CHAPTER 8

MEASURE THEORETIC ENTROPY

In this chapter we shall show how to associate to a measure preserving transformation an important quantity called the measure theoretic entropy. This gives important information on the dynamics of the map (cf. Chapter 12) and is useful in classifying measure preserving transformations.

The essential results are contained in sections 8.1-8.3. If one accepts Sinai's result on strong generators (Lemma 8.8) without proof then sections 8.4-8.8 will only be required again in chapter 12.

8.1 Partitions and conditional expectations

Let $\alpha = \{A_i\}_{i \in I}$ be a countable measurable partition of the probability space $(X, \mathcal{B}, \mu)$, i.e.

(i) $X = \cup_i A_i$ (up to a set of zero μ-measure), and
(ii) $A_i \cap A_j = \emptyset$ for $i \neq j$ (up to a set of zero μ-measure).

Definition. We define the information function $I(\alpha) : X \to \mathbb{R}$ by

$$I(\alpha)(x) = -\sum_i \log \mu(A_i) \chi_{A_i}(x),$$

i.e. $I(\alpha)(x) = -\log \mu(A_i)$ if $x \in A_i$.

Consider a *sub*-sigma-algebra $\mathcal{A} \subset \mathcal{B}$; then we can define a measure space $(X, \mathcal{A}, \mu)$ with respect to the smaller sigma-algebra. For any $f \in L^1(X, \mathcal{B}, d\mu)$ we can define a measure on the measure space $(X, \mathcal{A}, \mu)$ by $\mu_{\mathcal{A}}(A) = \int_A f d\mu$, for $A \in \mathcal{A}$. Clearly, $\mu_{\mathcal{A}} << \mu$ (where μ is here defined on $\mathcal{A}$).

Definition. By the Radon-Nikodym theorem there is a unique function $E(f|\mathcal{A}) := \frac{d\mu_{\mathcal{A}}}{d\mu} \in L^1(X, \mathcal{A}, d\mu)$ which is called the *conditional expectation.*

Since in general $\mathcal{A}$ is *strictly* contained in $\mathcal{B}$ then $E(f|\mathcal{A})$ may be very different from f, since it must be measurable on a smaller sigma-algebra. For example, if $\mathcal{A} = \{X, \emptyset\}$ then $E(f|\mathcal{A})$ is the constant function $\int f d\mu$.

The main properties of $E(f|\mathcal{A})$ are

(i) $\int_A E(f|\mathcal{A}) d\mu = \int_A f d\mu$ for all $A \in \mathcal{A}$,
(ii) $E(f|\mathcal{A}) \in L^1(X, \mathcal{A}, d\mu)$,

(The first two properties are just the definition repeated.),

(iii) if $f \in L^1(X, \mathcal{B}, \mu)$ and $g \in L^\infty(X, \mathcal{A}, d\mu)$ then $E(fg|\mathcal{A}) = gE(f|\mathcal{A})$,
(iv) if $f \in L^1(X, \mathcal{B}, \mu)$ and $\mathcal{A}_2 \subset \mathcal{A}_1 \subset \mathcal{B}$ then $E(E(f|\mathcal{A}_1)|\mathcal{A}_2) = E(f|\mathcal{A}_2)$,
(v) if $f \in L^1(X, \mathcal{B}, \mu)$ then $|E(f|\mathcal{A})| \le E(|f||\mathcal{A})$, and if $f, g \in L^2(X, \mathcal{B}, d\mu)$ and $\frac{1}{p} + \frac{1}{q} = 1$ then $E(|fg||\mathcal{A}) \le E(|f|^p|\mathcal{A})^{1/p} E(|g|^q|\mathcal{A})^{1/q}$,
(vi) If T preserves μ then $E(f|\mathcal{A})T = E(f \circ T|T^{-1}\mathcal{A})$, where $T^{-1}\mathcal{A} = \{T^{-1}A : A \in \mathcal{A}\}$.

Parts (iii)-(vi) are a trivial exercise from the definitions (cf. [5, p. 10]).

DEFINITION. Given any sub-sigma-algebra $\mathcal{A} \subset \mathcal{B}$ we can define the *conditional information function* $I(\alpha|\mathcal{A}) : X \to \mathbb{R}$ by

$$I(\alpha|\mathcal{A})(x) = -\sum_i \log \mu(A_i|\mathcal{A})(x)\chi_{A_i}(x)$$

where we write $\mu(A_i|\mathcal{A})(x) = E(\chi_{A_i}|\mathcal{A})(x)$ (called the *the conditional measure*).

Assume that we "know" the position of the point x relative to $\mathcal{A}$, then $I(\alpha|\mathcal{A})$ is an indicator of how much additional information we get from "knowing" the position of the point x relative to the partition α.

The following properties all come directly from the definition.

LEMMA 8.1.

(1) *When* $\mathcal{A} = \{\emptyset, X\}$ *then* $E(A_i|\{\emptyset, X\})(x) = \mu(A_i)$ *and* $I(\alpha|\{\emptyset, X\})(x) = I(\alpha)(x)$.
(2) *If* $T : X \to X$ *preserves the measure* μ *then* $I(\alpha|\mathcal{A})(Tx) = I(T^{-1}\alpha|T^{-1}\mathcal{A})(x)$ *(where* $T^{-1}\alpha = \{T^{-1}A_i\}$*).*
(3) *If* $\alpha \subset \mathcal{A}$ *then* $I(\alpha|\mathcal{A}) = 0$ *almost everywhere.*

We can associate to each partition α the sigma-algebra $\hat{\alpha}$ generated by α.

DEFINITION. Given two partitions α, β we define their *refinement*

$$\alpha \vee \beta = \{A_i \cap B_i : A_i \in \alpha, B_j \in \beta\}.$$

Given two sigma-algebras $\mathcal{A}, \mathcal{B}$ we denote by $\mathcal{A} \vee \mathcal{B}$ the sigma-algebra generated by $\{A \cap B : A \in \alpha, B \in \beta\}$.

The following lemma will prove very useful throughout this chapter.

LEMMA 8.2 (BASIC IDENTITY FOR INFORMATION). *Given partitions α, β and a third γ with associated sigma-algebra $\hat{\gamma}$ we have that*

$$I(\alpha \vee \beta|\hat{\gamma})(x) = I(\alpha|\hat{\beta} \vee \hat{\gamma})(x) + I(\beta|\hat{\gamma})(x)$$

(almost everywhere).

PROOF. Observe that for any function $g \in L^1(X, \mathcal{B}, \mu)$ we have that

$$E(g|\hat{\gamma})(x) = \sum_{C \in \gamma} \chi_C \frac{\int_C g(x) d\mu}{\mu(C)}.$$

In particular, for $B \in \beta$ we can set $g(x) = \chi_B(x)$ and then we get

$$\mu(B|\hat{\gamma})(x) = \sum_{C \in \gamma} \chi_C(x) \frac{\mu(B \cap C)}{\mu(C)}$$

and therefore

$$I(\beta|\hat{\gamma})(x) = -\sum_{C \in \gamma, B \in \beta} \chi_{C \cap B}(x) \log\left(\frac{\mu(B \cap C)}{\mu(C)}\right). \tag{8.1}$$

The partition $\beta \vee \gamma$ (with elements of the form $B \cap C$ with $B \in \beta, C \in \gamma$) gives that

$$I(\alpha|\hat{\gamma} \vee \hat{\beta})(x) = -\sum_{C \in \gamma, B \in \beta, A \in \alpha} \chi_{A \cap B \cap C}(x) \log\left(\frac{\mu(A \cap B \cap C)}{\mu(B \cap C)}\right). \tag{8.2}$$

Adding (8.1) and (8.2) gives that

$$\begin{aligned}
& I(\beta|\hat{\gamma})(x) + I(\alpha|\hat{\gamma} \vee \hat{\beta})(x) \\
&= -\sum_{C \in \gamma, B \in \beta, A \in \alpha} \chi_{A \cap B \cap C}(x) \left(\log\left(\frac{\mu(B \cap C)}{\mu(C)}\right) + \log\left(\frac{\mu(A \cap B \cap C)}{\mu(B \cap C)}\right)\right) \\
&= -\sum_{C \in \gamma, B \in \beta, A \in \alpha} \chi_{A \cap B \cap C}(x) \log\left(\frac{\mu(A \cap B \cap C)}{\mu(C)}\right) \\
&= I(\alpha \vee \beta|\hat{\gamma})(x).
\end{aligned}$$

This completes the proof.

■

If α and β are partitions we write $\alpha < \beta$ if every element of α is a union of elements of β. In this case $\alpha \vee \beta = \beta$.

COROLLARY 8.2.1. If $\alpha < \beta$ then $I(\alpha|\hat{\gamma})(x) \leq I(\beta|\hat{\gamma})(x)$.

8.2 The entropy of a partition

DEFINITION. We define the *entropy* of the partition α by

$$H(\alpha) = \int I(\alpha)(x) d\mu(x) = -\sum_{A\in\alpha} \mu(A)\log\mu(A).$$

Given a partition α and a sub-sigma-algebra $\mathcal{A} \subset \mathcal{B}$ we define the *conditional entropy* by $H(\alpha|\mathcal{A}) = \int I(\alpha|\mathcal{A})(x)d\mu(x)$.

LEMMA 8.3.

(1) *When* $\mathcal{A} = \{\emptyset, X\}$ *then* $H(\alpha|\{\emptyset, X\}) = H(\alpha)$.
(2) *If* $T : X \to X$ *preserves the measure* μ *then* $H(\alpha|\mathcal{A}) = H(T^{-1}\alpha|T^{-1}\mathcal{A})$.
(3) *If* $\alpha \subset \mathcal{A}$ *then* $H(\alpha|\mathcal{A}) = 0$.
(4) *Given* α, γ *we have* $H(\alpha|\hat{\gamma}) \leq H(\alpha)$.

PROOF. Parts (1), (2) and (3) follow by integrating the corresponding results for the information function in Lemma 8.1.

For part (4) we have

$$\begin{aligned}
H(\alpha|\gamma) &= -\sum_{A\in\alpha, C\in\gamma} \mu(A\cap C)\log\left(\frac{\mu(A\cap C)}{\mu(C)}\right) \\
&= -\sum_{C\in\gamma}\mu(C)\left[\sum_{A\in\alpha}\frac{\mu(A\cap C)}{\mu(C)}\log\left(\frac{\mu(A\cap C)}{\mu(C)}\right)\right] \\
&\leq -\sum_{A\in\alpha}\left[\sum_{C\in\gamma}\mu(A\cap C)\right]\log\left[\sum_{C\in\gamma}\mu(A\cap C)\right] \\
&\leq -\sum_{A\in\alpha}\mu(A)\log\mu(A) = H(\alpha)
\end{aligned}$$

since for fixed $A \in \alpha$ we can bound

$$-\sum_{C\in\gamma}\frac{\mu(A\cap C)}{\mu(C)}\log\left(\frac{\mu(A\cap C)}{\mu(C)}\right) \leq -\left[\sum_{C\in\gamma}\mu(A\cap C)\right]\log\left[\sum_{C\in\gamma}\mu(A\cap C)\right]$$

using concavity of $t \mapsto -t\log t$. (A more general result appears in Lemma 8.13.)

■

LEMMA 8.4 (BASIC IDENTITY FOR ENTROPY). *Given partitions* α, β *and a third* γ *with associated sigma-algebra* $\hat{\gamma}$ *we have that*

$$H(\alpha\vee\beta|\hat{\gamma}) = H(\alpha|\hat{\beta}\vee\hat{\gamma}) + H(\beta|\hat{\gamma}).$$

COROLLARY 8.4.1 ("MONOTONICITY" OF ENTROPY FOR PARTITIONS). *Given two partitions α, β with $\alpha < \beta$ we have that $H(\alpha|\hat{\gamma}) \leq H(\beta|\hat{\gamma})$ (and, in particular $H(\alpha) \leq H(\beta)$)*

With the next definition, we begin to re-introduce measure preserving transformations.

DEFINITION. Assume that $T : X \to X$ preserves μ. Given a partition $\alpha = \{A_i\}$ we write

$$\vee_{i=0}^{n-1} T^{-i}\alpha = \{A_{r_0} \cap T^{-1}A_{r_1} \cap \cdots \cap T^{-(n-1)}A_{r_{n-1}} : A_{r_i} \in \alpha, i = 0, \ldots, n-1\}.$$

NOTATIONAL COMMENT. Frequently it proves convenient to drop the circumflex (hat) over $\hat{\gamma}$. Thus if we write $H(\alpha|\beta)$, say, we understand this to mean $H(\alpha|\hat{\beta})$.

For $n \geq 1$ we can write $H_n(\alpha) = H(\vee_{i=0}^{n-1} T^{-i}\alpha)$. By the above estimates we have that

$$\begin{aligned} H_{n+m}(\alpha) &= H\left(\vee_{i=0}^{n+m-1} T^{-i}\alpha\right) \\ &= H\left(\vee_{i=0}^{n-1} T^{-i}\alpha\right) + H\left(\vee_{i=n}^{n+m-1} T^{-i}\alpha | \vee_{i=0}^{n-1} T^{-i}\alpha\right) \\ &\leq H\left(\vee_{i=0}^{n-1} T^{-i}\alpha\right) + H\left(\vee_{i=n}^{n+m-1} T^{-i}\alpha\right) \\ &= H_n(\alpha) + H_m(\alpha). \end{aligned} \tag{8.3}$$

Thus the sequence $H_n(\alpha)$, $n \geq 1$, is subadditive (which shows that the limit in the following definition exists).

DEFINITION. We define the *entropy of the partition* α relative to the transformation $T : X \to X$ as the limit $h(T, \alpha) = \lim_{n \to +\infty} \frac{H_n(\alpha)}{n}$.

Notice that in particular from (8.3) we have that $0 \leq h(T, \alpha) \leq H(\alpha)$. The following result gives an equivalent characterization.

PROPOSITION 8.5 (ALTERNATIVE DEFINITION OF $h(T, \alpha)$).

$$h(T, \alpha) = \lim_{n \to +\infty} H(\alpha | \vee_{i=1}^{n-1} T^{-i}\alpha).$$

(N.B. Sometimes it is convenient to write this limit as $H(\alpha | \vee_{i=1}^{\infty} T^{-i}\alpha)$.)

PROOF. Using Lemma 8.4 we see that

$$\begin{aligned} H(\vee_{i=0}^{n-1} T^{-i}\alpha) &= H(\alpha | \vee_{i=1}^{n-1} T^{-i}\alpha) + H(\vee_{i=0}^{n-2} T^{-i}\alpha) \\ &= H(\alpha | \vee_{i=1}^{n-1} T^{-i}\alpha) + H(\alpha | \vee_{i=1}^{n-2} T^{-i}\alpha) + H(\vee_{i=0}^{n-3} T^{-i}\alpha) \\ &\cdots \\ &= \sum_{r=2}^{n} H(\alpha | \vee_{i=1}^{r-1} T^{-i}\alpha) + H(\alpha). \end{aligned}$$

We then see that

$$\lim_{n\to+\infty}\frac{1}{n}H(\vee_{i=0}^{n-1}T^{-i}\alpha)=\lim_{n\to+\infty}H(\alpha|\vee_{i=1}^{n-1}T^{-i}\alpha)$$

as required (since if $a_n\to a$ for any sequence of real numbers then $\frac{a_1+\ldots+a_n}{n}\to a$).

■

The entropy of the transformation relative to two different partitions is described by the following inequality.

LEMMA 8.6. *For finite entropy partitions α,β of X we have that*

$$h(T,\alpha)\le h(T,\beta)+H(\alpha|\beta).$$

PROOF. Since $\left(\vee_{i=0}^{n-1}T^{-i}\alpha\right)\vee\left(\vee_{i=0}^{n-1}T^{-i}\beta\right)>\vee_{i=0}^{n-1}T^{-i}\alpha$ we have that

$$\begin{aligned}H\left(\vee_{i=0}^{n-1}T^{-i}\alpha\right)&\le H\left(\left(\vee_{i=0}^{n-1}T^{-i}\alpha\right)\vee\left(\vee_{i=0}^{n-1}T^{-i}\beta\right)\right)\\&=H\left(\vee_{i=0}^{n-1}T^{-i}\beta\right)+H\left(\vee_{i=0}^{n-1}T^{-i}\alpha|\vee_{i=0}^{n-1}T^{-i}\beta\right)\end{aligned}$$

(where we use Lemma 8.4, with γ being the trivial partition, for the last line). We next estimate

$$\begin{aligned}&H\left(\vee_{i=0}^{n-1}T^{-i}\alpha|\vee_{i=0}^{n-1}T^{-i}\beta\right)\\&=H\left(\alpha|\vee_{i=0}^{n-1}T^{-i}\beta\right)+H\left(\vee_{i=1}^{n-1}T^{-i}\alpha|\alpha\vee\left(\vee_{i=0}^{n-1}T^{-i}\beta\right)\right)\\&\le H(\alpha|\beta)+H\left(\vee_{i=1}^{n-1}T^{-i}\alpha|\vee_{i=1}^{n-1}T^{-i}\beta\right)\\&\le H(\alpha|\beta)+H\left(\vee_{i=0}^{n-2}T^{-i}\alpha|\vee_{i=0}^{n-2}T^{-i}\beta\right).\end{aligned}$$

Proceeding inductively gives us

$$H\left(\vee_{i=0}^{n-1}T^{-i}\alpha|\vee_{i=0}^{n-1}T^{-i}\beta\right)\le nH(\alpha|\beta).$$

Finally, we see that

$$\frac{1}{n}H\left(\vee_{i=0}^{n-1}T^{-i}\alpha\right)\le\frac{1}{n}H\left(\vee_{i=0}^{n-1}T^{-i}\beta\right)+H(\alpha|\beta).$$

Letting $n\to+\infty$ gives the correct inequality.

■

COROLLARY 8.6.1. *For finite entropy partitions α,β of X we have that*

$$|h(T,\beta)-h(T,\alpha)|\le H(\beta|\alpha)+H(\alpha|\beta).$$

PROOF. By interchanging α and β in Lemma 8.6 we get that $h(T,\beta)\le h(T,\alpha)+H(\beta|\alpha)$.

8.3 The entropy of a transformation

Consider a measure preserving transformation $T : X \to X$ on a probability space $(X, \mathcal{B}, \mu)$. We want to associate to this a numerical invariant. We start from the definition of the entropy relative to a partition α and then remove the dependence on α by taking a supremum.

DEFINITION. We define the *measure theoretic entropy* of $T : X \to X$ for the probability space $(X, \mathcal{B}, \mu)$ by $h_\mu(T) = \sup_{\{\alpha : H(\alpha) < +\infty\}} h(T, \alpha)$.

We write the measure μ as a subscript not only to remind us that there is an ambient measure, but also to distinguish the notation from that of topological entropy in chapter 3.

As one might imagine, it can be very difficult to compute the measure theoretic entropy from the definition given. We now want to describe a very important method of *practical* computation. We begin with a result which replaces the supremum in the definition of the measure theoretic entropy with a limit.

LEMMA 8.7 (ABRAMOV). *Let $\beta_1 \subset \beta_2 \subset \ldots \subset \beta_k \subset \mathcal{B}$ be an increasing sequence of partitions with $H(\beta_k) < +\infty$, $\forall k \geq 1$; and such that $\cup_n \beta_k$ generates the sigma-algebra $\mathcal{B}$. Then $h_\mu(T) = \lim_{k \to +\infty} h(T, \beta_k)$.*

We shall return to the proof in section 8.7.

The following definition gives us a way to generate the increasing partitions.

DEFINITION. We say that a partition α with $H(\alpha) < +\infty$ is called a *strong generator* for the probability space $(X, \mathcal{B}, \mu)$ if $\vee_{i=0}^{\infty} T^{-i}\alpha = \mathcal{B}$.
If T is invertible, then we say that a partition α with $H(\alpha) < +\infty$ is called a *generator* for the probability space $(X, \mathcal{B}, \mu)$ if $\vee_{i=-\infty}^{\infty} T^{-i}\alpha = \mathcal{B}$

LEMMA 8.8 (SINAI). *If α is a (strong) generator then $h_\mu(T) = h(T, \alpha)$.*

We shall return to the proof in section 8.7. Before developing the theory needed to prove these two results we shall use them to compute the measure theoretic entropy of some simple examples.

Example 1 (doubling map). Let $X = \mathbb{R}/\mathbb{Z}$, $\mathcal{B}$ denote the Borel sigma-algebra, and μ the Haar-Lebesgue measure. We let $T : X \to X$ be the doubling map $T(x) = 2x \pmod 1$. Let $\alpha = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$; then observe that

$$\alpha \vee T^{-1}\alpha = \left\{ \left[0, \frac{1}{4}\right), \left[\frac{1}{4}, \frac{1}{2}\right), \left[\frac{1}{2}, \frac{3}{4}\right), \left[\frac{3}{4}, 1\right) \right\}$$

and more generally,

$$\vee_{i=0}^{n-1} T^{-i}\alpha = \left\{ \left[\frac{i}{2^n}, \frac{i+1}{2^n} \right) \ : i = 0, \ldots, 2^n - 1 \right\}.$$

We can now calculate

$$\begin{aligned} H(\vee_{i=0}^{n-1} T^{-i}\alpha) &= -\sum_{i=0}^{2^n-1} \mu\left(\left[\frac{i}{2^n}, \frac{i+1}{2^n} \right) \right) \log\left(\left[\frac{i}{2^n}, \frac{i+1}{2^n} \right) \right) \\ &= -\sum_{i=0}^{2^n-1} \left(\frac{1}{2^n} \right) \log\left(\frac{1}{2^n} \right) \\ &= -2^n \left(\frac{1}{2^n} \right) \log\left(\frac{1}{2^n} \right) \\ &= n \log 2. \end{aligned}$$

Thus we see that $\frac{1}{n} H(\vee_{i=0}^{n-1} T^{-i}\alpha) = \log 2$ and thus letting $n \to +\infty$ gives that $h_\mu(T) = \log 2$.

Example 2 (rotations on the circle). Let $X = \mathbb{R}/\mathbb{Z}$, let $\mathcal{B}$ denote the Borel sigma-algebra, and let μ be the Haar-Lebesgue measure. We let $T : X \to X$ be the rotation $T(x) = x + a \pmod 1$ for some fixed values $a \in \mathbb{R}$.

First assume that $a = \frac{p}{q}$ is a rational number. For any partition β we see that $T^{-q}\beta = \beta$. Thus

$$\vee_{k=0}^{nq-1} T^{-k}\beta = \vee_{k=0}^{q-1} T^{-k}\beta$$

and so in particular

$$\begin{aligned} h_\mu(T, \beta) &= \lim_{n \to +\infty} \frac{1}{qn} H\left(\vee_{k=0}^{nq-1} T^{-k}\beta \right) \\ &= \lim_{n \to +\infty} \frac{1}{qn} H\left(\vee_{k=0}^{q-1} T^{-k}\beta \right) \\ &= 0. \end{aligned}$$

Thus the measure theoretic entropy of any partition is zero, and thus the measure theoretic entropy of the transformation, $h_\mu(T) = \sup_{H(\beta)<+\infty} h(T, \beta) = 0$.

Next, assume that a is irrational. We let $\beta = \{[0, \frac{1}{2}), [\frac{1}{2}, 1)\}$. Since the sequence $\frac{1}{2} + na \pmod 1$ is dense in the unit circle (Weyl's theorem) we see that the partition is (strong) generating. Moreover, we see that $\mathcal{B} = \vee_{k=1}^{\infty} T^{-k}\beta$. As we observed before

$$h(T, \beta) = \lim_{n \to +\infty} H(\beta | \vee_{k=1}^{n-1} T^{-k}\beta) = 0$$

and therefore $h_\mu(T) = h(T, \beta) = 0$.

As one might imagine, a similar method applies to rotation on tori $\mathbb{R}^n/\mathbb{Z}^n$, $n \geq 2$.

Example 3 (Markov measures). Let $\sigma : X \to X$ denote a subshift of finite type

$$X = \{(x_n) \in \prod_{-\infty}^{+\infty} \{1, \dots, k\} : A_{x_n x_{n+1}} = 1\}$$

where $A = (A_{ij})$ is a $k \times k$ matrix with entries either zero or unity.

We associate to this a $k \times k$ stochastic matrix $P = (P_{ij})$ (cf. Example 4 in section 7.3) and let $p = (p_1, \dots, p_k)$ be the left eigenvector associated to the left eigenvalue unity.

The partition $\alpha = \{[1], \dots [k]\}$ for X is generating. Let $\sigma : X \to X$ denote the shift transformation. The refined partition $\vee_{k=0}^{n-1} T^{-k}\alpha$ consists of "cylinder" sets of the form

$$[z_0, \dots, z_{n-1}] = \{(x_n)_{n\in\mathbb{Z}} \in X : x_i = z_i, 0 \leq i \leq n-1\}$$

where $z_i \in \{1, \dots, k\}$.

By the definition of the Markov measure μ associated to P we have that

$$\mu([z_0, \dots, z_{n-1}]) = p_{z_0} P_{z_0 z_1} P_{z_1 z_2} \dots P_{z_{n-2} z_{n-1}}.$$

We explicitly compute:

$$\begin{aligned}
&H(\vee_{k=0}^{n-1} T^{-k}\alpha) \\
&= - \sum_{[z_0,\dots,z_{n-1}]} \mu([z_0, \dots, z_{n-1}]) \log \mu([z_0, \dots, z_{n-1}]) \\
&= - \sum_{[z_0,\dots,z_{n-1}]} p_{z_0} P_{z_0 z_1} P_{z_1 z_2} \dots P_{z_{n-2} z_{n-1}} \log \left(p_{z_0} P_{z_0 z_1} P_{z_1 z_2} \dots P_{z_{n-2} z_{n-1}} \right) \\
&= - \sum_{[z_0,\dots,z_{n-1}]} p_{z_0} P_{z_0 z_1} P_{z_1 z_2} \dots P_{z_{n-2} z_{n-1}} (\log p_{z_0} + \log P_{z_0 z_1} + \log P_{z_1 z_2} \\
&\qquad\qquad + \log P_{z_2 z_3} + \dots + \log P_{z_{n-2} z_{n-1}}) \\
&= - \sum_{i=1}^{k} p_i \log p_i - (n-1) \sum_{i,j=1}^{k} p_i P_{ij} \log P_{ij}
\end{aligned}$$

(where we use that $pP = p$ and P is stochastic).

Therefore we see that

$$h_\mu(T) = \lim_{n\to+\infty} \frac{1}{n} H(\vee_{k=0}^{n-1} T^{-k}\alpha) = - \sum_{i,j=1}^{k} p_i P_{ij} \log P_{ij}.$$

In the special case that $X = X_k = \prod_{-\infty}^{+\infty}\{1, \dots, k\}$ we can define a "Bernoulli measure" from a probability vector $(p_1, \dots, p_k)$ (i.e. $(p_1 + \dots + p_k = 1)$ by

$$\mu([z_0, \dots, z_{n-1}]) = p_{z_0}p_{z_1}p_{z_2}\dots p_{z_{n-1}}.$$

The measure theoretic entropy in this case is $h_\mu(T) = \sum_{i=1}^k p_i \log p_i$. For example, where $X = X_2$ and $p = (\frac{1}{2}, \frac{1}{2})$ we have that $h_\mu(T) = \log 2$. Where $X = X_3$ and $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ we have that $h_\mu(T) = \log 3$, etc.

8.4 The increasing martingale theorem

We now begin to develop some of the machinery need to prove the results in section 8.3.

We know that if $f \in L^1(X, \mathcal{B}, \mu)$ and $\mathcal{A} \subset \mathcal{B}$ is a sub-sigma-algebra then we can associate the conditional expectation $E(f|\mathcal{A}) \in L^1(X, \mathcal{A}, \mu)$. The increasing martingale theorem describes how $E(f|\mathcal{A})$ depends on the sigma-algebra $\mathcal{A}$. This is crucial in understanding the corresponding behaviour of the information function and thus the measure theoretic entropy.

The following simple lemma is very useful.

LEMMA 8.9. *If $(X, \mathcal{B}, \mu)$ is a probability space and if $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \dots \subset \mathcal{B}_N \subset \mathcal{B}$ are sigma-algebras and $\lambda > 0$ then if we let*

$$E = \{x \in X\colon \max_{1 \le n \le N} E(f|\mathcal{B}_n)(x) > \lambda\}$$

then we have the upper bound on its measure $\mu(E) \le \frac{1}{\lambda}\int |f| d\mu$ with $f \in L^1(X, \mathcal{B}, \mu)$.

PROOF. Without loss of generality we can assume that $f \ge 0$ (otherwise we replace f by $\max\{f(x), 0\}$). We can partition $E = E_1 \cup \dots \cup E_N$ where

$$E_n = \{x \in X : E(f|\mathcal{B}_n)(x) > \lambda, E(f|\mathcal{B}_i)(x) \le \lambda, i = 1, 2, \dots, n-1\}$$

(and observe that $E_n \in \mathcal{B}_n$); then $E_i \cap E_j = \emptyset$ for $i \ne j$. We then write

$$\int_E f d\mu = \sum_{n=1}^N \int_{E_n} f d\mu = \sum_{n=1}^N \int_{E_n} E(f|\mathcal{B}_n) d\mu \ge \sum_{n=1}^N \lambda\mu(E_n) = \lambda\mu(E).$$

Thus $\mu(E) \le \frac{1}{\lambda}\int f d\mu = \frac{1}{\lambda}\int |f| d\mu$.

■

REMARK. This is very similar to the Chebyshev inequality for $f \in L^1(X, \mathcal{B}, \mu)$ and $\lambda > 0$ which says that $\mu\{x \in X : f(x) > \lambda\} \le \frac{\int |f| d\mu}{\lambda}$.

This brings us to the main result of this section.

THEOREM 8.10 (INCREASING MARTINGALE THEOREM). *Let $f \in L^1(X, \mathcal{B}, \mu)$. Assume that $\mathcal{B}_1 \subset \mathcal{B}_2 \subset \ldots \subset \mathcal{B}_n \subset \ldots \subset \mathcal{B}$ is an increasing sequence of sigma-algebras and that the union $\cup_{n=1}^{\infty} \mathcal{B}_n$ generates $\mathcal{B}$ (written $\mathcal{B}_n \to \mathcal{B}$). Then $E(f|\mathcal{B}_n) \to f$ in $L^1(X, \mathcal{B}, \mu)$ and $E(f|\mathcal{B}_n)(x) \to f(x)$, almost everywhere.*

PROOF. The theorem is clearly true on the subspace $\cup_{k=1}^{\infty} L^1(X, \mathcal{B}_k, \mu)$ since if $g \in L^1(X, \mathcal{B}_k, \mu)$ then $E(g|\mathcal{B}_n) = g$ for $n \geq k$. Moreover, this subspace is dense in $L^1(X, \mathcal{B}, \mu)$ in the L^1 norm.

Given an arbitrary $f \in L^1(X, \mathcal{B}, \mu)$ we can choose $\epsilon > 0$ and $g \in L^1(X, \mathcal{B}_k, \mu)$, say, with $\int |f - g| d\mu < \epsilon$. We then see that for any $n \geq k$ we have that

$$\begin{aligned}
&\int |E(f|\mathcal{B}_n) - f| d\mu \\
&\leq \int |E(f|\mathcal{B}_n) - E(g|\mathcal{B}_n)| d\mu + \int |E(g|\mathcal{B}_n) - g| d\mu + \int |g - f| d\mu \\
&\leq 2 \int |g - f| d\mu
\end{aligned}$$

(where $\int |E(g|\mathcal{B}_n) - g| d\mu = 0$ since $E(g|\mathcal{B}_n) = g$ and we use that $E(.|\mathcal{B}_n)$ is a contraction on $L^1(X, \mathcal{B}, \mu)$). In particular, $\limsup_{n\to+\infty} \int |E(f|\mathcal{B}_n) - f| d\mu \leq 2\epsilon$. Since $\epsilon > 0$ is arbitrary, we see that we have L^1 convergence.

To show that we also have almost everywhere convergence, we argue as follows:

$$\begin{aligned}
&\mu\{x \in X : \limsup_{n\to+\infty} |E(f|\mathcal{B}_n)(x) - f(x)| > \epsilon^{1/2}\} \\
&\leq \mu\{x \in X : \limsup_{n\to+\infty} (|E((f-g)|\mathcal{B}_n)(x) - (f-g)(x)| \\
&\qquad\qquad + |E(g|\mathcal{B}_n)(x) - g(x)|) > \epsilon^{1/2}\} \\
&\leq \mu\{x \in X : \limsup_{n\to+\infty} |E((f-g)|\mathcal{B}_n)(x)| + |(f-g)(x)| > \epsilon^{1/2}\} \\
&\leq \mu\{x \in X : \limsup_{n\to+\infty} |E((f-g)|\mathcal{B}_n)(x)| > \frac{1}{2}\epsilon^{1/2}\} \\
&\quad + \mu\{x \in X : |(f-g)(x)| > \frac{1}{2}\epsilon^{1/2}\} \\
&\leq 2\left(\frac{1}{\frac{1}{2}\epsilon^{1/2}}\right) \int |f - g| d\mu \leq 2\left(\frac{1}{\frac{1}{2}\epsilon^{\frac{1}{2}}}\right) \epsilon \leq 4\epsilon^{1/2}
\end{aligned}$$

(where we have used Lemma 8.9 and the Chebyshev inequality). Since $\epsilon > 0$ is arbitrary this shows almost everywhere convergence.

■

REMARK. There is a corresponding "decreasing martingale theorem", but we shall not need it.

8.5 Entropy and sigma-algebras

We want to apply the increasing martingale theorem to the information functions. First we need a simple technical lemma.

LEMMA 8.11. *If α is a partition with $H(\alpha) < +\infty$ and we have sub-sigma-algebras $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \ldots \subset \mathcal{B}$ then*

$$\int \left(\sup_{n \geq 1} I(\alpha|\mathcal{A}_n) \right) d\mu \leq H(\alpha) + 1$$

(and, in particular, $f(x) = \sup_{n\geq 1} I(\alpha|\mathcal{A}_n)(x) \in L^1(X, \mathcal{B}, \mu)$).

PROOF. We can write

$$\int f(x) d\mu(x) = \int_0^\infty F(t) dt \tag{8.4}$$

where $F(t) = \mu\{x \in X : f(x) > t\}$, provided the right hand side of (8.1) is finite.

We can write

$$\begin{aligned} F(t) &= \mu\{x \in X : \sup_{n\geq 1} I(\alpha|\mathcal{A}_n)(x) > t\} \\ &= \mu \left\{ x \in X : \sup_{n\geq 1} \left(-\sum_{A\in\alpha} \chi_A(x) \log \mu(A|\mathcal{A}_n)(x) \right) > t \right\} \\ &= \sum_{A\in\alpha} \mu \left(A \cap \{x \in X : \sup_{n\geq 1} (-\log \mu(A|\mathcal{A}_n)(x)) > t\} \right) \end{aligned}$$

(since the sets A are disjoint). However, we can simplify this by writing

$$\begin{aligned} &\{x \in X : \sup_{n\geq 1} (-\log \mu(A|\mathcal{A}_n)(x)) > t\} \\ &= \{x \in X : \inf_{n\geq 1} (\log \mu(A|\mathcal{A}_n)(x)) < -t\} \\ &= \{x \in X : \inf_{n\geq 1} (\mu(A|\mathcal{A}_n)(x)) < e^{-t}\} = \cup_{n\geq 1} A_n \end{aligned}$$

where $A_n = \{x \in X : \mu(A|\mathcal{A}_n)(x) < e^{-t}$ and $\mu(A|\mathcal{A}_i)(x) \geq e^{-t}$, for $i = 1, \ldots, n-1\}$ are disjoint sets. If we write

$$F(t) = \sum_{A\in\alpha} \mu\left(A \cap (\cup_{n\geq 1} A_n)\right) = \sum_{A\in\alpha} \sum_{n\geq 1} \mu\left(A \cap A_n\right)$$

then we can use the estimates

$$\mu(A \cap A_n) = \int_{A_n} \chi_A d\mu = \int_{A_n} E(\chi_A|\mathcal{A}_n) d\mu \leq \int_{A_n} e^{-t} d\mu = e^{-t} \mu(A_n).$$

We now have two possible upper bounds on the same summation:

$$\sum_{n\geq 1} \mu(A \cap A_n) \leq \sum_n e^{-t}\mu(A_n) = e^{-t} \quad \text{and} \quad \sum_{n\geq 1} \mu(A\cap A_n) \leq \mu(A).$$

Therefore $F(t) \leq \sum_{A\in\alpha} \min\{e^{-t}, \mu(A)\}$. Finally, we can use this bound to estimate

$$\begin{aligned}\int_0^\infty F(t)dt &\leq \int_0^\infty \left(\sum_{A\in\alpha} \min\{e^{-t}, \mu(A)\}\right) dt \\ &= -\sum_{A\in\alpha}\left(\mu(A)\log\mu(A) - \int_{-\log\mu(A)}^\infty e^{-t}dt\right) \\ &= -\sum_{A\in\alpha}(\mu(A)\log\mu(A) - \mu(A)) \\ &= H(\alpha) + 1.\end{aligned}$$

■

We are now in a position to prove the following crucial result.

THEOREM 8.12. *If α is a partition with $H(\alpha) < +\infty$ and $\mathcal{A}_1 \subset \mathcal{A}_2 \subset \ldots \to \mathcal{B}$ is an increasing sequence of sub-sigma-algebra then $I(\alpha|\mathcal{A}_n)(x) \to I(\alpha|\mathcal{B})(x)$ almost everywhere and in L^1. Thus $H(\alpha|\mathcal{A}_n) \to H(\alpha|\mathcal{B})$ as $n \to +\infty$.*

PROOF. By Theorem 8.10 $\mu(A|\mathcal{A}_n) \to \mu(A|\mathcal{B})$ almost everywhere, for any $A \in \alpha$. This implies that $I(\alpha|\mathcal{A}_n)(x) \to I(\alpha|\mathcal{B})(x)$ almost everywhere.

By Lemma 8.11 we have that $I(\alpha|\mathcal{A}_n)$ are dominated by the integrable function $\sup_{n\geq 1} I(\alpha|\mathcal{A}_n)(x)$. Thus by Lebesgue's dominated convergence theorem we have that $I(\alpha|\mathcal{A}_n)(x) \to I(\alpha|\mathcal{B})(x)$ in L^1 (i.e. $\int |I(\alpha|\mathcal{A}_n)(x) - I(\alpha|\mathcal{B})(x)|d\mu(x) \to 0$ as $n \to +\infty$).

Integrating shows the corresponding result for measure theoretic entropy i.e. $H(\alpha|\mathcal{A}_n)(x) \to H(\alpha|\mathcal{B})(x)$ as $n \to +\infty$.

■

Using Theorem 8.12 we can extend the basic identities (Lemma 8.2 and Lemma 8.4) to arbitrary sub-sigma-algebras $\mathcal{C} \subset \mathcal{B}$, i.e.

(1) for the information functions

$$I(\alpha \vee \beta|\mathcal{C}) = I(\alpha|\hat{\beta} \vee \mathcal{C}) + I(\beta|\mathcal{C})$$

(and, in particular, $I(\alpha|\mathcal{C}) \geq I(\beta|\mathcal{C})$ if $\alpha > \beta$),

(2) for the measure theoretic entropy

$$H_\mu(\alpha \vee \beta|\mathcal{C}) = H_\mu(\alpha|\hat{\beta} \vee \mathcal{C}) + h_\mu(\beta|\mathcal{C})$$

(and, in particular, $H_\mu(\alpha|\mathcal{C}) \geq H_\mu(\beta|\mathcal{C})$ if $\alpha > \beta$).

This only requires that we choose partitions γ such that $\hat{\gamma} \to \mathcal{C}$ and apply the theorem (and this is a basic property of Lebesgue spaces).

8.6 Conditional entropy

We want to consider how changing the sigma-algebra (with the same partition) affects the conditional measure theoretic entropy. The following lemmas are useful.

LEMMA 8.13. *Assume that $f \in L^1(X, \mathcal{B}, \mu)$ and $0 \le f(x) \le 1$ a.e. and that $\mathcal{A} \subset \mathcal{B}$ is a sub-sigma-algebra. Let $\psi : [0,1] \to \mathbb{R}$ be a concave function (i.e. $\psi(\alpha x + (1-\alpha)y) \ge \alpha\psi(x) + (1-\alpha)\psi(y)$); then $\psi(E(f|\mathcal{A})) \ge E(\psi(f)|\mathcal{A})$.*

PROOF. First consider the case of simple functions $f(x) = \sum_{i=1}^n b_i \chi_{B_i}$, where $\{B_1, \dots, B_n\}$ is a partition for X.

By linearity of $E(.|\mathcal{A})$,

$$E(f|\mathcal{A})(x) = \sum_{i=1}^n b_i E(\chi_{B_i}|\mathcal{A})(x) = \sum_{i=1}^n b_i \mu(B_i|\mathcal{A})(x)$$

and observe that $\sum_{i=1}^n \mu(B_i|\mathcal{A})(x) = 1$. We can compute

$$\psi\left(E(f|\mathcal{A})\right) \ge \sum_{i=1}^n \psi(b_i)\mu(B_i|\mathcal{A}) = E(\sum_{i=1}^n \psi(b_i)\chi_{B_i}|\mathcal{A}) = E(\psi(f)|\mathcal{A}). \quad (8.5)$$

For an arbitrary function $f \in L^1(X, \mathcal{B}, \mu)$ we can choose a monotonically increasing sequence of step functions f_k increasing to f a.e. Since $E(\cdot|\mathcal{A})$ takes positive functions to positive functions we see that $E(f_k|\mathcal{A}) \to E(f|\mathcal{A})$. We can now take limits in (8.5) to get that $\psi\left(E(f|\mathcal{A})\right) \ge E(\psi(f)|\mathcal{A})$, as required.

∎

The following is a simple application of this result.

PROPOSITION 8.14. *If β is a partition and $\mathcal{A}_2 \subset \mathcal{A}_1 \subset \mathcal{B}$ are sub-sigma-algebras then $H(\beta|\mathcal{A}_1) \le H(\beta|\mathcal{A}_2)$.*

(The corresponding result for information functions may not be true).

PROOF. For each $B \in \beta$ we fix the choice $f = \mu(B|\mathcal{A}_1)$. We then fix $\psi(t) = -t\log(t)$ for $0 < t \le 1$ (and $\psi(0) = 0$). We then have that $\psi(f) = -\mu(B|\mathcal{A}_1)\log\left(\mu(B|\mathcal{A}_1)\right)$ and the lemma (Jenson's inequality) gives that

$$-E\left(\mu(B|\mathcal{A}_1)\log\left(\mu(B|\mathcal{A}_1)\right)|\mathcal{A}_2\right) \le -\mu(B|\mathcal{A}_2)\log\mu(B|\mathcal{A}_2).$$

Integrating both sides with respect to μ (and summing over $B \in \beta$) gives that $H(\beta|\mathcal{A}_1) \le H(\beta|\mathcal{A}_2)$.

∎

8.7 Proofs of Lemma 8.7 and Lemma 8.8

We can now use the results from the preceding sections to supply the omitted proofs of Lemmas 8.7 and 8.8.

PROOF OF LEMMA 8.7. We know that $h(T, \beta_n) \leq h(T, \beta_m) + H(\beta_n | \hat{\beta}_m)$ for $n, m \geq 1$. Moreover, if $m \geq n$ then $\beta_n \subset \beta_m$ and so $H(\beta_n | \beta_m) = 0$. Thus $h(T, \beta_n)$ is monotonically decreasing and so converges.

For any partition α with $H(\alpha) < +\infty$ we can start with the inequality $h(T, \alpha) \leq h_\mu(T, \beta_n) + H(\alpha | \beta_n)$ (proved in a previous lemma). By a corollary to the increasing martingale theorem we know that $H(\alpha | \beta_n) \to H(\alpha | \mathcal{B}) = 0$ (since $\alpha \subset \mathcal{B}$). We conclude that

$$h(T, \alpha) \leq \limsup_{n \to +\infty} h(T, \beta_n).$$

Taking the supremum over all such α gives that

$$h_\mu(T) = \sup_\alpha h(T, \alpha) \leq \lim_{n \to +\infty} h(T, \beta_n).$$

Clearly, $h_\mu(T) \geq h(T, \beta_n)$ for $n \geq 1$. Thus

$$h_\mu(T) \geq \sup_n h(T, \beta_n) \geq \lim_{n \to +\infty} h(T, \beta_n)$$

and the proof is complete.

■

Before moving on to the proof of Lemma 8.8, we recall the following useful fact.

A FACT ABOUT LEBESGUE SPACES. For Lebesgue spaces a necessary and sufficient condition for $\beta_n \to \mathcal{B}$ is that there exists a set of zero measure $N \subset X$ such that for $x, y \in X - N$ (with $x \neq y$) there exist $n \geq 1$ and $B \in \beta_n$ such that $x \in B$ but $y \notin B$.

PROOF OF LEMMA 8.8. This is an application of Abramov's result where we take $\beta_n = \vee_{i=-n}^{n} T^{-i}\alpha$ or $\beta_n = \vee_{i=0}^{n} T^{-i}\alpha$, as appropriate. We then have that

$$\begin{aligned} h(T, \beta_k) &= H(\beta_k | \vee_{i=1}^{\infty} T^{-i}\beta_k) \\ &= H(\vee_{i=1}^{k-1} T^{-i}\alpha | \vee_{i=1}^{\infty} T^{-i}\alpha) \\ &= h(T, \alpha) \end{aligned}$$

and we need only let $k \to +\infty$.

■

8.8 Isomorphism

Entropy is very important in the classification of measure preserving transformations. We begin with a definition.

DEFINITION. Let $(X_i, \mathcal{B}_i, \mu_i)$, for $i = 1, 2$, be probability spaces; then an *isomorphism* between measure preserving transformations $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ is a map $\phi : X_1 \to X_2$ such that

(1) ϕ is a bijection (after removing sets of zero measure, if necessary),
(2) both ϕ and ϕ^{-1} are measurable (i.e. $\phi^{-1}\mathcal{B}_2 \subset \mathcal{B}_1$ and $\phi\mathcal{B}_1 \subset \mathcal{B}_2$),
(3) $\mu_1(\phi^{-1}B) = \mu_2(B)$ for $B \in \mathcal{B}_2$ (also $\mu_2(\phi B) = \mu_1(B)$ for $B \in \mathcal{B}_1$),
(4) $\phi \circ T_1 = T_2 \circ \phi$.

THEOREM 8.15. *Entropy is an isomorphism invariant (i.e. if $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ are isomorphic then $h_{\mu_1}(T_1) = h_{\mu_2}(T_2)$).*

PROOF. Let α be a partition for X_2; then clearly $\phi^{-1}\alpha = \{\phi^{-1}(A) : A \in \alpha\}$ is a partition for X_1. Moreover, the properties for ϕ imply that

$$\frac{1}{n}H\left(\vee_{i=0}^{n-1} T_2^{-i}\alpha\right) = \frac{1}{n}H\left(\vee_{i=0}^{n-1} T_1^{-i}\phi^{-1}\alpha\right).$$

Letting $n \to +\infty$ gives that $h_{\mu_1}(T_1, \phi^{-1}\alpha) = h_{\mu_2}(T_2, \alpha)$. Taking the supremum over all partitions α (with $H(\alpha) < +\infty$) gives the result.

■

EXAMPLE (BERNOULLI SHIFTS). Using the formula for the entropy of a Bernoulli measure from section 8.3 we see that the shifts

$$\begin{cases} \sigma : \prod_{n\in\mathbb{Z}}\{1,2\} \to \prod_{n\in\mathbb{Z}}\{1,2\} \text{ with probability vector } (\frac{1}{2}, \frac{1}{2}), \\ \sigma : \prod_{n\in\mathbb{Z}}\{1,2\} \to \prod_{n\in\mathbb{Z}}\{1,2,3\} \text{ with probability vector } (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}) \end{cases}$$

have entropies $\log 2$ and $\log 3$, respectively. Therefore they are *not* isomorphic.

REMARK. Let us consider a slightly different situation where we drop the assumption that ϕ is a bijection. That is, if we consider $(X_i, \mathcal{B}_i, \mu_i)$, for $i = 1, 2$, to be probability spaces then a *factor map* between $T_1 : X_1 \to X_1$ and $T_2 : X_2 \to X_2$ satisfies

(1) $\phi(X_1)$ is equal to X_2 (after removing a set of zero μ_2-measure, if necessary),
(2) ϕ is measurable,
(3) $\mu_1(\phi^{-1}B) = \mu_2(B)$ for $B \in \mathcal{B}_2$,
(4) $\phi \circ T_1 = T_2 \circ \phi$.

In this case it is easy to see that $h_\mu(T_1) \geq h_\mu(T_2)$.

We usually say that T_1 is an *extension* of T_2 or that T_2 is an *factor* of T_1.

8.9 Comments and references

Our development of entropy has followed the lines of Parry's treatment in [2]. It is possible to reduce some of the analysis if we accept working only with countable sigma-algebras [5].

References for some more advanced topics we have omitted include [3, §7.5] (Krieger's generator theorem), [3, §7.6] (Ornstein's isomorphism theorem), and [1, §10.7] (Keane-Smorodinsky finitary Isomorphism Theorem).

References

1. I. Cornfold, S. Fomin and Y. Sinai, *Ergodic Theory*, Springer, Berlin, 1982.
2. W. Parry, *Topics in Ergodic Theory*, C.U.P., Cambridge, 1981.
3. D. Rudolph, *Fundamentals of Measurable Dynamics*, O.U.P., Oxford, 1990.
4. Y. Sinai, *Topics in Ergodic Theory*, P.U.P., Princeton N.J., 1994.
5. P. Walters, *An Introduction to Ergodic Theory*, Springer, Berlin, 1982.