
Keith Ball
The second law of thermodynamics

Joule and Carnot studied ways to improve the efficiency of steam engines.

Is it possible for a thermodynamic system to move from state A to state B without any net energy being put into the system from outside?

A single experimental quantity, dubbed entropy, made it possible to decide the direction of thermodynamic changes.
The second law of thermodynamics

The entropy of a closed system increases with time.

The second law applies to all changes: not just thermodynamic.

Entropy measures the extent to which energy is dispersed: so the second law states that energy tends to disperse.
The second law of thermodynamics

Closed systems become progressively more featureless.

We expect that a closed system will approach an equilibrium with maximum entropy.
Information theory

Shannon showed that a noisy channel can communicate information with almost perfect accuracy, up to a fixed rate: the capacity of the channel.

The (Shannon) entropy of a probability distribution: if the possible states have probabilities $p_1, p_2, ..., p_n$ then the entropy is

$$H = - \sum_{i=1}^{n} p_i \log_2 p_i$$

Entropy measures the number of (YES/NO) questions that you expect to have to ask in order to find out which state has occurred.
Information theory

You can distinguish $2^k$ states with $k$ (YES/NO) questions.

If the states are equally likely, then this is the best you can do.

It costs $k$ questions to identify a state from among $2^k$ equally likely states.
Information theory

It costs $k$ questions to identify a state from among $2^k$ equally likely states.

It costs $\log_2 n$ questions to identify a state from among $n$ equally likely states: to identify a state with probability $1/n$.

<table>
<thead>
<tr>
<th>Probability</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$1/n$</td>
<td>$\log_2 n$</td>
</tr>
<tr>
<td>$p$</td>
<td>$\log_2 (1/p)$</td>
</tr>
</tbody>
</table>
## The entropy

<table>
<thead>
<tr>
<th>State</th>
<th>Probability</th>
<th>Questions</th>
<th>Uncertainty</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_1$</td>
<td>$p_1$</td>
<td>$\log_2 (1/p_1)$</td>
<td>$p_1 \log_2 (1/p_1)$</td>
</tr>
<tr>
<td>$S_2$</td>
<td>$p_2$</td>
<td>$\log_2 (1/p_2)$</td>
<td>$p_2 \log_2 (1/p_2)$</td>
</tr>
<tr>
<td>$S_3$</td>
<td>$p_3$</td>
<td>$\log_2 (1/p_3)$</td>
<td>$p_3 \log_2 (1/p_3)$</td>
</tr>
</tbody>
</table>

Entropy = $p_1 \log_2 (1/p_1) + p_2 \log_2 (1/p_2) + p_3 \log_2 (1/p_3) + ...$
Continuous random variables

For a random variable $X$ with density $f$ the entropy is

$$\text{Ent}(X) = \int f \log f$$

The entropy behaves nicely under several natural processes: for example, the evolution governed by the heat equation.
If the density $f$ measures the distribution of heat in an infinite metal bar, then $f$ evolves according to the heat equation:

$$\frac{!f}{!t} = f''$$

The entropy increases:

$$\frac{!}{!t} \left( f \log f \right) = \frac{f'^2}{f} \# 0$$

**Fisher information**
The central limit theorem

If $X_i$ are independent copies of a random variable with mean 0 and finite variance, then the normalized sums

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

converge to a Gaussian (normal) with the same variance.

Most proofs give little intuition as to why.
The central limit theorem

Among random variables with a given variance, the Gaussian has largest entropy.

**Theorem** (Shannon-Stam) If $X$ and $Y$ are independent and identically distributed, then the normalized sum

\[
\frac{X + Y}{\sqrt{2}}
\]

has entropy at least that of $X$ and $Y$. 
Idea

The central limit theorem is analogous to the second law of thermodynamics: the normalized sums

\[ S_n = \frac{1}{\sqrt{n}} \prod_{i=1}^{n} X_i \]

have increasing entropy which drives them to an “equilibrium” which has maximum entropy.
**Problem:** (folklore or Lieb (1978)).

Is it true that $\text{Ent}(S_n)$ increases with $n$?

Shannon-Stam shows that it increases as $n$ goes from 1 to 2 (hence 2 to 4 and so on). Carlen and Soffer found uniform estimates for entropy jump from 1 to 2.

It wasn’t known that entropy increases from 2 to 3.

The difficulty is that you can't express the sum of 3 independent random variables in terms of the sum of 2: you can't add 3/2 independent copies of $X$. 
The Fourier transform?

The simplest proof (conceptually) of the central limit theorem uses the FT. If $X$ has density $f$ whose FT is $!$ then the FT of the density of $! X_i$ is $! n$.

The problem is that the entropy cannot easily be expressed in terms of the FT. So we must stay in real space instead of Fourier space.
Example:

Suppose $X$ is uniformly distributed on the interval between 0 and 1. Its density is:

When we add two copies the density is:
For 9 copies the density is a spline defined by 9 different polynomials on different parts of the range.

The central polynomial (for example) is:

\[
3 \left( 857291 - 5027400 \, x + 12800340 \, x^2 - 18438840 \, x^3 + 16391970 \, x^4 - 9185400 \, x^5 + 3163860 \, x^6 - 612360 \, x^7 + 51030 \, x^8 \right) / 4480
\]

and its logarithm is?
The second law of probability

A new variational approach to entropy gives quantitative measures of entropy growth and proves the “second law”.

**Theorem** (Artstein, Ball, Barthe, Naor) If $X_i$ are independent copies of a random variable with finite variance, then the normalized sums

$$\frac{1}{\sqrt{n}} \prod_{i=1}^{n} X_i$$

have increasing entropy.
**Starting point:** used by many authors. Instead of considering entropy directly, we study the Fisher information:

\[
J(X) = \frac{f'^2}{f}
\]

Among random variables with variance 1, the Gaussian has the **smallest** Fisher information, namely 1.

The Fisher information should decrease as a process evolves.
The connection (we want) between entropy and Fisher information is provided by the Ornstein-Uhlenbeck process (de Bruijn, Bakry and Emery, Barron).

Recall that if the density of $X^{(t)}$ evolves according to the heat equation then

$$\frac{\partial}{\partial t} \text{Ent}(X^{(t)}) = J(X^{(t)})$$

The heat equation can be solved by running a Brownian motion from the initial distribution. The Ornstein-Uhlenbeck process is like Brownian motion but run in a potential which keeps the variance constant.
The Ornstein-Uhlenbeck process

A discrete analogue:

You have $n$ sites, each of which can be ON or OFF. At each time, pick a site (uniformly) at random and switch it.

$$X^{(t)} = \text{(number on)} - \text{(number off)}.$$
The Ornstein-Uhlenbeck process

A typical path of the process.
The Ornstein-Uhlenbeck evolution

The density evolves according to the modified diffusion equation:

\[
\frac{!f}{!t} = f'' + (xf)'
\]

From this:

\[
! \frac{\text{Ent}(X^{(t)})}{!t} = J(X^{(t)}) \quad 1
\]

As \( t \rightarrow \infty \), the evolutes approach the Gaussian of the same variance.
The entropy gap can be found by integrating the information gap along the evolution.

\[ \text{Ent}(G) - \text{Ent}(X^{(0)}) = \mathcal{J}(X^{(t)}) \cdot 1 \]

In order to prove entropy increase, it suffices to prove that the information

\[ J(n) = J \frac{n!}{\text{\#}\sqrt{n}} \sum_{i=1}^{n} X_i \]

**decreases** with \( n \).

It was known (Blachman-Stam) that \( J(2) \cdot J(1) \).
Main new tool: a variational description of the information of a marginal density.

If $w$ is a density on $\mathbb{R}^n$ and $e$ is a unit vector, then the marginal in direction $e$ has density

$$ h(t) = \int_{te + \langle e \rangle^c} w $$
Main new tool:

The density $h$ is a marginal of $w$ and

$$J(h) = \frac{h'(t)^2}{h(t)} = \frac{h'(t)^2}{h(t)} ! h''(t) = h(! \log h)''$$

The integrand is non-negative if $h$ has concave logarithm.

Densities with concave logarithm have been widely studied in high-dimensional geometry, because they naturally generalize convex solids.
The Brunn-Minkowski inequality

Let $A(x)$ be the cross-sectional area of a convex body at position $x$.

Then $\log A$ is concave.

The function $A$ is a marginal of the body.
The Brunn-Minkowski inequality

We can replace the body by a function with concave logarithm. If $w$ has concave logarithm, then so does each of its marginals.

If the density $h$ is a marginal of $w$, the inequality tells us something about $(! \log h)$" in terms of $\text{Hess}(! \log w)$
The Brunn-Minkowski inequality

If the density $h$ is a marginal of $w$, the inequality tells us something about $\left( ! \log h \right)^2$ in terms of $\text{Hess}(! \log w)$.

We rewrite a proof of the Brunn-Minkowski inequality so as to provide an explicit relationship between the two. The expression involving the Hessian is a quadratic form whose minimum is the information of $h$.

This gives rise to the variational principle.
The variational principle

**Theorem** If \( w \) is a density and \( e \) a unit vector then the information of the marginal in the direction \( e \) is

\[
J(h) = \int \frac{h'(t)^2}{h(t)} dt = \min \int \frac{\text{div}(pw)^2}{w} dt
\]

where the minimum is taken over vector fields \( p \) satisfying \( !p, e \neq 1 \).
Technically we have gained because \( h(t) \) is an integral: not good in the denominator.

The real point is that we get to choose \( p \). Instead of choosing the optimal \( p \) which yields the intractable formula for information, we choose a non-optimal \( p \) with which we can work.
Proof of the variational principle.

\[ h(t) = \int_{e(t)} \omega \]

so

\[ h'(t) = \int_{e(t)} \nabla \cdot (p \omega) \]

If \( p, e \) satisfies \# 1 at each point, then we can realise the derivative as

\[ h'(t) = \int_{e(t)} \nabla \cdot (p \omega) \]

since the part of the divergence perpendicular to \( e \) integrates to 0 by the Gauss-Green (divergence) theorem.
Hence

\[ \frac{\partial^2 h'(t)}{\partial t^2} - \left( \frac{\partial}{\partial \theta} \text{div}(pw) \right)^2 dt = \frac{\partial}{\partial \theta} \text{div}(pw)^2 dt. \]

There is equality if

\[ \text{div}(pw) = \frac{h'(t)}{h(t)} w. \]

This divergence equation has many solutions: for example we might try the electrostatic field solution. But this does not decay fast enough at infinity to make the divergence theorem valid.
\[ \text{div}(pw) = \frac{h'(t)}{h(t)} w \]

The right solution for \( p \) is a flow in the direction of \( e \) which transports between the probability measures induced by \( w \) on hyperplanes perpendicular to \( e \).

For example, if \( w \) is 1 on a triangle and 0 elsewhere, the flow is as shown. (The flow is irrelevant where \( w = 0 \).)
The second law of probability

Theorem If $X_i$ are independent copies of a random variable with variance, then the normalized sums

$$\frac{1}{\sqrt{n}} \prod_{i=1}^{n} X_i$$

have increasing entropy.