The Shannon-McMillan theorem (AEP) for quantum sources and related topics

I.Bjelakovic, T.K., A. Szkola,
R.Siegmund-Schultze

Motivation

- Transfer of fundamental theorems of classical information theory to quantum information theory
- In a wider context: how a quantum ergodic theory and quantum dynamical system theory looks like

The classical Shannon-McMillan-(Breiman) theorem

- Given (Σ,μ,σ), Σ sequence space over finite alphabet, μ ergodic measure, σ shift-transformation, $\Sigma \ni x$, $x(n) = (x_1, x_2, x_3, \ldots, x_n)$
- a.s. for ergodic μ: the individual information rate equals the average information rate

$$\lim_{n\to\infty} \frac{-\log\mu\big(x(n)\big)}{n} = h_\mu \qquad \left(= \lim_{n\to\infty} \frac{-1}{n} \sum_{w\in\Sigma^{(n)}} \mu(w)\log\mu(w) \right)$$

- This is a law of large numbers under very mild assumptions

Typical subspaces and data compression

Reformulation in terms of typical subspaces:

there is a family of typical sets $\left\{T_n \subset \Sigma^{(n)}\right\}$ s.t.

$T_{n+1}(n) \supset T_n$ (filtration property) and

$\mu(T_n) \to 1$ and

$\frac{1}{n}\log \# T_n \to h_\mu$ and

$\forall \varepsilon > 0$ one has: $\mu(w \in T_n) \le e^{-n(h_\mu - \varepsilon)}$ for $n > n_0(\varepsilon)$

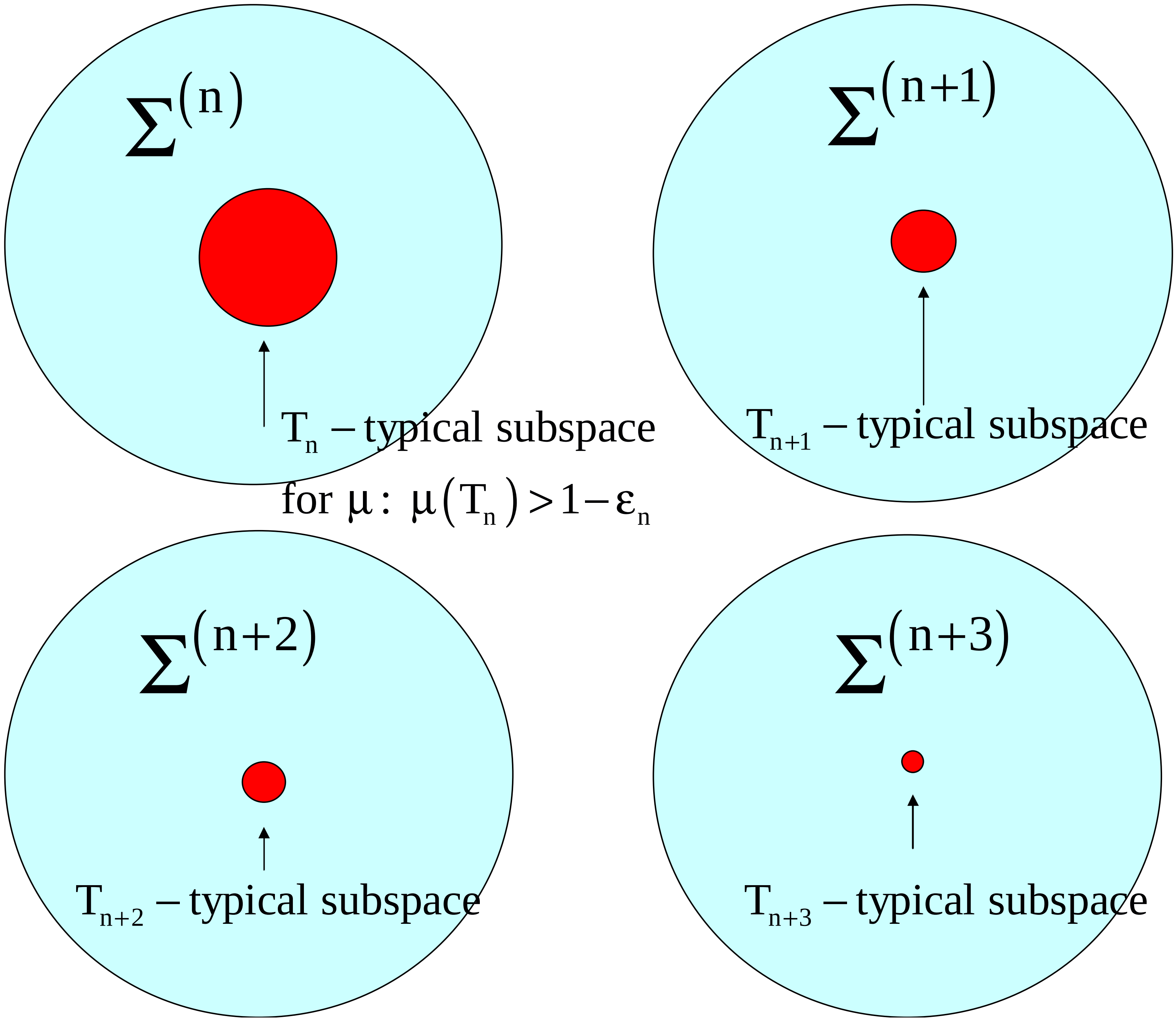
furthermore for any family $\{B_n\}$ s.t.

$\limsup \frac{1}{n} \log(\# B_n) < h_\mu$ it follows that

$\mu(B_n) \to 0$ (strong converse)

in other words: to cover a positive fraction of the whole space one needs asymptotically

e^{nh_μ} cylinder-sets of lengt n


$\Sigma^{(n)}$
T_n – typical subspace
for $\mu: \mu(T_n) > 1 - \varepsilon_n$
$\Sigma^{(n+1)}$
T_{n+1} – typical subspace
$\Sigma^{(n+2)}$
T_{n+2} – typical subspace
$\Sigma^{(n+3)}$
T_{n+3} – typical subspace

Application to data compression:

given a typical long symbol sequence $\left(x_1, x_2, \ldots\ldots, x_n\right) \in \{0,1\}^n$

Codebook: typical words of length k, $k < \frac{1}{h_\mu} \log_2 n$

there are about 2^{kh_μ} typical words $\Rightarrow$ Codebooksize □ n and

kh_μ bits needed to specify a word from the codebook

Splitting: $$\underbrace{\underbrace{x_1, x_2, \ldots, x_k}_{\text{code with } kh_\mu \text{ bits}} \; \underbrace{x_{k+1}, \ldots, x_{2k}}_{\text{code}} \; \ldots\ldots\ldots\ldots \; \underbrace{x_{k(n-1)+1}, \ldots, x_n}_{\text{code}}}_{\frac{n}{k} \text{ codewords}}$$

(only $o(n)$ fraction of blocks does not belong to the codebook)

$\Rightarrow \frac{n}{k} \cdot kh_\mu = nh_\mu$ bits needed to code the whole sequence ($h_\mu \in [0,1]$)

The quantum setting

A : matrix-algebra over Hilbert space $H=\mathbb{C}^{\kappa}$ (C^*-algebra)

A_x: copy of A at site x

$$A^{\infty} = \text{norm-closure of } \bigcup_n \left\{ A^n := \bigotimes_{x\in\{1,2,\ldots,n\}} A_x \right\}$$

σ: shift transformation

φ: positive, normed, linear functional on A^{∞} (measure)

φ invariant: $\varphi = \varphi \circ \sigma$

φ ergodic: φ is extremal among the invariant functionals for $\varphi_n := \varphi\big|_{A^n}$ there is a density matrix D_n

s.t. $\varphi(a) = \mathrm{tr}(D_n a)$ and $D_n = \mathrm{tr}_{n+1} D_{n+1}$ (consistency)

(tr_{n+1} partial trace with respect to site $n+1$)

entropy: $S(\varphi_n) = -\mathrm{tr}(D_n \log D_n)$ (von Neumann)

entropy rate: $s(\varphi) = \lim\limits_{n\to\infty} \frac{1}{n} S(\varphi_n)$

covering exponent: $\beta(\varepsilon)$:

$\lim\limits_{n\to\infty} \frac{1}{n} \min\{\log \mathrm{tr} P : P$ projector from A^n s.t. $\varphi(P) > 1-\varepsilon\}$

The quantum Shannon-McMillan theorem

(Ref.: Inventiones Mathematica, 2003)

Let φ be an ergodic state on A^{∞} $\Rightarrow$

$\exists$ family of orthogonal projectors $\{Q_n \in A^n\}$ s.t.:

i) $\varphi(Q_n) \to 1$ and $\lim\limits_{n\to\infty} \frac{1}{n} \log \operatorname{tr}(Q_n) = s(\varphi)$

ii) for any sequence of minimal projectors $\{p_n < Q_n\} \Rightarrow$

$$\frac{-1}{n} \log \varphi(p_n) \to s(\varphi)$$

iii) for any sequence of projectors $\{Q'_n \in A^n\}$ s.t.

$$\lim_{n\to\infty} \frac{1}{n} \log \operatorname{tr}(Q'_n) < s(\varphi) \Rightarrow \varphi(Q'_n) \to 0$$

Comments

- The theorem holds for $\mathbb{Z}^{\nu}$ - lattices as well
- Covering exponent is for all $\varepsilon > 0$: $\beta(\varepsilon) = s(\varphi)$
- The typical projectors (subspaces) can be explicitly constructed from the eigenspaces of D_n corresponding to eigenvalues of order $e^{-ns(\varphi)}$
- The relation between the typical subspaces for different n is still unclear
- Extensions to other group actions are possible
- The typical subspaces can be chosen to be universal (not depending on φ but only on $s(\varphi)$) due to a result by Kalchenkov

History

- Josza&Schumacher: typical subspace theorem for product states (Bernoulli case, 1996)
- Petz&Mosonyi: weak version of the Shannon-McMillan under the assumption of complete ergodicity (2001) and strong form for Gibbs states (with Hiai, 1993)
- Neshveyev&Størmer: Shannon-McMillan for finitely generated C*-algebras but only tracial states (2002)
- Datta&Shuchov: Shannon-McMillan for spin lattices with restrictions on the interaction (2002)

Extensions

A pointwise variant (Shannon-McMillan-Breiman):

Let φ be an ergodic state on A^{∞} $\Rightarrow$

$\forall \varepsilon > 0$ $\exists$ family of orthogonal projectors $\{Q_{n,\varepsilon} \in A^{n}\}$ s.t. for $n > n(\varepsilon)$:

i) $\varphi\left(Q_{n,\varepsilon}\right) > 1-\varepsilon$ and $\lim_{n\to\infty}\frac{1}{n}\log \mathrm{tr}\left(Q_{n,\varepsilon}\right) < s\left(\varphi\right)+\varepsilon$

ii) for any sequence of minimal projectors $\{p_n < Q_{n,\varepsilon}\} \Rightarrow$

$$\frac{-1}{n}\log\varphi\left(p_n\right) < s\left(\varphi\right)-\varepsilon$$

iii) $R[\mathrm{tr}_{n+1}(Q_{n+1,\varepsilon})] = Q_{n,\varepsilon}$ (here $R[.]$ is the range projector)

- The relation between the typical projectors for different ε is unclear
- For abelian algebras (classical case) the above theorem is equivalent to the Shannon-McMillan-Breiman theorem

A theorem for the relative entropy

(I.Bjelakovich, R.Siegmund-Schultze)

Relative entropy of two states ω and τ on finite dimensional algebra:

$$S(\omega,\tau) := \begin{cases} \mathrm{tr}\left(D_\omega\left(\log D_\omega - \log D_\tau\right)\right) & \text{for } \mathrm{supp}\,\omega \le \mathrm{supp}\,\tau \\ \infty & \text{otherwise} \end{cases}$$

Relative entropy rate:

ψ an invariant state and φ an invariant product state on A^∞

$$s(\psi,\varphi) = \lim_{n\to\infty} \frac{1}{n} S(\psi_n,\varphi_n) \quad \left(\varphi_n := \varphi|_{A^n}\right)$$

Relative exponent:

$\beta_{\varepsilon,n}(\psi,\varphi) := \min\{\log\varphi_n(Q) : Q \in A^n$, projector s.t. $\psi_n(Q) > 1-\varepsilon\}$

For ψ an ergodic state and φ an invariant product state on $A^\infty \Rightarrow$

$$\lim_{n\to\infty}\frac{1}{n}\beta_{\varepsilon,n}(\psi,\varphi) = s(\psi,\varphi) \text{ for } \forall\varepsilon > 0$$

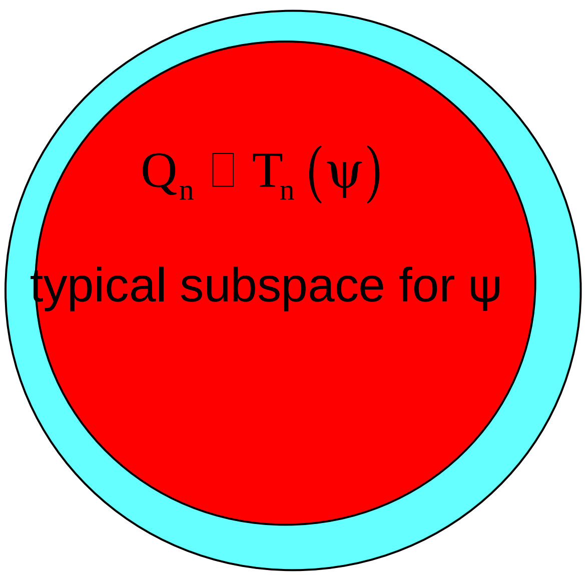
equivalently for typical subspace projectors $\{Q_n\}$ of $\psi \Rightarrow$

$$e^{-n(s(\psi,\varphi)+\varepsilon)} \le \varphi(Q_n) \le e^{-n(s(\psi,\varphi)-\varepsilon)} \quad ; n > n(\varepsilon)$$

Relative entropy typical and untypical subspaces

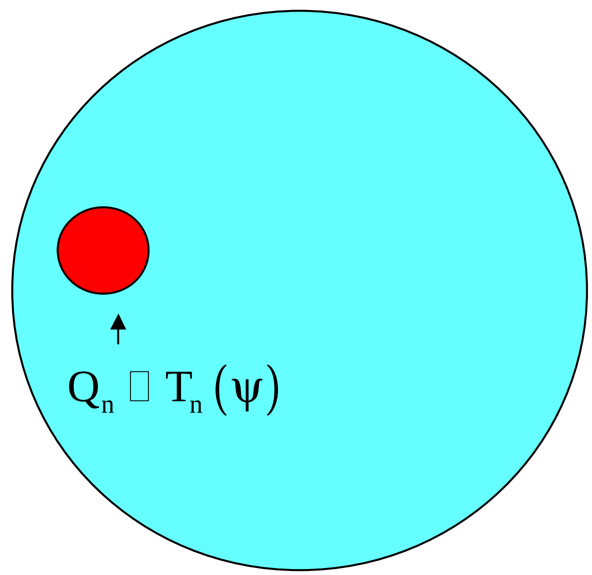
From ψ point of view:

$$\psi(Q_n) > 1 - \varepsilon$$



From φ point of view:

$$\varphi(Q_n) \approx e^{-ns(\psi,\varphi)}$$



- Complete analogy to the classical case
- The proof is similar to the one for the Shannon-McMillan theorem but more technical involved
- New simple proof of the monotonicity of the relative entropy can be derived from this result
- Starting point for developing a large deviation theory (Sanov’s theorem)

Proof strategy

Idea: want to use abelian approximations to lift the classical results to the quantum case

Natural candidate:

algebra B_n generated by the eigenspace projectors of D_n (density matrix corresponding to $\varphi_n := \varphi\big|_{\{1,..,n\}}$)

$$\underbrace{\underbrace{A \otimes A \otimes \ldots\ldots \otimes A}_{B_n \subset A^n} \underset{\otimes}{\otimes} \underbrace{A \otimes A \otimes \ldots\ldots \otimes A}_{B_n \subset A^n} \underset{\otimes}{\otimes} \ldots\ldots \underset{\otimes}{\otimes} \underbrace{A \otimes A \otimes \ldots\ldots \otimes A}_{B_n \subset A^n}}_{B_n^m \to B_n^\infty}$$

$\left(B_n^\infty, \varphi^{(B)}, \sigma^*\right)$ is an abelian system

σ^* corresponds to σ^n on A^∞

$\left(B_n^{\infty}, \varphi^{(B)}, \sigma^*\right)$ is isomorphic to a classical system $\left(\Sigma_{B_n}, \mu, \sigma\right)$

$$s(\varphi) \leq \frac{1}{n} s\left(\varphi^{(B)} \middle| \sigma^*\right) = \frac{1}{n} h_\mu \leq s(\varphi) + \varepsilon(n)$$

What can be said about the ergodic properties of $\left(\Sigma_{B_n}, \mu, \sigma\right)$?

μ is ergodic under the assumption of complete ergodicity of φ (Petz)

In the general case μ splits into at most $k|n$ ergodic components. All components are isomorphic under some shift-power and have the same entropy. To prove this one needs an ergodic decomposition theorem for $\left(A^{\infty}, \varphi, \sigma^n\right)$:

i) $\left(A^{\infty}, \varphi, \sigma^{n}\right)$ splits into $1 \leq k \leq n$ ergodic components $\left\{\varphi^{(i)}\right\}_{1 \leq i \leq k}$

ii) $k \mid n$ and $\varphi^{(i)} = \varphi^{(i-1)} \circ \sigma$

iii) $s\left(\varphi^{(i)} \middle| \sigma^{n}\right) = s\left(\varphi^{(j)} \middle| \sigma^{n}\right) = ns\left(\varphi\right)$

finite size entropy estimation:

iv) $\forall \eta > 0$ and $n \rightarrow \infty \Rightarrow s\left(\varphi\right) \leq \frac{1}{n} S\left(\varphi^{(i)}\Big|_{A^{n}}\right) \leq s\left(\varphi\right) + \eta$

for almost every ergodic component

Next step : combining the different levels of approximation

Lemma :

given a sequence of probability measures (μ_n)

over finite alphabets (B_n) s.t.

a) $\frac{1}{n}\log\# B_n \leq C < \infty$

b) $\frac{1}{n}H(\mu_n) \to h$

c) $\limsup\left(\beta_{\varepsilon,n} := \frac{1}{n}\min\left\{\log\#\Omega : \mu_n(\Omega) > 1-\varepsilon\right\}\right) \leq h$ for $\forall\varepsilon > 0$

$\Rightarrow$

$\Rightarrow$ for $\forall \varepsilon > 0$

i) $\lim_{n\to\infty} \frac{1}{n} \beta_{\varepsilon,n} = h$

ii) $\mu_n \{ a \in B_n : \mu_n(a) > e^{-n(h-\varepsilon)} \} \to 0$

iii) $\mu_n \{ a \in B_n : \mu_n(a) < e^{-n(h+\varepsilon)} \} \to 0$

take $h = s(\varphi)$ and B_n as the index set of the projectors corresponding to the eigenspaces of D_n and apply the results about the ergodic decomposition and mix everything carefully!

For the proof of the relative entropy theorem one needs simultaneous good abelian approximations of the states ψ and φ

A coding application

(I.Bjelakovich, A.Szkoła)

Question:

Is the projection onto the typical subpaces a quantum operation with asumptotic fidelity 1?

A quantum channel is a trace preserving completely positive map from $B(H) \to B(H')$, H: finite dimensional Hilbertspace

Compression scheme $\{C^{(n)}, D^{(n)}\}$ for

stationary quantum source $(A^{\infty}, \varphi, \sigma) \cong \{A^n = B(H^{\otimes n}), D_n\}$

$C^{(n)} : B(H^{\otimes n}) \to B(H^{(n)} \subset H^{\otimes n}) \quad D^{(n)} : B(H^{(n)}) \to B(H^{\otimes n})$

Fidelity of two density matrices ρ and τ:

$$F(\rho,\tau) = \mathrm{tr}\left(\sqrt{\sqrt{\rho} \circ \tau \circ \sqrt{\rho}}\right)$$

(generalizes the overlap $\left|\langle\psi|\xi\rangle\right|$ of vectors in a Hilbert space)

$$1 - F(\rho,\tau) \le \frac{1}{2}\mathrm{tr}\left|\rho - \tau\right| \le \sqrt{1 - \left(F(\rho,\tau)\right)^2}$$

Compression rate:

$$R_C := \limsup \frac{\log \dim H^{(n)}}{n}$$

How large is $F\left(D_n, D^{(n)} \circ C^{(n)} \circ D_n = D_n^{'}\right)$ for given R_C and large n?

Theorem:

i) there is a compression scheme $\{C^{(n)}, D^{(n)}\}$ with

$R_C = s(\varphi)$ s.t. $\lim_{n\to\infty} F(D_n, D_n^{'}) = 1$

ii) any scheme with $R_C < s(\varphi)$ satisfies $\lim_{n\to\infty} F(D_n, D_n^{'}) = 0$

similar statements hold for stronger versions of fidelity

(entanglement fidelity, ensemble fidelity)

Open problems

- Stronger pointwise theorem
- Estimation of entropy
- Universal coding schemes (unknown source)
- Lempel-Ziv type coding
- Rate distortion
- Coding theorems for different channels
- Large deviations, Sanov‘s theorem
- Isomorphism classes etc. (are q-Bernoulli systems completely classified by the entropy?)