Six Lectures on Probabiliy, Symmetry, Linearity. October 2014, Jussieu. (Unedited)

Misha Gromov

January 28, 2015

Contents

1 Perspectives 1

2 Naive Mathematician's Entropy. 12

3 Representation of infinite probability spaces X by sets of finite partitions of X and Kolmogorov's dynamical entropy . 53

4 Simplex of Probability "Laws" and Fisher Metric. 71

5 Mendelian Next Generation Map. 90

6 Von Neumann's "Hilbertization" and Strong Subadditivity of Entropy. 115

7 Measures Defined via Cohomology with Applications to the Morse Spectra and Parametric Packing Problem. 145

8 Two Words On Large Deviations. 179

1 Perspectives

.

What is Randomness?

What is Probabiliy?

Standpoints and contexts:

(Non-exhaustive list of 10 items.)

◦ History of the idea of probability.

• Use and misuse of Metaphors of randomness.

• Psychology of randomness.

• Natural evolution and human history.

○ Statistics in physics, in astronomy and in formal genetics.

○ Probabilistic reasoning in combinatorics and in geometry; randomization of categories.

○ Categorisation and generalisation of measure theory and of probability.

□ Molecular evolution.

□ Statistical analysis of natural languages.

□ Learning languages and learning mathematics.

Two related questions:

What is entropy?

What makes "information" in the cells and in the brains non-Shannon?

Origin of Probability Theory.

Gambling.

Rituparna, a king of Ayodhya said 5 000 years ago:

I of dice possess the science and in numbers thus am skilled.

(Cardano, Galileo, Pascal, Huygence, Bernoulli...)

Brownian Motion.

... small compound bodies...
are set in perpetual motion
by the impact of invisible blows...

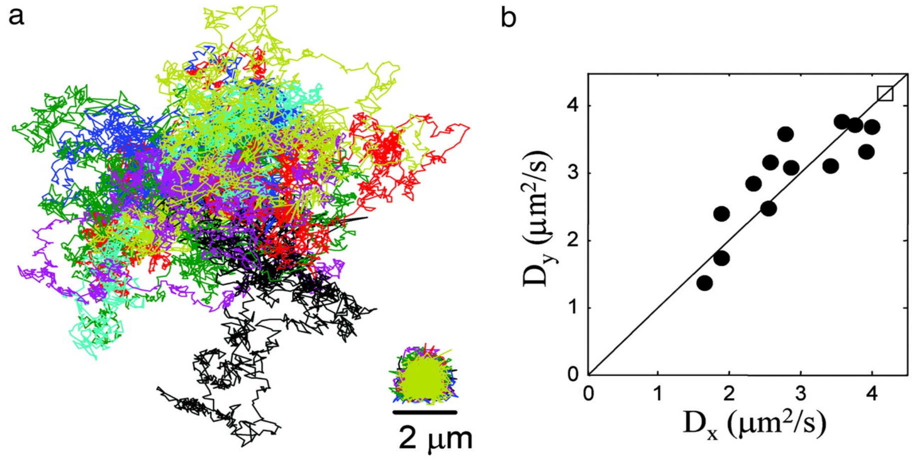
.

The movement mounts up
from the atoms
and gradually emerges
to the level of our senses.

Said by...

Titus Lucretius in 50 BCE.

Translated to numbers by
Thiele (1880), Bachelier (1900),



EINSTEIN (1905), SMOLUCHOWSKI (1906), WIENER (1923).

Is, as Maxwell believed,

The true logic of this world the calculus of probabilities?.

Are

all the mathematical sciences founded on relations between physical laws and laws of numbers?

SYMMETRY ENHANCED BY RANDOMNESS.

"Classical probability" depends on

(quasi)invariant Haar(-like) measures, such as the infinite product measure in the space X of binary sequences, denoted

$$\{\circ_{\frac{1}{2}}, \bullet_{\frac{1}{2}}\}^{\mathbb{N}}, \ \mathbb{N} = 1, 2, 3, ...,$$

that are functions from $\mathbb{N}$ to the set $\{\circ, \bullet\}$ with both values $\circ$ and $\bullet$ being equiprobable.

Besides the transitive action of the compact group $\mathbb{Z}_2^{\mathbb{N}}$, this X is acted by the infinite "permutation" group of the set $\mathbb{N}$; in fact the number structure in $\mathbb{N}$ is immaterial at this point – any countable set S in place of $\mathbb{N}$ will do.

The *"Haar symmetry"* may be not apparent in the space

$$X = \{\circ_p, \bullet_{1-p}\}^S$$

for $p \neq 1/2$, but this space has a

kind of *"tensorization symmetry"*: X can be decomposed in a variety of ways into Cartesian product of several spaces isomorphic to this very X. In fact, every partition of the (countable) set S into infinite subsets induces such a decomposition of X.

This kind of "symmetry" is pronounced in in Gaussian measures such as measures *Brownian Wiener measures* and up to a lesser degree in more general Gibbsian (Boltzmann's) measures as it underlies the *Bernoulli Approximation Theorem* and *Boltzmann entropy* as we shall see below.

More recently, Schram discovered *a conformally (quasi)invariant* measures in spaces of curves in Riemann surfaces that are parametrised by Brownian measures.

Most likely, there are no *mathematically significant* probability measures that would be *"fully asymmetric"*.

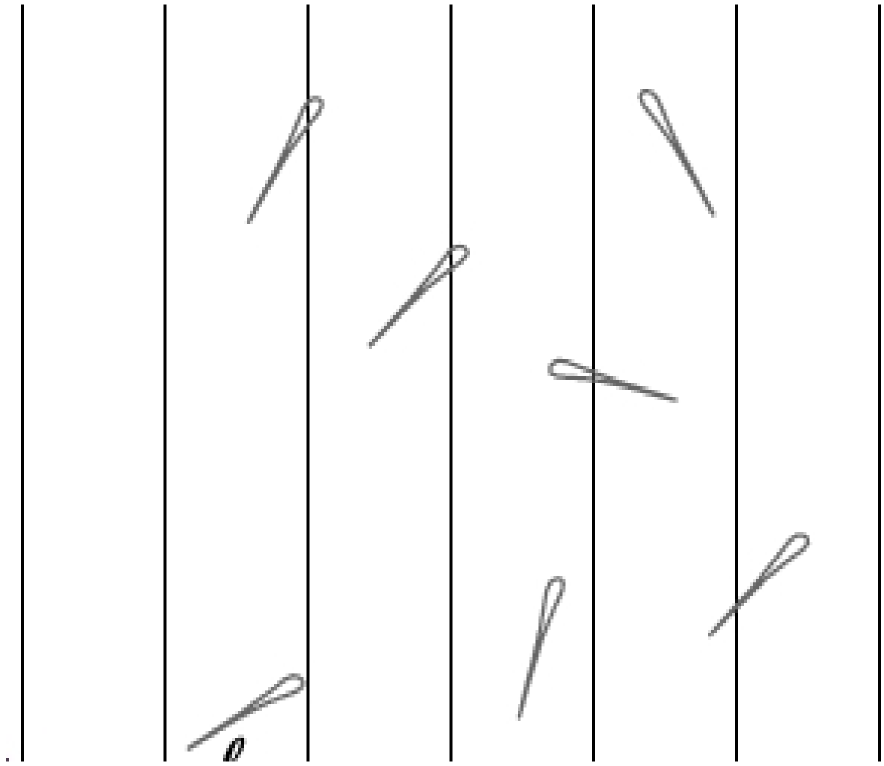
And if one can not *postulate* equiprobability and/or "parametrise randomness by independence" then the classical probability does not work in the "real world".

Probability and Measure.

According to Kolmogorov (1933) (and Buffon's needle example of 1733) "random events" are represented by subdomains Y in the square $\blacksquare$;

probability of Y *equals* $area(Y)$.

Kolmogorov's mode of thought (as that in André Weil's *Foundations*



of Algebraic Geometry 1946) is manifestly set theoretic one.

On the other hand, Boltzmann's ideas in statistical mechanics are naturally embraced by the language of *non-standard analysis* and by Grothendieck's style category theoretic formalism.

Also, the idea of probability in

languages and in *mathematics of learning* deviates from Kolmogorov-Buffon ■.

The notion of a probability of a sentence is an entirely useless one, under any interpretation of this term [that you find in 20th century text-books].

Naum Chomsky.

In fact, "linguistic events" such as sentences, are members of category theoretic like "anssembles" that are structurally quite different from what one encounters in statistical mechanics and in physics in general. And probabilities of these "events" are not number valued functions but rather something akin to functors from "categories of linguistic events" to similar, but significantly simpler/smaller, combinatorial category, such, for instance, as the category of (*weighted*) *trees*.

⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆⋆

Alternatives to Classical Probability and Entropy.

1. Entropy via Grothendieck Semigroup.

2. Probability spacers as covariant functors.

3. Large deviations and Non-Standard analysis for classical and quantum entropies.

4. Linearized Measures, Probabilities and Entropies.

5. Combinatorial Probability with Limited Symmetries.

We discuss the above issues at length

in the following articles posted on http://www.ihes.fr/ gromov/ in section "Recent"

Quotations and Ideas,

Ergostructures, Ergodic and the Universal Learning Problem,

Math Currents in the Brain,

In a Search for a Structure.

2 Naive Mathematician's Entropy.

... *pure thought can grasp reality...*
. Albert Einstein.

...exceedingly difficult task of our time is to work on the construction

of a new idea of reality.... .

Wolfgang Pauli.

The introduction of the cipher 0 or the group concept was general nonsense...

Alexander Grothendieck.

A "physisit's system" $\mathcal{S}$, e.g. "supported by a crystal", is an infinite ensemble of "infinitely small" *mutually equal* "states". The logarithm of the properly normalised number of these states is *(mean statistical Boltzmann) entropy of* $\mathcal{S}$.

The "space of states" of $\mathcal{S}$ is NOT a mathematician's "set", it is "something" that depends on a class of *mutually equivalent* imaginary experimental protocols.

Finite Measure Spaces.

A finite measure space $P = \{p\}$ is a finite set, called a *background of P* or *a supporting* set of P, denoted $set(P)$, of "atoms" p with a positive function $set(P) \to \mathbb{R}_+$, denoted $p \mapsto |p| > 0$, thought of as weights/masses of atoms.

If one gives a "name" to this background, say $I = set(P)$, then one writes $P = \{p_i\}_{i \in I}$ where p_i denote not "atoms themselves" but rather their weights.

(Should one allow "atoms" p with $|p| = 0$? Classical answer is: "not necessary"; but "massless atoms" can not be swept under the rug in quantum probability.)

$|P| = \sum_p |p|$: the (total) *mass* of P.

If $|P| = 1$, then P is called a *probability space.*

Scaling and Normalization. The multiplicitive group $\mathbb{R}_+^\times$ acts on "the set of all" measure spaces P by what we call λ-*scaling*, $\lambda \in \mathbb{R}_+^\times$,

$$P \mapsto \lambda \cdot P = \{\lambda p\} \text{ for all } p \in P.$$

If $\lambda = 1/|P|$, then this is called *normalisation* of P that turns a finite measure space P into a *probability* space.

In statistics – and this is essential – "probabilities" are defined (if at all) via *ratios* $n_i : n_j$ of numbers of occurrences of "events" $i \in I$, where the normalising factor $1/\sum n_i$ (and even the set I) is often unavailable.

Thus, ”probability distributions on I” reside in the *projective space* $(\mathbb{R}^I \smallsetminus 0)/\mathbb{R}^\times_+$ rather than in the Euclidean space $\mathbb{R}^I$ itself.

Reductions. A map $P \overset{f}{\to} Q$ (that is a shorthand for $set(P) \overset{f}{\to} set(Q)$) is a *reduction* if the *q-fibers* $P_q = f^{-1}(q) \subset P$ satisfy $|P_q| = |q|$ for all $q \in Q$. Q itself in this case is called *a reduction of* P.

Since the space Q is uniquely determined by P and by the map between sets, $f : set(P) \to set(Q)$, this Q may be called *the* f*-reduction*, sometimes just *a reduction of* P.

(Imagine Q is an ”apparatus” for observing P. What you see of P is a ”reduced picture” of what ”filters” through the ”windows” of Q.)

Conditioning. The group $(\mathbb{R}_+^\times)^{set(Q)}$ of functions

$$\lambda : set(Q) \to \mathbb{R}_+^\times,\ q \mapsto \lambda_q,$$

acts on the fibers $P_q \subset P$ of reductions $f : P \to Q$ by λ_q-*scaling*,

$$P_q \mapsto \lambda_q P_q.$$

If $\lambda(q) = 1/|P_q|$ (where usually P is a probability space), then this fiberwise normalisation is called *conditioning of P associated to f.*

Notice that this conditioning *does not depend* on the measure space structure Q in the set $set(Q)$. The definition makes sense for all (surjective) maps from measure spaces P to (finite) sets J (with $J = set(Q)$ in our case), where such maps $f : P \to J$ may be called *J-partitions* of P into *fibers/slices* $P_j = f^{-1}(j) \subset$

P.

Cartesian Products:

$P \times Q = \{(p, q)\}$ with the masses

$$|(p, q)| = |p| \cdot |q|.$$

(Think of $P \times Q$ as a joint system with *non-interacting* components P and Q.)

$\mathcal{P}$: Category of finite probability measure spaces and reductions.

Why $P \overset{f}{\to} Q$ rather than "simply" $P \succ Q$?

Physically, "$\succ$" per se is meaningless, it must be implement by a particular operation f. (In fact, "protocols of attaching Q to P", make 2-*category*.)

Notationally, one may write $ent(f)$ but not $ent(\succ)$.

However, many concept of probability theory can not be expressed purely in the language of the category $\mathcal{P}$. For instance, the above defined "*conditioning of* P" needs maps from probability spaces to "*bare sets*".

⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆⋆

Let us make sense of :

Entropy is a number equal the "logarithm of the number of states" of ???

⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆⋆

Expected/Desirable Properties of Entropy.

Additivity for non-Interacting Systems.

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

Symbolically:

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

Pictorially:

$$\overbrace{\underbrace{\cdots\cdots\cdots\cdots\cdots\cdots}_{[1]} \qquad\qquad\qquad \underbrace{\cdots\cdots\cdots\cdot}_{[2]}}^{[1 \qquad\quad 2]}$$

Subadditivity for Joint Interacting systems.

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

or

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

$$\overbrace{\underbrace{\cdots\cdots\cdots\cdots\cdots\cdots}_{[1]} \; \underbrace{\cdots\cdots\cdots\cdot}_{[2]}}^{[12]}$$

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

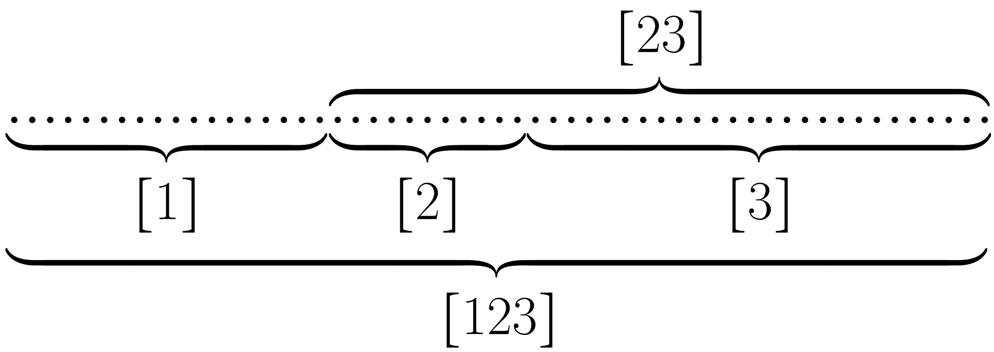
is less intuitive than simple "subadditivity".

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

or

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

(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.

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

○○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○ ○

If Q is a reduction of P then

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

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

The entropy of a finite probability space is related to the cardinalty of the underlying set by the inequality

$$ent(P) \leq \log |set(P)|,$$

where the equality takes place if and only if the measure is equidistributed (homogeneous) on $set(P)$, i.e. if all atoms $p \in P$ have equal weights $|p| = |set(P)|^{-1}$.

(However simple, this links $\mathcal{P}$ with another world – the category of finite sets.)

The very existence of entropy with

all these properties (that we prove below) harbours unexpected mathematical resources. For instance, it implies the following refinement of the (non-sharp) *isoperimetric inequality*

$$vol_n(Y)^{k-1} \leq vol(\partial Y)^k$$

for all measurable subsets Y in the Euclidean space $\mathbb{R}^k$.

LOOMIS-WHITNEY INEQUALITY.

Among all subsets $Y \subset \mathbb{R}^k$ with given measures of the projections to the k coordinate hyperplanes, the maximal measure is achieved by rectangular solids

(+ subsets obtained from them by measurable transformations of $\mathbb{R}^k$ preserving the coordinate line partitions).

This follows from the strong subadditivity and $ent(P) \leq log|set(P)|$.

In the first non-trivial case of $k = 3$, the above inequality for Y, denoted

$$Y = [123]_{\square} \subset \mathbb{R}^3,$$

says that

$$vol^2[123]_{\square} \leq$$

$$area[12]_{\blacksquare} \cdot area[13]_{\blacksquare} \cdot area[23]_{\blacksquare}$$

for the three planar coordinate projections

$[ij]_{\blacksquare} \subset \mathbb{R}^2,\ i,j = 1,2,3,$ of $[123]_{\square}$.

The proof of this starts with an obvious approximation argument that reduces the desired inequality between volume and areas to the corresponding inequality between cardinalities (denoted |...|) of finite subsets in product sets,

$Y_{appr} = [123]_{:::} \subset R = R_1 \times R_2 \times R_3$,

and their "binary" coordinate projections to $R_1 \times R_2$, $R_1 \times R_3$ and $R_2 \times R_3$:

$|[123]_{:::}|^2 \leq |[12]_{::}| \cdot |[13]_{::}| \cdot |[23]_{::}|$.

Then the latter inequality is proven by applying

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

to the probability space $[123]$ with $set[123] = [123]_{:::}$ and with the equidistributed measure on this set $[123]_{:::}$, where one uses the *equality*

$$ent[123] = \log |[123]_{:::}|,$$

along with the *inequalities*

$ent[ij] \geq log|[ij]_{::}|, \ i, j = 1, 2, 3, i \neq j$.

for the measure projections (reductions) $[ij]$ of $[123]$ to $R_i \times R_j$.

OPEN QUESTION. The Loomis Whitney inequality is invariant under the group S_k of the coordinate permutations of $\mathbb{R}^k$ and of Cartesian power sets R^k in general.

Are there similar inequalities for other symmetry groups?

For instance, is there an *orthogonally invariant* version of the Loomis-Whitney inequality in $\mathbb{R}^k$ that would imply the *sharp* isoperimetric inequality?

Linearized Loomis-Whitney. The Loomis-Whitney inequality for coordinate projections of a subset Y in the Cartesian product of finite sets, $Y \subset R = R_1 \times R_2 \times \ldots \times R_k$ can be formulated in terms of the $(k+1)$-linear form $\Phi_Y(z, x_1, x_2, ..., x_k)$ on functions $z : R \to \mathbb{R}$, $x_i : R_i \to \mathbb{R}$, that is

$$\Phi_Y(z, x_1, ..., x_k) = \sum_{(r_1,...,r_k)\in Y} z(r_1, ...r_k) \cdot x_1(r_1) \cdot ... \cdot x_k(r_k) d\mu.$$

Then the so written inequality easily (this is an exercise) generalises to all multilinear forms Φ. For instance, the algebraic counterpart of the $3D$ Loomis -Whitney reads:

The bilinear forms associated with an arbitrary 4-linear form Φ (over any field) satisfy:

$$|\Phi(x_1, x_2 \otimes x_3 \otimes x_4)|^2 \leq$$
$$|\Phi(s_1 \otimes x_2, x_3 \otimes x_4)| \cdot |\Phi(x_1 \otimes x_3, x_2 \otimes x_4)| \cdot$$
$$\cdot |\Phi(x_1 \otimes x_4, x_2 \otimes x_3)|$$

where

$$|\Phi(\ldots , \ldots)|$$

denote *ranks* of these bilinear forms.

(See *Entropy and Isoperimetry for Linear and non-Linear Group Actions*, in section "Recent" on my webpage.)

Grothendieck Semigroup, Bernoulli isomorphism and Entropy.

Homogeneous Spaces.

Categorically, P is *homogenous* if the morphisms $P \to Q$ that are invariant under the group of automorphisms of P factor through $P \to \bullet$ for terminal objects $\bullet \in \mathcal{P}$.

Concretely: *all atoms* $p \in P$ *have equal masses* $|p|$.

Reductions $f : P \to Q$ between

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

Entropy of a homogeneous P is *defined* as

$$ent(P) = \log |set(P)|.$$

Asymptotic Equivalence.

Injective correspondence between probability spaces:

$$P \supset P' \ni p \overset{\pi}{\leftrightarrow} q \in Q' \subset Q.$$

$$|p : q| = \max(p/q, q/p)$$

$$M = \min(|set(P)|, |set(Q)|).$$

Additive Distance:

$$|P - Q|_\pi = |P \smallsetminus P'| + |Q \smallsetminus Q'|.$$

Normilaised Multiplicative Distance:

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

Total Distance:

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

$$DIST(P,Q) = \inf_\pi dist_\pi(P,Q).$$

Definition. $\{P_N\}$ *is asymptotically equivalent to* $\{Q_N\}$ *for* $N = 1,2,3,... \to \infty$, if

$$DIST(P_N, Q_N) \to 0.$$

Bernoulli Topological Semigroup. $Ber(\mathcal{P})$ is the set of the asymptotic equivalence classes of $\{P^N\}$ for all $P \in \mathcal{P}$, where the product in $Ber(\mathcal{P})$ is induced by the Cartesian product $P \times Q$ and the topology by $DIST(P^N, Q^N)_{N\to\infty}$.

Definition: Boltzmann entropy of P: *Bernoulli class of* P *in* $Ber(\mathcal{P})$.

The Law of Large Numbers $\Rightarrow$

Approximation/Equipartition Theorem. The sequence of Cartesian powers P^N admits *Bernoulli approximations for all finite probability spaces* P.

By definition, this means that, for each P, there exists a sequence H_N of *homogeneous* probability spaces that is *asymptoticly equivalent* to the sequence of the Cartesian powers P^N.

This is a reformulation of the Law of Large Numbers applied to the (real valued) function $p \mapsto \log p$ (random variable) on the probability space P. (One can avoid logarithms if one is comfortable with the law of large numbers for functions with values in the *multiplicative group* $\mathbb{R}_+^\times$, where the relevant function is the tautological $\mathbb{R}_+ \ni p \mapsto p \in \mathbb{R}_+^\times$.)

Permutation Symmetry of Bernoulli subsets.

The supporting sets $set(H_N)$ of these homogenyous H_N naturally come up as subsets in the Cartesian power set $(set(P))^N$ where $(set(P))^N$ is acted upon by the symmetric group Sym_N that permutes N factors of this Cartesian product.

This action preserves the power measure of P^N and the *Bernoulli* *subsets*

$$B_N = set(H_N) \subset (set(P))^N$$

can be chosen (by no means uniquely) Sym_N-invariant as well.

In fact all proofs of the law of large numbers deliver invariant subsets $B_N \subset (set(P))^N$ such that the corresponding homogeneous spaces H_N, with equidistributed probability measure on the sets B_N that serve as $set(H_N)$, Bernoulli approximate P^N.

Moreover, this approximation issues from the inclusions $\pi_N : H_N \subset P^N$, that is

$$dist_{\pi_N}(H_N, P^N) \to 0 \text{ for } N \to \infty.$$

Thus one may justifiable say that

the sequences $B_N \subset set(P)^N$

Bernoulli approximate P^N.

It seem unquestionable, that the S_N-symmetry is an indispensable property of Bernoulli approximation; yet, it plays no role at the present moment.

Bernoulli Entropy:

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

This definition allows an effortless reduction of basic properties of entropy, such as subbadiitivity, to the corresponding inequalities for homogeneous spaces where these inequalities are obvious.

It is also clear that

"Boltzmann = Bernoulli": *The homomorphism $H \mapsto |set(H)|$*

for homogeneous spaces $H \in \mathcal{P}$ *extends to a topological isomorphism of the Bernoulli semigroup* $Ber(\mathcal{P})$ *onto the multiplicative semigroup* $\mathbb{R}^{\times}_{\geq 1}$ *of real numbers* ≥ 1.

Corollary: "Boltzmann Formula":

$$ent(P) = -\sum_{p \in P} |p| \log |p|$$

(that may be misguidedly taken for the definition of entropy).

In other words, $\{P^N\}$ is asymptotically equivalent to $\{Q^N\}$ if and only if $\sum_p |p| \log |p| = \sum_q |q| \log |q|$.

(There is nothing "misguided" about physisists'

$$ent(P) = -K \sum_p |p| \log |p|.$$

This formula, where K is the unit conversion constant, numerically links microworld on the $10^{-9\pm1}m$ scale with what we see with the "naked eye".)

On Sharpness of Entropy Inequalities. The proof of the subadditivity and/or of the bound $ent(P) \leq \log|P|$ via Bernoulli approximation theorem does not tell you what are the equality cases in these inequalities; one has resort to an analytic use of Boltzmann's formula.

Is there a more conceptual proof of sharpness of these inequalities away from the standard equality cases?

Relative Entropy $ent_\lambda(\mu)$. This is classically defined for pairs of measures λ and μ on some space X, where $\mu = f \cdot \lambda$ for a λ-measurable function f on X, such that the integral $\int f|\log f|d\lambda$ converges.

If the function $f = f(x)$ is *constant* on its *support*

$$S = supp(f) = \{x \in X\}_{f(x)\neq 0} \subset X,$$

then one postulates:

$$ent_\lambda(\mu) = \log|S| =_{def} \log\lambda(S),$$

where

$$\log|S| = \log\mu(S) - \mu(S)^{-1}\int_S \log f d\mu,$$

since f equals $\mu(S)/|S|$ on S.

In general, if $f(x)$, $x \in S$, is *non-constant*, the entropy of $\mu = f\cdot\lambda$ is defined, at least for *probability* measures μ, as earlier via the corresponding Bernoulli approximation (law of large numbers) for the tensorial powers $\mu^{\otimes N}$ on the Cartesian power spaces

$$(X^{\times N}, \lambda^{\otimes N}) \text{ for } N \to \infty.$$

And if $\mu = f \cdot \lambda$ is a *probability* measure, that is if $\int f d\lambda = 1$, then the traditional Boltzmann formula for the resulting entropy reads

$$ent_{\lambda}(\mu) = \int_S \log \frac{1}{f} d\mu = \int_S f \log \frac{1}{f} d\lambda$$

If λ and μ are measures on a finite set I with atoms of weights

$$q_i = \lambda(i) \text{ and } p_i = \mu(i)$$

then we write $ent_Q(P)$ for $ent_{\lambda}(\mu)$ and observe that if $q_i = 1$ for all i, then

$$ent_Q(P) \ \textit{equals} \ \ ent(P).$$

And if $Q = \{q_i\}$ is a homogeneous *probability* space with all atoms of weights $q_i = 1/m$, $m = |I|$, then

$$ent_Q(P) = ent(P) - \log m.$$

In general, for all background measures λ, the basic properties of $ent_{\lambda}(\mu)$

follow from those of $ent(P)$ since general measure spaces (X, λ) (without atoms) can be approximated by finite ones with equidistributed measures on them.

The beloved physicists' example is where λ is the *Liouville*(-Lebesgue-Haar) infinite(!) measure on $X = \mathbb{R}^{2n}$.

In the information theory, one is keen on λ *and* μ being both *probability* measures. In this case, obviously,

$$ent_\lambda(\mu) \leq 0$$

and $-ent_\lambda(\mu)$ is regarded as (*the Kullback-Leibler*) *information divergence* between λ and μ.

Mellin's $\zeta_P(s)$. Additivity of the entropy under Cartesian products is shared by many other invariants of $P = \{p\}$ that are symmetric functions of the weights $|p|$, $p \in P$.

In a way, all such functions are encompassed by the *Mellin transform* of $P = \{p\}$, that is

$$\zeta_P(s) = \sum_{p \in P} |p|^s.$$

Since the function ζ_P itself is multipilcative

$$\zeta_{P \times Q}(s) = \zeta_P(s) \cdot \zeta_Q(s),$$

$\log \zeta_P(s)$, that is defined for all real s, is additive and the integrals

$$\int \log \zeta_P(s) \phi(s) ds$$

are additive for $P \times Q$ as well.

But in order to extract entropy of P from ζ_P one has to differentiate rather than to integrate:

$$-ent(P) = \sum_{p\in P} |p| \log |p| = \frac{d}{ds}\zeta_P(s)_{s=1},$$

where the additivity of the value of $\frac{d}{ds}\zeta_P(s)$ at $s = 1$ for Cartesian product of (only!) *probability* spaces follows from multiplicativity of ζ and the relation

$$\zeta_P(1) = \sum_{p\in P} |p| = 1.$$

On Riemann's $\zeta(s) = \zeta_{\mathbb{N}}(-s)$*.* The multiplicativity of $\zeta_P(-s)$ for the infinite measure space

$$P = \mathbb{N} = \{1, 2, 3, 4, 5, ...\},$$

that decomposes into the (restricted) Cartesian product of the "prime measure spaces"

$$P_2 = \{1, 2, 4, 8, 16, ..., \},$$

$$P_3 = \{1, 3, 9, 27, 81, ...\},$$
$$P_5 = \{1, 5, 25, 125, 625, ...\},$$
$$P_7 = \{1, 7, 49, 343, 2401, ...\},$$
$$P_{11} = \{1, 11, 121, 1331, 14641, ...\},$$
$$...,$$

is known as

Euler's product formula.

Question. Is there anything interesting in counterparts to entropy for categories of spaces $X = \{x_i\}_{i \in I}$ where x_i are elements of a *topological semiring* (some monodical category?) R different from real numbers?

Arithmetically attractive R would be fields of p-adic numbers and/or adelic rings. But a more realistic possibility, motivated by Mendelian dynamics, is where R is *a trun-*

cated polynomial ring with its the exponential and logarithmic functions similar to those for real numbers.

Also one wonders whether it is worthwhile to replace finite sets I by something more substantial, e.g. by some "algebraic varieties" with constructive R-valued functions on them.

Grothendieck-Bernoulli Group and (Relative) Entropy of Reductions.

Grothendieck group $Gr(\mathcal{P})$ is generated by symbols $[f]$ for all reductions $f : P_1 \to P_2$ with the relations

$[f_1 \circ f_2]_{Gr} = [f_1]_{Gr} + [f_2]_{Gr}$

and

$[f \times id_Q]_{Gr} = [f]_{Gr}$ for the identity morphisms $id_Q : Q \to Q$ of all $Q \in \mathcal{P}$, where

$f \times id_Q : P_1 \times Q \to P_2 \times Q$.

As usual, $[P]_{Gr}$ is defined for spaces (objects) P as

$[P_{Gr} \to \bullet]$ for $P_{Gr} \to \bullet$ being the morphisms to the (terminal) monoatomic space.

$DIST$ naturally extends from objects to morphisms (reductions) in $Gr(\mathcal{P})$ in a natural way: such a distance between $f_1 : P_1 \to Q_1$ and $f_2 : P_2 \to Q_2$ must be implemented by pairs of partially defined correspondences $\pi : P_1 \leftrightarrow P_2$ and $\chi : Q_1 \leftrightarrow Q_2$, such that the obvious diagram commutes. (One could replace "*commute*" by "*commute up-to* ε" and add this ε to $DIST$.)

We keep the same notation for the

so extended $DIST$ and for the corresponding (*non* translation invariantt) metric induced on the group $Gr(\mathcal{P})$.

The Grothendieck-Bernoulli group. Divide the group $Gr(\mathcal{P})$ by the *Bernoulli equivalence relation*

$$g_1 \underset{\infty}{\sim} g_2 \Leftrightarrow DIST(N{\cdot}g_1, N{\cdot}g_2) \underset{N\to\infty}{\to} 0$$

and denote the resulting qutient space with the induced group structure by $GroBer(\mathcal{P})$.

Since Cartesian products of reductions $f_1 \times f_2 : P_1 \times P_2 \to Q_1 \times Q_2$ decompose as

$$P_1{\times}P_2 \overset{f_1\times id_{P_2}}{\to} Q_1{\times}P_2 \overset{id_{Q_1}\times f_2}{\to} Q_1{\times}Q_2$$

one has:

$$[f^N]_{Gr} = N\cdot[f]_{Gr};$$

This allows a use of the *Law of large*

numbers that (obviously) yields the *Bernoulli approximation theorem for reductions* as well as for spaces. It follows that

the group $GroBer(\mathcal{P})$ is isomorphic to the multiplicative group $\mathbb{R}_+^\times$ with the image of $\mathcal{P}$ consisting of numbers ≥ 1.

A posteriori,

$$ent(f : P \to Q) = ent(P) - ent(Q).$$

But the categorical definition of $ent(f)$ (unlike the above numerical formula) generalises to reductions f between *countable infinite* measure spaces

$$P = \{p_i\} \text{ and } Q = \{q_j\},$$

$$\sum_i |p_i| = \sum_j |q_j| = 1, i, j = 1, 2, 3, \dots .$$

Even if these spaces have *infinite*

entropies:

$$\sum_i |p_i| log |p_i| = \sum_j |q_j| \log |q_j| = -\infty,$$

$ent(f : P \to Q)$ may remain *finite* nevertheless.

$\rightrightarrows \rightrightarrows \rightrightarrows \rightrightarrows \rightrightarrows \rightrightarrows \rightrightarrows \rightrightarrows \rightrightarrows \rightrightarrows \rightrightarrows \rightrightarrows$

Fans and $\vee$-Categories. 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".

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.)

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

Bernoulli Functoriality Question. As we mentioned above, the Bernoulli approximation theorem extends from objects to morphisms f in $\mathcal{P}$ (that was needed to approximation of Cartesian powers f^N by sequences of reductions between homogeneous spaces).

More generally, let $\{f_i\}$, $i \in I$, be a finite set of reductions between some objects in $\mathcal{P}$. Do they admit homogeneous Bernoulli approximations ϕ_{iN} of all f_i^N, such that

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

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

injectivity (minimality) of $f_{i_\nu}: P \to Q_\nu \Rightarrow$ injectivity (minimality) of $\phi_{i_\nu N}: H_N \to H_{i_\nu N}$?

Probably, this is not always possible but what little we need for our present purposes, is true and does directly follow from the Law of large numbers.

Shannon Inequalities.

Let $Q = \vee_i P_i$ be a (non-canonical) minimal, hence injective, fan $\{Q \to P_i\}$. (This *injectivity* is a category theoretic way of saying that

$set(Q) = \times_i set(P_i)$ as we did earlier.)

Injectivity implies subadditivity:

$$ent(\vee_i P_i) \leq \sum_i ent(P_i),$$

as well as

subadditivity for reductions $f_i : P_i \to Q_i$,

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

Since

$$ent(f_i) = ent(P_i) - ent(Q_i) = ent(Q_i \vee P_i) - ent(Q_i),$$

this amounts to

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

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:

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

The subadditivity of entropy for reduction implies strong subbaditivity:

since $P \vee P = P$ and

$$(P_1 \vee P_2) \vee (P_2 \vee P_3) = P_1 \vee P_2 \vee P_3,$$

the joint reduction of

$f_{12} : P_1 \vee P_2 \to P_2$

and

$f_{23} : P_2 \vee P_3 \to P_2$

is

$$f_{12} \vee f_{23} : P_1 \vee P_2 \vee P_3 \to P_2.$$

Hence,

$$ent(P_1 \vee P_2 \vee P_3) - ent(P_2) \le$$
$$ent(P_1 \vee P_2) - ent(P_2) +$$
$$ent(P_2 \vee P_3) - ent(P_2).$$

⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆ ⋆

3 Representation of infinite probability spaces X by sets of finite partitions of X and Kolmogorov's dynamical entropy .

Spaces over $\mathcal{P}$. A space $\mathcal{X}$ *over* the category $\mathcal{P}$ of finite probability spaces 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 sets of measure zero) $f : X \to P$ 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 defines a space over $\mathcal{P}$.

(Strictly speaking, 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 $\mathcal{X}(P)$.)

$\vee$-*Categories and Measure Spaces.* Recall that an *x-fan over a set of objects* $\{b_i\}$ in a category is a set I of morphisms $f_i : x \to b_i$, $i \in I$, in this category, where an *a*-fan $f'_i : a \to b_i$ is said to *lie between*

x and $\{b_i\}$ and a itself is said to be *between* x *and* b_i, if there is a morphism $g : x \to a$ such that $f'_i \circ g = f_i$ for all $i \in I$.

Definition. An $\mathcal{X}$ over the the category $\mathcal{P}$ of finite measure spaces is called *a measure space* if $\mathcal{P}\backslash^{\mathcal{X}}$ a $\vee$-*category*, that is if every $\mathcal{X}$-fan over finitely many $P_i \in \mathcal{P}$ admits a $Q \in \mathcal{P}$ between $\mathcal{X}$ and $\{P_i\}$.

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

$$\vee_i g_i : \vee_i P_i \to \vee_i P'_i$$

for given reductions $g_i : P_i \to P'_i$. Observe that this $\vee = \vee_{\mathcal{X}}$ 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).

Resolution of Infinite Spaces $\mathcal{X}$. Say that a set $P_\infty = \{P_i\}$ of objects $P_i \in \mathcal{P}$ *resolves* a $Q \in \mathcal{P}^{\backslash \mathcal{X}}$

that lies under $\mathcal{X}$

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

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

("Physically" speaking, observation performed by P_i contain the full infermation about $\mathcal{X}$.)

Example: 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.)

Sub-Examples. (a) A product $\mathcal{X} = \prod_{s\in S} P_s$ is called *minimal* if a Q in $\mathcal{P}^{\setminus\mathcal{X}}$ 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.

(b) The classical *Lebesgue-Kolmogorov product* $X = \prod_{s\in S} P_s$ is also a product in this sense, where the resolution property depends on the following (obvious in the present form)

Lebesgue's density lemma.

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

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

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

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

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

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

ENTROPY SUMMATION FORMULA.

The entropies of reductions between probability spaces, $f : R \to P$, *satisfy*

$$ent(f) = \sum_p |p| \cdot ent(|R_p|^{-1} R_p).$$

(Recall that $|S|^{-1}S$, denote the probability space obtained by normalisation the measure space S, e.g. $S = R_p$.)

This formula is obvious for reductions f between homogeneous spaces and the general case follow by Bernoulli approximation theorem for reductions. Alternatively, one can derive it by a two line computation from Boltzman's formula.

Exercise. Reformulate the definition of Lebesgue's integral in the present terms and reprove its basic properties.

Normalisation and Symmetry.

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

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

To be specific, let Δ_N be finite sets of transformations δ of the space $\mathcal{X}$ represented by self-maps of the (small) category $\mathcal{P}\backslash^{\mathcal{X}}$ (which causes no logical problem in the present case).

These δ act on finite spaces $P \in \mathcal{P}\backslash^{\mathcal{X}}$ and we 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 $|\Delta_N|$ denotes the cardinality of the set Δ_N (and where we pass to a subsequence if the limit does not exist).

Since a single P, and even all of $\bigvee_{\delta \in \Delta_N} \delta(P)$ may not suffice to fully "resolve" $\mathcal{X}$, 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

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

therefore, when "renormalized by size" of Δ_N, the corrspomding $\vee$-products satisfy the same relations by the relative Shannon inequality. Thus,

$$\frac{Ent_{\vee Q_j} - Ent_{P_i}}{|\Delta_N|} \leq \varepsilon_i \underset{i\to\infty}{\to} 0,$$

where

$$Ent_{P_i} = ent\left(\bigvee_{\delta\in\Delta_N}\delta(P_i)\right),$$

and

$$Ent_{\vee Q_j} = ent\left(\bigvee_{\delta\in\Delta_N}(\delta(P_i)\vee\delta(Q_j))\right)$$

Now we see that adjoining $Q_1, Q_2, ..., Q_j$ 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:

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

Kolmogorov's 1958 Theorem for the Bernoulli (Shift) 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 shift action of $\mathbb{Z}$ on it.

If the probability 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, $i = 1, 2, \dots$. 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 and solvable) groups Γ (that satisfy a generalized "$(N+i)/N \to 1$, $N \to \infty$" property) where it also shows that the entropy is *decreasing* under Γ-reductions:

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.

Also recall that a countable group Γ is called *amenable* if it admits an exhaustion by finite subsets Δ_N such that the cardinalities of their group products with all finite subset $\Delta \subset \Gamma$ satisfy

$$\frac{|\Delta_N \cdot \Delta|}{|\Delta_N|} \underset{N\to\infty}{\to} 1,$$

where $A \cdot B = \{a \cdot b\}_{a\in A, b\in B}$.)

Questions. Have the logic of entropy for "Bernoulli (and more general) crystals" $P^{\mathbb{Z}^3}$, been apparent to physicists all along?

Why was it blocked by a blind spot in mathematicians' mind's eye?

What are other blind spots in our minds?

Ultralimits and Sofic Groups. Kolmogorov argument does not apply to non-amenabale groups. In fact entropy is *not* necessarily decreasing under Γ-reductions for non-amenable, e.g. free, groups Γ.

But it was shown by Lewis Bowen a few years ago that

Γ-*isomorphism between Bernoulli systems* $P_1^\Gamma \leftrightarrow P_2^\Gamma$ *implies that* $ent(P_1) = ent(P_2)$ *for many non-amnable groups* Γ, *including, for example, free groups.*

In fact, Bowen proved this for all *sofic groups* Γ.

Question. Is there a category theoretic definition of dynamic entropy that would automatically yield Kolmogorob's and Bowen's entropies?

Definition of "sofic". Sofic groups

Γ can be defined as subgroups of (properly defined) *metric ultra limits* of finite groups.

In simple words, this means the existence of "*ε-approximate actions*" of these Γ on finite sets X that spells out as follows.

Maps $\phi_1, \phi_2 : X \to Y$ are said to *ε-agree* or *agree up-to ε* if the subset of those $x \in X$ where $\phi_1(x) \neq \phi_2(x)$ has cardinality $\leq \varepsilon|X|$ for $|X|$ denoting, as earlier, the cardinality of X.

Then a map A from Γ to the set of self-maps $X \to X$ is called *an ε-monomorphism* on $\Delta \subset \Gamma$ if

• the maps $A(\delta_1 \cdot \delta_2) : X \to X$ and $A(\delta_1) \circ A(\delta_2) : X \to X$ *agree up to ε for all* $\delta_1, \delta_2 \in \Delta$.

• $A(id) : X \to X$, $id \in \Gamma$, *agrees* *up to* ε *with the identity map* $X \to X$.

• The map A is *one-to one on* Δ *up to* ε: that is the cardinality of the set of those $x \in X$ where $A(\delta)(x) = x$ is $\leq \varepsilon|X|$ for all $\delta \in \Delta$ except for $\delta = id \in \Gamma$.

Now, a countable group Γ is *sofic* if,

given a finite set $\Delta \subset \Gamma$ *and an* $\varepsilon > 0$*, there exists a finite set* $X = X(\Delta, \varepsilon)$ *and a map* A *from* Γ *to the set of self maps* $X \to X$ *which is an* ε*-monomorphism on* Δ.

Amenable as well as *residually finite* groups are, obviously, sofic, but the full extend of "sofic" remains unclear.

(Recall that Γ is residually finite if it admits a faithful isometric action on a compact metric space. Notice that all subgroups of linear groups, e.g. free groups, are residually finite.)

Probably, a predominant majority of infinite groups are *non-sofic*, but the existence of a single non-sofic group remains problematic.

4 Simplex of Probability "Laws" and Fisher Metric.

Probability spaces P represented by probability measures, sometimes called "laws", on the same background finite set $I = set(P)$ of cardinality m can be visualised as points in the Euclidean simplex

$$\triangle = \triangle(I) \subset \mathbb{R}^I = \mathbb{R}^m, m = |I|,$$

defined, in these notation, by

$$p_i \geq 0, \ \sum_{i \in I} p_i = 1.$$

In this terms, objects of the category $\mathcal{P}$ of finite probability spaces and reductions becomes simplices $\triangle$ with markings $P \in \triangle$ and morphisms are simplicial maps between sumplices, say $\triangle_1 = \triangle(I_1)$ and $\triangle_2 = \triangle(I_2)$,

$$f : (\triangle_1 \ni P_1) \to (\triangle_2 \ni P_2),$$

such that $f(P_1) = P_2$.

(The category $\mathcal{S}$ of unmarked simplices $\triangle$ and simplical maps can be regarded as a particular *implementation* or *materialisation* of the category $\mathcal{I}$ of finite sets I. This is supposed to say something more than that $\mathcal{S}$ is canonically equivalent to I.)

The simplex $\triangle(I)$ can be reconstructed "combinatorially" from I as

the limit of the $\frac{1}{N}$-rescaled finite spaces that are quotients of I^N by the symmetric group Sym_N.

In fact, the set I^N/Sym_N can be seen as a subset in the the integer lattice $\mathbb{Z}^m \subset \mathbb{R}^m$ that equals the intersection of this lattice with the scaled simplex $N \cdot \triangle(I) \subset \mathbb{R}^m$.

This symmetrisation tremendously diminishes (simplifies?) the power set I^N: its cardinality drops from $M = m^N$ to $M_s = \frac{(m+N-1)!}{m!(N-1)!}$.

On limits of Bernoulli sets $B_N = set(H_N) \subset I^N = (setP)^N$

projected to

$$\triangle(I) = \lim_{N\to\infty} \frac{1}{N}(I^N/Sym_N).$$

Given a sequence of Bernoulli subsets $B_N \subset I^N$ that support Bernoulli homogeneous approximating spaces H_N of P^N, we project these sets to the set I^N/Sym_N embedded into $\mathbb{Z}^m \subset \mathbb{R}^m$, $m = |I|$, and then send them to the simplex $\triangle(I) \subset \mathbb{R}^m$ by $\frac{1}{N}$-scaling.

Let

$$\lim_{N\to\infty} \frac{1}{N} B_N/Sym_N \subset \triangle(I)$$

be the set of the limit points of the subsets B_N/Sym_N so imbedded into $\triangle(I)$ and let $\underline{B}_{sym} = \underline{B}_{sym}(P)$ be the intersection of these limit sets for all Bernoulli sequences B_N approximating P^N.

It follows from the Boltzmann formula, that $\underline{B}_{sym}(P)$ equal the intersection of $\triangle(I)$ with an affine hyperplane $T_P \subset \mathbb{R}^M$ that contains the point $\{p_i\} \in \triangle(I)$ that represent our probability space P and such that this T_P is tangent to the level E_P of the function $\sum_{i\in I} p_i \log p_i$ that contain $\{p_i\}$. Thus, E_P and T_P are defined by the equations

$$\sum_{i\in I} x_i \log x_i = -ent(P)$$

and

$$\sum_{i\in I} x_i \log p_i = -ent(P).$$

(If all $p_i = 1/m$ then $\underline{B}_{sym}(P) = \{1/m, ...1/m\}$ as well.)

Logarithmic Rate Decay Formula.

The symmetrized Cartesian powers P^N/Sym_N of spaces $P = \{p_i\}$, $i \in I$, that are reductions of P^N by the quotient map $I^N \to I^N/Sym_n$, can be similarly seen in $\Delta(I)$.

Namely, let $P_\triangle$ be the measure on $\mathbb{R}^m$, $m = |I|$, supported on the vertex set of $\triangle(I) \subset \mathbb{R}^m$ with masses p_i assigned to these vertices and $P_\triangle{}^{\otimes N}$ be its tensorial power that is the measure on $\mathbb{R}^{mN} = (\mathbb{R}^m)^N$ – that represents P^N.

Let $\sigma : \mathbb{R}^{mN} \to \mathbb{R}^m$ be the *averaging map*:

$$\sigma(x_1, ..., x_N) \mapsto \frac{x_1 + ... + x_N}{N},$$

$$x_1, ..., x_N \in \mathbb{R}^m.$$

and let $P_\blacktriangle{}^N = \sigma_*(P_\triangle{}^{\otimes N})$ be the push forward of the measure $(P_\triangle)^{\otimes N}$ by σ to $\triangle(I) \subset \mathbb{R}^m$.

This $P_\blacktriangle{}^N$, that, in fact, equals the $\frac{1}{N}$-scaled N*-th convolution power* of Δ_P, serves as a representation of P^N/Sym_N in the simplex $\triangle(I)$.

The asymptotics of the values of the measures $P_\blacktriangle^N$ on "infinitesimally small" neighbourhoods $U_{\{q_i\}} \subset \triangle(I)$ of the points $\{q_i\}_{i\in I} \in \triangle(I) \subset \mathbb{R}^m$ encoded by the
(*logarithmic*) *rate* (*decay*) *function*,

$$rate_P : \triangle(I) \to \mathbb{R}_+$$

defined as

$$rate_P\{q_i\} = \inf_{U_{\{q_i\}}} \lim_{N\to\infty} \frac{1}{N} \log P_\blacktriangle^N(U_q),$$

where the infimum is taken over all neighbourhoods of $\{q_i\} \in \triangle(I)$ and where the limit (almost) obviously exists.

This rate, as was shown by Boltzmann, equals

minus relative entropy $ent_P\{q_i\}$,

$$\inf_{U_{\{q_i\}}} \lim_{N\to\infty} \frac{1}{N} \log P^N_{\blacktriangle}(U_q) = -ent_P\{q_i\},$$

where the points

$\{q_i\} \in \triangle(I) \subset \mathbb{R}^I = \mathbb{R}^m$, $m = |I|$,

are regarded as probability measures on the set I that supports the measure $\{p_i\}$ of P when it comes to $ent_P\{q_i\}$.

Verification of this asymptotic formula, (probably, known to Euler) is trivial modulo

Boltzmann's

$$ent_P\{q_i\} = \sum_i q_i \log p_i/q_i$$

and rough *Stirling's*

$$\frac{1}{M} \log M! = \log M - 1 + o(1)$$

applied to the multinomial coefficients $\frac{N!}{\prod_i M_i!}$, $\sum_i M_i = N$.

But everything can be seen directly with our Bernoullian definition of entropy, say for homogeneous $P = \{p_i = 1/m\}$, $m = |I|$, where $ent_P(Q) = ent(Q) - \log m$ is defined via the law of large numbers applied to the function

$$i \mapsto \log q_i$$

on the probability space $Q = \{q_i\}$, that says that one counts only those "states" $(i_1, ..., i_N) \in I^N$ where

$\sum_i q_i \log q_i$ *is close to* $-ent(Q)$.

In fact, the function $-rate$ is often taken for a mathematical definition of entropy corresponding to *"logarithm of the proportion of the number of micro-states"*, where these "micro-states" are those comprising "macro-states" U_q.

Also Boltzmann's formula

$$rate_P\{q_i\} = \sum_i q_i \log q_i/p_i$$

can be seen as an antecedent of *the large deviation theory*, that we briefly explain in section 8.

Cartesian multiplication of probability spaces,

$$(P_1, ..., P_N) \mapsto P_1 \times .... \times P_N,$$

can be seen in the light of the *Segre (tensor product) map*

$$\mathbb{R}^{m_1} \times ... \times \mathbb{R}^{m_k} \times ... \times \mathbb{R}^{m_N} \to \mathbb{R}^M =$$

$$= \mathbb{R}^{m_1} \otimes \mathbb{R}^{m_k} \otimes ... \otimes \mathbb{R}^{m_N}$$

that is restricted to the simplices

$$\triangle_k = \triangle(I_k) \subset \mathbb{R}^{m_k},\ m_k = |I_k|,$$

$$\triangle_1 \times ... \times \triangle_k \times ... \times \triangle_N \to \triangle \subset \mathbb{R}^M,$$

$$\triangle = \triangle(I_1 \times ... \times I_k \times ... \times I_N),$$

$$M = m_1 \cdot m_k \cdot ... \cdot m_N.$$

In the particular case where all factors are mutually equal, the restriction of the Segre map to the diagonal is called *Veronese N-power map*,

$$V_N : \mathbb{R}^m \to \mathbb{R}^{M_s} = (\mathbb{R}^m)^{\circledS N} \subset \mathbb{R}^M = (\mathbb{R}^m)^{\otimes N}$$

where $(\mathbb{R}^m)^{\circledS N}$ denotes the space of *symmetric* N-tensors.

(The dimension

$$M_s = \frac{(m+N-1)!}{m!(N-1)!} < (m+N)^m,$$

is much smaller than $M = m^N$ for large $N >> m$.)

The space $(\mathbb{R}^m)^{\circledS N}$ can be seen as (it is isomorphic to) the space of homogeneous polynomials of degree N in m variables, where the

Veronese map sends polynomials of degree one (linear forms) $\sum_i p_i x_i$ to $(\sum_i p_i x_i)^N$ that are polynomials of degree N.

(One may think these maps are too simple to deserve special names, but if one looks at them closer, one realises that their geometry, e.g. the properties of the metrics $dist_N$ on $\mathbb{R}^m$ induced by these maps, are far from being apparent.)

So far, the role of logarithms in the definition of entropy was purely cosmetic – we were not obliged to use the isomorphism $\log : \mathbb{R}_+^\times \to \mathbb{R}^+$ and could remain in the multiplicative group $\mathbb{R}_+^\times$.

But $ent(P)$ in the present picture, that is seen as a particular function on the simplex $\triangle(I) \subset \mathbb{R}^I$, has some remarkable properties due to its logarithmic origin. Namely

The Fisher Information metric associated to the entropy function, that is the Riemannian metric h on $\triangle(I)$ defined as the

Hessian of minus entropy,

$$h = Hess(e),\ e = e(P) = \sum_{i \in I} p_i \log p_i$$

(unbelivably!) has *constant positive* curvature (where semipositivity of h allows from subadditivity of entropy).

In fact the *Euclidean moment map*

$$M_{\mathbb{R}} : \mathbb{R}^I \to \mathbb{R}^I,$$

for

$$M_{\mathbb{R}} : \{x_i\} \to \{p_i = x_i^2\}$$

is, up to 1/4-factor, an *isometry* from the positive "quadrant" of the unit Euclidean sphere onto $(\triangle(I), h)$.

Proof. Start by observing that the strict convexity of $p \log p$ implies *positive definiteness* of the quadratic form $h = Hess(e)$ that makes it, indeed, a Riemannian metric.

In fact, this metric written in p_i-coordinates in $\mathbb{R}^I$ reads:

$$h = \sum_i \frac{d^2(p_i \log p_i)}{d^2 p_i} dp_i^2 =$$

$$\sum_i dp_i^2 / p_i,$$

for $\frac{d^2(p \log p)}{d^2 p} = 1/p$.

This agrees up to 1/4 factor with the Riemannian metric induced by

$M_{\mathbb{R}}^{-1}$ from the Euclidean metric:

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

QED.

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

$$M : \mathbb{C}^I \to \mathbb{R}_+^I = \mathbb{C}^I/\mathbb{T}^I$$

for

$$M : z_i \to z_i\bar{z}_i$$

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 $\mathbb{C}P(I) \to \triangle(I)$.)

Fisher Metric via Information Divergence. The information divergence $\delta = \delta(Q, P) = -ent_Q(P)$

between $P = \{p_i\}$ and $Q = \{q_i\}$ is a *smooth* function on $\triangle(I)\times\triangle(I)$. Since this function is *non-positive* and it vanishes on the diagonal, the differential of this d vanishes on the diagonal as well.

Hence, the second differential of δ is defined on the diagonal, where it can be seen as a family of, necessarily positive semidefinite, quadratic forms $D_P^2(\delta(Q,P))$ on the tangent spaces $T_q(\triangle(I))$, that is a quadratic differential form, say h_δ on $\triangle(I)$, and since

$$-ent_Q(P) = \sum_i p_i(\log p_i - \log q_i)$$

by Boltzmann formula, this δ equal the Fisher metric $h = Hess(e)$ for $e = \sum p_1 \log p_i$.

The derivation of Fisher metric

via the second differential $D^2(\delta)$ has the advantage of relying only on the *smooth structure* in the simplex $\triangle(I)$, while the definition of $Hess(e)$ depends on the *affine structure* in $\triangle(I)$.

Fisher Metric (almost) without Logarithms. The space of measures on the simplex $\triangle(I)$ is naturally acted upon by the multiplicative group $\mathbb{R}^{\times}_{>0}$.

Pick up an invariant Riemannian metric g on $\mathbb{R}^{\times}_{>0}$, and thus define Riemannian metrics, that are positive differential quadratic forms, in the spaces of positive weights $p_i \in \mathbb{R}_{>0}$, $i \in I$, on the vertices of $\triangle(I)$.

Denote these forms by $h_i = h_{g,i}$ and observe that the weighted sum

$h_g = \sum_i p_i h_i$ is a Riemannian metric on the simplex $\triangle(I)$ where this $\triangle(I)$ is regarded as the space of I-tuples of positive weights $\{p_i\}$.

Since, obviously, $g = C_g dp^2/p$ for the standard additive p-parametrization of the positive real line $R_{>0}$, the form h_g equals Fisher's $h = \sum_i dp_i^2/p_i$ times a constant C_g, where it is also clear that $C_g = 1$ if g is the Riemannian metric for which $dist_g(r_1, r_2) = |\log r_1 - \log r_2|$.

Remark. The above construction (going back to Fisher? to Rao?) delivers Rinannian metrics in a variety of *(moduli) spaces of geometric structures*" that contain "measures" among their "constituents".

The two prominent examples are *spaces of symplectic structures* (forms) on a smooth manifolds and spaces of Riemannian metrics/forms, where the latter for metrics of constant negative curvature on surfaces is called the *Weil-Petersson metric.*

Question. Is there an algebra-geometric setting where the entropy and the classical probability theory in general, would find there proper place along with the moment map from $\mathbb{C}P^n$ to the n-simplex?

Reference. Frédéric Barbaresco, *Koszul Information Geometry and Souriau Geometric Temperature/Capacity of Lie Group Thermodynamics.*

****

5 Mendelian Next Generation Map.

In 1908, Godfrey Harold Hardy sent a *"letter to the editor of Science"*, entitled:

"Mendelian proportions in a mixed population", where he writes:

"...suppose that ... mating may be regarded as random, that the sexes are evenly distributed among the three varieties, and that all are equally fertile. A little mathematics of the multiplication-table type is enough to show that in the next generation the numbers will be as $(p+q)^2 : 2(p+q)(q+r) : (q+r)^2$*, or as* $p_1 : 2q_1 : r_1$*.*

The interesting question is in what circumstances will this distribution be the same as that in the genera-

tion before? It is easy to see that the condition for this is $q^2 = pr$. *And since* $q_1^2 = p_1 r_1$, *whatever the values of p, q, and r may be, the distribution will in any case continue unchanged after the second generation.*"

(This nine lines, unknowingly to Hardy, firmly associated his name with *genetics* – one of the greatest scientific discoveries of all times, that has surpassed everything he has achieved in pure mathematics.)

A single formula expressing what Hardy says, called

Castle-Hardy-Weinberg Law,
reads:

$$\frac{\left(A^2 + AB\right)^2}{\left(A^2 + AB\right) \cdot \left(AB + B^2\right)} = \frac{A^2}{AB}$$

for $A = p + q$ and $B = q + r$.

This "multiplication-table type" identity expresses *idempotence*, customary called

equilibrium property,

(of an instance) of the Mendelian *next generation* map G that, in the simple case considered by Hardy, sends the projective plane into itself:

$$p : 2q : r \overset{G}{\mapsto} p_1 : 2q_1 : r_1,$$

for

$$\begin{aligned} p_1 &= (p + q)^2 \\ q_1 &= (p + q)(r + q) \\ r_1 &= (q + r)^2, \end{aligned}$$

where the Castle-Hardy-Weinberg law, often called *equilibrium principle*, (undobtfully known to Mendel) ascertains that

$$G \circ G = G.$$

(Such a relation is unusual, almost paradoxical, for *polynomial and rational* transformations G. This contributed to befuddlement of biologists faced with counterintuitive and ideologically unacceptable for many of them predictions of Mendelian inheritance theory.)

In biological terms, the map G acts on *allele* distributions in populations O of *diploid* organisms. If these alleles, that are variants of a particular gene present in a population, are market by indices $i \in I$

(in Hardy's case there are two alleles), then the set of these distributions can be represented by "vectors" $\{n_i\}$ for n_i being the number of i-th alleles present in O.

More realistically, one deals with ratios $n_i : n_j$, since the numbers n_i themselves are usually unknown. Thus the space of distribution lies in the projective space

$$P^{m-1} = P^I = (\mathbb{R}^I \smallsetminus 0)/\mathbb{R}^\times, \ m = |I|.$$

Every *diploid* organism $o \in O$ contains *two* sets of genes, and the distribution of occurrences of pairs (i, j) of alleles can be written as a matrix $\{r_{ij}\}$, $i, j \in I$. (This matrix is often symmetric, e.g. it is so in the Hardy case, but this not essential at the moment.)

The next generation map G associated with *"random matings"* between members of the population acts on matrices $R = \{r_{ij}\}$ by substituting each (i,j)-entry r_{ij} by the product (*Segre embedding*) of the sums of the entries in the i-row and the j-column:

$$\{r_{ij}\} \overset{G}{\mapsto} \{r_{ij}^{chld}\}$$

for

$$\{r_{ij}^{chld}\} = \sum_i r_{ij} \cdot \sum_j r_{ij},$$

where "*child*" is for "*children*".

Then, by pure algebra (see below), the distribution of genotypes (at a single locus)

achieves equilibrium

after the *first* round of reproduction, i.e. the projectivized map G is *idempotent*:

$$G(G(R)) = const \cdot G(R)$$

for

$$const = \sum_{ij} r_{ij}.$$

It is also clear that the image of G consists of *matrices of rank one* and, when restricted to *symmetric matrices*, this G *retracts* quadratic polynomials onto *the Veronese variety* (we return to "Veronese" in the next section) where, observe the fibers of these retraction are *affine* subspaces in the space of polynomials.

Also G preserves the hyperplane where $\sum_{ij} r_{ij} = 1$ and, in the case of real entries, the positivity of r_{ij} is also preserved.

Then, non-surprisingly, G is *entropy increasing*:

$$ent(G(R)) \geq ent(R),$$

where – this is immediate with Boltzmann's formula,

$$[ent(G(R)) = ent(R)] \Leftrightarrow$$
$$\Leftrightarrow [G(R) = R].$$

Questions. Is there a purely "Mendelian" proof of $ent(G(R)) \geq ent(R)$ and/or of "$\Rightarrow$"?

Let us describe the above in more general invariant terms.

$\langle\rangle$-Spaces.

These are linear spaces A with distinguished non-zero linear functions $a \mapsto \langle a \rangle$ on them.

Examples. (a) Spaces of distributions $\{r_i\}$ are endowed with $\langle\{r_i\}\rangle =$

$\sum_i r_i$.

(b) Linear operators with finite ranks have their *traces* for $\langle\rangle$ and quadratic forms on *Hilbert spaces also have traces for* $\langle\rangle$. (This suggests a non-commutative counterpart to Mendelian formalism.)

(c) Homologies of topological spaces X with coefficients in a filed come with distinguished zero dimensional cohomology classes that represent constant functions (0-cochains) on X, that serve as $\langle o\rangle$ on these homologies.

Tensor products $A \bigotimes B$ of $\langle\rangle$-spaces come with $\langle\rangle$-structures. for

$$\langle a \otimes b\rangle = \langle a\rangle \cdot \langle b\rangle.$$

Besides, there are $\langle\rangle$-natural linear maps from $C = A \bigotimes B$ to its tensorial components:

$$E_A : C \to A \text{ and } E_B : C \to B$$

that are the linear extensions of the (bilinear) maps

$$a \otimes b \mapsto \langle b \rangle a \text{ and } a \otimes b \mapsto \langle a \rangle b$$

In these terms the "next generation map" G from C to itself is expressed as

$$G(c) = E_A(c) \otimes E_B(c),$$

where the equilibrium property reads:

$$G \circ G(c) = \langle c \rangle G(c).$$

Proof. Clearly, $c' = G(c)$ is a monomial, say $c' = a' \bigotimes b'$ that is sent by G to

$$\langle b' \rangle a' \bigotimes \langle a' \rangle b' = \langle b' \rangle \langle a' \rangle a' \bigotimes b',$$

where $\langle b' \rangle \langle a' \rangle = \langle c' = G(c) \rangle = \langle c \rangle^2$ since G is a quadratic $\langle\rangle$ respecting map. QED

(This is what becomes of Hardy's "multiplication table" in the linear algebraic language.)

The above generalises to multiple tensor products of $\langle\rangle$-spaces, $\bigotimes_{l \in L} A_l$ for an arbitrary finite set L.

Such a product can be seen as a subspace in *the polynomial algebra* $A^* = A^*(X)$ on the Euclidian space X that is the sum (Cartesian product) $\bigoplus_{l \in L} X_l$ of the linear spaces X_l dual to A_l: the product $\bigotimes_{l \in L} A_l$ is identified with the set of homogeneous polynomials of degree 1 in each x_l-variable where $\langle a \rangle$

is represented by the value $a(x_0)$ at some vector $x_0 \in X$.

For instance, $x_0 = \mathbf{1} = (1, 1, ..., 1)$ for distribution spaces. But since one can go from one vector to another by a parallel translation of X and translations induce automorphisms of the algebra $A^*(X)$, the choice of x_0 makes no difference and we stick to $x_0 = 0$ in X. (This suppresses multinomial formulae common in traditional expositions of G.)

Given a subset $K \subset L$ we associate to it the coordinate projection P_K from X to the coordinate plane $X_K = \bigoplus_{l \in K} X_l \subset X$ and denote by $E_K = P_K^*$ the induced endomorphism of the algebra A^*, where $E_K(a(x_l))$ is obtained by equating all x_l in a with $l \in L - K$ to zero.

Since P_K are *commuting idempotents* so are E_K for all $K \subset L$, where E associated to the empty set sends A^* to the constants. Given a collection $\mathcal{K}$ of subsets $K \subset L$ we define $E_{\mathcal{K}}$ as the (polynomial) product of E_K for all $K \in \mathcal{K}$, i.e. $E_{\mathcal{K}}(a) = \prod_{K \in \mathcal{K}} E_K(a)$.

Since the multiplicative semigroup of polynomials is commutative and the maps E_K are endomorphisms, the transformations $E_{\mathcal{K}}$ are *multiplicative* endomorphisms of A^* (but not additive ones for more than one K in $\mathcal{K}$). Since all E_K commute, so do $E_{\mathcal{K}}$ and the composition of $E_{\mathcal{K}}$'s is expressible in terms of intersections of the underlying subsets $K \subset L$ by the simple rule:

$$E_{\mathcal{K}} \circ E_{\mathcal{K}'} = \prod_{K \in \mathcal{K}, K' \in \mathcal{K}'} E_{K \cap K'}$$

that follows from the similar rule for the composition of the maps P_K's.

Equilibrating Maps. If $\mathcal{K}$ is made of d *non-intersecting* non-empty subsets, e.g. $\mathcal{K}$ is a partition of L into d subsets, then $E = E_{\mathcal{K}}$ (that correspond to our old G, is called an *equilibrating* map of *degree d*. Equilibrating maps obviously satisfy:

(**A**) *Quasi-idempotence.*

$$E \circ E(a) = a(0)^{d^2 - d} E(a)$$

where d denotes the degree of E and where the exponent $d^2 - d$ corresponds to the presence of $d^2 - d$ *empty* intersections between different subsets $K_1, ..., K_d$ in L underlying E.

(**B**) *Polynomiality.* Equilibrating maps preserves subspaces $A^{\leq k} \subset A^*$ of polynomials of degree $\leq k$ in each variable. Thus A^* is representable as a union of *finite dimensional E-invariant* subspaces and if $\mathcal{K}$ is made of d subsets $K \subset L$ then the corresponding equilibrating map E is

a polynomial map of degree d

on each linear space $A^{\leq k}$.

(**C**) *Linearizability.* Let us regard $A^{\leq k}$ as the algebra of

k-truncated polynomials

that is *a quotient* of (rather than a subspace in) A^* obtained by adding the relations $x_l^{k+1} = 0$ to A^*.

Then the maps $E_{\mathcal{K}}$ (not only equilibrating ones) act on this algebra as multiplicative endomorphisms; they can be "simultaneously linearized" with the exponential map,

$$exp(a) = 1 + a + \frac{1}{2}a^2 + \frac{1}{6}a^3 + ...,$$

that isomorphically maps the *additive group* of k-truncated polynomials to the *multiplicative group* of k-truncated polynomials satisfying $a(0) > 0$.

(**D**) *Retraction to Veronese.* It follows from (A) (and also from (C)) that each equilibrating map $E = E_{\mathcal{K}}, \mathcal{K} = (K_1, K_2, ..., K_d)$, *retracts* the *normalizing hyperplane*

$$A^\times = A^\times(X) \subset A^*$$

defined by $a(0) = 1$ to the *Segre-*

Veronese product set

$V = V_E = E(A^\times) = A_1^\times \cdot A_2^\times \cdot \ldots \cdot A_d^\times \subset A^\times$

for $A_i^\times = A^\times(X_{K_i})$, that is the set of products of d polynomials $a_i \in A_i^\times$, where composition of E's corresponds to intersection of V's:

$$V_{E \circ E'} = V_E \cap V_{E'}.$$

The fibers $E^{-1}(v) \subset A^\times$ are *affine* subspaces: they are, obviously, equal the fibers of the additive counterpart to $E = E_{\mathcal{K}}$, that is $E_{K_1} + \ldots + E_{K_d}$, where $K_i \subset L$ are the constituents of $\mathcal{K} = (K_1, \ldots, K_d)$.

(**E**) *$\mathcal{G}$-equivariance.* The equilibrating maps E *commute* with the group $\mathcal{G}$ of linear transformations of X preserving the decomposition $X = \bigoplus_{l \in L} X_l$ that naturally act on polynomials. (For example, the Veronese varieties are $\mathcal{G}$-invariant.) In particular, All E commute with the *scaling transformation* Λ corresponding to $x \mapsto \lambda x$ in X which fixes constant polynomials, e.g. $1 \in A^{\times}$, and has other eigenvalues equal $\lambda, \lambda^2, \lambda^3$, etc. Thus, for $\lambda > 1$, the transformation Λ *expands* $A^{\times}$ with the fixed point 1 and so *global* properties of maps commuting with Λ, e.g. of equilibrating maps and linear combinations of these, can be derived from the corresponding *local* ones at the fixed point 1 of Λ by transporting all points close to 1 by applying Λ^{-N} with large $N \to \infty$.

Remark on Λ*-equivariant maps.* Let A be a linear space (e.g. $A^{\times} \bigcap A^{\leq k}$ with the constant polynomial 1 taken for the origin) with a linear transformation Λ, where A splits into n eigenspaces of Λ with the corresponding eigenvalues $\lambda, \lambda^2, ..., \lambda^n$, where λ is *not* a root of unity, e.g. $\lambda > 1$. It is easy to see that every smooth transformation F of A commuting with Λ is a polynomial map of degree at most n; a transformation F is invertible (necessarily by a *polynomial* transformation) if and only if its differential $D_0(F)$ at 0 is invertible; transformations F with $D_0(F) = 1$ make a nilpotent Lie group. For example, *all* iterates F^j are polynomials of degrees bounded by *the same* n that, non-surprisingly, admit explicit (albeit complicated) expression in terms of n.

Observe that neither topology, nor positivity, were entering our account of Mendelian dynamics. But all this comes back with the following

Robbins-Geiringer Convergence Property.

Let $E_1, ..., E_m : A^* \to A^*$ be equilibrating maps defined on the polynomial algebra A^* on a linear space X. These E_i, as we know, retract the normalising hyperplane $A^\times \subset A^*$ to the respective Veronese varieties

$$E_i \to V_i = V_{E_i}.$$

Let

$$V = V_E = E(A^\times) = \bigcap_i V_i$$

be the Veronese variety

$$V = E(A^\times) = \bigcap_i V_i$$

of the composition

$$E_\times = E_1 \circ E_2 \circ \ldots \circ E_m$$

Let $F = c_1 E_1 + c_2 E_2 + \ldots + c_m E_m$ be a *convex combination* with *strictly positive* coefficients $c_1, c_2, \ldots, c_m$.

Then the iterates $F^1 = F$, $F^2 = F \circ F^1, \ldots, F^j = F \circ F^{j-1}, \ldots$ *on* $A^\times$ *converge to the above equilibrating map* $E_\times : A^\times \longrightarrow V \subset A^\times$, *where the convergence is uniform and exponentially fast on the compact subsets in* $A^\times \bigcap A^{\leq k}$ *for all* $k = 1, 2, \ldots$.

Proof. Since $c_1 + c_2 + \ldots + c_n = 1$ and since all E_i fix the Veronese variety $V = E_\times(A^\times) = \bigcap_i V_i$ of the composition $E_\times = E_1 \circ E_2 \circ \ldots \circ E_m$, so does F and for the same reason F sends each (affine!) fiber $E_\times^{-1}(v)$ into itself.

The differentials D_1 of E_i on $A^\times$ at 1 have all their eigenvalues≤ 1 where the equalities are achieved on the vectors tangent to the corresponding Veronese varieties $V_i = E_i(A^\times)$, because E_i are smooth *retractions* to V_i (and where the eigenvalues equal 0 tangentially to their respective fibers).

The differential of F equals the convex combination of those of E_i; if we assume all $c_i > 0$, we conclude that all eigenvalues of the differential $D(F)$ on V on the tangent vectors transversal to V are< 1, since the tangent space to V equals the intersection of those to V_i. (Tangentially to V the eigenvalues of $D(F)$ equal 1 since V is fixed under F.) In other words, the differential $D(F)$ *strictly contracts* the tangent vectors at V that are transversal to V. It follows that F also contracts some neighborhood $U \subset A^{\times}$ of V; therefore, each point $v \in U$ exponentially fast approaches V under iterates of F. In fact, the F-orbit of v converges to $E_{\times}(v) \in V$ since F preserves the fibers of E.

This local property obviously globalizes with the expanding transformation Λ from the above (**E**) and impilies the convergence of F^j. QED.

Crossover and Recombination. The above convergence property can be applied to distributions of genomes of a population where this genomes undergo *recombination by crossover* under random matings, where it is called

Robbins-Geiringer Asymptotic Equilibrium Theorem.

The gamete probability distributions $a(X_i)$ *of the genomes of populations* X_0, X_1, X_2,..., X_i,... *resulting from consecutive rounds of random matings and recombinations, converge to equilibrium. for* $i \to \infty$.

This is explained in detail in *Mendelian Mendelian Dynamics and Sturtevant's Paradigm*" that can be found on my web page.

On Entropy. This theorem can be also seen by observing that each round of random mating increases the entropy of the distributions of alleles by a definite amount.

The above mathematical scheme, provides a highly idealised outline of statistics of real life genomes, the true complexity of which is still far from being fully understood or even adequately formalised. (See *Logic of Chance* by Eugene V. Koonin for the present state of art.)

But, amazingly, using this kind of mathematics, Mendel, unprecedentedly in science, extracted *non-trivial structural information*, such as *disctretnes* of genomic information and *diploidy* of certain organisms, from raw statistical data.

⊛⊠ ⊛⊠ ⊛⊠ ⊛⊠ ⊛⊠ ⊛⊠ ⊛⊠ ⊛⊠

6 Von Neumann's "Hilbertization" and Strong Subadditivity of Entropy.

The orthogonal symmetry of the Fisher metric disclosed by the moment map suggests an orthogonally invariant extension of entropy.

Start with an orthogonally invariant counterpart of the "simplex probability measures".

This simplex

$$\triangle(I) \subset \mathbb{R}^I = \bigtimes_{i\in I} \mathbb{R}$$

equals the convex hull of the set I realised in $\mathbb{R}^I$ by an orbit of the permutation group $Sym(I)$ that acts on I.

Replace I by the real projective space P^{n-1} acted upon by the orthogonal group $O(n)$, that is "*maximally $O(n)$-homogeneous*" space in the same sense as I is "*maximally Sym(I)-homogeneous*".

This space P^{n-1} embeds to the Euclidean space $\mathbb{R}^{\frac{(n+1)(n+2)}{2}}$ via the (*Segre*)-*Veronese squaring map* from the linear space $L = \mathbb{R}^n$ to the symmetric tensorial square of L.

$$V : L \to L^{\circledS 2} = \mathbb{R}^{\frac{(n+1)(n+2)}{2}}.$$

If L is realised by linear functions l on the dual space $L^\perp (= \mathbb{R}^n)$, then $L^{\circledS 2}$ identifies with the space of quadratic functions (forms) on this $L^\perp$ and V acts by the ordinary squaring of functions, $l \overset{V}{\mapsto} l^2$.

Since the squaring map is symmetric under $l \leftrightarrow -l$, its restriction to the unit sphere $S^{n-1} \subset \mathbb{R}^n = L$ factors via a map from the projective space $P^{n-1} = S^{n-1}/\pm$

$$P^{n-1} \to L^{\circledS 2}$$

where the image equals an orbit of the orthogonal group $O(n)$ naturally acting on

$$L^{\circledS 2} = (\mathbb{R}^n)^{\circledS 2} = \mathbb{R}^{\frac{(n+1)(n+2)}{2}}.$$

In terms of quadratic forms, this image, called *Veronese variety*

$$P_{Ver} = P_{Ver}^{n-1} \subset L^{\circledS 2} = \mathbb{R}^{\frac{(n+1)(n+2)}{2}}$$

consists of quadratic forms l on the Euclidean space $\mathbb{R}^n$, such that

$rank(l) = 1$ and $trace(l) = 1$, where the latter equality shows that P_{Ver} is contained in an affine hyperplane $H \subset L^{\circledS 2}$.

Observe that H is invariant under the action of the orthogonal group $O(n)$ with the centre of mass $o \in H$

of P_{Ver} being fixed. This o represent the normalised background Euclidean form, that is

$$o = \frac{1}{n} \sum_{i=1,\ldots n} x_i^2.$$

Example. The 1-dimensional Veronese is the ordinary circle in the plane but the 2-dimensional Veronese variety P^2_{Ver} that is situated in $H = \mathbb{R}^5$ is harder to visualise.

Density States. Define *states* P on a Euclidean space S as positive semidefinite quadratic forms on S, where *density* (*probability*) *states* P are distinguished by the condition $trace(P) = 1$.

Every density state equals the convex combination of *pure* (*Veronese*) *states* that have ranks equal one.

Every pure state is vanishes on a hyperplane in S; we say it is *supported* on the line normal to this hyperplane in S, where the (projective) space of these lines represents the Veronese variety.

If $S = \mathbb{R}^I$ for a finite set I, then pure states supported on the coordinate lines, that are $p_i x_i^2$, $i \in I$, correspond to atomic measures of weight p_i supported on $i \in I$, while general measures $\{p_i\}$ correspond to the states $\sum_{i \in I} p_i x_i^2$.

Thus, our states provide an orthogonally invariant extension of the concept of a finite measure space.

"Unitary" Remark. Quantum mechanics, lives in the world of *complex*, rather than real Hilbert spaces. But this is non-essential for what

we do here.

States P can be seen as a measure-like functions on linear 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, where the result does not depend on the basis by *the Pythagorean theorem*. (Without this theorem neither Hilbert spaces nor Quantum mechanics would be possible.) The total mass of P is denoted

$$|P| = P(S) = trace(P).$$

Observe that

$$P(T_1 \oplus T_2) = P(T_1) + P(T_2)$$

for *mutually orthogonal* subspaces T_1 and T_2 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

$$P(T_1 \otimes T_2) = P_1(T_1) \cdot P_2(T_2)$$

for all $T_1 \subset S_1$ and $T_2 \subset S_2$.

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 background Hilbert/Euclidean form of S, but the states that are homogeneous on their *supports* that are, by definition, the orthogonal complements of the *null-spaces* $0(P) \subset S$, constitute the respectable space of *all* linear subspaces in S.

Von Neumann Entropy. There are several equivalent definitions of this entropy $ent(P)$.

1. *"Minimalistic" Definition.* States P evaluated on vectors from finite or countable subsets $\Sigma \subset S$ define measures on these Σ denoted $\underline{P}|\Sigma$.

Then define

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

for $\Sigma \subset S$ running over all *full orthonormal frames* in S.

(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 the space of frames 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, as we know, is a concave function on the simplex $\triangle(\Sigma) \subset \mathbb{R}^{\Sigma}$ of probability measures on the set Σ and infima of familes of concave functions are concave.

2. *"Spectral" Entropy.*

The entropy of P was defined by von Neumann as the classical entropy of the *spectral measure* of P.

That is $ent(P)$ equals $ent(\underline{P}|\Sigma)$ for a frame $\Sigma = \{s_i\}$ that *diagonalizes* the quadratic 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.

Concavity of entropy is non-obvious with this definition – it was proven in 1968 by Lanford and Robinson, but is clear that

the spectrally defined entropy is additive under tensor products of *states:*

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

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. *ent_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.)

$$\underline{P}(\Sigma) = (\underline{p}_1, ..., \underline{p}_n),\ \underline{p}_i = P(s_i),$$

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

Symmetrization by Averaging.

Groups G that lineally act on S also naturally act on states P that are after all are quadratic forms on S. A priori, such an action may change the mass (trace) of P, but this mass $|P|$, is, obviously, kept unchanged for *isometric* actions that preserve the background Euclidean/Hilbert form in S.

If G is compact, we average states P acted by G over G, by integrating over G, where the resulting averaged state is denoted $G * P$,

$$G * P = \int_G g(P)d(g),$$

for dg being the normalised (i.e. probability) Haar measure on G. The state $G{*}P$ is, clearly, G-invariant,

$$G * (G * (P)) = P$$

and $|G * P| = |P|$ for isomeric actions of G.

Symmetrization by averaging is also possible for finite probability spaces $P = \{p_i\}$, $i \in I$, with finite groups G acting on the set $I = set(P)$.

For instance, let $f : I \to J$ be a surjective map and let G act on I, such that this action preserves the fibbers $I_j = f^{-1}(j) \subset I$, $j \in J$, and is transitively on every one of these fibers.

Then the averaged measure space $G * P$ on I faithfully represents the f-reduction $Q = \{q_j\}$ of P, since the $G * P$-masses of all atoms in all I_j, $j \in J$, equal $q_j/|I_j|$

for $q_j = \sum_{i \in I_j} p_i$ and $|I_j| = card(I_j)$.

Consequently,

$$ent(G*P) = ent(Q) + \sum_{j \in J} q_j \log |I_j|$$

that reduces to

$$ent(G * P) = ent(Q) + \log m$$

if all fibers I_j have equal cardinalities m.

Canonical Reductions.

Let $S = S_1 \otimes S_2$, and let $G = G_1$ be an isometry group of S_1 that naturally acts on $S_1 \otimes S_2$.

If the linear isometric action of G on S_1 is *irreducible*, or (obviously) equivalently, it admits *no invariant state* except for scalar multiples of the background Euclidean/Hilbert state/form, then there is a one-to-one correspondence between G_1-invariant states Q on S and states P_2 on S_2.

In fact, this correspondence is defined by the condition

$$Q = P_{\langle 1 \rangle} \otimes P_2$$

where $P_{\langle 1 \rangle}$ denotes the normalised background Euclidean/Hilbert state/form on S_1; if Q is a density state then the normalising factor equals $1/dim(S_1)$ that makes $P_{\langle 1 \rangle}$ a density state.

(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.)

Strong Subadditivity of von Neumann Entropy.

Theorem: *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

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

This was conjectured in 1968 by

Lanford and Robinson who established simple subadditivity

$ent(P_{12}) \leq ent(P_1) + ent(P_2).$

as a (simple) corollary of concavity of the von Neumann entropy.

The strong subadditivity was proven by Lieb and Ruskai in 1973 with *operator convexity techniques.*

Strong Subadditivity follows from the following

Double Average Inequaity. Let H and G be compact groups of isometric linear transformations of a Euclidean space S and let P be a density state (positive semidefinite quadratic form with trace one) on S.

If the actions of H and G commute, then the von Neumann en-

tropies of the G- and H-averages of P satisfy

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

This inequality (trivially) implies strong subadditivity when it is applied to the actions of the orthogonal (full isometry) groups $H = O(S_1)$ and $G = O(S_2)$ on $S = S_1 \otimes S_2 \otimes S_3$.

(The double average inequality may look significantly more general than strong subadditivity but the former follows from the latter by a simple argument.)

Non-Standard Proof of the double average inequality.

The relative Shannon inequality

(that is not fully trivial) for *measures* reduces by our argument (that goes back to Boltzmann-Gibbs and Bernoulli) to a *trivial* intersection property of *subsets in a finite set.* Let us explain how this woks for the von Neumann entropy.

Recall that 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) quadratic form P vanishes. We denote this support by $0^{\perp}(P)$ and write $rank(P)$ for $dim(0^{\perp}(P))$.

Observe that

$$(\Leftrightarrow) \quad P(T) = |P| \Leftrightarrow T \supset 0^{\perp}(P)$$

for all linear subspaces $T \subset S$.

A state P is called *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 the satisfy what we call

Localization Inequality:

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

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.

Corollary. The double average inequality 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.

Proof. The double average inequality 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 the proof of desired inequality is reduces 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 localisation inequality, 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 the general double average inequality for general density states.

Since tensorial powers $P^{\otimes N}$ of all states P "converge" to "ideal sub-homogeneous states" $P^{\otimes\infty}$ by Bernoulli's theorem, the above argument, applied to these "ideal states" $P^{\otimes\infty}$, yields the desired inequality for all P, where "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.

I must confess at this point that a suitable formalism that would make the above 100% rigorous has not been worked out and one needs to check all "non-standard" points one by one. This takes a couple of pages instead of, as it should be, a couple of lines.

(These "two pages" can be found

in *"In a Search for a Structure, Part 1: On Entropy"*, where they are written in a "semi-non-standard" language of *Bernoulli sequences* of density states.)

Question. Is "non-standard Euclidean geometry" worth pursuing?

Some arguments in favour of this are presented in my article *"... On Entropy"* but new results and/or specific questions are badly needed.

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 quadratic 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 quadratic 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

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

On Algebraic Inequalities. Besides "hilbertization" some Shannon-like inequalities admit linearization, where the first non-trivial instance of this is the *linearized Loomis-Whitney isoperimetric inequality* that we have met earlier.

More generally, let X_i, $i \in I$, be vector spaces over some field and denote by X_J, $J \subset I$, the tensor product of X_i over J, i.e.

$$X_J = \otimes_{i \in J} X_i.$$

Define the *J-reduction* $Y_J \subset X_J$ of a linear subspace $Y_I \subset X_I$ as the minimal subspace in X_J, such that $Y_J \otimes X_{I \smallsetminus J}$ contains Y_I.

To see it better, take the tensor product $Z_I = X_I \otimes Y'$ for some linear space Y' take a vector $z = Z_I$ and let

$$z_J : X'_{I \smallsetminus J} \otimes Y' \to X_J$$

be the homomorphism corresponding to z under the canonical isomorphism

$$X_I \otimes Y' = Hom(X'_{I \smallsetminus J} \otimes Y', X_J),$$

where $X'_{I \smallsetminus J}$ denotes the linear dual of $X_{I \smallsetminus J}$.

Then the J-reduction of the image $Y \subset X_I$ of z_I equals the image of z_J in X_J; thus, $rank(Y_J) = rank(z_J)$.

Tensorial Reduction Inequality. *The ranks of the J-reductions of every linear subspace $Y \subset X_I = \otimes_{i \in I} X_i$ satisfy*

$$\prod_{J \subset I} (rank(Y_J))^{\alpha_i} \geq rank(Y)$$

for an arbitrary *partition of unity* $\{\alpha_J\}$ on I, that is an assignment of a non-negative number α_J to each

$J \subset I$, such that $\sum \alpha_J \chi_J = 1$, where $\chi_J : I \to \{0,1\} \subset \mathbb{R}$ denote the characteristic (indicator) functions of the subsets $J \subset I$.

This algebraic inequality easily reduces to the corresponding combinatorial inequality known as *Shearer Lemma*, that bounds the cardinalities of finite subsets $S \subset R^I$ in terms of its projections $S_J \subset R^J$ by

$$\prod_{J \subset I} |S_J|^{\alpha_i} \geq |S|,$$

where this combinatorial inequality (or rather a refined entropic version of it) follows via the *inclusio-exclusion principle* from *strong subadditivity of entropy* similarly the *Loomis-Whitney isoperimetric inequality* that corresponds to the case where

$\alpha(J) = 1/(|I| - 1)$ if $|J| = |I| - 1$ and $\alpha(J) = 0$ otherwise.

(See my *"Entropy and Isoperimetry for Linear and non-Linear Group Actions"*.)

Question. What is the full range of such inequalities?

7 Measures Defined via Cohomology with Applications to the Morse Spectra and Parametric Packing Problem.

Entropy serves for the study of "ensembles" $\mathcal{A} = \mathcal{A}(X)$ of (finitely or infinitely many) particles in a space X, e.g. in the Euclidean 3-space by

$U \mapsto ent_U(\mathcal{A}) = ent(\mathcal{A}_{|U}),\ U \subset X,$

that assigns the *entropies* of *the U-reductions $\mathcal{A}_{|U}$ of $\mathcal{A}$*, to all bounded

open subsets $U \subset X$. In the physicists' parlance, this entropy is

"the logarithm of the number of the states of $\mathcal{E}$

that are effectively observable from U",

We want to replace *"effectively observable number of states"* by

"the number of effective degrees of freedom of ensembles of moving particles".

Let us prepare the topological language for expressing this idea.

Graded Ranks, Poincare Polynomials and Ideal Valued Measures.

The images as well as kernels of (co)homology homomorphisms that are induced by continuous maps are *graded* Abelian groups and their ranks are properly represented not by individual numbers but by *Poincaré polynomials*, $\sum_i rank_i t^i$.

The set function $U \mapsto \mathrm{Poinc}_U$ that assigns Poincaré polynomials to subsets $U \subset A$, (e.g. $U = A_r$) has some measure-like properties that become more pronounced for the set function

$$A \supset U \mapsto \mu^*(U) \subset H^*(A;\Pi),$$

$$\mu^*(U) = Ker\big(H^*(A;\Pi) \to H^*(A \smallsetminus U;\Pi)\big),$$

where Π is an Abelian (homology coefficient) group.

By elementary topology,

$\mu^*(U)$ is **additive** *for the sum-of-subsets in $H^*(A;\Pi)$ and* **super-multiplicative** *for the the $\smile$-product of ideals in the case Π is a commutative ring*:

$$\mu^*(U_1 \cup U_2) = \mu(U_i) + \mu^*(U_2)$$

for *disjoint* open subsets U_1 and U_2 in A, and

$$\mu^*(U_1 \cap U_2) \supset \mu^*(U_1) \smile \mu^*(U_2)$$

for all open $U_1, U_2 \subset A$

"Θ-Measures" $|\mu_\Theta(U)|_{\mathbb{F}}$.

We shall use below (co)homology with coefficient in some *field* $\mathbb{F}$ and, given a linear subspace $\Theta \subset H^*(A;\mathbb{F})$, write

$$\mu_\Theta(U) = \Theta \cap Ker_{A \smallsetminus U}$$

for $Ker_{A \smallsetminus U}$ denoting the kernel of the homomorphism

$$H^*(A; \mathbb{F}) \to H^*(A \smallsetminus U; \mathbb{F})$$

and denote the rank of $\mu_\Theta(U)$, by $|\mu_\Theta(U)|_{\mathbb{F}}$.

Homology Measures in Cartesian powers.

Poincare polynomials that encode ranks of (co)homologies with coefficients in a field $\mathbb{F}$, are *multiplicative for Cartesian products of spaces* and Poincare polynomials of Cartesian power spaces $A = X^I$ are power polinomolas,

$$Poinc_{A;\mathbb{F}} = (Poinc_{X;\mathbb{F}})^N, \ N = |I|.$$

for all finite sets I and all fields $\mathbb{F}$.

The asymptotics of coefficients of polynomials P^N for $N \to \infty$ (that are ranks of the cohomologies of X^N for $P = Poinc_{X^N}$) are seen with Boltzmann's logarithmic rate decay formula.

Besides, the *tensorial reduction inequality* (linearizied of Shearer's) from the previous section provides

a bound on the ranks of the homology inclusion homorphisms of open subsets $U \subset X^I$ in terms of such ranks of the coordinate projections $U_J \subset X^J$ from U to X^J, $J \subset I$, that is

$$\prod_{J \subset I} r_J^{\alpha(J)} \geq r_I,$$

where $\alpha(J)$, $J \subset I$, is a positive function on subsets $J \subset I$, such *that*

$$\sum_{J \subset I} \alpha(J)\chi_J = 1,$$

for the $\{0,1\}$*-characteristic (indicator) function of the subsets* $J \subset I$*.*

This inequality fails to be true for the ranks of homomorphisms $H_*(U) \to H_*(X^J)$, but, possibly, something can be recovered with a suitable *positivity condition*, e.g. where X is an algebraic variety, where $U \subset X^I$ is a subvariety and where "positivity classes" are those *representable by subvarieties.* And if the varieties in question are defined over $\mathbb{C}$ one may use some positivity on differential forms coming from *the Hodge theory.*

In fact, to seems some cones of

"Hodge positive" (harmonic) forms on (infinite dimensional?) algebraic varieties may serve as a basis for a generalised probability theory.

There also nontrivial inequalities between the "measures" μ_Θ of systems of subsets in $A = X^N$, e.g. as follows.

Separation inequality in the N-torus.

Let $U_1, U_2 \subset \mathbb{T}^N$ be *non-intersecting* (closed or open) subsets and let

$$\Theta_1 = H^{n_1}(\mathbb{T}^N; \mathbb{F}),$$
$$\Theta_2 = H^{n_2}(\mathbb{T}^n; \mathbb{F})$$

for $n_i \leq N/2$, $i = 1, 2$, and some field $\mathbb{F}$. Then

$$|\mu_{\Theta_1}(U_1)|_{\mathbb{F}} \cdot |\mu_{\Theta_2}(U_2)|_{\mathbb{F}} \leq c \cdot |\Theta_1|_{\mathbb{F}} \cdot |\Theta_2|_{\mathbb{F}}$$

for $c = n_1 n_2 / N^2$.

Observe that

$$\left|\Theta_i = \wedge^{n_i}\mathbb{F}\right|_{\mathbb{F}} = \binom{N}{n_i}.$$

About the Proof. This is reduced by a simple ordering argument in the Grassmann algebra $H^*(\mathbb{T}^N;\mathbb{F})$ to the special case of U_1 and U_2 being *monomial subsets*, i.e. unions of coordinate subtori in $\mathbb{T}^N$ where this amounts to a combinatorial inequality due to Matsumoto and Tokushige. (See part 2 of my *Singularities, Expanders and Topology of Maps*.)

Question. What is the "full set" of (asymptotic) inequalities between cohomology measures of (finite?) systems of open subsets $U_k \subset X^N$ with a given pattern (nerve) of intersections between U_k?

Homotopy Spectra.

Let A be a topological space and $E : A \to \mathbb{R}$ a continuous real valued function, that is thought of as an energy $E(a)$ of states $a \in A$ or as a Morse-like function on A.

The subsets

$$A_r = A_{\leq r} = E^{-1}(\infty, r] \subset A, \ r \in \mathbb{R},$$

are called the (closed) *r-sublevels of E.*

A number $r_\circ \in \mathbb{R}$ is said *to lie in the homotopy spectrum of E* if the homotopy type of A_r undergoes an *essential*, that is *irreversible, change* as r passes through the value $r = r_\circ$.

Quadratic Example. Let A be an infinite dimensional projective space and E equal the ratio of two quadratic functionals. More specifically, let E_{Dir} be the Dirichlet function(al) on differentiable functions $a = a(x)$ normalised by the L_2-norm on a compact Riemannian manifold X,

$$E_{Dir}(a) = \frac{||da||^2_{L_2}}{||a||^2_{L_2}} = \frac{\int_X ||da(x)||^2 dx}{\int_X a^2(x) dx}.$$

The eigenvalues $r_0, r_1, r_2, ..., r_n, ...$ of E_{Dir} (i.e. of the corresponding Laplace operator) are *homotopy essential* since the rank of the inclusion homology homomorphism

$$H_*(A_r; \mathbb{Z}_2) \to H_*(A; \mathbb{Z}_2)$$

strictly increases (for $* = n$) as r passes through r_n.

Volume as Energy. Besides Dirichlet's there are other natural "energies" on spaces A of continuous maps between Riemannian manifolds, $a : X \to R$, such as

the k-volume of the pullback of a subset $R_0 \subset R$,

$$a \mapsto vol_k(a^{-1}(R_0)),$$

$$k = dim(R_0)+(dim(X)-dim(R)).$$

(This volume may be understood as the corresponding *Hausdorff measure* but if $k \geq 2$ or as *the Minkovski measure.*)

Notice that only the topology of the range space R enters this definition, but often some symmetry group of the pair (R, R_0) is essential. For instance if $R = \mathbb{R}$ and $k = dim(X) - 1$ then one works with the (infinite projective) space of non-zero continuous functions $a : X \to \mathbb{R}$ divided by the involution

$a \leftrightarrow -a$.

A more sophisticated version of the above is the k-volume function on the

space $C_k(X;\Pi)$ *of* k*-dimensional* Π*-cycles in a Riemannian manifold* X,

where Π is an Abelian group with a norm-like function on it, e.g. $\Pi = \mathbb{Z}$ or $\Gamma = \mathbb{Z}_p = \mathbb{Z}/p\mathbb{Z}$.

These spaces of (rectifiable) cycles with natural (flat) topologies are homotopy equivalent to products of Eilenberg-MacLane spaces that have quite rich homology structures that makes the homotopy spectra of the volume energies on these spaces,

$$E = vol_k : C_k(X;\Pi) \to \mathbb{R}_+,$$

quite non-trivial. (See *Minimax problems related to cup powers and Steenrod squares* by Larry Guth.)

Packing Energy. Let X be a *metric space* and $A = A_N(X)$ be the *set of subsets* $a \subset X$ of finite cardinality N. Let

$$\rho(a) = \min_{x,y \in a, x \neq y} dist(x,y)$$

and define *packing energy* as

$$E_N(a) = \frac{1}{\rho(a)}$$

for the energy of a.

Sublevel $A_{1/r}$ of this energy represent packings of X by r-balls.

(There is nothing special about $\frac{1}{\rho}$ – one could use, instead of $\frac{1}{\rho}$, an obituary positive monotone decreasing function in ρ.)

Permutation Symmetry and

Fundamental Group.

The space $A_N(X)$ of (unordered!) N-tuples of points in X can be seen as the quotient space of the space $X^{I_{inj}} \subset X^I$ of injective maps of a set I of cardinality N into X by the permutation group $S_N = Sym(I)$,

$$A_N = X^{I_{inj}}/Sym(I),\ card(I) = N.$$

This suggest a G-equivariant setting for the homotopy spectrum for energy functions $E(x_1, x_2, ..., x_N)$ on X^I that are invariant under subgroups $G \subset S_N$, where even for fully symmetric E it may be profitable to use subgroups $G \subsetneq S_N$ containing only special permutations.

Since the action of $S_N = Sym(I)$ on $X^{I_{inj}}$, (unlike the corresponding action of $Sym(I)$ on the Cartesian power X^I) is *free* the group S_N is seen in the fundamental group of $A_N(X)$, provided, for instance, X is a connected manifold of dimension $n \geq 2$. And if X equals the Euclidean n-space, the n-ball or the n-sphere, for $n \geq 3$, then

the fundamental group $\pi_1(A_N(X))$ *is isomorphic to the permutation group* S_N *and the main contribution to the homotopy complexity of the space* $A = A_N(X)$ *comes from this fundamental group.*

Numbers and Orders.

The role of real numbers in the concept of "homotopy essential spectrum" reduces to indexing the subsets $A_r \subset A$ according to their *order by inclusion:* $A_{r_1} \subset A_{r_2}$ for

$r_1 \leq r_2$.

Homotopy "spectra" make sense for functions $X \to R$ where R is in an arbitrary *partially ordered* set, where it is convenient to assume that R is a lattice i.e. it admits inf and sup.

Additivity, that is the most essential feature of physical energy, becomes visible only for spaces A that split as $A = A_1 \times A_2$ for$E(a_1, a_2) = E(a_1) + E(a_2)$.

Induced Energy $E_\circ$ on Category $\mathcal{H}_\circ(A)$

Let $\mathcal{S}$ be a class of topologicl spaces S and let $\mathcal{H}_\circ(A) = \mathcal{H}_\circ(A; \mathcal{S})$ be the category where the objects are homotopy classes of continuous maps $\phi : S \to A$ and morphisms are homotopy classes of maps $\psi_{12} : S_1 \to S_2$, such that the corresponding triangular diagrams are (homotopy) commutative, i.e. the composed maps $\phi_2 \circ \psi_{12} : S_1 \to A$ are homotopic to ϕ_1.

Extend functions $E : A \to \mathbb{R}$ from A to $\mathcal{H}_\circ(A)$ as follows. Given a continuous map $\phi : S \to A$ let

$$E(\phi) = E_{max}(\phi) = \sup_{s \in S} E \circ \phi(s),$$

denote by $[\phi] = [\phi]_{hmt}$ the homotopy class of ϕ. and set

$$E_\circ[\phi] = E_{mnmx}[\phi] = \inf_{\phi \in [\phi]} E(\phi).$$

In other words,
$E_\circ[\phi] \leq e \in \mathbb{R}$ *if and only if the map* $\phi = \phi_0$ *admits a homotopy of maps*
$\phi_t : S \to A$, $0 \leq t \leq 1$, *such that*

ϕ_1 *sends* S *to the sublevel* $A_e = E^{-1}(-\infty, e] \subset A$.

Definition. *The covariant (homotopy)* $\mathcal{S}$*-spectrum of* E is the set of values $E_\circ[\phi]$ for some class $\mathcal{S}$ of (homotopy types of) topological spaces S and (all) continuous maps $\phi : S \to A$.

For instance, one may take for $\mathcal{S}$ the set of homemorphism classes of countable (or just finite) cellular spaces. In fact, the set of sublevels A_r, $r \in \mathbb{R}$, themselves is sufficient for most purposes.

Cohomotopy $\mathcal{S}$*-Spectra on* $\mathcal{H}^\circ(A)$.

Now, instead of $\mathcal{H}_\circ(A)$ we extend E to the category $\mathcal{H}^\circ(A)$ of homotopy classes of maps $\psi : A \to T$,

$T \in \mathcal{S}$, by defining $E^{\circ}[\psi]$ as the supremum of those $r \in \mathbb{R}$ for which the restriction map of ψ to A_r,

$$\psi_{|A_r} : A_r \to T,$$

is *contractible* that is expressed in writing as $[\psi] = 0$.

(In some cases, e.g. for maps ψ into discrete spaces T such as Eilenberg-MacLane's $K(\Pi; 0)$, *"contractible"*, must be replaced by *"contractible to a marked point serving as zero"* in T.)

Then the set of the values $E^{\circ}[\psi]$, is called *the contravariant homotopy (or cohomotopy) $\mathcal{S}$-spectrum of E.*

For instance, if $\mathcal{S}$ is comprised of the *Eilenberg-MacLane $K(\Pi, n)$-spaces*, $n = 1, 2, 3, \ldots$, then this is called *the* Π-*cohomology spectrum of* E.

Relaxing Contractibility via Cohomotopy Operations. Let us express "contractible" in writing as $[\psi] = 0$, let $\sigma : T \to T'$ be a continuous map and let us regard the (homotopy classes of the) compositions of σ with $\psi : A \to T$ as an *operation* $[\psi] \overset{\sigma}{\mapsto} [\sigma \circ \psi]$.

Then define $E^\circ[\psi]_\sigma \geq E^\circ[\psi]$ by maximising over those r where $[\sigma \circ \psi_{|A_r}] = 0$ rather than $[\psi_{A_r}] = 0$.

Supermultiplicativity of "cohomology measures" $\mu^*(U)$ in "spectral" terms.

The supermultiplicativity of μ^* for intersections of susbsets $D_{r,i} = E_i^{-1}(r, \infty) \subset A$ can be equivalently expressed in terms of cohomomoly spectra as follows.

[min $\smile$]-*Inequality.* Let

$$E_1, ..., E_i, .., E_N : A \to \mathbb{R}$$

be continuous functions/energies and let $E_{min} : A \to \mathbb{R}$ be the minimum of these,

$$E_{min}(a) = \min_{i=1,...,N} E_i(a),\ a \in A.$$

Let $h_i \in H^{k_i}(A; \Pi)$ be cohomology classes, where Π is a commutative ring, and let

$$h_\smile \in H^{\sum_i k_i}(A; \Pi)$$

be the $\smile$-product of these classes,

$$h_\smile = h_1 \smile ... \smile h_i \smile ... \smile h_N.$$

Then

$$E^*_{min}(h_\smile) \geq \min_{1=1,...,N} E^*_i(h_i).$$

Consequently, the value of the "*total energy*"

$$E_{\Sigma} = \sum_{i=1,\dots,N} E_i : A \to \mathbb{R}$$

on this cohomology class $h_{\smile} \in H^*(A;\Pi)$ is bounded from below by

$$E^*_{\Sigma}(h_{\smile}) \geq \sum_{i=1,\dots,N} E^*_i(h_i).$$

On $\wedge$*-Product.* The (obvious) proof of [min $\smile$] relies on locality of the $\smile$-product that, in homotopy theoretic terms, amounts to factorisation of $\smile$ via $\wedge$ that is the *smash product* of (marked) Eilenberg-MacLane spaces that represent cohomology, where, recall, the *smash product* of spaces with marked points, say $T_1 = (T_1, t_1)$ and $T_2 = (T_2, t_2)$ is

$$T_1 \wedge T_2 = T_1 \times T_2 / T_1 \vee T_2$$

where the factorisation "$/T_1 \vee T_2$" means "*with the subset* $(T_1 \times t_2) \cup (t_1 \times T_2) \subset T_1 \times T_2$ *shrunk to a point*" (that serves to mark $T_1 \wedge T_2$).

In fact, general cohomotopy "measures" and spectra defined with maps $A \to T$ satisfy natural (obviously defined) counterparts/generalizations [min ⌣].

Pairing Inequality
for Cohomotopy Spectra.

Let A_1, A_2 and B be topological spaces and let

$$A_1 \times A_2 \overset{\circledast}{\to} B$$

be a continuous map where we write

$$b = a_1 \circledast a_2 \text{ for } b = \circledast(a_1, a_2).$$

For instance, composition $a_1 \circ a_2$:

$X \to Z$ of morphisms $X \overset{a_1}{\to} Y \overset{a_2}{\to} Z$ in a topological category defines such a map between sets of morphisms,

$$mor(X \to Y) \times mor(Y \to Z) \overset{\circledast}{\to} mor(X \to Z).$$

A more relevant example for us is the following

Cycles × packings.

Here,

A_1 is a space of locally diffeomorphic maps $U \to X$ between manifolds U and X,

A_2 is the space of cycles in X with some coefficients Π,

B is the space of cycles U with the same coefficients,

$\circledast$ stands for *"pullback"*

$$b = a_1 \circledast a_2 =_{def} a_1^{-1}(a_2) \in B.$$

This U may equal the disjoint unions of N manifolds U_i that, in the spherical packing problems, would go to balls in X; since we want these balls *not to intersect*, we take the space of *injective* maps $U \to X$ for A_1.

Explanatory Remarks. (a) Our "*cycles*" are defined as *subsets* in relevant manifolds X and/or U with Π-*valued functions* on these subsets.

(b) In the case of *open* manifolds, we speak of cycles with *infinite supports*, that, in the case of compact manifolds with boundaries or of open subsets $U \subset X$, are, essentially, *cycles modulo the boundaries* ∂X.

(c) *"Pullbacks of cycles"* that preserve their codimensions are defined, following Poincaré, for a wide class of smooth *generic* (not necessarily equividimensional) maps $U \to X$. (This is spelled out in my article *Manifolds: Where Do We Come From?...* .)

Let h^T be a (preferably non-zero) cohomotopy class in B, that is a homotopy class of non-contractible maps $B \to T$ for some space T, (where "*cohomotopy*" reads "*cohomology*" if T is an Eilenberg-MacLane space) and let

$$h^{\circledast} = \circledast \circ h^T : [A_1 \times A_2 \to T]$$

be the induced class on $A_1 \times A_2$, that is the homotopy class of the composition of the maps $A_1 \times A_2 \overset{\circledast}{\to}$

$B \overset{h^T}{\to} T$.

(We do not always notationally distinguish *maps* and *homotopy classes* of maps.)

Let h_1 and h_2 be homotopy classes of maps $S_1 \to A_1$ and $S_2 \to A_2$ for some spaces S_i, $i = 1, 2$,

(In the case where h^T is a *cohomology* class, these h_i may be replaced by *homology* – rather than homotopy – classes represented by these maps.)

Compose the three maps,

$$S_1 \times S_2 \overset{h_1 \times h_2}{\to} A_1 \times A_2 \overset{\circledast}{\to} B \overset{h^T}{\to} T,$$

and denote the homotopy class of the resulting map $S_1 \times S_2 \to T$ by

$$[h_1 \circledast h_2]_{h^T} = h^{\circledast} \circ (h_1 \times h_2) : [S_1 \times S_2 \to T]$$

Let $\chi = \chi(e_1, e_2)$ be a function in two real variables that is monotone unceasing in each variable. Let $E_i : A_i \to \mathbb{R}$, $i = 1, 2$, and $F : B \to \mathbb{R}$ be (energy) functions on the spaces A_1, A_2 and B, such that the $\circledast$-pullback of F to $A \times B$ denoted

$$F^{\circledast} = F \circ \circledast : A_1 \times A_2 \to \mathbb{R}$$

satisfies

$$F^{\circledast}(a_1, a_2) \leq \chi(E(a_1), E(a_2)).$$

In other words, the $\circledast$-image of the product of the sublevels

$(A_1)_{e_1} = E_1^{-1}(-\infty, e_1) \subset A_1$ and $(A_2)_{e_2} = E_2^{-1}($

is contained in the f-sublevel $B_f = F^{-1}(-\infty, f) \subset B$ for $f = \chi(e_1, e_2)$,

$$\circledast\big((A_1)_{e_1} \times (A_2)_{e_2}\big) \subset B_{f=\chi(e_1,e_2)}.$$

$\circledast$-Pairing Inequality.

Let $[h_1 \circledast h_2]_{h^T} \neq 0$, that is *the* *composed map*

$$S_1 \times S_2 \to A_1 \times A_2 \to B \to T$$

is non-contractible. Then the values of E_1 and E_2 on the homotopy classes h_1 and h_2 are *bounded from below* in terms of a lower bound on $F^\circ[h^T]$ as follows.

$$\chi(E_{1\circ}[h_1], E_{2\circ}[h_2]) \geq F^\circ[h^T].$$

In other words

$$\big(E_{1_\circ}[h_1] \leq e_1\big) \& \big(E_{2\circ}[h_2] \leq e_2\big) \Rightarrow \big(F^\circ[h^T] \leq \chi(e_1, e_2)\big)$$

for all real numbers e_1 and e_2; thus,

upper *bound* $E_1^\circ[h_1] \leq e_1$ +
lower *bound* $F^\circ[h^T] \geq \chi(e_1, e_2)$
yield
upper *bound* $E_2^\circ[h_2] \geq e_2$,

where, observe, E_1 and E_2 are interchangeable in this relation.

First "Proof". Unfold the definitions.

Second "Proof". Look at the $h^\circledast$-*spectral line* in the (e_1, e_2)-plane

$$\Sigma_{h^\circledast} = \partial\Omega_{h^\circledast} \subset \mathbb{R}^2$$

where $\Omega_{h^\circledast} \subset \mathbb{R}^2$ consists of the pairs $(e_1, e_2) \in \mathbb{R}^2$ such that the restriction of $h^\circledast$ to the Cartesian product of the sublevels $A_{1e_1} = E_1^{-1}(-\infty, e_1) \subset A_1$ and $A_{2e_2} = E_2^{-1}(-\infty, e_2) \subset A_2$ vanishes,

$$h^\circledast_{|A_{1e_1} \times A_{2e_2}} = 0.$$

This pairing was used by Larry Guth in "*Minimax problems related to cup powers and Steenrod squares*" for lower bounds on the homology spectrum of volume functions in spaces of $\mathbb{Z}_2$-cycles but it also provide some, albeit limited, information on parametric sphere packings.

(Critical points of packing energy are studied in

Min-type Morse theory for configuration spaces of hard spheres

by Yuliy Baryshnikov, Peter Bubenik, and Matthew Kahle, and in

Computational topology for configuration spaces of hard disks

by Gunnar Carlsson, Jackson Gorham, Matthew Kahle, Jeremy Mason.)

Also such pairing applies to *parametric symplectic packings.* (See *From Symplectic Packing to Algebraic Geometry and Back.* by Paul Biran for a survey of non-parametric case.)

All of the above concerns *finite*

systems of balls and is not directly applicable to infinite systems, such as the following.

Let $I \subset \mathbb{R}^n$ be a countable subset and let A be the space of maps $a : I \to \mathbb{R}^n$ that move ponts by bounded amount,

$$\sup_{i \in I} dist(a(i), i) \leq C = C(a) < \infty.$$

This space carries a natural topology and it is acted upon by the Group $\mathcal{G}$ generated by

(1) isometric transformations of $\mathbb{R}^n$;

(2) transformations $I \to I$ that belong to A.

We are still infinitely far from solution of the following

HOMOTOPY PACKING PROBLEM.

What is the $\mathcal{G}$-equaivariant (co)homotopy, in particular cohomology, spectrum of the packing energy

$$E : a \mapsto \sup_{i \neq j \in I} \frac{1}{dist(a(i), a(j))}?$$

8 Two Words On Large Deviations.

A conceptual, albeit rather computational, proof of Boltzmann's rate formula follows from a general expression of *the Legendre transform* of *logarithmic vanishing rates* of certain sequences $\Phi = \{\phi_N\}$ of probability measures ϕ_N on a topological linear space X defined by

$$rate_\Phi\{x\} = \inf_{U_x} \lim_{N \to \infty} \frac{1}{N} \log \phi_N(U_x),$$

where the infimum is taken over all neighbourhoods $U_x \in X$ of $x \in X$.

One defines this $rate_\Phi(x)$ only

for those Φ for which *this limit exists* and one usually assumes that the function $rate_\Phi(x)$ is *concave* ("$\frown$') at those points $x \in X$ where $rate_\Phi(x) > 0$.

A typical example of such Φ is the $\frac{1}{N}$-scaled sequence of convolution powers ϕ^{*N} of a single measure ϕ on X, say with compact support in X, where our "scaling" is induced by the homotheties of X, that are $x \mapsto x/N$.

Then by the

Donsker-Varadahn formula

(going back to to Boltzmann and Gibbs in certain cases)

the Legendre transform

of the function $rate_\Phi(x)$, denoted $rate^\perp_\Phi(y)$, $y \in Y$, where Y is the linear dual of X, is expressible by the (what is now-a-days called "tropical") limit of *the Laplace transforms* of the measures ϕ_N,

$$rate^{\perp}_{\Phi}(y) =$$

$$-\lim \frac{1}{N} \log \int_X \exp N\langle y, x\rangle d\phi_N$$

where $\langle y, x\rangle$ denotes the value of linear functionals y on $x \in X$ and where $\int_X ...d\phi$ stand for integration against a measure ϕ on X.

If ϕ_N equal the $1/N$-scaled convolution N-powers of a ϕ, then this simplifies since

$$\frac{1}{N} \log \int_X \exp N\langle y, x\rangle d\phi_N =$$

$$\log \int_X \exp\langle y, x\rangle d\phi$$

for all N.

Legendre Transform. Given a smooth function $f(x)$ in a domain

U of a linear – let it be finite dimensional at this point – space X, regard the differentials $d_x(f)$, $x \in X$ as vectors in the linear dual $Y = X^\perp$ and denote by $L : U \to Y$ the map $L : x \mapsto y = d_x(f)$.

Now we assume that $U \subset X$ is *convex* and the function $f(x)$ is *concave.* In this case, the map L is one-to-one and the "L-transport" $f^\perp = f \circ L^{-1}$ of f from $U \subset X$ to $U^\perp = L(U) \subset Y$ is called the *Legendre transform of f.*

It is amazing, albeit (almost) obvious after being stated, that this transform is involutive:

$$(f^\perp)^\perp = f.$$