(a simple introduction to classical and)

Quantum Information Theory

Salvatore Cosentino
CMAT & DM, UMinho

"QUANTUM DAYS", Braga, April 11-12, 2019.

Some motivations

Understanding quantum mechanics

Understanding statistical mechanics of quantum systems

Control large quantum systems

. . . as, for example, quantum computers

Cryptography

. . .

Sources & References

J. Preskill, *Quantum Computation*, Lectures notes, Caltech, 2018.

M.A. Nielsen & I.L. Chuang, *Quantum Computation and Quantum Information*, Cambridge, 2000.

M.M. Wilde, *Quantum Information Theory*, Cambridge, 2013.

M.M. Wilde, *From Classical to Quantum Shannon Theory*, arXiv, 2016.

Messages and communication

Information theory is the creation of Claude Shannon[1],

whose insights were later made rigorous by Khinchin, McMillan, Breimann, ...

It deals with messages and communication.

The two main questions that he addressed and solved are

How much can we compress a message without losing its meaning

and

How much redundancy must we incorporate into a message in order to reliably transmitting it through a noisy channel?

[1] C.R. Shannon, A mathematical theory of communication, *The Bell System Technical Journal* **27** (1948), 379-423 and 623-656.

Roman inscriptions & graffiti

HELVIVM SABINVM AEDILEM D(IGNVM) R(EI) P(VBLICAE) V(IRVM) B(ONVM) O(RO) V(OS) F(ACIATIS)

```
"Please elect Elvio Sabino as a aedile, worthy of the state, a good one"
```

Our languages are redundant!

Brasilians say

`PORTUGAU`

while you say

`PRTGL`

They also say

`OI! TUDO JOIA?`

and you say

`Q TL?`

Messages

According to Wiener, reasonable models of messages/languages are stochastic processes, families

$$X_1\, X_2\, X_3\, \dots$$

of random variables X_k, parametrized by time $k \in \mathbb{N}$, with values in some finite set/alphabet, as for example

$$\mathsf{X} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \dots, \mathtt{z}\}$$

A realization of the processes is a finite or infinite word

$$x_1 x_2 x_3 \dots$$

in the letters of the alphabet, i.e. with $x_k \in \mathsf{X}$, as for example

```
"Ha em Lisboa um pequeno numero de restaurantes ou
casas de pasto ..."
```

Classical probability

The law of a random variable X with a finite number of values, say

$$|\mathsf{X}| = d$$

is a probability vector

$$p = (p_{\mathtt{a}}, p_{\mathtt{b}}, p_{\mathtt{c}}, \ldots, p_{\mathtt{z}})$$

of non-negative numbers

$$p_x \geq 0$$

with sum

$$p_{\mathtt{a}} + p_{\mathtt{b}} + p_{\mathtt{c}} + \cdots + p_{\mathtt{z}} = 1$$

i.e. a point in the unit simplex

$$\Delta^{d-1} \subset \mathbb{R}^d_+$$

Sources as Bernoulli trials

A very naive model of a source emitting a message is Bernoulli trials: independent copies of a fixed random variable X.

This means that the probability of observing/producing a finite word of lenght n is a product

$$\mathrm{Prob}(x_1 x_2 \dots x_n) = p_{x_1}\, p_{x_2}\, \dots\, p_{x_n}$$

Physicists call it a (classical) ensemble

$$\{x\,,\, p_x\}$$

Mathematicians speak of Cartesian products $\mathcal{X}^n$ of a finite probability space

$$\mathcal{X} = (\mathsf{X}, p)$$

Shannon's uncertainty function

The main character of classical Information Theory is Shannon's uncertainty function

$$\boxed{H(X) := -\sum_{x \in \mathsf{X}} p_x \log p_x}$$

of the source X with values in X and law p.

```
"You should call it entropy, for two reasons.  In
the first place your uncertainty function has been
used in statistical mechanics under that name, so
it already has a name.  In the second place, and
more important, no one really knows what entropy
really is, so in a debate you will always have the
advantage."
```

John von Neumann to Claude Shannon

Unicyclist

Boltzmann's entropy

According to Boltzmann's epitaph

$$\boxed{S = k \,\log W}$$

the entropy of a macroscopic system is proportional to the logarithm of the "thermodynamische Wahrscheinlichkeit" W, the number of microscopic states compatible with the macroscopic state of the system.

The magic, or mistery, is that this formula is not a definition, but an equality between two apparently different things!

His insight is that the Clausius' entropy[2], the thermodynamical potential measured according to

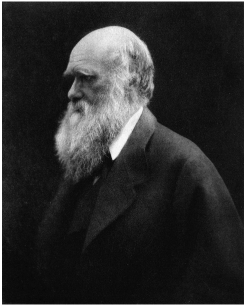
$$\Delta S = \int \frac{\delta Q}{T}$$

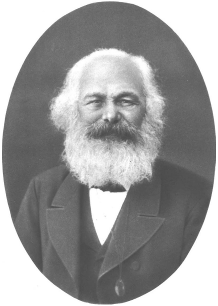
has actually a statistical/probabilistic interpretation.

[2] R. Clausius, Ueber verschiedene für die Anwendung bequeme Formen der Hauptgleichungen der mechanischen Wärmetheorie, *Annalen der Physik* **125** (1865), 353-400,

Wien 1944 - Duino 1906

(mais barbudos dos anos '60)





Shannon entropy as uncertainty

The entropy is bounded by

$$0 \leq H(X) \leq \log d$$

It is maximal $= \log d$ if all the d letters are equally probable, and it is minimal $= 0$ when one of the letters has total probability.

The unit is a bit, a random variable X taking values in $\{0, 1\}$ with uniform probability, which has entropy (taking base 2 logarithms)

$$H(X) = 1$$

The entropy is subadditive

$$H(XY) \leq H(X) + H(Y)$$

with equality holding iff X and Y are independent.

This allows to define the entropy rate of a process (X_k) as

$$H := \lim_{n \to \infty} \frac{1}{n} H(X_1 X_2 \ldots X_n)$$

Microscopic versus macroscopic

If you throw a dice, or measure the squared speed of a few molecules of gas, you don't see anything interesting.

Probability show itself as an asymptotic/macroscopic observable.

For example, if you throw a large number, say $n \sim 10^4$, of dices with d faces, and count the number N_n of times that you obtain one of them, you see the law of large numbers and the central limit theorem

$$N_n \simeq p\, n \pm \sqrt{pq}\,\sqrt{n}$$

with $p = 1/d$ and $q = 1 - p$.

Also, if you measure the mean squared speed of something like 10^{23} molecules of gas, you see the Maxwell-Boltzmann distribution

$$\alpha\, e^{-\beta v^2}$$

Entropy & typical words

Similarly, entropy is an asymptotic observable which shows itself when you take a global look at long words.

Typical words are those $w \in \mathsf{X}^n$ with a number of letters dictated by the law of large numbers

$$\left|\ \text{letters}\, x\, \text{contained in the word}\ \ w \in \mathsf{X}^n \right|\ \sim n\, p_x$$

It happens that typical words have roughly all the same probability

$$\mathrm{Prob}\,(\text{typical word}) \sim p_{\mathtt{a}}^{np_{\mathtt{a}}}\, p_{\mathtt{b}}^{np_{\mathtt{b}}}\, p_{\mathtt{c}}^{np_{\mathtt{c}}}\ \cdots\ p_{\mathtt{z}}^{np_{\mathtt{z}}} = 2^{-nH}$$

and the set of typical words has cardinality

$$|\text{typical words}| \sim 2^{nH}$$

Therefore, the set of typical words has almost total probability

$$\mathrm{Prob}\,(\text{typical words }) \sim 1$$

More precisely, with epsilons and deltas

For any $\delta > 0$, one defines the space of δ-typical words

$$\mathsf{T}^n_\delta \subset \mathsf{X}^n$$

as the set of those words $x_1 x_2 \dots x_n$ of lenght n having probability

$$2^{-n(H+\delta)} \leq |\mathrm{Prob}(x_1 x_2 \dots x_n)| \leq e^{-n(H-\delta)}$$

and prove that for any $\varepsilon > 0$, as small as we want, we can take the lenght n so large that

$$\mathrm{Prob}\;(\mathsf{T}^n_\delta) \geq 1 - \varepsilon$$

and

$$(1 - \varepsilon)\, 2^{n(H-\delta)} \leq |\mathsf{T}^n_\delta| \leq 2^{n(H+\delta)}$$

Entropy & compression rate

If the entropy is maximal, all the

$$|\mathsf{X}^n| = 2^{n \log d}$$

words of lenght n are equally probable, therefore typical.

Otherwise, almost all the probability is concentrated in an exponentially smaller set of typical words, with cardinality

$$\sim 2^{nH}$$

and all those words have roughly the same probability.

We can transmit all of them using words of lenght m in the same alphabet if

$$2^{nH} \sim 2^{m \log d}$$

Therefore, we can achieve a compression rate

$$R = \frac{m}{n} \sim \frac{H}{\log d}$$

Noiseless coding theorem

Noiseless coding theorem. *The maximal compression rate of a source using* d *letters and having entropy* H *is the relative entropy*

$$E = \frac{H}{\log d}$$

Namely, one can achieve any compression rate $R < E$*, and no compression rate* $R > E$*, with almost no loss of information in the limit where the lenght of the message go to infty.*

Asymptotic equipartition property

The core of Shannon's argument is the existence of typical words.

While it is easy for Bernoulli trials, it is a deep result for other correlated stochastic processes (X_n), as Markov chains.

Shannon-McMillan-Breiman theorem. *Let (X_n) be a stationary ergodic process with entropy rate H. Then*

$$-\frac{1}{n}\log p(X_1, X_2, \ldots, X_n) \to H$$

a.s. and in L^1.

Modern proofs use the ergodic theorem and the martingale convergence theorem.

Mutual information

Subadditivity of the entropy and monotonicity

$$H(XY) \geq H(X)$$

(which does not hold in the quantum context!)

suggest to define the conditional entropy of Y given X as

$$H(Y|X) := H(XY) - H(X)$$

which is ≥ 0.

The mutual information is the symmetric difference

$$\begin{aligned} I(X;Y) &:= H(X) + H(Y) - H(XY) \\ &= H(Y) - H(Y|X) \end{aligned}$$

which is minimal $= 0$, when X and Y are independent,
and maximal $= H(X) = H(Y)$, when X and Y are deterministically correlated, say $Y = f(X)$.

Noisy channels

$$\xrightarrow{\text{message}} \boxed{\texttt{ENCODER}} \xrightarrow{X_1X_2\ldots X_n} \boxed{\texttt{CHANNEL}} \xrightarrow{Y_1Y_2\ldots Y_n} \boxed{\texttt{DECODER}} \xrightarrow{\text{mxssage}}$$

Messages $t_1t_2\ldots t_m$ of lenght m are encoded in sequences of lenght n

$$x_1x_2\ldots x_n$$

The channel produces a (possibly) corrupted output, which are other sequences

$$y_1y_2\cdots y_n$$

according to certain conditional probabilities

$$p(y|x)$$

The output is finally decoded to produce a received message $s_1s_2\ldots s_m$, hopefully not so different from the original message.

The transmission rate is

$$R = \frac{m}{n}$$

(which is clearly $R \leq 1$).

Noisy channel coding theorem

Since each of the $\sim 2^{nH(X)}$ typical X-words may be corrupted in a number $\sim 2^{nH(Y|X)}$ of typical Y-words, we may reliably transmit messages with rate R if

$$2^{nH(X)}\, 2^{nH(Y|X)} \leq 2^{nH(Y)}$$

i.e. if

$$2^m \sim 2^{nR} \sim 2^{nH(X)} \leq \frac{2^{nH(Y)}}{2^{nH(Y|X)}} = 2^{nI(Y;X)} \leq 2^{nC}$$

Noisy channel coding theorem. *The maximal transmission rate of a discrete memoryless noisy channel is the capacity*

$$C := \sup_{\textit{law of } X} I(Y;X)$$

One can reliably transmit information at any rate $R < C$, and cannot reliably transmit information at any rate $R > C$.

Hidden symmetries of the entropy

The Boltzmann/Shannon function

$$B(p) := \sum_{k=1}^{d} p_k \log p_k$$

defined on the unit simplex $\Delta^{d-1} := \{p : \sum p_k = 1\} \subset \mathbb{R}_+^d$, has very poor symmetries, just the symmetric group S_d, permuting the vertices.

However, its Hessian defines the Fischer (Riemannian) metric

$$\sum_{k=1}^{n} \frac{dp_k^2}{p_k}$$

on Δ^{d-1}, which happens to be much more symmetric. Indeed, as observed by Gromov[3], it is a metric with constant sectional curvature !

To unveil this fact, and discover the hidden symmetries of classical probability, we must change coordinates, "blow up" the simplex, ...

[3] M. Gromov, In a Search for a Structure, Part 1: On Entropy (2013).

Blow up

The square map

$$\rho_k \mapsto p_k = \rho_k^2$$

sends $\mathbb{S}_+^{d-1} \to \Delta^{d-1}$, and the pull-back of the Fisher metric is just 4 times the Euclidean metric of the sphere, since

$$\frac{dp_k\, dp_k}{p_k} = 4\, d\sqrt{p_k}\, d\sqrt{p_k} = 4\, d\rho_k\, d\rho_k$$

We may then add phases, hence consider probability densities

$$z_k = \rho_k e^{i\varphi_k}$$

The square map extends naturally to the map

$$z_k \mapsto p_k = |z_k|^2$$

sending $\mathbb{C}^d \to \mathbb{R}_+^d$.

The Euclidian metric on $\mathbb{S}_+^{d-1}$ extends to the Fubini-Study Kahler metric on the projective Hilbert space $\mathbb{C}^d/\mathbb{C}^\times$.

Quantum world

This complexification led us to the world of Quantum Mechanics [4] [5] [6].

We have now a Hilbert space

$$\mathcal{H} \approx \mathbb{C}^d$$

with its linear and Hermitian structures,

and the full unitary group $U(n)$ of its symmetries.

So, how does complex/quantum probability looks like?

[4] H. Weyl, *Gruppentheorie und Quantenmechanik*, Leipzig, 1928.
[5] P.A.M. Dirac, *The principles of quantum mechanics*, Oxford, 1930.
[6] J. von Neumann, *Mathematishe Grundlagen der Quantenmechanik*, Berlin, 1932 [*Mathematical Foundations of Quantum Mechanics*, Princeton, 1955]

Interference

If something can happen in two mutually exclusive ways, with probabilities p and q, classical probabilities add

$$p + q$$

In the quantum/complex world we introduced phases.

When we add two probability densities like

$$\alpha = \sqrt{p}\, e^{i\theta} \qquad \text{and} \qquad \beta = \sqrt{q}\, e^{i\phi}$$

and then compute the square modulus of $\alpha + \beta$, we get interference

$$|\alpha + \beta|^2 = p + q + 2\sqrt{pq}\, \cos(\theta - \phi)$$

For example, this explains interference patterns in the double-slit experiment.

Quantum probability

Atomic measures (vertices of the unit symplex) extend to rays

$$\mathbb{C}\,|\psi\rangle \subset \mathcal{H}$$

that physicists call pure states and, since we want to exploit the linear structure of the Hilbert space, identify with rank-one projectors

$$P_\psi = |\psi\rangle\,\langle\psi|$$

or with the corresponding quadratic form $|\varphi\rangle \mapsto \langle\varphi|P_\psi|\varphi\rangle$.

Non-atomic measures extend to convex combinations of pure states

$$\rho = p_\psi P_\psi + p_\varphi P_\varphi + \ldots$$

with $p_\psi + p_\varphi + \cdots = 1$, that physicists call mixed states.

Observe that $\mathrm{Tr}(\rho) = 1$, since this is the trace of each P_ψ.

States & measures

A state ρ, mixed or pure (for a mathematician, a self-adjoint positive operator with unit trace), assigns a probability

$$p = \langle e_1 | \rho | e_1 \rangle + \langle e_2 | \rho | e_2 \rangle + \dots$$

to each subspace $\mathcal{E} \subset \mathcal{H}$, where $e_1, e_2, \dots$ is any orthonormal basis of $\mathcal{E}$.

Therefore assigns a (classical) probability vector

$$\mathcal{E}_1 \oplus \mathcal{E}_2 \oplus \dots \mapsto (p_1, p_2, \dots)$$

to each orthogonal direct sum decomposition

$$\mathcal{H} = \mathcal{E}_1 \oplus \mathcal{E}_2 \oplus \dots$$

Physicists call them projection valued measures (PVM) or von Neumann projective measurements.

Superposition principle

According to the superposition principle, if it is possible to prepare a system in both states $|\psi\rangle$ and $|\phi\rangle$, then it is also possible to prepare the system in the superposition

$$\alpha\,|\psi\rangle + \beta\,|\phi\rangle \ ,$$

with arbitrary complex coefficients α and β.

Two states are distinguishable if there exists some (possibly ideal) experience that let us decide whether the system is in one or the other state. This is codified by the notion of orthogonality.

Thus, states of a quantum system belong to a Hilbert space $\mathcal{H}$, a complex linear space equipped with an inner product $\langle\phi|\psi\rangle$.

Actually, states are rays $\mathbb{C}\,|\psi\rangle$ in $\mathcal{H}$, since a global factor, or phase if we consider only unitary states, is not observable

$$|\psi\rangle \sim e^{i\theta}\,|\psi\rangle$$

Qu(antum)bits

The smallest non-trivial quantum system is described by the Hilbert space

$$\mathcal{H} \approx \mathbb{C}^2$$

We may call $|0\rangle$ and $|1\rangle$ the elements of an orthonormal basis, so that a generic state is a superposition

$$|\psi\rangle = \alpha\,|0\rangle + \beta\,|1\rangle$$

with complex coefficients α and β.

These are the units, the building blocks, of quantum computers. As such, they are called qubits[7].

A concrete example is polarization of photons, which may be left or right polarized, hence may be in one of the orthogonal states

$$|\circlearrowleft\rangle \qquad \text{or} \qquad |\circlearrowright\rangle$$

[7] B. Schumacher, Quantum coding, *Physical Review A* **51** (1995), 2738-2747.

Observables and observations

Observables are self-adjoint linear operators defined on $\mathcal{H}$.

An observable A has a spectral resolution

$$A = \sum_k \alpha_k \, |\alpha_k\rangle \langle\alpha_k|$$

with real eigenvalues α_k and corresponding unitary eigenstates $|\alpha_k\rangle$.

Observation of the observable A on the unitary state $|\psi\rangle = \sum_k \psi_k \, |\alpha_k\rangle$ will give one (and only one) of the possible values α_k's, with probability

$$p_k = |\psi_k|^2 = |\, \langle\alpha_k|\psi\rangle \,|^2 \, .$$

The mean value of the observable A in the unitary state $|\psi\rangle$ is

$$\langle A\rangle_\psi = \langle\psi|A|\psi\rangle = \sum_k \alpha_k \, |\psi_k|^2 \, .$$

Projection valued measurements

Following von Neumann, we may think that a measurement is an orthogonal direct sum decomposition

$$\mathcal{H} = \mathcal{E}_1 \oplus \mathcal{E}_2 \oplus \cdots \oplus \mathcal{E}_n$$

(the proper spaces of an observable).

Equivalently, a family of pairwise orthogonal projections E_k of $\mathcal{H}$ onto the $\mathcal{E}_k$'s, satisfying $\sum_k E_k = I$.

If a system is in the unitary state $|\psi\rangle$, the probability to observe the outcome associated to the subspace $\mathcal{E}_k$ is equal to the squared norm

$$p_k = \| E_k \, |\psi\rangle \, \|^2 = \langle\psi|E_k|\psi\rangle$$

of its projection.

If such observation occurs, then the state of the system collapses to the normalized state proportional to $E_k \, |\psi\rangle$.

Randomness & disturbance

Intrinsic randomness of Q.M.: the observed value of an observable is one of its possible values α_k with certain probabilities (which are all we can compute).

Once the value α_k of the observable A is observed, the state of the system collapses from an initial state $|\psi\rangle$ to the eigenvector/state $|\alpha_k\rangle$ corresponding to the observed value.

The collapse is ascribed to the interaction of the quantum system with a classical macroscopic device.

Thus, to get information from a quantum system we must disturb it!

Linearity of Q.M. is far from intuitive!

For example, one may ask, following Shrödinger, what is the meaning of a superposition like

$$|\mathrm{cat}\rangle = \frac{1}{\sqrt{2}}\,|\mathrm{dead}\rangle + \frac{1}{\sqrt{2}}\,|\mathrm{alive}\rangle$$

The mainstream interpretation holds that such a state is indeed possible, but highly improbable.

The cat interacts continuously with the world around it (other cats, rats, children, granmothers, ...), so she is constantly "measured" by macroscopic devices, and therefore collapsed in one and only one of the two states.

Dynamics

Dynamics also is linear.

The time evolution of an isolated quantum systems is given by a group of unitary operators

$$U_t = e^{-itH/\hbar}$$

where H is the Hamiltonian, an observable which plays the role of the energy, and $\hbar \simeq 1.055 \times 10^{-34}$ J·s is the reduced Planck constant.

The state at time t of a system which has been prepared in the state $|\psi(0)\rangle$ at time 0 is therefore

$$|\psi(t)\rangle = e^{-itH/\hbar} |\psi(0)\rangle$$

which is the solution of the Schrödinger equation

$$i\hbar \frac{d}{dt} |\psi\rangle = H \, |\psi\rangle$$

with initial condition $|\psi(0)\rangle$.

Multiparticle systems & tensor products

A consequence of the superposition principle and the probabilistic interpretation of the square modulus of the coefficients, is that multiparticle systems are described by tensor products

$$\mathcal{H}_X \otimes \mathcal{H}_Y \otimes \dots$$

of the state spaces $\mathcal{H}_X$, $\mathcal{H}_Y$, ... of their components.
A basis of the tensor product is made o products $|x_i\rangle \otimes |y_j\rangle \otimes \dots$ of basis states of each factor, and inner products (of pure tensors) are products $\langle x|x'\rangle \cdot \langle y|y'\rangle \cdots$

The dimension of tensor products grows exponentially with the number of components. For example, the Hilbert space of a few hundreds qubits has a dimension

$$2^{300} \sim 10^{90}$$

which is much larger than the estimated number 10^{80} of baryons in the Universe!

Quantum computers

Ideally, a Quantum Computer works with a certain number n of qubits. It is prepared in some initial/input state

$$|\psi\rangle = \sum_{x_n \dots x_2 x_1 = 0}^{2^n - 1} \psi_{x_n \dots x_2 x_1} |x_n \dots x_2 x_1\rangle$$

in the tensor product $\mathcal{H}^{\otimes n}$ of $\mathcal{H} \approx \mathbb{C}^2$, where $x_k \in \{0, 1\}$ and

$$|x_n \dots x_2 x_1\rangle := |x_n\rangle \otimes \cdots \otimes |x_2\rangle \otimes |x_1\rangle$$

The initial state is acted upon by a certain number of gates, which are unitary operators U_k acting on one, two, or more qubits, producing a final/output state

$$|\psi_f\rangle = \dots U_k \dots U_2 U_1 \; |\psi\rangle$$

We can eventually measure some or all of the qubits, and therefore collapse the state and get classical information.

We cannot observe states ...

In classical Information Theory, we get for granted that we may read bits and get all the information they carry.

Clearly, at least in principle, we may prepare a quantum system in any quantum state $|\psi\rangle$.

(for example letting light passing through a polarized lens)

However, an observer with no a priori knowledge cannot infer the prepared state $|\psi\rangle$ from her measurements!

She may test whether the state is $|\varphi\rangle$ or not, the fidelity being

$$F := |\langle\psi|\varphi\rangle|^2$$

Or, given many copies of the unknown state $|\psi\rangle$, she may measure its projections on some orthonormal frame $|\varphi_1\rangle, |\varphi_2\rangle, \ldots$, i.e. observe the probabilities/frequencies

$$p_k = |\langle\psi|\varphi_k\rangle|^2$$

. . . or acquire informations without disturbing!

Suppose we want to distinguish between the two states $|\psi\rangle$ and $|\varphi\rangle$ of a system (described by the Hilbert space) $\mathcal{H}$ without disturbing the system.

We couple them with some fixed state $|0\rangle$ of a second system $\mathcal{H}_R$, and apply a unitary transformation U to the composite system $\mathcal{H} \otimes \mathcal{H}_R$, sending

$$|\psi\rangle \otimes |0\rangle \mapsto |\psi\rangle \otimes |a\rangle \qquad \text{and} \qquad |\varphi\rangle \otimes |0\rangle \mapsto |\varphi\rangle \otimes |b\rangle$$

Unitarity forces

$$\langle\psi|\varphi\rangle = \langle\psi|\varphi\rangle \cdot \langle a|b\rangle$$

If $|\psi\rangle$ and $|\varphi\rangle$ are not orthogonal, $|a\rangle$ and $|b\rangle$ represent the same state!

(this may be a resource in quantum cryptography!)

Heisenberg uncertainty principle

If the two observable A and B does not commute, i.e. if

$$[A, B] := AB - AB \neq 0$$

then they are not simultaneously diagonalizable.

This is the case of position and momentum operators, defined as

$$(Qf)(x) := x\, f(x) \qquad (Pf)(x) := -i\hbar \frac{\partial f}{\partial x}(x)$$

which satisfy $[P, Q] = i\hbar I$.

As discovered by Heisenberg[8] this implies a lower bound on the product of the standard deviations, known as Heisenberg uncertainty principle

$$\Delta P \cdot \Delta Q \geq \hbar/2$$

---

[8] W. Heisenberg, Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, *Zeitschrift für Physik*, **43** (1927) 172-198.

Entropic uncertainty principle

Suppose we prepare the system in a state ρ, e.g. a pure state $|\psi\rangle\langle\psi|$.

Non-commuting observables A and B define different orthogonal (eigenspaces) decompositions corresponding to the eigenvectors $|\alpha_k\rangle$ and $|\beta_k\rangle$, respectively, to which the state ρ associates different probability vectors p and q, respectively.

Heisenberg uncertainty principle is a consequence of the stronger[9] [10] [11]

Entropic uncertainty inequality.

$$H(p) + H(q) \geq \log(1/c) + S(\rho)$$

where $c = \sup |\langle\alpha_i|\beta_j\rangle|^2$.

[9] I.I. Hirschman Jr., A note on entropy, *Amer. J. of Math.* **79** (1957), 152-156.

[10] D. Deutsch, Uncertainty in quantum measurement, *Phys. Rev. Lett.* **50** (1983), 631-633

[11] H. Maassen and J. B. M. Uffink, Generalized entropic uncertainty relations, *Phys. Rev. Lett.* **60**. (1988), 1103-1106.

No cloning

Roughly speaking, cloning means producing a state $|\psi\rangle \otimes |\psi\rangle$ out of a state $|\psi\rangle$. The reverse operation is called deleting.

In classical computation, we get for granted that we can clone or delete (but, according to Landauer principle, this costs some entropy/energy!)

There is a conflit between linearity and reversibility of Q.M. and cloning or deleting, which are non-linear and irreversible operations!

> No-cloning theorem. *There exists no unitary operator U on $\mathcal{H} \otimes \mathcal{H}$ s.t.*
>
> $$U(|\psi\rangle \otimes |\varphi\rangle) = e^{i\alpha(\psi,\varphi)} \, |\psi\rangle \otimes |\psi\rangle$$
>
> *for all normalized states $|\psi\rangle$ and $|\varphi\rangle \in \mathcal{H}$ and some phases $\alpha(\psi, \varphi)$.*

It is clear that if $U_t = e^{-itH}$ is a unitary operator perfoming cloning, the time reversal $U_t^\dagger = e^{itH}$ performs deleting, and viceversa. Therefore, the no-cloning theorem is also a no-deleting theorem.

Cloning conflits with linearity

Suppose we have a linear operator which is able to clone both the states $|\psi\rangle$ and $|\varphi\rangle$, i.e. to produce the pure states

$$|\psi\rangle \otimes |\psi\rangle \qquad \text{and} \qquad |\varphi\rangle \otimes |\varphi\rangle$$

out of them.

We may apply it to the superposition

$$\alpha\,|\psi\rangle + \beta\,|\varphi\rangle$$

Linearity would give the state

$$\alpha\,|\psi\rangle \otimes |\psi\rangle + \beta\,|\varphi\rangle \otimes |\varphi\rangle \ .$$

This is clearly different from cloning the superposition, i.e. from

$$(\alpha\,|\psi\rangle + \beta\,|\varphi\rangle) \otimes (\alpha\,|\psi\rangle + \beta\,|\varphi\rangle)$$

Cloning conflits with unitarity

Suppose we have a unitary operator cloning the two states $|\psi\rangle$ and $|\varphi\rangle$ (e.g. after coupling/tensoring the two with a fixed state $|0\rangle$).

Then the inner product

$$\langle\psi|\varphi\rangle$$

should be equal to the inner product between

$$|\psi\rangle \otimes |\psi\rangle \qquad \text{and} \qquad |\varphi\rangle \otimes |\varphi\rangle$$

which is

$$\langle\psi|\varphi\rangle^2$$

But, according to the Cauchy-Schwartz inequality, the identity

$$\langle\psi|\varphi\rangle = \langle\psi|\varphi\rangle^2$$

happens only when the states are equal (since proportional vectors define the same states) or when the states are orthogonal.

Entanglement

What makes quantum probability so weird[12] is the phenomenon called entanglement (`entrelaçamento`) by Shrödinger [13].

The archetypal example is the state

$$|\psi\rangle = \frac{1}{\sqrt{2}}\,|0\rangle \otimes |1\rangle + \frac{1}{\sqrt{2}}\,|1\rangle \otimes |0\rangle$$

Observation of one of the particles produces the collapse of the global wave function, and therefore determines the state of the other particle!

[12] `"If you think you understand quantum mechanics, you don't understand quantum mechanics"`, according to Feynman.

[13] E. Schrödinger, Discussion of Probability Relations Between Separated Systems, *Proceedings of the Cambridge Philosophical Society* **31** (1935), 555-563 and (1936), 446-451.

Non-local correlations

Einstein called it spooky action (`ação fantasmagórica`) at a distance.

With one of his famous Gedankenexperiment[14], he tried to illustrate non-completeness of (the Copenhagen interpretation of) Q.M.

Two particles in an entangled state, like

$$|\psi\rangle = \frac{1}{\sqrt{2}} |0\rangle \otimes |1\rangle + \frac{1}{\sqrt{2}} |1\rangle \otimes |0\rangle$$

may be separated by a huge distance, and yet, observation/collapse of one of the particles determines instantaneously the state of the other!

Nowadays we know, thanks to Bell's inequalities[15], that such non-local correlations cannot be explained with the existence of (classical) hidden variables.

And they are observed!

---

[14] A. Einstein, N. Rosen and B. Podolsky, Can Quantum-Mechanical Description of Physical Reality be Considered Complete? *Phys. Rev.*. **47** (1935), 777-780.

[15] J. Bell, On the Einstein-Poldolsky-Rosen paradox, *Physics* **1** (1964), 195-200.

Teleportation

If Maria and João share an entangled state, say an EPR pair

$$\frac{1}{\sqrt{2}} |0\rangle \otimes |0\rangle + \frac{1}{\sqrt{2}} |1\rangle \otimes |1\rangle$$

(i.e., she owns the first qubit and he owns the second qubit),

Maria may send a (unknown to her!) third quantum state

$$\alpha |0\rangle + \beta |1\rangle$$

in her possess to João, transmitting only 2 bits with a classical channel.

Thus,

$$\boxed{1 \text{ EPR } + 2 \text{ bits } \geq 1 \text{ qubit}}$$

Teleportation protocol

Indeed[16], the joint state

$$(\alpha |0\rangle + \beta |1\rangle) \otimes \left(\frac{1}{\sqrt{2}} |0\rangle \otimes |0\rangle + \frac{1}{\sqrt{2}} |1\rangle \otimes |1\rangle \right)$$

is (proportional to) a sum of 4 orthogonal states

$$(|00\rangle \pm |11\rangle) \otimes (\alpha |0\rangle \pm \beta |1\rangle) \quad \text{and} \quad (|10\rangle \pm |01\rangle) \otimes (\beta |0\rangle \pm \alpha |1\rangle)$$

With a projective measurement on the first two qubits (which she owns), Maria may collapse the state into one of the 4 possibilities.

She may communicate, using 2 bits through a classical channel, the collapsed state, and João may use this information to apply the appropriate unitary transformation and recostruct the original state

$$\alpha |0\rangle + \beta |1\rangle$$

from his qubit.

[16] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, W. K. Wootters, Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels, *Phys. Rev. Lett.* **70** (1993), 1895-1899.

Superdense coding

Conversely, Maria and João may use a shared EPR pair, and 1 qubit to encode 2 classical bits.

She performs the unitary transformation corresponding to the 2 bits that she wants to communicate,

and he measures the state of his qubit

Thus,

$$\boxed{1 \text{ EPR } + 1 \text{ qubit } \geq 2 \text{ bits}}$$

Ensembles & density matrices

According to von Neumann, a statistical ensemble/mixed state
of unitary states $|\psi_k\rangle$ (not necessarily orthogonal or even independent)
with (classical) probabilities p_k
is conveniently described by a density operator/matrix

$$\rho = \sum_k p_k \, |\psi_k\rangle \, \langle\psi_k|$$

The mean value of the observable A on the state ρ is

$$\langle A\rangle = \mathrm{Tr}(\rho A)$$

Abstractly, a density operator is a positive semi-definite self-adjoint operator with unit trace.

For example, unpolarized light is described by the mixed state

$$\frac{1}{2}\,|\circlearrowleft\rangle\,\langle\circlearrowleft| + \frac{1}{2}\,|\circlearrowright\rangle\,\langle\circlearrowright| \quad \approx \quad \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Density for entangled states

As observed by Landau, density operators also appear naturally when we describe a subsystem of an entangled system.

If we have a pure state

$$|\psi\rangle \in \mathcal{H}_X \otimes \mathcal{H}_R$$

and an observable A acting on $\mathcal{H}_X$, then the mean vaue of the extended operator $A \otimes I$ on the state $|\psi\rangle$ is

$$\langle\psi|A \otimes I|\psi\rangle = \mathrm{Tr}(\rho_X A)$$

if we define the mixed state

$$\rho_X := \mathrm{Tr}_R(|\psi\rangle \langle\psi|)$$

Purification

Conversely, any mixed state in $\mathcal{H}_X$, as

$$\rho_X = \sum_x p_x \left|\psi_x\right\rangle \otimes \left\langle\psi_x\right|$$

may be seen as above, as the marginal of a pure state

$$\left|\psi\right\rangle = \sum_x \sqrt{p_x} \left|\psi_x\right\rangle \otimes \left|r_x\right\rangle$$

called purification of ρ_X, in a larger system $\mathcal{H}_X \otimes \mathcal{H}_R$,
where the $\left|r_x\right\rangle$'s form an orthonormal basis of $\mathcal{H}_R$.

Indeed,

$$\rho_X = \mathrm{Tr}_R \left(\left|\psi\right\rangle \left\langle\psi\right|\right)$$

(remember the passage from classical to quantum probability!)

Marginals & POVM

More generally, if we have a (possibly mixed) state

$$\rho_{XY}$$

in the Hilbert space $\mathcal{H}_X \otimes \mathcal{H}_Y$ of a composite system,
and observe quantities depending only on the first subsystem $\mathcal{H}_X$, we may just consider the marginal

$$\rho_X := \mathrm{Tr}_Y(\rho_{XY})$$

which is a density matrix on $\mathcal{H}_X$.

Similarly, a projective measurement

$$\mathcal{H}_X \otimes \mathcal{H}_Y = \mathcal{E}_1 \oplus \mathcal{E}_2 \oplus \dots$$

as seen from (the system described by the Hilbert space) $\mathcal{H}_X$,
is a positive-operator valued measure (POVM), that is, a family $\mathcal{F} = \{F_k\}$ of positive-semidefinite self-adjoint operators F_k such that

$$\sum_k F_k = I$$

von Neumann entropy

Following a gedankenexperiment (computing the work needed to separate a bipartite gas using semi-permeable walls ...), von Neumann showed that the thermodinamical entropy of an ensemble of quantum states described by the density matrix ρ must be defined according to

$$S(\rho) := -\mathrm{Tr}(\rho \log \rho)$$

If the p_k's are the (nonnegative) eigenvalues of $\rho = \sum_k p_k |\psi_k\rangle \langle\psi_k|$, with $\sum_k p_k = 1$, this is simply

$$S(\rho) = -\sum_k p_k \log p_k$$

Thus, the von Neumann entropy of a state ρ is the Shannon entropy of the spectrum (spectral measure) of ρ.

little Jancsi, “enfant prodige”

von Neumann's is the physical entropy!

The state ρ which maximizes the entropy $S(\rho)$ once fixed the energy, the mean value $E = \mathrm{Tr}(H\rho)$ of the Hamiltonian, is the Gibbs state

$$\rho = \frac{1}{Z(\beta)}\, e^{-\beta H}$$

where $\beta = 1/T$ and the partition function is

$$Z(\beta) := \mathrm{Tr}(e^{-\beta H})$$

The Gibbs state may be rewritten $\rho = e^{-\beta(H-F)}$, where the free energy is

$$F(\beta) := -T \log Z(\beta)$$

Thus

$$F = E - TS$$

and therefore it is minimized by the Gibbs state.

Elementary properties of the von Neumann entropy

If $\mathcal{H} \approx \mathbb{C}^d$, the von Neumann entropy is bounded by

$$0 \leq S(\rho) \leq \log d$$

It is minimal $= 0$ iff ρ is a pure state $\rho = |\psi\rangle \langle\psi|$,
and it is maximal $= \log d$ when $\rho = \frac{1}{d} I$.

It is unitarily invariant,

$$S(U \rho U^\dagger) = S(\rho)$$

It is subadditive

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B)$$

with equality when $\rho_{AB} = \rho_A \otimes \rho_B$.

Stranger properties of the von Neumann entropy

The von Neumann entropy is not monotone !

All we can say is the triangle inequality [17]

$$|S(\rho_A) - S(\rho_B)| \leq S(\rho_{AB})$$

For example, the entropy of a (pure) entangled state ρ_{AB} is zero, while the entropy of the marginals $\rho_A = \mathrm{Tr}_B(\rho_{AB})$ and $\rho_B = \mathrm{Tr}_A(\rho_{AB})$ are equal (since, by the Schmidt decomposition, marginals of a pure state share the same spectrum) and positive.

There follows that conditional entropies may be negative !

And also suggests that entangled states can be used to store information non-locally !

[17] H. Araki and E.H. Lieb, Entropy inequalities, *Comm. Math. Phys.*. **18** (1970), 160-170.

Entropy for EPR pair

For example, we may consider our favourite entangled state, the EPR pair

$$\frac{1}{\sqrt{2}}|0\rangle \otimes |1\rangle + \frac{1}{\sqrt{2}}|1\rangle \otimes |0\rangle$$

which is described by a density matrix ρ_{AB} which is the rank-one projector onto the pure state, therefore with zero entropy.

The marginals are

$$\rho_A = \rho_B = \frac{1}{2}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and have entropy

$$S(\rho_A) = \log 2$$

corresponding to one classical bit.

In the words of Schrödinger: `"the best possible knowledge of a whole does not necessarily include the best possible knowledge of all its parts".`

Information carried by qubits

If we send one bit, the receiver may observe it.

On the other side, if we send one qubit, say

$$\alpha |0\rangle + \beta |1\rangle$$

the receiver, performing just one measurement, has no direct access to the coordinates α and β.

How many bis are contained in a qubit ?

More generally, how much classical information can be transmitted sending quantum states ?

Sending bits using qubits

Maria sends a sequence $x_1 x_2 \dots x_n$ drawn from a quantum ensemble $\{\rho_x, p_x\}$, with density

$$\rho = \sum_{x \in \mathcal{X}} p_x \, \rho_x$$

João, who knows the sender's ensemble, performs measures $\mathcal{E} = \{E_y\}$ on the received quantum states ρ_x , and get as output a realization $y_1 y_2 \dots y_n$ of a random variable $Y = \{y, p_y\}$. Here

$$p_y = \sum_x p_x \, p(y|x)$$

and

$$p(y|x) = \mathrm{Tr}(\rho_x E_y)$$

Information gain & accessible information

Before the measurements, João ignorance about the signal is

$$H(X)$$

(since he knews the ensemble!).

After the measurements, his ignorance is reduced to

$$H(X|Y) = H(XY) - H(Y)$$

Thus, his information gain is

$$I(X;Y) = H(XY) - H(Y) - H(X)$$

The accessible information is the maximal information gain

$$\mathrm{Acc}(\mathcal{R}) := \max_{\mathcal{E}} I(X;Y)$$

over all possible measurements $\mathcal{E}$.

Holevo bound

If the states ρ_x's are mutually orthogonal, they can be distinguished by a measurement, and therefore we are in a classical situation. The accessible information is

$$\mathrm{Acc}(\mathcal{R}) = S(\rho) = H(X)$$

However, if the states ρ_x overlap, the best we can say is

Holevo theorem. *The accessible information is bounded above by*

$$\mathrm{Acc}(X) \leq \chi(\rho)$$

where the Holevo information[18] is

$$\chi(\rho) := S(\rho) - \sum_{x\in\mathcal{X}} p_x\, S(\rho_x)$$

[18] A.S. Holevo, Bounds for the quantity of information transmitted by a quantum communication channel, *Problems of Information Transmission* **9** (1973), 177-183.

Bits contained in qubits

Thus, if we use pairwise orthogonal pure states ρ_x's, we may send all the classical information contained in n bits using n qubits.

On the other side, this is the best we can do, since

$$\chi(\rho) \leq S(\rho) \leq \log |\mathsf{X}|$$

Therefore,

n qubits $\leq$ n bits *The accessible information contained in n qubits is not larger than the classical information contained in n bits!*

on the proof of Holevo bound

Holevo bound depends on a nontrivial fact, strong subadditivity [19]

$$S(\rho_{ABC}) + S(\rho_C) \leq S(\rho_{AC}) + S(\rho_{BC})$$

which is equivalent to both

$$S(\rho_A|\rho_{BC}) \leq S(\rho_A|\rho_B)$$

and

$$I(\rho_A;\rho_{BC}) \geq I(\rho_A;\rho_B)$$

both obvious in the classical case, where conditional entropies are all non-negative!

[19] E.H. Lieb, M.B. Ruskai, Proof of the Strong Subadditivity of Quantum Mechanical Entropy, *J. Math. Phys.*. **14** (1973), 1938-1941.

Quantum sources

A source $\{x, p_x\}$ sends her messages using as letter quantum states $|x\rangle$, with $x \in \mathsf{X}$, in some state space $\mathcal{H} \approx \mathbb{C}^d$.

Thus, she sends the messages $x_1 x_2 \dots x_n$ as quantum states

$$|x_1\rangle \otimes |x_2\rangle \otimes \dots |x_n\rangle$$

chosen according to the mixed state $\rho^{\otimes n}$ where

$$\rho = \sum_{x \in \mathsf{X}} p_x \, |x\rangle \langle x|$$

is the mixed state describing the quantum ensenble $\{p_x, |x\rangle\}$:

Compression rate

To save space or computational resources, we want to encode or store the messages using a minimal number of qubits, allowing, if necessary, block-coding.

Clearly, if the $|x\rangle$'s are pairwise orthogonal, we can distinguish them with a projective measurement, and we are in a classical situation treated by Shannon noiseless coding theorem.

Otherwise, we want to compress the message using, possibly, less, say

$$nR \quad \text{qubits}$$

(the natural unit), hence coding messages in a Hilbert space

$$\mathcal{H} \approx \mathbb{C}^{2^{nR}}$$

The compression rate is R, the number of qubits per letter.

von Neumann entropy & typical subspaces

The density matrix ρ has a spectral resolution

$$\rho = \sum_{y=1}^{d} q_y \left|y\right\rangle \left\langle y\right|$$

where $q = (q_1, q_2, \ldots, q_d)$ is the law of a random variable Y having Shannon entropy

$$S(\rho) = H(q)$$

which is clearly smaller that $H(p)$.

Quantum states

$$\left|y_1\right\rangle \otimes \left|y_2\right\rangle \otimes \cdots \otimes \left|y_n\right\rangle$$

corresponding to typical n-sequences $y_1 y_2 \ldots y_n$ of the r.v. Y span a typical subspace $\mathcal{T}^n \subset \mathcal{H}^{\otimes n}$ of dimension

$$\dim(\mathcal{T}^n) \sim 2^{n\,S(\rho)}$$

in general much smaller than $\dim(\mathcal{H}^{\otimes n}) = 2^{n \log d}$.

Schumacher compression

Schumacher's[20] [21] idea consists in encoding only the messages which project onto $\mathcal{T}^n$, thus using $\sim nR$ qubits.

If T^n denotes the orthogonal projection onto the typical subspace $\mathcal{T}^n$, it follows from classical theory that this happens with almost total probability, i.e.

$$\mathrm{Tr}\left(T^n\,\rho^{\otimes n}\right) \simeq 1$$

We then decode the message and get some (in general mixed) state $\sigma_{\mathbf{x}} \in \mathcal{H}^{\otimes n}$ in the original state space.

There follows from Shannon noiseless coding theorem that the average fidelity

$$\overline{F} = \sum_{\mathbf{x}\in \mathsf{X}^n} p_{\mathbf{x}}\, \langle \mathbf{x}|\sigma_{\mathbf{x}}|\mathbf{x}\rangle$$

can be made arbitrarily near to one for sufficiently large n.

---

[20] B. Schumacher, Quantum coding, *Phys. Rev. A* **51** (1995), 2738-2747.

[21] R. Jozsa and B. Schumacher, A new proof of the quantum noiseless coding theorem, *J. Modern Optics.* **41** (1994), 2343-2349.

Noiseless coding theorem

Thus, the von Neumann entropy also measures the minimal number of qubits per letter necessary to reliably encode a message made of quantum states.

Schumacher compression theorem. *The maximal compression rate of a source sending states $|x\rangle$ with probabilities p_x is the von Neumann entropy*

$$S(\rho)$$

of the mixed state $\rho = \sum p_x |x\rangle \langle x|$.

Quantum noisy channels

Here is where things become interesting,

What is "quantum capacity"?

and hard.

Thanks

Obrigado!