Probability/Topology – Synopsis of lecture 4

Misha Gromov,

September 24, 2024

Given a set (space) X and a group (semigroup) G of transformations $T : X \to X$, *postulate* the inequality $\sum_1^n mes_G Y_i \le 1$, $Y_i \subset X$, if there exits $g_i \in G$, such that $\bigcap_1^n g_i(Y_i) = \varnothing$.

More generally, one may have a class (category) of ("probability") spaces X and transformations (maps) $g : X_1 \to X_2$, where the above applies to subsets $Y_i \subset X_i$ and maps $g_i : X_i \to X$, or alternatively to maps $g_i : X \to X_i$, where instead of images one takes the pullbacks of subsets $Y_i \in X_i$.

Examples

I. **Finite Sets:** *X is a finite set with equiprobable points (atoms) and with permutation groups acting on them. Or, more generally the two categories $\mathcal{F}_{inj}$ and $\mathcal{F}_{surj}$ of one-to one and of onto maps of finite sets.*

Everything else radiates from this example.

A. X is a finite dimensional liner space, G is the group of linear transformations and $Y_i \subset X$ are linear subspaces.. e/g over the field $\mathbb{R}$ of real numbers. Here 0 replaces $\varnothing$ and

$$mes_G(Y) = dim(Y)/dim(X).$$

This "measure" agrees with that in **I** via the *functor* from $\mathcal{F}$ to the category $\mathcal{FL}$ linear spaces over $\mathbb{R}$ for $X \rightsquigarrow \mathbb{R}^X = \{maps\} : X \to \mathbb{R}$.

B. *Standard (Lebesgue- Rokhlin) probability spaces*[1] X, i.e. those isomorphic to $[0,\delta] \sqcup \{x_i\}$, $\sum_i mes(x_i) = 1\delta$ can be represented as limits of objects in $\mathcal{F}_{surj}$. For instance countable probability spaces $\sqcup\{x_i\}$, $\sum mes)x_i) = 1$ re limits of those with rational $p_i = mes(x_i)$ where rational spaces are quotient spaces of objects in $\mathcal{F}_{surj}$ defined with subgroups of permutation groups commuting with maps $X_1 \to X_2$.

Then all probability spaces are characterized by their maps to finite ones defined by finite partitions, where atomless spaces, which can be partitioned in many ways into subsets with equal measures, can be directly represented by maps to objects from $\mathcal{F}_{surj}$.

We shall rigorously explain this later along with detailed proofs of following three theorems.

(1) *Equipartition Form of the Law of Large Numbers for Powers of Finite Probability Spaces.* Define the measure on the product set $\{z_{ij}\} = \{x_i\} \times \{y_j\}$ by

$$mes(z_{ij} = (x_i, y_j)) = mes(x_i) \cdot mes(y_j).$$

[1]

"Most masses in the power spaces for large N

$$X^N = \underbrace{X \times .... \times X}_{N}$$

is distributed over atoms with approximately equal masses"

In terms of the *non-standard (Leibniz-Robinson) analysis*, if N is an infinitely large (non-stanfdrd) number then $\{x_i\}^N$ is probability space with N equal atoms.

***Hoeffding's Inequality**: the exponential tail bound for general independent random variables* : Let f_i be be bounded, e.g. take values in the interval $[-1,1]$. Then the average of f_i

$$A_N = \frac{1}{N}\sum_{i=1}^{N} f_i$$

exponentially sharply concentrates near its expectation $\mathbf{E}(A_N)$,

$$\mathbf{Prob}\{|A_N - \mathbf{E}(A_N)| \geq c\} \leq 2\exp -\frac{Nc^2}{2}.$$

Proof. Let $\mathbf{E}(f_i) = 0$ and write

$$\mathbf{Prob}\{A_N \geq c\} = \mathbf{Prob}\{\exp \lambda A_N \geq \exp \lambda c\} \leq (\exp -\lambda c)\mathbf{E}(A_N),$$

which, by independence of $\exp \frac{\lambda c}{N} f_i$, is equal to

$$(\exp -\lambda c)\prod_i \mathbf{E}(\exp \frac{\lambda c}{N} f_i) \text{ for all } \lambda > 0,$$

a, where

$$\mathbf{E}(\exp \frac{\lambda c}{2N} f_i) \leq \exp \frac{(\lambda c)^2}{2N^2},$$

since $|f_i| \leq 1$ and $\mathbf{E}(f_i) = 0$.

Then the proof follows with a good choice of λ.

(2) *Loomis-Whitney Isoperimetric Inequality.* Let X_i, $i = 1, ...n$, be standard probability spaces, e.g. all isomorphic to $[0,1]$, let

$$Y_{\hat{i}} \subset X_1 \times ... \times X_{i-1} \times X_{i+1} \times ... \times X_n$$

be measurable subsets and let $Y \subset \times_1^n X_i$ be the intersection of the pullbacks of the obvious maps $P_i : \times_1^n X_i \to X_1 \times ... \times X_{i-1} \times X_{i+1} \times ... \times X_n$ Then

$$mes(Y) \leq \left(\prod_1^n mes(Y_{\hat{i}})\right)^{\frac{1}{n-1}}.$$

****

Idea of the Proof. A simple approximation argument reduces this to the case of finite X_i with equal atoms. If the maps P_i in this case are "homogeneous", i.e. have equal non-empty pullbacks, $card(P_i^{-1}(y_1)) = card(P_i^{-1}(y_2))$, unless one of the two cardinalities is zero, then the inequality is obvious. Then an application of (1) to pushforward measures by P_i shows that the maps

$$P_i^N : (\bigtimes_1^n X_i)^N \to X_1^N \times ... \times X_{i-1}^N \times X_{i+1}^N \times ... \times X_n^N$$

for infinite N are "homogeneous" and the proof follows.[2]

Corollary: Euclidean log-Isoperimetric inequality with non-sharp constant. *The n-volumes of smooth bounded domains $V \subset \mathbb{R}^n$ are bounded by the $(n-1)$-volumes of their boundaries:*

$$vol_n(V) \leq C_n vol_{n-1}(\partial V)^{\frac{n}{n-1}},$$

where $C_n = \frac{1}{2}$.

Indeed, the Lebesgue measures of the orthogonal projections $p_h(V) \subset g$ of V to the hyperplanes $h \subset \mathbb{R}^n$ are bounded by $\frac{1}{2} vol_{n-1}(\partial V)$ and the geometric means of the products the measures of such projections to the coordinates hyperplanes are bounded by their arithmetic means.

More interestingly, by projecting to all hyperplanes $h \subset \mathbb{R}^n$ and averaging over the Grassmann manifold $H = Gr_{n-1}(\mathbb{R}^n) = \mathbb{R}P^{n-1}$, one concludes that

$$\log vol(V) \leq \frac{1}{mes(H)} \int \log mes(p_h(V)) dh + \log C_n,$$

where $C_n = 1/2$.

Unsolved Problem. What is the sharp value of this C_n?

(3) *Linearized Loomis-Whiteny.* Let $\mathcal{X}_i$, $i \in I = \{1, ..., n\}$, be finite dimensional vector spaces over some field $\mathbb{F}$ and denote by $\mathcal{X}_J$, $J \subset I$ the tensor product of $\mathcal{X}_i$ for $i \in J$, i.e. $\mathcal{X}_J = \otimes_{i \in J} \mathcal{X}_i$. Let $\mathcal{Y}_{\hat{i}} \subset \mathcal{X}_{I \smallsetminus \{i\}}$ be linear subspaces and let $\mathcal{Y} = \bigcap_1^n \mathcal{Y}_{\hat{i}} \otimes \mathcal{X}_i$.

Then

$$dim(\mathcal{Y}) \leq \left(\prod_1^n (dim(\mathcal{Y}_{\hat{i}}) \right)^{\frac{1}{n-1}}.$$

Proof. Represent $\mathcal{X}_i$ by spaces of $\mathbb{F}$-functions on finite sets X_i, $card(X_i) = dim(\mathcal{X}_i)$ and observe that there exists a subset

$$Y \subset X_1 \times ...... \times X_n,$$

such that the restrictions of functions on $X_1 \times ...... \times X_n$, from $\mathcal{Y}$ to functions on $Y_{\hat{i}}$ establishes an *isomorphism* between $\mathcal{Y}$ and the space $\mathbb{F}^Y$ of functions on Y, where, of course,

$$card(Y) = dim(\mathcal{Y}).$$

Let

$$Y_{\hat{i}} \subset X_1 \times ... \times X_{i-1} \times X_{i+1} \times ... \times X_n$$

be the projections of Y from $X_1 \times ... \times X_n$ to $X_1 \times ... \times X_{i-1} \times X_{i+1} \times ... \times X_n$ for

$$X_1 \times ... \times X_n, X_1 \times ... \times X_{i-1} \times X_{i+1} \times ... \times X_n,$$

and observe that the restrictions maps from $\mathcal{Y}_{\hat{i}}$ to $\mathbb{F}^{Y_{\hat{i}}}$ are *surjective.*

[2]We shall see later how this homogeneity enters the "topological" proof of the strong subadditivity of the *von Neumann entropy.*

Thus, $card(Y_{\hat{i}}) \leq dim(\mathcal{Y}_{\hat{i}})$ and the proof follows, since

$$dim(\mathcal{Y}) = card(Y) \leq \prod_i card(Y_{\hat{i}})^{\frac{1}{n-1}}$$

by the Loomis-Whitney inequality.

Let is go back (by the *adjoint functor*) from linear spaces over a field $\mathbb{F}$ to sets

linear space $X \rightsquigarrow$ same X with the linear structure forgotten.

Now, in the category of sets, linear subspaces spaces $Y \subset X$ have negligibly small "measures". For instance, if $\mathbb{F}$ is a finite field with q elements, then the set theoretic probability measure, of X is q^{m-n}, which is much smaller than the "linear probability" $\frac{m}{n}$. If $Y_1, Y_2 \subset X$ are subspaces of positive codimension, then, except for the case $q = 2$ and $codim(Y_i) = 1$, $i = 1, 2$ and $g_1, g_2 : X \to X$ two permutations, then the subsets $g_1(Y_1)$ and $g_2(Y_2)$ has almost no chance to intersect in X, and the little chance the had disappears in the limit for $q \to \infty$.

It is obvious, but to make sure let is compute, this is easy, the probability p of $Y_1 \cap g(Y_2) \neq \emptyset$ for a random permutation g and $q \to \infty$ and see that ...

this p does not go to zero!

In fact, if $dim(Y_1) + dim(Y_2) > dim(X)$,or if $dim(Y_1) + dim(Y_2) = dim(X)$ and $dim(Y_1) \neq dim(Y_2)$ then $p \to 1$ and if if $dim(X) - 2dim(Y_1) = 2dim(Y_2)$, then $p \to 1 - \frac{1}{e}$[3]

Thus, the sets $Y \subset X$ with $\log card(Y) = \varepsilon \log card X$ have "probability measures" ε with respect to random permutations of X which is the same as the linear "probability measure" of linear subspaces of this cardinality.

This is simple and the corresponding topological phenomenon is not hard either:

the "probability measure" that of projective subspaces in $\mathbb{R}P^n$ (and also in $\mathbb{C}P^n$ with respect to the group of homeomorphisms, even with respect to all self-homotopy equivalences, is equal to that with respect to the group of projective transformations.

But the question remains: Is there a common reason for this similarity?[4]

Boltzmann-Shannon Entropy

Let $X = \{x_i\}$ be a finite probability space where the measures of the atoms x_i are denoted $p_i = |x_i|$, $p_i \geq 0$, $\sum_i p_i = 1$, and let us define the entropy, denoted $ent(X)$, that is a function of p_i, which must have the following properties

$\bullet_{add}$

$$ent(X \times Y) = ent(X) + ent(Y)$$

for all finite probability spaces;

$\bullet_{cont}$ the entropy is continuous in p_i;

$\bullet_{card}$ if all non-vanishing p_i are mutually equal, then

$$ent(X) = card(X_{>0}),$$

where $X_{>0} \subset X$ is the set of x_i with non-vanishing weights p_i.

Exercise. Show that the logarithms of sums of powers of p_i,

$$e_d(X) = \log \sum_i |x_i|^d,$$

---

[3] Please, check it! I tend to make mistakes in calculations.

[4] This is reminiscent of how algebraic properties of (algebraic varieties over) the field of complex numbers follow from these for finite fields.

and the logarithms of the products of these, e.g. $\log\sum_{ij}|x_i||x_j|$, satisfy $\bullet_{add}$ and $\bullet_{cont}$.

Keeping this in mind and guided by a physicist's

"entropy = logarithm of the number of states"

let $ent_\varepsilon(X)$, $0<\varepsilon<1$ be

the minimum of the logarithms of the cardinalities of subsets $X_\varepsilon\subset X$ *with* $mes(X_\varepsilon)\geq\varepsilon$ and

$$ent(X)=\lim_{\varepsilon\to1}\limsup_{N\to\infty}\frac{1}{N}ent_\varepsilon(X^N),\ 0<\varepsilon<1.$$

Remark. This "limsup" definition is physically meaningless, since it gives no possibility of computing this entropy. But $\frac{1}{N}ent_\varepsilon(X^N)$ converges by the law of large numbers and by the proof of this "law" converges controllably fast.

Exercise. Verify $\bullet_{add}$, $\bullet_{cont}$ and $\bullet_{card}$ for the so defined entropy.

Monotonicity under Reduction. A map between finite probability spaces $f:X\to Y$ is a *reduction* if $mes_Y(f^{-1})=mes_X(x)$ for all atoms $x\in X$.

Every reduction is a composition of "gluing pairs of atoms":

$$f_{ij}:\{x_1,...x_i,...x_j,...x_k,...\}\mapsto\{x_1,...y_{ij},...x_k,...\},$$

where

$$x_i,x_j\overset{f_{ij}}{\mapsto}y_{ij},\ |y_{ij}|=|x_i|+|x_j|.$$

It simple but instructive to check that $ent(Y)\leq ent(X)$ for Y being a reduction of X.

Shannon Inequality due to Boltzmann. Let $f_i:X\to Y_i$, $i=1,...n$. be reductions, such that the map $f=(f_1,...,f_n):X\to Y_1\times....\times Y_n$ (this is *not* a reduction) is *injective*. Then

$$ent(X)\leq\sum_i ent(Y_i)=ent(X_1\times...\times Y_n)$$

The proof follows from the law of large number formulated in functorial terms as follows.

Bernoulli Approximation. The "distance" between X_1 and X_2 is $\leq\varepsilon$ if there exist subsets $Y_1\subset X_1$ and $Y_2\subset X_2$ with $mes(Y_i)\geq1-\varepsilon$, $i=1,2$, and a bijection

$$Y_1\overset{f}{\leftrightarrow}Y_2,$$

where the corresponding atoms $y_1\overset{f}{\leftrightarrow}y_2$ satisfy:

$$(1-\varepsilon)^M\leq\frac{|y_1|}{|y_2|}\leq(1+\varepsilon)^M\text{ for }M=\log card(Y_1)=\log card(Y_2).$$

Concavity. Shannon inequality applied to $\{p_{ij}\}$ for $j=1,2$ implies concavity of the entropy:

$$ent\left\{\frac{p_{i,1}+p_{i.2}}{2}\right\}\geq\frac{1}{2}(ent\{p_{i,1}\}+ent\{p_{i,2}\}).$$

Homogeneous Spaces and Homogeneous Reductions (Morphisms). A finite measure space is homogeneous if all $|x_i|=p_i\neq0$ are mutually equal.

A reduction $f : X \to Y$ between homogeneous spaces is homogeneous if the cardinality $card(f^{-1}(y))$ doesn't depend on y for all y in the image $f(X) \subset Y$.

Functorial Law of Large Numbers. Let $f : X \to Y$ be a reduction between finite probability spaces. Then, for all $\varepsilon > 0$ and all sufficiently large $N \geq N(\varepsilon)$, there exist *homogeneous* spaces X_N and Y_N which are *ε-close to the power spaces* X^N and Y^N and such that the map $f_N : X_N \to Y_N$, which correspond to f is also *homogeneous.*

Thus, for instance, the general Shannon inequality reduces to the (trivial) homogeneous case. Another corollary is the Boltzmann formula (first written down by Max Planck)

$$ent(X) = \sum_i p_i \log p_i$$

as well as the "local formula":

The mass of a typical random (multi)atom $\chi \in X^N$ satisfies

$$\sqrt[N]{|\chi|} = \exp -ent(X) + o(1) \textit{ for } N \to \infty.$$

The Shannon inequality written as $ent(X_{12}) \leq ent_{(}X_1) + ent(X_2)$. can be refined to the **Relative Shannon Inequality.**

$$ent(X_{123}) \leq ent(X_{23}) + ent(X_{13}) - ent(X_3).$$

+++++++++++++++++++++++++++++++=

STATES INSTEAD OF SPACES

Let S be a (countable in the present case) set S of "sites" s, e.g. $S = \mathbb{Z}^3$, where the sites are position of molecules in a crystal.

States (which are essentially the same as the above X) are probability measures on the powers S^i.

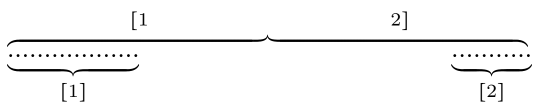
Additivity for non-Interacting states P.
Systems.

$$ent(P_1 \times P_2) = ent(P_1) + ent(P_2).$$

Symbolically:

$$ent[1 \qquad\qquad 2] = ent[1] + ent[2]$$

Pictorially:

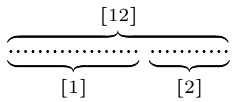


Subadditivity for Joint
Interacting systems.

$$ent(P_1 \vee P_2) \leq ent(P_1) + ent(P_2)$$

or

$$ent[12] \leq ent[1] + ent[2]$$



This "$\vee$" is *not a canonically defined* operation; correct notation would be "$\vee_\rho$" where ρ is a particular "relation/interaction" between P_1 and P_2.

For instance, if P_1 and P_2 *do not interact*, then $P_1 \vee P = P_1 \times P_2$; if P_1 and P_2 are *related by a reduction* $P_1 \overset{\rho}{\to} P_2$ then, by definition, $P_1 \vee_\rho P_2 = P_1$.

Formally, one may define $P_1 \vee P_2$ as a probability space Q, such that

$$set(Q) \subset set(P_1) \times set(P_2)$$

and such that the coordinate projections $Q \to P_1$ and $Q \to P_2$ are reductions.

The following

Strong Subadditivity of Entropy is less intuitive than simple "subadditivity".

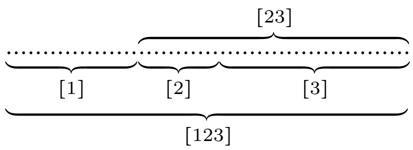
$$ent(P_1 \vee P_2 \vee P_3) + ent(P_2) \le$$
$$ent(P_1 \vee P_2) + ent(P_2 \vee P_3),$$

or

$$ent[123] + ent[2] \le ent[12] + ent[23]$$

(According to our definition of "$\vee$",

$set[123] \subset set[1] \times set[2] \times set[3]$ where the coordinate projections $[ijk] \to [ij] \to [i]$ are reductions.)



Corollary: Entropic Loomis -Whitney Inequaliy

$$2 \cdot ent[123] \le ent[12] + ent[23] + ent[13].$$

∘∘ ∘ ∘ ∘ ∘ ∘ ∘ ∘ ∘ ∘ ∘ ∘ ∘ ∘ ∘

If Q is a reduction of P then

$$ent(Q) \le ent(P).$$

(This seems a most natural property but it fails to be true in the quantum case.)

a (stochastic) state is a probability measure on S.

Then each subset $\subset S$

****

Von Neumann Entropy

Given a (complex or real) Euclidian space S, where we are not tied up to a particular orthogonal basis for defining entropy: one basis of orthonormal vectors 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.

Definitions of the Von Neumann Entropy.

(1) *Minimalistic Definition.* Extracte 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) *Spectral definition.* 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

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

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) *ε-Definition.* Denote by $\mathcal{T}_\varepsilon = \mathcal{T}_\varepsilon(S)$ the set of the linear subspaces $T \subset S$ such that $P(T) \geq (1-\varepsilon)P(S))$ 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 $P(s_j) \geq P(s_k)$ for all $j \in J$ and $k \in I \setminus J$.

(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

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

for all states P on S.

correct: 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

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

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:

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

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):

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

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 $(*)$, 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

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

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

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

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

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

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

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

.

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 $\lambda_1 \le \lambda_2 ..., \lambda_i, ...$. By Pythagorean theorem, $dim(U)^{-1} \sum_i \lambda_i = 1 - B(T)$; therefore the span U_ε of those u_i where $\lambda_i \ge 1 - \varepsilon$ 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 $a_* - b_* \sim c_* \ge 0$.

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 •, 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,

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

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

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

In particular,

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

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

Thus, $|C(u)|_\star \leq c|D(u)|_\star$ 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

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

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

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

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 $\check{}\Upsilon_{ij} : \underline{P}_i \times \underline{P}_j \to \mathbb{C}$, 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

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

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 [?] and the most transparant one to Effros [?].

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 [?] and in [?], 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 [?] 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 [?] and he also pointed out to me to the paper [?] 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 [?] 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 [?]. 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 [?],[?] for how it goes along different lines.)

****

More on the laws of large numbers: cubes, spheres, Maxwell distribution, etc

P Levy isoperim inequality

Concentration

Large dimension of linear subspaces corresponding to subsets with large measure spaces Intersections of random subsets similar to that of linear subspaces in vector ver finite fields

"Large homological topology" $\implies$ large probability

Isoperimetric inequality

concentration

Visual diameter

maps $S^n \to \mathbb{R}^m$

A space Y with an automorphism group, G

Largeness of subsets $X_1, X_2 \subset$ as tendency of X_1 intersect $g(X_2)$