"Dimension" brings to mind the idea of "space", but the word "large", doesn’t belong with mathematics. It isn’t our business to know how many grains make a heap.

Yet, we ask:

Which spaces are qualified as large dimensional?

To get an idea, look at:

A. Finite dimensional Euclidean spaces $\mathbb{R}^N$.

B. Finite combinatorial power spaces, such as $\{\bullet,\circ\}^N$.

Motivated by the Life on Earth examples (see below), we declare that "large" begins somewhere around $N = 100$ and may go up to $10^{10} - 10^{12}$, maybe even up to $10^{15}$, while dimensions $N$ which are well above $10^{15}$, say starting with $10^{18}$, are too close to infinity from our perspective to be called "finite".1

Then the second question arises:

What is so special about these spaces, what are properties characteristic for these $N$?

A pronounced feature of $N$-dimensional power spaces $X$ with large $N$ is tendency of functions on $X$ to be nearly constant.

For instance, the angle (function) between two random unit vectors in $\mathbb{R}^N$ is almost constant: close to $90^\circ$ with overwhelming probability for large $N$.

This is most conspicuous for very large $N$, greater than $10^{12}$, and if you wish $> 10^{18}$, which is common in statistical mechanics, where propensity of functions (macroscopic observables) to be constant is articulated as the identity average = typical.

1Live systems, unlike physical ones, are numberphobic. We shall see later on that virtually all meaningful entities $N$, be they large or small, enter the scene of Life in non-numerical gowns.

In fact, logic of Life, brings into question the common convention of $X^N$ defined as the set of "strings" $x_1,\ldots,x_N$, which depends on an ad hoc representation of numbers $N$ by sets $\{1,\ldots,N\}$. 
which means that the values of functions on these spaces tend to concentrate near their averages.\(^2\)

A mathematician is able to This concentration of measure phenomenon is ubiquitous in mathematics and in mathematical physics, we return to it in section????, but it is not always there. For instance, in lotteries of Life, where the winners take all, exactly the opposite happens:

*instead of concentration of functions \(f\) in the domains of their values, one sees Dirac-like \(\delta\)-concentrations of \(f\) in their domains of definition. This makes averages of such functions \(f\) as far from typical as it is conceivably possible.*

However, none of these two concentrations is visible, at least not directly, in functions, often called features of (almost) anything related to Life. These features come ablaze in billions colors, nothing you can call "constant" or "approximately constant".

And albeit the domains \(X\) of definitions and domains \(Y\) of values of these functions/features \(f\) are associated with power spaces, these \(X\) and (not always) \(Y\) have intricate internal structures and relevant \(f : X \rightarrow Y\) are far from being plain and simple.

**But what are these spaces? What are functions on these spaces – features of their members?**

How about this?

1. **The space of all live things on Earth.**

But is it a space in any sense? Does it possess any geometry? Can one attach a number to this "space" and justifiably call it "dimension"?

A possible (but not the only) way to bring this "space" to the dominion of Math is to view it as a (randomised?) quotient space of another space.

2. **The space of genomes of individual organisms on Earth,**

where the latter can be regarded as

\[ 2_n \text{ a subset of the set of the space of finite strings in four symbols}.^3 \]

This is still far from the true definition (if such a definition exists at all), where the main reason for this is a difficulty with a proper mathematical interpretation of "a" in the above "a quotient" and "a set".

But regardless of what this "proper" is – this we shall discuss later – one can safely say that the dimension of the "space" of (significantly different individual) organisms on Earth is in the range \(10^4 - 10^9\).\(^4\)

The difficulty faced by a mathematician in studying, or even in defining "spaces" of organisms and genomes, besides their size and complexity, is their accidentally: these "spaces" are come as end-points of a single, possibly non-representative, branch of a grand random process: biological evolution.

\(^2\)Without this the statistics would be inapplicable in physics, since average is what is amenable to a mathematical evaluation and typical is what is observed in an experiment.

\(^3\)Customary, these are A, C, G, and T for the nucleotides: Adenine, Cytosine, Guanine and Thymine making DNA.

\(^4\)The length of genomes of certain viruses goes below \(10^3\), humans have almost \(3 \times 10^9\)-long genomes and the genomes of some amoeba like creatures may reach close to \(10^{11}\). But most of "information" carried by genomes, especially by the long ones, is, apparently, erased by the "quotient map" from genomes to organisms. Probably, every organism (class of organisms?) can be identified and adequately described – modulo stochastic variations – by \(10^4\)-\(10^6\) (not necessarily numerical) parameters – the (mainly physiological) features of this organism.
Thus, a mathematician should either turn to the larger "space" of all conceivable organisms and/or genomes or, on the contrary, to focus on representative fragments of these spaces.

The most studied such fragment, which is located near the boundary of Life with the physical world, is

3. The space of proteins.

This, similarly to the "space of organisms", can be seen as a quotient of a larger but more accessible space:

4. The space of polypeptide chains,

that is a sequence space in 20 letters:

polypeptides which make proteins are (hetero)polymeric chains of length \( N \), roughly, between 30 and 30 000 \(^6\) composed of 20 (sometimes 21) basic amino acids.

The arrow

\[ \text{polypeptides} \rightarrow \text{proteins} \]

is physically implemented by the process of protein folding, which takes place in the polypeptide configuration spaces.

Mathematically, the configuration space \( C_P \) for a polypeptide \( P \), is a domain in the torus of dimension \((2+\sigma)N\), where \( N \) is the length of \( P \) and \( \sigma \) is, roughly, the average number of the side chains in the amino acids in \( P \).

Folding of a polypeptide to a protein in a water environment can be modelled by a randomised gradient descent for the energy function \( E_P \) on \( C_P \) defined by the mutual physical/chemical interactions between the residues in \( P \) as well as their interaction with the water molecules.

Albeit the principles of the protein folding (essentially, the shape of the energy landscape in \( C_P \)), unlike how it is with the arrow genomes to organisms, are, at least in general terms, understood, the protein folding problem in most respects remains unresolved.

But the true biological problem, which is more subtle and more interesting than the (essentially physical/mathematical) folding problem, concerns not individual spaces \((C_P, E_P)\), but their totality parametrised by the space \( \mathcal{P} \) of polypeptide (sequences) \( P \), where the present day \( \mathcal{P}_{\text{now}} \) can be seen as a set of quasistationary points of the evolutionary dynamics acting on \( \mathcal{P} \), which, up to some extent, may be represented by a protein fitness landscape in \( \mathcal{P} \).

---

5"Conceivable" and "mathematically expressible" are synonymous for a mathematician.

6Short polypeptide chains, even if they serve some functions in cells, are, somewhat arbitrarily, called peptides.

7In the course of polymerisation – synthesis of polypeptides – amino acids are slightly curtailed; what remains of them in polypeptides are called amino acid residues.

8There is also an opposite arrow as well, \( \text{proteins} \rightarrow \text{polypeptides} \), since proteins "remember" the order of amino acids in them: strictly speaking the protein space is bigger than that of polypeptides. But a working protein is as little aware of this order as an organism of the order of nucleotides in its chromosomal DNA.

9According to the orthodox Darwinism, evolution is adequately described by the fitness function defined as the relative reproduction rate, similarly to how a physical system is run by a single energy function. But when you look at this "relative" with an open mind ready to accept the ubiquitous numberphobicity of evolutionary biology, you realise that what come out of "rate" is not a mere number but an elaborate structural entity and you will see a new mathematical picture of the evolutionary landscape much richer in colors than what is offered by the model(s) of the (neo)classical Darwinism.
"Space" is an attractive concept, but does it apply to all Life's children. Are, for instance:

5. the space of states of the mind
and
6. the space of states of the brain

which are so close and dear to us, true spaces? There is no simple answer (some non-answers are given in section ???) but there are several bona fide spaces which contains traces of the above which we shall discuss in detail later. Among these are:

7. memory spaces, including Kanerva model as an example.
and
8. weight spaces of neural networks with composed functions on them.

Also Life has several beautiful grandchildren spaces, such as

9. spaces of natural languages and sentences in these
and
10. space of mathematical ideas,

where even the tiny fragments of the latter:

11. spaces of chess positions and chess games,

hide more charming surprises (we shall demonstrate them later), than a traditional mathematical picture of these spaces shows.

Contents

1 Probability, Symmetry, Concentration. 5
   1.1 Bernouilli-Boltzmann-Hoeffding Inequality and Hamming Geom-
        try ......................................................... 5
   1.2 Kanerva Memory ............................................ 10
   1.3 Balls, Spheres, Gaussian Cubes, Maxwell Distributions, Compar-
        ison of Concentrations and the Archimedean Moment Maps, . . 12
   1.4 Boltzmann and Entropy .................................... 16

2 ........ 16
3 ........ 16
4 ........ 16
5 ........ 16
6 Selected References 16
1 Probability, Symmetry, Concentration.

1.1 Bernoulli-Boltzmann-Hoeffding Inequality and Hamming Geometry

The first (?) recorded instance of what is now called "measure concentration" was the Law of Large Numbers stated by Cardano (1501 – 1576) in qualitative form:

the typical value of the sum of many independent random variables is (relatively) close to the sum of expectations of these variables,

which was was proved by Jacob Bernoulli almost two hundred years later for independent equidistributed \((0, 1)\) variables.  

In this case the sum of \(\Sigma_N\) of these takes the values \(0, 1, 2, \ldots, N\), where the probability \(p_i\) of \(\Sigma_N = i\), for \(i = 0, 1, \ldots, N\), equals the \(i\)th binomial coefficient normalised by \(2^{-N}\) in order to have \(p_0 + p_1 + \ldots + p_N = 1\). Thus, \(p_i\) are defined by

\[
(1 + x)^N = 2^N (p_0 + p_1 x + \ldots + p_N x^N), \text{ i.e. } p_i = 2^{-N} \binom{N}{i}.
\]

They say, it took 20 years for Bernoulli to prove his theorem, but today it is effortlessly derived from the (Hilbertian) Pythagorean theorem as follows.

Let \(F\) be the (Hilbert) space of functions \(f\) on a measure space \(M\) (e.g. a finite set of atoms with unit weights), where the distance is defined by the formula

\[
dist^2(f, g) = \int_M (f - g)^2 d\mu.
\]

Let \(f_1, f_2, \ldots, f_N\) be mutually orthogonal (i.e. \(\int_M (f_i \cdot f_j) = 0\)) functions.

Then the norm (i.e the distance from the zero function) of the sum of these functions satisfies

\[
\|f_1 + f_2 + \ldots + f_N\|^2 = \|f_1\|^2 + \|f_2\|^2 + \ldots + \|f_N\|^2.
\]

(It is hard to appreciate the greatness of this formula – familiarity breeds contempt. Yet, try to rationally explain why nothing of the kind holds true if the exponent "2" is replaced by any other number.\(^1\))

Back to Bernoulli, think of random variables as functions \(f\) on a probability space \(M\).

This may (justifiably) strike you as artificial but there is a distinguished space \(M\) where the action takes place, namely the set \((0, 1)^N \subset \mathbb{R}^N\) with the \(2^{-N}\) weights assigned to all points in this set, where the \(i\)th random variable is given by the projection of \((0, 1)^N\) to the \(i\)th Euclidean coordinate.

These, of course, are non-orthogonal, but we can render them such by subtracting the constant functions equal \(1/2\) from all of them and, to save notation, we multiply each of them by 2.

Now the resulting, (still independent!) variables, call them \(f_i, i = 1, 2, \ldots, N\), take values \(-1\) and 1 and, since they are independent and have zero means,

\(^{10}\)I haven’t check what and how was originally proven by Bernoulli.

\(^{11}\)A physicist would say this is so because the Nature has chosen the exponent "2" to relate energy to velocity, but a mathematician would maintain that this choice was forced on Nature by the Pythagorean theorem.
they are orthogonal. (This can be seen directly for our \( f_i \).) Therefore, by the Pythagorean theorem the norm of their average
\[
A_N = \frac{1}{N} (f_1 + \ldots + f_N)
\]
satisfies
\[
\|A_N\| = \left\| \frac{f_1 + \ldots + f_N}{N} \right\| = \sqrt{\frac{\|f_1\|^2 + \ldots + \|f_N\|^2}{N^2}} = \frac{1}{\sqrt{N}},
\]
since
\[
\|f_1\| = \|f_2\| = \ldots = \|f_N\| = 1.
\]

Now, obviously, if a random variable \( A \), seen (naturally or unnaturally) as a function on a probability space \( M \), have small norm it must be small on the most part of \( M \).

Indeed if \( A \) were \( > c \) on a subset in \( M \) of measure \( > \epsilon \) then its norm would satisfy
Markov (Chebyshev-Bienaymé) Inequality (in reverse).
\[
\|A\| > c\sqrt{\epsilon}.
\]
Thus,

the probability of \( A_N \) being \( > c \) is bounded by

Bernoilli Inequality
\[
\text{Prob}\{|A_N| > c\} < \frac{1}{Nc^2}.
\]

Remarks and Observations. (a) There is a sharper bound on this \( A_N = \frac{1}{N} \sum_i f_i \), which goes under the name of Hoeffding’s inequality.
\[
\text{Prob}\{|A_N| > c\} < \exp\left(-\frac{cN^2}{2}\right),
\]

inequality the proof of which fully relies on the independence of \( f_i \). (The above proof of \( \text{Prob}\{|A_N| > c\} < \frac{1}{Nc^2} \) uses only orthogonality of \( f_i \).)

Namely, independence of \( f_i \) implies that for \( \lambda \exp f_i \) for all \( \lambda \), where then the multiplicativity property of expectations, denoted \( E(\ldots) = \int_{X\ldots} \ldots d\mu \), applies:
\[
E\left(\exp \lambda \sum_i f_i\right) = E\left(\prod_i (\lambda \exp f_i)\right) = \prod_i E(\lambda \exp f_i).
\]

This, in conjunction with the (obvious) Markov inequality yields the proof. (Do it yourself or consult [Vershynin])

(b) All of the above applies to the sums of general independent variables, while in the present case (going back to Boltzmann, I guess, if not to Pascal) the inequality \( \text{Prob}\{|A_N| > c\} < \exp\left(-\frac{cN^2}{2}\right) \) amounts to the following bound on the sum of the first \( k \) binomial coefficients \( \binom{N}{i} = \frac{n!}{i!(N-i)!} \).
\[
2^{-N} \sum_{i=0}^{k} \binom{N}{i} \leq \exp\left(-2\alpha^2 N\right)
\]
for all $k \leq N/2$ and

$$\alpha = \frac{1}{2} - \frac{k}{N}$$

(It is instructive to think of the numbers $p_i = 2^{-N} \binom{N}{i}$ as probability weights of the $N+1$ atoms of the quotient space of the binary power set $\{0,1\}^N$ divided by the permutation group $\Pi_N$. In fact, other interesting probability spaces come this way, when a (homogeneous or non-homogeneous) space is divided by a group of its automorphisms. Examples of these are the spaces of unitary matrices with the Haar measures divided by conjugation and moment maps resulting from factorisation of their symmetry groups.)

(d) Bernoulli’s and Hoeffding’s inequalities say in geometric terms that the majority of vertices of a high dimensional $N$-cube

$$\Box^N = [0,1]^N$$

is located near the hyperplane $H_0$ and which passes through the center of the cube,

$$\left( \frac{1}{2}, \ldots, \frac{1}{2} \right) \in [0,1]^N,$$

and which is normal to the principal diagonal, that is the line between the opposite vertices

$$\left( 0, \ldots, 0 \right) \text{ and } \left( 1, \ldots, 1 \right)$$

in the cube.

In fact, the percentage of points that lie within the Euclidean distance $> c\sqrt{N}$ from $H_0$ exponentially decays for all $c > 0$ and $N \to \infty$.

Similarly the geometric interpretation of the Hoeffding’s inequality for the random variables uniformly distributed on the unit segments, shows that if $N > \frac{1}{c^2}$ then

almost all Euclidean volume/measure of the unit $N$-cube is contained in the band around $H_0$ of width $c\sqrt{N}$.

Here, it is worth mentioning that

The Euclidean diameter of the unit cube is $\sqrt{N}$.

(However obvious it is amusing to meet the same $\sqrt{N}$ as in the proof of the Bernoulli theorem.)

**On Random Walk and Random Spread.**

Bernoulli’s Pythagorean $\sqrt{N}$ is most vividly demonstrated by random walks, e.g in the integer lattice $\mathbb{Z}^k \subset \mathbb{R}^k$, where a typical path of length $N$ has the (Euclidean) diameter about $\sqrt{N}$.

This suggests that a typical configuration of a molecular chain $C_N$ of $N$ monomers, e.g. a polypeptide chain freely floating in (hot) water, must have the diameter about $N^\alpha$ for $\frac{1}{2} + \varepsilon_1 \leq \alpha \leq 1 - \varepsilon_2$.

The rational behind this is that $C_N$ looks like a path of a self avoiding random walk in $\mathbb{Z}^3$, where the self-avoidance (or self-repulsion) condition would make

\[\text{This condition says, in effect, that the relevant probability measure is supported on the set of all non-self-intersecting paths in the lattice graph of } \mathbb{Z}^3, \text{ and all these paths are assigned equal weights.}\]
$C_N$ spread/stretch on the average more than for the unrestricted random walk – conjecturally $\varepsilon_1 > 0$ – yet, not enough to make $C_N$ virtually straight, i.e. $\varepsilon_2$ should be $> 0$.\footnote{13}

Amazingly, none of the two inequalities $\varepsilon_{1,2} > 0$ has been proved,\footnote{14} and neither one knows (this seems easier) if the typical diameter of $C_N$ is equal to the average one.\footnote{15}

Concentrated? Yes, but where?

There are $2^{n-1}$ hyperplanes in the cube $[0,1]^N \subset \mathbb{R}^N$ geometrically indistinguishable from $H_0$: one hyperplane $H$ for each pair of opposite vertices. The band of width $\varepsilon\sqrt{N}$ around every of one of these $H$ contains most of cube volume, if $N >> \varepsilon^{-1}$.

It follows, that the intersections of such bands around several $H$, if this "several" is significantly smaller than $2^N$, contain most of the cube. But if $\varepsilon < e^{-1}$, $e = 2.718...$, then the intersection of all these bands in the cube around all these $H$, carries only $\delta^N$-fraction of the volume of the whole cube $[0,1]^N$ for $\delta$ only negligibly greater than $\varepsilon e$.

In fact this intersection is contained in the "diamond" of "radius" $\varepsilon N/2$ around the center of the cube, that is isometric to the standard $\diamond \subset \mathbb{R}^N$, defined by

$$\diamond = \left\{ x_1,...x_i,...x_N \right\}_{\sum_i |x_i| \leq \varepsilon N},$$

the volume of which is $\frac{(\varepsilon N)^N}{N!} \approx (\varepsilon e)^N$.\footnote{16} (Isn’t it amazing that the diamond of radius $N/6$, which looks pretty large, say for $N > 600$, has negligibly smaller volume than that of the unit cube?)

But despite what probability is shouting in you ear:

THE MEASURE OF THE CUBE IS STUCK TO ITS BOUNDARY

you can’t dismiss the center of the $N$-cube as something non-essential.

\footnote{13}{This is expected for the self avoiding random walk in $\mathbb{Z}^2$ and in $\mathbb{Z}^3$, while the higher dimensional random walks are oblivious of the self-avoidance condition.}
\footnote{14}{Lectures on Self-Avoiding Walks by Roland Bauerschmidt, Hugo Duminil-Copin, Jesse Goodman, and Gordon Slade, https://www.ihes.fr/~duminil/publi/saw_lecture_notes.pdf}
\footnote{15}{Conceivably (but unlikely) 49\% of chains have diameters approximately $\sqrt{N}$ and another 49\% are close to $N$.}
\footnote{16}{Since only a small part of our "diamond" is contained in the cube, the volume of this part must be significantly (?) smaller than $\frac{(\varepsilon N)^N}{N!} \approx (\varepsilon e)^N$.}
What one probabilistically perceives of a high dimensional object is very sensitive to the position of the observer.

Exercise. Figure out how much of the mass of the unit $N$-dimensional cube $[0,1]^N$ is concentrated near its $k$-dimensional faces. That is, evaluate the volumes of the $\rho$-neighbourhoods $U^k(\rho) \subset [0,1]^N$ of the unions of the $k$-faces in $[0,1]^N$. ( Relevant $\rho$ are $\varepsilon \sqrt{N}$)

Hint. It is instructive to start with such an evaluation for the cubical lattice, where is also useful to look at the dual lattice and where the computation simplifies if instead of the Euclidean/Pythagorean norm $\sum_i x_i^2$ one takes $\sup_i |x_i|$ for the norm.

**Hamming Metric**

The Hamming distance between two elements in a product space\(^{17}\)

$$X = \bigtimes_i F_i,$$

such as the power spaces $F^I$, e.g. for $F = \{0,1\}$, equals, by definition, 

*the number of those $i$ in the set $I$, where $a_i \neq b_i$.\(^{17}\)*

For example , if $I$ is a one point set and $X$ has no nontrivial product structure, then

$$\text{dist}_{\text{Ham}}(x,y) = 1 \text{ whenever } x \neq y.$$

Exercise. Check the triangle inequality for the Hamming distance, observe that the Hamming diameter of $X = F^I$ equals the cardinality of $I$,

$$\sup_{x,y \in X} \text{dist}_{\text{Ham}}(x,y) = |I|,$$

and show that every two points $x, y \in X = F^I$ with $\text{dist}_{\text{Ham}}(x,y) = d$ can be joined by a chain of immediate neighbour points $z_k \in X$,

$$x = z_1, ... z_k, ... z_d = y, \text{ dist}_{\text{Ham}}(z_k, z_{k+1}) = 1.$$

**Hamming Concentration.** The most essential feature of this metic is the concentration of the Hamming distance function which follows from the Bernoulli-Boltzmann-Hoeffding inequality which, in terms of $\text{dist}_{\text{Ham}}$, says that

*for majority of the pairs $(x,y) \in X \times X$, the distance $\text{dist}_{\text{Ham}}(x,y)$ is close to the mean distance that is half cardinality of the set $I$.\(^{17}\)*

For instance, the distances of 98% of "$\{0,1\}$-strings" $x \in \{0,1\}^{1000}$ from a given $x_0 \in X$ are in the range:

$$\text{dist}_{\text{Ham}}(x,x_0) = 500 \pm 37 \approx 500 \pm 0.135 \cdot 500,$$

where there are only 70 points out of 1000 within this distance from $x_0$.\(^{18}\)

---

\(^{17}\)The Hamming metric is most commonly (but not exclusively) used for binary spaces $F^I$, where $F$ is a two element set and where $\text{dist}_{\text{Ham}}$ reasonably well quantifies the concepts of similarity between "$\{0,1\}$-strings" $\{f_i\} \in X$.

\(^{18}\)Here, $A = B \pm C$ means $|A - B| \leq C$. 

9
Hamming Connectivity. The balls of radii 250 contain only negligible proportions of \( X \), namely less than \( \exp^{-125} < 0.000 000 000 000 000 000 000 000 000 \) of all \( 2^{1000} \) points in \( X \).

Yet all point \( x \) in \( X = \{0, 1\}^{1000} \) can be reached from any given point \( x_0 \) in four 250-long steps

\[
x_0 \leftrightarrow x_1 \leftrightarrow x_2 \leftrightarrow x_3 \leftrightarrow x_4 = x.
\]

Exercise. Show that the there are more than \( 2^{1500} \) and less than \( 2^{2000} \) triples \((x_1, x_2, x_3)\) that can serve in such chains between given \( x_0 \) and \( x \) in \( X \).

1.2 Kanerva Memory

Imagine, conceivable (potential?) memory items \( x \) being described by their features \( \phi \) the list \( \Phi \) of which is known to you beforehand.\(^{19}\) (This is unrealistic, but let it go.)

This means, our \( x \) are represented by \{yes, no\}-valued functions on a (fixed) set \( \Phi \) and the set \( X \) of all conceivable memory items is equated with the binary \{yes, no\}-space,

\[
X = \{\text{yes, no}\}^\Phi.
\]

We denote the value of such a function \( x \) at \( \phi \) by

\[
x \circ \phi = \phi \circ x,
\]

that is "yes" if \( x \) has feature "\( \phi \)" and "no" otherwise.

Then, the reason for this will become clear below, we represent yes by +1 and no by −1 and, often, call our functions \( \phi \mapsto x \circ \phi \) strings or \( \pm 1 \)-strings, despite the fact that there is no natural order in \( \Phi \). (This makes \( x \circ \phi \) look as a kind of a numerical scalar product.)

The realistic number \( N \) of features in \( \Phi \) may range, according to Kanerva, from 100 to 10 000 which make the cardinality \(|X|\) of the space \( X \) quote large , more than \( 10^{30} \) already for \( N = 100 \) and truly enormous, \( > 10^{100} \), for \( N > 330 \).

No realistic memory is large enough to encode all these items, but we don’t need it anyway.

All we want is to be able to encode any single item and then to continue encoding up to, say \( 10^9 \) of these, which is quite satisfactory, at least from a human point of view:

if you record one item each second 8 hours a day every day, such a memory will suffice for more than 90 years.\(^{20}\)

And a reliably registering of this amount of information needs only a few billion memory locations.

KMM. Below is a (slightly mathematised) description of a memory model suggested by Kanerva, called KMM, that, despite its shortcomings, displays certain features of the human memory (we discuss pros and cons in section ???)

\(^{19}\)We write "\( \Phi \)" instead of "\( I \)" not to be tempted to think of this set \( \Phi \) as \( \{1, 2, ..., |\Phi|\} \).

\(^{20}\)(Approximately) 365.25 days of the full turn of Earth around the sun make,

\[31 557 600 \text{ seconds}.\]
Registers, Counters and their Contents. Let the "hardware" of KMM be represented by a set $R$ of registers, say of cardinality $|R| = 10^9$, where each register $r$ consists of a set of counters and where each counter corresponds to a feature $\phi \in \Phi$.

According to this correspondence, the counters from an $r \in R$ are denoted $\phi_r$ and the set of these by $\Phi_r$.

The memory content of each counter is an integer $m(\phi_r)$. Thus, the full memory kept in all registers is an integer valued function $m$ on the product set,

$$m : \Phi \times R \rightarrow \mathbb{Z},$$

where we agree that "no recorded memory" is represented by zeros in the corresponding counters.

These numerical functions $m$, unlike $R$ and $\Phi_r$, are modified when new items enter the memory according to the rules described below.

Vicinity Structure of KMM. The main architectural attribute, which allows recording memory items from $X$ in $R$, is a subset $D \subset X \times R$, where the inclusion $(x, r) \in D$ reads as "$x$ and $r$ are $D$-neighbours", or "in $D$-vicinity, of each other".21

Memory Recording in KMM. Whenever a new item $x$ enters the memory, the numbers $x = \phi, \phi \in \Phi$, are added to the contents of the counters $\phi_r \in \Phi_r = \Phi$ for all registers $r$ in the $D$-vicinity of $x$.

For instance, if, originally, all registers were set on zero, then $x$ is recorded, exactly as it is, in all $D$-neighbour registers $r$ of $x$.

However, as we add more and more memory items, the sets of $D$-neighbours of different $x$ may start overlapping and some registers will contain sums of several $\pm 1$-strings.

Reading from Memory in KMM. We want to decide if our memory have earlier recorded an item $x$, or, it contains an item similar to $x$.

For this we introduce a cut-off operation $\sigma(\phi) \mapsto \bar{\sigma}(\phi)$ on functions $\sigma$ on $\Phi$, such that the result of this cut-off is an item $x \in X$ regarded as a $\pm 1$-function. We agree (this is negotiable) that

$$\bar{\sigma}(\phi) = +1 \text{ if } \sigma(\phi) > 0,$$

$$\bar{\sigma}(\phi) = -1 \text{ if } \sigma(\phi) < 0,$$

$$\bar{\sigma}(\phi) = \pm 1 \text{ randomly with probabilities } 1/2, \text{ if } \sigma(\phi) = 0.$$

Granted such a cut-off, let is construct the following (memory) search transformation on the set $X$,

$$S : X \rightarrow X,$$

where $S(x)$ is defined in two steps.

1. Add the contents $m_r = m(\phi_r), \phi_r \in \Phi_r = \Phi$, of the counters $\phi_r$ for those $r \in R$ which lie in the $D$-vicinity, say $V_x \subset R$, of $x$; this makes sense, since all sets $\Phi_r$ of counters in all registers $r$ are identified with $\Phi$.

2. Regard the resulting sum as a function on $\Phi$,

$$\sigma = \sigma(\phi) = \sum_{r \in V_x} m_r$$

21Think of "$D$-vicinity" as shorthand for "distance between $x$ and $r$ is less than $D$".
and let 
\[ S(x) \circ \phi = \hat{\sigma}(\phi). \]
If \( S(x) = x \) we conclude that \( x \) was introduced to the memory at some point. More generally, we regard all 
fixed points of the iterated maps 
\[ S^k = S \circ \ldots \circ S : X \rightarrow X \]
for moderate \( k \), say for \( k \leq 5 \), as the items recorded by the memory or at least as approximations to the actually recorded items.

The concentration and connectivity properties of the Hamming space allow (see [Kanerva]) a particular choice of the vicinity structure \( D \), which brings certain properties of \( \mathcal{KMM} \) close to those of the human memory.

We return to this in section ???. Here, we conclude by bringing reader’s attention to two features of \( \mathcal{KMM} \), which suggest a broader mathematical perspective on a class of such memory models.

A. \( \mathcal{KMM} \) is distributive: the information encoding an individual item is contained in several registers of \( \mathcal{KMM} \).

B. \( \mathcal{KMM} \) is, up to certain extent, dynamic: reading from \( \mathcal{KMM} \) relies on iterations of transformations on the set of (possible) memory items.

1.3 Balls, Spheres, Gaussian Cubes, Maxwell Distributions, Comparison of Concentrations and the Archimedean Moment Maps.

The concentration properties of \( N \)-Dimensional Euclidean balls and spheres are similar to those of cubes, where "roundness" of balls renders proofs easier.

The "law of large numbers" for balls. The volume of \( B^N \) is concentrated near the equatorial subball 
\[ B^{n-1} = \{x_1, \ldots, x_n\}_{x_1=0} \subset B^N. \]

In fact, \( B^{n-1} \subset B^N \) equals the pullback of zero under the projection of \( B^N \) to the \( x_1 \)-coordinate line, where the push-forward of the Euclidean (Lebesgue) measure\(^{22}\) of \( B^N \) is given (up to a constant coefficient) by the (density) function
\[ \left( \sqrt{1-x^2} \right)^N, \quad x \in [-1,1], \]
which is sharply concentrated near zero for large \( N \), since the function \( \sqrt{1-x^2} \) is strictly concave (the second derivative is strictly negative, \( \frac{d^2(\sqrt{1-x^2})}{dx^2} < 0 \)) with maximum at zero.

Similarly,

---

\(^{22}\)The push-forward, sometimes called projection of a measure \( \mu \) on \( X \) under a (continuous) map from \( X \) to \( Y \) is a measure \( \nu \) on \( Y \), such that \( \nu(U) = \mu(f^{-1}(U)) \) for all open subsets \( U \subset Y \), where \( f^{-1}(U) \subset X \) is the \( f \)-pullback of \( U \).
the measures of the spheres are exponentially concentrated near their equators.

This follows from the corresponding concentration property for balls, because

the measures of the balls are concentrate at their boundary spheres:

almost all mass of the unit \( N \)-ball

\[
B^N = B^n(1) = \{x_1, \ldots, x_N : x_1^2 + \ldots + x_N^2 \leq 1 \} \subset \mathbb{R}^N
\]

is concentrated near the boundary sphere

\[
S^{N-1} = \{x_1, \ldots, x_N : x_1^2 + \ldots + x_N^2 = 1 \} \subset \partial B^N.
\]

In fact, the concentric balls \( B^N(1-\delta) \subset B^N(1) \) have exponentially smaller volumes,

\[
\text{vol} \left( B^N(1-\delta) \right) = (1-\delta)^N \text{vol} \left( B^N(1) \right).
\]

The concentration for spheres can be also seen directly by observing that the spherical distance function \( x \mapsto \text{dist}_{S^{N-1}}(x, x_0) \) pushes forward the spherical measure to the segment \([0, \pi]\), where it’s density function is equal, up to a scaling constant, to \( \sin^{N-2}(x) \), which is sharply concentrated at \( x = \pi/2 \) for large \( N \).

Alternatively, (a weaker) concentration follows by (obvious) Markov’s inequality) from the following evaluation of the expectation (average) of squared coordinate \( x_1^2 \) on \( S^{N-1} \subset \mathbb{R}^N \):

\[
\mathbf{E}(x_1^2) = \frac{1}{N} \mathbf{E} \left( \sum_{i=1}^N x_i^2 \right) = \frac{1}{N},
\]

since \( \mathbf{E}(x_i^2) = \mathbf{E}(x_1^2) \) for all \( i = 1, 2, \ldots, N \) and \( \sum_i x_i^2 = 1 \) on \( S^{N-1} \).

Exercise. Show that the Euclidean volumes (Lebesgue measures) of all domains \( U \subset B^N(1) \) are concentrated at their boundaries:

the proportion of the measure of a \( U \subset B^N(1) \) in the set of points in \( U \) with distance \( \geq \delta \) from \( \partial U \) is \( \leq |1-\delta|^N \), where the equality holds only for \( U = B^N(1) \).

This is easy; what is non-trivial is the validity of the same inequality

\[
\frac{\text{vol} \{ x \in U : \text{dist}(x, \partial U) \geq \delta \}}{\text{vol}(U)} \leq |1-\delta|^N
\]

for all domains \( U \subset \mathbb{R}^N \) with \( \text{vol}(U) \leq \text{vol} \left( B^N(1) \right) \).
This, which is called the (Euclidean) isoperimetric inequality, must be on the list of ten (may be five) greatest theorems in geometry (in all of mathematics)?

Balls and spheres, unlike cubes, are not products of other spaces but they are closely related to the Gaussian cube, that is the power space

\[
(\mathbb{R}, G(x)dx)^N = \left(\mathbb{R}^N, \prod_{i=1}^N G(x_i)dx_i\right)
\]

where

\[
G(x) = \exp\left(-\frac{x^2}{2}\right)
\]

and where the Cartesian powers \(G^N = \prod_{i=1}^N G(x_i)dx_i\) of this measure, as well as the \((a, b)\)-scaled measures \(aG^N(bx)\), are also called Gaussian.

Although obvious, it is remarkable that, despite being a product space, this "cube" is fully rotationally symmetric:

the measure \(G^N\) is invariant under the orthogonal group.

The following two relations between the Gaussian and spherical measures, which probably, were known to Laplace, make the basis of Maxwell’s kinetic gas theory of 1860s.

1. **Approximation of the Euclidean measure in the ball by the Gaussian one.** Let

\[
\beta_n = \text{vol}(B^N(\sqrt{N}))
\]

and

\[
\gamma = (2\pi)^{1/2} = \int_{\mathbb{R}^N} \exp\left(-\frac{x^2}{2}\right)dx
\]

Then the normalised volumes of all domains \(U \subset B^N(\sqrt{N})\) are approximately equal to their normalised Gaussian measures:

\[
\beta_N^{-1}\text{vol}(U) = \gamma^{-1} \int_{\mathbb{R}^N} e^{-\sum_i \frac{x_i^2}{2}}dx_1...dx_N \leq \varepsilon_N \rightarrow 0.
\]

This relation, shows, in particular, that the concentration property of the Euclidean measure on the balls (hence on spheres) follows from the law of large numbers for Gaussian random variables, where, observe, this law is obvious, since the sums of independent Gaussian variables are again Gaussian.

This means, geometrically, that the projections of the measure

\[
G^N = e^{-\sum \frac{x_i^2}{2}}dx_1...dx_N
\]

to all lines in \(\mathbb{R}^N\) are Gaussian, which is obvious due to the rotational symmetry of the measure \(G^N\).

23The numero uno theorem in mathematics is the Pythagorean \(a^2 + b^2 = c^2\), while Bernoulli’s law of the large numbers competes with Leibniz’ \(1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + ... = \frac{\pi}{4}\) for being the second greatest one.
2. The push forward of the normalised Lebesgue measure (Euclidean volume) in the ball $B^N(\sqrt{N})$ under the coordinate projection on the $x_1$-axes converges, for $N \to \infty$, to the normalised Gaussian measure

$$\gamma^{-1} \exp -\frac{x_1^2}{2} dx_1.$$ 

This, for instance, allows a reduction of concentration properties of the Gaussian measures in $\mathbb{R}^n$ to that of the spherical one as we shall explain below.

The proof of both 1 and 2, if you think a little, are childishly simple – do it yourself or consult [Metric Properties].

**Compression of Measures**

Given two random variables, think of them as as probability measures $\mu$ and $\nu$ on the real line $\mathbb{R}$ and say that $\mu$ is *more concentrated* or *less spread* than $\nu$, in writing

$$\mu \prec \text{conc} \nu,$$

if $\mu$ is obtained by a compression of $\nu$ that is a distance non-increasing map $\mathbb{R} \to \mathbb{R}$, where "obtained" means being the push-forward under this map.

**Example.** Let $\mu$ be the Lebesgue measure on the segment $[0, 1]$ and $\nu = a(x)dx$ is a measure on $\mathbb{R}$ with a continuous density function $a$. Then (obviously) $\mu$ is more concentrated than $\nu$, if and only if $\sup_x a(x) \leq 1$.

It is not literally true that if independent random variables $\mu_i$ are more concentrated than, also mutually independent, $\nu_i$ then the sums of $\mu_i$ are more concentrated than these of $\nu_i$, because the compression maps, being non-linear, don’t commute with summation.

However, this becomes true with the laws of large numbers extended to non-linear functions of random variables by proving the measure concentration property for relevant classes of functions on power space, for which we refer to [Concentration of Measure], [Concentration Property] and [Metric Structures].

**Archimedean Map.**

More than 2 000 years ago, Archimedes proved that (what we now call) the push-forward of the spherical measure under a coordinate projection $S^2 \to [-1, 1]$ is

*equal, up to a scale factor to the Lebesgue measure*,\(^{24}\) which, this was explained to me by Michael Atiyah many years ago, is an instance of a *moment map*, where the leading example directly generalising Archimedes' map is as follows.

\(^{24}\)This is uncontested the greatest number two theorem in geometry and, arguably, in all of math.
Map the complex space $\mathbb{C}^{N+1}$ to the positive "quadrant" $\mathbb{R}^{N+1}$ by

$$A = A_N : (z_0, ..., z_N) \mapsto (|z_0|^2, ..., |z_N|^2)$$

and observe that the unit sphere $S^{2N+1} \subset \mathbb{C}^N$ goes to the (probability) $N$-simplex

$$\Delta^N = \{x_i \geq 0\}_i x_i \in \mathbb{R}^{N+1}.$$  

The remarkable feature of this $A$, which follows by the Pythagorean theorem applied to the differential $dA$, is that

the map $A$ sends the spherical measure to a multiple of the Lebesgue measure on the simplex.

(The Archimedean case follows by observing that $A_1 : S^3 \to \Delta^1$ factors via the Hopf map $S^3 \to S^2 \to \Delta^1 = [0, \sqrt{2}]$.)

The existence of such an $A$ shows, in particular, that

the Lebesgue measure on the regular $N$-simplex $\Delta^N$ is concentrated at least as much as the spherical measure on $S^{2N+1}$.

(In fact, $\Delta^N$ is much more concentrated than $S^{2N+1}$, see [Metric Structures] and references therein.)

1.4 Boltzmann and Entropy

2 ........

3 ........

4 ........

5 ........

6 Selected References


(more to come)