variables on the line, *do* satisfy a spectral gap inequality with a uniform constant. Moreover, the Gaussian random variable satisfies the best possible spectral gap inequality so one expects the constant to improve as the random variable is repeatedly convolved with itself. There may be scope here for understanding long-term entropy behaviour, even for random variables that do not themselves satisfy a Poincaré inequality. Before we begin the analysis of the information of marginals we collect some of the standard facts alluded to above.

2 Preliminaries concerning information

Throughout this article, X will be a random variable with mean 0, variance 1 and density f and G will be a standard Gaussian with density $g : x \mapsto 1/\sqrt{2\pi}e^{-x^2/2}$. The Fisher information of G is

$$\int (g')^2/g = 1/\sqrt{2\pi} \int x^2 e^{-x^2/2}\, dx = 1.$$

Now as long as f is absolutely continuous

$$\begin{aligned} 0 &\leq \int \left(f'(x)/f(x) - x\right)^2 f(x)\, dx \\ &= \int (f')^2/f - 2\int x f'(x) + \int x^2 f(x) \\ &= J(X) - 2\int f + 1 = J(X) - 1 \end{aligned}$$

where $J(X)$ is the Fisher information of X. Thus among random variables with variance 1, the Gaussian has the least information.

Given a random variable X with density f and variance 1, we may run an Ornstein-Uhlenbeck process starting with X and the ordinate at time t having the same distribution as $\sqrt{e^{-t}}X + \sqrt{1-e^{-t}}G$ where G is a Gaussian with variance 1. Call the density of this random variable, f_t. For each $t > 0$, f_t will be strictly positive, continuously differentiable, and with the information

$$J_t = \int (f_t')^2 / f_t$$

finite.

Moreover f_t satisfies the modified diffusion equation

$$\frac{\partial}{\partial t} f_t = f_t'' + (x f_t)' .$$

Let the spatial differential operator on the right be denoted L. Then the time derivative of the entropy is

$$\begin{aligned} -\frac{\partial}{\partial t} \int f_t \log f_t &= -\int L f_t . \log f_t - \int L f_t \\ &= -\int L f_t . \log f_t \\ &= -\int f_t'' \log f_t - \int (x f_t)' \log f_t \\ &= \int (f_t')^2 / f_t + \int x f_t' \\ &= \int (f_t')^2 / f_t - 1 = J_t - 1 \end{aligned}$$

where J_t is the Fisher information of the random variable at time t.

Thus the derivative of the entropy is the gap between the information of the ordinate and that of the Gaussian limit. To obtain the result of Bakry-Emery and Barron, that the entropy gap is the integral of the information gap

$$\mathrm{Ent}(G) - \mathrm{Ent}(X) = \int_0^\infty (J_v - 1)\, dv$$

therefore boils down to a continuity result about the semigroup. A streamlined proof of this, under the assumption just that X has finite entropy can be found in [6]. Since we want to apply Theorem 2 along the semigroup, we need to know that the Poincaré constant does not deteriorate. (As we shall see below, it strictly improves along the semigroup). This observation seems to be well known, but we include a proof for the sake of completeness.

Proposition 1 *Let X, Y be independent random variables such that for every smooth function s,* $\mathrm{Var}[s(X)] \le A \mathbb{E} s'(X)^2$ *and* $\mathrm{Var}[s(Y)] \le B \mathbb{E} s'(Y)^2$. *Then for every $\lambda \in (0, 1)$ one has*

$$\mathrm{Var}[s(\sqrt{\lambda} X + \sqrt{1-\lambda} Y)] \le (\lambda A + (1-\lambda) B) \mathbb{E} s'(\sqrt{\lambda} X + \sqrt{1-\lambda} Y)^2.$$

Proof: Let P, Q denote the laws of X, Y respectively. Then

$$
\begin{aligned}
&\mathrm{Var}[\, s\,(\sqrt{\lambda}X + \sqrt{1-\lambda}Y)] \\
&\quad = \int s^2(\sqrt{\lambda}x + \sqrt{1-\lambda}y)\, dP(x)dQ(y) - \left(\int s(\sqrt{\lambda}x + \sqrt{1-\lambda}y)\, dP(x)dQ(y)\right)^2 \\
&\quad = \int \left[\int s^2(\sqrt{\lambda}x + \sqrt{1-\lambda}y)dQ(y) - \left(\int s(\sqrt{\lambda}x + \sqrt{1-\lambda}y)dQ(y)\right)^2\right] dP(x) \\
&\qquad + \int \left(\int s(\sqrt{\lambda}x + \sqrt{1-\lambda}y)dQ(y)\right)^2 dP(x) - \left(\int s(\sqrt{\lambda}x + \sqrt{1-\lambda}y)\, dP(x)dQ(y)\right)^2 \\
&\quad \le B(1-\lambda)\int s'(\sqrt{\lambda}x + \sqrt{1-\lambda}y)^2 dP(x)dQ(y) + A\int \left(\frac{\partial}{\partial x}\int s(\sqrt{\lambda}x + \sqrt{1-\lambda}y)dQ(y)\right)^2 dP(x) \\
&\quad \le (\lambda A + (1-\lambda)B)\mathbb{E}s'(\sqrt{\lambda}X + \sqrt{1-\lambda}Y)^2.
\end{aligned}
$$

■

From these remarks we immediately see that Theorem 3 above follows from Theorem 2.

3 The transportation argument

The aim of this section is to explain the transportation proof of the Brunn-Minkowski-Prékopa-Leindler inequality and why it is a "local" argument. We start with a function $w : \mathbb{R}^2 \to [0, \infty)$ which is log-concave and for each x we set

$$h(x) = \int w(x, y)\, dy.$$

We want to show that h is log-concave. For our purposes this means that for each a and b,

$$h\left(\frac{a+b}{2}\right) \ge h(a)^{1/2} h(b)^{1/2}. \tag{2}$$

Now define new functions as follows: for each y let

$$
\begin{aligned}
f(y) &= w(a, y) \\
g(y) &= w(b, y) \\
m(y) &= w\left((a+b)/2, y\right)
\end{aligned}
$$

We want to show that

$$\int m \ge \left(\int f\right)^{1/2} \left(\int g\right)^{1/2}$$

and from the log-concavity of w we know that for any y and z,

$$\begin{aligned} m\left((y+z)/2\right) &= w\left((a+b)/2,(y+z)/2\right) \\ &\geq w(a,y)^{1/2}w(b,z)^{1/2} \\ &= f(y)^{1/2}g(z)^{1/2}. \end{aligned} \tag{3}$$

By homogeneity, we may assume that $\int f = \int g = 1$ and we may also assume that f and g are smooth and strictly positive. Choose a fixed smooth density on the line, say $k : \mathbb{R} \to [0,\infty)$. We *transport* the measure defined by k to that defined by f and g in the following sense. We let $T, S : \mathbb{R} \to \mathbb{R}$ be the monotone increasing functions with the property that for each z,

$$\int_{-\infty}^{z} k(u)\,du = \int_{-\infty}^{S(z)} f(u)\,du$$

and

$$\int_{-\infty}^{z} k(u)\,du = \int_{-\infty}^{T(z)} g(u)\,du.$$

Equivalently, for each z,

$$k(z) = S'(z)f(S(z)) = T'(z)g(T(z)).$$

Then

$$\begin{aligned} 1 = \int k &= \int S'(z)^{1/2}T'(z)^{1/2}f(S(z))^{1/2}g(T(z))^{1/2} \\ &\leq \int S'(z)^{1/2}T'(z)^{1/2}m\left((S(z)+T(z))/2\right) \\ &\leq \int \left(\frac{S'(z)+T'(z)}{2}\right) m\left((S(z)+T(z))/2\right) \\ &= \int m \end{aligned}$$

where the first of the two inequalities is from (3) and the second is the arithmetic/geometric mean inequality. So $\int m \geq 1$ as required.

Our first aim in this article is to replace the inequality (2) with an identity relating the second derivative of $\log h$ to the Hessian of $\log w$. So we wish to examine the inequality when a and b are close together. In that case the functions f, g and m used in the proof are all close together. Thus, if we choose the density k to be the appropriate multiple of m, the transportation functions S and T will both be close to the identity function on $\mathbb{R}$:

$$S(z) \approx T(z) \approx z.$$

Therefore, the inequality

$$m\left((S(z)+T(z))/2\right) \geq f(S(z))^{1/2} g(T(z))^{1/2}$$

used in the proof, which depends upon the log-concavity of w, involves only values of w at points near to one another. So, in the limit, it will convert into an expression involving the Hessian of $\log w$. The resulting identity is the subject of the next section.

4 The basic formula

The aim in this section is to derive (informally) a formula for the Fisher information of a marginal in terms of the original density. We are given a density $w : \mathbb{R}^2 \to (0,\infty)$ and for each x we set

$$h(x) = \int_{\mathbb{R}} w(x,y)\,dy.$$

We want an expression for

$$\frac{h'^2}{h} - h'' = h\left(-\log h\right)''$$

at each point. To simplify the notation, let's fix this point at 0. For each value of x the function

$$y \mapsto w(x,y)/h(x)$$

is a density. Now for each value of x other than 0, transport the density at 0 to the one at x by a map $y \mapsto T(x,y)$. Thus, for each x and y we have

$$\frac{w\left(x, T(x,y)\right).\partial_y T(x,y)}{h(x)} = \frac{w(0,y)}{h(0)}. \tag{4}$$

If we take logarithms, differentiate twice with respect to x, set $x=0$, and use the fact that $T(0,y)=y$, we obtain a relationship which can be rearranged to give

$$\begin{aligned}
\frac{h'(0)^2}{h(0)^2} - \frac{h''(0)}{h(0)} &= \left(\partial_x \partial_y T|_{x=0}\right)^2 + \\
&+ \left(1, \partial_x T|_{x=0}\right).\mathrm{Hess}(-\log w)|_{x=0}\left(1, \partial_x T|_{x=0}\right)^* \\
&- \frac{1}{w(0,y)}\partial_y\left[w(0,y).\partial_{xx} T\right] \\
&= p'(y)^2 + (1, p(y))\,\mathrm{Hess}(-\log w)|_{x=0}\,(1, p(y))^* \\
&- \frac{1}{w(0,y)}\partial_y\left[w(0,y)\partial_{xx} T\right]
\end{aligned}$$

where $p : \mathbb{R} \to \mathbb{R}$ is given by

$$p(y) = \partial_x T(x, y)|_{x=0}. \tag{5}$$

Now multiply by $w(0, y)$ and integrate with respect to y. The last term in the above makes no contribution because we integrate a derivative. Since $\int w(0, y) = h(0)$ we end up with

$$\begin{aligned} \frac{h'(0)^2}{h(0)} - h''(0) &= \int w.(p')^2 \, dy + \\ &+ \int w. (1, p) \operatorname{Hess}(-\log w)|_{x=0} (1, p)^* \, dy. \end{aligned} \tag{6}$$

The connection with the Prékopa-Leindler inequality is now clear, since if w is log-concave then $\operatorname{Hess}(-\log w)$ is positive semi-definite and the right side of (6) is non-negative while the left side is $h(0)(-\log h)''(0)$.

The function p can be written directly in terms of w by differentiating equation (4) just once with respect to x. We get

$$\frac{d}{dy}\left(w(0, y)p(y)\right) = w(0, y)\frac{h'(0)}{h(0)} - \partial_x w(x, y)|_{x=0}. \tag{7}$$

The solution satisfying appropriate boundary conditions at $\pm\infty$ is

$$p(y) = \frac{1}{w(0, y)} \left(\frac{h'(0)}{h(0)} \int_{-\infty}^{y} w(0, v) \, dv - \int_{-\infty}^{y} \partial_x w(x, v)|_{x=0} \, dv \right).$$

Now, replacing the point 0 by a general one x we obtain formula (1).

The basic formula *Let $w : \mathbb{R}^2 \to (0, \infty)$ be integrable and let h be the marginal given by,*

$$h(x) = \int w(x, y) \, dy.$$

Then, under appropriate regularity conditions, for each x

$$\begin{aligned} \frac{h'(x)^2}{h(x)} - h''(x) &= \int w.(\partial_y p)^2 \, dy + \\ &+ \int w. (1, p) \operatorname{Hess}(-\log w) (1, p)^* \, dy \end{aligned} \tag{8}$$

where p is given by

$$p(x, y) = \frac{1}{w(x, y)} \left(\frac{h'(x)}{h(x)} \int_{-\infty}^{y} w(x, v) \, dv - \int_{-\infty}^{y} \partial_x w(x, v) \, dv \right). \tag{9}$$

The right side of (8) is a quadratic form in p. As explained in the introduction, it is natural to ask whether its minimum occurs at the particular function p given by (9). The Euler-Lagrange equation for the minimum simplifies to

$$\frac{\partial}{\partial y}\frac{1}{w(x,y)}\left(\frac{\partial}{\partial y}w(x,y)p(x,y)+\partial_x w(x,y)\right)=0$$

and this is satisfied by the function p. Hence the inequality in (1).

5 The inequality for information

The aim in this section is to give a proof of the inequality for information contained in Theorem 1.

Theorem (Variational inequality for the information) *Let $w:\mathbb{R}^2\to(0,\infty)$ be a continuously twice differentiable density with*

$$\int\frac{|\nabla w|^2}{w}\ ,\ \int|\partial_{xx}w|\ <\infty.$$

Let h be the marginal density given by,

$$h(x)=\int w(x,y)\,dy$$

and let $J(h)$ be its Fisher information. Then,

$$\begin{aligned} J(h) \ &\le\ \int w.(\partial_y p)^2\,dxdy+\\ &+\ \int w.\,(1,p)\,\mathrm{Hess}(-\log w)\,(1,p)^*\ dxdy \end{aligned}$$

where $p:\mathbb{R}^2\to\mathbb{R}$ is any differentiable function for which the terms $\frac{p\partial_x w\partial_y w}{w}$, $p\partial_{xy}w$ and $p^2\partial_{yy}w$ are integrable.

Proof Plainly we may assume that

$$\int w(\partial_y p)^2\ ,\ \int\frac{(\partial_y w)^2}{w}p^2\ <\infty.$$

For such p we have that for almost every x,

$$\int\partial_y\left[w(x,y).p(x,y)\right]\,dy=0.$$

The conditions on w ensure that for every x

$$h'(x) = \int \partial_x w(x,y)\, dy.$$

Hence

$$\begin{aligned}
\int \frac{h'(x)^2}{h(x)}\, dx &= \int \frac{\left(\int \partial_x w\, dy\right)^2}{\int w\, dy} \\
&= \int \frac{\left(\int \left[\partial_x w + \partial_y (w.p)\right]\, dy\right)^2}{\int w\, dy} \\
&\leq \int\int \frac{1}{w} \left(\partial_x w + \partial_y [w.p]\right)^2\, dydx
\end{aligned}$$

by the Cauchy-Schwarz inequality. The integrand on the right can be expanded and rearranged to give

$$\partial_{xx} w + w(\partial_y p)^2 + w.(1,p)\mathrm{Hess}(-\log w)(1,p)^* + \partial_y \left(p^2 \partial_y w + 2p\partial_x w\right).$$

The conditions on w ensure that $\int\int \partial_{xx} w = 0$. The conditions on p ensure that the last term also integrates to zero.

■

6 The variational construction of a test function

In this section we shall turn our attention to convolutions and estimate the jump in Fisher information for sums of random variables satisfying a spectral gap inequality. Let X be a random variable on the line with mean 0 and variance 1. Assume that X has a positive, continuously twice differentiable density f, for which

$$\int \frac{(f')^2}{f}, \int |f''| < \infty.$$

Let Y be an independent copy of X and $w : \mathbb{R}^2 \to (0,\infty)$ be the joint density given by $(u,v) \mapsto f((u+v)/\sqrt{2}) f((u-v)/\sqrt{2})$ so that the density of $(X+Y)/\sqrt{2}$ is

$$h : u \mapsto \int w(u,v)\, dv.$$

By the preceding theorem, the Fisher information of $(X+Y)/\sqrt{2}$ is at most

$$\int w.(\partial_v p)^2\, dvdu + \int w.\,(1,p)\,\mathrm{Hess}(-\log w)\,(1,p)^*\, dvdu.$$

for any suitable p. If we make the change of variables $x = (u+v)/\sqrt{2}$, $y = (u-v)/\sqrt{2}$, rewrite $p(u,v)$ as $q(x,y)$ and put H for the Hessian matrix we get

$$\frac{1}{2}\int f(x)f(y)\left(\partial_x q - \partial_y q\right)^2\, dxdy +$$

$$+\frac{1}{2}\int f(x)f(y)\left(1+q, 1-q\right) H \left(1+q, 1-q\right)^*\, dxdy. \tag{10}$$

Our task is to find a function q for which we can obtain a good estimate on this expression. The choice $q = 0$ leaves us with $\int f(-\log f)''$ which is the information for the original random variable X. So we need to choose a function q which improves upon this. We were led to the method which follows by studying a particular convolution semigroup: the Gamma semigroup. This semigroup consists of positive random variables X_t, indexed by a positive real parameter t, where X_t has density

$$x \mapsto e^{-x}x^{t-1}/\Gamma(t).$$

For every s and t, independent copies of X_s and X_t have a sum whose distribution is the same as X_{s+t}. In particular, for each t it is easy to describe the distribution of the sum of 2 independent copies of X_t. It turns out that for these random variables, the optimal function q is always given by

$$q(x,y) = \frac{x-y}{x+y}.$$

As t increases and the random variables look more and more Gaussian, the mass of the random variables concentrates in regions where x and y are large, so the denominator of q is essentially constant in the places where the action occurs. This suggested that it might be possible to replace q by an appropriate multiple of $x - y$ and this turns out to work well.

This in turn suggested that in the general situation, it might be sufficient to consider a function q of the form $q(x,y) = r(x) - r(y)$ for some function r of one variable. For such functions, the formula (10) simplifies considerably and it becomes clear that one loses very little by simplifying further and choosing q of the form $q(x,y) = r(x)$. In this case, if we put $\phi = -\log f$, the upper estimate for the Fisher information of the convolution, (10), becomes

$$J + 1/2\int f(r')^2 + 1/2\int f\phi'' r^2 + J/2\int fr^2 + \int f\phi'' r - J\int fr$$

where J is the information of the original random variable

$$J = \int f\phi''$$

and the regularity conditions on q will be satisfied as long as fr^2, $f(\phi')^2r^2$ and $f\phi''r^2$ are integrable. The information *change* as a result of convolution is thus at most

$$T(r) = 1/2\int f(r')^2 + 1/2\int f\phi''r^2 + J/2\int fr^2 + \int f\phi''r - J\int fr$$

and our aim is to choose the function r so as to make this quantity as negative as possible. The argument given below ignores questions of existence, regularity and integrability of solutions to the Sturm-Liouville problems that arise. All these issues are dealt with in the appendix.

At first sight this may look slightly odd if ϕ'' can take negative values but under the assumptions, we may integrate by parts to get

$$\int f(r')^2 + \int f\phi''r^2 = \int f(r' - \phi' r)^2 \tag{11}$$

so the objective functional cannot be too small.

The Euler-Lagrange equation for this minimisation problem is

$$-(fr')' + f(J + \phi'')r = f(J - \phi'') \tag{12}$$

Integrating this Sturm-Liouville equation respectively against r and 1 we obtain

$$\int f(r')^2 + \int f\phi''r^2 + J\int fr^2 = J\int fr - \int f\phi''r$$

and

$$J\int fr + \int f\phi''r = 0.$$

This shows for an appropriately integrable solution of equation (12)

$$T(r) = -J\int fr.$$

In order to estimate this quantity it is more convenient to obtain a differential equation for a new function s whose derivative is r. By writing $f' = -f\phi'$, and dividing by f, we can rewrite equation (12) as

$$-r'' + \phi'r' + \phi''r + Jr = J - \phi''.$$

If $r = s'$ then we can integrate the equation to get

$$-s'' + \phi's' + Js = Jx - \phi' \tag{13}$$

(apart from a constant of integration).

It remains to show that $T = -J \int fs'$ is bounded well away from 0. Multiplying (13) by f we get

$$-(fs')' + Jfs = Jfx + f' \tag{14}$$

Now equation (14) is the Euler-Lagrange equation for a different minimisation problem in which the objective function is

$$Q(s) = \frac{1}{2}\int f(s')^2 + \frac{J}{2}\int f\left(s - x + \phi'/J\right)^2. \tag{15}$$

Set $A = \int fs'$ and $B = \int fsx$. Integrating equation (14) against s and x respectively we get

$$\int fs'^2 + J\int fs^2 = J\int fsx - \int fs' = JB - A \tag{16}$$

and

$$\int fs' + J\int fsx = J\int fx^2 - \int f = J - 1$$

since f is the density of a random variable with variance 1. Hence

$$A + JB = J - 1. \tag{17}$$

From these equations one gets that the objective (15) is precisely A and the aim is to obtain a good lower bound for (15). Now suppose that f satisfies a Poincaré inequality with constant c. The constant of integration was chosen above so that the condition $\int f(x)x = 0$ (on the mean of the random variable) ensures that $\int fs = 0$. This means that the Poincaré inequality applies to s and

$$c\int fs^2 \leq \int f(s')^2.$$

Thus we get

$$\begin{aligned} A &= \frac{1}{2}\int f(s')^2 + \frac{J}{2}\int f\left(s - x + \phi'/J\right)^2 \\ &\geq \frac{1}{2}\int f\left(cs^2 + J\left(s - x + \phi'/J\right)^2\right) \\ &\geq \frac{1}{2}\frac{cJ}{c+J}\int f\left(-x + \phi'/J\right)^2 = \frac{cJ}{2(c+J)}\frac{J-1}{J}. \end{aligned}$$

This in turn gives, using $J \geq 1$

$$T = -JA \leq -\frac{cJ}{2(c+J)}(J-1) \leq -\frac{c}{2+2c}(J-1).$$

7 Appendix

Here we give detailed proofs for the arguments in the preceding section involving the calculus of variations.

First we notice that for the purpose of proving Theorem 2, we may restrict attention to densities f with the following properties: f is C^∞ and everywhere strictly positive, and f, f', f'' have a strong decay at infinity (like $1/t^2$ or even e^{-Kt} for some constant $K > 0$). To achieve this we replace f by the density f_t of $\sqrt{t}X + \sqrt{1-t}G$, for t very small. We have seen that f_t also satisfies a Poincaré inequality with constant c. Letting t tend to zero in the result for f_t, will provide the required inequality for f. It is plain that f_t is smooth and strictly positive. Its decay at zero is related to the tails of the density f. Indeed, up to a scaling factor, f_t is the convolution of f with a Gaussian function. More generally, let h be a function on $\mathbb{R}$ with $\int |h| = M$ and $|h(x)| \leq Ce^{-Bx^2}$. One has

$$\begin{aligned}\left|\int f(x-y)h(y)dy\right| &\leq \int_{|y|\leq|x|/2} f(x-y)|h(y)|dy + \int_{|y|>|x|/2} f(x-y)|h(y)|dy \\ &\leq M\int_{|y|>|x|/2} f + Ce^{-Bx^2/4}.\end{aligned}$$

So we obtain a bound for f determined by the tails of its distribution. The fact that $\int fx^2 = 1$ ensures a bound of the form $A/|x|^2$, but the spectral gap inequality is known to guarantee sub-exponential tails (see e.g. [11]). Therefore $|f * h(x)| \leq A_1 e^{-A_2|x|}$ where the constants depend on B, C, M and the spectral gap c. Applying this in turn for h equal to g, g' and g'' (where $g(x) = e^{-Kx^2}$) shows that f_t and its first and second derivatives are bounded by exponential functions. This in turns justifies the use of relations like $\int f'' = 0$.

From now on we assume that f has the above properties. Next we observe a consequence of the spectral gap inequality satisfied by f. Recall that for every smooth function s, we have assumed $\int fs^2 \leq C\int fs'^2 + (\int fs)^2$. This makes sense as soon as $\int fs^2$ and $\int fs'^2$ are finite. We shall work in $H^1(f)$, the Sobolev space with weight f of locally integrable functions s for which $\int fs^2 < \infty$ ($s \in L_2(f)$) and $\int fs'^2 < \infty$ where s' is the distributional derivative of s. This is a Hilbert space, equipped with the norm $\|s\|^2_{H^1(f)} = \int fs^2 + \int fs'^2$. By the density of smooth functions, the Poincaré inequality is valid for $s \in H^1(f)$. We point out a useful fact: if s is smooth and $s' \in L_2(f)$, then the assumption on f ensures that $s \in L_2(f)$ as well, and that the inequality holds for s. We give a proof of this. For each $n \geq 1$ let $s_n = \xi_n - \int f\xi_n$, where $\xi_n(t) = s(t)$ if $t \in [-n, n]$,

$\xi_n(t) = s(n)$ if $t > n$ and $\xi_n(t) = s(-n)$ if $t < -n$. Clearly s_n is in $H^1(f)$, $\int f s_n'^2 \leq \int f s'^2$ and by the spectral gap inequality, $\int f s_n^2 \leq C \int f s'^2$. Therefore the sequence (s_n) is in a ball of $H^1(f)$, which is compact for the weak topology. So the on subsequence has a weak limit s_∞ in $H^1(f)$. However by construction, the limit of s_n' in the $L_2(f)$ sense is s'. Thus $s' = s_\infty'$ in the distributional sense. This implies that s is equal to $s_\infty \in H^1(f)$ plus a constant function (also in $H^1(f)$). So $s \in H^1(f)$ and the Poincaré inequality is valid for s.

We are now ready to give details of the argument in Section 6. We recall some classical facts from the calculus of variations:

Proposition 2 *Let w be a strictly positive smooth function on $\mathbb{R}$. Let $\varphi_0 \in H^1(w)$ be also smooth. For $\varphi \in H^1(w)$, let*

$$F(\varphi) = \int w\varphi'^2 + J\int w(\varphi - \varphi_0)^2$$

1) *The infimum $\inf\limits_{\varphi \in H^1(w)} F(\varphi)$ is achieved by a function $\rho \in H^1(w)$.*
2) *For every function $\varphi \in H^1(f)$, one has*

$$\int w\rho'\varphi' + J\int w(\rho - \varphi_0)\varphi = 0.$$

3) *The minimum of the functional is equal to*

$$F(\rho) = J\int w\varphi_0^2 - J\int w\varphi_0\rho.$$

4) *The function ρ is in C^∞ and satisfies the Euler-Lagrange equation*

$$(-w\rho')' + Jw(\rho - \varphi_0) = 0.$$

Proof: Let $m \geq 0$ be the infimum of the functional and $(\varphi_n)_{n\geq 1}$ be a minimizing sequence with $F(\varphi_n) \leq m + 1/n$. Clearly $\int w\varphi_n'^2 \leq m+1$ and $\|\varphi_n\|_{L_2(w)} \leq \|\varphi_0\|_{L_2(w)} + \|\varphi_n - \varphi_0\|_{L_2(w)} \leq \|\varphi_0\|_{L_2(w)} + \sqrt{(m+1)/J}$. So the sequence is in a bounded subset of $H^1(w)$, which is weakly relatively compact. Therefore there is a subsequence with a weak limit $\rho \in H^1(w)$. The functional F is clearly norm-continuous in $H^1(w)$. Since it is also convex, it is weakly continuous and $F(\rho) = m$ is the minimum. To prove the second part of the proposition, one looks at the behaviour around the minimum. For every $\varphi \in H^1(w)$ and $t \in \mathbb{R}$, $F(\rho) \leq F(\rho + t\varphi)$. As $t \to 0$ this yields the relation stated in

2). This applied for $\varphi = \rho$ gives the formula for $F(\rho)$ given in 3). The relation given in 2) applied for smooth compactly supported functions shows that ρ satisfies the equation $(-w\rho')' + Jw(\rho - \varphi_0) = 0$ in the distributional sense. This equation can be rearranged as

$$\rho'' + \frac{w'}{w}\rho' - J\rho = -J\varphi_0.$$

The theory of regularity for uniformly elliptic second order linear equations can be used to show that ρ is a strong solution (see e.g. [8]). We first apply L_p regularity theory on open segments I : the second term in the equation is bounded by continuity so it is in $L^2(I)$ (with Lebesgue measure. It follows from the the theory that ρ has two derivatives in $L^2(I')$ for every smaller open interval I'. In particular $\rho \in H^1(I')$, so by the Sobolev injection, ρ is Hölder continuous. It is therefore locally bounded, and Schauder theory then applies (see [8, Theorem 6.2 page 85]). Thus ρ is infinitely differentiable, and the Euler-Lagrange equation is satisfied in the strong sense. ■

This proposition applies directly to the minimization of the quantity 15. It ensures that the minimum is achieved for a smooth s satisfying equation 14, and is equal to A. More effort is needed for the quantity

$$T(r) = 1/2\int f(r')^2 + 1/2\int f\phi'' r^2 + J/2\int fr^2 + \int f\phi'' r - J\int fr.$$

Our aim is to show that the infimum of T on, (say) smooth compactly supported functions, is well below zero. For such functions, we may rewrite $T(r)$ as:

$$T(r) = -J + \frac{1}{2}\int f(r+1)'^2 + \frac{1}{2}\int f\phi''(r+1)^2 + \frac{J}{2}\int f(r-1)^2.$$

Integrating by parts as in (11), which is justified for functions that are eventually constant, we get

$$T(r) = -J + \frac{1}{2}\left(\int \frac{1}{f}(f(r+1))'^2 + J\int \frac{1}{f}(f(r+1) - 2f)^2\right).$$

So by the density of functions of the form $f(r+1)$ we are reduced to the study of the infimum problem

$$\inf_{\varphi \in H^1(1/f)} \int \frac{1}{f}(\varphi')^2 + J\int \frac{1}{f}(\varphi - 2f)^2.$$

By the proposition above, there is a smooth minimizer $\rho \in H^1(1/f)$. Let r be such that $\rho = f(r+1)$. The minimum of the functional T is given in terms of ρ by statement 3). In terms of r it is equal to $-J\int fr$. Moreover, the Euler-Lagrange equation satisfied by

ρ translates into equation (12) for r. Now choose the primitive σ of r for which there is no additional constant in equation (13). We are done if we can prove that this σ is the function s obtained as a minimizer and satisfying the same equation. To do so, note that $\int f\sigma'^2 = \int fr^2 < \infty$, so by the previous remarks on the Poincaré inequality, $\sigma \in H^1(f)$. Finally, s and σ are two critical points of the same strictly convex functional on $H^1(f)$: so they coincide and the argument is complete.

Acknowledgements: We would like to thank Hatem Zaag for useful discussions. This research was carried out while the second and third named authors where visiting University College London.

References

[1] D. Bakry and M. Emery. Diffusions hypercontractives. In *Séminaire de Probabilités XIX*, number 1123 in Lect. Notes in Math., pages 179–206. Springer, 1985.

[2] A. R. Barron. Entropy and the central limit theorem. *Ann. Probab.*, 14:336–342, 1986.

[3] A. R. Barron and O. Johnson. Fisher information inequalities and the central limit theorem. preprint.

[4] N. M. Blachman. The convolution inequality for entropy powers. *IEEE Trans. Info. Theory*, 2:267–271, 1965.

[5] L. D. Brown. A proof of the central limit theorem motivated by the Cramer-Rao inequality. In Kalliampur et al., editor, *Statistics and Probability: Essays in Honor of C. R. Rao*, pages 314–328, Amsterdam, 1982. North-Holland.

[6] E. A. Carlen and A. Soffer. Entropy production by block variable summation and central limit theorem. *Commun. Math. Phys.*, 140(2):339–371, 1991.

[7] I. Csiszar. Informationstheoretische Konvergenzbegriffe im Raum der Wahrscheinlichkeitsverteilungen. *Publications of the Mathematical Institute, Hungarian Academy of Sciences*, VII, Series A:137–157, 1962.

[8] D. Gilbarg and N. Trudinger. *Elliptic partial differential equations of second order*. Springer-Verlag, Berlin, second edition, 1983.

[9] R. Henstock and A. M. Macbeath. On the measure of sum sets. (I) the theorems of Brunn, Minkowski and Lusternik. *Proceedings of the London Math. Soc.*, 3:182–194, 1953.

[10] S. Kullback. A lower bound for discrimination information in terms of variation. *IEEE Trans. Info. Theory*, 4:126–127, 1967.

[11] M. Ledoux. Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXIII*, number 1709 in Lecture Notes in Math., pages 120–216, Berlin, 1999. Springer.

[12] Ju. V. Linnik. An information theoretic proof of the central limit theorem with Lindeberg conditions. *Theory Probab. Appl.*, 4:288–299, 1959.

[13] M.S̃. Pinsker. Information and information stability of random variables and processes. Holden-Day, San Francisco, 1964.

[14] G. Pisier. *The volume of convex bodies and Banach space geometry.* Cambridge University Press, Cambridge, 1989.

[15] C. E. Shannon and W. Weaver. *The mathematical theory of communication.* University of Illinois Press, Urbana, IL, 1949.

[16] R. Shimizu. On Fisher's amount of information for location family. In Patil et al., editor, *A modern course on statistical distributions in scientific work*, Boston, MA, 1974. D. Reidel.

[17] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Info. Control*, 2:101–112, 1959.

Keith M. Ball,
Department of Mathematics,
University College London,
Gower Street,
London WC1 6BT,
United Kingdom.
E-mail: kmb@math.ucl.ac.uk

F. Barthe,
Laboratoire d'Analyse et de Mathématiques Appliquées

CNRS UMR8050-Universités de Marne-la-Vallée et de Créteil Val-de-Marne,
Cité Descartes, Champs-sur-Marne,
77454 Marne-la-Vallée, cedex 2,
France.
E-mail: barthe@math.univ.mlv.fr

A. Naor,
Institute of Mathematics,
Hebrew University,
Jerusalem 91904,
Israel.
E-mail: naor@math.huji.ac.il