Quantum ML-randomness and the Shannon-McMillan-Breiman Theorem

André Nies, University of Auckland
Joint works with Volkher Scholz, Marco Tomamichel

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

Algorithmic questions in dynamical systems,
Toulouse, March 2018

- The Shannon-McMillan-Breiman Theorem recovers the entropy of an ergodic measure from almost every single point. I will explain the classical theorem, obtained 1949-1959, and its algorithmic versions based on randomness, obtained from 2009 (Hochman, Hoyrup).

- I will explain quantum bits (qubits), and finite sequences of them. Because of entanglement, one needs density operators, which are statistical superpositions of finite sequences of the same length.

- Infinite sequences of qubits are formalised by coherent sequences of density operators. They are states on a suitable C^* algebra.

- Introduce an algorithmic notion of randomness for states (with Volkher Scholz, arXiv:1709.08422, 2017). Work towards an algorithmic version of the SMB theorem in the quantum setting. Do the i.i.d. case. (With Tomamichel.)

Quantum bits and sequences of quantum bits

Quantum bits

- A classical bit can be in states $0, 1$. Write them as $|0\rangle$, $|1\rangle$.
- A qubit is a physical system with two classical states. E.g.
  - polarisation of photon horizontal/vertical,
  - hydrogen atom with electron in basic/excited state.
- A qubit is in a superposition of the two classical states:

  $$\alpha \mid 0\rangle + \beta \mid 1\rangle,$$

  $\alpha, \beta \in \mathbb{C}$, $|\alpha|^2 + |\beta|^2 = 1$. E.g. $\alpha = 2/\sqrt{5}, \beta = -i/\sqrt{5}$.
- Measurement of a qubit w.r.t. standard basis $|0\rangle, |1\rangle$ yields 0 with probability $|\alpha|^2$, and 1 with probability $|\beta|^2$.
- Measurement forces the system to settle on a classical state.

Hilbert spaces and their tensor products

- The state of a physical system is represented by a vector in a finite-dimensional Hilbert space.
- $\langle a|b\rangle$ denotes the inner product of vectors a, b, linear in the second component and antilinear in the first.
- For systems A, B, the tensor product $A \otimes B$ is a Hilbert space that represents the combined system.
- One defines an inner product on $A \otimes B$ by

$$\langle a \otimes b|c \otimes d\rangle = \langle a|c\rangle\langle b|d\rangle.$$

Finite sequences of quantum bits

- Mathematically, a qubit is simply a unit vector in $\mathbb{C}^2$. The (pure) state of a system of n qubits is a unit vector in the tensor power
$$(\mathbb{C}^2)^{\otimes n} := \underbrace{\mathbb{C}^2 \otimes \ldots \otimes \mathbb{C}^2}_{n}.$$
- We denote the standard basis of $\mathbb{C}^2$ by $|0\rangle, |1\rangle$. The standard basis of $(\mathbb{C}^2)^{\otimes n}$ is given by n-bit strings: it consists of vectors
$$|a_1 \ldots a_n\rangle := |a_1\rangle \otimes \ldots \otimes |a_n\rangle.$$
- The state of the system of n qubits is a linear superposition of them. Example: $n = 2$, "maximally entangled" state
$$\tfrac{1}{\sqrt{2}}|00\rangle + \tfrac{1}{\sqrt{2}}|11\rangle.$$

Mixed states, or density operators

So far we have considered "pure" states $|\psi\rangle$ viewed as unit vectors in $(\mathbb{C}^2)^{\otimes n}$. Let $|\psi\rangle\langle\psi|$ denote the orthogonal projection onto the subspace spanned by $|\psi\rangle$, fixing $|\psi\rangle$.

A mixed state is a convex linear combination

$$\textstyle\sum_{i=1}^{2^n} p_i |\psi_i\rangle\langle\psi_i|$$

for pairwise orthogonal pure states ψ_i. E.g. $\frac{1}{3}|0\rangle\langle 0| + \frac{2}{3}|1\rangle\langle 1|$ is a mixed state where $n = 1$.

Recall that for an operator S on A, the trace is

$$\mathsf{Tr}(S) = \text{sum of diagonal of } S = \text{sum of eigenvalues of } S.$$

A mixed state is the same as a positive Hermitean operator S on $(\mathbb{C}^2)^{\otimes n}$ with $\mathsf{Tr}(S) = 1$ (aka density operator).

Partial trace $T_B \colon L(A \otimes B) \to L(A)$

Recall: Given systems (finite dimensional Hilbert spaces) A, B, the tensor product $A \otimes B$ is a Hilbert space that represents the combined system. $L(A)$ denotes the space of the linear operators on A.

The partial trace T_B is the unique linear operator $L(A \otimes B) \to L(A)$ such that for $R \in L(A), S \in L(B)$, we have $T_B(R \otimes S) = R \cdot \mathsf{Tr}(S)$.

- Example: Let $A = B = \mathbb{C}^2$. The partial trace T_B corresponds to deleting the last qubit. E.g. $T_B(|10\rangle\langle 10|) = |1\rangle\langle 1|$.
- Let's consider again the EPR state $\frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, now viewed as projection β in $L(A \otimes B)$.
- We have $T_B(\beta) = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|$ which is a mixed state!

  Mixed states are necessary to be able to delete a qubit.

Erasing last qubit of a finite sequence

Recall: $(\mathbb{C}^2)^{\otimes n}$ has a base consisting of the vectors $|\sigma\rangle$, for σ a string of n classical bits.

Let us arithmetically describe the partial trace operation $T_n \colon M_{2^{n+1}} \to M_{2^n}$ (“erase the last qubit”).

Bit strings are seen as numbers written in “reverse” binary: the least significant digit on the left.

For a $2^{n+1} \times 2^{n+1}$ matrix $M = (a_{\sigma r, \tau s})$ where $|\sigma|, |\tau| = n$, r, s are bits, $N = T_n(M)$ is given by the $2^n \times 2^n$ matrix

$$b_{\sigma,\tau} = a_{\sigma 0, \tau 0} + a_{\sigma 1, \tau 1}.$$

Infinite coherent sequences of density operators

M_{2^n} denotes the set of $2^n \times 2^n$ matrices over $\mathbb{C}$. We have a partial trace operation $T_n \colon M_{2^{n+1}} \to M_{2^n}$ ("erase the last qubit").

"Quantum Cantor space" $S(M_\infty)$ consists of the sequences $(\rho_n)_{n\in\mathbb{N}}$ of density operators in M_{2^n} such that $T_n(\rho_{n+1}) = \rho_n$ for each n.

- This is the set of states (positive linear functionals of norm 1) on the computable C^* algebra $M_\infty = \lim_n M_{2^n}$, known as the CAR algebra (for "canonical anticommutation relations").
- $S(M_\infty)$ is compact in the weak-$*$ topology (the weakest topology for which all the maps $\rho \to \rho(a)$, $a \in M_\infty$, are continuous), and has a convex structure.
- It also has dynamics (shift map) T, erasing the first qubit.

Embed Cantor space into quantum Cantor space

Recall: Given a $2^{n+1} \times 2^{n+1}$ matrix $M = (a_{\sigma r, \tau s})$ where $|\sigma|, |\tau| = n$, and r, s are bits. $N = T_n(M)$ is given by the $2^n \times 2^n$ matrix

$$b_{\sigma,\tau} = a_{\sigma 0, \tau 0} + a_{\sigma 1, \tau 1}.$$

- A classical bit sequence Z turns into $(\rho_n)_{n \in \mathbb{N}}$ where the bit matrix $B = \rho_n \in M_{2^n}$ satisfies $b_{\sigma,\tau} = 1 \iff \sigma = \tau = Z \upharpoonright n$.
- If all the ρ_n are diagonal matrices, we describe a measure on Cantor space. Classical bit sequences are Dirac measures.

Shannon-McMillan-Breiman theorem

1949 — 1955 — 1959

see Logic Blog 2017, linked from my web site, for more background

A.e. convergence of empirical entropy

Let $\mathbb{A}$ be an alphabet, and let ω denote an element of $\mathbb{A}^\infty$. The "log-likelihood" random variables are defined by

$$h_n(\omega) = -\frac{1}{n} \log_2 \mu[\omega \restriction n.]$$

The SMB theorem says that the entropy of an ergodic measure can be seen from almost every trajectory ω as the limit of the (normalised) empirical entropies $h_n(\omega)$.

It relies on a lemma that holds for T-invariant measures in general:

Let μ be an invariant measure for the shift operator T on the space $\mathbb{A}^\infty$. Then for μ-a.e. x, $h(x) = \lim_n h_n(x \restriction n)$ exists.

Ergodic measure

Recall that μ is ergodic iff every μ-integrable function f with $f \circ T = f$ is constant μ-a.s. An equivalent condition: for $u, v \in \mathbb{A}^*$,

$$\lim_n \frac{1}{n} \sum_{k=0}^{n-1} \mu([u] \cap T^{-k}[v]) = \mu[u]\mu[v].$$

Bernoulli measure on $\mathbb{A}^\infty$ is ergodic (and in fact, strongly mixing). For ergodic μ, the entropy $h(\mu)$ is defined as $\lim_n H_n(\mu)$, where

$$H_n(\mu) = -\frac{1}{n} \sum_{|w|=n} \mu[w] \log \mu[w].$$

- In other words, $H_n(\mu) = \mathbb{E}_\mu h_n$.
- $H_{n+1}(\mu) \leq H_n(\mu) \leq 1$ so that the limit exists.

The SMB theorem

“For ergodic measures, μ-a.s. the empirical entropy equals $h(\mu)$.”

Theorem (SMB theorem)

Let μ be an ergodic invariant measure for the shift operator T on the space $\mathbb{A}^\infty$. For μ-a.e. ω we have $\lim_n h_n(\omega) = h(\mu)$.

Proof: We already know that for μ-a.e. ω, $h(\omega) = \lim_n h_n(\omega \upharpoonright n)$ exists.

- Now, first one checks that since μ is T-invariant, we have $h(T\omega) \leq h(\omega)$ for each ω.
- Next, from the Poincare recurrence theorem it follows that $B = \{x \colon h(Tx) < q < h(x)\}$ is a null set for each q (because we can't return to B outside a null set). So h is actually invariant: $h(T\omega) = h(\omega)$ for μ-a.e. ω.
- Also, $h \in L^1(\mu)$ by the dominated convergence theorem. So if μ is ergodic then $h(\omega)$ has some constant value, for μ-a.s. ω.

Instructive: a direct proof of SMB in the i.i.d. case

Let μ be a Bernoulli measure on $\mathbb{A}^\infty$, giving probability p_i to the event that symbol a_i is in a particular position, $\mathbb{A} = \{a_1, \ldots, a_k\}$. For μ-a.e. ω we have $\lim_n h_n(\omega) = h(\mu)$.

- By $k_{i,n}(\omega)$ we denote the number of occurences of the symbol a_i in $\omega \restriction n$.
- By independence, we have

$$h_n(\omega) = -\frac{1}{n} \log \prod_i p_i^{k_{i,n}(\omega)} = -\frac{1}{n} \sum_i k_{i,n}(\omega) \log p_i.$$

- By the strong law of large numbers, for μ-a.e. ω, $k_{i,n}(\omega)/n$ converges to p_i.

An algorithmic version of the SMB theorem

We now assume that we can compute $\mu[u]$ uniformly from a string u. This holds e.g. for the Bernoulli measure when the p_i are all computable reals.

Thm (Hochman 2009, Hoyrup 2012)
Let μ be a computable ergodic invariant measure for the shift operator T on the space A^∞.
If ω is μ-Martin-Löf random, then $\lim_n h_n(\omega) = h(\mu)$.

Quantum SM(B?) theorem

2004, recent

Again see Logic Blog 2017 for more background

Recall quantum setting

- We only consider the analog of the case of binary alphabet.
- "Quantum Cantor space" consists of the state set $\mathcal{S}(M_\infty)$, which is a convex, compact, connected set.
- Given a finite sequence of qubits, "deleting" a particular one generally results in a statistical superposition of the remaining ones.
- This is why $\mathcal{S}(M_\infty)$ consists of coherent sequences of density matrices in $M_{2^n}(\mathbb{C})$ (which formalise such superpositions) rather than just of sequences of unit vectors in $(\mathbb{C}^2)^{\otimes n}$.
- Dynamics is given by shift operator (which we interpret as deleting the first qubit).

Notation: given state ρ, write ρ_n for $\rho \restriction M_{2^n}$

Quantum Shannon-McMillan theorem

- Bjelakovich et al.(2004) provided a quantum version of the Shannon-McMillan theorem, building on work of Hiai and Petz.

  (They worked with bi-infinite sequences, which makes little difference here, as a stationary process is given by its marginal distributions on the places from 0 to n, for all n.)
- They first convert the classical SM theorem into an equivalent form which doesn't directly mention measure; rather, they have "chained typical sets" which are coherent sequences of Shannon's typical sets. This is then transferred to the quantum setting.
- The reason they avoided the full Breiman version is that on $\mathcal{S}(M_\infty)$ there has been so far no reasonable way to say "for almost every".

Quantum ML-randomness

Let μ be a computable shift-invariant state on M_∞.

- A quantum Σ^0_1 set has the form $G = \langle p_n \rangle_{n\in\mathbb{N}}$, where $p_n \in M_{2^n}$ is a uniformly computable projection with algebraic matrix entries, and $p_n \le p_{n+1}$. For a state ρ, let $G(\rho) = \sup_n \mathsf{Tr}(\rho_n p_n) = \sup_n \rho(p_n)$ (the “set” is $[0,1]$-valued).
- A quantum Martin-Löf test relative to μ is a uniformly computable sequence $\langle G_r \rangle_{r\in\mathbb{N}}$ of quantum Σ^0_1 sets such that $\mu(G_r) \le 2^{-r}$.
- State ρ fails test $\langle G_r \rangle_{r\in\mathbb{N}}$ at level $\delta > 0$ if $G_r(\rho) > \delta$ for each r. Else ρ passes at level δ.
- State ρ is quantum Martin-Löf random if it passes each test $\langle G_r \rangle_{r\in\mathbb{N}}$ at each level: $\inf_r G_r(\rho) = 0$.

Fact. There is a universal quantum ML-test relative to μ.

Quantum ML-randomness: facts and examples

The tracial state τ is random, even though it corresponds to the uniform measure on Cantor space and hence, from a different point of view, is computable.

Theorem (with Scholz)

- Every ML-random bit sequence is quantum ML-random.
- We can generalise this to measures ψ (i.e. diagonal states): if $\psi(G_m) \to_m 0$ for each classical ML-test $\langle G_m \rangle_{m \in \mathbb{N}}$ then ψ is quantum ML random.

Following my suggestion, Tejas Bhojraj (a student of Joseph Miller in Madison) showed that the quantum analog of Solovay tests yields the same notion. Using this, he showed that the quantum ML-random states form a convex set. (What are the extreme points?)

Effective quantum SMB theorem?

A state μ on M_∞ is called ergodic if it is an extreme point on the convex set of shift invariant states.

- The von Neumann entropy of a density matrix S is $H(S) = -\mathsf{Tr}(S \log S)$.
- By concavity of $\log$ the following exists:

$$h(\mu) = \lim \frac{1}{n} H(\mu_n)$$

Conjecture (and some special cases are known)

Let μ be an ergodic computable state on M_∞. Let ρ be a state that is quantum ML-random with respect to μ. Then

$$h(\mu) = -\lim \frac{1}{n} \mathsf{Tr}(\rho_n \log \mu_n).$$

A pathway for settling the conjecture

$$h(\mu) = -\lim \frac{1}{n} \mathsf{Tr}(\rho_n \log \mu_n) \quad ???$$

Go through more and more general cases for both ρ and μ.

- The computable state μ can be the uniform measure, i.i.d. but quantum, a computable ergodic measure, and finally any computable ergodic state.
- The random state ρ can be a bit sequence that is ML random w.r.t. μ, a μ-random measure on $2^{\mathbb{N}}$, and finally any state that is quantum ML-random relative to μ.
- The combination that: ρ is a bit sequence, and μ a measure, is the effective version of classical SMB theorem.

The case of i.i.d μ but general μ-random ρ

- To say that μ (seen as a sequence of RV's) is i.i.d. means that for some fixed computable $v \in S(M_2)$, i.e. a 2×2 density matrix, we have $\mu \restriction M_{2^n} = v^{\otimes n}$.
- Note that the partial trace removes the "final v", so this "infinite tensor power $v^{\otimes \infty}$" indeed can be seen as a computable state on M_∞.
- There is a computable unitary $u \in M_2$ such that uvu^* is diagonal, with p, $1 - p$ on the diagonal, p is computable.
- The von Neumann entropy is $h(\mu) = -p \log p - (1 - p) \log(1 - p)$.

Theorem

For i.i.d. μ we have $-\lim_n \frac{1}{n} \mathsf{Tr}(\rho_n \log \mu_n) = h(\mu)$.

Preparation of proof: diagonalise

For i.i.d. μ we have $-\lim_n \frac{1}{n}\mathsf{Tr}(\rho_n \log \mu_n) = h(\mu)$.

- qML(μ)-randomness is closed under the unitary of M_∞ which is obtained applying conjugation by u^* "qubit-wise".
- So replacing ρ by its conjugate we may as well assume that the 2×2 density matrix v is diagonal.

Fix $\delta > 0$. Let $P_{n,\delta}$ be the projection in M_{2^n} corresponding to the set of bitstrings with empirical entropy close to $h(\mu)$:

$$\{x \colon |x| = n \wedge |-\tfrac{1}{n}\log \mu[x] - h(\mu)| \leq \delta\}.$$

$P^{\perp}_{n,\delta}$ corresponds to the other strings.

Preparation of proof: Chernoff bounds

For i.i.d. μ we have $-\lim_n \frac{1}{n}\mathsf{Tr}(\rho_n \log \mu_n) = h(\mu)$.

- Since μ is a product measure, $\log \mu[x]$ is a sum of n independent random variables looking at the single bits of x, and the expectation of $-\frac{1}{n}\log \mu[x]$ is $h(\mu)$.
- The usual Chernoff bound yields $\mu(P_{n,\delta}^{\perp}) \leq 2\exp(-2n\delta^2)$.
- Let $G_{m,\delta} = \bigcup_{n>m} P_{n,\delta}^{\perp}$ where these projectors are now viewed as clopen sets in Cantor space, so that $G_{m,\delta}$ determines a classical μ-ML-test.
- Since ρ is qML random w.r.t. μ, we have $\lim_m G_{m,\delta}(\rho) = 0$.

The case of i.i.d μ but general μ-random ρ

For i.i.d. μ we have $\lim_n -\frac{1}{n}\mathsf{Tr}(\rho_n \log \mu_n) = h(\mu)$.

Fix $\delta > 0$. We insert the identity term $I_{2^n} = P_{n,\delta}^{\perp} + P_{n,\delta}$ between the two factors. By linearity of the trace, the limit equals

$$\lim_n -\frac{1}{n}\mathsf{Tr}(\rho_n P_{n,\delta}^{\perp} \log \mu_n) + \lim_n -\frac{1}{n}\mathsf{Tr}(\rho_n P_{n,\delta} \log \mu_n).$$

We look separately at both resulting limits.

Left hand side

$$\underbrace{\lim_n -\frac{1}{n}\mathsf{Tr}(\rho_n P^{\perp}_{n,\delta}\log\mu_n)} + \lim_n -\tfrac{1}{n}\mathsf{Tr}(\rho_n P_{n,\delta}\log\mu_n).$$

For positive operators A, B we have $\mathsf{Tr}(AB) \leq ||A||_1 \cdot ||B||_\infty$ where $||A||_1$ is the sum of the eigenvalues, and $||B||_\infty$ is their maximum.

$$-\tfrac{1}{n}\mathsf{Tr}(\rho_n P^{\perp}_{n,\delta}\log\mu_n) \leq \tfrac{1}{n}||P^{\perp}_{n,\delta}\rho_n P^{\perp}_{n,\delta}||_1 \log\mu_n||_\infty.$$

- $||\frac{1}{n}\log\mu_n||_\infty$ has a bound that depends only on p
- for large enough n we have $||P^{\perp}_{n,\delta}\rho_n P^{\perp}_{n,\delta}||_1 \leq 2\delta$ since ρ passes the quantum ML test.

So the left hand limit is 0.

Right hand side

$$\lim_n -\tfrac{1}{n}\mathsf{Tr}(\rho_n P^{\perp}_{n,\delta} \log \mu_n) + \underbrace{\lim_n -\frac{1}{n}\mathsf{Tr}(\rho_n P_{n,\delta} \log \mu_n)}.$$

We show that the right hand limit is in the interval

$$[h(\mu)(1-2\delta), h(\mu)+\delta].$$

This goes to $h(\mu$ with $\delta \to 0$, as required.

- Use that μ_n is a diagonal matrix in M_{2^n} with $p^k(1-p)^{n-k}$ in the position (σ,σ), where the binary string σ of length n has k 0s.
- By definition of $P_{n,\delta}$, $||P_{n,\delta}(-\frac{1}{n}\log \mu_n) - h(\mu)P_{n,\delta}||_\infty \leq \delta$.

(See 2017 Logic Blog, page 16, for the full calculation.)

Random states satisfy the law of large number

The strong law of large numbers says that for $B(p)$ distributed i.i.d. random variables (X_i), $\frac{1}{n}\sum_0^{n-1} X_i$ goes to p almost surely.

Let μ be an i.i.d computable state with eigenvalues $p, 1-p$, and let ρ be quantum ML-random relative to μ.
For $i < n$ let $S_{n,i}$ be the subspace of $\mathbb{C}^{2^n}$ generated by those vectors $|\sigma\rangle$ with $\sigma_i = 1$. We have

$$\lim_n \frac{1}{n}\sum_{i<n} \mathsf{Tr}(\rho_n S_{n,i}) = p$$

($S_{n,i}$ is identified with its orthogonal projection).