In a Search for a Structure, Part 1: On Entropy.

Misha Gromov

June 25, 2013

Contents

Abstract

Mathematics is about "interesting structures". What make a structure interesting is an abundance of interesting problems; we study a structure by solving these problems.

The worlds of science, as well as of mathematics itself, is abundant with gems (germs?) of simple beautiful ideas. When and how may such an idea direct you toward beautiful mathematics?

I try to present in this talk a few suggestive examples.

1 States, Spaces, Crystals and Entropy.

What is the "number of states" of a (classical as opposed to quantum) system, $\mathcal{S}$, e.g. of a crystal?

A "naive physicist's" system is an infinite ensemble of infinitely small *mutually equal* "states", where you know nothing about what these states are but you believe they are equal because of *observable symmetries* of $\mathcal{S}$. The number of these states, although infinite, can be assigned a specific numerical value by comparing with another system taken for a "unit of entropy"; moreover, this number can be *measured experimentally* with a predictable error.

The logarithm of this number is called *(mean statistical Boltzmann) entropy of* $\mathcal{S}$, [14].

What is the "space of states" of $\mathcal{S}$? No such thing exists – would be our physicists answer (unless he/she is a Shroedinger's cat). Even the "number of states" – the value of entropy – may depend on accuracy of your data. This $\mathcal{S}$ is not a "real thing", nor is it a mathematician's "set", it is "something" that depends on a class of *mutually equivalent* imaginary experimental protocols.

This, probably, appeared gibberish to mathematicians of the late 19th century, when Boltzmann developed his concept of entropy, and even of the early 20th century, when Lebesgue (1902) and Kolmogorov (1933) expressed the ideas of measure and probability in the language of the set theory of Kantor (1873).

But now-a-days this "gibberish" (almost) automatically translates to the language of *non-standard analysis* (Abraham Robinson, 1966) [15], [18] and even easier to that of *category theory* (Eilenberg-MacLane-Steenrod-Cartan-Grothendieck's, 1945-1957).

For instance, our physists' description of a crystal (see below) amounts to Kolmogorov's theorem on *dynamic entropy of Bernoulli shifts* (1958) (that was originally motivated by Shannon's information theory, 1948). To see this, you just need to realize that "something" of a physicist, is a *covariant functor* from a suitable "category of protocols" to the category of sets – outcomes of experiments; all you have to do afterwards is to follow the guidelines prescribed by the *syntax* of category theory.

(Arguably, the category language, some call it "abstract", reflects mental undercurrents that surface as our "intuitive reasoning"; a comprehensive mathematical description of this "reasoning", will be, probably, even farther removed from the "real world" than categories and functors.)

Think of $\mathcal{S}$ is an "ensemble" of molecules located at some sites/points in the 3-dimensional Euclidean space, denoted $s \in S \subset \mathbb{R}^3$, e.g. at the *integer* points $s \in S = \mathbb{Z}^3 \subset \mathbb{R}^3$, (i.e. $s = (n_1, n_2, n_3)$ such that n_1, n_2, n_3 are integers) where each molecule may occupy finitely many, say k_s, different states, e.g. levels of energy characterized by k_s different colors. Represent such a molecule by a finite set of cardinality k_s and declare the Cartesian product of all these finite sets over $s \in S$ to be *the space of pure states of* $\mathcal{S}$; accordingly, let the product of numbers $\prod_{s \in S} k_s = \exp \sum_s \log k_s$ (which should be properly normalized for infinite S) represent the number of states of $\mathcal{S}$.

If, however, molecules exchange pure states exceptionally rarely, we shall see only one state, the perceived entropy will be zero: a state counts only if it visited by molecules with a definite frequency with emission/absorbtion of energy at the change of states that can be registered by experimantal devices.

If the molecules at all sites visit their individual states with equal relative frequencies $1/k_s$, then $\sum_s \log k_s$ is, indeed, a fair measure of entropy, provided the molecules *do not interact.* Yet, if particles do interact, where, for example, two neighbors in S only reluctuntly display the same color, then $\mathcal{S}$ will have fewer observable states. How to account for this?

And, remember, you have no direct access to the "total space of states", you do not observe individual molecules, you do not know, a priori, how molecules interact (if at all) and you do not even know what the numbers k_s are.

What you have at your disposal are certain devices – "state detectors", call them P, that are also "physical systems" but now with relatively few, say n_P, "pure states" in them. You may think of a P as a plate with an array of n_P windows that are sensitive to different "colors". When you "attach" P to $\mathcal{S}$ (you do not have to know the physical nature of this "attachment") you may see flashes of lights in these windows. But you yourself are color blind and you do not know beforehand if two windows have identical or different "colors". All you can do is to count the numbers of flashes in the windows at various (small or large) time intervals.

Moreover, given two P, you do not know if they have identical colors of their respective windows or different ones; yet, if a window P_2 is moved along S, by a symmetry, that is a group element $\gamma \in \Gamma = \mathbb{Z}^3$, then you *assume* that P_2 is "the same" as P_1.

You assign a number $|p|$ to each window p in a P attached to $\mathcal{S}$ that is the relative frequency of flashes in P; thus, $\sum_p |p| = 1$ for all windows. Then you postulate that "entropy of $\mathcal{S}$ perceived by P" call it $ent(P)$, is given by $ent(P) = -\sum_{p\in P}|p|\log|p|$ and you assume that the probability (relative frequency) of observing a sequence of flashes in given windows $p_1, p_2, ..., p_N \in P$ at consecutive time intervals is roughly $\exp(-N \cdot ent(P))$ for all sequences $p_1, p_2, ..., p_N$ of windows where flashes in this order *do* "realistically occur". (You can not experimentally verify this – the number $\exp(N \cdot ent(P))$ may be smaller then n_P^N but it is still huge.)

If you attach two plates P_1 and P_2 with n_{P_1} and n_{P_2} windows, you regard the pair as a new plate (state detector), denoted $P_1 \vee P_2$ with $n_{P_1} \cdot n_{P_2}$ windows. You count the numbers of flashes in the pairs of windows $(p_1 \in P_1, p_2 \in P_2)$ and thus define/determine the entropy $ent(P_1 \vee P_2)$.

A possible mathematical representation of a "state detector" P *attached* to $\mathcal{S}$ is a *finite measurable partition* $\sqcup_p X_p$ of a *measure space* $X = (X, \mu)$, i.e. $X = \sqcup_p X_p$, where $\mu(X) = |P|$, $\mu(X_p) = |p|$ and where $P_1 \vee P_2$ becomes $\sqcup_{p_1,p_2} X_{p_1} \cap X_{p_2}$.

But a *precise* definition of this is heavy: X is not *quite* a set but "something" associated with a *σ-algebra* Σ of *all* (more than continuum!) its subsets; to express this *rigorously* one needs the language of the Zermelo-Fraenkel set theory.

In mathematical practice, one takes a specific *model* of X, that is a topological space with a Borel measure on it, where X *is* represented by a set. This is similar to representation of vectors by n-tuples of numbers with a chosen coordinate system in a linear space.

On the other hand one can define "measure spaces" without introducing a particular set theoretic model as follows.

Finite Measure Spaces. A finite measure space $P = \{p\}$ is a finite set with a positive function denoted $p \mapsto |p| > 0$. We think of it as a set of atoms p that are one point sets with positive masses $|p|$ attached to them. We denote by $|P| = \sum_p |p|$ the (total) *mass* of P. If $|P| = 1$, then P is called a *probability space.*

We manipulate with spaces P as with their underlying sets, denoted $set(P)$, in-so-far as it does lead to confusion. For example, we speak of subsets $P' \subset P$, with mass $|P'| = \sum_{p\in P'} |p|$ and of *maps* $P \to Q$ that are maps $set(P) \to set(Q)$, etc.

Reductions and $\mathcal{P}$. Following physicists, we call a map $P \overset{f}{\to} Q$ a *reduction* if the *q-fibers* $P_q = f^{-1}(q) \subset P$ satisfy $|P_q| = |q|$ for all $q \in Q$. We also express this by saying that Q *is a reduction of* P. (Think of Q as a "plate with windows" through which you "observe" P. What you see of the states of P is what "filters" through the windows of Q.)

We use the notation $\mathcal{P}$ for *the category* with objects P and *reductions taken for morphisms.*

All morphisms in this category are epimorphisms, $\mathcal{P}$ looks very much as a partially ordered set (with $P \succ Q$ corresponding to reductions $f : P \to Q$ and few, if any, reductions between given P and Q) but we treat it for a time being as a general category.

Why Category? There is a subtle but significant conceptual difference between writing $P \succ Q$ and $P \overset{f}{\to} Q$. Physically speaking, there is no a priori given "attachment" of Q to P, an abstract "$\succ$" is meaningless, it must be implement by a particular operation f. (If one keeps track of "protocol of attaching Q to P", one arrives at the concept of 2-*category*.)

The f-notation, besides being more precise, is also more flexible. For example one may write $ent(f)$ but not $ent(\succ)$ with no P and Q in the notation.

Spaces over $\mathcal{P}$. A space $\mathcal{X}$ *over* $\mathcal{P}$ is, by definition, a *covariant functor* from $\mathcal{P}$ to the category of sets, where the value of $\mathcal{X}$ on $P \in \mathcal{P}$ is denoted $\mathcal{X}(P)$.

For example, if X is an ordinary measure space, then the corresponding $\mathcal{X}$ assigns the sets of (classes of) measure preserving maps (modulo ...) $f : X \to P$ to all $P \in \mathcal{P}$.

In general, an element f in the set $\mathcal{X}(P)$ can be regarded as a morphism $f : \mathcal{X} \to P$ in a category $\mathcal{P}^{\backslash \mathcal{X}}$ that is obtained by augmenting $\mathcal{P}$ with an object corresponding to $\mathcal{X}$, such that every object, in $\mathcal{P}^{\backslash \mathcal{X}}$ receives *at most one* (possibly none) morphism from $\mathcal{X}$. Conversely, every category extension written of $\mathcal{P}$ with such an object[1]defines a space over $\mathcal{P}$.

$\vee$-*Categories and Measure Spaces.* Given a set I of morphisms $f_i : x \to b_i$, $i \in I$, in a category, we call these x-*fans over* $\{b_i\}$, say that an a-fan $f'_i : a \to b_i$ lies *between* x and $\{b_i\}$ if there is a morphism $g : x \to a$ such that $f'_i \circ g = f_i$ for all $i \in I$. To abbreviate we may say "a between x and b_i".

Call $\mathcal{P}^{\backslash \mathcal{X}}$ a $\vee$-*category* if every $\mathcal{X}$-fan over finitely many $P_i \in \mathcal{P}$ admits a $Q \in \mathcal{P}$ between $\mathcal{X}$ and $\{P_i\}$.

Definition. An $\mathcal{X}$ over the category $\mathcal{P}$ of finite measure spaces P, is called a *measure space* if $\mathcal{P}^{\backslash \mathcal{X}}$ is a $\vee$-category.

Minimal Fans and Injectivity. An x-fan over b_i in a category is called *minimal* if every a between x and $\{b_i\}$ is isomorphic to x. (More precisely, the arrow $x \to a$ that implements "between" is an isomorphism.)

It is obvious that every $\mathcal{X}$-fan over finitely many finite measure spaces $P_i \in \mathcal{P}$ in a $\vee$-category over $\mathcal{P}$ admits a $Q \in \mathcal{P}$ between $\mathcal{X}$ and $\{P_i\}$, such that the corresponding Q-fan over P_i is minimal. This Q, when seen as an object in $\mathcal{P}$ is unique up to an isomorphism; the same Q is unique up to a *canonical isomorphism* in $\mathcal{P}^{\backslash X}$. We call this $\vee$-*(co)product* of P_i in $\mathcal{P}^{\backslash X}$ and write: $Q = \vee_i P_i$.

This product naturally/functorially extends to morphisms g in $\mathcal{P}^{\backslash \mathcal{X}}$, denoted

$$\vee_i g_i : \vee_i P_i \to \vee_i P'_i \text{ for given reductions } g_i : P_i \to P'_i.$$

Observe that the $\vee$-product is defined (only) for those objects and morphisms in $\mathcal{P}^{\backslash \mathcal{X}}$ that lie under $\mathcal{X}$.

An essential feature of *minimal* fans, say $f_i : Q \to P_i$, a feature that *does not depend* on $\mathcal{X}$ (unlike the $\vee$-product itself) is the *injectivity* of the corresponding (set) map from Q to the Cartesian product $\prod_i P_i$ (that, in general, is *not* a reduction).

Let us express the idea of "number of states" and/or of entropy – logarithm of this number, in *rigorous* terms by reformulating *Bernoulli's law of large numbers* (1713) in the language of $\mathcal{P}$ as follows.

Cartesian product: $P \times Q$ is the is the set of pairs of atoms (p, q) that are given the weights $|p| \cdot |q|$ and denoted $pq = (p, q)$. (This corresponds to observing

[1]This, as was pointed out to me by Thomas Riepe, is uncarefully written. In order to have the "*at most one*" property each $P \in \mathcal{P}$, must appear in the category $\mathcal{P}^{\backslash \mathcal{X}}$ in several "copies" indexed by the set $\mathcal{X}(P)$.

non-interacting "somethings" with P and Q.) The maps $pq \mapsto p$ and $pq \mapsto q$ are called *Cartesian projections* $P \times Q \to P, Q$.

Notice that, say $pq \mapsto p$, is a reduction *only* if Q is a *probability* space. In general, one may rescale/normalize the spaces and make these maps reductions. Such a rescaling, being a non-trivial symmetry, is a significant structure in its own right; for example, the group of families of such "rescalings" leads the amazing orthogonal symmetry of the *Fisher metric* (see section 2); you are not supposed to say "rescale" and forget about it.)

Homogeneous Spaces. A finite measure space P is called *homogenous* if all atoms $p \in P$ have equal masses $|p|$. (Categorically speaking, all morphisms $P \to Q$ that are invariant under the group of automorphisms of P factor through $P \to \bullet$ for terminal objects $\bullet \in \mathcal{P}$, that are monoatomic spaces.)

Entropy of a homogeneous P is defined as the logarithm of the cardinality of $set(P)$, that is $ent(P) = \log|set(P)|$.

Observe that reductions $f : P \to Q$ between homogeneous spaces (non-canonononically) *split*, that is P decomposes into Cartesian product $P = P' \times Q$ where the projection $P \to Q$ equals f.

$dist_\pi(P,Q)$ *and Asymptotic Equivalence.* Let P and Q be finite measure spaces, and let $\pi : P \to Q$ be an *injective correspondence* that is a partially defined bijective map defined on a subset $P' \subset P$ that is bijectively sent to $Q' \subset Q$. Let us introduce a numerical measure of deviation of π from being an isomorphism. To simplify, we assume P and Q are *probability spaces*, i.e. $|P| = |Q| = 1$, otherwise, *normalize* them by $p \mapsto p/|P|$ in P and $q \mapsto q/|Q|$ and denote

$$|p : q| = \max(p/q, q/p) \text{ for } q = \pi(p) \text{ and } M = \min(|set(P)|, |set(Q)|).$$

Let

$$|P - Q|_\pi = |P \setminus P'| + |Q \setminus Q'|,$$

$$|\log P : Q|_\pi = \sup_{p \in P'} \frac{\log|p : q|}{\log M}, \text{ where } 0/0 =_{def} 0,$$

and

$$dist_\pi(P,Q) = |P - Q|_\pi + |\log P : Q|_\pi.$$

Call sequence of injective correspondences $\pi_N : P_N \to Q_N$ an *asymptotic equivalence* if

$$dist_{\pi_N}(P_N, Q_N) \underset{N\to\infty}{\to} 0$$

and say that two sequences of finite measure spaces P_N and Q_N are *asymptotically equivalent* if there exists an asymptotic equivalence $\pi_N : P_N \to Q_N$.

The law of large numbers applied to the random variable $p \to \log p$ on P, can be stated as follows.

Bernoulli Approximation Theorem.[2] *The sequence of Cartesian powers P^N of every $P \in \mathcal{P}$ admits an asymptoticly equivalent sequence H_N of homogeneous spaces.*

Such a sequence H_N is called a *homogeneous Bernoulli approximation* of P^N.

[2] This is often called *asymptotic equipartition property.*

Bernoulli Entropy. *This is defined as*

$$ent(P) = \lim_{N\to\infty} N^{-1} \log |set(H_N)|$$

for a homogeneous sequence H_N that is asymptotically equivalent to P^N.

Entropy can be also defined without an explicit use of Bernoulli theorem as follows.

Call probability spaces P_1 and P_2 *Bernoulli equivalent* if the power sequences P_1^N and P_2^N are asymptotically equivalent. The set $Ber(\mathcal{P})$ of the classes of probability spaces $P \in \mathcal{P}$ under this equivalence carries a natural structure of commutative semigroup corresponding to the Cartesian product $P \times Q$ as well as a topology for the metric $\limsup_{N\to\infty} dist_{\pi_N}(P^N, Q^N)$.

Boltzmann entropy. This, by definition, is the Bernoulli class of P in $Ber(\mathcal{P})$.

A posteriori, the law of large numbers shows that this is equivalent to Bernoulli's definition:

two finite probability spaces P and Q are Bernoulli equivalent if and only if they have equal Bernoulli's entropies.

More precisely,

There is a the topological isomorphism of the Bernoulli (Grothendieck) semigroup $Ber(\mathcal{P})$ onto the multiplicative semigroup $\mathbb{R}^{\times}_{\geq 1}$ of real numbers ≥ 1 that extend the homomorphism $H \mapsto |set(H)|$ for homogeneous spaces $H \in \mathcal{P}$.

The Bernoulli-Boltzmann entropy is then recaptured by composing this isomorphism with $\log : \mathbb{R}^{\times}_{\geq 1} \to \mathbb{R}_{+}$. (The mathematical significance of this log is not apparent until you give a close look at the Fisher metric.)

Boltzmann Formula:

$$ent(P) = -\sum_{p\in P} |p| \log |p| \quad \text{for all finite probability spaces } P = \{p\} \in \mathcal{P}.$$

If $|P| \neq 1$, then

$$ent(P) = -\sum_{p} \frac{|p|}{|P|} \log \frac{|p|}{|P|} = |P|^{-1} \left(-\sum_{p} |p| \log |p| \right) + \log |P|.$$

This is obvious with Bernoulli's approximation theorem but the original $ent(P) = -K \sum_p |p| \log |p_i|$, where K is the unit conversion constant, is by no means obvious: it provides a non-trivial link between microworld on the $10^{-9\pm1}m$ scale with what we see with the naked eye.

Bernoulli-Boltzmann's definition (unlike $-\sum_p |p| \log |p|$) fully and rigorously expresses the idea that entropy equals the logarithm of the "number of mutually equal states encoded/detected by P" and, thus, makes essential properties of entropy quite transparent. (There is also an information-theoretic rendition of Boltzmann's argument, often presented as a "bits bargaining" between "Bob and Alice". Probably, it is understandable by those who is well versed in the stock marked.) For example, one immediately sees the following

$(\log n)$-*Bound*: $ent(P) \leq \log |set(P)|$ with the equality for homogenous spaces with n equal atoms, since the powers P^N "Bernoulli converge" to measures with "nearly equal" atoms on *subsets* $S_N \subset set(P)^N$, that have cardinalities $|S_N| \leq |set(P)|^N$ and where $\frac{1}{N}\log|S_N| \to ent(P)$.

(Text-book proofs, where $-\sum|p|\log|p|$ is taken for the *definition* of entropy, rely on convexity of $x\log x$. In fact, this convexity follows from the law of large numbers, but the *sharpness* of the $\log n$-bound, that is the implication

$$ent(P) = \log|set(P)| \Rightarrow P \text{ is homogeneous},$$

is better seen with $\sum|p|\log|p|$, where real analyticity of $\log x$ implies sharpness of this $(\log n)$-inequality. Also Boltzmann's formula implies *continuity* of entropy as a function of $|p| \leq 0$, $p \in P$.)

Functorial Bernoulli. The law of large numbers not only (trivially) yields Bernoulli approximation of objects (finite measure spaces) in $\mathcal{P}$, but also approximation of reduction (morphisms) in $\mathcal{P}$. Namely,

given a reduction $f : P_1 \to P_2$, there exists a sequence of reductions $\phi_N : H_{1N} \to H_{2N}$, where H_{1N} and H_{2N} are homogeneous Bernoulli approximations of P_1^N and of P_2^N.

We call this *a homogenous Bernoulli approximation of the Cartesian powers $f^N : P_1^N \to P_2^N$ of f.*

The existence of such approximation immediately implies, for example, that

Entropy is monotone decreasing under reductions: if P_2 is a reduction of P_1 then $ent(P_2) \leq ent(P_1)$; in particular, $ent(P \vee Q) \geq ent(P)$ for all P and Q in $\mathcal{P}^{\backslash\mathcal{X}}$ under $\mathcal{X}$.

Let $\{f_i\}$, $i \in I$ be a finite set of reductions between some objects P in $\mathcal{P}$. Ideally, one would like to have to have homogeneous Bernoulli approximations ϕ_{iN} of all f_i^N, such that

$$[BA]_1 \qquad [f_i = f_j \circ f_k] \Rightarrow [\phi_{iN} = \phi_{jN} \circ \phi_{kN}],$$

and such that injectivity/minimality of all fans is being preserved, i.e.

$$[\mathrm{BA}]_2 \qquad \text{minimality of } f_{i_\nu} : P \to Q_\nu \Rightarrow \text{minimality of } \phi_{i_\nu N} : H_N \to H_{i_\nu N}.$$

Probably, this is not always possible (I have no specific counterexample), but one can achieve this with the following weaker assumption on the approximating sequences.

Call a sequence $B_N = \{b_N\}$ of finite measure spaces *Bernoulli* if it is ε_N-*homogeneous* for some sequence $\varepsilon_N \underset{N\to\infty}{\to} 0$. This means that the atoms b_N in all B_N satisfy:

$$\frac{1}{N}|\log|b_N| + \log|set(B_N)|| \leq \varepsilon_N + \frac{1}{N}\log|B_N|.$$

A Bernoulli approximation of P^N is a Bernoulli sequence B_N that is asymptotically equivalent to P^N; accordingly, one defines Bernoulli approximation ϕ_N of powers f^N of reductions $f : P \to Q$.

Now it is easy to see (as in the slice removal lemma from [11]) that the above $\{f_i\}$ do admit Bernoulli (not necessarily homogeneous) approximations that satisfy $[\mathrm{BA}]_1$ and $[\mathrm{BA}]_2$.

Shannon Inequality. If a fan $\phi_i : H_0 \to H_i$ of homogeneous spaces is minimal/injective, i.e. the Cartesian product map $\times_i\phi_i : H_0 \to \times_i H_i$ is injective, then, obviously, $|set(H_0)| \leq \prod_i |set(H_i)|$ and $ent(H_0) \leq \sum_i ent(H_i)$.

This, applied to a (ε_N-homogeneous) Bernoulli approximation of a minimal/injective fan $P_0 \to P_i$ of arbitrary finite measure spaces, shows that $ent(P_0) \le \sum_i ent(P_i)$.

In particular, if $P_i \in \mathcal{P}^{\backslash \mathcal{X}}$ lie under $\mathcal{X}$ (e.g. being represented by finite partitions of an ordinary measure space X), then

$$ent(\vee_i P_i) \le \sum_i ent(P_i).$$

The above argument, going back to to Boltzmann and Gibbs, is a translation of a naive physicist's reasoning to mathematical language. In fact, this $\vee$, physically speaking, is a kind of a sum, the result of pooling together the results of the joint entropy count by all P_i.

If all P_i positioned far away one from another on your, say, crystal, then you assume (observe?) that flashes of lights are (essentially) independent: $\vee_i P_i = \prod_i P_i$ and $ent(\vee_i P_i) = \sum_i ent(P_i)$.

In general however, the observable states may constrain one another by mutual interaction; then, there are less states to observe and $ent(\vee_i P_i) < \sum_i ent(P_i)$ in agreement with experiment.

Relative Entropy. Since the fibers $G_h = \phi^{-1}(h) \subset G$, $h \in H$, of a reduction $\phi : G \to H$ between homogeneous spaces have equal cardinalities, one may define entropy of ϕ by $ent(\phi) = \log|set(G_h)|$, where, obviously, this entropy satisfies: $ent(\phi) = ent(G) - ent(H)$.

Then one defines *relative Boltzmann's entropy* $ent(f)$ of a reduction $f : P \to Q$ between arbitrary finite measure spaces via a homogeneous Bernoulli approximation $\phi_N : G_N \to H_N$ of f^N as

$$ent(f) = \lim_{N\to\infty} N^{-1} ent(\phi_N).$$

Alternatively, one can do it in more abstract fashion with the *relative (Grothendieck) Bernoulli semigroup* $\overrightarrow{Ber}(\mathcal{P})$ generated by classes $[f]$ of asymptotic equivalence of reductions $f \in \mathcal{P}$ with the addition rule $[f_1 \circ f_2] = [f_1] + [f_2]$ (compare [1], [16]).

Relative Shannon inequality. It is clear with Bernoulli approximation as in the absolute case that reductions $f_i : P_i \vee Q_i \to Q_i$ in $\mathcal{P}^{\backslash \mathcal{X}}$ for spaces P_i and Q_i under $\mathcal{X}$ satisfy:

$$ent(\vee_i f_i) \le \sum_i ent(f_i).$$

Since $ent(f_i) = ent(Q_i \vee P_i) - ent(Q_i)$, this is equivalent to

$$ent(\vee_i (Q_i \vee P_i)) - ent(\vee_i Q_i) \le \sum_i [ent(Q_i \vee P_i) - ent(Q_i)].$$

Alternatively, one can formulate such an inequality in terms of minimal/injective fans of reductions $P \to Q_i$, $i = 1, 2, ..., n$, coming along with (cofans of) reductions $Q_i \to R$, such that the obvious diagrams commute:

$$ent(P) + (n-1)ent(R) \le \sum_i ent(Q_i).$$

Another pleasant, albeit obvious (with Bernoulli), feature of the relative entropy of reductions $f : P \to Q$ between *probability spaces* is the representation of $ent(f)$ by the convex combination of the entropies of the q-fibers $P_q = f^{-1}(q) \subset P$, $q \in Q$.

Summation Formula: $ent(f) = \sum_q |q| \cdot ent(P_q)$.

Remark. The above definition of $ent(f)$ is applicable to $f : P \to Q$, where P and Q are *countable* probability (sometimes more general) spaces. possibly, with $ent(P) = \infty$ and $ent(Q) = \infty$ where the formula $ent(f) = ent(P) - ent(Q)$ serves as a *definition* of the difference between these two infinities.

Resolution of Infinite Spaces $\mathcal{X}$. Let $\mathcal{P}^{\backslash\mathcal{X}}$ be the $\vee$-category associated with $\mathcal{X}$ and let us formalize the notion of "equivalent protocols" of our physicist with sequences $P_\infty = \{P_i\}$ of finite objects in $\mathcal{P}^{\backslash\mathcal{X}}$, i.e. of finite measure spaces. Say that P_∞ *resolves* a finite measure space $Q \in \mathcal{P}^{\backslash\mathcal{X}}$ that lies under $\mathcal{X}$ if there is no eventual gain in state detection if you include Q into your protocol:

$$ent(Q \vee P_i) - ent(P_i) \le \varepsilon_i \underset{i\to\infty}{\to} 0.$$

If P_∞ resolves all Q, then, by definition, it is a *resolution* of $\mathcal{X}$.

Infinite Products. Say that $\mathcal{X}$ is *representable* by a (usually countable) *Cartesian product* $P_s \in \mathcal{P}^{\backslash\mathcal{X}}$, $s \in S$, briefly, $\mathcal{X}$ *is a Cartesian product* $\prod_{s\in S} P_s$, if the *finite* Cartesian products $\Pi_T = \prod_{s\in T} P_s$, $s \in T$, *lie under* $\mathcal{X}$ *for all finite subsets* $T \subset S$ and if these Π_T *resolve* $\mathcal{X}$, namely, some sequence Π_{T_i} resolves $\mathcal{X}$. (The subsets $T_i \subset S$ exhaust S in this case.)

Examples. A product $\mathcal{X} = \prod_{s\in S} P_s$ is called *minimal* if a Q in $\mathcal{P}^{\backslash\mathcal{X}}$ lies under $\mathcal{X}$ if and only if it lies under some finite product Π_T. For instance, all Q under the minimal Cartesian power $\{\frac{1}{2}, \frac{1}{2}\}^S$ are composed of dyadic atoms.

The classical *Lebesgue-Kolmogorov product* $X = \prod_{s\in S} P_s$ is also a product in this sense, where the resolution property is a reformulation of *Lebesgue density theorem*, where translation

Lebesgue's density$\Rightarrow$ resolution

goes with the following evident property of relative entropy:

Let $P \overset{f}{\leftarrow} R \to Q$ be a *minimal* R-fan of reductions, let $P' \in P$ be a subspace, denote by $R_{p'} = f^{-1}(p') \subset R$, $p' \in P'$, the p'-fibers of f and let $M_{II}(p')$ be the mass of the *second greatest* atom in $R_{p'}$.

If

$$|P \smallsetminus P'| \le \lambda \cdot |P| \quad and \quad M_{II}(p') \le \lambda |R_{p'}|$$

for some (small) $0 \le \lambda < 1$ *and all* $p' \in P'$, *then*

$$ent(f) \le (\lambda + \varepsilon) \cdot |set(Q)| \; for \; \varepsilon = \varepsilon(\lambda) \underset{\lambda\to 0}{\to} 0.$$

(Secretly, $\varepsilon \le \lambda \cdot (1 - \log(1 - \lambda))$ by Boltzmann formula.)

To see this, observe that $ent(R_p) \le |set(R_p| \le |set(Q|$ for all $p \in P$, that $ent(R_{p'}) \le \varepsilon \underset{\lambda\to 0}{\to} 0$ by continuity of entropy for $M_{II}(p') \to 0$ and conclude by using summation formula.

Normalization and Symmetry. All of the above properties of entropy of finite spaces appear in Shannon's information theory. (Probably, this was known to Boltzmann and Gibbs who had never explicitly formulated something so physically obvious.) Granted these, we can now understand what "naive physicist" was trying to say.

Infinite systems/spaces $\mathcal{X}$ have infinite entropies that need be renormalized, e.g. with some "natural" approximation of $\mathcal{X}$ by finite spaces P_N, such that

$$\text{"}ent(\mathcal{X}:size)\text{"} = \lim_{N\to\infty} \frac{ent(P_N)}{\text{"}size\text{"}(P_N)}.$$

The simplest case where "size" makes sense is when you state detector P_N consists of, say k, "identical parts"; then you may take k for the size of P_N. Physically speaking, "identical" means "related by symmetries" of $\mathcal{X}$ with detectors attached to it, e.g. for $\mathcal{X}$ corresponding to a crystal $\mathcal{S}$.

With this in mind, take a finite P and apply several (eventually many) symmetry transformations δ of $\mathcal{X}$ to P (assuming these symmetries exist), call the set of these transformation Δ_N, denote by $|\Delta_N|$ its cardinality and let

$$ent_P(\mathcal{X}:\Delta_\infty) = \lim_{N\to\infty} |\Delta_N|^{-1} ent\left(\bigvee_{\delta\in\Delta_N} \delta(P)\right)$$

for some sequence Δ_N with $|\Delta_N| \to \infty$ where a sublimit (let it be physically meaningless) will do if there is no limit. (Caution: transformations of categories are functors, not maps, but you can easily define them as maps in $\mathcal{P}^{\backslash\mathcal{X}}$.) The answer certainly will depend on $\{\Delta_N\}$ but what concerns us at the moment is dependence on P. A single P, and even all $\bigvee_{\delta\in\Delta_N} \delta(P)$ may not suffice to fully "resolve" $\mathcal{X}$. So we take a resolution $P_\infty = \{P_i\}$ of $\mathcal{X}$ (that, observe, has nothing to do with our transformations) and define

$$ent(\mathcal{X}:\Delta_\infty) = ent_{P_\infty}(\mathcal{X}:\Delta_\infty) = \lim_{i\to\infty} ent_{P_i}(\mathcal{X}).$$

This, indeed, *does not depend* on P_∞. If $Q_\infty = \{Q_i\}$ is another resolution (or any sequence for this matter), then the entropy contribution of each Q_j to P_i, that is the difference $ent(P_i \vee Q_j) - ent(P_i)$ is smaller than $\varepsilon_i = \varepsilon(j,i) \underset{i\to\infty}{\to} 0$ by the above definition of resolution.

Since δ are automorphisms, the entropies do not change under δ-moves and

$$ent(\delta(P_i) \vee \delta(Q_j)) - ent(\delta(P_i)) = ent(P_i \vee Q_j) - ent(P_i) \le \varepsilon_i;$$

therefore, when "renormalized by size" of Δ_N, the corrspomding $\vee$-products satisfy the same inequality:

$$|\Delta_N|^{-1}\left(ent\left[\bigvee_{\delta\in\Delta_N}(\delta(P_i)\vee\delta(Q_j))\right] - ent\left[\bigvee_{\delta\in\Delta_N}\delta(P_i)\right]\right) \le \varepsilon_i \underset{i\to\infty}{\to} 0$$

by the relative Shannon inequality.

Now we see that adding $Q_1, Q_2, ..., Q_j$ to P_∞ does not change the above entropy, since it is defined with $i \to \infty$ and adding all of Q_∞ does not change it either. Finally, we turn the tables, resolve P_j by Q_i and conclude that P_∞ and Q_∞, that represent "equivalent experimental protocols", give us the same entropy:

$$ent_{P_\infty}(\mathcal{X}:\Delta_\infty) = ent_{Q_\infty}(\mathcal{X}:\Delta_\infty),$$

as our physicist has been telling us all along.

Kolmogorov Theorem for Bernoulli Systems. Let P be a finite probability space and $X = P^{\mathbb{Z}}$. This means in our language that the corresponding $\mathcal{X}$ is representable by a Cartesian power $P^{\mathbb{Z}}$ with the obvious (Bernoulli) action of $\mathbb{Z}$ on it.

If spaces $P^{\mathbb{Z}}$ and $Q^{\mathbb{Z}}$ are $\mathbb{Z}$-equivariantly isomorphic then $ent(P) = ent(Q)$.

Proof. Let P_i denote the Cartesian Power $P^{\{-i,...0,...i\}}$, let $\Delta_N = \{1,...N\} \subset \mathbb{Z}$, observe that

$$\bigvee_{\delta\in\Delta_N} \delta(P_i) = P^{\{-i,...,i+N\}}$$

and conclude that $ent(\bigvee_{\delta\in\Delta_N} \delta(P_i)) = (N+i)ent(P)$ for all, $i = 1,2,...$. Therefore,

$$ent_{P_i}(X:\Delta_\infty) = \lim_{N\to\infty} N^{-1} ent(\bigvee_{\delta\in\Delta_N} \delta(P_i)) = \lim_{N\to\infty} \frac{N+i}{N} ent(P) = ent(P)$$

and

$$ent(X:\Delta_\infty) = \lim_{i\to\infty} ent_{P_i}(X:\Delta_\infty) = ent(P).$$

Similarly, $ent(Q^{\mathbb{Z}}:\Delta_\infty) = ent(Q)$ and since $P^{\mathbb{Z}}$ and $Q^{\mathbb{Z}}$ are $\mathbb{Z}$-equivariantly isomorphic, $ent(P^{\mathbb{Z}}:\Delta_\infty) = ent(Q^{\mathbb{Z}}:\Delta_\infty)$; hence $ent(P) = ent(Q)$. QED.

Discussion (A) The above argument applies to all *amenable* (e.g. Abelian) groups Γ (that satisfy a generalized "$(N+i)/N \to 1$, $N\to\infty$" property) where it also shows that

if Q^Γ is a Γ-reduction of P^Γ then $ent(Q) \leq ent(P)$.

("Reduction" means that Q^Γ receives a Γ-equivariant measure preserving map from P^Γ that is a natural transformation of functors represented by the two Γ-spaces.)

To our "naive physicist's" surprise, the invariance of entropy for "Bernoulli crystals", was accepted by mathematicians not around 1900 but in 1958 (see [13] for how it was going after 1958).

Had it taken so long because mathematicians were discouraged by the lack of "rigor" in physicists' reasoning? But had this been already known to physicists, rigor or no rigor? (A related result– the existence of thermodynamic limit for a physically significant class of systems, was published by Van Hove in 1949, but no self-respecting physicist, even if he/she understood it, would not care/dare to write anything about one-dimensional systems like $P^{\mathbb{Z}}$ with no interaction in them.)

Maybe, the simplicity of Kolmogorov's argument and an apparent inevitability with which it comes along with translation of "baby-Boltzmann" to "baby-Groethendieck" is illusory. An "entropy barrier" on the road toward a conceptual proof (unlike the "energy barrier" surrounding a "hard proof") may remain unnoticed by one who follows the marks left by a pathfinder that keep you on the track through the labyrinth under the "mountain of entropy".

All this is history. The tantalizing possibility suggested by entropy – this is the main reason for telling the story – is that there may be other "little somethings" around us the mathematical beauty of which we still fail to recognize because we see them in a curved mirror of our preconceptions.

Ultralimits and Sofic Groups. In 1987, Ornstein and Weiss constructed Γ-equivariant continuous surjective, hence measure preserving, *group homomorphisms* $A^\Gamma \to (A\times A)^\Gamma$, for all *free non-cyclic* groups Γ and all finite Abelian groups A.

It is unknown, in general, when there is such a continuous surjective (injective, bijective) Γ-equivariant homomorphism $A^\Gamma \to B^\Gamma$ for given Γ and compact (e.g. finite) groups A and B but many significant examples of "entropy increasing" Γ-reductions for non-amenable groups Γ are constructed in [2], and a general result of this kind is available [7] for groups Γ that contain *free subgroups.* For example,

if $\Gamma \supset F_2$, *then there exists a* Γ*-reduction* $P_1^\Gamma \to P_2^\Gamma$ *for all finite probability spaces* P_1 *and* P_2 *except for the trivial case where* P_1 *consists of a single atom.* (In view of [2], this is likely to be true for all non-amenable groups.)

But, amazingly, this was shown by Bowen in 2010,

a Γ*-isomorphism between Bernoulli systems* $P_1^\Gamma \leftrightarrow P_2^\Gamma$ *implies that* $ent(P_1) = ent(P_2)$ *for a class of non-amenable groups* Γ*, including, for example, all residually finite groups such as free groups.*

One can arrive at this class of groups, they are called *sofic* following Weiss (2000), by implementing "naive physicist's reasoning", (probably close to what Boltzmann had in mind) in terms of non-standard analysis, namely, by completing $\mathcal{P}$ not with "projective-like limit spaces" $\mathcal{X}$ but with "non-standard spaces" that are objects in a non-standard model $\mathcal{P}^*$ of the $\mathbb{R}$-valued first order language of $\mathcal{P}$ that can be represented as an ultra limit (or ultra product) of $\mathcal{P}$ as it is done by Pestov in [19].

Roughly, objects P in $\mathcal{P}^*$ are, collection of N atoms of weights $|p|$ where N is an infinitely large non-standard integer, $|p|$ are positive infinitesimals and where the sum $\sum_P |p|$ is an ordinary real number. Then sofic groups are defined as subgroups of automorphism groups of such spaces.

These groups seem rather special, but there is no example at the moment of a countable non-sofic group. Probably, suitably defined random groups [17] are non-sofic. On the other hand, there may be a meaningful class of "random Γ-spaces" parametrized by the same probability measure (space) as random Γ.

In 2010, Bowen introduced a spectrum of sofic entropies (with some properties reminiscent of *von-Neimann entropy*) and proved, in particular, that

minimal/injective fans of reduction of Bernoulli systems $P^\Gamma \to Q_i^\Gamma$, $i = 1, 2, ..., n$, *for sofic groups* Γ *satisfy Shannon's inequality*

$$ent(P) \le \sum_{i=1,\dots,n} ent(Q_i).$$

Moreover, let $n = 2$ and let $Q_1^\Gamma \to R^\Gamma$, $Q_2^\Gamma \to R^\Gamma$ be reductions, such that the "cofan" $Q_1^\Gamma \to R^\Gamma \leftarrow Q_2^\Gamma$ is *minimal* (i.e. there is no Γ-space R_o between Q_i^Γ and R^Γ) and such that the diamond diagram with four arrows, $P^\Gamma \rightrightarrows Q_i^\Gamma \rightrightarrows R^\Gamma$, $i = 1, 2$, *commutes.*

Then the four Γ-systems in this diamond satisfy the *Relative Shannon Inequality*, that is also called *Strong Subadditivity*:

$$ent(P) + ent(R) \le ent(Q_1) + ent(Q_2).$$

(This and everything else we say about sofic entropies, was explained to me by Lewis Bowen.)

It may be non-surprising that Shannon inequalities persist in the sofic Γ-spaces categories, since Shannon's inequalities were derived by Bernoulli approximation from the implication $[A$ injects into $B] \Rightarrow |A| \le |B|$, rather than from [$B$ surjects onto A] $\Rightarrow |A| \leq |B|$, but it is unknown if these inequalities are true/false for a single non-sofic group Γ, assuming such a Γ exists.

To get an idea why "injective" rather than "surjective"plays a key role in the sofic world, look at a self-map of a set, say $f : A \to A$. If A is finite, then

$$[f \text{ is non-injective}] \Leftrightarrow [f \text{ is non-surjective}],$$

but it is not so anymore if we go to some kind of (ultra)limit:

(i) non-injectivity says that an equation, namely, $f(a_1) = f(a_2)$, *does admit* a non-trivial solution. This is stable under (ultra)limits and, supported by a counting argument (of the "numbers of pairs of pure states" that come together under fans of reductions), seems to underly some of Bowen's entropies.

(ii) non-surjectivity says that another equation, $f(a_1) = b$, *does not always admit* a solution. New solutions may appear in the (ultra)limit.

(Computationally speaking, you may have a simple rule/algorithm for finding solutions of $f(a_1) = f(a_2)$ with $a_1 \neq a_2$, e.g. for a polynomial, let it be even linear, self-map of a vector space over $\mathbb{F}_p$ but it may be harder to obtain an effective description of the image $f(A) \subset A$ and even less so of its complement $A \setminus f(A)$.)

Questions. Is there a description/definition of (some of) sofic entropies in categorical terms?

More specifically, consider the category of Lebesgue probability Γ-spaces $\mathcal{X}$ for a given countable group Γ and let $[\mathcal{X} : \Gamma]$ be the Grothendieck (semi) group generated by Γ-reductions f with the relations $[f_1 \circ f_2] = [f_1] + [f_2]$, where, as usual, the Γ-spaces themselves are identified with the reductions to one point spaces. How large is this semigroup? When is it non-trivial? Which part of it is generated by the Bernoulli shifts?

If we do not require any continuity property, this group may be huge; some continuity, e.g. under projective limits of reductions (such as *infinite* Cartesian products), seems necessary, but it is unclear what should be a Γ-counterpart of the the asymptotic equivalence.

Also some extra conditions, e.g. additivity for Cartesian products: $[f_1 \times f_2] = [f_1] + [f_2]$, or at least, $[f^N] = N[f]$ for Cartesian powers may be needed.

Besides, the semigroup $[\mathcal{X} : \Gamma]$ must(?) carry a partial order structure that should satisfy (at least some of) the inequalities that hold in $\mathcal{P}$, e.g. the above Shannon type inequalities for minimal/injective fans. (I am not certain if there are entropy inequalities for more complicated diagrams/quivers in $\mathcal{P}$ that do not *formally* follow from Shannon inequalities, but if there are any, they may be required to hold in $[\mathcal{X} : \Gamma]$.)

The most naive entropy invariant that should be expressible in terms $[\mathcal{X} : \Gamma]$ is the infimum of entropies of generating partitions, or rather, the infimum of $(\log M)/N$, such that the Cartesian power $(\mathcal{X}^N, \Gamma)$ is isomorphic to the Bernoulli action of Γ on the *topological* infinite power space $Y = \{1, 2, ..., M\}^\Gamma$ with *some* Γ-invariant measure Borel probability on Y (that is not necessarily the Cartesian power measure).

One may expect (require?) the (semi)group $[\mathcal{X} : \Gamma]$ to be functorial in Γ, e.g. for equivariant reductions $(\mathcal{X}_1, \Gamma_1) \to (\mathcal{X}_2, \Gamma_2)$ for homomorphisms $\Gamma_1 \to \Gamma_2$ and/or for several groups Γ_i acting on an $\mathcal{X}$, in particular, for Bernoulli shifts on P^Δ for Γ_i transitively acting on a countable set Δ.

2 Fisher Metric and Von Neumann Entropy.

Let us ponder over Boltzmann's function $e(p) = \sum_i p_i \log p_i$. All our inequalities for the entropy were reflections of the convexity of this $e(p)$, $p = \{p_i\}$, $i \in I$, on the unit simplex $\triangle(I)$, $\sum_i p_i = 1$, in the positive cone $\mathbb{R}^I_+ \subset \mathbb{R}^I$.

Convexity translates to the language of calculus as positive definiteness of Hessian $h = Hess(e)$ on $\triangle(I)$; following Fisher (1925) let us regard h as a Riemannian metric on $\triangle(I)$.

Can you guess how the Riemannian space $(\triangle(I), h)$ looks like? Is it metrically complete? Have you ever seen anything like that?

In fact, the Riemannian metric h on $\triangle(I)$ has *constant* sectional curvature, where the *real moment* map $M_\mathbb{R} : \{x_i\} \to \{p_i = x_i^2\}$ is, up to 1/4-factor, an *isometry* from the positive "quadrant" of the unit Euclidean sphere onto $(\triangle(I), h)$. Unbelievable! Yet this trivially follows from $(p \log p)'' = 1/p$, since the Riemannian metric induced by $M_\mathbb{R}^{-1}$ at $\{p_i\}$ equals

$$\sum_i (d\sqrt{p_i})^2 = \sum_i dp_i^2/4p_i.$$

This $M_\mathbb{R}$ extends to the (full) *moment map*

$$M : \mathbb{C}^I \to \mathbb{R}^I_+ = \mathbb{C}^I/\mathbb{T}^I \text{ for } M : z_i \to z_i\overline{z}_i$$

where $\mathbb{T}^I$ is the n-torus naturally acting on $\mathbb{C}^I$ and where the the restriction of M to the unit sphere in $\mathbb{C}^I \to \mathbb{R}^I_+$ factors through the complex projective space $\mathbb{C}P(I)$ of complex dimension $|I| - 1$ that sends $\mathbb{C}P(I) \to \triangle(I)$.

This tells you what you could have been feeling all along: the cone $\mathbb{R}^I_+$ is ugly, it breaks the Euclidean/orthogonal symmetry of $\mathbb{R}^I$ – the symmetry is invisible (?) in the category $\mathcal{P}$ unless we write down and differentiate Boltzmann's formula.

Now we have the orthogonal symmetry, even better the unitary symmetry of $\mathbb{C}^I$, and may feel proud upon discovering the new world where entropy "truly" lives. Well..., it is not new, physicists came here ahead of us and named this world "quantum". Yet, even if disappointed, we feel warm toward Nature who shares with us the idea of mathematical beauty.

We are not tied up to a particular orthogonal basis for defining entropy anymore, we forget the coordinate space $\mathbb{C}^I$ that we regard as a Hilbert space S, where one basis of orthonormal vectors $\{s\} \subset S$ is as good as another.

An "atomic measure", or *a pure state* P in S is a (complex) line in S with a positive real number $|p|$ attached to it. In order to be able to add such measures, we regard P it as positive definite Hermitian form of rank one that vanishes on the orthogonal complement to our line, and such that P equals $|p|$ on the unit vectors in this line.

Accordingly, (non-atomic) *states* P *on* S are defined as convex combinations of pure ones. In other words, a quantum state P on a Hilbert space S is a non-zero semipositive Hermitian form on S (that customary is represented by a semipositive self adjoint operator $S \to S$) that we regard as a real valued quadratic function on S that is invariant under multiplication by $\sqrt{-1}$. (In fact, one could forget the $\mathbb{C}$-structure in S and admit all non-negative quadratic function $P(s)$ as states on S.)

We may think of a state P as a "measure" on subspaces $T \subset S$, where the "P-mass" of T, denoted $P(T)$, is the sum $\sum_t P(t)$, where the summation is taken over an orthonormal basis $\{t\}$ in T. (This does not depend on the basis by the Pythagorean theorem.) The total mass of P is denoted $|P| = P(S)$; if $|P| = 1$ then P is called a *density* (instead of probability) *state*.

Observe that

$$P(T_1 \oplus T_2) = P(T_1) + P(T_2) \text{ for } \textit{orthogonal} \text{ subspaces } T_1 \text{ and } T_2 \text{ in } S$$

and that the *tensor product* of states P_1 on S_1 and P_2 on S_2, that is a state on $S_1 \otimes S_2$, denoted $P = P_1 \otimes P_2$, satisfies

$$P(T_1 \otimes T_2) = P_1(T_1) \cdot P_2(T_2) \text{ for all } T_1 \subset S_1 \text{ and } T_2 \subset S_2.$$

If $\Sigma = \{s_i\}_{i \in I} \subset S$, $|I| = dim(S)$ is an orthonormal basis in S then the set $\underline{P}(\Sigma) = \{P(s_i)\}$ is a finite measure space of mass $|\underline{P}(\Sigma)| = |P|$. Thus, P defines a map from the space $Fr_I(S)$ of full orthonormal I-frames Σ in S (that is a principal homogeneous space of the unitary group $U(S)$) to the Euclidean $(|I|-1)$-simplex of measures of mass $|P|$ on the set I, that is $\{p_i\} \subset \mathbb{R}_+^I, \sum_i p_i = |P|$.

Classical Example. A finite measure space $\underline{P} = \{\underline{p}\}$ defines a quantum state on the Hilbert space $S = \mathbb{C}^{set(P)}$ that is the diagonal form $P = \sum_{\underline{p} \in \underline{P}} |\underline{p}| z_{\underline{p}} \overline{z}_{\underline{p}}$.

Notice, that we excluded spaces with zero atoms from the category $\mathcal{P}$ in the definition of classical measure spaces with no(?) effect on the essential properties of $\mathcal{P}$. But one needs to keep track of these "zeros" in the quantum case. For example, there is a unique, up to a scale homogeneous state, on S that is the Hilbert form of S, but the states that are homogeneous on their *supports* (normal to $0(S)$) constitute a respectable space of all linear subspaces in S.

Von Neumann Entropy. There are several equivalent definitions of $ent(P)$ that we shall be using interchangingly.

(1) The "minimalistic" definition is given by extracting a single *number* from the classical entropy *function* on the space of full orthonomal frames in S, that is $\Sigma \mapsto ent(\underline{P}(\Sigma))$, by taking the infimum of this functions over $\Sigma \in Fr_I(S)$, $|I| = dim(S)$,

$$ent(P) = \inf_{\Sigma} ent(\underline{P}(\Sigma)).$$

(The supremum of $ent(\underline{P}(\Sigma))$ equals $\log dim(S)$. In fact, there always exists a full orthonomal frame $\{s_i\}$, such that $P(s_i) = P(s_j)$ for all $i, j \in I$ by Kakutani-Yamabe-Yujobo's theorem that is applicable to all continuous function on spheres. Also, the average of $ent(\underline{P}(\Sigma))$ over Fr_I is close to $\log dim(S)$ for large $|I|$ by an easy argument.)

It is immediate with this definition that

the function $P \mapsto ent(P)$ is concave on the space of density states:

$$ent\left(\frac{P_1 + P_2}{2}\right) \geq \frac{ent(P_1) + ent(P_2)}{2}.$$

Indeed, the classical entropy is a concave function on the simplex of probability measures on the set I, that is $\{p_i\} \subset \mathbb{R}_+^I, \sum_i p_i = 1$, and infima of familes of concave functions are concave.

(2) The traditional "spectral definition" says that the von Neumann entropy of P equals the classical entropy of the *spectral measure* of P. That is $ent(P)$ equals $\underline{P}(\Sigma)$ for a frame $\Sigma = \{s_i\}$ that *diagonalizes* the Hermitian form P, i.e. where s_i is P-orthogonal to s_j for all $i \neq j$.

Equivalently, "spectral entropy" can be defined as the (obviously unique) *unitary invariant extension* of Boltzmann's entropy from the subspace of classical states to the space of quantum states, where "unitary invariant" means that $ent(g(P)) = ent(P)$ for all unitary transformations g of S.

If concavity of entropy is non-obvious with this definition, it is clear that *the spectrally defined entropy is additive under tensor products of states:*

$$ent(\otimes_k P_k) = \prod_k ent(P_k),$$

and if $\sum_k |P_k| = 1$*, then the direct sum of* P_k *satisfies*

$$ent(\oplus_k P_k) = \sum_{1\le k\le n} |P_k| ent(P_k) + \sum_{1\le k\le n} |P_k| \log |P_k|,$$

This follows from the corresponding properties of the classical entropy, since tensor products of states correspond to Cartesian products of measure spaces:

$$\underline{(P_1 \otimes P_2)}(\Sigma_1 \otimes \Sigma_2) = \underline{P}_1(\Sigma_1) \times \underline{P}_2(\Sigma_2)$$

and the direct sums correspond to disjoint unions of sets.

(3) Let is give yet another definition that will bring together the above two.

Denote by $\mathcal{T}_\varepsilon = \mathcal{T}_\varepsilon(S)$ the set of the linear subspaces $T \subset S$ such that $P(T) \ge (1-\varepsilon)P(S))$ and define

$$ent_\varepsilon(P) = \inf_{T\in\mathcal{T}_\varepsilon} \log dim(T).$$

By *Weyl's variational principle*, the supremum of $P(T)$ over all n-dimensional subspaces $T \subset S$ is achieved on a subspace, say $S_+(n) \subset S$ spanned by n mutually orthogonal *spectral vectors* $s_j \in S$, that are vectors from a basis $\Sigma = \{s_i\}$ that diagonalizes P. Namely, one takes s_j for $j \in J \subset I$, $|J| = n$, such that $P(s_j) \ge P(s_k)$ for all $j \in J$ and $k \in I \smallsetminus J$.

(To see this, orthogonally split $S = S_+(n) \oplus S_-(n)$ and observe that the P-mass of every subspace $T \subset S$ *increases* under the transformations $(s_+, s_-) \to (\lambda s_+, s_-)$ that eventually, for $\lambda \to +\infty$, bring T to the span of spectral vectors.)

Thus, this ent_ε equals its classical counterpart for the spectral measure of P.

To arrive at the actual entropy, we evaluate ent_ε on the tensorial powers $P^{\otimes N}$ on $S^{\otimes N}$ of states S and, by applying the law of large numbers to the corresponding Cartesian powers of the spectral measure space of P, conclude that

the limit

$$ent(P) = \lim_{N\to\infty} \frac{1}{N} ent_\varepsilon(P^{\otimes N})$$

exists and it equals the spectral entropy of P *for all* $0 < \varepsilon < 1$. (One may send $\varepsilon \to 0$ if one wishes.)

It also follows from Weyl's variational principle that the ent_ε-definition agrees with the "minimalistic" one. (It takes a little extra effort to check that $ent(\underline{P}(\Sigma))$ is *strictly* smaller than $\lim \frac{1}{N} ent_\varepsilon(P^{\otimes N})$ for *all non-spectral* frames Σ in S but we shall not need this.)

Unitary Symmetrization and Reduction. Let $d\mu$ be a Borel probability measure on the group $U(S)$ of the unitary transformation of S, e.g. the normalized Haar measure dg on a compact subgroup $G \subset U(S)$.

The *μ-average of* P of a state P on S, that is called the *G-average* for $d\mu = dg$ is defined by

$$\mu * P = \int_G (g * P) d\mu \text{ for } (g * P)(s) =_{def} P(g(s)).$$

Notice that $ent(\mu * P) \geq ent(P)$ by concavity of entropy and that the G-average of P, denoted $G * P$, equals the (obviously unique) G-invariant state on S such that $G * P(T) = P(T)$ for all G-invariant subspaces $T \subset S$. Also observe that the μ-averaging operator commutes with tensor products:

$$(\mu_1 \times \mu_2) * (P_1 \otimes P_2) = (\mu_1 * (P_1)) \otimes (\mu_2 * (P_2)).$$

If $S = S_1 \otimes S_2$, and the group $G = G_1$ equals $U(S_1)$ that naturally acts on S_1 (or any G *irreducibly* acting on S_1 for this matter), then there is a one-to-one correspondence between G_1-invariant states on S and states on S_2. The state P_2 on S_2 that corresponds to $G_1 * P$ on S is called the *canonical reduction of P to S_2* . Equivalently, one can define P_2 by the condition $P_2(T_2) = P(S_1 \otimes T_2)$ for all $T_2 \subset S_2$.

(Customary, one regards states as selfadjoint operators O on S defined by $\langle O(s_1), s_2 \rangle = P(s_1, s_2)$). The reduction of an O on $S_1 \otimes S_2$, to an operator, say, on S_2 is defined as the S_1*-trace* of O that does not use the Hilbertian structure in S.)

Notice that $|P_2| = P_2(S_2) = |P| = P(S)$, that

$$(*) \qquad ent(P_2) = ent(G * P) - \log dim(S_1)$$

and that the canonical reduction of the tensorial power $P^{\otimes N}$ to $S_2^{\otimes N}$ equals $P_2^{\otimes N}$.

Classical Remark. If we admit zero atoms to finite measure spaces, then a classical reduction can be represented by the push-forward of a measure $\underline{P}$ from a Cartesian product of sets, $\underline{S} = \underline{S}_1 \times \underline{S}_2$ to $\underline{P}_2$ on $\underline{S}_2$ under the coordinate projection $\underline{S} \to S_2$. Thus, canonical reductions generalize classical reductions. ("Reduction by G-symmetrization" with non-compact, say amenable G, may be of interest also for Γ-dynamical spaces/systems, for instance, such as P^Γ in the classical case and $P^{\otimes\Gamma}$ in the quantum setting.)

A novel feature of "quantum" is a possible increase of entropy under reductions (that is similar to what happens to sofic entropies of classical Γ-systems for non-amenable groups Γ).

For example if P is a *pure* state on $S \otimes T$ (entropy=0) that is supported on (the line generated by) the vector $\sum_i s_i \otimes t_i$ for an orthonormal bases in S and in T (here $dim(S) = dim(T)$), then, obviously, the canonical reduction of P to T is a homogenous state with entropy= $\log dim(T)$. (In fact, every state of P on a Hilbert space T equals the canonical reduction of a pure state on $T \otimes S$ whenever $dim(S) \geq dim(T)$, because every Hermitian form on T can be represented as a vector in the tensor product of T with its Hermitian dual.)

Thus a singe classically indivisible "atom" represented by a pure state on $S \otimes T$ may appear to the observer looking at it through the kaleidoscope of quantum windows in T as several (equiprobable in the above case) particles.

On the other hand, the Shannon inequality remains valid in the quantum case, where it is usually formulated as follows.

Subadditivity of von Neimann's Entropy (Lanford-Robinson 1968). The entropies of the canonical reductions P_1 and P_2 of a state P on $S = S_1 \otimes S_2$ to S_1 and to S_2 satisfy

$$ent(P_1) + ent(P_2) \geq ent(P).$$

Proof. Let Σ_1 and Σ_2 be orthonormal bases in S_1 and S_2 and let $\Sigma = \Sigma_1 \times \Sigma_2$ be the corresponding basis in $S = S_1 \times S_2$. Then the measure spaces $P_1(\Sigma_1)$ and $P_2(\Sigma_2)$ equal classical reductions of $P(\Sigma)$ for the Cartesian projections of Σ to Σ_1 and to Σ_2. Therefore,

$$ent(P(\Sigma_1 \times \Sigma_2)) \leq ent(P(\Sigma_1)) + ent(P(\Sigma_1))$$

by Shannon inequality, while

$$ent(P) \leq ent(P(\Sigma_1 \times \Sigma_2))$$

according to our minimalistic definition of von-Neimann entropy,

Alternatively, one can derive subadditivity with the ent_ε-definition by observing that

$$ent_{\varepsilon_1}(P_1) + ent_{\varepsilon_2}(P_2) \geq ent_{\varepsilon_{12}}(P) \text{ for } \varepsilon_{12} = \varepsilon_1 + \varepsilon_2 + \varepsilon_1\varepsilon_2$$

and applying this to $P^{\otimes N}$ for $N \to \infty$, say with $\varepsilon_1 = \varepsilon_2 = 1/3$.

Concavity of Entropy Versus Subadditivity. There is a simple link between the two properties.

To see this, let P_1 and P_2 be density states on S and let $Q = \frac{1}{2}P_1 \oplus \frac{1}{2}P_2$ be their direct sum on $S \oplus S = S \otimes \mathbb{C}^2$. Clearly, $ent(Q) = ent(P) + \log 2$

On the other hand, the canonical reduction of Q to S equals $\frac{1}{2}(P_1 + P_2)$, while the reduction of Q to $\mathbb{C}^2 = \mathbb{C} \oplus \mathbb{C}$ is $\frac{1}{2} \oplus \frac{1}{2}$.

Thus, concavity follows from subadditivity and the converse implication is straightforward.

Here is another rendition of subadditivity.

Let compact groups G_1 and G_2 unitarly act on S such that the two actions commute and the action of $G_1 \times G_2$ on S is irreducible, then

$$(\star) \qquad ent(P) + ent((G_1 \times G_2) * P) \leq ent(G_1 * P) + ent(G_2 * P)$$

for all states P on S.

This is seen by equivariantly decomposing S into the direct sum of, say n, tensor products:

$$S = \bigoplus_k (S_{1k} \otimes S_{2k}), \ k = 1, 2, ...n,$$

for some unitary actions of G_1 on all S_{1k} and of G_2 on S_{2k} and by observing that $(\star)$ is equivalent to subbaditivity for the reductions of P on these tensor products.

Strong Subadditivity and Bernoulli States. The inequality $(\star)$ generalizes as follows.

Let H and G be compact groups of unitary transformations of a finite dimensional Hilbert space S and let P be a state (positive semidefinite Hermitian form) *on S. If the actions of H and G commute, then the von Neumann entropies of the G- and H-averages of P satisfy*

$$(\star\star)\qquad ent(G*(H*P))-ent(G*P)\leq ent(H*P)-ent(P).$$

Acknowledgement. This was stated in the earlier version of the paper for *non-commuting* actions with an indication of an argument justifying it. But Michael Walter pointed out to me that if P is G-invariant, then, in fact, one has the opposite inequality:

$$ent(G*(H*P))-ent(G*P)\geq ent(H*P)-ent(P).$$

Also he formulated the following (correct) version of $(\star\star)$ for non-commuting actions (that follows by the argument similar to that for the derivation of concavity of entropy from subadditivity):

$$ent(G*(H*P))-\int_H ent(G*(h*P)dh\leq ent(H*P)-ent(P).$$

The inequality $(\star\star)$, applied to the actions of the unitary groups $H=U(S_1)$ and $G=U(S_2)$ on $S=S_1\otimes S_2\otimes S_3$, is equivalent, by the above $(\star)$, to the following

Strong Subadditivity of von Neumann Entropy (Lieb-Ruskai, 1973). Let $P=P_{123}$ be a state on $S=S_1\otimes S_2\otimes S_3$ and let P_{23}, P_{13} and P_3 be the canonical reductions of P_{123} to $S_2\otimes S_3$, to $S_1\otimes S_3$ and to S_3.

Then

$$ent(P_3)+ent(P_{123})\leq ent(P_{23})+ent(P_{13}).$$

Notice, that the action of $U(S_1)\times U(S_2)$ on S is a *multiple* of an irreducible representation, namely it equals N_3-multiple, $N_3=dim(S_3)$, of the action of $U(S_1)\times U(S_2)$ on $S_1\otimes S_2$. This is why one needs $(\star\star)$ rather than $(\star)$ for the proof.

The relative Shannon inequality (that is not fully trivial) for *measures* reduces by Bernoulli-Gibbs' argument to a *trivial* intersection property of *subsets in a finite set.* Let us do the same for the von Neumann entropy.

The *support* of a state P on S is the orthogonal complement to the *null-space* $0(P)\subset S$ – the subspace where the (positive semidefinite) Hermitian form P vanishes. We denote this support by $0^\perp(P)$ and write $rank(P)$ for $dim(0^\perp(P))$.

Observe that

$$(\Leftrightarrow)\qquad P(T)=|P|\Leftrightarrow T\supset 0^\perp(P)$$

for all linear subspaces $T\subset S$.

A state P is *sub-homogeneous*, if $P(s)$ is constant, say equal $\lambda(P)$, on the unit vectors from the support $0^\perp(P)\in S$ of P. (These states correspond to subsets in the classical case.)

If, besides being sub-homogeneous, P is a *density* state, i.e. $|P|=1$, then, obviously, $ent(P)=-\log\lambda(P)=\log\dim(0^\perp(P))$.

Also observe that if P_1 and P_2 are sub-homogeneous states such that $0^\perp(P_1) \subset 0^\perp(P_2)$, then

$$(/\geq/) \qquad\qquad P_1(s)/P_2(s) \leq \lambda(P_1)/\lambda(P_2)$$

for all $s \in S$ (with the obvious convention for 0/0 applied to $s \in 0(P_2)$).

if a sub-homogeneous state Q equals the G-average of some (not necessarily sub-homogeneous)*state P, then* $0^\perp(Q) \supset 0^\perp(P)$*).*

Indeed, by the definition of the average, $Q(T) = P(T)$ for all G-invariant subspaces $T \subset S$. Since $Q(0^\perp(Q)) = Q(S) = P(S) = P(0^\perp(Q))$ and the above $(\Leftrightarrow)$ applies.

Trivial Corollary. The inequality $(\star\star)$ holds in the case where all four states: P, $P_1 = H * P$, $P_2 = G * P$ and $P_{12} = G * (H * P)$, are sub-homogeneous.

Trivial Proof. The inequality $(\star\star)$ translates in the sub-homogeneous case to the corresponding inequality between the values of the states on their respective supports:

$$\lambda_2/\lambda_{12} \leq \lambda/\lambda_1,$$

for $\lambda = \lambda(P)$, $\lambda_1 = \lambda(P_1)$, etc. and proving the sub-homogeneous $(\star\star)$ amounts to showing that the implication

$$(\leq\Rightarrow\leq) \qquad\qquad \lambda \leq c\lambda_1 \Rightarrow \lambda_2 \leq c\lambda_{12}$$

holds for all $c \geq 0$.

Since $0^\perp(P) \subset 0^\perp(P_1)$, the inequality $\lambda \leq c\lambda_1$ implies, by the above $(/\geq/)$, that $P(s) \leq cP_1(s)$ for *all* s, where this integrates over G to $P_2(s) \leq cP_{12}(s)$ for *all* $s \in S$.

Since $0^\perp(P_2) \subset 0^\perp(P_{12})$, there exists *at least one* non-zero vector $s_0 \in 0^\perp(P_2) \cap 0^\perp(P_{12})$ and the proof follows, because $P_2(s_0)/P_{12}(s_0) = \lambda_2/\lambda_{12}$ for such an s_0.

"Nonstandard" Proof of $(\star\star)$ *in the General Case.* Since tensorial powers $P^{\otimes N}$ of all states P "converge" to "ideal sub-homogeneous states" $P^{\otimes\infty}$ by Bernoulli's theorem, the "trivial proof", applied to these ideal $P^{\otimes\infty}$, yields $(\star\star)$ for all P.

If "ideal sub-homogeneous states" are understood as objects of a *non-standard model* of the first oder $\mathbb{R}$-language of the category of finite dimensional Hilbert spaces, then the trivial proof applies in the case where the action of G and of H commute, where the role of "commute" is explained later on.

In truth, one does not need for the proof the full fledged "non-standard" language – everything can be expressed in terms of infinite families of ordinary states; yet, this needs a bit of additional terminology that we introduce below.

From now on, our states are defined on finite dimensional Hilbert spaces S_N, that make a countable family, denoted $S_* = \{S_N\}$, where where N are members of a countable set $\mathcal{N}$, e.g. $\mathcal{N} = \mathbb{N}$ with some non-principal ultra filter on it. This essentially means that what we say about S_* must hold for infinitely many N.

Real numbers are replaced by families/sequences of numbers, say $a_* = \{a_N\}$, where we may assume, using our ultrafilter, that the limit a_N, $N \to \infty$, always exists (possibly equal $\pm\infty$). This means, in simple terms, that we are allowed to pass to convergent subsequences as often as we wish to. We write $a_* \sim b_*$ if the corresponding sequences have equal limits.

If P_* and Q_* are states on S_*, we write $P_* \sim Q_*$ if $P_*(T_*) \sim Q_*(T_*)$ for all linear subspaces $T_* \subset S_*$. This signifies that $\lim P_N(T_N) = \lim Q_N(T_N)$ for all $T_N \subset S_N$ and some subsequence of $\{N\}$.

Let us formulate and prove the counterpart of the above implication $P(T) = |P| \Rightarrow T \supset 0^\perp(P)$ for *sub-homogeneous density states* P_*.

Notice that $P_*(T) \sim |P_*|$ *does not* imply that $T_* \supset 0^\perp(P_*)$; yet, it *does imply* that

- *there exists a state* $P'_* \sim P_*$, *such that* $T_* \supset 0^\perp(P'_*)$.

Proof. let U_* be the support of P_* and let $\Pi_* : U_* \to T_*$ be the normal projection. Then the *sub-homogeneous density* state Π'_* with the support $\Pi_*(U_*) \subset T_*$ (there is only one such state) is the required one by a trivial argument.

To complete the translation of the "nonstandard" proof of $(\star\star)$ we need a few more definitions.

Multiplicative Homogeneity. Let $Ent_* = \{Ent_N\} = \log dim(S_N)$ and let us normalize positive (multiplicative) constants (scalars) $c = c_* = \{c_N\} \geq 0$ as follows,

$$|c|_\star = |c_*|^{\frac{1}{Ent_*}}.$$

In what follows, especially if "$\star$" is there, we may omit "$*$".

A state $B = B_* = \{B_N\}$ is called $\star$*-homogenous*, if $|B(s_1)|_\star \sim |B(s_2)|_\star$ for all spectral vectors $s_1, s_2 \in 0^\perp(B) \subset S_*$, or, equivalently, if the (unique) sub-homogenous, state B' for which $0^\perp(B') = 0^\perp(B)$ and $|B'| = |B|$ satisfies $|B'(s)|_\star \sim |B(s)|_\star$ for all unit vectors $s \in 0^\perp(B)$.

Since the number $|B'(s)|$, $s \in 0^\perp(B')$ is independent of $s \in 0^\perp(A')$, we may denote it by $|B|_\star$.

Let B be a $\star$-homogeneous density state with support $T = 0^\perp(B)$ and A a sub-homogenous density state with support $U = 0^\perp(A)$.

If $A(T) \sim B(T) = 1$ *Then there exist a linear subspace* $U' \subset U$ *such that*

$$|dim(U')/dim(U)| \sim 1$$

and

$$|B(s)|_\star \sim |B|_\star \text{ for all unit vectors } s \in U'.$$

.

Proof. Let $\Pi_T : U \to T$ and $\Pi_U : T \to U$ be the normal projections and let u_i be the eigenvectors of the (self-adjoint) operator $\Pi_U \circ \Pi_T : U \to U$ ordered by their eigenvalues $\lambda_1 \leq \lambda_2 ..., \lambda_i, ...$. By Pythagorean theorem, $dim(U)^{-1} \sum_i \lambda_i = 1 - B(T)$; therefore the span U_ε of those u_i where $\lambda_i \geq 1 - \varepsilon$ satisfies $|dim(U_\varepsilon)/dim(U)| \sim 1$ for all $\varepsilon > 0$; any such U_ε can be taken for U'.

•• *Corollary.* Let $\mathcal{B}$ be be a finite set of $*$-homogeneous density states B on S_*, such that $A(0^\perp(B)) \sim 1$ for all $B \in \mathcal{B}$. Then there exists a unit vector $u \in U = 0^\perp(A)$, such that $|B(u)|_\star \sim |B|_\star$ for all $B \in \mathcal{B}$.

This is shown by the obvious induction on cardinality of $\mathcal{B}$ with U' replacing U at each step.

Let us normalize entropy of $A_* = \{A_N\}$ by setting

$$ent_\star(A_*) = ent(A_*)/Ent_* = \left\{ \frac{ent(A_N)}{\log dim(S_N)} \right\}$$

and let us call a vector $s \in S_*$ *Bernoulli* for a density state A_* on S_*, if $\log|A(s)|_\star \sim -ent_\star(A)$.

A density state A on S_* is called *Bernoulli* if there is a subspace U, called a *Bernoulli core of* A, spanned by some spectral Bernoulli vectors of A, such that $A(U) \sim 1$.

For example, all s in the support of a $\star$-homogeneous density state A are Bernoulli.

More significantly, the families of tensorial powers, $A_* = \{P^{\otimes N}\}$ on $S_* = \{S^{\otimes N}\}$, are Bernoulli for all density states P on S by Bernoulli's law of large numbers.

Multiplicative Equivalence and Bernoulli Equivalence. Besides the relation $A \sim B$ it is convenient to have its multiplicative counterpart, denoted $A \overset{\star}{\sim} B$, which signifies $|A(s)|_\star \sim |B(s)|_\star$ for all $s \in S_*$.

Bernoulli equivalence relation, on the set of density states on S_* is defined as the span of $A \sim B$ and $A \overset{\star}{\sim} B$. For example, if $A \sim B$, $B \overset{\star}{\sim} C$ and $C \sim D$, then A is Bernoulli equivalent to D.

Observe that

Bernoulli equivalence is stable under convex combinations of states.

In particular, if $A \overset{\star}{\sim} B$, then $G * A \overset{\star}{\sim} G * B$, for all compact groups G of unitary transformations of S_* (i.e. for all sequences G_N acting on S_N.)

This Bernoulli equivalence is similar to that for (sequences of) classical finite measure spaces and the following two properties of this equivalence trivially follow from the classical case via Weyl variational principle. (We explain this below in "non-standard" terms.)

(1) *If A is Bernoulli and B is Bernoulli equivalent to A then B is also Bernoulli. Thus, A is Bernoulli if and only if it is Bernoulli equivalent to a sub-homogeneous state on S_*.*

(2) *If A is Bernoulli equivalent to B then $ent_\star(A) \sim ent_\star(B)$.*

We write $a_* \gtrsim b_*$ for $a_N, b_N \in \mathbb{R}$, if $a_* - b_* \sim c_* \geq 0$.

If B is a Bernoulli state on S_* and A is a density state, write $A \prec B$ if B admits a Bernoulli core T, such that $A(T) \sim 1$.

This relation is invariant under equivalence $A \sim A'$, but *not* for $B \sim B'$. Neither is this relation transitive for Bernoulli states.

Main Example. If B equals the G-average of A for some compact unitary transformation group of S_*, then $A \prec B$.

Indeed, by the definition of average, $B(T) = A(T)$ for all G-invariant subspaces T. On the other hand, if a G-invariant B is Bernoulli, then it admits *a G-invariant* core, since the set of spectral Bernoulli vectors is G-invariant and all unit vectors in the span of spectral Bernoulli vectors are Bernoulli.

Main Lemma. Let A, B, C, D be Bernoulli states on S_*, such that $A \prec B$ and $A \prec D$ and let G be a compact unitary transformation group of S_*.

If $C \sim G * A$ and $D = G * B$ and if A is sub-homogeneous, then

$$ent_\star(B) - ent_\star(A) \gtrsim ent_\star(C) - ent_\star(D).$$

Proof. According to $\bullet$, there is a state $A' \sim A$, such that it support $0^\perp(A')$ is contained in some Bernoulli core of B, and since our assumptions and the conclusion are invariant under equivalence $A \sim A'$. we may assume that $U = 0^{\perp}(A)$ itself is contained in a Bernoulli core of B.

Thus,

$$A(s) \le c^{Ent_*} B(u) \text{ for all } c > \exp(ent(B) - ent(A)) \text{ and all } s \in S_*$$

Also, we may assume that $C = G * A$ since averaging and $ent_\star$ are invariant under the $\sim$-equivalence.

Then $C = G * A$ and $D = G * B$ also satisfy

$$C(s) \le c^{Ent_*} D(s) \text{ for all } s \in S_*.$$

In particular,

$$C(u) \le c^{Ent_*} D(u) \text{ for a common Bernoulli vector, } u \text{ of } C \text{ and } D$$

where the existence of such a $u \in U$ is ensured by $\bullet\bullet$.

Thus, $|C(u)|_\star \le c|D(u)|_\star$ for all $c > \exp(ent_\star(B) - ent_\star(A))$. Since C and D are Bernoulli, $ent_\star(C) \sim -\log|C(u)|_\star$ and $ent_\star(D) \sim -\log|D(u)|_\star$; hence

$$ent_*(D) - ent_\star(C) \le c \text{ for all } c \le ent_\star(B) - ent_\star(A)$$

that means $ent_\star(B) - ent_\star(A) \gtrsim ent_\star(C) - ent_\star(D)$. QED.

Proof of $(\star\star)$. Let P be a density state on a Hilbert space S, let G and H be unitary group acting on S, and let us show that

$$ent(G * (H * P)) - ent(G * P) \le ent(H * P) - ent(P)$$

assuming that G and H *commute*.

In fact, all we need is that the state $G * (H * P)$ equals the K-average of P for some group K, where $K = G \times H$ serves this purpose in the commuting case.

Recall that the family $\{P^{\otimes N}\}$ on $S_* = \{S_N = S^{\otimes N}\}$ is Bernoullian for all P on S, and the averages, being tensorial powers themselves, are also Bernoullian.

Let $A_* = \{A_N\}$ be the subhomogeneous state S_* that is Bernoulli equivalent to $P^{\otimes N}$, where, by the above, their averages remains Bernoullian. (Alternatively, one could take $A_N^{\otimes M}$, say, for $M = 2^N$.)

Since both states B and D are averages of A in the commuting case, $A \prec B$ and $A \prec D$; thus the lemma applies and the proof follows.

On the above (1) and (2). A density state P on S is fully characterized, up to unitary equivalence, by its *spectral distribution function* $\Psi_P(t) \in [0,1]$, $t \in [0, dim(S)]$, that equals the maximum of $P(T)$ over linear subspaces $T \subset S$ of dimension n for integer n, and that is linearly interpolated to $t \in [n, n+1]$.

By Weyl's variational principal this Ψ equals its classical counterpart, where the maximum is taken over *spectral* subspaces T.

The ε-entropy and Bernoullian property, are easily readable from this function and so the properties (1) and (2) follow from their obvious classical counterparts, that we have used, albeit implicitly, in the definition of the classical Bernoulli-Boltzmann's entropy.

Nonstandard Euclidean/Hilbertian Geometry. Entropy constitute only a tiny part of asymptotic information encoded by Ψ_{A_N} in the limit for $N \to \infty$, where there is no problem with passing to limits since, obviously, Ψ are *concave* functions. However, most of this information is lost under "naive limits" and one has to use limits in the sense of nonstandard analysis.

Furthermore, individual Ψ do not tell you anything about mutual positions between different states on S_*: joint Hilbertian geometry of several states is determined by the complex valued functions, kind of (scattering) "matrices", say $\Upsilon_{ij} : \underline{P}_i \times \underline{P}_j \to \mathbb{C}$, where the "entries" of Υ_{ij} equal the scalar products between unit spectral vectors of P_i and of P_j. (There is a *phase ambiguity* in this definition that becomes significant if there are multiple eigenvalues.)

Since these Υ_{ij} are *unitary* "matrices" in an obvious sense, the corresponding $\Sigma_{ij} = |\Upsilon_{ij}|^2$ define *bistochastic correspondences* (customary represented by matrices) between respective spectral measure spaces.

(Unitarity imposes much stronger restrains on these matrices than mere bistochasticity. Only a minority of bistochastic matrices, that are called *unistochastic*, have "unitary origin". In physics, if I get it right, experimentally observable unistochasticity of scattering matrices can be taken for evidence of unitarity of "quantum universe".)

Moreover, the totality of "entries" of "matrices" Υ_{ij}, that is the full array of scalar products between all spectral vectors of all P_i, satisfy a stronger positive definiteness condition.

At the end of the day, everything is expressed by scalar products between unit spectral vectors of different P_i and the values of P_i on their spectral vectors; non-standards limits of arrays of these numbers fully describe the nonstandard geometry of finite sets of non-standard states on nonstandard Hilbert spaces.

Reformulation of Reduction. The entropy inequalities for canonical reductions can be more symmetrically expressed in terms of *entropies of bilinear forms* $\Phi(s_1, s_2)$, $s_i \in S_i$ i=1,2, where the entropy of a Φ is defined as the entropy of the Hermitian form P_1 on S_1 that is induced by the linear map $\Phi'_1 : S_1 \to S'_2$ from the Hilbert form on the linear dual S'_2 of S_2, where, observe, this entropy equal to that of the Hermitian form on S_2 induced by $\Phi'_2 : S_2 \to S'_1$.

In this language, for example, subadditivity translates to

Araki-Lieb Triangular Inequality (1970). The entropies of the three bilinear forms associated to a given 3-linear form $\Phi(s_1, s_2, s_3)$ satisfy

$$ent(\Phi(s_1, s_2 \otimes s_3)) \leq ent(\Phi(s_2, s_1 \otimes s_3)) + ent(\Phi(s_3, s_1 \otimes s_3)).$$

Discussion. Strong subadditivity was conjectured by Lanford and Robinson in 1968 and proved five years later by Lieb and Ruskai with *operator convexity techniques.*

Many proofs are based on an easy reduction of strong subadditivity to the *trace convexity* of the *operator function* $e(x, y) = x \log x - x \log y$. The shortest present day proof of this trace convexity is due to Ruskai [21] and the most transparant one to Effros [9].

On the other hand, this was pointed out to me by Mary Beth Ruskai (along with many other remarks two of which we indicate below), there are by now other proofs of SSA, e.g. in [12] and in [20], which do not use trace convexity of $x \log x - x \log y$.

1. In fact, one of the two original proofs of SSA did not use the trace convexity of $x \log x - x \log y$ either, but relied on the concavity of the map $x \mapsto$

$trace\left(e^{y+\log x}\right)$ as it is explained in [22] along with H. Epstein's elegant proof that $e^{y+\log x}$ *is* a trace concave function in x.

2. The possibility of deriving SSA from the trace concavity of $e^{y+\log x}$ was independently observed in 1973 by A. Uhlmann who also suggested a reformulation of SSA in terms of group averages.

Recently, Michael Walter explained to me that our "Bernoullian" proof is close to that in [20] and he also pointed out to me to the paper [8] where the authors establish asymptotics of *recoupling coefficients* for tensor products of representations of permutation groups. This refines the Bernoulli theorem and, in particular, directly implies the SSA inequality.

Sharp convexity inequalities are circumvented in our "soft" argument by exploiting the "equalizing effect" of Bernoulli theorem that reduces evaluation of sums (or integrals) to a point-wise estimate. Some other operator convexity inequalities can be also derived with Bernoulli approximation, but this method is limited (?) to the cases that are stable under tensorization and it seems poorly adjusted to identification of states where such inequalities become equalities.

(I could not find a simple "Bernoullian proof" of the trace convexity of the operator function $x\log x - x\log y$, where such a proof of convexity of the ordinary $x\log x - x\log y$ is as easy as for $x\log x$.)

There are more powerful "equalization techniques" that are used in proofs of "classical" geometric inequalities and that involve elliptic PDE, such as solution of *Monge-Kantorovich transportation problem* in the proof of *Bracamp-Lieb refinement of the Shannon-Loomis-Whitney-Shearer inequality* (see [3] and references therein) and invertibility of some *Hodge operators on toric Kähler manifolds* as in the analytic rendition of Khovanski-Teissier proof of the Alexandrov-Fenhcel inequality for mixed volumes of convex sets [10]. It is tempting to to find "quantum counterparts" to these proofs.

Also it is desirable to find more functorial and more informative proofs of "natural" inequalities in geometric (monoidal?) categories. (See [4],[23] for how it goes along different lines.)

On Algebraic Inequalities. Besides "unitarization" some Shannon inequalities admit linearization, where the first non-trivial instance of this is the following *linearized Loomis-Whitney* $3D$*-isoperimetric inequality* for ranks of bilinear forms associated with a 4-linear form $\Phi = \Phi(s_1, s_2, s_3, s_4)$ where we denote $|...| = rank(...)$:

$$|\Phi(s_1, s_2\otimes s_3\otimes s_4)|^2 \le |\Phi(s_1\otimes s_2, s_3\otimes s_4)|\cdot|\Phi(s_1\otimes s_3, s_2\otimes s_4)|\cdot|\Phi(s_1\otimes s_4, s_2\otimes s_3)|$$

This easily reduces (see [11]) to the original Loomis-Whitney inequality and also can proven directly with Bernoulli tensorisation.

But the counterpart to the strong subadditivity – the relative Shannon inequality:

$$|\Phi(s_1, s_2\otimes s_3\otimes s_4)|\cdot|\Phi(s_4, s_1\otimes s_2\otimes s_3)| \le |\Phi(s_1\otimes s_2, s_3\otimes s_4)|\cdot|\Phi(s_1\otimes s_3, s_2\otimes s_4)|$$

(that is valid with $\exp ent(...)$ instead of $|...|$) fails to be true for general Φ. (The obvious counterexamples can be taken care of with suitable Bernoulli-like-core stabilized ranks, but this, probably, does not work in general.)

Such "rank inequalities" are reminiscent of inequalities for spaces of sections and (cohomologies in general) of positive vector bundles such e.g. as in the

Khovanski-Teissier theorem and in the Esnault-Viehweg proof of *the sharpened Dyson-Roth lemma*, but a direct link is yet to be found.

Apology to the Reader. Originally, Part 1 of "Structures" was planned as about a half of an introduction to the main body of the text of my talk at the European Congress of Mathematics in Kraków with the sole purpose to motivate what would follow on "mathematics in biology". But it took me several months, instead of expected few days, to express apparently well understood simple things in an appropriately simple manner.

Yet, I hope that I managed to convey the message: the mathematical language developed by the end of the 20th century by far exceeds in its expressive power anything, even imaginable, say, before 1960. Any *meaningful* idea coming from science can be fully developed in this language. Well..., actually, I planned to give examples where a new language was needed, and to suggest some possibilities. It would take me, I naively believed, a couple of months but the experience with writing this "introduction" suggested a time coefficient of order 30. I decided to postpone.

3 Bibliography.

References

[1] J. Baez, T. Fritz, and T. Leinster, A Characterization of Entropy in Terms of Information Loss, Entropy, 13, pp. 1945-1957 (2011).

[2] K. Ball, Factors of i.i.d. processes with nonamenable group actions, www.ima.umn.edu/ kball/factor.pdf, 2003.

[3] F. Barthe, On a reverse form of the Brascamp-Lieb inequality, Invent. Math. 134 (1998), no. 2, 335-361.

[4] Philippe Biane , Luc Bouten , Fabio Cipriani, Quantum Potential Theory, Springer 2009.

[5] Lewis Bowen, A new measure conjugacy invariant for actions of free groups. Ann. of Math., vol. 171, No. 2, 1387-1400, (2010).

[6] Lewis Bowen, Sofic entropy and amenable groups, to appear in Ergodic Theory and Dynam. Systems.

[7] Lewis Bowen, Weak isomorphisms between Bernoulli shifts, Israel J. of Math, Volume 183, Number 1, 93-102 (2011).

[8] Matthias Cristandl, Mehmet Burak Sahinoglu, Michael Walter, Recoupling Coefficients and Quantum Entropies arXiv:1210.0463 (2012).

[9] Edward G. Effros, A matrix convexity approach to some celebrated quantum inequalities. PNAS vol. 106 no. 4, PP 1006-1008, (2009).

[10] M. Gromov, Convex sets and Kähler manifolds. Advances in Differential Geometry and Topology, ed. F. Tricerri, World Scientific, Singapore, pages 1-38, 1990.

[11] M Gromov, Entropy and Isoperimetry for Linear and non-Linear Group Actions. Groups Geom. Dyn. 2, No. 4, 499-593 (2008).

[12] Michal Horodecki, Jonathan Oppenheim, Andreas Winter, Quantum state merging and negative information, CMP 269, 107 (2007).

[13] A Katok, Fifty years of entropy in dynamics: 1958-2007. J. of Modern Dyn. 1 (2007), 545-596.

[14] Oscar E. Lanford, Entropy and equilibrium states in classical statistical mechanics, Lecture Notes in Physics, Volume 20, pp. 1-113, 1973.

[15] P.A. Loeb, Measure Spaces in Nonstandard Models Underlying Standard Stochastic Processes. Proc. Intern. Congr. Math. Warsaw, p.p. 323-335, 1983

[16] M. Marcolli, R. Thorngren, Thermodynamic Semirings, arXiv.org ¿ math ¿ arXiv:1108.2874.

[17] Yann Ollivier, A January 2005 invitation to random groups. Ensaios Matemticos [Mathematical Surveys], 10. Sociedade Brasileira de Matemtica, Rio de Janeiro, 2005.

[18] A. Ostebee, P. Gambardella and M. Dresden, Nonstandard approach to the thermodynamic limit. I, Phys. Rev. A 13, p.p. 878-881, 1976.

[19] Vladimir G. Pestov, Hyperlinear and sofic groups: a brief guide. Bull. Symbolic Logic Volume 14, Issue 4, 449-480 Bull.2008.

[20] R. Renner, Security of Quantum Key Distribution. arXiv:quant-ph/0512258, (2005).

[21] Mary Beth Ruskai, Another Short and Elementary Proof of Strong Subadditivity of Quantum Entropy, arXiv:quant-ph/0604206v1 27 Apr 2006.

[22] Mary Beth Ruskai, Inequalities for quantum entropy: A review with conditions for equality, Journal of Mathematical Physics, Volume 43:58, Issue 9 (2002).

[23] Erling Stormer, Entropy in operator algebras, In Etienne Blanchard; David Ellwood; Masoud Khalkhali; Mathilde Marcolli; Henri Moscovici Sorin Popa (ed.), Quanta of Maths. AMS and Clay Mathematics Institute. 2010.

[24] B. Weiss, Sofic groups and dynamical systems. The Indian Journal of Statistics Special issue on Ergodic Theory and Harmonic Analysis 2000, Volume 62, Series A, Pt. 3, pp. 350-359. .