Chapter 2
Conditional Measure-Theoretic Entropy

The basic entropy theory from Chapter 1 will become a more powerful and flexible tool after we extend the theory from partitions to σ-algebras.

However, infinite sub-σ-algebras do not share all the properties of finite ones, and in particular the correspondence between partitions and sub-σ-algebras is less straightforward. If ξ is a countably infinite partition, then $\sigma(\xi)$ is in general an uncountable σ-algebra. However, σ-algebras of the form $\sigma(\xi)$ where ξ is a countable partition are rather special, and should not be confused with the much larger class of countably-generated σ-algebras.

2.1 Conditional Measures, Factors, and Invariant σ-Algebras

The material in this chapter (and the later use of conditional entropy) uses some less standard material in measure theory, so in order to make this volume reasonably self-contained we collect the results needed from [52]. Throughout this section, let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system, where $(X, \mathscr{B}, \mu)$ is a *Borel probability space*. We recall that a Borel probability space is a dense Borel subset X of a compact metric space $\overline{X}$, with a probability measure μ defined on the restriction of the Borel σ-algebra $\mathscr{B}$ to X. We remark only that this restriction in generality could be avoided in most statements by using the conditional expectation instead of conditional measures. However, we feel that the intuition of this more geometric approach provided by conditional measures outweighs the slight restriction in generality.

If $\mathscr{A} \subseteq \mathscr{B}$ is a countably generated sub-σ-algebra in $(X, \mathscr{B}, \mu)$, then there exists a family of conditional measures $\{\mu_x^{\mathscr{A}} \mid x \in X\}$ which decompose μ into probability measures that give full measure to the corresponding atom $[x]_{\mathscr{A}}$, where the atom

$$[x]_{\mathscr{A}} = \bigcap_{x \in A \in \mathscr{A}} A$$

is the smallest element of $\mathscr{A}$ containing x. In a sense which we will make precise below, $\mu_x^{\mathscr{A}}$ describes μ restricted to the atom $[x]_{\mathscr{A}}$ in a way that makes sense even if the atoms are null set. We also note that if $\mathscr{A} = \sigma(\xi)$ for a partition ξ, then the atoms are just the elements of the partition and the conditional measures are simply the renormalized restriction of μ to the atoms. However, if $\mathscr{A}$ is a general countably-generated σ-algebra, then X decomposes[†] into uncountably many atoms $[x]_{\mathscr{A}}$ for $x \in X$ that typically have zero measure for μ, and so the construction of the conditional measures requires some care.

2.1.1 Defining properties of conditional measures

We refer to [52, Ch. 5] for the details of this construction, but recall the main results and properties of these measures. We begin by recalling the conditional expectation as this will in turn characterize the conditional measures.

Theorem 2.1 (Conditional expectation). *Let $(X, \mathscr{B}, \mu)$ be a probability space, and let $\mathscr{A} \subseteq \mathscr{B}$ be a sub-σ-algebra. Then there is a map*

$$E_\mu(\cdot \big| \mathscr{A}) : L^1(X, \mathscr{B}, \mu) \longrightarrow L^1(X, \mathscr{A}, \mu),$$

called the conditional expectation, that satisfies the following properties.

(1) *For $f \in L^1(X, \mathscr{B}, \mu)$, the image function $E_\mu(f \big| \mathscr{A})$ is characterized almost everywhere by the two properties*
   - *$E_\mu(f \big| \mathscr{A})$ is $\mathscr{A}$-measurable;*
   - *for any $A \in \mathscr{A}$, $\int_A E_\mu(f \big| \mathscr{A})\,\mathrm{d}\mu = \int_A f\,\mathrm{d}\mu$.*

(2) *$E_\mu(\cdot \big| \mathscr{A})$ is a linear operator of norm 1. Moreover, $E_\mu(\cdot \big| \mathscr{A})$ is positive (that is, $E_\mu(f \big| \mathscr{A}) \geqslant 0$ almost everywhere whenever $f \in L^1(X, \mathscr{B}, \mu)$ has $f \geqslant 0$).*

(3) *For $f \in L^1(X, \mathscr{B}, \mu)$ and $g \in L^\infty(X, \mathscr{A}, \mu)$,*

$$E_\mu(gf \big| \mathscr{A}) = g E_\mu(f \big| \mathscr{A})$$

almost everywhere.

(4) *If $\mathscr{A}' \subseteq \mathscr{A}$ is a sub-σ-algebra, then*

[†] In that sense a countably generated σ-algebra also gives rise to a (potentially uncountable) measurable partition of X. For that reason some authors also speak of a measurable partition when discussing a countably generated σ-algebra. However, for us the word partition will always mean a finite or countable decomposition of X.

$$E_\mu\left(E_\mu(f\big|\mathscr{A})|\mathscr{A}'\right) = E_\mu(f\big|\mathscr{A}')$$

almost everywhere.

The conditional expectation $E_\mu(\cdot\big|\mathscr{A})$ may be thought of as the natural projection map from $L^1(X,\mathscr{B},\mu)$ to its closed subspace $L^1(X,\mathscr{A},\mu)$. In fact the conditional expectation can be constructed (by continuous continuation) using the orthogonal projection from the Hilbert space $L^2(X,\mathscr{B},\mu)$ to its closed subspace $L^2(X,\mathscr{A},\mu)$ (see [52, Sect. 5.1]). When the underlying measure μ is understood we will also simply write $E(\cdot\big|\mathscr{A}) = E_\mu(\cdot\big|\mathscr{A})$ for the conditional expectation.

We say that two σ-algebras $\mathscr{A}$ and $\mathscr{C}$ are equivalent modulo μ, denoted

$$\mathscr{A} \underset{\mu}{=} \mathscr{C},$$

if for any $A \in \mathscr{A}$ there exists $C \in \mathscr{C}$ with $\mu(A \triangle C) = 0$ and for any $C \in \mathscr{C}$ there exists $A \in \mathscr{A}$ with $\mu(A \triangle C) = 0$. Similarly, partitions ξ and η are equivalent modulo μ,

$$\xi \underset{\mu}{=} \eta,$$

if the σ-algebras they generate are equal modulo μ (see Exercise 2.1.1).

Using the above we can now state the defining property of the conditional measures.

Theorem 2.2 (Conditional measure). *Let $(X,\mathscr{B},\mu)$ be a Borel probability space, and $\mathscr{A} \subseteq \mathscr{B}$ a σ-algebra. Then there exists an $\mathscr{A}$-measurable conull set $X' \subseteq X$ and a system $\{\mu_x^{\mathscr{A}} \mid x \in X'\}$ of measures on X, referred to as* conditional measures*, with the following properties.*

(1) *The measure $\mu_x^{\mathscr{A}}$ is a probability measure on X with*

$$E(f\big|\mathscr{A})(x) = \int f(y)\,\mathrm{d}\mu_x^{\mathscr{A}}(y) \tag{2.1}$$

almost everywhere for all $f \in \mathscr{L}^1(X,\mathscr{B},\mu)$. In other words, for any function[†] *$f \in \mathscr{L}^1(X,\mathscr{B},\mu)$ we have that $\int f(y)\,\mathrm{d}\mu_x^{\mathscr{A}}(y)$ exists for all x belonging to a conull set in $\mathscr{A}$, that on this set*

$$x \longmapsto \int f(y)\,\mathrm{d}\mu_x^{\mathscr{A}}(y)$$

depends $\mathscr{A}$-measurably on x, and that

$$\int_A \int f(y)\,\mathrm{d}\mu_x^{\mathscr{A}}(y)\,\mathrm{d}\mu(x) = \int_A f\,\mathrm{d}\mu$$

[†] Notice that we are forced to work with genuine functions in $\mathscr{L}^1$ in order to ensure that the right-hand side of (2.1) is defined. As we said before, $\mu_x^{\mathscr{A}}$ may be singular to μ.

for all $A \in \mathscr{A}$.

(2) *If* $\mathscr{A}$ *is countably-generated, then* $\mu_x^{\mathscr{A}}([x]_{\mathscr{A}}) = 1$ *for all* $x \in X'$, *where*

$$[x]_{\mathscr{A}} = \bigcap_{x \in A \in \mathscr{A}} A$$

is the atom of $\mathscr{A}$ *containing* x; *moreover* $\mu_x^{\mathscr{A}} = \mu_y^{\mathscr{A}}$ *for* $x, y \in X'$ *whenever* $[x]_{\mathscr{A}} = [y]_{\mathscr{A}}$.

(3) *Property* (1) *uniquely determines* $\mu_x^{\mathscr{A}}$ *for almost every* $x \in X$. *In fact, property* (1) *for a dense countable set of functions in* $C(\overline{X})$ *uniquely determines* $\mu_x^{\mathscr{A}}$ *for almost every* $x \in X$.

(4) *If* $\mathscr{C}$ *is any* σ*-algebra with* $\mathscr{A} \underset{\mu}{=} \mathscr{C}$, *then* $\mu_x^{\mathscr{A}} = \mu_x^{\mathscr{C}}$ *almost everywhere.*

This is simply [52, Th. 5.14] and we only note that the conditional measures can be constructed from the conditional expectation applied to sufficiently many continuous functions on $\bar{X}$ (as in Property (3)). The defining Property (1) in Theorem 2.2 we sometimes abbreviate by writing

$$\mu = \int \mu_x^{\mathscr{A}} \, \mathrm{d}\mu$$

and saying that $\mu_x^{\mathscr{A}}$ depends on $x \in X$ in an $\mathscr{A}$-measurable way.

We recall from [52, Lemma 5.17] that for any σ-algebra $\mathscr{A} \subseteq \mathscr{B}$ and any probability measure μ there exists a countably generated σ-algebra

$$\mathscr{C} \underset{\mu}{=} \mathscr{A}.$$

This allows us to switch using Property (4) to $\mathscr{C}$ and then to talk about its atoms using Property (2).

We note that for a countably generated σ-algebra $\mathscr{A}$ the atoms are indeed elements of the σ-algebra: If $\mathscr{C}$ is a countably generating set of $\mathscr{A}$ (which in general will be uncountable), then we can generate a countable algebra $\mathscr{C}'$ out of $\mathscr{C}$ and obtain that

$$[x]_{\mathscr{A}} = [x]_{\mathscr{C}'} = \bigcap_{x \in C \in \mathscr{C}'} C \in \mathscr{A}$$

since the latter intersection is a countable intersection.

We finally also recall from [52, Sect. 5.3] the following example which clarifies why we should think of Theorem 2.2 as a grand generalization of Fubini's theorem.

Example 2.3. Let $X = [0,1]^2$ and $\mathscr{A} = \mathscr{B} \times \{\varnothing, [0,1]\}$. Theorem 2.2 says that any Borel probability measure μ on X can be decomposed into vertical components in the following sense: the conditional measures $\mu_{(x_1,x_2)}^{\mathscr{A}}$ are defined on the line segments $\{x_1\} \times [0,1]$, and these sets are precisely the atoms of $\mathscr{A}$. Moreover,

$$\mu(B) = \int_X \mu^{\mathscr{A}}_{(x_1,x_2)}(B)\,\mathrm{d}\mu(x_1,x_2). \tag{2.2}$$

Here $\mu^{\mathscr{A}}_{(x_1,x_2)} = \nu_{x_1}$ does not depend on x_2, so (2.2) may be written as

$$\mu(B) = \int_{[0,1]} \nu_{x_1}(B)\,\mathrm{d}\overline{\mu}(x_1)$$

where $\overline{\mu} = \pi_*\mu$ is the measure on $[0,1]$ obtained by the projection

$$\begin{aligned}\pi : [0,1]^2 &\longrightarrow [0,1]\\ (x_1,x_2) &\longmapsto x_1.\end{aligned}$$

In this section we have considered functions in $\mathscr{L}^1$; we refer to Exercise 2.1.2 for the non-negative measurable case.

2.1.2 Structural properties of conditional measures

As the conditional measures are often singular to the original measures, null sets for μ require some additional care. However, the following general remark helps to a large extent. If $\mathscr{A} \subseteq \mathscr{B}$ is a sub-σ-algebra, and $N \in \mathscr{B}$ is a null set for μ, then $\mu^{\mathscr{A}}_x(N) = 0$ for μ-almost every $x \in X$. This is a simple consequence of Theorem 2.2(1), which we will use frequently without explicit reference.

We recall from [52, Prop. 5.20] that conditional measures 'commute' with refinement in the following sense. This is simply a formulation of Property (4) in Theorem 2.1 for conditional expectation.

Proposition 2.4 (Double conditioning). *Let $(X,\mathscr{B},\mu)$ be a Borel probability space, and let*

$$\mathscr{A}' \subseteq \mathscr{A} \subseteq \mathscr{B}$$

be countably-generated sub-σ-algebras. Then $[z]_{\mathscr{A}} \subseteq [z]_{\mathscr{A}'}$ for $z \in X$, and for almost every $z \in X$ the conditional measures for the measure $\mu^{\mathscr{A}'}_z$ with respect to $\mathscr{A}$ are given for $\mu^{\mathscr{A}'}_z$-almost every $x \in [z]_{\mathscr{A}'}$ by $(\mu^{\mathscr{A}'}_z)^{\mathscr{A}}_x = \mu^{\mathscr{A}}_x$.

The next result, taken from [52, Cor. 5.24] describes how conditional measures behave with respect to measure-preserving maps.

Lemma 2.5 (Push-forward of conditional measures). *Let*

$$\phi\colon (X,\mathscr{B}_X,\mu) \to (Y,\mathscr{B}_Y,\nu)$$

be a measure-preserving map between Borel probability spaces, and let $\mathscr{A}$ be a sub-σ-algebra of $\mathscr{B}_Y$. Then

$$\phi_*\mu^{\phi^{-1}\mathscr{A}}_x = \nu^{\mathscr{A}}_{\phi(x)}$$

for μ-almost every $x \in X$.

We note that the above follows from the formula

$$E_\mu(f \circ \phi \big| \phi^{-1}\mathscr{A}) = E_\nu(f \big| \mathscr{A}) \circ \phi$$

for all $f \in L^1(Y, \mathscr{B}_Y, \nu)$, which in turn is easy to check from the characterizing properties of the conditional expectation.

2.1.3 Factors and σ-Algebras

A very common way in which a σ-algebra may arise is from a measurable map $\phi : (X, \mathscr{B}_X) \to (Y, \mathscr{B}_Y)$. Indeed from ϕ we may immediately define the σ-algebra $\mathscr{A} = \phi^{-1}\mathscr{B}_Y$.

Given a general sub-σ-algebra $\mathscr{A} \subseteq \mathscr{B}_X$ it is interesting to note that one can always find a map ϕ for which

$$\mathscr{A} \underset{\mu}{=} \phi^{-1}\mathscr{B}_Y.$$

Indeed one can use $\phi(x) = \mu_x^{\mathscr{A}}$, which takes values in the compact space

$$Y = \mathscr{M}(\bar{X})$$

consisting of probability measures on the ambient compact metric space $\bar{X}$. This can be taken further to give a proof of the following theorem; we refer to [52, Th. 6.5] for the details.

Theorem 2.6 (Factors). *Let $(X, \mathscr{B}_X, \mu)$ be a Borel probability space, and let T be a measure-preserving transformation on X. Assume furthermore that there is a strictly invariant sub-σ-algebra $\mathscr{A} = T^{-1}\mathscr{A} \subseteq \mathscr{B}_X$. Then there is a measure-preserving system $(Y, \mathscr{B}_Y, \nu, S)$ on a Borel probability space and a factor map $\phi : X \to Y$ with $\mathscr{A} = \phi^{-1}\mathscr{B}_Y$ modulo μ. If T is invertible then S may be chosen to be invertible.*

2.1.4 Ergodic Decomposition

Another way in which a σ-algebra $\mathscr{A} \subseteq \mathscr{B}$ will arise in our discussions is the following example. In a measure-preserving system $(X, \mathscr{B}, \mu, T)$ we may define the σ-algebra $\mathscr{E} = \{B \in \mathscr{B} \mid \mu(T^{-1}B \triangle B) = 0\}$ of invariant sets. The notion of conditional measures for this σ-algebra leads to a proof of the ergodic decomposition, which we describe now.

Theorem 2.7 (Ergodic decomposition). *Let $T : (X, \mathscr{B}, \mu) \to (X, \mathscr{B}, \mu)$ be a measure-preserving map of a Borel probability space. Then there is a Borel probability space $(Y, \mathscr{B}_Y, \nu)$ and a measurable map $y \mapsto \mu_y$ for which*

- *μ_y is a T-invariant ergodic probability measure on X for almost every y, and*
- *$\mu = \int_Y \mu_y \,\mathrm{d}\nu(y)$.*

Moreover, we can require that the map $y \mapsto \mu_y$ is injective, or alternatively set $(Y, \mathscr{B}_Y, \nu) = (X, \mathscr{B}, \mu)$ and $\mu_x = \mu_x^{\mathscr{E}}$, where $\mathscr{E}$ is the σ-algebra of T-invariant sets, and $\mu_x^{\mathscr{E}}$ denotes the conditional measure of μ at x with respect to $\mathscr{E}$.

Invariance of $\mu_x^{\mathscr{E}}$ under T almost surely follows from the push-forward formula for conditional measures (Lemma 2.5), but to see almost sure ergodicity some more work is needed (which uses the pointwise ergodic theorem as a tool and as a characterization of ergodicity). We refer to [52, Th. 6.2] for a proof.

Exercises for Section 2.1

Exercise 2.1.1. Let $(X, \mathscr{B}, \mu)$ be a Borel probability space, and let $\mathscr{A}$ and $\mathscr{C}$ be countably generated sub-σ-algebras of $\mathscr{B}$. Show that $\mathscr{A} \underset{\mu}{=} \mathscr{C}$ if and only if there exists a null set N such that the restrictions of $\mathscr{A}$ and $\mathscr{C}$ to $X \smallsetminus N$ are identical.

Exercise 2.1.2. Let $(X, \mathscr{B}, \mu)$ be a Borel probability space, let $\mathscr{A} \subseteq \mathscr{B}$ be a sub σ-algebra, and $f \geqslant 0$ a measurable function. Show that the definition of $E(f|\mathscr{A})(x) = \int f \,\mathrm{d}\mu_x^{\mathscr{A}}$ still satisfies Theorem 2.1(1). Show that if $g \geqslant 0$ is $\mathscr{A}$-measurable, then $E(gf\big|\mathscr{A}) = gE(f\big|\mathscr{A})$ almost everywhere, as in Theorem 2.1(3).

2.2 Conditional Entropy

We now start to generalize the definitions and basic results from the previous chapter by allowing σ-algebras instead of just countable partitions.

Definition 2.8. Let $(X, \mathscr{B}, \mu)$ be a Borel probability space and let $\mathscr{A}, \mathscr{C} \subseteq \mathscr{B}$ be countably-generated sub-σ-algebras. The *information function of $\mathscr{C}$ given (the information of) $\mathscr{A}$* with respect to μ is defined by

$$I_\mu\left(\mathscr{C}\big|\mathscr{A}\right)(x) = -\log \mu_x^{\mathscr{A}}\left([x]_{\mathscr{C}}\right).$$

Moreover, the conditional entropy of $\mathscr{C}$ given $\mathscr{A}$,

$$H_\mu\left(\mathscr{C}\big|\mathscr{A}\right) = \int I_\mu\left(\mathscr{C}\big|\mathscr{A}\right)(x)\,\mathrm{d}\mu(x),$$

is defined to be the average of the information.

We will see later that $x \longmapsto I_\mu\left(\mathscr{C}\middle|\mathscr{A}\right)(x)$ is measurable so that $H_\mu\left(\mathscr{C}\middle|\mathscr{A}\right)$ is well-defined (see Proposition 2.10(2)). Assuming this for the moment we assemble a few basic properties that are easy to see.

(1) We have
$$I_\mu\left(\mathscr{C}\middle|\mathscr{A}\right) = I_\mu\left(\mathscr{A}\vee\mathscr{C}\middle|\mathscr{A}\right)$$
almost everywhere, since $[x]_{\mathscr{A}\vee\mathscr{C}} = [x]_{\mathscr{A}} \cap [x]_{\mathscr{C}}$ and $\mu_x^{\mathscr{A}}([x]_{\mathscr{A}}) = 1$ by Theorem 2.2.

(2) If $\mathscr{C} \underset{\mu}{=} \mathscr{C}'$ are countably generated σ-algebras, then there is a null set N such that
$$[x]_{\mathscr{C}}\smallsetminus N = [x]_{\mathscr{C}'}\smallsetminus N$$
for all $x \in X$. Since $\mu_x^{\mathscr{A}}(N) = 0$ for almost every x by Theorem 2.2, we deduce that
$$I_\mu\left(\mathscr{C}\middle|\mathscr{A}\right) = I_\mu\left(\mathscr{C}'\middle|\mathscr{A}\right)$$
almost everywhere.

(3) Similarly, if $\mathscr{A} \underset{\mu}{=} \mathscr{A}'$ then
$$\mu_x^{\mathscr{A}} = \mu_x^{\mathscr{A}'}$$
almost everywhere, and once again $I_\mu(\mathscr{C}\big|\mathscr{A}) = I_\mu(\mathscr{C}\big|\mathscr{A}')$ almost everywhere.

(4) Finally, notice that if $\mathscr{N} = \{X, \varnothing\}$ is the trivial σ-algebra, then
$$I_\mu(\mathscr{C}\big|\mathscr{N})(x) = I_\mu(\mathscr{C})(x) = -\log\mu([x]_{\mathscr{C}})$$
and
$$H_\mu(\mathscr{C}\big|\mathscr{N}) = H_\mu(\mathscr{C}) = \int I_\mu(\mathscr{C})\,\mathrm{d}\mu$$
is infinite unless $\mathscr{C} = \sigma(\xi)$ is the σ-algebra generated by a countable partition with finite entropy.

Just as in the case of partitions discussed on page 12, the information function of $\mathscr{C}$ given $\mathscr{A}$ at x is a measure of how much additional information is revealed by finding out which atom $[\cdot]_{\mathscr{C}}$ contains x starting from the knowledge of the atom $[x]_{\mathscr{A}}$. This informal description may help to motivate the following discussion, and the reader may find it helpful to find similar informal descriptions of the technical statements below.

Example 2.9. If $\mathscr{C} = \sigma(\xi)$ and $\mathscr{A} = \sigma(\eta)$ for countable partitions ξ and η, then the conditional measure satisfies $\mu_x^{\mathscr{A}} = \frac{1}{\mu([x]_\eta)}\mu|_{[x]_\eta}$, and so
$$I_\mu\left(\sigma(\xi)\middle|\sigma(\eta)\right)(x) = -\log\frac{\mu([x]_\xi \cap [x]_\eta)}{\mu([x]_\eta)},$$

and

$$H_\mu\left(\sigma(\xi)\big|\sigma(\eta)\right) = -\sum_{\substack{P\in\xi,\\ Q\in\eta}} \mu(P\cap Q)\log\frac{\mu(P\cap Q)}{\mu(Q)}.$$

Thus, in this case the definition of $H_\mu(\xi\big|\eta)$ seen in (1.1) and Definition 1.6 for the case of countable partitions is recovered.

Motivated by the above example we will not distinguish between a partition η and the σ-algebra $\sigma(\eta)$ that is generates. We will also write $\eta \vee \mathscr{A}$ or $\mathscr{A} \vee \eta$ as an abbreviation for $\sigma(\eta) \vee \mathscr{A}$ when η is a countable partition and $\mathscr{A}$ is a sub σ-algebra.

2.2.1 Dependence on the sub σ-Algebra whose Information is Measured

In order to justify the definition of $H_\mu(\mathscr{C}\big|\mathscr{A})$ we need to know that $I_\mu(\mathscr{C}\big|\mathscr{A})$ is a measurable function. In addition to this, we shall see that both the information and the entropy are monotone and continuous (in a suitable sense) with respect to the σ-algebra whose information is being computed.

Proposition 2.10 (Monotonicity and Continuity). *Let $(X,\mathscr{B},\mu)$ be a Borel probability space with sub-σ-algebras $\mathscr{A},\mathscr{C},\mathscr{C}_n \subseteq \mathscr{B}$ for $n\in\mathbb{N}$, and assume that $\mathscr{C},\mathscr{C}_n$ are countably generated for all $n \geqslant 1$. Then*

(1) *the map $x \mapsto I_\mu(\mathscr{C}\big|\mathscr{A})(x)$ is measurable;*
(2) *if $\mathscr{C}_1 \subseteq \mathscr{C}_2$ then $I_\mu(\mathscr{C}_1\big|\mathscr{A}) \leqslant I_\mu(\mathscr{C}_2\big|\mathscr{A})$; and*
(3) *if $\mathscr{C}_n \nearrow \mathscr{C}$ is an increasing sequence of σ-algebras then*

$$I_\mu(\mathscr{C}_n\big|\mathscr{A}) \nearrow I_\mu(\mathscr{C}\big|\mathscr{A})$$

and

$$H_\mu(\mathscr{C}_n\big|\mathscr{A}) \nearrow H_\mu(\mathscr{C}\big|\mathscr{A})$$

as $n\to\infty$.

Proof. Property (2) follows from the fact that $[x]_{\mathscr{C}_2} \subseteq [x]_{\mathscr{C}_1}$ for $x \in X$ and the definition. For (1), first consider the case where $\mathscr{C} = \sigma(\xi)$ is generated by a countable partition $\xi = \{P_1, P_2, \dots\}$. In this case

$$I_\mu\left(\mathscr{C}\big|\mathscr{A}\right)(x) = \begin{cases} -\log\mu_x^{\mathscr{A}}(P_1) \text{ for } x\in P_1\in\xi, \\ -\log\mu_x^{\mathscr{A}}(P_2) \text{ for } x\in P_2\in\xi, \\ \qquad\vdots \end{cases}$$

so $I_\mu(\mathscr{C}\big|\mathscr{A})$ is measurable. The general case of a countably-generated σ-algebra $\mathscr{C} = \sigma\left(\{C_1, C_2, \dots\}\right)$ follows by defining the sequence (ξ_n) of partitions to have the property that $\mathscr{C}_n = \sigma(\xi_n) = \sigma(\{C_1,\dots,C_n\}) \nearrow \mathscr{C}$ and then applying (3), which we now prove.

Let $\mathscr{C}_n \nearrow \mathscr{C}$ be an increasing sequence of countably-generated σ-algebras. Then the atoms $[x]_{\mathscr{C}_n}$ are shrinking as $n \to \infty$,

$$[x]_{\mathscr{C}} = \bigcap_{n \geqslant 1} [x]_{\mathscr{C}_n},$$

and so

$$\mu_x^{\mathscr{A}}\left([x]_{\mathscr{C}_n}\right) \searrow \mu_x^{\mathscr{A}}\left([x]_{\mathscr{C}}\right)$$

which gives (3) by the definition of $I_\mu(\mathscr{C}\big|\mathscr{A})$ and monotone convergence. □

2.2.2 Dependence on the Given Sub-σ-algebra

We now turn to the properties of information and entropy with respect to the given σ-algebra — that is, properties of the function

$$\mathscr{A} \longmapsto I_\mu(\mathscr{C}\big|\mathscr{A})$$

for fixed $\mathscr{C}$. This is more delicate than the corresponding properties for the function $\mathscr{C} \mapsto I_\mu(\mathscr{C}\big|\mathscr{A})$ for fixed $\mathscr{A}$, considered above. In particular, for most of the following discussion we will assume that $\mathscr{C} = \sigma(\xi)$ for some partition ξ.

The next lemma gives an alternative description of conditional entropy, by showing that $H_\mu(\mathscr{A}\big|\mathscr{C})$ is the average of the entropies $H_{\mu_x^{\mathscr{A}}}(\mathscr{C})$ (notice that this result coincides with the definition (1.1) if $\mathscr{A}$ and $\mathscr{C}$ are σ-algebras generated by countable partitions).

Lemma 2.11 (Conditional entropy equals an average). *Let $(X, \mathscr{B}, \mu)$ be a Borel probability space, with $\mathscr{C}$ and $\mathscr{A}$ sub-σ-algebras of $\mathscr{B}$. Then*

$$H_\mu\left(\mathscr{C}\big|\mathscr{A}\right) = \int H_{\mu_x^{\mathscr{A}}}(\mathscr{C})\,\mathrm{d}\mu(x),$$

where $H_{\mu_x^{\mathscr{A}}}(\mathscr{C})$ is infinite unless $\mathscr{C}$ agrees modulo $\mu_x^{\mathscr{A}}$ with a σ-algebra generated by a countable partition of finite entropy with respect to $\mu_x^{\mathscr{A}}$.

Proof. By monotonicity and continuity of the entropy function with respect to the first σ-algebra (Proposition 2.10) and monotone convergence, it is enough to check the first statement for $\mathscr{C} = \sigma(\xi)$ for a finite partition ξ. In this case the characterizing properties of conditional measures (Theorem 2.2) show that

$$\begin{aligned}\int \mu_x^{\mathscr{A}}(P)\log\mu_x^{\mathscr{A}}(P)\,\mathrm{d}\mu(x) &= \int\left(\int \mathbb{1}_P(y)\,\mathrm{d}\mu_x^{\mathscr{A}}(y)\right)\log\mu_x^{\mathscr{A}}(P)\,\mathrm{d}\mu(x)\\ &= \int\int \mathbb{1}_P(y)\log\mu_y^{\mathscr{A}}(P)\,\mathrm{d}\mu_x^{\mathscr{A}}(y)\,\mathrm{d}\mu(x)\\ &= \int \mathbb{1}_P(x)\log\mu_x^{\mathscr{A}}(P)\,\mathrm{d}\mu(x)\end{aligned}$$

for any $P\in\xi$. Therefore,

$$\begin{aligned}\int H_{\mu_x^{\mathscr{A}}}(\xi)\,\mathrm{d}\mu &= -\int\sum_{P\in\xi}\mu_x^{\mathscr{A}}(P)\log\mu_x^{\mathscr{A}}(P)\,\mathrm{d}\mu(x)\\ &= -\int\sum_{P\in\xi}\mathbb{1}_P(x)\log\mu_x^{\mathscr{A}}(P)\,\mathrm{d}\mu(x)\\ &= \int I_\mu\left(\xi\middle|\mathscr{A}\right)\,\mathrm{d}\mu = H_\mu\left(\xi\middle|\mathscr{A}\right).\end{aligned}$$

Finiteness of $H_{\mu_x^{\mathscr{A}}}(\mathscr{C})$ for a general σ-algebra $\mathscr{C}$ was discussed on page 54. □

Working with finite partitions, the above allows us to prove continuity of entropy with respect to the given σ-algebra.

Proposition 2.12 (Continuity). *Let ξ be a finite or countable partition of a Borel probability space $(X,\mathscr{B},\mu)$, and let $\mathscr{A}_n\nearrow\mathscr{A}_\infty$ be an increasing (or $\mathscr{A}_n\searrow\mathscr{A}_\infty$ a decreasing) sequence of sub-σ-algebras of $\mathscr{B}$. Then*

$$I_\mu\left(\xi\middle|\mathscr{A}_n\right)\longrightarrow I_\mu\left(\xi\middle|\mathscr{A}_\infty\right)$$

and if ξ is a finite partition then also

$$H_\mu\left(\xi\middle|\mathscr{A}_n\right)\longrightarrow H_\mu\left(\xi\middle|\mathscr{A}_\infty\right).$$

Proof. Let $\xi=\{P_1,P_2,\dots\}$ be a finite or countable partition. Then the increasing martingale theorem [52, Th. 5.5] (resp. the decreasing martingale theorem [52, Th. 5.8]) implies for μ-almost every $x\in P_n$ that

$$\begin{aligned}I_\mu\left(\xi\middle|\mathscr{A}_n\right)(x) &= -\log\mu_x^{\mathscr{A}_n}(P_n)\\ &= -\log E_\mu\left(\mathbb{1}_{P_n}\middle|\mathscr{A}_n\right)(x)\\ &\longrightarrow -\log E_\mu\left(\mathbb{1}_{P_n}\middle|\mathscr{A}_\infty\right)(x) \qquad (\text{as } n\longrightarrow\infty)\\ &= I_\mu\left(\xi\middle|\mathscr{A}_\infty\right)(x).\end{aligned}$$

Assume now that ξ is a finite partition. By Lemma 2.11 we also have

$$H_\mu\left(\xi\middle|\mathscr{A}_n\right)=\int H_{\mu_x^{\mathscr{A}_n}}(\xi)\,\mathrm{d}\mu(x),$$

and, by the martingale theorems,

$$\begin{aligned} H_{\mu_x^{\mathscr{A}_n}}(\xi) &= -\sum_{P\in\xi} \mu_x^{\mathscr{A}_n}(P)\log \mu_x^{\mathscr{A}_n}(P) \\ &\longrightarrow -\sum_{P\in\xi} \mu_x^{\mathscr{A}_\infty}(P)\log \mu_x^{\mathscr{A}_\infty}(P) = H_{\mu_x^{\mathscr{A}_\infty}}(\xi) \end{aligned}$$

as $n\to\infty$, since the function ϕ in (1.6) is continuous for $x\geqslant 0$. Since

$$H_{\mu_x^{\mathscr{A}_n}}(\xi)\leqslant \log|\xi|<\infty$$

for μ-almost every x, dominated convergence gives

$$\lim_{n\to\infty} H_\mu\left(\xi\middle|\mathscr{A}_n\right) = \int H_{\mu_x^{\mathscr{A}_\infty}}(\xi)\,\mathrm{d}\mu = H_\mu\left(\xi\middle|\mathscr{A}_\infty\right),$$

which concludes the proof. □

With the above we can generalise the addition formula and monotonicity properties for entropy to σ-algebras.

Proposition 2.13 (Additivity and Monotonicity). *Let $(X,\mathscr{B},\mu)$ be a Borel probability space, and let $\mathscr{A},\mathscr{C}_1$, and $\mathscr{C}_2$ be countably-generated sub-σ-algebras of $\mathscr{B}$. Then we have*

(1) *the addition formulas*

$$\begin{aligned} I_\mu\left(\mathscr{C}_1\vee\mathscr{C}_2\middle|\mathscr{A}\right) &= I_\mu\left(\mathscr{C}_1\middle|\mathscr{A}\right)+I_\mu\left(\mathscr{C}_2\middle|\mathscr{C}_1\vee\mathscr{A}\right), \\ H_\mu\left(\mathscr{C}_1\vee\mathscr{C}_2\middle|\mathscr{A}\right) &= H_\mu\left(\mathscr{C}_1\middle|\mathscr{A}\right)+H_\mu\left(\mathscr{C}_2\middle|\mathscr{C}_1\vee\mathscr{A}\right) \end{aligned}$$

for the information function and the entropy,

(2) *the monotonicity property $H_\mu\left(\mathscr{C}_2\middle|\mathscr{C}_1\vee\mathscr{A}\right)\leqslant H_\mu\left(\mathscr{C}_2\middle|\mathscr{A}\right)$ of entropy,*

(3) *and so also $H_\mu\left(\mathscr{C}_1\vee\mathscr{C}_2\middle|\mathscr{A}\right)\leqslant H_\mu\left(\mathscr{C}_1\middle|\mathscr{A}\right)+H_\mu\left(\mathscr{C}_2\middle|\mathscr{A}\right)$.*

PROOF. Pick sequences of finite partitions (ξ_ℓ), (η_m), and (ζ_n) with

$$\xi_\ell\nearrow\mathscr{C}_1,\ \eta_m\nearrow\mathscr{C}_2,\ \text{and }\zeta_n\nearrow\mathscr{A}.$$

For (1), we first apply continuity with respect to the given σ-algebra (Proposition 2.12) and take $n\to\infty$ to deduce from

$$\begin{aligned} I_\mu\left(\xi_\ell\vee\eta_m\middle|\zeta_n\right) &= -\log\frac{\mu([x]_{\xi_\ell\vee\eta_m\vee\zeta_n})}{\mu([x]_{\zeta_n})} \\ &= -\log\frac{\mu([x]_{\xi_\ell\vee\zeta_n})}{\mu([x]_{\zeta_n})}-\log\frac{\mu([x]_{\xi_\ell\vee\eta_m\vee\zeta_n})}{\mu([x]_{\xi_\ell\vee\zeta_n})} \\ &= I_\mu\left(\xi_\ell\middle|\zeta_n\right)+I_\mu\left(\eta_m\middle|\xi_\ell\vee\zeta_n\right) \end{aligned}$$

that

$$I_\mu(\xi_\ell\vee\eta_m\big|\mathscr{A}) = I_\mu(\xi_\ell\big|\mathscr{A})+I_\mu(\eta_m\big|\xi_\ell\vee\mathscr{A}).$$

Next we use continuity with respect to both σ-algebras (Proposition 2.10 and 2.12) to take $\ell \to \infty$ and see that

$$I_\mu(\mathscr{C}_1 \vee \eta_m \big| \mathscr{A}) = I_\mu(\mathscr{C}_1 \big| \mathscr{A}) + I_\mu(\eta_m \big| \mathscr{C}_1 \vee \mathscr{A}).$$

Finally we use continuity with respect to the first σ-algebra (Proposition 2.10) and take $m \to \infty$ to see that

$$I_\mu(\mathscr{C}_1 \vee \mathscr{C}_2 \big| \mathscr{A}) = I_\mu(\mathscr{C}_1 \big| \mathscr{A}) + I_\mu(\mathscr{C}_1 \big| \mathscr{C}_1 \vee \mathscr{A}),$$

which proves both statements in (1) after taking the integral.

For (2) we use again the sequence of partitions as above and recall that

$$H_\mu\left(\eta_m \big| \xi_\ell \vee \zeta_n\right) \leqslant H_\mu\left(\eta_m \big| \zeta_n\right)$$

Using continuity of entropy (Proposition 2.12) we let $\ell \to \infty$ and $n \to \infty$ and get

$$H_\mu\left(\eta_m \big| \mathscr{C}_1 \vee \mathscr{A}\right) \leqslant H_\mu\left(\eta_m \big| \mathscr{A}\right)$$

Now we let $m \to \infty$ and use continuity of entropy (Proposition 2.10), which proves (2).

Combining (1) and (2) we also obtain (3). □

Even though we already discussed the continuity properties of the entropy function with respect to the given σ-algebra, we will now give a second proof for the increasing case which also gives dominated convergence for the information function.

Proposition 2.14 (Continuity and dominated convergence). *Let ξ be a countable partition of the Borel probability space $(X, \mathscr{B}, \mu)$ with $H_\mu(\xi) < \infty$. Let $\mathscr{A}_n \nearrow \mathscr{A}_\infty$ be an increasing sequence of σ-algebras. Then*

$$\int \sup_{n \geqslant 1} I_\mu\left(\xi \big| \mathscr{A}_n\right) \, \mathrm{d}\mu < \infty, \tag{2.3}$$

$$I_\mu\left(\xi \big| \mathscr{A}_n\right) \to I_\mu\left(\xi \big| \mathscr{A}_\infty\right) \tag{2.4}$$

almost everywhere and in L^1_μ, and

$$H_\mu\left(\xi \big| \mathscr{A}_n\right) \searrow H_\mu\left(\xi \big| \mathscr{A}_\infty\right) \tag{2.5}$$

as $n \to \infty$.

Proof. We note that (2.4) we already have shown in Proposition 2.12 and that the monotonicity claim in (2.5) follows from Proposition 2.13. Thus, by the dominated convergence theorem, it remains to show (2.3). For this we let

$$I^* = \sup_{n \geqslant 1} I_\mu\left(\xi \big| \mathscr{A}_n\right),$$

so in particular $I^* \geqslant 0$. Write $\mathbb{R}^+$ for the non-negative reals and m for Lebesgue measure restricted to $\mathbb{R}^+$. Let

$$F(t) = \mu\left(\{x \in X \mid I^*(x) > t\}\right);$$

then

$$\int I^* \,\mathrm{d}\mu = \int_{X\times\mathbb{R}^+} \mathbb{1}_{\{(x,t)\mid I^*(x)>t\}} \,\mathrm{d}(\mu\times m) = \int_0^\infty F(t)\,\mathrm{d}t$$

by applying Fubini's theorem twice. Now

$$\begin{aligned} F(t) &= \mu\Big(\bigsqcup_{P\in\xi} \big\{\, x \in P \mid \sup_{n\geqslant 1}\big(-\log \mu_x^{\mathscr{A}_n}(P)\,\big) > t \,\big\}\Big) \\ &= \sum_{P\in\xi} \mu\left(\{x \in P \mid \inf_{n\geqslant 1} \mu_x^{\mathscr{A}_n}(P) < \mathrm{e}^{-t}\}\right) \\ &= \sum_{P\in\xi}\sum_{n=1}^\infty \mu\big(\underbrace{\{x\in P \mid \mu_x^{\mathscr{A}_n}(P) < \mathrm{e}^{-t} \text{ but } \mu_x^{\mathscr{A}_m}(P) \geqslant \mathrm{e}^{-t} \text{ for } m<n\}}_{=P\cap A_n(P)}\big), \end{aligned}$$

where we define

$$A_n(P) = \{x \mid \mu_x^{\mathscr{A}_n}(P) < \mathrm{e}^{-t} \text{ but } \mu_x^{\mathscr{A}_m}(P) \geqslant \mathrm{e}^{-t} \text{ for } m<n\} \in \mathscr{A}_n,$$

for $P \in \xi$ and a fixed $t \geqslant 0$, and also used that the sets $A_1(P), A_2(P), \dots$ are all disjoint. Using the definition of $A_n(P)$ we obtain

$$\mu(P\cap A_n(P)) = \int_{A_n(P)} \mathbb{1}_P \,\mathrm{d}\mu = \int_{A_n(P)} E\big(\mathbb{1}_P \big| \mathscr{A}_n\big)\,\mathrm{d}\mu < \mathrm{e}^{-t}\mu(A_n).$$

Then (each $A_n(P)$ depends also on t)

$$F(t) = \sum_{P\in\xi}\sum_{n=1}^\infty \mu\big(P\cap A_n(P)\big) \leqslant \sum_{P\in\xi} \min\{\mu(P), \mathrm{e}^{-t}\}, \tag{2.6}$$

which by integrating yields

$$\begin{aligned} \int I^*\,\mathrm{d}\mu = \int_0^\infty F(t)\,\mathrm{d}t &\leqslant \sum_{P\in\xi}\int_0^\infty \min\{\mu(P),\mathrm{e}^{-t}\}\,\mathrm{d}t \\ &= \sum_{P\in\xi} -\mu(P)\log\mu(P) + \mu(P) \\ &= H_\mu(\xi) + 1. \end{aligned}$$

□

2.2.3 Extremal values of conditional entropy

Both zero conditional entropy and maximal conditional entropy have important characterizations which we will discuss here.

We start by giving a natural interpretation of zero conditional entropy. For a σ-algebra $\mathscr{A} \subseteq \mathscr{B}$ and a set $C \in \mathscr{B}$ we write

$$C \underset{\mu}{\in} \mathscr{A}$$

if there exists a set $A \in \mathscr{A}$ such that $\mu(C \triangle A) = 0$. For σ-algebras $\mathscr{A}, \mathscr{C} \subseteq \mathscr{B}$ we write

$$\mathscr{C} \underset{\mu}{\subseteq} \mathscr{A}$$

if for any $C \in \mathscr{C}$ we have $C \underset{\mu}{\in} \mathscr{A}$. We note that if $\mathscr{C}$ is countably generated the latter implies that there exists one null set $N \in \mathscr{B}$ such that for all $C \in \mathscr{C}$ there exists a set $A \in \mathscr{A}$ with $C \smallsetminus N = A \smallsetminus N$.

Proposition 2.15 (Zero Entropy). *Let $(X, \mathscr{B}, \mu)$ be a Borel probability space, with $\mathscr{C}$ and $\mathscr{A}$ a pair of countably-generated sub-σ-algebras of $\mathscr{B}$. Then*

$$H_\mu\left(\mathscr{C} \middle| \mathscr{A}\right) = 0$$

if and only if

$$\mathscr{C} \underset{\mu}{\subseteq} \mathscr{A}.$$

Proof. First notice that $H_\mu\left(\mathscr{C} \middle| \mathscr{A}\right) = 0$ if and only if $\mu_x^{\mathscr{A}}\left([x]_{\mathscr{C}}\right) = 1$ for almost every $x \in X$.

Assume now that

$$\mathscr{C} \underset{\mu}{\subseteq} \mathscr{A}$$

and that $\mathscr{C}$ is generated by the countable algebra $\{C_1, C_2, \dots\}$. Then there exists a μ-null set N such that for every C_i there exists an $A_i \in \mathscr{A}$ for which $A_i \smallsetminus N = C_i \smallsetminus N$. This gives $[x]_{\mathscr{A}} \smallsetminus N \subseteq [x]_{\mathscr{C}}$ for all $x \in X$. Hence

$$\mu_x^{\mathscr{A}}\left([x]_{\mathscr{C}}\right) \geqslant \mu_x^{\mathscr{A}}\left([x]_{\mathscr{A}} \smallsetminus N\right) = 1$$

for μ-almost every $x \in X$.

Assume now that $\mu_x^{\mathscr{A}}\left([x]_{\mathscr{C}}\right) = 1$ almost everywhere. Then for $C \in \mathscr{C}$ we deduce that

$$\mu_x^{\mathscr{A}}(C) = \mathbb{1}_C(x),$$

first for almost every $x \in C$ and by using $X \smallsetminus C$ then for almost every $x \notin C$. It follows that

$$E\left(\mathbb{1}_C \middle| \mathscr{A}\right) = \mathbb{1}_C.$$

This shows that $C \underset{\mu}{\in} \mathscr{A}$, and we conclude that $\mathscr{C} \underset{\mu}{\subseteq} \mathscr{A}$. □

By analogy with the case of two partitions (see Exercise 1.1.3), we say that two σ-algebras $\mathscr{A}, \mathscr{C} \subseteq \mathscr{B}$ are *independent*, denoted† by $\mathscr{A} \perp \mathscr{C}$, if

$$\mu(A \cap C) = \mu(A)\mu(C) \tag{2.7}$$

for all $A \in \mathscr{A}$ and all $C \in \mathscr{C}$. We say that a partition ξ is independent to $\mathscr{C}$ if $\sigma(\xi) \perp \mathscr{C}$ and will again write $\xi \perp \mathscr{C}$. We also say that a measurable set A is independent to $\mathscr{C}$ if (2.7) holds for A and all $C \in \mathscr{C}$.

By monotonicity of entropy (Proposition 2.13(2)) the following gives a characterization of maximal conditional entropy.

Proposition 2.16 (Maximal entropy). *Let $\mathscr{C}$ be a countably-generated σ-algebra, ξ a countable partition with finite entropy, and $P \in \mathscr{B}$ be any measurable set in a Borel probability space $(X, \mathscr{B}, \mu)$. Then ξ is independent to $\mathscr{C}$ if and only if*

$$H_\mu(\xi \big| \mathscr{C}) = H_\mu(\xi).$$

Also a measurable set P is independent to $\mathscr{C}$ if and only if

$$\mu_x^{\mathscr{C}}(P) = \mu(P) \tag{2.8}$$

for almost every $x \in X$.

Proof. We consider first the case of a measurable set P and suppose that (2.8) holds, and let $C \in \mathscr{C}$. Then

$$\mu(P \cap C) = \int_C \mathbb{1}_P \,\mathrm{d}\mu = \int_C \underbrace{E_\mu\left(\mathbb{1}_P \big| \mathscr{C}\right)}_{=\mu_x^{\mathscr{C}}(P)} \mathrm{d}\mu = \mu(P)\mu(C).$$

Now assume (2.7) holds for $P = A$ and all $C \in \mathscr{C}$. Then the constant function $f(x) = \mu(P)$ satisfies

$$\int_C f \,\mathrm{d}\mu = \mu(P)\mu(C) = \mu(P \cap C) = \int_C \mathbb{1}_P \,\mathrm{d}\mu$$

for all $C \in \mathscr{C}$ and is clearly $\mathscr{C}$-measurable. It follows that

$$f(x) = E\left(\mathbb{1}_P \big| \mathscr{C}\right) = \mu_x^{\mathscr{C}}(P)$$

almost everywhere with respect to μ.

For the first statement of the proposition we let ξ be a countable partition with finite entropy. Then, using the function ϕ from (1.6) and Lemma 2.11 we obtain

† Strictly speaking this should be written $\perp_\mu$ to reflect the dependence of this notion on the measure, but where we use independence of partitions or of σ-algebras the measure will always be clear from the context.

$$\begin{aligned} H_\mu\left(\xi \middle| \mathscr{C}\right) &= \int H_{\mu_x^{\mathscr{C}}}(\xi)\,\mathrm{d}\mu \\ &= -\sum_{P\in\xi}\int \phi\left(\mu_x^{\mathscr{C}}(P)\right)\,\mathrm{d}\mu \\ &\leqslant -\sum_{P\in\xi}\phi\Bigg(\underbrace{\int \mu_x^{\mathscr{C}}(P)\,\mathrm{d}\mu}_{=\mu(P)}\Bigg) \qquad \text{(by Lemmas 1.3 and 1.4)} \\ &= H_\mu(\xi). \end{aligned}$$

By strict convexity of ϕ equality occurs here if and only if

$$\mu(P) = \mu_x^{\mathscr{C}}(P)$$

for almost every x and all $P \in \xi$, and this holds if and only if

$$\mu(P \cap C) = \mu(P)\mu(C)$$

for all $P \in \xi, C \in \mathscr{C}$ by the first part of the proposition. □

Exercises for Section 2.2

Exercise 2.2.1. Show that Proposition 2.12 does not hold for an arbitrary σ-algebra $\mathscr{C}$ and sequence of σ-algebras $\mathscr{A}_n \nearrow \mathscr{A}_\infty$, by finding an example for which $H_\mu\left(\mathscr{C}\middle|\mathscr{A}_n\right)$ does not converge to $H_\mu\left(\mathscr{C}\middle|\mathscr{A}_\infty\right)$ as $n \to \infty$.

Exercise 2.2.2. Analyse where our proof of (2.3) fails in the case of a decreasing sequence of σ-algebras.

Exercise 2.2.3. Let ξ be a countable partition of finite entropy in a Borel probability space $(X, \mathscr{B}, \mu)$. Let $\mathscr{A}_n \searrow \mathscr{A}_\infty$ be a decreasing sequence of sub σ-algebras of $\mathscr{B}$. Show that $H_\mu(\xi\middle|\mathscr{A}_n) \nearrow H_\mu(\xi\middle|\mathscr{A}_\infty)$ as $n \to \infty$.

2.3 Conditional Entropy of a Transformation

For brevity we introduce the notation ξ_a^b for $\bigvee_{i=a}^{b} T^{-i}\xi$ whenever $a < b$ belong to $\mathbb{N}_0 \cup \{\infty\}$ (or $\mathbb{Z} \cup \{-\infty, \infty\}$ in the invertible case) and the map is clear from the context. We will use this notation also for σ-algebras.

Recall from Lemma 1.12 that for partitions ξ and η we have the invariance property

$$H_\mu\left(\xi\middle|\eta\right) = H_\mu\left(T^{-1}\xi\middle|T^{-1}\eta\right).$$

This extend to general σ-algebras as follows.

Lemma 2.17 (Invariance). *Let $(X,\mathscr{B},\mu,T)$ be a measure-preserving system on a Borel probability space and let $\mathscr{A},\mathscr{C}$ be sub σ-algebras of $\mathscr{B}$. Then*

$$H_\mu\left(\mathscr{C}\middle|\mathscr{A}\right)=H_\mu\left(T^{-1}\mathscr{C}\middle|T^{-1}\mathscr{A}\right)$$

and

$$I_\mu\left(\mathscr{C}\middle|\mathscr{A}\right)\circ T=I_\mu\left(T^{-1}\mathscr{C}\middle|T^{-1}\mathscr{A}\right). \tag{2.9}$$

PROOF. It is enough to show (2.9), and for this we have

$$\begin{aligned}
I_\mu\left(\mathscr{C}\middle|\mathscr{A}\right)(Tx) &= -\log\mu_{Tx}^{\mathscr{A}}\left([Tx]_{\mathscr{C}}\right)\\
&= -\log\left(T_*\mu_x^{T^{-1}\mathscr{A}}\right)\left([Tx]_{\mathscr{C}}\right) && \text{(by Lemma 2.5)}\\
&= -\log\mu_x^{T^{-1}\mathscr{A}}\left(T^{-1}[Tx]_{\mathscr{C}}\right)\\
&= -\log\mu_x^{T^{-1}\mathscr{A}}\left([x]_{T^{-1}\mathscr{C}}\right)=I_\mu\left(T^{-1}\mathscr{C}\middle|T^{-1}\mathscr{A}\right)(x).
\end{aligned}$$

□

For later developments, it will be useful to discuss the entropy of T with respect to a given sub-σ-algebra $\mathscr{A}$, so that when measuring the entropies of the repeated experiment ξ we will always assume that the information of $\mathscr{A}$ is given. In this context we will always assume that $\mathscr{A}$ is *strictly invariant* in the sense that $T^{-1}\mathscr{A}=\mathscr{A}$. The first step is to show that the sequence of (conditional) entropies of the repeated experiments is sub-additive. Given $m,n\geqslant 1$, we may use subadditivity of entropy (Proposition 2.13(3)), $T^{-n}\mathscr{A}=\mathscr{A}$, and invariance of entropy (Lemma 2.17) to obtain

$$\begin{aligned}
H_\mu\left(\xi_0^{m+n-1}\middle|\mathscr{A}\right) &\leqslant H_\mu\left(\xi_0^{n-1}\middle|\mathscr{A}\right)+H_\mu\left(T^{-n}\xi_n^{m+n-1}\middle|T^{-n}\mathscr{A}\right)\\
&= H_\mu\left(\xi_0^{n-1}\middle|\mathscr{A}\right)+H_\mu\left(\xi_0^{m-1}\middle|\mathscr{A}\right).
\end{aligned}$$

So the sequence $\left(H_\mu\left(\xi_0^{n-1}\middle|\mathscr{A}\right)\right)_{n\geqslant 1}$ is sub-additive in the sense of Lemma 1.13, justifying the claimed convergence in Definition 2.18.

Definition 2.18. Let $(X,\mathscr{B},\mu,T)$ be a measure-preserving system on a Borel probability space. If $\mathscr{A}$ is a sub-σ-algebra of $\mathscr{B}$ with $T^{-1}(\mathscr{A})=\mathscr{A}$ (that is, a strictly invariant sub σ-algebra) then the *conditional entropy of T given $\mathscr{A}$* is

$$h_\mu(T\big|\mathscr{A})=\sup_{\xi:H_\mu(\xi)<\infty}h_\mu(T,\xi\big|\mathscr{A})$$

where

$$h_\mu(T,\xi\big|\mathscr{A})=\lim_{n\to\infty}\frac{1}{n}H_\mu\left(\xi_0^{n-1}\middle|\mathscr{A}\right)=\inf_{n\geqslant 1}\frac{1}{n}H_\mu\left(\xi_0^{n-1}\middle|\mathscr{A}\right)$$

for any countable partition ξ.

We note that clearly $h_\mu(T,\xi\big|\mathscr{A}) \leqslant h_\mu(T,\xi)$ and $h_\mu(T\big|\mathscr{A}) \leqslant h_\mu(T)$ for any countable partition ξ with finite entropy and any strictly invariant σ-algebra $\mathscr{A}$. Just as in Chapter 1, we need to develop the basic properties of conditional entropy.

Proposition 2.19 (Basic properties). *Let $(X,\mathscr{B},\mu,T)$ be a measure-preserving system on a Borel probability space, let ξ and η be countable partitions of X with finite entropy, and let $\mathscr{A} = T^{-1}\mathscr{A} \subseteq \mathscr{B}$ be a strictly invariant sub σ-algebra. Then properties* (1)–(3) *of Proposition* 1.17 *and* (1)–(4) *of Proposition* 1.18 *also hold for the entropy conditioned on $\mathscr{A}$. Moreover, we have the following formulas.*

(1) **(Future formula)** *The dynamical entropy can be expressed as the entropy of the partition given the future of the partition, i.e.*

$$h_\mu(T,\xi\big|\mathscr{A}) = \lim_{n\to\infty} H_\mu\left(\xi\big|\xi_1^n \vee \mathscr{A}\right) = H_\mu\left(\xi\big|\xi_1^\infty \vee \mathscr{A}\right).$$

(2) **(Additivity)** *If T is invertible, the dynamical entropy is additive in the sense that*

$$\begin{aligned} h_\mu\left(T,\xi\vee\eta\big|\mathscr{A}\right) &= h_\mu(T,\xi\big|\mathscr{A}) + h_\mu\left(T,\eta\big|\xi_{-\infty}^\infty \vee \mathscr{A}\right) \\ &= h_\mu(T,\xi\big|\mathscr{A}) + H_\mu\left(\eta\big|\eta_1^\infty \vee \xi_{-\infty}^\infty \vee \mathscr{A}\right). \end{aligned}$$

Proof. The proofs of the generalization of Proposition 1.17–1.18 to conditional entropies follow the same lines as these proofs, we will not repeat them here.

(1): For any $n \geqslant 1$,

$$\begin{aligned} \frac{1}{n}H_\mu(\xi_0^{n-1}\big|\mathscr{A}) &= \frac{1}{n}\big(H_\mu(\xi\big|\xi_1^{n-1}\vee\mathscr{A}) + H_\mu(T^{-1}\xi_0^{n-2}\big|T^{-1}\mathscr{A})\big) \\ &= \frac{1}{n}\big(H_\mu(\xi\big|\xi_1^{n-1}\vee\mathscr{A}) + H_\mu(T^{-1}\xi\big|\xi_2^{n-1}\vee T^{-1}\mathscr{A}) \\ &\qquad + \cdots + H_\mu(T^{-(n-1)}\xi\big|T^{-(n-1)}\mathscr{A})\big) \\ &= \frac{1}{n}\sum_{j=0}^{n-1} H_\mu(\xi\big|\xi_1^j\vee\mathscr{A}) \underset{\text{Cesàro}}{\longrightarrow} H_\mu(\xi\big|\xi_1^\infty\vee\mathscr{A}). \end{aligned}$$

In the last line we first used invariance of entropy (Lemma 2.17), then continuity of entropy (Proposition 2.14), and finally the fact that the Cesàro averages of a convergent sequence (a_m) converge to the limit $\lim_{m\to\infty} a_m$.

(2): By splitting the entropy up in a similar way to the argument in (1), but in a different order, we get

$$\begin{aligned} h_\mu(T,\xi\vee\eta\big|\mathscr{A}) &= \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\xi_0^{n-1}\big|\mathscr{A}\right)+\lim_{n\to\infty}\frac{1}{n}H_\mu\left(\eta_0^{n-1}\big|\xi_0^{n-1}\vee\mathscr{A}\right)\\ &= h_\mu(T,\xi\big|\mathscr{A})+\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}H_\mu\left(T^{-j}\eta\big|\eta_{j+1}^{n-1}\vee\xi_0^{n-1}\vee\mathscr{A}\right)\\ &= h_\mu(T,\xi\big|\mathscr{A})+\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1}H_\mu\left(\eta\big|\eta_1^{n-1-j}\vee\xi_{-j}^{n-1-j}\vee\mathscr{A}\right)\\ &\geqslant h_\mu(T,\xi\big|\mathscr{A})+H_\mu\left(\eta\big|\eta_1^{\infty}\vee\xi_{-\infty}^{\infty}\vee\mathscr{A}\right) \end{aligned}$$

by monotonicity of entropy with respect to the given σ-algebra (Proposition 2.13).

On the other hand, given $\varepsilon>0$ there exists N such that

$$H_\mu\left(\eta\big|\eta_1^{N}\vee\xi_{-N}^{N}\vee\mathscr{A}\right)<H_\mu\left(\eta\big|\eta_1^{\infty}\vee\xi_{-\infty}^{\infty}\vee\mathscr{A}\right)+\varepsilon,$$

by continuity of entropy (Proposition 2.14). Therefore, we can again use monotonicity for any n and $j\in[N,n-N-1]$ to get

$$H_\mu\left(\eta\big|\eta_1^{n-1-j}\vee\xi_{-j}^{n-1-j}\vee\mathscr{A}\right)\leqslant H_\mu\left(\eta\big|\eta_1^{\infty}\vee\xi_{-\infty}^{\infty}\vee\mathscr{A}\right)+\varepsilon.$$

Since for large enough n the contribution of the other terms, which is at most $2NH_\mu(\eta)$, will be smaller than $n\varepsilon$, the reverse inequality follows by taking $n\to\infty$. □

Recall that the Kolmogorov–Sinaĭ Theorem (Theorem 1.21) gives a method to compute the entropy of a measure-preserving system given a generator. The next result is a useful extension to this, which is also applicable in situations where a generator is difficult to find.

Theorem 2.20 (Kolmogorov–Sinaĭ for sequences of partitions). *Suppose that $(X,\mathscr{B},\mu,T)$ is a measure-preserving system on a Borel probability space. If (ξ_k) is an increasing sequence of partitions (that is, if $\xi_k\subseteq\sigma(\xi_{k+1})$ for all $k\geqslant 1$) of finite entropy with the property that*

- $\mathscr{B}\underset{\mu}{=}\bigvee_{k=1}^{\infty}(\xi_k)_0^{\infty}$ *or*
- $\mathscr{B}\underset{\mu}{=}\bigvee_{k=1}^{\infty}(\xi_k)_{-\infty}^{\infty}$ *if T is invertible,*

then

$$h_\mu(T)=\sup_k h_\mu(T,\xi_k)=\lim_{k\to\infty}h_\mu(T,\xi_k).$$

More generally, under the same hypothesis, if $\mathscr{A}=T^{-1}\mathscr{A}\subseteq\mathscr{B}$ is a strictly invariant sub-σ-algebra, then

$$h_\mu(T\big|\mathscr{A})=\sup_k h_\mu(T,\xi_k\big|\mathscr{A})=\lim_{k\to\infty}h_\mu(T,\xi_k\big|\mathscr{A}).$$

We may abbreviate the assumptions to the theorem by saying that the *sequence of partitions* (ξ_k) *is generating* (with respect to T).

PROOF OF THEOREM 2.20. Let $\mathscr{A}$ and (ξ_k) be as in the theorem, and let ξ be another countable partition of finite entropy. Then

$$h_\mu(T,\xi\big|\mathscr{A}) \leqslant h_\mu(T,(\xi_n)_0^n\big|\mathscr{A}) + H_\mu\left(\xi\big|(\xi_n)_0^n \vee \mathscr{A}\right)$$

for any $n \geqslant 1$ by the continuity bound in Proposition 1.17(3) conditioned on $\mathscr{A}$. Here the first term

$$h_\mu(T,(\xi_n)_0^n\big|\mathscr{A}) = h_\mu(T,\xi_n\big|\mathscr{A})$$

by Proposition 1.18(1), and the second term

$$H_\mu\left(\xi\big|(\xi_n)_0^n \vee \mathscr{A}\right) \longrightarrow 0$$

as $n \to \infty$ by continuity of entropy (Proposition 2.14) and the assumption about the sequence (ξ_k). □

We end this section by showing how the entropy of a measure-preserving transformation splits into two terms with respect to a given factor.(13)

Corollary 2.21 (Abramov–Rokhlin formula). *Let* $\mathsf{X} = (X,\mathscr{B}_X,\mu,T)$ *be an invertible measure-preserving system on a Borel probability space, and let* $\phi : \mathsf{X} \to (Y,\mathscr{B}_Y,\nu,S)$ *be a factor map. Then*

$$h_\mu(T) = h_\nu(S) + h_\mu(T\big|\mathscr{A}) \tag{2.10}$$

where the factor Y *is identified with the corresponding strictly invariant sub-*σ*-algebra* $\mathscr{A} = \phi^{-1}\mathscr{B}_Y \subseteq \mathscr{B}_X$. *Moreover,*

$$h_\nu(S) = \sup\big\{h_\mu(T,\eta) \mid \eta \subseteq \mathscr{A} \text{ is a countable partition of finite entropy}\big\}.$$

PROOF. For the final statement, let $\eta \subseteq \mathscr{B}_Y$ be a countable partition. Then $H_\nu(\eta) = H_\mu(\phi^{-1}\eta)$ so that, in particular, η has finite entropy if and only if $\phi^{-1}\eta$ has finite entropy. Using this also for $\bigvee_{i=0}^{n-1} S^{-i}\eta$ instead of η, we see that

$$h_\nu(S,\eta) = h_\mu(T,\phi^{-1}\eta),$$

which implies the last statement.

Turning to the proof of (2.10), pick sequences of finite partitions (η_m) and (ξ_n) with $\eta_m \nearrow \mathscr{B}_Y$ and $\xi_n \nearrow \mathscr{B}_X$. By the additivity of dynamical entropy (Proposition 2.19(2)),

$$\begin{aligned} h_\mu(T,\xi_n \vee \phi^{-1}\eta_m) &= h_\mu(T,\phi^{-1}\eta_m) + H_\mu\left(\xi_n\big|(\xi_n)_1^\infty \vee (\phi^{-1}\eta_m)_{-\infty}^\infty\right) \\ &= h_\nu(S,\eta_m) + H_\mu\left(\xi_n\big|(\xi_n)_1^\infty \vee (\phi^{-1}\eta_m)_{-\infty}^\infty\right) \end{aligned} \tag{2.11}$$

Assume first that the left-hand side of (2.10) is finite. By the Kolmogorov–Sinaĭ theorem applied for T (Theorem 2.20), for any fixed $\varepsilon > 0$ and all large enough n we have

$$h_\mu(T) - \varepsilon \leqslant h_\mu(T, \xi_n) \leqslant h_\mu(T, \xi_n \vee \phi^{-1}\eta_m) \leqslant h_\mu(T).$$

Combining this with (2.11) gives

$$h_\mu(T) - \varepsilon \leqslant h_\mu(S, \eta_m) + H_\mu\left(\xi_n \big| (\xi_n)_1^\infty \vee (\phi^{-1}\eta_m)_{-\infty}^\infty\right) \leqslant h_\mu(T).$$

Let $m \to \infty$. Applying the Kolmogorov–Sinaĭ theorem to S (Theorem 2.20) and continuity of entropy (Proposition 2.14) we obtain

$$h_\mu(T) - \varepsilon \leqslant h_\nu(S) + H_\mu\left(\xi_n \big| (\xi_n)_1^\infty \vee \mathscr{A}\right) \leqslant h_\mu(T).$$

Moreover, by the future formula for entropy (Proposition 2.19(1)) and the Kolmogorov–Sinaĭ theorem applied for T conditioned on $\mathscr{A}$ (Theorem 2.20) we have

$$\lim_{n\to\infty} H_\mu\left(\xi_n \big| (\xi_n)_1^\infty \vee \mathscr{A}\right) = \lim_{n\to\infty} h_\mu(T, \xi_n \big| \mathscr{A}) = h_\mu(T \big| \mathscr{A}),$$

which proves (2.10).

The case of $h_\mu(T) = \infty$ is similar. Note that in this case the left-hand side of (2.11) will become arbitrarily large as $n \to \infty$. If $h_\nu(S) = \infty$, there is nothing to prove. So assume that the first term on the right-hand side of (2.11) stays bounded as $m \to \infty$. As the second term converges to $h_\mu(T, \xi_n \big| \mathscr{A})$ as $m \to \infty$, it follows from (2.11) that $h_\mu(T \big| \mathscr{A}) = \infty$. □

2.4 The Pinsker Algebra

Studying a measure-preserving system via its factors can have far-reaching consequences (a striking instance of this is Furstenberg's proof of Szemerédi's theorem; see [52, Ch. 7]). In particular, it is natural to ask if a measure-preserving system can always be decomposed into a zero-entropy system and a system with the property that every non-trivial factor has positive entropy. This turns out to be too much to ask, but a partial answer in the same direction is afforded by the *Pinsker algebra* [165], which we now wish to study.

Definition 2.22. Let $(X, \mathscr{B}, \mu, T)$ be an invertible measure-preserving system on a Borel probability space. The *Pinsker algebra* of T is

$$\begin{aligned}\mathscr{P}(T) &= \{B \in \mathscr{B} \mid h_\mu(T, \{B, X \smallsetminus B\}) = 0\} \\ &= \left\{ B \in \mathscr{B} \mid B \underset{\mu}{\in} \bigvee_{n=1}^{\infty} T^{-n}\{B, X \smallsetminus B\} \right\}. \end{aligned} \tag{2.12}$$

The formulation of $\mathscr{P}(T)$ in (2.12) follows from the future formula for entropy and the characterization of zero conditional entropy (Proposition 2.19(1) and Proposition 2.15) and may be described as follows: the Pinsker algebra comprises those sets with the property that knowledge of whether the orbit of a point lies in the set or not in all of the future determines whether it lies in the set in the present.

Proposition 2.23 (Pinsker factor). *The Pinsker algebra $\mathscr{P}(T)$ is a strictly invariant sub-σ-algebra and so defines via Theorem* 2.6 *the* Pinsker factor *of T. This factor*

$$\mathsf{X}_{\mathscr{P}} = (X_{\mathscr{P}}, \mathscr{P}(T), \mu, T_{\mathscr{P}})$$

has zero entropy and is maximal with respect to that property in the following sense. If

$$\mathsf{Y} = (Y, \mathscr{A}, \nu, S)$$

is another factor of X with zero entropy then Y is a factor of $\mathsf{X}_{\mathscr{P}}$. Moreover, the relative entropy of T given $\mathscr{P}(T)$ coincides with the entropy of T,

$$h_\mu(T) = h_\mu(T \big| \mathscr{P}(T)). \tag{2.13}$$

Proof. To see that $\mathscr{P}(T)$ is a σ-algebra, let $\{B_i \mid i \geqslant 1\}$ be a collection of sets in $\mathscr{P}(T)$ and write $\xi_i = \{B_i, X \smallsetminus B_i\}$ for the associated partitions. If $Q \in \bigvee_{i=1}^{\infty} \xi_i$ and $\eta = \{Q, X \smallsetminus Q\}$, then for any $\varepsilon > 0$ there is an n such that

$$H_\mu\Big(\eta \big| \bigvee_{i=1}^{n} \xi_i\Big) < \varepsilon.$$

It follows, by the continuity bound in Proposition 1.17(3), that

$$\begin{aligned} h_\mu(T, \eta) &\leqslant h_\mu\Big(T, \bigvee_{i=1}^{n} \xi_i\Big) + H_\mu\Big(\eta \big| \bigvee_{i=1}^{n} \xi_i\Big) \\ &\leqslant \sum_{i=1}^{n} h_\mu(T, \xi_i) + \varepsilon = \varepsilon, \end{aligned}$$

so $\eta \subseteq \mathscr{P}(T)$. For the T-invariance, note first that $h_\mu(T, \xi) = h_\mu(T, T^{-1}\xi)$ for any partition with finite entropy, which implies that $T^{-1}\mathscr{P}(T) \subseteq \mathscr{P}(T)$. The converse holds as well, since any $B \in \mathscr{P}(T)$ agrees by (2.12) modulo μ with some element of

$$\bigvee_{n=1}^{\infty} T^{-n}\{B, X\smallsetminus B\} = T^{-1}\bigvee_{n=0}^{\infty} T^{-n}\{B, X\smallsetminus B\} \subseteq T^{-1}\mathscr{P}(T).$$

By Theorem 2.6 this defines now a factor $\mathsf{X}_{\mathscr{P}} = (X_{\mathscr{P}}, \mathscr{P}(T), \mu, T_{\mathscr{P}})$ of the measure-preserving system $(X, \mathscr{B}, \mu, T)$. By the definition of $\mathscr{P}(T)$ and by the last statement in the Abramov–Rokhlin formula (Corollary 2.21), the entropy $h_\mu(T_{\mathscr{P}})$ of the Pinsker factor vanishes. Now (2.13) follows from Corollary 2.21 (or the following Lemma 2.24). If Y is any factor of X with zero entropy, then every finite partition of the corresponding T-invariant σ-algebra $\mathscr{A}$ must have zero entropy. Hence the whole σ-algebra $\mathscr{A}$ must be measurable with respect to $\mathscr{P}(T)$ modulo μ. By Theorem 2.6 (applied to $\mathscr{A} \subseteq \mathscr{P}(T)$), it follows that there is a factor map $\mathsf{X}_{\mathscr{P}} \to \mathsf{Y}$. □

The next lemma (which refines (2.13) and is needed in the next section) encapsulates once again the idea that the factor $\mathscr{P}(T)$ has no entropy, and is maximal with respect to this property.

Lemma 2.24 (Conditioning on the Pinsker algebra). *Let $(X, \mathscr{B}, \mu, T)$ be a measure-preserving system on a Borel probability space, and let $\mathscr{A} \subseteq \mathscr{B}$ be a strictly invariant sub-σ-algebra. Then for any partition ξ with finite entropy,*

$$h_\mu(T, \xi\big|\mathscr{A}) = h_\mu(T, \xi\big|\mathscr{P}(T) \vee \mathscr{A}).$$

Proof. Let $\eta \subseteq \mathscr{P}(T)$ be a partition with $H_\mu(\eta) < \infty$ (so $h_\mu(T, \eta) = 0$) and let ξ be a partition with $H_\mu(\xi) < \infty$. Then

$$\begin{aligned} h_\mu(T, \xi\big|\mathscr{A}) &\leqslant h_\mu(T, \xi \vee \eta\big|\mathscr{A}) \\ &= h_\mu(T, \eta\big|\mathscr{A}) + h_\mu\left(T, \xi\big|\eta_{-\infty}^{\infty} \vee \mathscr{A}\right) \leqslant h_\mu(T, \xi\big|\mathscr{A}) \end{aligned}$$

by additivity of dynamical entropy (Proposition 2.19(2)), and monotonicity with respect to the given σ-algebra. By the future formula (Proposition 2.19(1)) this gives

$$h_\mu(T, \xi\big|\mathscr{A}) = H_\mu\left(\xi\big|\xi_1^{\infty} \vee \eta_{-\infty}^{\infty} \vee \mathscr{A}\right).$$

By choosing an increasing sequence of such partitions (η_n), which generate the Pinsker algebra in the sense that $\eta_n \nearrow \mathscr{P}(T)$, the lemma follows from the continuity of entropy with respect to the given σ-algebra (see Proposition 2.14). □

A partition ξ is called non-trivial if it contains two sets of positive measure, or equivalently if $H(\xi) > 0$. The Pinsker algebra singles out a natural class of measure-preserving transformations — those for which all non-trivial factors have positive entropy.

Definition 2.25. An invertible measure-preserving system $(X, \mathscr{B}, \mu, T)$ is said to have *completely positive entropy* or to be a *K-automorphism* if we have $h_\mu(T, \xi) > 0$ for any non-trivial partition ξ.

Notice that $(X, \mathscr{B}, \mu, T)$ has completely positive entropy if and only if $\mathscr{P}(T) \underset{\mu}{=} \mathscr{N}_X = \{X, \varnothing\}$ is the trivial σ-algebra.

2.4.1 Tail σ-algebras

For any σ-algebra $\mathscr{A}$ (and similarly for partitions) we will refer to the σ-algebra

$$\bigcap_{n=0}^{\infty} \mathscr{A}_n^{\infty} = \bigcap_{n=0}^{\infty} \bigvee_{i=n}^{\infty} T^{-i}(\mathscr{A})$$

as the *tail σ-algebra* or *tail field* of $\mathscr{A}$. Below we will frequently use the characterization of zero conditional entropy in Proposition 2.15. As in that characterization containment modulo μ is used, it is convenient to interpret the intersection of σ-algebras as in the above definition also modulo μ, meaning that A belongs to the tail σ-algebra of $\mathscr{A}$ if A is an element of $\mathscr{A}_n^{\infty}$ modulo μ for all n. However, we note that for decreasing sequences of σ-algebras, as for instance in the definition of the tail, this is not necessary by the following argument.

Suppose $\mathscr{C}_n \searrow \mathscr{C}_\infty$ and suppose for some measurable B there exists for all n some $C_n \in \mathscr{C}_n$ with $\mu(B \triangle C_n) = 0$ (i.e. B equals C_n modulo μ). Then we note that $\bigcup_{k \geqslant n} C_k \in \mathscr{C}_n$ also equals B modulo μ. Defining the limes superior of these sets we see that

$$C_\infty = \limsup_{n \to \infty} C_n = \bigcap_n \bigcup_{k \geqslant n} C_k = \bigcap_{n \geqslant n_0} \bigcup_{k \geqslant n} C_k$$

for all $n_0 \geqslant 1$. Therefore, C_∞ belongs to $\mathscr{C}_\infty = \bigcap_n \mathscr{C}_n$ and also equals B modulo μ. Hence the intersection defined modulo μ is equivalent to the intersection of σ-algebras.

The tail of partitions with finite entropy is directly related to the Pinsker σ-algebra as we will now discuss.

Proposition 2.26 (Pinsker and tail σ-algebras). *For any invertible measure-preserving transformation T of a Borel probability space $(X, \mathscr{B}, \mu)$,*

$$\mathscr{P}(T) \underset{\mu}{=} \bigvee_{\xi : H_\mu(\xi) < \infty} \bigcap_{n=0}^{\infty} \xi_n^{\infty}.$$

Proof of Proposition 2.26. Let ξ be a partition with $H_\mu(\xi) < \infty$, and let η be a finite partition measurable with respect to $\bigcap_{n=0}^{\infty} \xi_n^{\infty}$. Then $\eta \subseteq \xi_1^{\infty}$ and so

$$\begin{aligned}h_\mu(T,\xi) &= H_\mu\left(\xi\middle|\xi_1^\infty\right) = H_\mu\left(\xi\vee\eta\middle|(\xi\vee\eta)_1^\infty\right)\\ &= h_\mu(T,\xi\vee\eta) = h_\mu(T,\eta) + h_\mu\left(T,\xi\middle|\eta_{-\infty}^\infty\right)\\ &= h_\mu(T,\eta) + h_\mu(T,\xi),\end{aligned}$$

where the last equality holds since

$$h_\mu\left(T,\xi\middle|\eta_{-\infty}^\infty\right) = H_\mu\left(\xi\middle|\xi_1^\infty\vee\eta_{-\infty}^\infty\right)$$

and

$$\eta_{-\infty}^\infty \leqslant \xi_1^\infty.$$

It follows that $h_\mu(T,\eta)=0$ since $h_\mu(T,\xi)\leqslant H_\mu(\xi)<\infty$.

Conversely, if $\eta=\{Q,X\smallsetminus Q\}\subseteq\mathscr{P}(T)$ then

$$h_\mu(T,\eta) = 0 = H_\mu(\eta\big|\eta_1^\infty).$$

Using the characterization of zero conditional entropy in Proposition 2.15 we see that $\eta\underset{\mu}{\subseteq}\eta_1^\infty$. In particular,

$$\eta_0^\infty \underset{\mu}{=} \eta_1^\infty \underset{\mu}{=} \eta_n^\infty$$

for all $n\geqslant 1$, which implies that

$$Q \underset{\mu}{\in} \bigcap_{n=0}^\infty \eta_n^\infty$$

as required. □

If a generator is known, then the tail σ-algebra can be expressed in terms of the generator, giving the following strengthening of Proposition 2.26.

Theorem 2.27 (Tail of generator). *Let ξ be a finite entropy generator for an invertible measure-preserving transformation T. Then the Pinsker σ-algebra*

$$\mathscr{P}(T) \underset{\mu}{=} \bigcap_{n=1}^\infty \xi_n^\infty$$

equals the tail *of the generator ξ. In particular, the σ-algebra of invariant sets $\mathscr{E}$ is modulo μ a subset of ξ_1^∞.*

Example 2.28. Let $(X,\mathscr{B},\mu,\sigma)$ be the Bernoulli shift defined by the probability vector $(p_1,\dots,p_s)$, so that $X=\prod_{\mathbb{Z}}\{1,\dots,s\}$, $\mu=\prod_{\mathbb{Z}}(p_1,\dots,p_s)$, and $T=\sigma$ is the left shift. Recall that the state partition

$$\xi = \{[1]_0,[2]_0,\dots,[s]_0\}$$

is a generator. Since $\xi_{-k}^k\perp\xi_{k+1}^n$ for all $n>k$ we obtain $\xi_{-k}^k\perp\xi_{k+1}^\infty$. To see this, it suffices to show that the collection of measurable sets that are independent of ξ_{-k}^{k} forms a monotone class or reformulate the independence as maximality of entropy (Propositions 2.16), then apply continuity of entropy (Proposition 2.14) and finally apply Propositions 2.16 again. This also implies $\xi_{-k}^{k} \perp \mathscr{P}(T)$ since $\mathscr{P}(T) \subseteq \xi_{k+1}^{\infty}$. As this holds for all $k \geqslant 1$ we obtain $\mathscr{B} \perp \mathscr{P}(T)$, which implies that

$$\mathscr{P}(T) \underset{\mu}{=} \mathscr{N} = \{X, \varnothing\}$$

since any element $P \in \mathscr{P}(T)$ is now independent to itself. Thus a Bernoulli shift has completely positive entropy.(14)

PROOF OF THEOREM 2.27. For any $n \geqslant 1$, the partition ξ_{-n}^{n} is a generator for T^{2n}, so by the Kolmogorov–Sinaĭ theorem (Theorem 2.20)

$$h_\mu(T^{2n}) = H_\mu\left(\xi_{-n}^{n} \middle| \xi_n^{\infty}\right) \tag{2.14}$$

and similarly

$$h_\mu(T^{2n} \big| \mathscr{P}(T)) = H_\mu\left(\xi_{-n}^{n} \middle| \xi_n^{\infty} \vee \mathscr{P}(T)\right). \tag{2.15}$$

On the other hand, by Proposition 2.23,

$$h_\mu(T^{2n}) = 2n h_\mu(T) = 2n h_\mu(T \big| \mathscr{P}(T)) = h_\mu(T^{2n} \big| \mathscr{P}(T)).$$

Thus, for any finite partition $\eta \subseteq \mathscr{P}(T)$ we have, by additivity of entropy (Proposition 2.13(1)) applied in two different ways, continuity of entropy (Proposition 2.14), and the combination of (2.14)–(2.15) that

$$\begin{aligned} H_\mu\left(\eta \middle| \xi_n^{\infty}\right) &= H_\mu\left(\xi_{-n}^{n} \vee \eta \middle| \xi_n^{\infty}\right) - H_\mu\left(\xi_{-n}^{n} \middle| \xi_n^{\infty} \vee \eta\right) \\ &= \underbrace{H_\mu\left(\eta \middle| \xi_n^{\infty} \vee \xi_{-n}^{n}\right)}_{<\varepsilon \text{ for large } n} + \underbrace{H_\mu\left(\xi_{-n}^{n} \middle| \xi_n^{\infty}\right) - H_\mu\left(\xi_{-n}^{n} \middle| \xi_n^{\infty} \vee \eta\right)}_{=0}. \end{aligned}$$

It follows that

$$H_\mu\left(\eta \middle| \xi_n^{\infty}\right) < \varepsilon$$

for all large enough n. Hence, by continuity of entropy (Proposition 2.12), and the characterization of vanishing of conditional entropy in Proposition 2.15, we have

$$\eta \underset{\mu}{\subseteq} \bigcap_{n=1}^{\infty} \xi_n^{\infty},$$

which gives

$$\mathscr{P}(T) \underset{\mu}{\subseteq} \bigcap_{n=1}^{\infty} \xi_n^{\infty}$$

since $\eta \subseteq \mathscr{P}(T)$ was an arbitrary finite partition. The opposite inclusion follows from Proposition 2.26. $\square$

A similar result holds for a sequence of partitions that generate under the transformation in the limit.

Theorem 2.29 (Tails for a generating sequence). *Let $(X, \mathscr{B}, \mu, T)$ be an invertible measure-preserving system on a Borel probability space and let (ξ_k) be an increasing sequence of partitions (that is, $\xi_k \subseteq \sigma(\xi_{k+1})$ for all $k \geqslant 1$) of finite entropy with the property that $(\xi_k)_{-\infty}^{\infty} \nearrow \mathscr{B}$ as $k \to \infty$. Then*

$$\mathscr{P}(T) \underset{\mu}{=} \bigvee_{k \geqslant 1} \bigcap_{n=1}^{\infty} (\xi_k)_n^{\infty}.$$

PROOF. Given any finite partition $\eta \subseteq \mathscr{P}(T)$ and $\varepsilon > 0$, choose k so large that $H_\mu\left(\eta \middle| (\xi_k)_{-\infty}^{\infty}\right) < \varepsilon$, and proceed as in the proof of Theorem 2.27. □

Exercises for Section 2.4

Exercise 2.4.1. Let $(X, \mathscr{B}, \mu, T)$ be an invertible measure-preserving system. Let ξ, η be two partitions with finite entropy. Show directly that

$$I_\mu\Big(\xi \Big| \xi_1^{\infty} \vee \bigcap_{n \geqslant 1} \eta_n^{\infty}\Big) = I_\mu(\xi \big| \xi_1^{\infty})$$

and interpret this as a relative independence of the elements of ξ and the tail of η when conditioned on ξ_1^{∞}.

Exercise 2.4.2. Let $(X, \mathscr{B}, \mu, T)$ be an invertible measure-preserving system. Show that we have $\mathscr{P}(T) = \mathscr{P}(T^n)$ for any $n \in \mathbb{Z} \smallsetminus \{0\}$.

Exercise 2.4.3. Complete the proof of Theorem 2.29 in greater detail.

Exercise 2.4.4 (Countable Lebesgue spectrum[(15)]). Let $\mathsf{X} = (X, \mathscr{B}, \mu, T)$ be an invertible measure-preserving system on a Borel probability space. Use the steps below to show that if X has completely positive entropy, then it has *countable Lebesgue spectrum*.
(a) Let ξ be a finite partition for X, and let $V \subseteq L^2(\xi_0^{\infty})$ be the orthogonal complement of $L^2(\xi_1^{\infty})$. Show that the subspaces $U_T^k V$ for $k \in \mathbb{Z}$ are mutually orthogonal and that

$$L^2(\xi_{-\infty}^{\infty}) = \bigoplus_{k \in \mathbb{Z}} U_T^k V \oplus L^2(\mathscr{T}),$$

where

$$\mathscr{T} = \bigcap_{n=0}^{\infty} \xi_n^{\infty}$$

is the tail σ-algebra of ξ.
(b) Notice[†] that (a) implies that the unitary operator U_T has pure Lebesgue spectrum on

$$\bigoplus_{k \in \mathbb{Z}} U_T^k V$$

[†] After looking up the spectral theorem if necessary, see [52, Th. B.4] or [53, Sec. 9.1].

with multiplicity $\dim V$.
(c) Let $A \in \xi_1^\infty$ be an atom of the σ-algebra (meaning that for any B in the σ-algebra generated by ξ_1^∞ we have $\mu(A \cap B) \in \{0, \mu(A)\}$). Show that $A \in \mathscr{T}$.
(d) Using (c), show that if $V \neq \{0\}$ then V is infinite-dimensional.
(e) If $(X, \mathscr{B}, \mu, T)$ has completely positive entropy, deduce that U_T has countable Lebesgue spectrum.
(f) Replacing the assumption of completely positive entropy with the assumption of positive entropy, generalize (e) to show that the unitary operator U_T restricted to the orthogonal complement of $L^2(\mathscr{P}(T))$ within $L^2(\mathscr{B})$ has countable Lebesgue spectrum.

2.5 Entropy and Disjointness

We start by recalling the following fundamental definitions, which were introduced into ergodic theory by Furstenberg [64] in an influential paper of 1967.

Definition 2.30. Let $\mathsf{X} = (X, \mathscr{B}_X, \mu, T)$ and $\mathsf{Y} = (Y, \mathscr{B}_Y, \nu, S)$ be invertible measure-preserving systems on Borel probability spaces. A measure ρ on

$$(X \times Y, \mathscr{B}_X \otimes \mathscr{B}_Y)$$

is a *joining* of the two systems if

- ρ is invariant under $T \times S$, and
- the projections of ρ onto the X and Y coordinates are μ and ν respectively.

The product measure $\mu \times \nu$ is called the *trivial joining*. The two systems X and Y are said to be *disjoint*, written as $\mathsf{X} \perp \mathsf{Y}$, if the trivial joining is the only joining between X and Y.

Notice that the trivial joining is always a joining. Moreover, disjointness is the strongest sense in which two measure-preserving systems can be unrelated. In particular, disjoint systems cannot have a non-trivial factor in common, because of the existence of the *relatively independent joining* over a common factor (we refer to [64] or [52, Sec. 6.5] for the details).

Also notice that spectral information about the systems can sometimes be used to prove disjointness. For example, a weakly mixing systems is always disjoint from a Kronecker system (see [64], [52, Ex. 6.5.4], or [53, Sec. 9.1.5]). As remarked in [64], entropy may also sometimes be used to show disjointness.

Theorem 2.31 (Disjointness via Entropy). *Let X and Y be invertible measure-preserving systems on Borel probability spaces, and suppose that X has zero entropy and Y has completely positive entropy. Then X and Y are disjoint.*

In the next corollary, we implicitly restrict attention to the class of invertible measure-preserving systems on Borel probability spaces.

Corollary 2.32. *Let* $\mathscr{D} = \{\mathsf{X} \mid \mathsf{X}$ *has zero entropy*$\}$ *be the class of* deterministic systems. *Then*

$$\mathscr{D}^{\perp} = \{\mathsf{Y} \mid \mathsf{Y} \perp \mathsf{X} \text{ for all } \mathsf{X} \in \mathscr{D}\}$$

is equal to the class $\mathscr{K} = \{\mathsf{Y} \mid \mathsf{Y}$ *has completely positive entropy*$\}$.

PROOF. By Theorem 2.31 we have $\mathscr{K} \subseteq \mathscr{D}^{\perp}$. For the reverse inclusion, assume that $\mathsf{Y} \in \mathscr{D}^{\perp}$, and let $\mathsf{Y}_{\mathscr{P}}$ denote the Pinsker factor of Y as defined by Proposition 2.23. Then by definition we have $\mathsf{Y}_{\mathscr{P}} \in \mathscr{D}$, and so must be disjoint from Y. The graph of the factor map

$$\pi_{\mathscr{P}} : \mathsf{Y} = (Y, \mathscr{B}_Y, \nu, S) \longrightarrow \mathsf{Y}_{\mathscr{P}} = (Y_{\mathscr{P}}, \mathscr{P}, \nu|_{\mathscr{P}}, S_{\mathscr{P}})$$

gives rise to a joining ρ defined by the property that

$$\rho(A \times B) = \nu(A \cap \pi_{\mathscr{P}}^{-1}(B))$$

for $A \in \mathscr{B}_Y$ and $B \in \mathscr{P}$. It follows that $\mathsf{Y}_{\mathscr{P}}$ must be the trivial factor, and hence $\mathsf{Y} \in \mathscr{K}$. □

PROOF OF THEOREM 2.31. Let $\mathsf{X} = (X, \mathscr{B}_X, \mu, T)$ and $\mathsf{Y} = (Y, \mathscr{B}_Y, \nu, S)$ be invertible measure-preserving systems on Borel probability spaces, assume that X has zero entropy and Y has completely positive entropy, and let ρ be a joining of X and Y. Write $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ for the projection maps onto X and Y.

Now let η be a finite partition of X into elements of $\mathscr{B}_X$, and let ξ be a finite partition of Y into elements of $\mathscr{B}_Y$. We wish to show that $\pi_X^{-1}\eta$ and $\pi_Y^{-1}\xi$ are independent with respect to ρ (see Exercise 1.1.3 and Proposition 2.16). As η and ξ are arbitrary, this will then imply that $\rho = \mu \times \nu$ is the trivial joining. Using additivity of dynamical entropy (Proposition 2.19(2)) in both of the possible ways, we obtain for the partition $\eta \times \xi = \pi_X^{-1}\eta \vee \pi_Y^{-1}\xi$ and the joining ρ the identities

$$\begin{aligned} h_\rho(T \times S, \eta \times \xi) &= \underbrace{h_\rho(T \times S, \pi_X^{-1}\eta)}_{=h_\mu(T,\eta)=0} + h_\rho(T \times S, \pi_Y^{-1}\xi \big| \pi_X^{-1}\eta_{-\infty}^{\infty}) \\ &= \underbrace{h_\rho(T \times S, \pi_Y^{-1}\xi)}_{=h_\nu(S,\xi)} + \underbrace{h_\rho(T \times S, \pi_X^{-1}\eta \big| \pi_Y^{-1}\xi_{-\infty}^{\infty})}_{\leqslant h_\rho(T\times S, \pi_X^{-1}\eta)=0}. \end{aligned}$$

Alternatively, we may rewrite this equality in the form

$$H_\nu\left(\xi \big| \xi_1^{\infty}\right) = H_\rho\left(\pi_Y^{-1}\xi \big| \pi_Y^{-1}\xi_1^{\infty} \vee \pi_X^{-1}\eta_{-\infty}^{\infty}\right), \tag{2.16}$$

by the future formula of entropy (Proposition 2.19(1)). This already expresses some form of the independence property we seek, but we still need to dispose of the ξ_1^{∞} term using the assumptions on T. Using additivity of entropy

(Proposition 2.13(1)) and invariance of entropy (Lemma 2.17), for any $n \geqslant 1$ we have

$$H_\nu\left(\xi_0^{n-1}\middle|\xi_n^\infty\right) = H_\nu\left(T^{-(n-1)}\xi\middle|\xi_n^\infty\right) + H_\nu\left(T^{-(n-2)}\xi\middle|\xi_{n-1}^\infty\right) + \cdots$$
$$+ H_\nu\left(\xi\middle|\xi_1^\infty\right) = nH_\nu\left(\xi\middle|\xi_1^\infty\right),$$

and similarly

$$H_\rho\left(\pi_Y^{-1}\xi_0^{n-1}\middle|\pi_Y^{-1}\xi_n^\infty \vee \pi_X^{-1}\eta_{-\infty}^\infty\right) = nH_\rho\left(\pi_Y^{-1}\xi\middle|\pi_Y^{-1}\xi_1^\infty \vee \pi_X\eta_{-\infty}^\infty\right)$$
$$= H_\nu\left(\xi_0^{n-1}\middle|\xi_n^\infty\right),$$

where we used (2.16) in the last step. Also notice that

$$\begin{aligned} H_\nu\left(\xi_0^{n-1}\middle|\xi_n^\infty\right) &= H_\nu\left(\xi\middle|\xi_n^\infty\right) + H_\nu\left(\xi_1^{n-1}\middle|\xi \vee \xi_n^\infty\right), \\ H_\nu\left(\xi\middle|\xi_n^\infty\right) &= H_\rho\left(\pi_Y^{-1}\xi\middle|\pi_Y^{-1}\xi_n^\infty\right) \\ &\geqslant H_\rho\left(\pi_Y^{-1}\xi\middle|\pi_Y^{-1}\xi_n^\infty \vee \pi_X^{-1}\eta_{-\infty}^\infty\right), \\ H_\nu\left(\xi_1^{n-1}\middle|\xi \vee \xi_n^\infty\right) &= H_\rho\left(\pi_Y^{-1}\xi_1^{n-1}\middle|\pi_Y^{-1}\xi \vee \pi_Y^{-1}\xi_n^\infty\right) \\ &\geqslant H_\rho\left(\pi_Y^{-1}\xi_1^{n-1}\middle|\pi_Y^{-1}\xi \vee \pi_Y^{-1}\xi_n^\infty \vee \pi_X^{-1}\eta_{-\infty}^\infty\right), \end{aligned}$$

and

$$H_\rho\left(\pi_Y^{-1}\xi\middle|\pi_Y^{-1}\xi_n^\infty \vee \pi_X^{-1}\eta_{-\infty}^\infty\right) + H_\rho\left(\pi_Y^{-1}\xi_1^{n-1}\middle|\pi_Y^{-1}\xi \vee \pi_Y^{-1}\xi_n^\infty \vee \pi_X^{-1}\eta_{-\infty}^\infty\right)$$
$$= H_\rho\left(\pi_Y^{-1}\xi_0^{n-1}\middle|\pi_Y^{-1}\xi_n^\infty \vee \pi_X^{-1}\eta_{-\infty}^\infty\right).$$

Together these relations force there to be equality in all of the inequalities above. Hence

$$\begin{aligned} H_\nu\left(\xi\middle|\xi_n^\infty\right) &= H_\rho\left(\pi_Y^{-1}\xi\middle|\pi_Y^{-1}\xi_n^\infty \vee \pi_X^{-1}\eta_{-\infty}^\infty\right) \\ &\leqslant H_\rho\left(\pi_Y^{-1}\xi\middle|\pi_X^{-1}\eta\right) \leqslant H_\rho\left(\pi_Y^{-1}\xi\right) = H_\nu\left(\xi\right). \end{aligned}$$

We now let $n \to \infty$. By our assumption on $\mathsf{Y} = (Y, \mathscr{B}_Y, \nu, S)$ its Pinsker σ-algebra is trivial and since the tail of ξ belongs to the Pinsker σ-algebra (Proposition 2.26) we know that

$$\bigcap_{n=1}^\infty \xi_n^\infty$$

coincides with the trivial σ-algebra modulo ν. Applying continuity of entropy with respect to the given σ-algebra (Proposition 2.12) we deduce that

$$H_\nu(\xi) = \lim_{n\to\infty} H_\nu\left(\xi\middle|\xi_n^\infty\right) \leqslant H_\rho\left(\pi_Y^{-1}\xi\middle|\pi_X^{-1}\eta\right) \leqslant H_\nu(\xi).$$

It follows that

$$H_\rho\left(\pi_Y^{-1}\xi \middle| \pi_X^{-1}\eta\right) = H_\rho\left(\pi_Y^{-1}\xi\right) = H_\nu(\xi),$$

and so $\pi_Y^{-1}\xi$ and $\pi_X^{-1}\eta$ are independent with respect to ρ by Proposition 2.16. As mentioned earlier, this implies that $\rho = \mu \times \nu$ and hence the theorem. □

Exercises for Section 2.5

Exercise 2.5.1. Let $\mathsf{X} = (X, \mathscr{B}_X, \mu, T)$ and $\mathsf{Y} = (Y, \mathscr{B}_Y, \nu, S)$ be invertible measure-preserving systems on Borel probability spaces with completely positive entropy, and let ρ be an ergodic joining of X and Y. Then ρ may not have completely positive entropy, and the conditional measures $\rho_{(x,y)}^{\mathscr{P}}$ of ρ for the Pinsker σ-algebra may not be invariant under $T \times S$ (see Exercise 2.5.2). Show that nonetheless we have

$$(\pi_X)_*\rho_{(x,y)}^{\mathscr{P}} = \mu$$

and

$$(\pi_Y)_*\rho_{(x,y)}^{\mathscr{P}} = \nu$$

for ρ-almost every $(x, y) \in X \times Y$.

Exercise 2.5.2. Let $A \in \mathrm{GL}_r(\mathbb{Z})$ be a *quasihyperbolic* matrix,(16) so that $T_A : \mathbb{T}^r \to \mathbb{T}^r$ is ergodic but there is a plane $V < \mathbb{R}^r$ with the property that $A|_V : V \to V$ is a rotation. Let $m_\triangle$ denote the Haar measure on the diagonally embedded torus $\mathbb{T}^r < \mathbb{T}^r \times \mathbb{T}^r$. Clearly $m_\triangle$ is an ergodic joining between T_A on $\mathbb{T}^r$ and itself. Let ρ denote the Lebesgue measure (normalized to be a probability measure) on a fixed circle in the plane V, and consider ρ as an invariant measure for T_A in $\mathbb{T}^r \cong \mathbb{T}^r \times \{0\}$. Show that $\mu = \rho * m_\triangle$ is an ergodic joining between T_A and itself. Show that the Pinsker factor of μ coincides with the Kronecker factor, and is isomorphic to $(\mathbb{T}^r, \mathscr{B}_{\mathbb{T}^r}, \rho, T_A)$.

2.6 Entropy and Convex Combinations

It is often useful to assume an invariant measure is ergodic and Theorem 2.7 shows how any invariant measure can be decomposed into ergodic components. In this section we show how entropy behaves with respect to generalized convex combinations, including the ergodic decomposition as a special case.

Theorem 2.33 (Entropy and convex combinations). *Let $(X, \mathscr{B}_X, \mu, T)$ be an invertible measure-preserving system on a Borel probability space, with ergodic decomposition*

$$\mu = \int_Y \mu_y \,\mathrm{d}\nu(y), \tag{2.17}$$

where $(Y, \mathscr{B}_Y, \nu)$ is some Borel probability space. Then

$$h_\mu(T) = \int_Y h_{\mu_y}(T)\,\mathrm{d}\nu(y) \tag{2.18}$$

and more generally

$$h_\mu(T,\xi\big|\mathscr{A}) = \int_Y h_{\mu_y}(T,\xi\big|\mathscr{A})\,\mathrm{d}\nu(y) \tag{2.19}$$

for any partition ξ with $H_\mu(\xi) < \infty$ and T-invariant sub σ-algebra $\mathscr{A} \subseteq \mathscr{B}$. The conclusion in (2.18) *also holds if* (2.17) *is any way of expressing μ as a generalized convex combination of invariant measures.*

We will give two related but slightly different proofs, the first for the ergodic decomposition exploits the construction of the ergodic decomposition using conditional measures with respect to the σ-algebra of invariant sets. The second proof uses the Abramov–Rokhlin formula (Corollary 2.21) to deal with any generalized convex combination of measures.

Proof of equations (2.18) and (2.19) using the Pinsker algebra. Recall from Theorem 2.7 that one way to construct the ergodic decomposition is to use $Y = X$ and $\mu_x = \mu_x^{\mathscr{E}}$ for $x \in X$, where

$$\mathscr{E} = \{E \in \mathscr{B}_X \mid T^{-1}E = E\}.$$

If $E \in \mathscr{E}$ then

$$T^{-1}\{E, X\smallsetminus E\} = \{E, X\smallsetminus E\}$$

so $E \in \mathscr{P}(T)$, the Pinsker algebra of T. If ξ is a partition with $H_\mu(\xi) < \infty$, then since conditioning on the Pinsker does not affect dynamical entropy (Lemma 2.24), since conditional entropy equals an average (Lemma 2.11), by the double conditioning formula (Proposition 2.4), and dominated convergence we have

$$\begin{aligned}
h_\mu(T,\xi\big|\mathscr{A}) &= h_\mu(T,\xi\big|\mathscr{E}\vee\mathscr{A})\\
&= \lim_{N\to\infty}\frac{1}{N}H_\mu\left(\xi_0^{N-1}\big|\mathscr{E}\vee\mathscr{A}\right)\\
&= \lim_{N\to\infty}\int\frac{1}{N}H_{\mu_x^{\mathscr{E}\vee\mathscr{A}}}\left(\xi_0^{N-1}\right)\,\mathrm{d}\mu(x)\\
&= \lim_{N\to\infty}\int\frac{1}{N}\int H_{\mu_y^{\mathscr{E}\vee\mathscr{A}}}\left(\xi_0^{N-1}\right)\,\mathrm{d}\mu_x^{\mathscr{E}}(y)\,\mathrm{d}\mu(x)\\
&= \lim_{N\to\infty}\int\frac{1}{N}\int H_{(\mu_x^{\mathscr{E}})_y^{\mathscr{E}\vee\mathscr{A}}}\left(\xi_0^{N-1}\right)\,\mathrm{d}\mu_x^{\mathscr{E}}(y)\,\mathrm{d}\mu(x)\\
&= \lim_{N\to\infty}\int\frac{1}{N}H_{\mu_x^{\mathscr{E}}}\left(\xi_0^{N-1}\big|\mathscr{A}\right)\,\mathrm{d}\mu(x)\\
&= \int h_{\mu_x^{\mathscr{E}}}(T,\xi\big|\mathscr{A})\,\mathrm{d}\mu(x).
\end{aligned}$$

Now take a generating sequence of finite partitions (ξ_n) with

$$\sigma(\xi_n)\nearrow\mathscr{B}_X$$

to deduce, using the Kolmogorov–Sinaǐ theorem (Theorem 2.20) and monotone convergence of the function

$$x \longmapsto h_{\mu_x^{\mathscr{A}}}(T, \xi_n)$$

for $n \to \infty$ in the integral above, that $h_\mu(T) = \int h_{\mu_x^{\mathscr{E}}}(T)\,\mathrm{d}\mu(x)$. □

Proof of (2.18) using the Abramov–Rokhlin formula. By assumption, we are given a Borel probability space $(Y, \mathscr{B}_Y, \nu)$ and a measurable function $y \mapsto \mu_y$, defined ν-almost everywhere, with μ_y a T-invariant Borel probability measure on $(X, \mathscr{B}_X)$, such that

$$\mu = \int_Y \mu_y \,\mathrm{d}\nu(y).$$

We define a probability measure ρ on $(Z, \mathscr{B}_Z) = (X \times Y, \mathscr{B}_X \otimes \mathscr{B}_Y)$ by

$$\rho = \int_Y \mu_y \times \delta_y \,\mathrm{d}\nu(y),$$

and note that if π_X, π_Y denote the projections onto the X and Y coordinates from Z, then $(\pi_X)_*\rho = \mu$ and $(\pi_Y)_*\rho = \nu$. We will use the Abramov–Rokhlin formula (Corollary 2.21) to compute the entropy of the map $T \times I_Y : Z \to Z$ sending (x, y) to $(T(x), y)$, in two different ways. Firstly, $T \times I_Y$ is an extension of the identity map on Y, so

$$h_\rho(T \times I_Y) = h_\nu(I_Y) + h_\rho(T \times I_Y \big| \mathscr{N}_X \times \mathscr{B}_Y) \tag{2.20}$$

where $\mathscr{N}_X = \{\varnothing, X\}$ is the trivial σ-algebra on X; secondly, $T \times I_Y$ is an extension of the map T on $(X, \mathscr{B}_X, \mu)$, so

$$h_\rho(T \times I_Y) = h_\mu(T) + h_\rho(T \times I_Y \big| \mathscr{B}_X \times \mathscr{N}_Y). \tag{2.21}$$

Clearly $h_\nu(I_Y) = 0$ as in Example 1.24. We claim that

$$h_\rho(T \times I_Y \big| \mathscr{B}_X \times \mathscr{N}_Y) = 0$$

for the following reason. Let (ξ_n) be an increasing sequence of partitions of X with $\sigma(\xi_n) \nearrow \mathscr{B}_X$, and similarly let (η_n) be an increasing sequence of partitions of Y with $\sigma(\eta_n) \nearrow \mathscr{B}_Y$. Then

$$\begin{aligned} h_\rho(T \times I_Y, \xi_n \times \eta_n \big| \mathscr{B}_X \times \mathscr{N}_Y) \leqslant\ & h_\rho(T \times I_Y, \xi_n \times \{\varnothing, Y\} \big| \mathscr{B}_X \times \mathscr{N}_Y) \\ & + h_\rho(T \times I_Y, \{\varnothing, X\} \times \eta_n \big| \mathscr{B}_X \times \mathscr{N}_Y) \end{aligned}$$

and both terms vanish by Proposition 2.15. It follows from (2.20) and (2.21) that

$$h_\mu(T) = h_\rho(T \times I_Y \big| \mathscr{N}_X \times \mathscr{B}_Y).$$

Now

$$\rho_{(x,y)}^{\mathscr{N}_X\times\mathscr{B}_Y} = \mu_y \times \delta_y \tag{2.22}$$

by the definition of ρ and [52, Prop. 5.19].

Let $\xi \subseteq \mathscr{B}_X$ and $\eta \subseteq \mathscr{B}_Y$ be finite partitions. Then

$$\begin{aligned} h_\rho(T\times I_Y, \xi\times\eta \big| \mathscr{N}_X\times\mathscr{B}_Y) &= \lim_{n\to\infty}\frac{1}{n}H_\rho\left(\xi_0^{n-1}\times\eta \big| \mathscr{N}_X\times\mathscr{B}_Y\right)\\ &= \lim_{n\to\infty}\frac{1}{n}H_\rho\left(\xi_0^{n-1}\times\{\varnothing, Y\} \big| \mathscr{N}_X\times\mathscr{B}_Y\right)\\ &= \lim_{n\to\infty}\int\frac{1}{n}H_{\mu_y}\left(\xi_0^{n-1}\right)\,\mathrm{d}\nu(y)\end{aligned}$$

since conditional entropy equals an average of entropies (Lemma 2.11) and by (2.22). By the dominated convergence theorem, we conclude that

$$h_\rho(T\times I_Y, \xi\times\eta \big| \mathscr{N}_X\times\mathscr{B}_Y) = \int_Y h_{\mu_y}(T,\xi)\,\mathrm{d}\nu(y).$$

Finally, taking sequences $\xi_n \nearrow \mathscr{B}$ and $\eta_n \nearrow \mathscr{B}_Y$, and using the Kolmogorov–Sinaĭ theorem (Theorem 2.20) and the monotone convergence theorem, we obtain

$$h_\mu(T) = h_\rho(T\times I_Y \big| \mathscr{N}_X\times\mathscr{B}_Y) = \int h_{\mu_y}(T)\,\mathrm{d}\nu(y)$$

as claimed. □

Exercises for Section 2.6

Exercise 2.6.1. Prove (2.18) in the non-invertible case by proving $h_\mu(T,\xi) = h_\mu(T,\xi\big|\mathscr{E})$ without referring to Section 2.4 (where invertibility was assumed).

2.7 An Entropy Calculation: Other Measures

†We return now to the specific automorphism of the torus discussed in Section 1.6. Recall that this is the map $T = T_A : \mathbb{T}^2 \to \mathbb{T}^2$ defined by

$$T : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} y \\ x+y \end{pmatrix} \pmod 1,$$

† Just as for Section 1.6, this section could be skipped if the reader wants to focus on the theoretical developments.

which is naturally associated to the matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$.

Behind the somewhat enigmatic details of the argument in Section 1.6 lies a simple idea reflected in the geometry of the action of T on suitable rectangles: it is contraction (respectively, expansion) along eigenspaces for eigenvalues of absolute value less (resp. greater) than one that contributes to the entropy. However, if we are considering an arbitrary invariant Borel measure μ, naturally the properties of the measure also play a role in computing $h_\mu(T)$. In this section, we will explain why $h_\mu(T)$ depends mainly on the properties of the conditional measures $\mu_x^{\mathscr{A}}$ for $\mathscr{A}$ as in (2.23) below as x varies in $\mathbb{T}^2$. This is a very special case of a profound theory.(17)

We let ξ be the partition of $\mathbb{T}^2$ as in Figure 1.3. Assuming that μ gives zero mass to the origin, the boundaries of the elements of ξ are null sets. To see this, recall that the boundaries of these rectangles are made of pieces of $\mathbb{R}\mathbf{v}^+$ and $\mathbb{R}\mathbf{v}^-$ through integer points. In $\mathbb{T}^2$ this means that these points either approach the origin in their backward orbit or in their forward orbit. In either case we can apply Poincaré recurrence (see [52, Th. 2.11]) to see that the boundary is a μ-null set. Hence we do not have to specify to which elements these boundaries belong. It is easy to derive the following from Lemma 1.33.

Lemma 2.34. *The atoms for the σ-algebra*

$$\mathscr{A} = \xi \vee T^{-1}\xi \vee \cdots = \bigvee_{i=0}^{\infty} T^{-i}\xi \tag{2.23}$$

are line segments parallel to $\mathbf{v}^-$ as long as the element of ξ containing them, and ξ is a generator under T.

Define the *stable subgroup* for T by

$$U^- = \mathbb{R}\mathbf{v}^-,$$

and the *unstable subgroup* of T by

$$U^+ = \mathbb{R}\mathbf{v}^+.$$

For $x \in \mathbb{R}^2$ the *stable manifold* through x is the coset $x + U^-$ and the *unstable manifold* is $x + U^+$. Finally, for $\delta > 0$ we let

$$B_\delta^{U^-}(x) = x + \bigl(U^- \cap B_\delta(0)\bigr)$$

denote the δ-neighbourhood of x inside the stable manifold. The δ-neighbourhood of x inside the unstable manifold, $B_\delta^{U^+}(x)$, is defined similarly. It is important to note that we consider the intersections like $U^- \cap B_\delta(0)$ in the covering space $\mathbb{R}^2$, while the translation by the point x is made in $\mathbb{T}^2$ to define finally a subset of $\mathbb{T}^2$.

Theorem 2.35 (Entropy and conditional measures). *Let $T = T_A$ be the automorphism of the 2-torus associated to the matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ and let μ be a T-invariant non-atomic probability measure on $\mathbb{T}^2$. Then for μ-almost every x we have*

$$h_{\mu_x^{\mathscr{E}}}(T) = \lim_{n\to\infty} \frac{-\log \mu_x^{\mathscr{A}}([x]_{T^n\mathscr{A}})}{n}. \tag{2.24}$$

That is, the limit on the right-hand side of (2.24) exists and equals the entropy of T with respect to the ergodic component $\mu_x^{\mathscr{E}}$ of μ at x. In particular, by Theorem 2.33,

$$h_\mu(T) = \int h_{\mu_x^{\mathscr{E}}}(T)\,\mathrm{d}\mu(x) = \int \lim_{n\to\infty} \frac{-\log \mu_x^{\mathscr{A}}[x]_{T^n\mathscr{A}}}{n}\,\mathrm{d}\mu(x).$$

Proof of Theorem 2.35. Define $f(x) = I_\mu(\xi\big| T^{-1}\mathscr{A})(x)$ for the generating partition ξ as in Lemma 1.33, and let $\mathscr{A}$ be as in (2.23). Then by invariance of the information function (Lemma 2.17)

$$\begin{aligned} f(T^{-1}x) &= I_\mu(\xi\big|T^{-1}\mathscr{A})(T^{-1}x) \\ &= I_\mu(T\xi\big|\mathscr{A})(x) \end{aligned}$$

and also

$$f(T^{-k}x) = I_\mu(T^k\xi\big|T^{k-1}\mathscr{A})(x)$$

for all $k \geqslant 0$. Thus additivity of the entropy function (Proposition 2.13) and $\xi \vee T^{-1}\mathscr{A} = \mathscr{A}$ gives

$$\begin{aligned} \frac{1}{n}\sum_{k=0}^{n-1} f(T^{-k}x) &= \frac{1}{n}\sum_{k=0}^{n-1} I_\mu(T^k\xi\big|T^{k-1}\mathscr{A})(x) \\ &= \frac{1}{n} I_\mu(\xi \vee T\xi \vee \cdots \vee T^{n-1}\xi\big|T^{-1}\mathscr{A})(x) \\ &= -\frac{n-1}{n}\cdot\frac{1}{n-1}\log \mu_x^{T^{-1}\mathscr{A}}\left([x]_{T^{n-1}\mathscr{A}}\right). \end{aligned}$$

By the pointwise ergodic theorem, it follows that the limit of the expression on the right of (2.24) exists and is equal to $E_\mu\left(I_\mu(\xi\big|T^{-1}\mathscr{A})\big|\mathscr{E}\right)(x)$. Next notice that $\mathscr{E} \subseteq \mathscr{A}$ by Theorem 2.27, since ξ is a generator. Therefore, by the double conditioning formula (Proposition 2.4),

$$\begin{aligned} I_\mu\left(\xi\big|T^{-1}\mathscr{A}\right)(y) &= -\log \mu_y^{T^{-1}\mathscr{A}}([y]_\xi) \\ &= -\log\left(\mu_x^{\mathscr{E}}\right)_y^{T^{-1}\mathscr{A}}([y]_\xi) = I_{\mu_x^{\mathscr{E}}}\left(\xi\big|T^{-1}\mathscr{A}\right)(y) \end{aligned}$$

for $\mu_x^{\mathscr{E}}$-almost every y and μ-almost every $x \in X$. Integrating this over y with respect to $\mu_x^{\mathscr{E}}$ now gives

$$E\left(I_\mu(\xi\big|T^{-1}\mathscr{A})\big|\mathscr{E}\right)(x) = h_{\mu_x^{\mathscr{E}}}(T,\xi) = h_{\mu_x^{\mathscr{E}}}(T)$$

and so the theorem. □

The dependence in Theorem 2.35 on the geometry of $[x]_{T^n\mathscr{A}}$ is slightly unsatisfactory. We know that $[x]_{T^n\mathscr{A}}$ is an interval in the set $x+U^-$ of size comparable to ρ^{-n}, but what we do not know is the position of x within that interval. Assuming for the moment that this does not have any influence on the entropy, we expect that the limit

$$s_x = \lim_{\delta\to 0} \frac{\log \mu_x^{\mathscr{A}}(B_\delta^{U^-}(x))}{\log\delta},$$

exists, and we may call s_x the *local dimension* of μ along the stable manifold of T. Moreover, $h_\mu(T)$ is then given by the logarithm of the expansion factor times the average dimension of μ along the stable manifold.† We will prove this extension in greater generality later.

We think of this quantity as a dimension as it tell us roughly at which power of δ the measure of a typical δ-ball decays as $\delta\to 0$. As a more formal justification of the terminology 'dimension' for s_x we show the following lemma.

Lemma 2.36 (Upper bound on dimension of measure). *For any finite Borel measure ν on $\mathbb{R}^d$ the function*

$$g(x) = \limsup_{\delta\to 0} \frac{\log\nu(B_\delta(x))}{\log\delta}$$

satisfies $g(x)\leqslant d$ for almost every $x\in\mathbb{R}^d$.

Proof. The lemma follows from the special case of measures supported on compact subsets of $\mathbb{R}^d$, and using an affine contraction it suffices to consider a finite measure supported on $[0,1]^d$.

Fix $\varepsilon\in(0,1)$, and let $A=\{x\in[0,1]^d \mid g(x) > d+\varepsilon\}$. It is enough to show that $\nu(A)=0$ if $\varepsilon>0$. Write

$$A_k = \left\{x \mid \nu\left(B_{2^{-k}}(x)\right) < 2^{d+1}2^{-k(d+\varepsilon)}\right\}.$$

If $x\in A$, then there exists a sequence $\delta_n\to 0$ with $\nu(B_{\delta_n}(x)) < \delta_n^{d+\varepsilon}$. Now define the integer sequence $k_n\to\infty$ with $2^{-k_n}\leqslant\delta_n < 2\cdot 2^{-k_n}$, so that $\nu(B_{2^{-k_n}}(x))\leqslant\delta_n^{d+\varepsilon} < 2^{d+1}2^{-(d+\varepsilon)k_n}$. Therefore,

$$A\subseteq \bigcap_{\ell\geqslant 1}\bigcup_{k\geqslant\ell} A_k.$$

† Because of the symmetry $h_\mu(T)=h_\mu(T^{-1})$ the same argument gives a similar result phrased in terms of the unstable manifolds.

If $B \subseteq [0,1]^d$ is a ball of diameter 2^{-k} with $A_k \cap B \neq \varnothing$ then the definition of A_k gives immediately

$$\nu(B) < 2^{d+1} 2^{-k(d+\varepsilon)}.$$

Since $[0,1]^d$ can be covered† by $\ll 2^{dk}$ many balls B, we conclude that

$$\nu(A_k) \ll 2^{kd} 2^{-k(d+\varepsilon)} = 2^{-\varepsilon k}$$

and therefore

$$\nu\left(\bigcup_{k \geqslant \ell} A_k\right) \ll \sum_{k \geqslant \ell} 2^{-\varepsilon k} = 2^{-\varepsilon \ell} \cdot \frac{1}{1 - 2^{-\varepsilon}} \to 0$$

as $\ell \to \infty$. It follows that $\nu(A) = 0$. □

Notes to Chapter 2

(13)(Page 67) The map T may be thought of as a skew-product construction, and the entropy formula is proved by Abramov and Rokhlin [4]. It is generalized to actions of countable amenable groups by Ward and Zhang [205].

(14)(Page 72) The converse is not true: there are measure-preserving systems with trivial Pinkser algebra that are not isomorphic to Bernoulli shifts. The distinction is a subtle one, and erroneous arguments that the two properties are the same were put forward by Wiener [209] among others. An uncountable family of non-isomorphic measure-preserving transformations with trivial Pinsker algebra, all with the same entropy, none of which is isomorphic to a Bernoulli shift, is constructed by Ornstein and Shields [153]. Smooth examples of this sort were constructed by Katok [97].

(15)(Page 74) Kolmogorov [108] considered the family of measure-preserving systems with completely positive entropy, and the (larger) family of automorphisms with countable Lebesgue spectrum as motivational examples for the new invariant of measure-theoretic entropy. Rokhlin [176] showed that completely positive entropy implies countable Lebesgue spectrum, and the question raised by Rokhlin of whether any invertible measure-preserving system can have Lebesgue spectrum of finite multiplicity remains open; Banach explictly raised the question of the existence of measure-preserving systems with Lebesgue spectrum of multiplicity one, as reported in Ulam's collection of problems [199, p. 76]. We refer to the survey of Katok and Thouvenot [99] for an overview of the spectral properties of unitary operators arising in ergodic theory. For infinite measure-preserving systems the situation is different. In [1] el Abdalaoui and Nadkarni give an example of an ergodic non-singular map whose unitary operator admits a Lebesgue component of multiplicity one in its spectrum, and we refer to that paper for a brief overview of several other results in similar directions.

(16)(Page 78) This means that no eigenvalue of A is a root of unity, and that at least one eigenvalue of A has unit modulus. This is only possible for $r \geqslant 4$, and an explicit example is

† The implicit constant depends on the choice of the norm on $\mathbb{R}^d$.

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & -4 & 2 & -4 \end{pmatrix}.$$

The implications of quasihyperbolicity for toral automorphisms (and its analog for compact connected group automorphisms) in relation to rigidity of invariant measures is studied by Lindenstrauss and Schmidt [127]. Other dynamical properties of quasihyperbolic toral automorphisms, particularly those connected to the existence of generators and the number and distribution of periodic points, were earlier studied by Lind [119].

(17)(Page 82) The material in this section is a very special case of the beginning of the theory of the entropy of diffeomorphisms developed by many researchers including Pesin [164], Ledrappier and Young [114], [115] and Mañé [130].