Bernoully Lecture, March 27, 2018: Alternative Probabilities.

Misha Gromov,

April 8, 2018

I. What is probability?
II. What is it good for?
III. What are limitations of the concept of probability?
IV. What alternatives are desirable and what is available?

Below are a few classical answers to I and II.

We define the art of conjecture, or stochastic art, as the art of evaluating as exactly as possible the probabilities of things...

Jacob Bernoulli

(the author of one of 10 greatest

theorems of all time).

Lavoisier echoes:

The art of drawing conclusions from experiments and observations consists in evaluating probabilities.

Laplace amplifies:

Probability theory is nothing but common sense reduced to calculation.

George Boole, a logician, put it somewhat differently:

Probability is expectation founded upon partial knowledge. A perfect acquaintance with all the circumstances affecting the occurrence of an event would change expectation into certainty, and leave nether room nor demand for a theory of probabilities.

This idea is expressed more pictorially by Bart Kosko:

The probability that the bowman's arrow hits the deer does not lie in the arrow or the deer. It lies in the bowman's mind.

And all this makes Jay Gould, an evolutionary biologist, lament:

Misunderstanding of probability may be the greatest of all impediments to scientific literacy.

One cannot help but conclude by saying:

a curious aspect of the probability theory is that everybody thinks he understands it.

Jacques Monod, *misquoted.*

(We selected certain quotes for this lecture because they appear interesting and thought provoking, never mind if what in them is wrong or meaningless. They are not intended to confirm any view or opinion but rather to invite the reader to ponder on the ideas of the authors of these quotes.)

A FEW WORDS ON HISTORY.

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

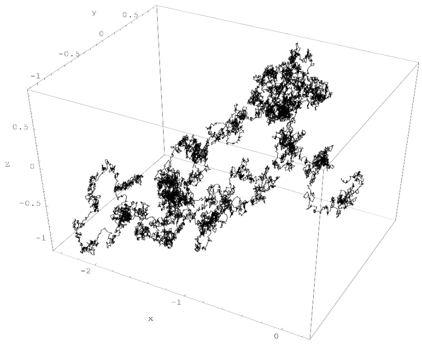
I of dice possess the science
and in numbers thus am skilled.

3000 years later, Titus Lucretius outlined a stochastic model of what was discovered and analysed by Jan Ingenhousz around 1785 and called for brevity

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.

Two thousand years have passed by and physicists and mathematicians (... Maxwell, Boltzmann, Gibbs, Einstein... , ... Cardano, Bernoulli, Laplace) have understood what Lucretius had in mind and worked out

the calculus of probabilities

which, in the words of Maxwell, constitutes [the] true logic of this world, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind.

Then this calculus was planted on the measure-theoretic soil (Borel, Kolmogorov) made ready by Cantor's set theory where it stretched to almost all regions of mathematics and physics.

In fact, probability displays the full beauty of its colors not so much at the core of the theory, but where its ideas intermingle with ideas from other fields.

And later several non–classical seedlings sprouted from the roots around the gorgeously branched tree of the classical probability:

non-commutative (quantum, free) probability(Von Neumann, Voiculescu)

algorithmic probability (Von Mises, Kolmogorov, Chaitin, Martin Löf...)

categorical probability theories.

Non-standard probability (Nelson, Loeb)

Fuzzy probability.

If somebody starts telling you he/she knows what *probability* is spit in the eye of this individual.

And it is not so much the immensity of the field which has been already developed, but rather the difficulty of an assessment of the limitations of the present day theory and discerning new directions – this is what makes "I KNOW" sound ridiculous.

Sample Questions.

Stochastic Generation of Bizarre Objects and Natural Conjectures.

Given a class/set $\mathcal{O}$ of "mathematical objects" ob it may be difficult to say much significant about all $ob \in \mathcal{O}$ and/or to single out interesting representatives ob in $\mathcal{O}$.

But it may be easy to pinpoint a natural probability measure μ or a class of such μ on $\mathcal{O}$, where the μ-random (μ-typical) objects $ob \in \mathcal{O}$ demonstrate their peacock's tail brightness.

For instance, little of substance can be said or *even asked* about geometry/topology of general subsets ob in the integer lattice $\mathbb{Z}^n \subset \mathbb{R}^n$.

But there are volumes of known and *conjectural* properties of *random ob* $\subset \mathbb{Z}^n$, coming under the heading of *percolation theory.*

Problem 1. Identify/generalise the logical mechanism(s) that empowers the idea of probability to generate new mathematical questions.

An effective construction of objects in $\mathcal{O}$ with typical properties, which are enjoyed by random/typical *ob*, may be difficult.

For instance, nobody has ever seen a *specific* subset in $\mathbb{Z}^2$ with essential features of a μ-*random* subset for any natural probability measure μ.

Problem 2. Rigorously formulate, prove (disprove?) and generalise this impossibility of deterministic imitation of properties of random objects.

Now let us formulate two concrete questions on the border of probability with two (quite elementary) mathematical structures which further illustrates limitations of our understanding of probability.

Linearised LW Inequality. Let $\Phi = \Phi(s_1, s_2, s_3, s_4)$ be a 4-linear form over some field and denote $|...| = rank(...)$. Then

the rank of the bilinear form

$$\Phi(s_1, s_2 \otimes s_3 \otimes s_4),$$

denoted

$$|1, 234| = rank(\Phi(s_1, s_2 \otimes s_3 \otimes s_4)),$$

is bounded by the ranks

$|12,34| = rank(\Phi(s_1 \otimes s_2, s_3 \otimes s_4))$,

$|13,24| = rank(\Phi(s_1 \otimes s_3, s_2 \otimes s_4))$

and

$|14,23| = rank(\Phi(s_1 \otimes s_4, s_2 \otimes s_3))$

as follows.

$|1,234|^2 \leq |12,34| \cdot |13,24| \cdot |14,23|$.

This can be reduced to the *3D Loomis-Whitney isoperimetric inequality* (and/or to the Shannon entropy inequality); also this can proven by applying the *Bernoulli law of large numbers* to the $\mathcal{N}$th tensorial power $\Phi^{\mathcal{N}\otimes}$, where $\mathcal{N}$ is a *nonstandard* (infinitely large) *integer.*

Question 1. Is there a link of the linearised LW inequality and similar "large numbers" phenomena with algebra-geometric inequalities concerning ranks of cohomology groups, e.g. in the Esnault-Viehweg proof of *the sharpened Dyson-Roth lemma*?

Fisher Metric. Recall (Archimedes, 287-212 BCE) the *real moment map* from the unit sphere $S^n \subset \mathbb{R}^{n+1}$ to the probability simplex $\triangle^n \subset \mathbb{R}^{n+1}$ for

$$(x_0, ..., x_n) \mapsto (p_0 = x_0^2, ..., p_n = x_n^2)$$

and observe following R. Fisher that the spherical metric (with constant curvature +1) thus transported to $\triangle^n$, call it ds^2 on $\triangle^n$, is equal, up to a scalar multiple, to the *Hessian of the entropy* $ent\{p_0, ..., p_n\} = -\sum_i p_i \log p_i$.

$$ds^2 = const \frac{\partial^2 ent(p_i)}{dp_i dp_j}.$$

If, accordingly, we take the "inverse Hessian" – a kind of double integral "$\int \int ds^2$" for the *definition* of entropy – we arrive at

Question 2. Are there *interesting* "entropies" associated to (real and complex) moment maps of general toric varieties? Is there a *meaningful* concept of "generalised probability" grounded in positivity encountered in algebraic geometry?

****

The calculus of probabilities, when confined within just limits, ought to interest, in an equal degree, the mathematician, the experimentalist, and the statesman.

Francois Arago.

Probability works fantastically well

in the classical physics, e.g. in the statistical mechanics and in its non-commutative version in the quantum physics as well, apparently due to the enormous symmetry of the systems it applies to.

Amusingly however, the famous *second law of thermodynamics*, which, apparently, is rooted in symmetry as much as in randomness, admits no mathematically satisfactory justification (formulation?) in the language of the classical probability theory.

Unable to justify this law logically, physicists glorify it poetically.

...the entropy of the universe tends to a maximum.

Rudolf Clausius (1865).

... though the energy itself is indestructible, the available part is liable to diminution by the action of certain natural processes, such as conduction and radiation of heat, friction, and viscosity.

James Clerk Maxwell.

.......a given system can never of its own accord go over into another equally probable state but into a more probable one.

Ludwig Boltzmann.

... Nature prefers the more probable states to the less probable because in nature processes take place in the direction of greater probability.

Max Planck (1903).

... if... the universe is in disagreement with Maxwell's equations– then so much the worse for Maxwell's equations. ... But if your theory is found to be against the second law of thermodynamics... – there is nothing for it but to collapse in deepest humiliation.

Arthur Eddington.

...the second law of thermodynamics has played in the history of science a fundamental role far beyond its original scope.

Ilya Prigogine.

Not only particles and fields of force had to come into being at the big bang, but the laws of physics themselves, and this by a process as higgledy-piggledy as genetic mutation or the second law of thermodynamics.

John Wheeler.

... classical thermodynamics... will never be overthrown, within the framework of applicability of its basic concepts.

Albert Einstein.

But let us be humble:

Every mathematician knows it is impossible to understand any elementary course in thermodynamics.

Vladimir Arnold.

The further we go from physics to the worlds of biology, psychology, linguistic, economics, ... the more we lose symmetry – the use of classical probabilistic concepts in heterogeneous environment becomes problematic.

This was articulated already by Claude Bernard who said around 1865 that

Averages confuse while aiming to unify and distort while aiming to simplify.

But in his 1866 paper Mendel demonstrated that probabilistic rendition of

the striking regularity with which the same hybrid forms always reappeared

in his experiments reveals the key players in the game of Life- – *the genes.*

Forty years later, in 1908, an aspect of Mendelian theory was (re)formulated in the 19th century probabilistic terms and became one of the most cited mathematical theorems (more than 1 000 000 for "Hardy-Weinberg" on Google), yet, unknown to majority of mathematicians.

A few years prior to Mendel's article, Darwin and Wallace proposed a description of evolution in terms of

random variations of organisms
+
potentially exponential growth of populations
+
cut-off of most of this growth by EXTINCTION,

called by Darwin natural selection. However, even nowadays, nobody can say for certain what random signifies.

Jacques Monod says in this regard:

The universe was not pregnant with life nor the biosphere with man. Our number came up in the Monte Carlo game....

and that.

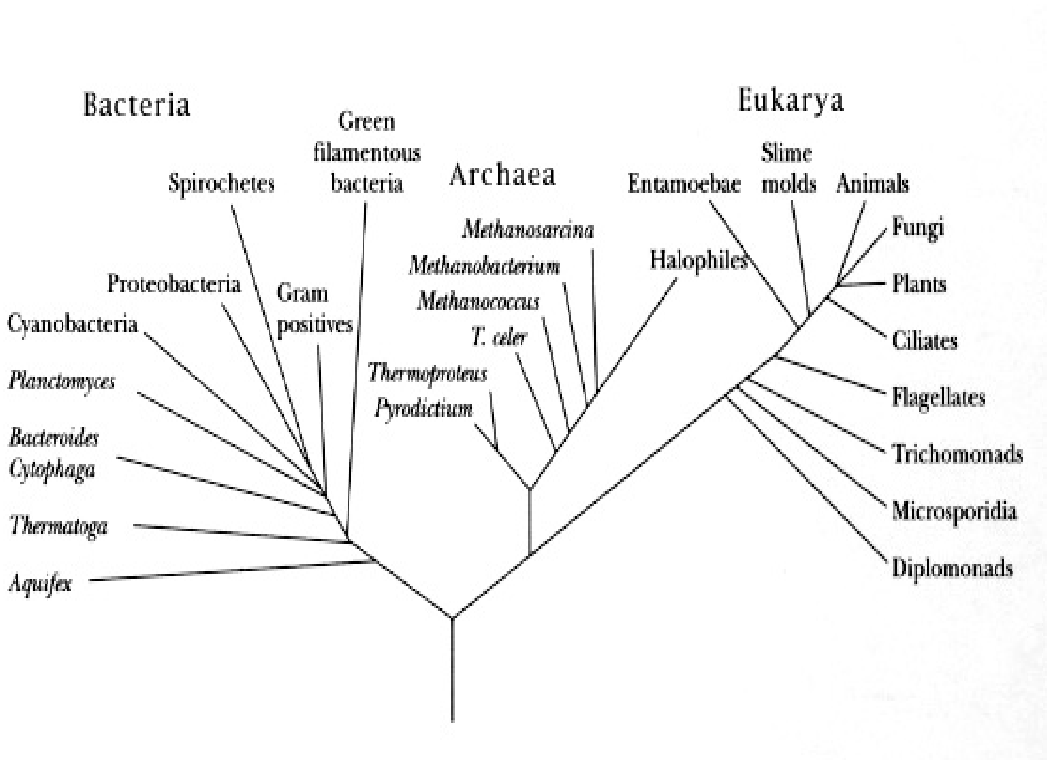
it is legitimate to view the irreversibility of evolution as an expression of the second law in the biosphere.

This may sound paradoxical, since

the directionality of the time arrow of Life (in the tiny islets where there is Life) is opposite to that of the rest of the physical universe:

thermodynamic dissipation homogenises whilst evolution diversifies.

But, possibly, what the two ther-



modynamic dissipation and biological evolution – have in common is (unknown) mathematics which supports both of them and may, in particular, provide an abstract reason for why

natural selection is a mechanism for generating an exceedingly high degree of improbability as Ronald Fisher remarks, But the usage of the word "probability may be, and often is, confusing, because, as Niels Bohr says

We are trapped by language to such a degree that every attempt to formulate insight is a play on words.

For instance if your probability is the one (tacitly and often unconsciously) accepted by physicists, you run into problems as Fred Hoyle – a possessor one of the finest minds of the 20th century – does when he says:

there are about two thousand enzymes, and the chance of obtaining them all in a random trial is only one part in 10 to the 40,000 power, an outrageously small....

It is therefore almost inevitable that our own measure of intelligence must reflect ... higher intelligences ... even to the limit of God ...

Definitely,

probability≠probability.

Also one can object to Hoyle as Richard Dawkins does:

I think the probability of a supernatural creator existing is very very low.

Fine... except nobody, even the super-natural creator herself, can make sense of this probability.

Joking apart, a quantitatively minded Hoyle, was refusing to unquestionably accept the highly improbable in his view idea that mere potentiality of exponential in the naked natural selection model of evolution was powerful enough to make the *stochastic gradient ascent* in the fitness landscape – *natural selection* in biologists' parlance – implement the observed evolution rate, given real life limitations on reproduction rates and population sizes.

What about linguistics? Psychology, Economics?

Having no idea of what, mathematically speaking, economics is we limit ourselves to quoting Nassim Nicholas Taleb:

If you hear a "prominent" economist using the word *equilibrium* or *normal distribution* ... just ... put a rat down his shirt.

Linguistics feels closer home and

we – mathematicians – may even venture out our own definition of *language*, something like

a probability measure on the set of strings of symbols from a finite set.

Smart and cute, isn't it? – especially in view of what linguists think about it.

probability of a sentence is an entirely useless [concept], under any known interpretation of this term.

Naum Chomsky.

Question. Should we listen to linguists and shut up with our definitions or could we try to think of something better?

Our only hope lies in the Chomskian *known* – we must come up with a new concept of probability.

Apparently, life sciences, broadly understood,: biology, psychology, linguistics, machine learning.... need probability $\mathbf{p}(\mathit{event})$, where $\mathbf{p}$ is NOT a number.

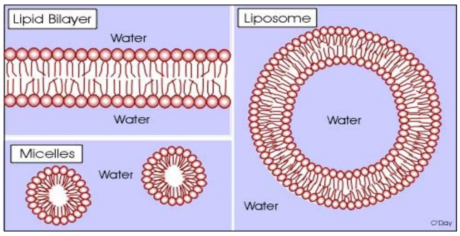
But...

... all the mathematical sciences are founded on relations between physical laws and laws of numbers, so that the aim of exact science is to reduce the problems of nature to the determination of quantities by operations with numbers.

This is what Maxwell says speaking of *Faraday's Lines of Force* (1856).

Well..., we shall be content to be alive rather than exact.

Conceivably, Life starts with *com-*



partmentalisation, such as formation of micelles – a process which is describable by means of the classical probability immersed into a beautifully intricate physical/mathematical structure and which, however, has not been studied much by pure mathematicians.

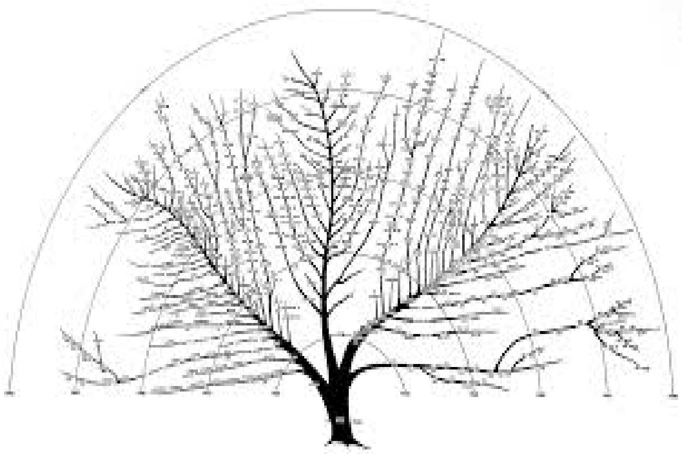
Next comes the basic biological instance of *self-organisation* – protein folding: a polypeptide chain P in a watery environment, driven by attraction/repulsion forces between residues and Brownian bombardment by water molecules, takes a definite 3D shape $P_{☻}$

Formally speaking, one has an energy function E on the configuration space $\mathcal{P}$ of chains P in the Euclidean 3-space, where the dimension N of $\mathcal{P}$ is roughly proportional to the number of residues in P.

The space $\mathcal{P}$ comes with (more or less) natural topology and measure structures; these endow the space $\mathcal{T}_P$ of sublevels of E into a *forest of weighted trees,* which play the role of numberical probability values p in the traditional statistical mechanics.

(The higher connectivity/homology invariants of these sublevels can be described in terms of *homological probability*, which, albeit mathematically attractive, doesn't seem to have a biological significance.)

The tree $\mathcal{T}_P$ doesn't come in isolation but as a member of an evolutionary family $\mathcal{F}(P)$ which makes a tree (modulo the horizontal gene transfer) in its own right.



Besides protein related structures, tree-like patterns can be seen in generative grammars of natural languages as well as in statistical descriptions of words in a corpus of a language, where their "meaning" is determined not by their "brute frequencies" but by distributions of their associations with other words.

And the above kinds of trees suggest a direction that may lead to "denumerification" (categorisation?) of the probability theory.

References.

Feller's

An Introduction to Probability Theory and Its Applications

probably, remains the best classical probability textbook.

Charles J. Geyer's

Radically Elementary Probability and Statistics.

http://www.stat.umn.edu/geyer/nsa/o.pdf

presents an exposition of foundation and application of probability from the standpoint of nonstandard

analysis in the spirit of Edward Nelson's book with a similar title.

Sérgio B.Volchan's historical survey

What Is a Random Sequence,

Journal The American Mathematical Monthly Volume 109, 2002 - Issue 1

https://www.maa.org/sites/default/files/pdf/
19-12/monthly046-063.pdf

contains the definition of algorithmic randomness a la Kolmogorov etc.

Noga Alon's and Joel H. Spencer's book

The Probabilistic Method

presents applications of classical probability theory to combinatorial problems.

There are several accounts of non-commutative/quantum/free probability theories on the web:

Quantum Probability Theory

by Miklós Rédei and Stephen J. Summers,

arXiv:quant-ph/0601158

Lecture Notes on Free Probability

by Vladislav Kargan

arXiv:1305.2611

and

Non-Commutative Probability Theory

by Paul D. Mitchener

www.mitchener.staff.shef.ac.uk/free.pdf

The following are two surveys on fuzzy theories.

Possibility Theory versus Proba-

bility Theory in Fuzzy Measure Theory

by Parul Agarwal and Dr. H.S. Nayal.

in Int. Journal of Engineering Research and Applications Vol. 5, Issue 5, (Part -2) May 2015, pp.37-43,

and

Fuzzy Probability Theory I: Discrete Case I.

by Burak Parlak and A. Cağri Tolga

in

Fuzzy Statistical Decision-Making, Theory and Applications

Springer 2016

Algebraisation of basic probabilistic concepts is in the texts

The Homological Nature of En-

tropy.
by Pierre Baudot and Daniel Bennequin
Entropy 2015, 17(5), 3253-3318,

Lectures on Algebraic Statistics
by Mathias Drton, Bernd Sturmfels, Seth Sullivant
https://math.berkeley.edu/ bernd/owl.pdf,

Algebraic Statistics for Computational Biology.
Edited by Lior Pachter and Bernd Sturmfels
yaroslavvb.com/papers/pachter-algebraic.pdf

Categories of probability spaces,
Melissa Lynn (2010)
www.math.uchicago.edu/ may/VIGRE/VIGR
and

Where Do Probability Measures Come From?

Posted by Tom Leinster
https://golem.ph.utexas.edu/category
/2014
/10/ where_do_probability_measures.html

The title of the following article by Peter Norvig,
http://norvig.com/chomsky.html,
is self-explanatory.

On Chomsky and the Two Cultures of Statistical Learning

Finally, I refer to my texts where the topics mentioned in this lecture are discussed.

Six Lectures on Probabiliy, Symmetry, Linearity.
http://www.ihes.fr/~gromov/PDF/
probability- huge-Lecture-Nov-2014.pdf

In a Search for a Structure, Part

1: On Entropy.
http://www.ihes.fr/ gromov/PDF/ structre- serch- entropy-july5-2012.pdf

Mendelian Dynamics and Sturtevant's Paradigm.
http://www.ihes.fr/ gromov/ topics/mendel-may31.pdf

Morse Spectra, Homology Measures and Parametric Packing Problems.
arXiv:1710.03616

Symmetry, Probabiliy, Entropy,
Entropy 2015,17, 1273-1277;
https://pdfs.semanticscholar.org/bd36/ da916b0cbbe5a591593c6da0e50245cad0c7.pdf