1 Proof of the ‘update’ proposition

Consider again an ergodic and atomless source $[A^\mathbb{Z}, \mu]$ and a Bernoulli source $[B^\mathbb{Z}, p^{\times \mathbb{Z}}]$ such that $h(\mu) \geq H(p)$. As in Lecture 27, let us say a stationary cade $\varphi : A^\mathbb{Z} \rightarrow B^\mathbb{Z}$ is $\delta$-good if

i) $\| (\varphi \ast \mu)_\ell - p^{\times \ell} \| < \delta$ for $\ell := [\delta^{-1}]$, and

ii) $h(\varphi \ast \mu) > H(p) - \delta$.

Using this terminology, we previously formulated a key ‘update’ proposition for approximate factor maps $A^\mathbb{Z} \rightarrow B^\mathbb{Z}$, and saw how it implies Sinai’s factor theorem. Here is that proposition again:

**Proposition 1.** For every $\varepsilon > 0$ there exists $\delta > 0$ such that the following holds. If $\varphi : A^\mathbb{Z} \rightarrow B^\mathbb{Z}$ is $\delta$-good, then for any $\eta > 0$ there exists an $\eta$-good code $\psi : A^\mathbb{Z} \rightarrow B^\mathbb{Z}$ such that $\mu\{\varphi_0 \neq \psi_0\} < \varepsilon$.

It only remains to prove this proposition. This is where we apply the following key tools acquired previously:

**Proposition 2 (Obtaining a $\mathcal{D}$-approximation and lifting it).** For each $\varepsilon > 0$ there exists $\delta > 0$ such that the following holds. If $\varphi : A^\mathbb{Z} \rightarrow B^\mathbb{Z}$ is $\delta$-good, then there is an ergodic joining $\lambda$ of $\mu$ and $p^{\times \mathbb{Z}}$ such that

\[ \lambda\{(x, y) : y_0 \neq \varphi_0(x)\} < \varepsilon. \]  

(1)

\[ \square \]
Proposition 3 (Weak containment with retention of entropy, joining version, special case). Let $\lambda$ be an ergodic joining of $\mu$ and $p^{\times\mathbb{Z}}$. For any $\ell \in \mathbb{N}$ and $\varepsilon > 0$ there is a stationary code $\psi : A^Z \longrightarrow B^Z$ such that
\[
\|\lambda_{\ell} - (\text{gr}(\psi, \mu))_{\ell}\| < \varepsilon \quad \text{and} \quad h(\psi_{\ast}\mu) > h(\nu) - \varepsilon.
\]

The assumption that $\varphi$ is $\delta$-good asserts that it is an approximate factor map from $\mu$ to $p^{\times\mathbb{Z}}$, in the sense of the vague topology and entropy. Given that $\varphi$ is such an approximate factor map, we proceed in two steps:

1. Apply Proposition 2 to obtain a joining $\lambda$ whose second marginal is exactly $p^{\times\mathbb{Z}}$, and which approximates $\varphi$ in the sense of (1), but which may not be a graphical joining.

2. Apply Proposition 3 to approximate $\lambda$ as closely as desired by the graphical joining of a new approximate factor map $\psi$.

It takes just a little extra work to check that all the approximations involved here behave correctly.

Proof of Proposition 1. Let $\varepsilon > 0$, let $\delta$ be given by Proposition 2 for this $\varepsilon$, and suppose that $\varphi : A^Z \longrightarrow B^Z$ is $\delta$-good.

Step 0. Since local sets generate the full $\sigma$-algebra of $A^Z$, there are local maps $\varphi'_0$ for which $\mu(\varphi'_0 \neq \varphi_0) \leq \delta$ arbitrarily small. If $\varphi' = (\varphi'_0 \circ S^n)_n$ is the sliding block code generated by such a $\varphi'_0$, then
\[
\overline{d}(\varphi'_0 \ast \mu, \varphi \ast \mu) \leq \mu(\varphi'_0 \neq \varphi_0).
\]
Since approximations in $\overline{d}$ control both the vague topology (Lemma 4 from Lecture 27) and entropy rate (by Fano’s inequality), it follows that we can find such an approximation $\varphi'$ which is still $\delta$-good.

Therefore, replacing $\varphi$ with $\varphi'$ if necessary, we may assume that $\varphi_0$ is a local map. The value of this assumption appears shortly.

Step 1 (corresponds to step 1 above). By Proposition 2, there is a joining $\lambda$ of $\mu$ and $p^{\times\mathbb{Z}}$ such that the set
\[
W := \{(x, y) : y_0 \neq \varphi_0(x)\}
\]
satisfies
\[ \lambda(W) < \varepsilon. \] (2)

**Step 2 (corresponds to step 2 above).**  Since \( \phi_0 \) is a local function, \( W \) is a local subset of \( A^\mathbb{Z} \times B^\mathbb{Z} \). This means it is closed and open for the product topology. Therefore any other joining \( \lambda' \) of \( \mu \) and \( p^x \mathbb{Z} \) which is sufficiently close to \( \lambda \) in the vague topology must also satisfy \( \lambda'(W) < \varepsilon \).

Now, for any \( \eta > 0 \), Proposition 3 gives an \( \eta \)-good code \( \psi : A^\mathbb{Z} \rightarrow B^\mathbb{Z} \) for which \( \text{gr}(\mu, \psi) \) is as close as we please to \( \lambda \) in the vague topology. In particular, we may choose \( \psi \) so that we still have
\[ \text{gr}(\mu, \psi)(W) = \mu\{\psi_0 \neq \varphi_0\} < \varepsilon, \]
as required. \( \Box \)

## 2 The residual Sinai factor theorem

In the proof of Sinai’s factor theorem, we started by choosing a \( \delta_1 \)-good code \( \varphi^1 : A^\mathbb{Z} \rightarrow B^\mathbb{Z} \) for some initial choice of \( \delta_1 \), and then we applied Proposition 1 repeatedly to construct a sequence of improvements \( \varphi^k : A^\mathbb{Z} \rightarrow B^\mathbb{Z} \). The code \( \varphi^k \) is \( \delta_k \)-good for some \( \delta_k \rightarrow 0 \) and which converge to the desired factor map \( \psi \) from \( \mu \) to \( p^x \mathbb{Z} \).

In that proof, it was important that we choose \( \delta_k \) small enough to force the convergence of \( \varphi^k \) to a limiting factor map \( \psi \). But by choosing even faster convergence, we can secure other properties of the limiting map as well. In particular, suppose we are given an ergodic joining \( \gamma \) of \( \mu \) and \( p^x \mathbb{Z} \) a priori, and also \( \eta > 0 \) and \( \ell \in \mathbb{N} \). Then:

- For any \( \eta > 0 \), if we choose the sequence \( \delta_k \) converging to zero fast enough, we can arrange that
  \[ \mu\{\varphi^1 \neq \psi\} \leq \sum_{i=1}^{\infty} \mu\{\varphi^i \neq \varphi^{i+1}\} < \eta/\ell; \] (3)

- By choosing \( \delta_1 < \eta \) and starting with an application of Proposition 3, we can arrange that
  \[ \|\text{gr}(\mu, \varphi^1)_\ell - \gamma_\ell\| < \eta. \] (4)
Now (3) and (4) together imply that
\[ \| \text{gr}(\mu, \psi)_\ell - \gamma_\ell \| < 2\eta. \]
Since \( \eta \) and \( \ell \) are arbitrary, we have proved the following.

**Theorem 4** (Residual Sinai factor theorem). If \( [A^\mathbb{Z}, \mu] \) is ergodic and atomless, and if \( h(\mu) \geq H(p) \), then the set of graphical joinings
\[ \{ \text{gr}(\mu, \psi) : \psi \text{ a factor map } [A^\mathbb{Z}, \mu] \rightarrow [B^\mathbb{Z}, p^{\times \mathbb{Z}}] \} \]
is vaguely dense within the set of all ergodic joinings of \( \mu \) and \( p^{\times \mathbb{Z}} \).

### 3 Ornstein’s theorem

The principal importance of Theorem 4 is in proving the following.

**Theorem 5** (Ornstein’s theorem). If \( [A^\mathbb{Z}, p^{\times \mathbb{Z}}] \) and \( [B^\mathbb{Z}, q^{\times \mathbb{Z}}] \) are Bernoulli shifts of equal entropy, then they are isomorphic.

**Lemma 6.** Let \( [A^\mathbb{Z}, \mu] \) be a source, \( B \) another alphabet, and \( \varphi : A^\mathbb{Z} \rightarrow B^\mathbb{Z} \) a stationary code. For every \( \varepsilon > 0 \) there is a vague neighbourhood \( U \) of the graphical joining \( \text{gr}(\mu, \varphi) \) for which the following holds: if \( \pi \) is any other stationary code \( A^\mathbb{Z} \rightarrow B^\mathbb{Z} \) satisfying \( \text{gr}(\mu, \pi) \in U \), then
\[ \mu\{\varphi_0 \neq \pi_0\} < \varepsilon. \]

**Proof.** As in Step 0 of the proof of Proposition 1, there exists a local map \( \varphi'_0 \) satisfying \( \mu\{\varphi_0 \neq \varphi'_0\} < \varepsilon. \) From this it follows that the set
\[ W := \{(x, y) : \varphi'_0(x) \neq y_0\} \]
satisfies \( \text{gr}(\mu, \varphi)(W) = \mu\{\varphi_0 \neq \varphi'_0\} < \varepsilon. \)

Let \( U \) be the set of all joinings which satisfy
\[ \lambda\{\varphi'_0(x) \neq y_0\} < \varepsilon. \]

Because \( \varphi'_0 \) is a local map, the set \( W \) is closed and open in the product topology, and so \( U \) is open in the vague topology (restricted to the set of all joinings).

If \( \text{gr}(\mu, \pi) \in U \), then the same calculation as above gives
\[ \mu\{\pi_0 \neq \varphi'_0\} = \text{gr}(\mu, \pi)(W) < \varepsilon, \]
and so \( \mu\{\pi_0 \neq \varphi_0\} < 2\varepsilon. \)
Proof of Theorem 5. Sinai’s factor theorem provides factor maps in each direction. We can improve on this observation by repeatedly applying Theorem 4 in alternating directions.

To explain this, let us write

\[ \tilde{\text{gr}}(q^{xZ}, \psi) := \int_{B^Z} \delta_{(\psi(y), y)} q^{xZ}(dy) \]

when \( \psi : B^Z \to A^Z \). This is just a graphical joining with the role of the two coordinate factors swapped.

Fix a summable sequence of error tolerances \( \varepsilon_k \), \( k \geq 1 \). Now we construct sequences of codes \( \varphi^k : A^Z \to B^Z \) and \( \psi^k : B^Z \to A^Z \) as follows.

First, let \( \varphi^1 : A^Z \to B^Z \) be any factor map between the Bernoulli shifts.

Let \( U_1 \) be a vague neighbourhood of \( \text{gr}(p^{xZ}, \varphi^1) \) as provided by Lemma 6 for the tolerance \( \varepsilon_1 \).

Now apply the residual Sinai factor theorem in the reverse direction: there is a factor map \( \psi^1 : r_{B^Z}, q^{xZ} \) such that \( \text{gr}(p^{xZ}, \varphi^1) \subseteq U_1 \). Apply Lemma 6 again to obtain a vague neighbourhood \( V_1 \) of \( \text{gr}(q^{xZ}, \psi^1) \) as in that lemma, for the tolerance \( \varepsilon_1 \) but for maps going from \( B^Z \) to \( A^Z \). Shrink \( V_1 \) if necessary so that \( V_1 \subseteq U_1 \).

Now we return to the forward direction \( A^Z \to B^Z \) and apply the residual Sinai theorem again. Continuing back-and-forth in this way, we obtain

i) sequences of factor maps

\[ \varphi^k : A^Z \to B^Z \quad \text{and} \quad \psi^k : B^Z \to A^Z, \quad k \geq 1, \]

ii) vaguely open sets of joinings

\[ U_1 \supseteq V_1 \supseteq U_2 \supseteq V_2 \supseteq \ldots \]

such that

\[ \text{gr}(p^{xZ}, \varphi^k) \in U_k \quad \text{and} \quad \tilde{\text{gr}}(q^{xZ}, \psi^k) \in V_k \quad \forall k. \quad (5) \]

Since the vague topology on \( \text{Prob}(A^Z \times B^Z) \) is metrizable as well as compact, we may also choose the sets \( U_k \) and \( V_k \) to shrink so fast that \( \bigcap_k \overline{U_k} = \bigcap_k \overline{V_k} \) is a singleton. (For instance, by insisting that \( U_k \) and \( V_k \) have diameter at most \( 2^{-k} \) in a suitable choice of metric.)

From the way we choose each \( U_k \) and \( V_k \) using Lemma 6, we have

\[ \text{gr}(p^{xZ}, \varphi^k), \text{gr}(p^{xZ}, \varphi^{k+1}) \in U_k \quad \implies \quad p^{xZ}\{\varphi_k^0 \neq \varphi_k^{k+1}\} < \varepsilon_k \]
for each $k$, and similarly $q^Z \{ \psi_0^k \neq \psi_0^{k+1} \} < \varepsilon_k$.

Since $\sum_k \varepsilon_k < \infty$, the Borel–Cantelli lemma gives limiting observables $\varphi_0$ and $\psi_0$ such that $\varphi_0^k \to \varphi_0$ and $\psi_0^k \to \psi_0$ in probability. Let $\varphi$ and $\psi$ be the stationary codes generated by $\varphi_0$ and $\psi_0$. As before, since each $\varphi^k$ and $\psi^k$ is a factor map between our Bernoulli processes, the same is true of the limiting processes $\varphi$ and $\psi$.

Finally, it follows from this convergence and (5) that
\[
\text{gr}(p^Z, \varphi) \in \bigcap_k U_k \quad \text{and} \quad \text{gr}(q^Z, \psi) \in \bigcap_k V_k.
\]

Since these two intersections are equal to the same singleton, we must have
\[
\text{gr}(p^Z, \varphi) = \text{gr}(q^Z, \psi).
\]

This is possible only if $\psi$ is an inverse to $\varphi$ almost everywhere. 

4 Finitely determined processes and beyond

Recall from Lecture 27 that a source $[A^Z, \mu]$ is \textbf{finitely determined} if for every $\varepsilon > 0$ there exist $\delta > 0$ and $\ell \in \mathbb{N}$ such that, if $\nu \in \text{Prob}^S(A^Z)$ satisfies
\[
\| \mu_\ell - \nu_\ell \| < \delta \quad \text{and} \quad h(\nu) > h(\mu) - \delta,
\]
then $\overline{d}(\mu, \nu) < \varepsilon$.

We saw in Lecture 27 that Bernoulli sources are finitely determined, and then this is the only property of those sources that we use in the subsequent proofs. Therefore:

Any finitely determined source is isomorphic to any Bernoulli source of the same entropy, and Sinai’s factor theorem is still true with any finitely determined source as the target.

Thus, finite determination is a characterization of those processes that are isomorphic to Bernoulli shifts. Starting with this, Ornstein and others found many further characterizations of ‘Bernoullicity’ which can be checked in a wide range of examples. These characterizations include the ‘very weak Bernoulli property’ (Ornstein), ‘extremality’ (Thouvenot) and the ‘almost blowing-up property’ (really a version of concentration of measure; Marton and Shields).

The first important consequence of this theory is the following:
Theorem 7. Any factor of a Bernoulli shift is isomorphic to a Bernoulli shift.

The following examples are also all isomorphic to Bernoulli shifts:

- The geodesic flow on the unit tangent bundle of a negatively curved compact Riemannian manifold, equipped with its volume form.

- The invariant measure on $A^\mathbb{Z}$ arising from any mixing stationary $A$-valued Markov process.

- The transformation of $\mathbb{T}^2$ with Lebesgue measure given by the matrix $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ (or any other element of $\text{SL}_2(\mathbb{Z})$ with eigenvalues off the unit circle).

There are many others besides.

5 Infinite entropy

An infinite-entropy Bernoulli shift is a system of the form $(A^\mathbb{Z}, p^x, S)$ where $(A, p)$ is a non-finite probability space of infinite Shannon entropy (for instance, $p$ could have a non-trivial atomless part).

Sinai’s and Ornstein’s theorems generalize to the setting of infinite entropy.

6 Completely positive entropy

A MPS $(X, \mu, T)$ has completely positive entropy if every nontrivial factor of it has strictly positive entropy.

This property has an alternative characterization for shift systems $(A^\mathbb{Z}, \mu, S)$ with a finite alphabet $A$: it holds if and only if the left tail $\sigma$-algebra

$$\bigcap_{n \geq 1} \sigma\text{-alg}(\alpha_m: m \leq -n)$$

is trivial, where $\alpha$ is the canonical $A$-valued process on $A^\mathbb{Z}$. This equivalence is due to Kolmogorov, and MPSs with completely positive entropy are sometimes called Kolmogorov or K-automorphisms.

Using this equivalence, it follows at once that Bernoulli shifts are K-automorphisms. It was unknown for many years whether this property actually characterizes systems that are isomorphic to Bernoulli shifts, but Ornstein disproved this: once
he had established his necessary and sufficient conditions for isomorphism to a Bernoulli shift, he was able to construct examples of K-automorphisms that violate those conditions.

By now a wide variety of non-Bernoulli K-automorphisms are known.

7 Splittings and relative Ornstein theory

Thouvenot introduced an important new direction for this theory. A factor map

$$\pi : (X, \mu, T) \rightarrow (Y, \nu, S)$$

of MPSs is relatively Bernoulli if there is another factor map

$$\varphi : (X, \mu, T) \rightarrow (B^\mathbb{Z}, p^\mathbb{Z}, S_B)$$

with image a Bernoulli shift such that the combined map

$$(\pi, \varphi) : X \rightarrow Y \times B^\mathbb{Z} : x \mapsto (\pi(x), \varphi(x))$$

is an isomorphism

$$(X, \mu, T) \rightarrow (Y \times B^\mathbb{Z}, \nu \times p^\mathbb{Z}, S \times S_B).$$

Then we also refer to the pair $$(\pi, \varphi)$$ as a splitting of $$(X, \mu, T)$$ into $$(Y, \nu, S)$$ and a Bernoulli factor.

Thouvenot generalized Ornstein’s theory by providing analogous necessary and sufficient conditions for a factor map such as $$\pi$$ to be relatively Bernoulli. This should be seen as a ‘relative version’ of Ornstein theory over the distinguished factor $$(Y, \nu, S)$$.

These ideas recently culminated in the following strengthening of Sinai’s factor theorem.

**Theorem 8** (The weak Pinsker theorem; A’17). If $$(X, \mu, T)$$ is ergodic, then for every $$\varepsilon > 0$$ it has a splitting into a Bernoulli factor and a factor of entropy less than $$\varepsilon$$.  

However, there are systems that one cannot split into a Bernoulli factor and a factor of entropy zero. Indeed, any non-Bernoulli K-automorphism is an example (exercise!). But there are others as well: there are systems that cannot be split into a K-automorphism and a factor of entropy zero. This was an old question of Pinsker, and was answered negatively by Ornstein in 1973. This is why the above is called the ‘weak Pinsker’ theorem. The idea to pursue this theorem, as well as its name, are due to Thouvenot.
8 An omission: Krieger’s generator theorem

Another very natural question about MPSs and entropy is the following: when is an abstract MPS \((X, \mu, T)\) isomorphic to a shift-system with a given alphabet \(A\)? Once again it turns out that entropy is the only relevant obstruction.

**Theorem 9** (Krieger’s generator theorem). *Let \(A\) be an alphabet of size \(k\). Then an ergodic MPS \((X, \mu, T)\) is isomorphic to a shift-system with alphabet \(A\) if and only if*

- either \(h(\mu, T) < \log |A|\)
- or \(h(\mu, T) = \log |A|\) and \((X, \mu, T)\) is isomorphic to the Bernoulli shift \((A^\mathbb{Z}, p^{\times \mathbb{Z}}, S)\), where \(p\) is the uniform distribution on \(A\).

The proof of Kreiger’s theorem uses many similar ideas to those that we met on route to Sinai’s and Ornstein’s theorems. But the technical details are substantial, and different enough that we do not cover them in this course.

**Tim Austin**

*Email: tim@math.ucla.edu*

*URL:* math.ucla.edu/~tim