

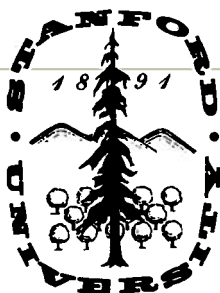
ON THE SIMILARITY OF THE ENTROPY POWER INEQUALITY AND THE BRUNN MINKOWSKI INEQUALITY

by
Max H. M. Costa
and
Thomas M. Cover

TECHNICAL REPORT NO. 48
SEPTEMBER 1983

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION GRANT
ECS82-11568

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



ON THE SIMILARITY OF THE ENTROPY POWER INEQUALITY
AND THE BRUNN MINKOWSKI INEQUALITY

by

Max H.M. Costa
Thomas M. Cover

TECHNICAL REPORT NO. 48

SEPTEMBER 1983

PREPARED UNDER THE AUSPICES
OF
NATIONAL SCIENCE FOUNDATION
GRANT ECS82-11568

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

On the Similarity of the Entropy Power Inequality and the Brunn Minkowski Inequality

Max H.M. Costa[†]
Thomas M. Cover^{††}

ABSTRACT

The entropy power inequality states that the effective variance (entropy power) of the sum of two independent random variables is greater than the sum of their effective variances. The Brunn Minkowski inequality states that the effective radius of the set sum of two sets is greater than the sum of their effective radii. We recast both these inequalities in a form that enhances their similarity. In spite of this similarity, there is as yet no common proof of the inequalities. Nevertheless, their intriguing similarity suggests that we may find new results relating to entropies from known results in geometry and vice versa. We present two applications of this reasoning. First we prove an isoperimetric inequality for entropy, showing that the spherical normal distribution minimizes the trace of the Fisher information matrix given an entropy constraint; just as a sphere minimizes the surface area given a volume constraint. Second, we prove a theorem involving the effective radii of growing convex sets.

[†] Stanford University. This work was partially supported by National Science Foundation Grant ECS82-11568 and by the Coordenacao de Aperfeicoamento de Pessoal de Nivel Superior (CAPES) and the Instituto de Pesquisas Espaciais (INPE) of Brazil.

^{††} Stanford University. This work was partially supported by National Science Foundation Grant ECS82-11568 and Joint Services Electronics Program DAAG-29-81-K-0057.

1. The Entropy Power Inequality.

Let a random variable X have a probability density function $f(x)$, $x \in \mathbb{R}$. Then its (differential) *entropy* $H(X)$ is defined by

$$H(X) = - \int f(x) \ln f(x) dx .$$

Shannon's entropy power inequality [1] states, for X and Y independent random variables having density functions, that

$$e^{2H(X+Y)} \geq e^{2H(X)} + e^{2H(Y)} . \quad (1)$$

We wish to recast this inequality. First we observe that a normal random variable $Z \sim \phi(z) = (1/\sqrt{2\pi\sigma^2}) e^{-z^2/2\sigma^2}$ with variance σ^2 has entropy

$$\begin{aligned} H(Z) &= - \int \phi \ln \phi \\ &= \frac{1}{2} \ln 2\pi e \sigma^2 . \end{aligned} \quad (2)$$

By inverting, we see that if Z is normal with entropy $H(Z)$, then its variance is

$$\sigma^2 = \frac{1}{2\pi e} e^{2H(Z)} . \quad (3)$$

Thus the entropy power inequality is an inequality between effective variances, where effective variance (entropy power) is simply the variance of the normal random variable having the same entropy.

The above equations allow rewriting of the entropy power inequality in the equivalent form

$$H(X+Y) \geq H(X' + Y') \quad (4)$$

where X' and Y' are independent normal with corresponding entropies $H(X') = H(X)$ and $H(Y') = H(Y)$. Verification of this restatement follows from the use of (1) to show that

$$\begin{aligned}
 \frac{1}{2\pi e} e^{2H(X+Y)} &\geq \frac{1}{2\pi e} e^{2H(X)} + \frac{1}{2\pi e} e^{2H(Y)} \\
 &= \frac{1}{2\pi e} e^{2H(X')} + \frac{1}{2\pi e} e^{2H(Y')} \\
 &= \sigma_{X'}^2 + \sigma_{Y'}^2 \\
 &= \sigma_{X'+Y'}^2 \\
 &= \frac{1}{2\pi e} e^{2H(X'+Y')}, \tag{5}
 \end{aligned}$$

where the penultimate equality follows from the fact that the sum of two independent normals is normal with a variance equal to the sum of the variances.

By the same line of reasoning, the entropy power inequality for independent random n -vectors \mathbf{X} and \mathbf{Y} , which is given by

$$e^{\frac{2}{n} H(\mathbf{X} + \mathbf{Y})} \geq e^{\frac{2}{n} H(\mathbf{X})} + e^{\frac{2}{n} H(\mathbf{Y})}, \tag{6}$$

can be recast as

$$H(\mathbf{X} + \mathbf{Y}) \geq H(\mathbf{X}' + \mathbf{Y}'), \tag{7}$$

where \mathbf{X}' , \mathbf{Y}' are independent multivariate normal random vectors with proportional covariance matrices and corresponding entropies.

2. The Brunn Minkowski Inequality.

Let A and B be two measurable sets in \mathbf{R}^n . The set sum $C = A + B$ of these sets may be written

$$C = \{ \mathbf{x} + \mathbf{y} : \mathbf{x} \in A, \mathbf{y} \in B \} . \quad (8)$$

Let $V(A)$ denote the volume of A . The Brunn Minkowski inequality [2],[3] states that

$$V^{1/n}(A + B) \geq V^{1/n}(A) + V^{1/n}(B) . \quad (9)$$

To recast this inequality we observe that an n -sphere S with radius r has volume

$$V(S) = c_n r^n . \quad (10)$$

Thus if S is a sphere with volume V , its radius is

$$r = (V / c_n)^{1/n} . \quad (11)$$

Hence the Brunn Minkowski inequality can be viewed as an inequality between the radii of the spherical equivalents of the sets.

Rewriting, using the model in (5), we have the following restatement of the Brunn Minkowski inequality:

$$V(A + B) \geq V(A' + B') \quad (12)$$

where A' and B' are spheres with volumes $V(A') = V(A)$ and $V(B') = V(B)$.

3. Comparisons.

Further similarity between the entropy power and Brunn Minkowski inequalities is revealed by inspection of the convolution. The entropy $H(X + Y)$ is a functional of the convolution, i.e.,

$$H(X + Y) = - \int f_Z \ln f_Z , \quad (13)$$

where

$$\begin{aligned} f_Z(z) &= (f_X * f_Y)(z) \\ &= \int f_X(t) f_Y(z - t) dt . \end{aligned}$$

Similarly, the volume $V(A + B)$ is a functional of a "convolution", i.e.,

$$V(A + B) = \int I_C(\mathbf{z}) d\mathbf{z} , \quad (14)$$

where $I_C(\mathbf{z})$, the indicator function for $C = A + B$ is given by

$$I_C(\mathbf{z}) = \max_{\mathbf{t}} I_A(\mathbf{t}) I_B(\mathbf{z} - \mathbf{t}) . \quad (15)$$

We note that $f_Z(z)$ is the L_1 norm of $h_z(t) = f_X(t) f_Y(z - t)$ and that $I_C(\mathbf{z})$ is the L_∞ norm of $h_z(\mathbf{t}) = I_A(\mathbf{t}) I_B(\mathbf{z} - \mathbf{t})$ and that L_1 and L_∞ are dual spaces.

Finally e^H , like V , is a measure of volume. For example, for all random variables X with support set A , we have $H(X) \leq \ln V(A)$, with equality if the probability density is uniform over A . Moreover, by the Asymptotic Equipartition Property, we know the volume of the set of ϵ -typical n -sequences (X_1, X_2, \dots, X_n) , X_i i.i.d. $\sim f(x)$ is equal to $e^{nH(X)}$ to first order in the exponent. Thus $e^{H(X)} = e^{nH(X)}$ is the volume of the typical set for $\mathbf{X} = (x_1, x_2, \dots, x_n)$. To suggest a link between the Gaussian distribution and spheres, we note that the ϵ -typical set of n -sequences (X_1, X_2, \dots, X_n) when X_i are i.i.d. according to a Gaussian distribution is given by a sphere.

These observations suggest not only that the two inequalities may be different manifestations of the same underlying idea, but that there may be a continuum of inequalities between L_1 and L_∞ with their respective natural definitions of volume.

In spite of the obvious similarity between the above inequalities, there is no apparent similarity of any of the known proofs of the Brunn Minkowski inequality [2],[3],[4] with the Stam and Blachman [5],[6] proofs of the entropy power inequality, nor have we succeeded in finding a new common proof. Nevertheless, the similarity of these inequalities suggests that we may find new results relating to entropies from known results in geometry and vice versa. We present two applications of this reasoning.

4. Isoperimetric Inequalities.

It is known that the sphere minimizes surface area for given volume. A proof follows immediately from the Brunn Minkowski inequality [4] as shown below. For "regular" sets A , the surface area $S(A)$ of A is given by

$$S(A) = \lim_{\epsilon \rightarrow 0} \frac{V(A + S_\epsilon) - V(A)}{\epsilon} \quad (16)$$

where S_ϵ is a sphere of radius $\epsilon > 0$. Using the Brunn Minkowski inequality we have

$$S(A) \geq \lim_{\epsilon \rightarrow 0} \frac{V(A' + S'_\epsilon) - V(A')}{\epsilon} = S(A'), \quad (17)$$

where A' is a sphere with volume $V(A') = V(A)$. Of course, $S'_\epsilon = S_\epsilon$ and $A' + S'_\epsilon$ is a sphere. Thus the surface area of A is greater than that of a sphere with the same volume.

We now proceed to perform the same steps on the entropy power. Noting that e^H , like V , is a measure of volume, we define $S(\mathbf{X})$, the "surface area" of a multivariate random variable \mathbf{X} , by

$$S(\mathbf{X}) = \lim_{\epsilon \rightarrow 0} \frac{e^{H(\mathbf{X} + \mathbf{Z}_\epsilon)} - e^{H(\mathbf{X})}}{\epsilon}, \quad (18)$$

where \mathbf{Z}_ϵ is Gaussian with covariance matrix ϵI . Thus $S(\mathbf{X})$ is the rate of change of the "volume" e^H when a small normal random variable is added. We may write

$$S(\mathbf{X}) = \lim_{t_0 \rightarrow 0} \frac{d}{dt} e^{H(\mathbf{X} + \mathbf{Z}_t)} \big|_{t=t_0}. \quad (19)$$

Assuming \mathbf{X} has a reasonably smooth density $p(\mathbf{x})$ we use the result proved in the Appendix to find

$$S(\mathbf{X}) = \frac{1}{2} E \frac{\|\nabla p\|^2}{p^2} e^{H(\mathbf{X})}. \quad (20)$$

where $\|\nabla p\|$ denotes the norm of the gradient of $p(\mathbf{x})$. Also, by the entropy power inequality (7), we have

$$e^{H(\mathbf{X} + \mathbf{Z}_\epsilon)} \geq e^{H(\mathbf{X}' + \mathbf{Z}_\epsilon)}, \quad (21)$$

where \mathbf{X}' is a Gaussian n -vector as before, with entropy $H(\mathbf{X}') = H(\mathbf{X})$. Thus, from (18) and (21), we obtain the following bound on $S(\mathbf{X})$:

$$S(\mathbf{X}) \geq \lim_{\epsilon \rightarrow 0} \frac{e^{H(\mathbf{X}' + \mathbf{Z}_\epsilon)} - e^{H(\mathbf{X}')}}{\epsilon} = S(\mathbf{X}'). \quad (22)$$

For the Gaussian vector \mathbf{X}' , the "surface area" is found from (20) to be

$$S(\mathbf{X}') = \frac{n \pi e e^{H(\mathbf{X}')}}{e^{\frac{2}{n} H(\mathbf{X}')}}. \quad (23)$$

Finally, recalling that $H(\mathbf{X}') = H(\mathbf{X})$, we combine (20), (22), and (23) to obtain the entropy analog of the isoperimetric inequality:

$$\frac{1}{n} E \frac{\|\nabla p\|^2}{p^2} \geq \left(\frac{e^{\frac{2}{n} H(\mathbf{X})}}{2 \pi e} \right)^{-1}. \quad (24)$$

Letting $J(X)$ denote the trace of the Fisher information matrix for the translation family of densities $\{f(\mathbf{x} - \theta)\}$, $\theta \in \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^n$, we have

$$J(X) = E \frac{\|\nabla p\|^2}{p^2}$$

with the corresponding isoperimetric inequality

$$J(\mathbf{X}) \geq J(\mathbf{X}') \quad (25)$$

where \mathbf{X}' is spherical normal with entropy $H(\mathbf{X}') = H(\mathbf{X})$.

This establishes that the Gaussian distribution minimizes the trace of the Fisher Information matrix given an entropy constraint. The scalar version of the above relation was proved in [5].

5. Concavity.

Here we present an example of how a result involving entropy may yield a conjecture in geometry. The concavity of entropy power with added Gaussian noise [7] suggests the following conjecture.

Conjecture: Let A be an arbitrary measurable set and S a sphere of unit radius in R^n . Then $V^n(A + tS)$ is a concave function of $t \geq 0$.

We shall prove that the conjecture is true if A is convex.

Theorem. If A is convex, and S is a sphere, then $V^n(A + tS)$ is a concave function of $t \geq 0$.

Proof. Let $\theta \in [0,1]$. Then we have

$$\begin{aligned} V^n(A + \theta S) &= V^n((1-\theta)A + \theta A + \theta S) \\ &\geq V^n((1-\theta)A) + V^n(\theta A + \theta S) \\ &= (1-\theta)V^n(A) + \theta V^n(A + S), \end{aligned} \tag{26}$$

where we used the convexity of the set A in the first line and the Brunn Minkowski inequality in the second line. Since A is an arbitrary convex set, scaling $A + tS$ completes the proof.

As far as we know, the conjecture is unresolved for arbitrary measurable sets in R^n .

6. Concluding Remark.

We still do not know if the Brunn Minkowski and Entropy Power inequalities have a common underlying idea leading to similar proofs. The inequalities look the same ((7) and (12)), conditions for equality are similar, and both involve measures of volume.

Acknowledgements

The authors wish to thank Brad Efron, Paul Shields, and Rudolf Ahlswede for their discussions.

Appendix

Let $\mathbf{X}_t = \mathbf{X} + \mathbf{Z}_t$ be the sum of a vector valued random variable \mathbf{X} having arbitrary density $p(\mathbf{x})$ and a spherical multivariate normal random variable \mathbf{Z}_t having covariance tI where I is the identity matrix. Then \mathbf{X}_t has density $p_t(\mathbf{x}_t)$ given by

$$p_t(\mathbf{x}_t) = \frac{1}{(2\pi t)^{n/2}} \int_{\mathbf{R}^n} p(\mathbf{x}) \exp \left(\frac{-\|\mathbf{x}_t - \mathbf{x}\|^2}{2t} \right) d\mathbf{x} . \quad (\text{A.1})$$

Due to the smoothing properties of the normal distribution we can differentiate the above expression inside the integral (the integrand is continuous and differentiable in t) to show that $p_t(\mathbf{x}_t)$ satisfies the diffusion (heat) equation

$$\frac{d}{dt} p_t(\mathbf{x}_t) = \frac{1}{2} \nabla^2 p_t(\mathbf{x}_t) , \quad (\text{A.2})$$

where

$$\nabla^2 p_t(\mathbf{x}_t) = \sum_{i=1}^n \frac{\partial^2}{\partial x_{(i)t}^2} p_t(\mathbf{x}_t) . \quad (\text{A.3})$$

Interchanging derivative and integral once more, we get

$$\begin{aligned} \frac{d}{dt} H(\mathbf{X}_t) &= - \int_{\mathbf{R}^n} \frac{d}{dt} p_t(\mathbf{x}_t) d\mathbf{x}_t - \int_{\mathbf{R}^n} \left(\frac{d}{dt} p_t(\mathbf{x}_t) \right) \log p_t(\mathbf{x}_t) d\mathbf{x}_t \\ &= 0 - \frac{1}{2} \int_{\mathbf{R}^n} (\nabla^2 p_t(\mathbf{x}_t)) \log p_t(\mathbf{x}_t) d\mathbf{x}_t . \end{aligned} \quad (\text{A.4})$$

We now recall Green's identity [8]: if $\phi(\mathbf{x})$ and $\psi(\mathbf{x})$ are twice continuously differentiable functions in \mathbf{R}^n and V is any set bounded by a piecewise smooth, closed, oriented surface S in \mathbf{R}^n , then

$$\int_V \phi \nabla^2 \psi dV = \int_S \phi \nabla \psi \cdot d\mathbf{s} - \int_V \nabla \phi \cdot \nabla \psi dV , \quad (\text{A.5})$$

where $\nabla\psi$ denotes the gradient of ψ , ds denotes the elementary area vector and $\nabla\psi \cdot ds$ is the inner product of these two vectors. This identity plays the role of integration by parts in \mathbf{R}^n .

To apply Green's identity to (A.4), we let V_r be the n -sphere of radius r centered at the origin and having surface S_r . Then we use Green's identity on V_r and S_r with $\phi(\mathbf{x}_t) = \log p_t(\mathbf{x}_t)$ and $\psi(\mathbf{x}_t) = p_t(\mathbf{x}_t)$ and take the limit as $r \rightarrow \infty$. The surface integral over S_r can be shown to vanish in the limit. Hence we obtain

$$\begin{aligned} \frac{d}{dt} H(\mathbf{X}_t) &= -\frac{1}{2} \int_{\mathbf{R}^n} \nabla p_t(\mathbf{x}_t) \cdot \nabla \log p_t(\mathbf{x}_t) d\mathbf{x}_t \\ &= \frac{1}{2} \int_{\mathbf{R}^n} \frac{\|\nabla p_t(\mathbf{x}_t)\|^2}{p_t(\mathbf{x}_t)} d\mathbf{x}_t, \end{aligned} \quad (\text{A.6})$$

as desired.

Equation (A.6) can also be written as

$$\frac{d}{dt} H(\mathbf{X}_t) = \frac{1}{2} E \frac{\|\nabla p_t\|^2}{p_t^2}. \quad (\text{A.7})$$

References

- [1] C.E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. Journal*, Vol. 27, pp. 623-656, Oct. 1948.
- [2] B. Grunbaum, *Convex Polytopes*. London: Interscience Publishers, 1967.
- [3] H. G. Eggleston, *Convexity*. Cambridge: The University Press, 1958.
- [4] H. Hadwiger and D. Ohmann, "Brunn-Minkowskischer Satz und Isoperimetrie", *Math. Zeitschr*, vol. 66, pp. 1-8, 1956.
- [5] A. J. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon", *Information and Control*, vol. 2, pp. 101-112, Jun. 1959.
- [6] N. M. Blachman, "The convolution inequality for entropy powers", *IEEE Trans. Inform. Theory*, vol. IT-11, pp.267-271, Apr. 1965.
- [7] M.H. Costa, "Interference Channels", Ph.D. thesis, Dept. of Elec. Eng., Stanford University, February 1983.
- [8] J. Amazigo and L. Rubinfeld, *Advanced Calculus and its Applications to the Engineering and Physical Sciences*; New York: Wiley, 1980.