version Dec.7, 1999

# CHAPTER 1
# MEASURE PRESERVING ENDOMORPHISMS

## §1.1 MEASURE SPACES AND MARTINGALE THEOREM

We assume that the reader knows basic elements of measure and integral theory. For a complete treatment see for example [Halmos] or [Billingsley, 1979]. We start with some basics to fix notation and terminology.

A family $\mathcal{F}$ of subsets of a set $X$ is said to be a *$\sigma$-algebra* if the following conditions are satisfied:

$$X \in \mathcal{F}, \tag{1.1.1}$$

$$A \in \mathcal{F} \;\Rightarrow\; A^c \in \mathcal{F} \tag{1.1.2}$$

and

$$\{A_i\}_{i=1}^{\infty} \subset \mathcal{F} \;\Rightarrow\; \bigcup_{i=1}^{\infty} A_i \in \mathcal{F} \tag{1.1.3}$$

It follows from this definition that $\emptyset \in \mathcal{F}$, that the $\sigma$-algebra $\mathcal{F}$ is closed under countable intersections and subtractions of sets. If (1.1.3) is assumed only for finite subfamilies of $\mathcal{F}$ then $\mathcal{F}$ is called an *algebra*. Fixed $\mathcal{F}$, elements of the $\sigma$–algebra $\mathcal{F}$ will be frequently called *measurable* sets. For any family $\mathcal{F}_0$ of subsets of $X$, we denote by $\sigma(\mathcal{F}_0)$ the minimal $\sigma$–algebra that contains $\mathcal{F}_0$ and call it the $\sigma$–algebra *generated* by $\mathcal{F}_0$.

A function on a $\sigma$-algebra $\mathcal{F}$, $\mu : \mathcal{F} \to [0,\infty]$, is said to be *$\sigma$-additive* if for any countable subfamily $\{A_i\}_{i=1}^{\infty}$ of $\mathcal{F}$ consisting of mutually disjoint sets, we have

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mu(A_i) \tag{1.1.4}$$

We say then that $\mu$ is a *measure*. If we consider in (1.1.4) only finite families of sets, we say $\mu$ is *additive*. The two notions: of *additive* and of *$\sigma$-additive*, make sense for a $\sigma$-algebra as well as for an algebra, provided in the algebra case in (1.1.4) that all $A_i$ and their union belong to $\mathcal{F}$. The simplest consequences of the definition of measure are the following:

$$\mu(\emptyset) = 0; \tag{1.1.5}$$

$$(1.1.6) \qquad \text{if } A, B \in \mathcal{F} \text{ and } A \subset B \text{ then } \mu(A) \le \mu(B);$$

$$(1.1.7) \qquad \text{if } A_1 \subset A_2 \subset \dots \text{ and } \{A_i\}_{i=1}^{\infty} \subset \mathcal{F} \text{ then } \mu(\bigcup_{i=1}^{\infty} A_i) = \sup_i \mu(A_i) = \lim_{i\to\infty} \mu(A_i).$$

We say that the triple $(X, \mathcal{F}, \mu)$ with a $\sigma$-algebra $\mathcal{F}$ and $\mu$ a measure on $\mathcal{F}$ is a *measure space*. In this book we will always assume, unless the opposite is stated, that $\mu$ is a *finite measure* that is $\mu : \mathcal{F} \to [0, \infty)$. By (1.1.6) it equivalently means that $\mu(X) < \infty$. If $\mu(X) = 1$, the triple $(X, \mathcal{F}, \mu)$ is called *probability space* and $\mu$ *probability measure*.

We say that $\phi \to I\!\!R$ is a measurable function, if $\phi^{-1}(J) \in \mathcal{F}$ for every interval $J \subset I\!\!R$ (compare Sec.2). We say that $\phi$ is $\mu$-integrable if $\int |\phi|\, d\mu < \infty$. We write $\phi \in L^1(\mu)$. More generally, for every $1 \le p < \infty$ we write $\int |\phi|^p\, d\mu)^{1/p} = ||\phi||_p$ and say $\phi$ belongs to $L^p(\mu) = L^p(X, \mathcal{F}, \mu)$. If $\inf_{\mu(E)=0} \sup_{X\setminus E} |\phi| < \infty$ we say $\phi \in L^\infty$ and denote the latter expression by $||\phi||_\infty$. $||\phi||_p, 1 \le p \le \infty$ are called $L^p$-norms of $\phi$. We usually identify in this Chapter functions which differ only on a set of $\mu$-measure 0. $L^p(X, \mathcal{F}, \mu)$'s after these identifications are Banach spaces.

??? We say that a property $q(x)$, $x \in X$, is satisfied for $\mu$ almost every $x \in X$ (abbr: a.e.), or $\mu$-a.e., if $\mu(\{x : q(x) \text{ is not satisfied }\}) = 0$. We can consider $q$ as a subset of $X$ with $\mu(X \setminus q) = 0$.

We shall often use in the book the following two facts.

**Monotone Convergence Theorem.** Suppose $\phi_1 \le \phi_2 \le ...$ is an increasing sequence of integrable, real-valued functions on a probability space $(X, \mathcal{F}, \mu)$. Then $\phi = \lim_{n\to\infty} \phi_n$ exists a.e. and $\lim_{n\to\infty} \int \phi_n\, d\mu = \int \phi\, d\mu$. (We allow $+\infty$'s here.)

and

**Dominated Convergence Theorem.** If $\phi_n, n \ge 1$ is a sequence of measurable real-valued functions on a probability space $(X, \mathcal{F}, \mu)$ and $|\phi_n| \le g$ for an integrable function $g$ and $\phi_n \to \phi$ a.e., then $\phi$ is integrable and $\lim_{n\to\infty} \int \phi_n\, d\mu = \int \phi\, d\mu$.

Recall now that if $\mathcal{F}'$ is a sub-$\sigma$-algebra of $\mathcal{F}$ and $\phi : X \to I\!\!R$ is a $\mu$-integrable function, then there exists a unique (mod 0) function usually denoted by $E(\phi|\mathcal{F}')$ such that $E(\phi|\mathcal{F}')$ is $\mathcal{F}'$-measurable and

$$(1.1.8) \qquad \int_A E(\phi|\mathcal{F}')\, d\mu = \int_A \phi\, d\mu$$

for all $A \in \mathcal{F}'$. $E(\phi|\mathcal{F}')$ is called *conditional expectation value* of the function $\phi$ with respect to the $\sigma$-algebra $\mathcal{F}'$. Sometimes we shall use for $E(\phi|\mathcal{F}')$ the simplified notation $\phi_{\mathcal{F}'}$.

For $\mathcal{F}$ generated by a finite partition $\mathcal{A}$ (cf. Sec.3), one can think of $E(\phi|\sigma(\mathcal{A})$ as constant on each $A \in \mathcal{A}$ equal to the average $\int_A \phi \, d\mu / \mu(A)$.

The existence of $E(\phi|\mathcal{F}')$ follows from famous Radon-Nikodym theorem, saying that if $\nu \ll \mu$, the both measures defined on the same $\sigma$-algebra $\mathcal{F}'$, ($\nu \ll \mu$ means $\nu$ *absolutely continuous* with respect to $\mu$, i.e. $\mu(A) = 0 \Rightarrow \nu(A) = 0$ for all $A \in \mathcal{F}'$), then there exists a unique (mod 0) $\mathcal{F}'$-measurable, $\mu$-integrable function $\Phi = d\nu/d\mu : X \to \mathbb{R}^+$ such that for every $A \in \mathcal{F}'$

$$\int_A \Phi \, d\mu = \nu(A).$$

To deduce (1.1.8) we set $\nu(A) = \int_A \phi \, d\mu$ for $\mathcal{A} \in \mathcal{F}'$. The trick is that we restrict $\mu$ from $\mathcal{F}$ to $\mathcal{F}'$.

If $\phi \in L^p(X, \mathcal{F}, \mu)$ then $E(\phi|\mathcal{F}' \in L^p(X, \mathcal{F}', \mu)$ for all $\sigma$-algebras $\mathcal{F}'$ with $L^p$ norms uniformly bounded. More precisely the operators $\phi \to E(\phi|\mathcal{F}')$ are linear projections from $L^p(X, \mathcal{F}, \mu)$ to $L^p(X, \mathcal{F}', \mu)$, with $L^p$-norms equal to 1 (see Exercise 0.).

We end this section with the following version of **Martingale Convergence Theorem**.

**Theorem 1.1.1.** If $(\mathcal{F}_n : n \geq 1)$ is either monotone increasing or monotone decreasing sequence of $\sigma$-algebras contained in $\mathcal{F}$, then for every $\phi \in L^p(\mu)$, $1 \leq p < \infty$

$$\lim_{n \to \infty} E(\phi|\mathcal{F}_n) = E(\phi|\mathcal{F}'), \quad \text{a.e. and in} \;\; L^p,$$

where $\mathcal{F}'$ is equal to either $\bigvee_{n=1}^{\infty} \mathcal{F}_n$ or $\bigcap_{n=1}^{\infty} \mathcal{F}_n$ respectively.

In the theorem above we denoted by $\bigvee_{n=1}^{\infty} \mathcal{F}_n$ the smallest $\sigma$-algebra containing $\bigcup_{n=1}^{\infty} \mathcal{F}_n$, the latter usually is not a $\sigma$-algebra, but only an algebra. Compare Sec.6 where complete $\sigma$-algebras of this form in Lebesgue space are considered.

**Remark 1.1.2.** For the existence of $\mathcal{F}'$ and the convergence in $L^p$ no monotonicity is needed. It is sufficient to assume that for every $A \in \mathcal{F}$ the limit $\lim E(1\!\!1_A|\mathcal{F}_n)$ in measure $\mu$ exists.

(Recall that $\psi_n$ is said to *converge in measure* $\nu$ to $\psi$ if for every $\varepsilon > 0$, $\lim_{n \to \infty} \mu(\{x \in X : |\psi_n(x) - \psi(x)| \geq \varepsilon\}) \to 0$.)

In this book we denote by $1\!\!1_A$ the indicator function of $A$, namely equal to 1 on $A$ and to 0 outside $A$.

We shall not provide here a proof of Theorem 1.1.1 in the full generality . Let us provide however a proof Theorem 1.1.1 (and Remark 1.1.2 in the case $\lim E(1\!\!1_A|\mathcal{F}_n) = 1\!\!1_A$) for the $L^2$-convergence for functions $\phi \in L^2(\mu)$. (This is the case sufficient for example to prove the important Lemma 1.8.6 later on in this chapter.)

For the increasing sequence $(\mathcal{F}_n)$ we have the equality $L^2(X, \mathcal{F}', \mu) = \overline{\bigcup_n L^2(X, \mathcal{F}_n, \mu)}$. Indeed, for every $B \in \mathcal{F}'$ there exists a sequence $B_n \in \mathcal{F}_n$, $n \geq 1$, such that $\mu(B \div B_n) \to 0$. ($B \div C = (B \setminus C) \cup (C \setminus B)$ is the *symmetric difference* of sets $B$ and $C$.)

This follows for example from Carathéodory's argument, see the note Theorem 1.7.2. We have $\mu(B)$ equal to the outer measure of $B$ constructed from $\mu$ restricted to the algebra $\bigcup_{n=1}^{\infty} \mathcal{F}_n$. In the Remark 1.1.2 case where we assumed $\lim E(1\!\!1_A|\mathcal{F}_n) = 1\!\!1_A$, this is immediate.

Hence $L^2(X, \mathcal{F}_n, \mu) \ni 1\!\!1_{B_n} \to 1\!\!1_B$ in $L^2(X, \mathcal{F}, \mu)$. Finally use the fact that every function $f \in L^2(X, \mathcal{F}', \mu)$ can be approximated in the space $L^2(X, \mathcal{F}', \mu)$ by the step functions, i.e. finite linear combinations of indicator functions. Therefore, since $E(\phi|\mathcal{F}_n)$ and $E(\phi|\mathcal{F}')$ are orthogonal projections of $\phi$ to $L^2(X, \mathcal{F}_n, \mu)$ and $L^2(X, \mathcal{F}', \mu)$ respectively (exercise) we obtain $E(\phi|\mathcal{F}_n) \to E(\phi|\mathcal{F}')$ in $L^2$.

For a decreasing sequence $\mathcal{F}_n$ use the equality $L^2(X, \mathcal{F}', \mu) = \bigcap_n L^2(X, \mathcal{F}_n, \mu)$.

## §1.2 MEASURE PRESERVING ENDOMORPHISMS, ERGODICITY

Let $(X, \mathcal{F}, \mu)$ and $(X', \mathcal{F}', \mu')$ be measure spaces. A transformation $T : X \to X'$ is said to be *measurable* if $T^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{F}'$. If moreover $\mu(T^{-1}(A)) = \mu'(A)$ for every $A \in \mathcal{F}'$, then $T$ is called *measure preserving*. If $(X, \mathcal{F}, \mu) = (X', \mathcal{F}', \mu')$ we call $T$ a *measure preserving endomorphisms*; we will say also that measure $\mu$ is $T$–invariant, or that $T$ preserves $\mu$.
If a measure preserving map $T$ is invertible and the inverse $T^{-1}$ is measurable, then clearly $T^{-1}$ is also measure preserving. Therefore $T$ is an *isomorphism* in the category of measure spaces. In the case of $(X, \mathcal{F}, \mu) = (X', \mathcal{F}', \mu')$ the transformation $T$ is called an *automorphism*.
We shall prove now the following very useful fact:

**Theorem 1.2.1.** (Poincaré Recurrence Theorem) If $T : X \to X$ is a measure preserving endomorphism, then for every mesurable set $A$

$$\mu\big(\{x \in A : T^n(x) \in A \ \text{ for infinitely many } n\text{'s}\}\big) = \mu(A).$$

**Proof.** Let

$$N = N(T, A) = \{x \in A : T^n(x) \notin A \ \forall n \geq 1\}.$$

We shall first show that $\mu(N) = 0$. Indeed, $N$ is measurable since $N = A \cap \big(\bigcap_{n \geq 1} T^{-n}(X \setminus A)\big)$. If $x \in N$, then $T^n(x) \notin A$ for all $n \geq 1$ and, in particular, $T^n(x) \notin N$ which implies that $x \notin T^{-n}(N)$ and consequently $N \cap T^{-n}(N) = \emptyset$ for all $n \geq 1$. Thus, all the sets $N$, $T^{-1}(N)$, $T^{-2}(N), \ldots$ are mutually disjoint since if $n_1 \leq n_2$, then

$$T^{-n_1}(N) \cap T^{-n_2}(N) = T^{-n_1}(N \cap T^{-(n_2 - n_1)}(N)) = \emptyset.$$

Hence

$$1 \geq \mu\left(\bigcup_{n=0}^{\infty} T^{-n}(N)\right) = \sum_{n=0}^{\infty} \mu(T^{-n}(N)) = \sum_{n=0}^{\infty} \mu(N).$$

Therefore $\mu(N) = 0$. Fix now $k \geq 1$ and put

$$N_k = \{x \in A : T^n(x) \notin A \ \forall n \geq k\}.$$

Then $N_k \subset N(T^k, A)$ and therefore from what have been proved above it follows that $\mu(N_k) \le \mu(N(T^k, A)) = 0$. Thus

$$\mu\big(\{x \in A : T^n(x) \in A \text{ for only finitely many } n\text{'s}\}\big) = 0.$$

The proof is finished. ♣

A measurable transformation $T : X \to X$ of a measure space $(X, \mathcal{F}, \mu)$ is said to be *ergodic* if for any measurable set $A$

$$\mu(T^{-1}(A) \div A) = 0 \quad \Rightarrow \quad \mu(A) = 0 \text{ or } \mu(X \setminus A) = 0$$

(Recall the notation $B \div C = (B \setminus C) \cup (C \setminus B)$.)

Note that we did not assume in the definition of ergodicity that $\mu$ is $T$-invariant (neither that $\mu$ is finite). Suppose that for every $E$ of measure 0 the set $T^{-1}(E)$ is also of measure 0 (in Ch.4 we call this property of $\mu$ with respect to $T$, *backward quasi-invariant*). Then in the definition of ergodicity one can replace $\mu(T^{-1}(A) \div A) = 0$ by $T^{-1}(A) = A$. Indeed having $A$ as in the definition one can define $A' = \bigcup_{n=0}^{\infty} \bigcap_{m=n}^{\infty} T^{-m}(A)$. Then $\mu(A') = \mu(A)$ and $T^{-1}(A') = A'$. If we assumed that the latter implies $\mu(A') = 0$ or $\mu(X \setminus A') = 0$, then $\mu(A) = 0$ or $\mu(X \setminus A) = 0$.

Let $\phi : X \to I\!R$ be a measurable function. For any $n \ge 1$ we define

$$S_n\phi = \phi + \phi \circ T + \ldots + \phi \circ T^{n-1} \tag{1.2.1}$$

Let $\mathcal{I} = \{A \in \mathcal{F} : \mu(T^{-1}(A) \div A) = 0\}$. We call it $\sigma$-algebra of $T$-invariant (mod 0) sets. Note that every $\psi : X \to I\!R$, measurable with respect to $\mathcal{I}$, is $T$-invariant (mod 0), namely $\psi \circ T = \psi$, but on a set of measure $\mu$ equal to 0. Indeed let $A = \{x \in X : \psi(x) \ne \psi \circ T(x)$ and suppose $\mu(A) > 0$. Hence there exists $a \in I\!R$ such that $A_a = \{x \in A : \psi(x) < a, \psi \circ T > a\}$ and $\mu(A_a) > 0$. (or a similar $A_a$ with reversed inequalities). Since $A_a \in \mathcal{I}$, there exists $E \subset A$ of measure 0, such that $T(A_a \setminus E) \subset A_a$, hence on $A_a \setminus E$ we have $\psi = \psi \circ T$. We arrived at a contradiction.

**Theorem 1.2.2.** (Birkhoff's Ergodic Theorem) If $T : X \to X$ is a measure preserving endomorphism of a probability space $(X, \mathcal{F}, \mu)$ and $\phi : X \to I\!R$ is an integrable function then

$$\lim_{n\to\infty} \frac{1}{n} S_n\phi(x) = E(\phi|\mathcal{I}) \text{ for } \mu\text{-a.e. } x \in X$$

We say that the *time average* exists for $\mu$-almost every $x \in X$.

In particular Theorem 1.2.2 yields for $T$ ergodic preserving $\mu$, that

$$\lim_{n\to\infty} \frac{1}{n} S_n\phi(x) = \int \phi \, d\mu, \quad \text{for } \mu\text{-a.e. } x \tag{1.2.1a}$$

We say that the *time average equals the space average.*

If $\phi = 1\!\!1_A$, the indicator function of a measurable set $A$, then we deduce that for a.e. $x$ the frequency of hitting $A$ by the forward trajectory equals to the measure (probability) of $A$, namely $\lim_{n\to\infty} \#\{0 \le j < n : T^j(x) \in A\}/n$, is equal to $\mu(A)$.
This means for example that if we choose a point in $X$ in a euclidean space at random its sufficiently long forward trajectory fills $X$ with the density being approximately the density of $\mu$ with respect to the Lebesgue measure, provided $\mu$ is equivalent to the Lebesgue measure.

On the figure below, Fig.1.2a, for a randomly chosen backward trajectory $x_j, j = 0, 1, ..., n,$, $T(x_k) = x_{k-1}$, for $T(x) = 2x^2 - 1$ (see Example 0.9), for the interval $[-1, 1]$ divided into $k = 100$ equal pieces, the graph of the function $-1 + 2t/100 \mapsto 100 \cdot \#\{0 \le j < n : -1 + 2t/100 \le x_j(x) < -1 + 2(t + 1)/100\}/n$ is plotted. It indeed resembles the graph of $1/\pi\sqrt{1 - x^2}$, Fig.1.2b, which is the density of the invariant probability measure equivalent to the length measure.

FIGURE 1.1, The density of an invariant measure for $T(x) = 2x^2 - 1$.

As a corollary of Birkhoff's Ergodic Theorem one can obtain von Neumann's Ergodic Theorem. It says that if $\phi \in L^p(\mu)$ for $1 \le p < \infty$, then the convergence to $E(\phi|\mathcal{I})$ holds in $L^p$. It is not difficult, see for example [Wa].

**Proof of Birkhoff's Ergodic Theorem.** Let $f \in L^1(\mu)$ and $F_n = \max\{\sum_{i=0}^{k-1} f \circ T^i : 1 \le k \le n\}$. Then for every $x \in X$, $F_{n+1}(x) - F_n(T(x)) = f(x) - \min(0, F_n(T(x))) \ge f(x)$ and is monotonously decreasing. Define

$$A = \{x : \sup_n \sum_{i=0}^{n} f(T^i(X)) = \infty\}$$

If $x \in A$ then $F_{n+1}(x) - F_n(T(x))$ monotonously decreases to $f(x)$ as $n \to \infty$. The Dominated Convergence Theorem then implies that

$$0 \le \int_A (F_{n+1} - F_n)\, d\mu = \int_A (F_{n+1} - F_n \circ T)\, d\mu \to \int f d\mu. \tag{1.2.2}$$

(We arrived at $\int_A f\, d\mu \ge 0$, which is a variant of so-called Maximal Ergodic Theorem.) Notice that $\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k \le F_n/n$, so outside $A$ we have

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k \le 0. \tag{1.2.3}$$

Therefore, if the conditional expectation value $f_{\mathcal{I}}$ of $f$ is negative a.e., that is if $\int_C f d\mu = \int_C f_{\mathcal{I}} d\mu < 0$ for all $C \in \mathcal{I}$ with $\mu(C) > 0$, then, since by definition $A \in \mathcal{I}$, (1.2.2) implies that $\mu(A) = 0$, and hence (1.2.3) holds a.e.. Now if we let $f = \phi - \phi_{\mathcal{I}} - \varepsilon$, then $f_{\mathcal{I}} = -\varepsilon < 0$. Note that $\phi_{\mathcal{I}} \circ T = \phi_{\mathcal{I}}$ implies

$$\frac{1}{n}\sum_{k=0}^{n-1} f \circ T^k = \Big(\frac{1}{n}\sum_{k=0}^{n-1} \phi \circ T^k\Big) - \phi_{\mathcal{I}} - \varepsilon.$$

So (1.2.3) yields

$$\limsup_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi\circ T^k \le \phi_{\mathcal{I}} + \varepsilon \quad \text{a.e.}$$

Replacing $\phi$ by $-\phi$ gives

$$\liminf_{n\to\infty} \frac{1}{n}\sum_{k=0}^{n-1} \phi\circ T^k \le \phi_{\mathcal{I}} - \varepsilon \quad \text{a.e.}$$

Thus $\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}\phi\circ T^k = \phi_{\mathcal{I}}$ a.e. ♣

Recall that on the opposite end to the absolute continuity (see Sec.1) there is the notion of singularity. Finite measures $\mu_1$ and $\mu_2$ on a $\sigma$-algebra $\mathcal{F}$ are called mutually *singular*, $\mu_1 \perp \mu_2$ if there exist disjoint sets $X_1, X_2 \in \mathcal{F}$ with $\mu_i(X_i) = 1$ for $i = 1, 2$.

**Theorem 1.2.3.** If $T : X \to X$ is a map measurable with respect to a $\sigma$-algebra $\mathcal{F}$ and if $\mu_1$ and $\mu_2$ are two different $T$-invariant probability ergodic measures on $\mathcal{F}$, then $\mu_1$ and $\mu_2$ are singular.

**Proof.** Since $\mu_1$ and $\mu_2$ are different, there exists a measurable set $A$ such that

$$\mu_1(A) \ne \mu_2(A) \tag{1.2.2}$$

By Theorem 1.2.2 (Birkhoff's Ergodic Theorem) applied to $\mu_1$ and $\mu_2$ there exist sets $X_1, X_2 \in \mathcal{F}$ such that for every $i = 1, 2$ and every $x \in X_i$

$$\lim_{n\to\infty} \frac{1}{n} S_n \mathbb{1}_A(x) = \mu_i(A)$$

and $\mu_i(X_i) = 1$. Thus in view of (1.2.2) the sets $X_1$ and $X_2$ are disjoint. The proof is finished. ♣

**Proposition 1.2.4.** If $T : X \to X$ is a measure preserving endomorphism of a probability space $(X, \mathcal{F}, \nu)$, then $\nu$ is ergodic if and only if there is no $T$-invariant probability measure on $\mathcal{F}$ absolutely continuous with respect to $\nu$ and distinct from $\nu$.

**Proof.** Suppose that $\nu$ is ergodic and $\mu$ is a $T$-invariant probability measure on $\mathcal{F}$ with $\mu \ll \nu$. Then $\mu$ is also ergodic. Otherwise $\mu(A), \mu(X \setminus A) > 0$ so $\nu(A), \nu(X \setminus A) > 0$ so $\nu$ would not be ergodic. Hence by Theorem 1.2.3 $\mu = \nu$.

Suppose in turn that $\nu$ is not ergodic and let $A \in \mathcal{F}$ be a $T$-invariant set such that $0 < \nu(A) < 1$. Then the conditional measure on $A$ is also $T$-invariant but simultaneously it is distinct from $\nu$ and absolutely continuous with respect to $\nu$. The proof is finished. ♣

Observe now that the space $M(\mathcal{F})$ of probability measures on $\mathcal{F}$ is a convex set i.e. the convex combination $\alpha\mu + (1-\alpha)\nu$, $0 \le \alpha \le 1$, of two such measures is again in $M(\mathcal{F})$. The subspace $M(\mathcal{F}, T)$ of $M(\mathcal{F})$ consisting of $T$-invariant measures is also convex.

Recall that a point in a convex set is said to be *extreme* if and only if it cannot be represented as a convex combination of two distinct points with corresponding coefficient $0 < \alpha < 1$. We shall prove the following.

**Theorem 1.2.5.** The ergodic measures in $M(\mathcal{F}, T)$ are exactly the extreme points of $M(\mathcal{F}, T)$.

**Proof.** Suppose that $\mu, \mu_1, \mu_2 \in M(\mathcal{F}, T)$, $\mu_1 \neq \mu_{,2}$ and $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$ with $0 < \alpha < 1$. Then $\mu_1 \neq \mu$ and $\mu_1 \ll \mu$. Thus in view of Proposition 1.2.4 mesure $\mu$ is not ergodic.

Suppose in turn that $\mu$ is not ergodic and let $A \in \mathcal{F}$ be a $T$-invariant set such that $0 < \mu(A) < 1$. Recall that given $B \in \mathcal{F}$ with $\mu(B) > 0$ the *conditional measure* $A \to \mu(A|B)$ is defined by $\mu(A \cap B)/\mu(B)$. Thus the conditional measures $\mu(\cdot|A)$ and $\mu(\cdot|A^c)$ are distinct, $T$-invariant and $\mu = \mu(A)\mu(\cdot|A) + (1 - \mu(A)\mu(\cdot|A^c)$. Consequently $\mu$ is not en extreme point in $M(\mathcal{F}, T)$. The proof is finished. ♣

In Section 8 we shall formulate a theorem on decomposition into ergodic components, that will better clear the situation. This will correspond the Choquet Theorem in functional analysis, see Ch.2.1.

## §1.3. ENTROPY OF PARTITION

Let $(X, \mathcal{F}, \mu)$ be a probability space. A *partition* of $(X, \mathcal{F}, \mu)$ is a subfamily (*a priori* may be uncountable) of $\mathcal{F}$ consisting of mutually disjoint elements whose union is $X$.
If $\mathcal{A}$ is a partition and $x \in X$ then the only element of $\mathcal{A}$ containing $x$ is denoted by $\mathcal{A}(x)$ or, if $x \in A \in \mathcal{A}$, by $A(x)$.

If $\mathcal{A}$ and $\mathcal{B}$ are two partitions of $X$ we define their *join*

$$\mathcal{A} \vee \mathcal{B} = \{A \cap B : A \in \mathcal{A},\ B \in \mathcal{B}\}$$

We write $\mathcal{A} \leq \mathcal{B}$ if and only if $\mathcal{B}(x) \subset \mathcal{A}(x)$ for every $x \in X$, which in other words means that each element of the partition $\mathcal{B}$ is contained in an element of the partition $\mathcal{A}$ or equivalently $\mathcal{A} \vee \mathcal{B} = \mathcal{B}$. We sometimes say in this case, that $\mathcal{B}$ is *finer* than $\mathcal{A}$ or that $\mathcal{B}$ is a *refinement* of $\mathcal{A}$.

Now we introduce the notion of entropy of a countable (this word includes in this book: finite) partition and we collect its basic elementary properties. Define the function $k : [0, 1] \to [0, \infty]$ putting

$$k(t) = \begin{cases} -t \log t & \text{for } t \in (0, 1] \\ 0 & \text{for } t = 0 \end{cases} \tag{1.3.1}$$

Check that the function $k$ is continuous. Let $\mathcal{A} = \{A_i : 1 \leq i \leq n\}$ be a countable partition of $X$, where $n$ is a finite integer or $\infty$. In the sequel we shall usually write $\infty$.

The *entropy* of $\mathcal{A}$ is the number

$$\mathrm{H}(\mathcal{A}) = \sum_{i=1}^{\infty} -\mu(A_i)\log\mu(A_i) = \sum_{i=1}^{\infty} k(\mu(A_i)) \tag{1.3.2}$$

If $\mathcal{A}$ is infinite, $\mathrm{H}(\mathcal{A})$ may happen to be infinite too.

Define $I(x) = I(\mathcal{A})(x) := -\log\mu(\mathcal{A}(x))$. This is called an information function. Intuitively $I(x)$ is an information on an object $x$ given by the experiment $\mathcal{A}$ in the logarithmic scale. Therefore the entropy in (1.3.2) is the integral (the average) of the information function.

Note that $\mathrm{H}(\mathcal{A}) = 0$ for $\mathcal{A} = \{X\}$ and that if $\mathcal{A}$ is finite, say consists of $n$ elements, then $0 \le \mathrm{H}(\mathcal{A}) \le \log n$ and $\mathrm{H}(\mathcal{A}) = \log n$ if and only if $\mu(A_1) = \mu(A_2) = \ldots = \mu(A_n) = 1/n$. This follows from the fact that the logarithmic function is strictly concave.

In this section we deal with only one fixed measure $\mu$. If however we need to consider more measures simultaneously (see for example Ch.2) we will rather use the notation $\mathrm{H}_\mu(A)$ for $\mathrm{H}(A)$.

Let $\mathcal{A} = \{A_i : i \ge 1\}$ and $\mathcal{B} = \{B_j : j \ge 1\}$ be two countable partitions of $X$. The *conditional entropy* $\mathrm{H}(\mathcal{A}|\mathcal{B})$ of $\mathcal{A}$ given $\mathcal{B}$ is defined as

$$\begin{aligned}\mathrm{H}(\mathcal{A}|\mathcal{B}) &= \sum_{j=1}^{\infty}\mu(B_j)\sum_{i=1}^{\infty} -\frac{\mu(A_i\cap B_j)}{\mu(B_j)}\log\frac{\mu(A_i\cap B_j)}{\mu(B_j)} \\ &= \sum_{i,j} -\mu(A_i\cap B_j)\log\frac{\mu(A_i\cap B_j)}{\mu(B_j)} \end{aligned}\tag{1.3.3}$$

The first equality, defining $\mathrm{H}(\mathcal{A}|\mathcal{B})$, can be viewed as follows: one considers each element $B_j$ as a probability space with conditional measure $\mu(A|B_j) = \mu(A)/\mu(B_j)$ for $A \in B_j$ and calculates the entropy of the partition of the set $B_j$ into $A_i \cap B_j$. Then one averages the result over the space of $B_j$'s. (This will be generalized in Def.1.8.3.)

For each $x$ denote $-\log\mu((\mathcal{A}(x)\cap\mathcal{B}(x))|\mathcal{B}(x))$ by $I(x)$ or $I(\mathcal{A}|\mathcal{B})(x)$. The second equality in (1.3.3) can be rewritten as

$$\mathrm{H}(\mathcal{A}|\mathcal{B}) = \int_X I(\mathcal{A}|\mathcal{B})\,d\mu. \tag{1.3.3a}$$

Note by the way that if $\tilde{\mathcal{B}}$ is the $\sigma$-algebra consisting of all unions of elements of $\mathcal{B}$ (i.e. generated by $\mathcal{B}$, then $I(x) = -\log\mu((\mathcal{A}(x)\cap\mathcal{B}(x))|\mathcal{B}(x)) = -\log E(1\!\!1_{\mathcal{A}(x)}|\tilde{\mathcal{B}})$, cf (1.1.8).

Note finally that for any countable partition $\mathcal{A}$ we have

$$\mathrm{H}(\mathcal{A}|\{X\}) = \mathrm{H}(\mathcal{A}). \tag{1.3.4}$$

.

Some futher basic properties of entropy of partitions are collected in the following.

**Theorem 1.3.1.** Let $(X, \mathcal{F}, \mu)$ be a probability space. If $\mathcal{A}$, $\mathcal{B}$ and $\mathcal{C}$ are countable partitions of $X$ then:

(a) $\quad \mathrm{H}(\mathcal{A} \vee \mathcal{B}|\mathcal{C}) = \mathrm{H}(\mathcal{A}|\mathcal{C}) + \mathrm{H}(\mathcal{B}|\mathcal{A} \vee \mathcal{C})$

(b) $\quad \mathrm{H}(\mathcal{A} \vee \mathcal{B}) = \mathrm{H}(\mathcal{A}) + \mathrm{H}(\mathcal{B}|\mathcal{A})$

(c) $\quad \mathcal{A} \le \mathcal{B} \;\Rightarrow\; \mathrm{H}(\mathcal{A}|\mathcal{C}) \le \mathrm{H}(\mathcal{B}|\mathcal{C})$

(d) $\quad \mathcal{B} \le \mathcal{C} \;\Rightarrow\; \mathrm{H}(\mathcal{A}|\mathcal{B}) \ge \mathrm{H}(\mathcal{A}|\mathcal{C})$

(e) $\quad \mathrm{H}(\mathcal{A} \vee \mathcal{B}|\mathcal{C}) \le \mathrm{H}(\mathcal{A}|\mathcal{C}) + \mathrm{H}(\mathcal{B}|\mathcal{C})$

(f) $\quad \mathrm{H}(\mathcal{A}|\mathcal{C}) \le \mathrm{H}(\mathcal{A}|\mathcal{B}) + \mathrm{H}(\mathcal{B}|\mathcal{C})$

**Proof.** Let $\mathcal{A} = \{A_n : n \ge 1\}$, $\mathcal{B} = \{B_m : m \ge 1\}$, and $\mathcal{C} = \{C_l : l \ge 1\}$. Without loosing generality we can assume that all these sets are of positive measure.
(a) By (1.3.3) we have

$$\mathrm{H}(\mathcal{A} \vee \mathcal{B}|\mathcal{C}) = -\sum_{i,j,k} \mu(A_i \cap B_j \cap C_k) \log \frac{\mu(A_i \cap B_j \cap C_k)}{\mu(C_k)}$$

But

$$\frac{\mu(A_i \cap B_j \cap C_k)}{\mu(C_k)} = \frac{\mu(A_i \cap B_j \cap C_k)}{\mu(A_i \cap C_k)} \frac{\mu(A_i \cap C_k)}{\mu(C_k)}$$

unless $\mu(A_i \cap C_k) = 0$. But then the left hand side vanishes and we need not consider it. Therefore

$$\begin{aligned}
\mathrm{H}(\mathcal{A} \vee \mathcal{B}|\mathcal{C}) &= -\sum_{i,j,k} \mu(A_i \cap B_j \cap C_k) \log \frac{\mu(A_i \cap C_k)}{\mu(C_k)} \\
&\quad -\sum_{i,j,k} \mu(A_i \cap B_j \cap C_k) \log \frac{\mu(A_i \cap B_j \cap C_k)}{\mu(A_i \cap C_k)} \\
&= -\sum_{i,k} \mu(A_i \cap C_k) \log \frac{\mu(A_i \cap C_k)}{\mu(C_k)} + \mathrm{H}(\mathcal{B}|\mathcal{A} \vee \mathcal{C}) \\
&= \mathrm{H}(\mathcal{A}|\mathcal{C}) + \mathrm{H}(\mathcal{B}|\mathcal{A} \vee \mathcal{C})
\end{aligned}$$

(b) Put $\mathcal{C} = \{X\}$ and apply (1.3.4) in (a).
(c) By (a)

$$\mathrm{H}(\mathcal{B}|\mathcal{C}) = \mathrm{H}(\mathcal{A} \vee \mathcal{B}|\mathcal{C}) = \mathrm{H}(\mathcal{A}|\mathcal{C}) + \mathrm{H}(\mathcal{B}|\mathcal{A} \vee \mathcal{C}) \ge \mathrm{H}(\mathcal{A}|\mathcal{C})$$

(d) Since the function $k$ defined by (1.3.1) is strictly concave, we have for every pair $i, j$

$$k\left(\sum_l \frac{\mu(C_l \cap B_j)}{\mu(B_j)} \frac{\mu(A_i \cap C_l)}{\mu(C_l)}\right) \ge \sum_l \frac{\mu(C_l \cap B_j)}{\mu(B_j)} k\left(\frac{\mu(A_i \cap C_l)}{\mu(C_l)}\right) \tag{1.3.5}$$

But since $\mathcal{B} \leq \mathcal{C}$, we can write above $C_l \cap B_j = C_l$, hence the left hand side equals

$$k\left(\frac{\mu(A_i \cap B_j)}{\mu(B_j)}\right) = -\frac{\mu(A_i \cap B_j)}{\mu(B_j)} \log \frac{\mu(A_i \cap B_j)}{\mu(B_j)}$$

Thus multiplying both sides of (1.3.5) by $\mu(B_j)$ and summing over $i$ and $j$ we get

$$\begin{aligned}
-\sum_{i,j} \mu(A_i \cap B_j) \log \frac{\mu(A_i \cap B_j)}{\mu(B_j)} &\geq -\sum_{i,j,l} \mu(C_l \cap B_j) \frac{\mu(A_i \cap C_l)}{\mu(C_l)} \log \frac{\mu(A_i \cap C_l)}{\mu(C_l)} \\
&= -\sum_{i,l} \mu(C_l) \frac{\mu(A_i \cap C_l)}{\mu(C_l)} \log \frac{\mu(A_i \cap C_l)}{\mu(C_l)}
\end{aligned}$$

or equivalently $\mathrm{H}(\mathcal{A}|\mathcal{B}) \geq \mathrm{H}(\mathcal{A}|\mathcal{C})$.

Formula (e) follows immediately from (a) and (d) and formula (f) can proved by a straightforward calculation (its consequences are discussed in Exercise 1.9). ♣

## §1.4. ENTROPY OF ENDOMORPHISM.

Let $(X, \mathcal{F}, \mu)$ be a probability space and let $T : X \to X$ be a measure preserving endomorphism of $X$. If $\mathcal{A} = \{A_i\}_{i \in I}$ is a partition of $X$ then by $T^{-1}\mathcal{A}$ we denote the partition $\{T^{-1}(A_i)\}_{i \in I}$. Note that for any countable $\mathcal{A}$

$$\mathrm{H}(T^{-1}\mathcal{A}) = \mathrm{H}(\mathcal{A}) \tag{1.4.1}$$

Later on we shall consider partitions $\bigvee_{i=0}^{n} T^{-i}\mathcal{A} = \mathcal{A} \vee T^{-1}(\mathcal{A}) \vee ... \vee T^{-n}(\mathcal{A})$. We shall sometimes use the notation $\mathcal{A}^n$. We shall write $\mathcal{A}_m^n$ for $\bigvee_{i=m}^{n} T^{-i}\mathcal{A}$).

We shall define the entropy of $T$ with respect to a countable (in particular finite) partition of finite entropy We start with the following well known elementary fact.

**Lemma 1.4.1.** If $\{a_n\}_{n=1}^{\infty}$ is a sequence of real numbers such that $a_{n+m} \leq a_n + a_m$ for all $n, m \geq 1$ then $\lim_{n\to\infty} a_n$ exists and equals $\inf_n a_n/n$. The limit could be $-\infty$, but if the $a_n$'s are bounded below, then the limit will be nonnegative.

**Proof.** Fix $m \geq 1$. Each $n \geq 1$ can be expressed as $n = km + i$ with $0 \leq i < m$. Then

$$\frac{a_n}{n} = \frac{a_{i+km}}{i+km} \leq \frac{a_i}{km} + \frac{a_{km}}{km} \leq \frac{a_i}{km} + \frac{ka_m}{km} = \frac{a_i}{km} + \frac{a_m}{m}$$

If $n \to \infty$ then also $k \to \infty$ and therefore $\limsup_{n\to\infty} \frac{a_n}{n} \leq \frac{a_m}{m}$. Thus $\limsup_{n\to\infty} \frac{a_n}{n} \leq \inf \frac{a_m}{m}$. Now the inequality $\inf \frac{a_m}{m} \leq \liminf_{n\to\infty} \frac{a_n}{n}$ finishes the proof. ♣.

Notice that there exists a subadditive sequence (i.e. satisfying $a_{n+m} \leq a_n + a_m$) such that the corresponding sequence $a_n/n$ is not eventually decreasing. Indeed, it suffices to observe that each sequence consisting of 1's and 2's is subadditive and to consider such a sequence having infinitely many 1's and 2's. If for an $n > 1$ we have $a_n = 1$ and $a_{n+1} = 2$ we have $\frac{a_n}{n} < \frac{a_{n+1}}{n+1}$.

**Exercise.** Prove that Lemma 1.4.1 remains true under the weaker assumptions that there exists $c \in R$ such that $a_{n+m} \le a_n + a_m + c$ for all $n$ and $m$.

**Lemma 1.4.2.** If $\mathcal{A}$ is a countable partition of $X$, then $\lim_{n\to\infty} \frac{1}{n}\mathrm{H}(\mathcal{A}^{n-1})$ exists.
**Proof.** Let $a_n = \mathrm{H}(\mathcal{A}^{n-1})$. Then for all $n, m \ge 1$, with the use of Theorem 1.4.1(e) and (1.4.1) we estimate

$$a_{n+m} = \mathrm{H}(\mathcal{A}^{n+m-1}) \le \mathrm{H}(\mathcal{A}^{n-1}) + \mathrm{H}(\mathcal{A}_n^{n+m-1}) = a_n + \mathrm{H}(\mathcal{A}^{m-1}) = a_n + a_m.$$

and apply Lemma 1.4.1. ♣

The limit whose existence has been shown in Lemma 1.4.2. is known as the *(measure–theoretic) entropy of $T$ with respect to the partition $\mathcal{A}$* and is denoted by $\mathrm{h}(T,\mathcal{A})$ or by $\mathrm{h}_\mu(T,\mathcal{A})$ if one wants to indicate the measure under consideration. Intuitively this means the limit rate of the growth of average (integral) information (in logarithmic scale), under consecutive experiments, for number of experiments tending to infinity. In the sequel we will need the following result improving slightly Lemma 1.4.2.

**Lemma 1.4.3.** The sequences $\frac{1}{n+1}\mathrm{H}(\mathcal{A}^n)$ and $\mathrm{H}\big(\mathcal{A}|\mathcal{A}_1^n\big)$ are monotone decreasing to $\mathrm{h}(T,\mathcal{A})$.
**Proof.** We first show by induction that

$$\mathrm{H}(\mathcal{A}^n) = \mathrm{H}(\mathcal{A}) + \sum_{j=1}^{n} \mathrm{H}(\mathcal{A}|\mathcal{A}_1^j) \tag{1.4.2}$$

If $n = 0$ it is tautology. Suppose it is true for $n-1 \ge 0$. Then with the use of Theorem 1.4.1(b) and (1.4.1) we obtain

$$\mathrm{H}(\mathcal{A}^n) = \mathrm{H}(\mathcal{A}_1^n \vee \mathcal{A}) = \mathrm{H}(\mathcal{A}_1^n) + \mathrm{H}(\mathcal{A}|\mathcal{A}_1^n) = \mathrm{H}(\mathcal{A}^{n-1}) + \mathrm{H}(\mathcal{A}|\mathcal{A}_1^n) = \mathrm{H}(\mathcal{A}) + \sum_{j=1}^{n} \mathrm{H}(\mathcal{A}|\mathcal{A}_1^j)$$

by the inductive assumption. Hence (1.4.2) holds for all $n$. Since, by Theorem 1.3.1 (d), the sequence
$\mathrm{H}\big(\mathcal{A}|\mathcal{A}_1^n\big)$, $n = 0, 1, \dots$ is monotone decreasing, the sequence of averages is also monotone decreasing to the same limit, furthermore it coincides with the sequence $\frac{1}{n+1}\mathrm{H}(\mathcal{A}^n)$ by (1.4.2). ♣

The basic elementary properties of the entropy $\mathrm{h}(T,\mathcal{A})$ are collected in the next theorem below. The standard proof (see for example [Wa]) based on Theorem 1.3.1 and formula (1.3.2) is left for the reader as an exercise.

**Theorem 1.4.4.** If $\mathcal{A}$ and $\mathcal{B}$ are countable partitions of finite entropy then

(a) $\mathrm{h}(T, \mathcal{A}) \le \mathrm{H}(\mathcal{A})$

(b) $\mathrm{h}(T, \mathcal{A} \vee \mathcal{B}) \le \mathrm{h}(T, \mathcal{A}) + \mathrm{h}(T, \mathcal{B})$

(c) $\mathcal{A} \le \mathcal{B} \;\Rightarrow\; \mathrm{h}(T, \mathcal{A}) \le \mathrm{h}(T, \mathcal{B})$

(d) $\mathrm{h}(T, \mathcal{A}) \le \mathrm{h}(T, \mathcal{B}) + \mathrm{H}(\mathcal{A}|\mathcal{B})$

(e) $\mathrm{h}(T, T^{-1}(\mathcal{A})) = \mathrm{h}(T, \mathcal{A})$

(f) If $k \ge 1$ then $\mathrm{h}(T, \mathcal{A}) = \mathrm{h}\big(T, \mathcal{A}^k\big)$

(g) If $T$ is invertible and $k \ge 1$ then $\mathrm{h}(T, \mathcal{A}) = \mathrm{h}\big(T, \bigvee_{i=-k}^{k} T^i(\mathcal{A})\big)$

Here is one more useful fact, stronger than Th.1.4.4 (c):

**Theorem 1.4.5.** If $T : X \to X$ is a measure preserving endomorphism of a probability space $(X, \mathcal{F}, \mu)$ and $\mathcal{A}$ and $\mathcal{B}_m, m = 1, 2, ...$ are countable partitions of finite entropy, and $\mathrm{H}(\mathcal{A}|\mathcal{B}_m) \to 0$ as $m \to \infty$, then $\mathrm{h}(T, \mathcal{A}) \le \liminf_{m\to\infty} \mathrm{h}(T, \mathcal{B}_m)$. In particular, for $\mathcal{B}_m := \mathcal{B}^m = \bigvee_{j=0}^{m} T^{-j}(\mathcal{B})$, one obtains $\mathrm{h}(T, \mathcal{A}) \le \mathrm{h}(T, \mathcal{B})$.

**Proof.** For every positive integer $m$,

$$\mathrm{h}(T, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n}\mathrm{H}(\mathcal{A}^{n-1}) \le \lim_{n\to\infty} \frac{1}{n}\Big(\mathrm{H}(\mathcal{A}^{n-1}|(\mathcal{B}_m)^{n-1}) + \mathrm{H}((\mathcal{B}_m)^{n-1})\Big)$$

$$\le \mathrm{H}(\mathcal{A}|\mathcal{B}_m) + \lim_{n\to\infty} \frac{1}{n}\mathrm{H}((\mathcal{B}_m)^{n-1}) = \mathrm{H}(\mathcal{A}|\mathcal{B}_m) + \mathrm{h}(T, \mathcal{B}_m).$$

Letting $m \to \infty$ this yields the first part of the assertion. For $\mathcal{B}_m = \mathcal{B}^m$, one can substitute in place of the last summand $\mathrm{h}(T, \mathcal{B}^m) = \mathrm{h}(T, \mathcal{B})$, by Theorem 1.4.4(f). ♣

The *(measure-theoretic) entropy of the endomorphism* $T : X \to X$ is defined as

(1.4.3) $$\mathrm{h}_\mu(T) = \mathrm{h}(T) = \sup_{\mathcal{A}}\{\mathrm{h}(T, \mathcal{A})\}$$

where the supremum is taken over all finite (or countable of finite entropy) partitions of $X$. See Exercise 12.

It is clear from the definition that the entropy of $T$ is an isomorphism invariant.

Later on (see Th.1.8.7, Remark 1.8.7", Corollary 1.8"' and Exercise 1.9') we shall discuss the cases where $\mathrm{H}(\mathcal{A}|\mathcal{B}_n) \to 0$ for every $\mathcal{A}$ (finite or of finite entropy). This will allow us to write $\mathrm{h}_\mu(T) = \lim_{m\to\infty} \mathrm{h}(T, \mathcal{B}_m)$ or $\mathrm{h}(T) = \mathrm{h}(T, \mathcal{B})$.

Let us end this Section with the following useful

**Theorem 1.4.6.** If $T : X \to X$ is a measure preserving endomorphism of a probability space $(X, \mathcal{F}, \mu)$ then

(a) $$\mathrm{h}(T^k) = k\mathrm{h}(T) \text{ for all } k \geq 1$$

(b) $$\text{If } T \text{ is invertible then } \mathrm{h}(T^{-1}) = \mathrm{h}(T)$$

**Proof.** (a) Fix $k \geq 1$. Since

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{H}\Big(\bigvee_{j=0}^{n-1} T^{-kj}\Big(\bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}\Big)\Big) = \lim_{n\to\infty} \frac{k}{nk}\mathrm{H}\Big(\bigvee_{i=0}^{nk-1} T^{-i}\mathcal{A}\Big) = k\mathrm{h}(T, \mathcal{A})$$

we have $\mathrm{h}\big(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}\big) = k\mathrm{h}(T, \mathcal{A})$. Therefore

$$k\mathrm{h}(T) = k \sup_{\mathcal{A} \text{ finite}} \mathrm{h}(T, \mathcal{A}) = \sup_{\mathcal{A}} \mathrm{h}\big(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}\big) \leq \sup_{\mathcal{B}} \mathrm{h}(T^k, \mathcal{B}) = \mathrm{h}(T^k) \tag{1.4.4}$$

On the other hand by Theorem 1.4.4(c) we get $\mathrm{h}(T^k, \mathcal{A}) \leq \mathrm{h}\big(T^k, \bigvee_{i=0}^{k-1} T^{-i}\mathcal{A}\big) = k\mathrm{h}(T, \mathcal{A})$ and therefore $\mathrm{h}(T^k) \leq k\mathrm{h}(T)$. The result follows from this and (1.4.4).

(b) In view of (1.4.1) for all finite partitions $\mathcal{A}$ we have

$$\mathrm{H}\Big(\bigvee_{i=0}^{n-1} T^{i}\mathcal{A}\Big) = \mathrm{H}\Big(T^{-(n-1)} \bigvee_{i=0}^{n-1} T^{i}\mathcal{A}\Big) = \mathrm{H}\Big(\bigvee_{i=0}^{n-1} T^{-i}\mathcal{A}\Big)$$

This finishes the proof. ♣

## §1.5. SHANNON-MCMILLAN-BREIMAN THEOREM.

Let $(X, \mathcal{F}, \mu)$ be a probability space, $T : X \to X$ be a measure preserving endomorphism of $X$ and $\mathcal{A}$ be a countable finite entropy partition of $X$.

**Lemma 1.5.1.** For each $n = 1, 2, ...$ let $f_n = I(\mathcal{A}|\mathcal{A}_1^n)$ and $f^* = \sup_{n\geq 1} f_n$. Then for each $\lambda$ and each $A \in \mathcal{A}$

$$\mu\{x \in A : f^*(x) > \lambda\} \leq e^{-\lambda}.$$

**Proof.** For each $A \in \mathcal{A}$ and $n = 1, 2, ...$ let $f_n^A = -\log E(1\!\!1_A|\mathcal{A}_1^n)$ and

$$B_n^A = \{x : f_1^A(x), ..., f_{n-1}^A(x) \leq \lambda, f_n(x) > \lambda\}.$$

Of course $f_n = \sum_{A\in\mathcal{A}} 1\!\!1_A f_n^A$.

Since $B_n^A \in \mathcal{F}(\mathcal{A}_1^n)$, the $sg$-algebra generated by $\mathcal{A}_1^n$,

$$\mu(B_n^A \cap A) = \int_{B_n^A} 1\!\!1_A \, d\mu = \int_{B_n^A} E(1\!\!1_A|\mathcal{A}_1^n) =$$

$$\int_{B_n^A} e^{-f_n^A}\, d\mu \le e^{-\lambda}\mu(B_n^A).$$

Therefore

$$\mu\{x \in A : f^*(x) > \lambda\} = \sum_{n=1}^{\infty} \mu(B_n^A \cap A) \le e^{-\lambda} \sum_{n=1}^{\infty} \mu(B_n^A) \le e^{-\lambda}.\clubsuit$$

**Corollary 1.5.2** The function $f^*$ is integrable.

**Proof.** Of course $\mu\{x \in A : f^* > \lambda\} \le \mu(A)$, so $\mu\{x \in A : f^* > \lambda\} \le \min\{\mu(A), e^{-\lambda}\}$. So

$$\int_X f^*\, d\mu = \int_0^\infty \mu\{f^* > \lambda\}\, d\lambda = \sum_A \in \mathcal{A} \int_0^\infty \mu\{x \in A : f^* > \lambda\}\, d\lambda$$

$$\le \sum_A \in \mathcal{A} \int_0^\infty \min\{\mu(A), e^{-\lambda}\}\, d\lambda = \sum_A \in \mathcal{A} \Big( \int_0^{-\log\mu(A)} \mu(A)\, d\lambda + \int_{-\log\mu(A)}^\infty e^{-\lambda}\, d\lambda \Big)$$

$$= \sum_A \in \mathcal{A} \Big( -\mu(A)(\log\mu(A)) + \mu(A) \Big) = \mathrm{H}(\mathcal{A}) + 1.\clubsuit$$

**Corollary 1.5.3.** $f_n$ converge a.e. and in $L^1$.

**Proof.** $E(\mathbb{1}_A|\mathcal{A}_1^n)$ is a martingale to which we can apply Theorem 1.1.1. This gives convergence a.e., hence convergence a.e of each $f_n^A$, hence $f_n$. Now convergence in $L^1$ follows from Corollary 1.5.2. and Dominated Convergence Theorem $\clubsuit$

**Theorem 1.5.4.** (Shannon-McMillan-Breiman) Suppose that $\mathcal{A}$ is a countable partition of finite entropy. Then there exist limits

$$f = \lim_{n\to\infty} I(\mathcal{A}|\mathcal{A}_1^n) \text{ and } f_{\mathcal{I}}(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)). \text{ for a.e. } x.$$

and

$$\lim_{n\to\infty} \frac{1}{n} I(\mathcal{A}^n) = f_{\mathcal{I}} \text{ a.e. and in } L^1. \tag{1.5.1}$$

Furthermore

$$\mathrm{h}(T, \mathcal{A}) = \lim_{n\to\infty} \frac{1}{n} \mathrm{H}(\mathcal{A}^{n-1}) = \int f_{\mathcal{I}}\, d\mu = \int f\, d\mu. \tag{1.5.2}$$

The limit $f$ will gain a new interpretation in (1.8.6), in the context of Lebesgue spaces, where the notion of information function $I$ will be generalized.

**Proof.** First note that $f_n = I(\mathcal{A}|\mathcal{A}_1^n)$ converge to an integrable $f$ by Corollary 1.5.3. (Caution: though integrals of $f_n$ decrease to the entropy, Lemma 1.4.3, it is usually not true that $f_n$ decrease.) Hence the a.e. convergence of time averages to $f_{\mathcal{I}}$ a.e. holds by Birkhoff's Ergodic Theorem. It will suffice to prove (1.5.1) since then (1.5.2), the second equality, holds by integration and the last equality by Birkhoff's Ergodic Theorem, the convergence in $L^1$.

Let us now establish some identities (compare Lemma 1.4.3). Let $\{\mathcal{A}_n : n \geq 0\}$ be a sequence of countable partitions. Then we have

$$I\left(\bigvee_{i=0}^{n} \mathcal{A}_i\right) = I\left(\mathcal{A}_0|\bigvee_{i=1}^{n} \mathcal{A}_i\right) + I\left(\bigvee_{i=1}^{n} \mathcal{A}_i\right) = I\left(\mathcal{A}_0|\bigvee_{i=1}^{n} \mathcal{A}_i\right) + I\left(\mathcal{A}_1|\bigvee_{i=2}^{n} \mathcal{A}_i\right) + .. + I(\mathcal{A}_n).$$

In particular, it follows from the above formula that for $\mathcal{A}_i = T^{-i}\mathcal{A}$, we have

$$\begin{aligned} I(\mathcal{A}^n) &= I(\mathcal{A}|\mathcal{A}_1^n) + I(T^{-1}\mathcal{A}|\mathcal{A}_2^n) + \ldots + I(T^{-n}\mathcal{A}) \\ &= I(\mathcal{A}|\mathcal{A}_1^n) + I(\mathcal{A}|\mathcal{A}_1^{n-1}) \circ T + \ldots I(\mathcal{A}) \circ T^n \\ &= f_n + f_{n-1} \circ T + f_{n-2} \circ T^2 + \ldots + f_0 \circ T^n, \end{aligned}$$

where $f_k = I(\mathcal{A}|\mathcal{A}_1^k)$, $f_0 = I(\mathcal{A})$. Now

$$\left|\frac{1}{n+1} I(\mathcal{A}^n) - f_{\mathcal{I}}\right| \leq \left|\frac{1}{n+1}\sum_{j=0}^{n}(f_{n-i} \circ T^i - f \circ T^i)\right| + \left|\frac{1}{n+1}\sum_{j=0}^{n} f \circ T^i - f_{\mathcal{I}}\right|.$$

Since by Birkhoff's Ergodic Theorem the latter term converges to zero both almost everywhere and in $L^1$ it suffices to prove that for $n \to \infty$

$$\frac{1}{n+1}\sum_{i=0}^{n} g_{n-i} \circ T^i \to 0 \text{ a.e. and in } L^1. \tag{1.5.3}$$

where $g_k = |f - f_k|$.

Now, since $T$ is measure preserving, for every $i \geq 0$

$$\int g_{n-i} \circ T^i d\mu = \int g_{n-i} d\mu.$$

Thus $\frac{1}{n}\sum_{i=0}^{n} \int g_{n-i} \circ T^i \, d\mu = \frac{1}{n}\sum_{i=0}^{n} \int g_{n-i} \, d\mu \to 0$, since $f_k = \to f$ in $L^1$ by Corollary 1.5.3. Thus we established the $L^1$ convergence in (1.5.3).

Now, let $G_N = \sup_{n>N} g_n$. Of course $G_N$ is monotone decreasing and by $g_n \to 0$ a.e. (Corollary 1.5.3) we get $G_N \searrow 0$ a.e.. Moreover, by Corollary 1.5.2, $G_0 \leq \sup_n f_n + f \in L_1$.

For $N < n$ we have

$$\begin{aligned} \frac{1}{n+1}\sum_{i=0}^{n} g_{n-i} \circ T^i &= \frac{1}{n+1}\sum_{i=0}^{n-N-1} g_{n-i} \circ T^i + \frac{1}{n+1}\sum_{i=n-N}^{n} g_{n-i} \circ T^i \\ &\leq \frac{1}{n+1}\sum_{i=0}^{n-N-1} G_N \circ T^i + \frac{1}{n+1}\sum_{i=n-N}^{n} G_0 \circ T^i. \end{aligned}$$

Hence, for $K_N = G_0 + G_0 \circ T + \ldots + G_0 \circ T^N$

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} g_{n-i} \circ T^i \le (G_N)_{\mathcal{I}} + \limsup_{n\to\infty} \frac{1}{n+1} K_N \circ T^{n-N} = (G_N)_{\mathcal{I}} \ \text{ a.e.},$$

where $(G_N)_{\mathcal{I}} = \lim n \to \infty \frac{1}{n+1} \sum_{i=0}^{n} G_N \circ T^i$ (Birkhoff's Ergodic Theorem).

Now $(G_N)_{\mathcal{I}}$ decreases with $N$ because $G_N$ decreases, and

$$\int (G_N)_{\mathcal{I}}\, d\mu = \int G_N d\mu \to 0$$

because $G_N$ are non-negative uniformly bounded by $G_0 \in L^1$ and tend to 0 a.e..

Hence $(G_N)_{\mathcal{I}} \to 0$ a.e.. Therefore

$$\limsup_{n\to\infty} \frac{1}{n+1} \sum_{i=0}^{n} g_{n-i} \circ T^i \to 0 \ \text{ a.e.}$$

establishing the missing a.e. convergence in (1.5.3). ♣

As an immediate consequence of (1.5.1) and 1.5.2) for $T$ ergodic, along with $f_{\mathcal{I}} = \int f_{\mathcal{I}}\, d\mu$, we get the following:

**Theorem 1.5.5** (Shannon-McMillan-Breiman, ergodic case) If $T : X \to X$ is ergodic and $\mathcal{A}$ is a countable partition of finite entropy, then

$$\lim_{n\to\infty} \frac{1}{n} I(\mathcal{A}^{n-1})(x) = \mathrm{h}_\mu(T, \mathcal{A}). \quad \text{for a.e. } x \in X$$

The left hand side can be viewed as a *local entropy* at $x$. The Theorem says that at a.e. $x$ the local entropy exists and is equal to the entropy (compare comments after (1.3.2) and Lemma 1.4.2).

## §1.6. LEBESGUE SPACES, MEASURABLE PARTITIONS AND CANONICAL SYSTEMS OF CONDITIONAL MEASURES.

Let $(X, \mathcal{F}, \mu)$ be a probability space. We consider only *complete* measures (probabilities), namely such that every subset of a set of measure 0 is measurable. If a measure is not complete we can always consider its completion, namely to include in the completion of $\mathcal{F}$ all sets $A$ such that there exists $B \in \mathcal{F}$ with $A \div B$ contained in a set in $\mathcal{F}$ of measure 0. Consider $\mathcal{A}$, an arbitrary partition of $X$, not necessarily countable nor consisting of measurable sets. By $\tilde{\mathcal{A}}$ we denote the sub $\sigma$-algebra of $\mathcal{F}$ consisting of those sets in $\mathcal{F}$ that are unions of whole elements (fibres) of $\mathcal{A}$. Note that $\tilde{\mathcal{A}} \supset \sigma(\mathcal{A})$ defined in Sec.1 (in case $\mathcal{A} \subset \mathcal{F}$) but the inclusion can be strict. Obviously $\tilde{\mathcal{A}} \supset \{\emptyset, X\}$.

**Definition 1.6.0.** The partition $\mathcal{A}$ is called *measurable* if it satisfies the following separation property.

(1.6.1) There exists a sequence $\mathbf{B}=\{B_n : n \geq 1\} \subset \tilde{\mathcal{A}}$ such that for any two $A_1, A_2 \in \mathcal{A}$ with $A_1 \neq A_2$ there is an integer $n \geq 1$ such that either

$$A_1 \subset B_n \text{ and } A_2 \subset X \setminus B_n$$

or

$$A_2 \subset B_n \text{ and } A_1 \subset X \setminus B_n$$

Since each element of the measurable partition $\mathcal{A}$ can be represented as an intersection of countably many elements $B_n$ or their complements, each element of $\mathcal{A}$ is measurable. Let us stress however that the measurability of all elements of $\mathcal{A}$ is not sufficient for $\mathcal{A}$ to be a measurable partition (see Exercise 1). The sequence $\mathbf{B}$ is called a *basis* for $\mathcal{A}$.

**Remark 1.6.0a.** A popular definition of an uncountable measurable partition $\mathcal{A}$ is that there exists a sequence of finite partitions (recall that this means: finite partitions into measurable sets) $\mathcal{A}_n, n = 0, 1, ...,$ such that $\mathcal{A} = \bigvee_{n=0}^{\infty} \mathcal{A}_n$. Here (unlike later on) the join $\bigvee$ is considered in the set-theoretic sense, i.e. as $\{A_{n_1} \cap A_{n_2} \cap ... : A_{n_i} \in \mathcal{A}_{n_i}, i = 1, ...\}$. Clearly it is equivalent to (1.6.1).

Notice that for any measurable map $T : X \to X'$ between probability measure spaces, if $\mathcal{A}$ is a measurable partition of $X'$, then $T^{-1}(\mathcal{A})$ is a measurable partition of $X$.

Now we pass to the very useful class of probability spaces: *Lebesgue spaces.*

**Definition 1.6.1.** We call a sequence $\mathbf{B}= (B_n : n \geq 1) \subset \mathcal{F}$, *basis* of $(X, \mathcal{F}, \mu)$ if the two following conditions are satisfied:

(i) (1.6.1) holds for $\mathcal{A} = \varepsilon$, the partition into points;

(ii) for any $A \in \mathcal{F}$ there exists a set $C \in \sigma(\mathbf{B})$ such that $C \supset A$ and $\mu(C \setminus A) = 0$.

(Recall, Sec.1, that $\sigma(\mathbf{B})$ denotes the smallest $\sigma$-algebra containing all $B_n \in \mathbf{B}$. Rohlin used the name Borel $\sigma$-algebra.) ,

$(X, \mathcal{F}, \mu)$ satisfying (i) and (ii) for a basis $\mathbf{B}$ is called *separable.*

Now let $\varepsilon = \pm 1$ and $B_n^{(\varepsilon)} = B_n$ if $\varepsilon = 1$ and $B_n^{(\varepsilon)} = X \setminus B_n$ if $\varepsilon = -1$. To any sequence of numbers $\varepsilon_n, n = 1, 2, ...$ there corresponds the intersection $\bigcap_{n=1}^{\infty} B_n^{(\varepsilon_n)}$. By (i) every such intersection contains no more than one point.

The space $(X, \mathcal{F}, \mu)$ is said to be *complete* with respect to a basis $\mathbf{B}$ if all the intersections $\bigcap_{n=1}^{\infty} B_n^{(\varepsilon_n)}$ are non-empty. The space $(X, \mathcal{F}, \mu)$ is said to be *complete* (mod 0) with respect to a basis $\mathbf{B}$ if $X$ can be included as a subset of full measure into a certain measure space $(\overline{X}, \overline{\mathcal{F}}, \overline{\mu})$ which is complete with respect to its own basis $\overline{\mathbf{B}} = (\overline{B}_n)$ satisfying $\overline{\mathrm{B}}_n \cap X = B_n$ for all $n$.

It turns out that a space which is complete (mod 0) with respect to its one bases is also complete (mod 0) with respect to its every other basis.

**Definition 1.6.2.** The space $(X, \mathcal{F}, \mu)$ complete (mod 0) with respect to one of its bases is called *Lebesgue space.*

Exercise: If $(X_1, \mathcal{F}_1, \mu_1)$ and $(X_2, \mathcal{F}_2, \mu_2)$ are two probability spaces with complete measures, such that $X_1 \subset X_2, \mu_2(X_2 \setminus X_1) = 0$ and $\mathcal{F}_1 = \mathcal{F}_2|_{X_1}, \mu_1 = \mu_2|_{\mathcal{F}_1}$ (where $\mathcal{F}_2|_{X_1} := \{A \cap X_1 : A \in \mathcal{F}_2\}$), then the first space is Lebesgue iff the second is.

It is not difficult to check that (see Exercise 3) that $(X, \mathcal{F}, \mu)$ is a Lebesgue space if and only if $(X, \mathcal{F}, \mu)$ is isomorphic to the unit interval (equipped with classical Lebesgue measure) together with countably many atoms.

**Theorem 1.6.3.** Assume that $T : X \to X'$ is a measurable injective map from a Lebesgue space $(X, \mathcal{F}, \mu)$ to a separable space $(X', \mathcal{F}', \mu')$ and pre-images of the sets of mesure 0 (or positive) are of measure 0 (resp. positive). Then the space $(Y', \mathcal{F}', \mu')$ is Lebesgue and $T^{-1}$ is a measurable map.

Remark that in particular a measurable, measure preserving, injective map between Lebesgue spaces is an isomorphism. If $X = X', \mathcal{F} \supset \mathcal{F}', \mathcal{F} \neq \mathcal{F}'$ and $X', \mathcal{F}', \mu'$ is separable, then the above implies that $(X, \mathcal{F}, \mu)$ is not Lebesgue.

Let now $(X, \mathcal{F}, \mu)$ be a Lebesgue space and $\mathcal{A}$ be a measurable partition of $X$. We say that a property holds for all almost all atoms of $\mathcal{A}$ if and only if the union of atoms for which it is satisfied is measurable, of full measure. The following fundamental theorem holds:

**Theorem 1.6.4.** For almost all $A \in \mathcal{A}$ there exists a Lebesgue space $(A, \mathcal{F}_A, \mu_A)$ such that the following conditions are satisfied:

(1.6.2) If $B \in \mathcal{F}$, then $B \cap A \in \mathcal{F}_A$ for almost all $A \in \mathcal{A}$.

(1.6.3) The function $X \to [0, 1]$, $x \mapsto \mu_{A(x)}(B \cap A(x))$ is $\mathcal{F}$-measurable for all $B \in \mathcal{F}$, where $A(x)$ is the element of $\mathcal{A}$ containing $x$.

(1.6.4) $\mu(B) = \int_X \mu_{A(x)}(B \cap A(x))\, d\mu(x)$

**Remark 1.** One can consider the quotient (factor) space $(X/\mathcal{A}, \mathcal{F}_{\mathcal{A}}, \mu_{\mathcal{A}})$ with $X/\mathcal{A}$ being just $\mathcal{A}$ and with $\mathcal{F}_{\mathcal{A}} = p(\tilde{\mathcal{A}})$ and $\mu_{\mathcal{A}}(B) = \mu(p^{-1}(B))$ for the projection $p(x) = A(x)$. It can be proved that the factor space is again a Lebesgue space. Then $x \mapsto \mu_A(x)(B \cap A(x))$ is $\mathcal{F}_{\mathcal{A}}$-measurable and the property 1.6.4 can be rewritten in the form

$$\mu(B) = \int_{X/\mathcal{A}} \mu_A(B \cap A)\, d\mu_{\mathcal{A}}(A). \tag{1.6.5}$$

**Remark 2.** If partition $\mathcal{A}$ is finite or countable, then the measures $\mu_A$ are just the conditional measures given by the formulas $\mu_A(B) = \mu(A \cap B)/\mu(A)$.

**Remark 3.** (1.6.4) can be rewritten for every $\mu$-integrable $\phi$, or non-negative $\mu$-measurable $\phi$ if we allow $+\infty$-ies, as

$$\int \phi\, d\mu = \int_X \Big( \int_{A(x)} \phi|_{A(x)}\, d\mu_{A(x)} \Big)\, d\mu(x).$$

This is a version of the Fubini Theorem.

The family of measures $\{\mu_A : A \in \mathcal{A}\}$ is called the *canonical system of conditional measures* with respect to the partition $\mathcal{A}$. It is unique (mod 0) in the sense that any other system $\mu'_A$ coincides with it for almost all atoms of $\mathcal{A}$.

The method of construction of the system $\mu_A$ is *via* conditional expectations values with respect to the $\sigma$-algebra $\tilde{\mathcal{A}}$. Having chosen a basis $(B_n)$ of the Lebesgue space $(X, \mathcal{F}, \mu)$, for every finite intersection

$$B = \bigcap_i B_{n_i}^{(\varepsilon_{n_i})} \tag{1.6.6}$$

one considers $\phi_B := E(1\!\!1_B | \mathcal{A})$, that can be considered as a function on the factor space $X/\mathcal{A}$, unique on a.e. $A \in \mathcal{A}$ such that for all $Z \in \tilde{\mathcal{A}}$

$$\mu(B \cap Z) = \int_{p(Z)} \phi_B(A)\, d\mu_{\mathcal{A}}(A).$$

Clearly $(B_n \cap A)$ is a basis for all $A$. It is not hard to prove that for a.e. $A$, for each $B$ from our countable family (1.6.6), $\phi_B(A)$ as a function of $B$ generates Lebesgue space on $A$. Uniqueness of $\phi_B$ yields additivity.

**Theorem 1.6.5.** If $T : X \to X'$ is a measurable map of a Lebesgue space $(X, \mathcal{F}, \mu)$ onto a Lebesgue space $(X', \mathcal{F}', \mu')$, then the induced map from $(X/\zeta, \mathcal{F}_\zeta, \mu_\zeta)$ for $\zeta = T^{-1}(\varepsilon)$, to $(X', \mathcal{F}', \mu')$ is an isomorphism.

**Proof.** This immediately follows from the fact that the quotient is a Lebesgue space and from Theorem 1.6.2. ♣

In what follows we consider partitions (mod 0), i.e. we identify two partitions if they coincide, restricted to a measurable subset of full measure. For these classes of equivalence we use the same notation $\leq, \geq$ as in Section 4. . They define a partial order. If $\mathcal{A}_\tau$ is a family of measurable partitions of a measure space (unlike in previous Sections the family may be uncountable), then by its product $\mathcal{A} = \bigvee_\tau \mathcal{A}_\tau$ we mean the measurable partition $\mathcal{A}$ defined by

(i) $\mathcal{A} \geq \mathcal{A}_\tau$ for every $\tau$;

(ii) if $\mathcal{A}' \geq \mathcal{A}_\tau$ for every $\tau$ and $\mathcal{A}'$ is measurable, then $\mathcal{A}' \geq \mathcal{A}$.

Similarly, replacing $\geq$ by $\leq$, we define the intersection $\bigwedge_\tau \mathcal{A}_\tau$.

The product and intersection exist in a Lebesgue space (i.e. the partially ordered structure is complete). They of course generalize the notions of Section 4. Clearly for a countable family of measurable partitions $\mathcal{A}_\tau$ the above $\bigvee$ and the set-theoretic one coincide (the assumption the space is Lebesgue and the reasoning (mod 0) is not needed). In Exercise 7 we give some examples.

There is a natural one-to-one correspondence between the measurable partitions (mod 0) of a Lebesgue space $(\mathcal{A}, \mathcal{F}, \mu)$ and the complete $\sigma$-subalgebras of $\mathcal{F}$, i.e. such $\sigma$-algebras

$\mathcal{F}' \subset \mathcal{F}$ that the measure $\mu$ restricted to $\mathcal{F}'$ is complete. This correspondence is defined by the assignment to each $\mathcal{A}$ the $\sigma$-algebra $\mathcal{F}(\mathcal{A})$ of all sets which coincide (mod 0) with the sets of $\tilde{\mathcal{A}}$ (defined at the beginning of this Section). To operations on the measurable partitions (mod 0) correspond operations on the corresponding $\sigma$-algebras. Namely , if $\mathcal{A}_\tau$ is a family of measurable partitions (mod 0), then

$$\mathcal{F}(\bigvee_\tau \mathcal{A}_\tau) = \bigvee_\tau \mathcal{F}(\mathcal{A}_\tau), \qquad \mathcal{F}(\bigwedge_\tau \mathcal{A}_\tau) = \bigwedge_\tau \mathcal{F}(\mathcal{A}_\tau).$$

Here $\bigwedge_\tau \mathcal{F}(\mathcal{A}_\tau) = \bigcap_\tau \mathcal{F}(\mathcal{A}_\tau)$ is the set-theoretic intersection of the $\sigma$-algebras, while $\bigvee_\tau \mathcal{F}(\mathcal{A}_\tau)$ is the set-theoretic intersection of all the $\sigma$-algebras which contain all $\mathcal{F}(\mathcal{A}_\tau)$.

For a monotone increasing (decreasing) sequence of measurable partitions $\mathcal{A}_n$ and $\mathcal{A} = \bigvee_n \mathcal{A}_n$ ($\mathcal{A} = \bigwedge_n \mathcal{A}_n$ respect.) we write $\mathcal{A}_n \nearrow \mathcal{A}$ (or $A_n \searrow \mathcal{A}$). In the language of measurable partitions of a Lebesgue space the Martingale Theorem 1.1.1 can be expressed as follows:

**Theorem 1.6.6.** If $\mathcal{A}_n \nearrow \mathcal{A}$ or $\mathcal{A}_n \searrow \mathcal{A}$, then for every integrable function $f$, $\mu$ a.s. $E(f|\mathcal{A}_n) \to E(f|\mathcal{A})$, where for $\mathcal{A}$ any measurable partition one writes $E(f|\mathcal{A})(x) := \int f|_{\mathcal{A}(x)}\, d\mu_{\mathcal{A}(x)}$.

**Proof.** By the definition of canonical system of conditional measures and the definition of conditional expectation value we have for every measurable partition $\mathcal{A}$ the identity $E(f|\mathcal{A}) = E(f|\mathcal{F}(\mathcal{A}))$. ♣

## §1.7 ROHLIN NATURAL EXTENSION

We shall prove here following very useful (see Ch.8.9)

**Theorem 1.7.1.** For every measure preserving endomorphism $T$ of a Lebesgue space $(X, \mathcal{F}, \mu)$ there exists a Lebesgue space $(\tilde{X}, \tilde{\mathcal{F}}, \tilde{\mu})$ with measure preserving transformations $\pi_n : \tilde{X} \to X, n \le 0$ satisfying $T \circ \pi_{n-1} = \pi_n$, which is an inverse limit of the system $... \xrightarrow{T} X \xrightarrow{T} X$.

Recall that in *category theory* [Lang, Ch.I], for a sequence (system) of objects and morphisms $... \xrightarrow{M_{n-1}} O_n \xrightarrow{M_n} ...... \xrightarrow{M_0} O_0$ an object $O$ equipped with morphisms $\pi_n : O \to O_n$ is called an *inverse limit* if $M_n \circ \pi_{n-1} = \pi_n$ and for every other $O'$ equipped with morphisms $\pi_n' : O' \to O_n$ satisfying $M_n \circ \pi_{n-1}' = \pi_n'$ there exists a unique morphism $M : O \to O$ such that $\pi_n \circ M = \pi_n'$ for every $n \le 0$. In particular, for $\pi_n' := M_n \circ \pi_{n-1} : O \to O_n$, there exists $M : O \to O$ such that $\pi_n \circ M = \pi_n' = M_n \circ \pi_{n-1}$ for every $n$. It is easy to see that $M$ is an automorphism.

Here objects are probability spaces or probability spaces with complete probabilities, and morphisms are measure preserving transformations or measure preserving transformations up to sets of measure 0. (We have thus multiple meaning of Theorem 1.7.1.)

Thus Theorem 1.7.1 produces a measure preserving automorphism $\tilde{T} : \tilde{X} \to \tilde{X}$ satisfying

$$\pi_n \circ \tilde{T} = T \circ \pi_{n-1} \tag{1.7.1}$$

for every $n \le 0$. This automorphism is called Rohlin's *natural extension* of $T$.

In the proof of the Theorem we shall use the following

**Theorem 1.7.2** (On Extension of Measure). Every probability measure $\nu$ ($\sigma$-additive) on an algebra $\mathcal{G}_0$ of subsets of a set $X$ can be uniquely extended to a measure on the $\sigma$-algebra $\mathcal{G}$ generated by $\mathcal{G}_0$

This Theorem can be proved with the use of the famous construction by Carathéodory [Carathéodory, Ch.V], namely by the construction of the outer measure: $\nu_e(A) = \inf \nu(B) : B \in \mathcal{G}_0, A \subset B$ for every $A \subset X$.

We say that $A$ is measurable (in Carathéodory's sense) if for every $E \subset X$ the outer measure $\nu_e$ satisfies $\nu_e(E) = \nu_e(E \cap A) + \nu_e(E \setminus A)$. The family of these sets appears to be a $\sigma$-algebra containing $\mathcal{G}_0$, hence containing $\mathcal{G}$.

For a general definition of outer measure and sketch of the theory see Ch.6.

**Proof of Theorem 1.7.1.** We start with producing inverse limit in the set-theoretic category: Consider for $\boldsymbol{Z}^-$, the set of all non-positive integers, the space

$$\tilde{X} = \{(x_n)_{n \in \boldsymbol{Z}^-} : T(x_n) = x_{n+1} \ \forall n < 0\}. \tag{1.7.2}$$

and $\pi_i : \tilde{X} \to X$ the projection to the $i$-th coordinate, $\pi_i((x_n)_{n\in\boldsymbol{Z}}) = x_i$

Now provide $\tilde{X}$ with a $\sigma$-algebra $\tilde{\mathcal{F}}$ and probability measure $\tilde{\mu}$, so that $(\tilde{X}, \tilde{\mathcal{F}}, \tilde{\mu})$ becomes the inverse limit.

Consider $\mathcal{G}_n = \pi_n^{-1}(\mathcal{F})$. Note that this is an increasing sequence of $\sigma$-algebras with growing $|n|$ because $\pi^{-1}(A) = \pi_{n-1}^{-1}(T^{-1}(A))$ for every $A \in \mathcal{F}$. Write $\tilde{\mathcal{F}}_0 = \bigcup_{n\le 0} \mathcal{G}_n$. This is an algebra. For every $A \in \mathcal{F}$ and $n \le 0$ define $\tilde{\mu}(\pi_n^{-1}(A)) := \mu(A)$. This is well-defined because if $C = \pi_n^{-1}(A_1) = \pi_m^{-1}(A_2)$ for $A_1, A_2 \in \mathcal{F}$ and $n < m$ then $A_1 = T^{-(m-n)}(A_2)$. Since $T$ preserves $\mu$, we have $\mu(A_1) = \mu(A_2)$.

The next step is to observe that $\tilde{\mu}$ is $\sigma$-additive on the algebra $\tilde{\mathcal{F}}_0$. For that we use the assumption $(X, \mathcal{F}, \mu)$ is a Lebesgue space[1]. We just assume that $X$ is a full Lebesgue measure subset of the unit interval, with classical Lebesgue measure and atoms, and the $\sigma$-algebra of Lebesgue measurable sets $\mathcal{F}$, see Exercise 3. Now it is sufficient to prove that for every decreasing sequence $C_i \in \tilde{\mathcal{F}}_0, i = 1, 2, \ldots$ if $\bigcap_i C_i = \emptyset$ then $\tilde{\mu}(C_i) \to 0$. Suppose to the contrary that there exists $\varepsilon > 0$ such that $\tilde{\mu}(C_i) \ge \varepsilon$ for every $i$. Passing to a subsequence and reindexing we can write $C_n = \pi_{-n}^{-1}(C_n'), n = 1, 2, \ldots$. We construct compact sets $D_n' \subset C_n'$ such that $\mu(C_n' \setminus D_n') \le \varepsilon 2^{-(n+1)}$ and $T|_{D_n'}$ is continuous for all $n$ (Lusin's Theorem, [Halmos, Sec.55]).

[1]this "detail" has been ovelooked in [CFS]

Write $\Pi = \prod_{-\infty}^{0} X = \{(x_n)_{n\in \boldsymbol{Z}^-} : x_n \in X\}$ for the cartesian product of the countable number of exemplars of $X$ with the product topology (compact by Tichonov's Theorem). Define $\tilde{X}^n := \{(x_i)_{i\in \boldsymbol{Z}^-} : T(x_i) = x_{i+1} \forall n \le i < 0\}$. Of course $\tilde{X} \subset \tilde{X}^n \subset \Pi$. Denote by $\pi^n$ the projection from $\tilde{X}^n$ to the $n$-th coordinate.

Then the sets $D_n = \bigcap_{i=1}^{n} (\pi^i)^{-1}(D_i')$ are compact and decreasing. They are non-empty because $\mu(\pi^n(D_n)) \ge \varepsilon/2$ by the construction, for all $n$. Therefore $\bigcap_n D_n$ is non-empty (Cantor theorem). Notice finally that $\bigcap_n D_n \subset \bigcap C_n$ so the latter set is non-empty. We have proved that $\tilde{\mu}$ is $\sigma$-additive on $\tilde{\mathcal{F}}_0$.

The final step of the construction is the extension of $\tilde{\mu}$ to the $\sigma$-algebra $\tilde{\mathcal{F}}$ generated by $\tilde{\mathcal{F}}_0$. It exists (and is unique) due to Theorem 1.7.2.

If we work in the category of complete measures we define the $\sigma$-algebra $\tilde{\mathcal{F}}$ as the completion (by subsets of sets of measure 0) of the $\sigma$-algebra generated by $\tilde{\mathcal{F}}_0$.

Thus the probability space $(\tilde{X}, \tilde{\mathcal{F}}, \tilde{\mu})$ has been constructed. We leave checking that it is indeed the inverse limit to the reader.

Let us prove that the probability space $(\tilde{X}, \tilde{\mathcal{F}}, \tilde{\mu})$ with completed $\tilde{\mu}$ is Lebesgue space. Let $(B_l)$ be a basis of $(X, \mathcal{F}, \mu)$. Denote by $\pi_n$ the projection of $\Pi$ to the $n$-th coordinate for all $n$. (We use the same symbol as for projections from $\tilde{X} \subset \Pi$ before. Recall also that projections from intermediate domains have been denoted by $\pi^n$.) Then clearly the family $\pi_n^{-1}(B_l)$ is a basis of the partition $\varepsilon$ in $\Pi$. The restrictions of $\pi_n^{-1}(B_l)$ to $\tilde{X}$ generate the $\sigma$-algebra $\tilde{\mathcal{F}}$ on $\tilde{X}$ discussed before (in the sense of Def.1.6.1 (ii)), because $(B_l)$ generates $\mathcal{F}$. We define

$$\tilde{\mu}_{\Pi,n}\Big(\bigcap_{i=-n}^{0} \pi_i^{-1}(C_i)\Big) := \mu\Big(\bigcap_{i=-n}^{0} T^{-(i-n)}(C_i)\Big),$$

for $C_i = \bigcap_{l=1}^{l(i)} B_l^{(\varepsilon_{l,i})}$ for all $n \le i \le 0$ and $\pm 1$ sequences $\varepsilon_{l,i}$ for $l = 1, 2, ..., l(i)$. It is easy to see that the sequence $\tilde{\mu}_{\Pi,n}$ is compatible on algebras: finite unions of $\bigcap_{i=-n}^{0} \pi_l^{-1}(C_i)$, namely $\tilde{\mu}_{\Pi,n+1}$ extends $\tilde{\mu}_{\Pi,n}$. (One says that this is a *compatible family of of finite-dimensional probability distributuions.*)But $\Pi$ is compact, hence $\tilde{\mu}_{\Pi,n}$ are $\sigma$-additive on the union of these algebras, hence extend to a measure ($\sigma$-additive) $\tilde{\mu}_\Pi$ on the $\sigma$-algebra $\tilde{\mathcal{F}}_\Pi$ generated by them (Kolmogoroff Theorem, see bibliographical notes). The restriction of $\tilde{\mu}_\Pi$ to $\tilde{X}$ coincides with $\tilde{\mu}$ on $\tilde{\mathcal{F}}$ by the uniqueness in Theorem 1.7.2. The restriction of $\tilde{\mathcal{F}}_\Pi$ is a $\sigma$-algebra so it contains $\tilde{\mathcal{F}}$. We shall know these $\sigma$-algebras coincide if we verify that $\tilde{X}$ is $\tilde{\mu}_\Pi$-measurable, i.e. $\tilde{X} \in \tilde{\mathcal{F}}_\Pi$.

Thus the assertion to be proved is that $\tilde{X} \in \tilde{\mathcal{F}}_\Pi$ and that $\tilde{X}$ is of full measure $\tilde{\mu}_\Pi$. This will prove that $(\tilde{X}, \tilde{\mathcal{F}}, \tilde{\mu})$ is complete (mod 0) with respect to $(B_l)$ restricted to $\tilde{X}$, hence it is Lebesgue space.

Recall that $\tilde{X} = \bigcap_n \tilde{X}^n$ and note that by Lusin Theorem for each $n$ there exist compact sets $D_{n,i} \subset \tilde{X}^n, i = 1, 2, ...$ such that $\tilde{\mu}_\Pi\Big(\Pi \setminus \bigcup_i D_{n,i}\Big) = 0$. Compact sets are measurable as their complementary open sets are countable unions of cylinders. ♣

**Remark 1.** $\tilde{X}$ can be interpreted as the space of all backward trajectories for $T$. The map $\tilde{T} : \tilde{X} \to \tilde{X}$ can be defined by the formula

$$\tilde{T}((x_n)_{n\in \boldsymbol{Z}^-}) = (..., x_{-2}, x_{-1}, x_0, T(x_0)). \tag{1.7.3}$$

$\tilde{X}$ could be defined in (1.7.2) as the space of full trajectories $\{(x_n)_{n\in\boldsymbol{Z}}; T(x_n) = x_{n+1}\}$. Then (1.7.3) is the *shift to the left*.

The formula (1.7.3) holds because $\tilde{T}$ defined by it, satisfies (1.7.1), and there holds uniqueness of $\tilde{T}$ satisfying (1.7.1)

**Remark 2.** Alternatively to Lusin Theorem argument above, we could find for $\tilde{X}^n$ sets $E_{n,i} \supset \tilde{X}^n$, with $\tilde{\mu}_\Pi(E_{n,i} \setminus \tilde{X}^n) \to 0$, which are unions of cylinders $\bigcap_{i=-n}^{0} \pi_i^{-1}(C_i)$. This agrees with the following general fact:

If a sequence of sets $\Sigma$ generates a $\sigma$-algebra $\mathcal{G}$ with a mesure $\nu$ on it (see Def.1.6.2 (ii)) then for every $A \in \mathcal{G}$ there exists $C \supset A$ with $\nu(C \setminus A) = 0$ such that $C \in \Sigma'_{d\sigma\delta}$, i.e. $C$ is a countable intersection of countable unions of finite intersections of sets belonging to $\Sigma$ or their complements. Exercise: Prove this general fact, using Caratheodory's outer measure constructed on measurable sets.

The construction via Lusin theorem presents $\tilde{X}$ as $\Sigma'_{d\sigma\delta\sigma\delta}$ set up to measure 0 (as compact sets are in $\Sigma'_{d\sigma\delta}$). So it is not the most economic.

**Remark 3.** Another way to prove Theorem 1.7.1 is to construct first $(\Pi, \tilde{F}_\Pi, \tilde{\mu}_{\mathrm{P}})$ on the infinite cartesian product $\Pi$, and next $(\tilde{X}, \tilde{F}, \tilde{\mu})$ as the restriction of the first probability space to $\tilde{X}$. We have chosen a different way in order to avoid in the construction the correctness of the definition of $\tilde{\mu}_n$'s in the finite products and the compatibility. We needed it only to prove that the inverse limit is Lebesgue.

We end this section with another version of Theorem 1.7.1. Let us start with

**Definition 1.7.3.** Suppose that $T$ is an automorphism of a Lebesgue space $(X, \mathcal{F}, \mu)$. Let $\zeta$ be a measurable partition. Assume it is forward invariant, namely $T(\zeta) \geq \zeta$, equivalently $T^{-1}(\zeta) \leq \zeta$. Then $\zeta$ is said to be *exhausting* if $\bigvee_{n\geq 0} T^n(\zeta) = \varepsilon$.

**Theorem 1.7.4.** For every measure preserving endomorphism $T$ of a Lebesgue space $(X, \mathcal{F}, \mu)$ there exists a Lebesgue space $(\tilde{X}, \tilde{\mathcal{F}}, \tilde{\mu})$ with an automorphism $\tilde{T}$, with a forward invariant exhausting measurable partition $\zeta$, such that $(X, \mathcal{F}, \mu) = (\tilde{X}/\zeta, \tilde{\mathcal{F}}_\zeta, \tilde{\mu}/\zeta)$ the factor space, cf.Sec.6, Remark 1, and $T$ is factor of $\tilde{T}$, namely $T \circ p = p \circ \tilde{T}$ for the projection $p : \tilde{X} \to X$.

**Proof.** Take $(\tilde{X}, \tilde{\mathcal{F}}, \tilde{\mu})$ and $\tilde{T}$ from Theorem 1.7.1. Set $\zeta := \pi_0^{-1}(\varepsilon)$. By (1.7.1) and $T^{-1}(\varepsilon) \leq \varepsilon$ we get $\tilde{T}^{-1}(\zeta) \leq \zeta$.

If $\varepsilon' = \bigvee_{n\geq 0} T^n(\zeta)$ is not the partition of $\tilde{X}$ into points, then $\tilde{T}/\varepsilon'$ is an automorphism of $(\tilde{X}/\varepsilon', \tilde{\mathcal{F}}_{\varepsilon'}, \tilde{\mu}_{\varepsilon'})$. Moreover if we denote by $p'$ the projection from $\tilde{X}$ to $\tilde{X}/\varepsilon'$ then we can write $\pi_{-n} = \pi'_{-n} \circ p'$ for some maps $\pi'_{-n}$ for every $n \geq 0$. By the definition of inverse limit $p'$ must have an inverse which is impossible.

The last part, that $\bigvee_{n\geq 0} T^n(\zeta)$ is the partition of $\tilde{X}$ into points, has also an immediate, not category theory, proof following directly from the form of $\tilde{X}$ in (1.7.2). Indeed for $n \geq 0$ $T^n(\zeta)$ at $\tilde{x} = (..., x_{-2}, x_{-1}, x_0$ is the $n$-th image of $\zeta$ at $\tilde{T}^{-n}(\tilde{x})$ i.e. at $(..., x_{-n-1}, x_{-n})$. So it is equal to $\{(..., x'_{-n-1}, x'_{-n}, ..., x_0) \in \tilde{X} : x_{-n} = x_{-n})\}$. Intersecting for $n \to \infty$ we obtain $\{\tilde{x}\}$.

## §1.8 GENERALIZED ENTROPY, CONVERGENCE THEOREMS.

This section contains generalizations of entropy notions introduced in Section 3 to the case of all measurable partitions. The triple $(X, \mathcal{F}, \mu)$ is assumed to be a Lebesgue space.

**Definition 1.8.1.** If $\mathcal{A}$ is a measurable partition of $X$ then its (generalized) entropy is defined as follows:

$$\mathrm{H}(\mathcal{A}) = \infty \text{ if } \mathcal{A} \text{ is not a countable partition (mod 0);}$$
$$H(\mathcal{A}) = \textstyle\sum_{A\in\mathcal{A}} -\mu(A)\log\mu(A) \text{ if } \mathcal{A} \text{ is a countable partition (mod 0).}$$

**Lemma 1.8.2.** If $\mathcal{A}_n$ and $\mathcal{A}$ are measurable partitions of $X$ and $\mathcal{A}_n \nearrow \mathcal{A}$, then $\mathrm{H}(\mathcal{A}_n) \nearrow \mathrm{H}(\mathcal{A})$.

**Proof.** Write $\mathrm{H}(\mathcal{A}) = \int I(\mathcal{A})\, d\mu$ where $I(\mathcal{A})(x) = -\log\mu(\mathcal{A}(x))$ is the information function (compare Sec.4, we set $\log 0 = -\infty$, hence $I(\mathcal{A})(x) = \infty$ if $\mu(\mathcal{A}(x)) = 0$, here). Write the same for $\mathcal{A}_n$. As $\mu(\mathcal{A}_n(x)) \searrow \mu(\mathcal{A}(x))$ for a.e. $x$, the convergence in the Lemma follows from Monotone Convergence Theorem. ♣

**Definition 1.8.3.** If $\mathcal{A}$ and $\mathcal{B}$ are two measurable partitions of $X$, then the (generalized) *conditional entropy* $\mathrm{H}(\mathcal{A}|\mathcal{B})$ of partition $\mathcal{A}$ subject to $\mathcal{B}$ is defined by the following integral

$$\mathrm{H}_\mu(\mathcal{A}|\mathcal{B}) = \int_{X/\mathcal{B}} \mathrm{H}_{\mu_B}(\mathcal{A}|B)\, d\mu_{\mathcal{B}}(B) \tag{1.8.1}$$

where $\mathcal{A}|B$ is the partition $\{A \cap B : A \in \mathcal{A}\}$ of $B$ and $\mu_B$ form a canonical system of conditional measures (Sec. 7). Choose a sequence of finite partitions $\mathcal{A}_n \nearrow \mathcal{A}$ (see Remark 1.6.0). The conditional entropy $\mathrm{H}_{\mu_B}(\mathcal{A}_n|B)$ is measurable as a function of $B$ in the factor space $(X/\mathcal{B}, \mathcal{F}_\mathcal{B}, \mu_\mathcal{B})$, hence of course as a function on $(X, \mathcal{F}, \mu)$, since it is a finite sum of measurable functions $B \mapsto -\mu_B(A\cap B)\log\mu_B(A\cap B)$. Since $\mathcal{A}_n|B \nearrow \mathcal{A}|B$ for a.e. $B$, we obtain, by using Lemma 1.8.2, that $\mathrm{H}_{\mu_B}(\mathcal{A}_n|B) \to \mathrm{H}_{\mu_B}(\mathcal{A}|B)$. Hence $\mathrm{H}_{\mu_B}(\mathcal{A}|B)$ is measurable, so our definition of $\mathrm{H}_\mu(\mathcal{A}|\mathcal{B})$ makes sense (we allow $\infty$'s here).

Of course (1.8.1) can be also written in the form

$$\int_X \mathrm{H}_{\mu_{\mathcal{B}(x)}}(\mathcal{A}|\mathcal{B}(x))\, d\mu(x), \tag{1.8.2}$$

with $\mathrm{H}_{\mu_B}(\mathcal{A}|B)$ understood as constant function on each $B$ (compare (1.6.4) versus (1.6.5)).

As in Sec.3 we can write

$$\mathrm{H}_\mu(\mathcal{A}|\mathcal{B}) = \int_X I(\mathcal{A}|\mathcal{B})\, d\mu, \tag{1.8.3}$$

where $I(\mathcal{A}|\mathcal{B})$ is the *conditional information function*:
$I(\mathcal{A}|\mathcal{B})(x) := -\log \mu_{\mathcal{B}(x)}(\mathcal{A}(x) \cap \mathcal{B}(x))$.

Indeed $I(\mathcal{A}|\mathcal{B})$ is non-negative and $\mu$-measurable as $\lim_{n\to\infty} I(\mathcal{A}_n|\mathcal{B})$ (a.e.), so (1.8.3) follows from (1.6.5a).

**Lemma 1.8.4.** If $\{\mathcal{A}_n : n \geq 1\}$ and $\mathcal{A}$ are measurable partitions, $\mathcal{A}_n \searrow \mathcal{A}$ and $\mathrm{H}(\mathcal{A}_1) < \infty$ then $\mathrm{H}(\mathcal{A}_n) \searrow \mathrm{H}(\mathcal{A})$.

**Proof.** The proof is similar to Proof of Lemma 1.8.2. ♣

**Theorem 1.8.5.** If $\mathcal{A}, \mathcal{B}$ are measurable partitions and $\{\mathcal{A}_n : n \geq 1\}$ is an increasing (decreasing and $\mathrm{H}(\mathcal{A}_1|\mathcal{B}) < \infty$) sequence of measurable partitions converging to $\mathcal{A}$, then

$$\lim_{n\to\infty} \mathrm{H}(\mathcal{A}_n|\mathcal{B}) = \mathrm{H}(\mathcal{A}|\mathcal{B}) \tag{1.8.4}$$

and the convergence is respectively monotone.

**Proof.** Applying Lemmas 1.8.2 and 1.8.4 we get the monotone convergence $\mathrm{H}_{\mu_B}(\mathcal{A}_n|B) \to \mathrm{H}_{\mu_B}(\mathcal{A}|B)$ for almost all $B \in X/\mathcal{B}$. Thus the integrals in the Definition 1.8.3 converge by the Monotone Convergence Theorem. ♣

**Theorem 1.8.6.** If $\mathcal{A}, \mathcal{B}$ are measurable partitions and $\{\mathcal{B}_n : n \geq 1\}$ is a decreasing (increasing and $\mathrm{H}(\mathcal{A}|\mathcal{B}_1) < \infty$) sequence of measurable partitions converging to $\mathcal{B}$, then

$$\lim_{n\to\infty} \mathrm{H}(\mathcal{A}|\mathcal{B}_n) = \mathrm{H}(\mathcal{A}|\mathcal{B}) \tag{1.8.5}$$

and the convergence is respectively monotone.

**Proof 1.** Assume first that $\mathcal{A}$ is finite (or countable with finite entropy). Then the a.e. convergence $I(\mathcal{A}|\mathcal{B}_n) \to I(\mathcal{A}|\mathcal{B})$ follows from Martingale Convergence Theorem (more precisely from Theorem 1.6.6), applied to $f = \mathbb{1}_A$, the indicator function for each $A \in \mathcal{A}$.

Now it is sufficient to prove $\sup_n I(\mathcal{A}|\mathcal{B}_n) \in L^1$ in order to use Dominated Convergence Theorem (compare Corollary 1.5.3) and (1.8.3). One can repeat Proofs of Lemma 1.5.1 (for increasing $\mathcal{B}_n$) and Corollary 1.5.2.

The monotonicity of the sequence $\mathrm{H}(\mathcal{A}|\mathcal{B}_n)$ relies on Theorem 1.3.d. However for infinite $\mathcal{B}_n$ one needs to approximate $\mathcal{B}_n$ by finite (or finite entropy) partitions. For details see [Rohlin 1967, Sec.5.12].

For $\mathcal{A}$ measurable, represent $\mathcal{A}$ as $\lim_{j\to\infty} \mathcal{A}_j$ for an increasing sequence of finite partitions $\mathcal{A}_j, j = 1, 2, ..$, next refer to Th.1.8.5. In the case of decreasing $\mathcal{B}_n$ the proof is straightforward. In the case of increasing $\mathcal{B}_n$ use

$$\mathrm{H}(\mathcal{A}|\mathcal{B}_n) - \mathrm{H}(\mathcal{A}_j|\mathcal{B}_n) = H(\mathcal{A}|(\mathcal{A}_j \vee \mathcal{B}_n)) \leq H(\mathcal{A}|(\mathcal{A}_j \vee \mathcal{B}_1)) \leq \mathrm{H}(\mathcal{A}|\mathcal{B}_1) - \mathrm{H}(\mathcal{A}_j|\mathcal{B}_1).$$

This implies that the convergence as $j \to \infty$ is uniform with respect to $n$, hence in the limit $\mathrm{H}(\mathcal{A}|\mathcal{B}_n) \to H(\mathcal{A}|\mathcal{B})$. ♣

**Proof 2.** For $\mathcal{A}$ finite (or countable with finite entropy) there is a simpler way to prove (1.8.5). We have for every $A \in \mathcal{A}$ by Theorem 1.1.1, the convergence in $L^2$ applied to $E(1\!\!1_A|\mathcal{F}(\mathcal{B}_n))$, hence the convergence in measure $\mu$ of $\mu_{\mathcal{B}_n(x)}(A \cap \mathcal{B}_n(x))$. By the continuity of the function $k(t) = -t\log t$, see Sec.3, this implies the convergence in measure $\mu$

$$k(\mu_{\mathcal{B}_n(x)}(A \cap \mathcal{B}_n(x))) \to k(\mu_{\mathcal{B}(x)}(A \cap \mathcal{B}(x))).$$

(We do not assume $x \in A$ here.) Summing over $A \in \mathcal{A}$ we obtain the convergence $\mathrm{H}_{\mu_{\mathcal{B}_n(x)}}(\mathcal{A}|\mathcal{B}_n(x)) \to \mathrm{H}_{\mu_{\mathcal{B}(x)}}(\mathcal{A}|\mathcal{B}(x))$ in measure $\mu$. These functions are uniformly bounded by $\log \#\mathcal{A}$ ( or by $\mathrm{H}(\mathcal{A})$ ) and non-negative, hence we get the convergence in $L^1$ and in consequence, due to (1.8.2), we obtain (1.8.5). (Note that we have not used the a.e. convergence in Th.1.1.1, but only the convergence in $L^2$ proved there.) ♣

Observe that we can rewrite now the definition of the entropy $\mathrm{h}_\mu(T,\mathcal{A})$ from Section 1.5 as

$$\mathrm{h}_\mu(T,\mathcal{A}) = \mathrm{H}(\mathcal{A}|\mathcal{A}^-), \quad \text{where } \mathcal{A}^- := \bigvee_{n=1}^{\infty} T^{-n}(\mathcal{A}). \tag{1.8.6}$$

A countable partition $\mathcal{B}$ is called *a countable generator* for an endomorphism of a Lebesgue space if $\mathcal{B}^m \nearrow \varepsilon$. Due to Theorem 1.8.6 we obtain the following facts useful in computing the entropy for concrete examples.

**Theorem 1.8.7.** (a) If $\mathcal{B}_m$ is a sequence of finite partitions of a Lebesgue space, such that $\mathcal{B}_m \nearrow \varepsilon$, then, for any endomorphism $T$ of the space, $\mathrm{h}(T) = \lim_{m\to\infty} \mathrm{h}(T,\mathcal{B}_m)$.

(b) If $\mathcal{B}$ is a countable generator of finite entropy for an endomorphism $T$ of a Lebesgue space, then $\mathrm{h}(T) = \mathrm{h}(T,\mathcal{B})$.

**Proof.** By Theorem 1.8.6 for every finite $\mathcal{A}$ we have $\lim_{m\to\infty} \mathrm{H}(\mathcal{A}|\mathcal{B}_m) = \mathrm{H}(\mathcal{A}|\varepsilon) = 0$, hence in view of Theorem 1.4.5, instead of $\sup_{\mathcal{A}}$ in the definition of $\mathrm{h}(T)$, it is sufficient in (1.4.3) to consider $\lim_{m\to\infty} \mathrm{h}(T,\mathcal{B}_m)$. This proves (a). Theorem 1.4.5 together with the definition of the generator prove also (b). ♣

**Remark 1.8.7'**. For $T$ an automorphism one considers two-sided countable (in particular finite) generator: $\bigvee_{n=-\infty}^{\infty} T^n(\mathcal{B}) = \varepsilon$. Then, as in the one-sided case, $\mathrm{H}(\mathcal{B})$ finite implies $\mathrm{h}(T) = \mathrm{h}(T,\mathcal{B})$.

**Remark 1.8.7".** In both Theorem 1.8.6 and Theorem 1.8.7(a) the assumption on the monotonicity of $\mathcal{B}_m$ can be weakened. Assume for example that $\mathcal{A}$ is finite and $\mathcal{B}_m \to \varepsilon$ in the sense that for every measurable $Y$, $E(1\!\!1_Y|\mathcal{B}_m) \to 1\!\!1_Y$ in measure, as in Remark 1.1.2. Then $\mathrm{H}(\mathcal{A}|\mathcal{B}_m) \to 0$, hence $\mathrm{h}(T) = \lim_{m\to\infty} \mathrm{H}(T,\mathcal{B}_m)$.

Indeed for $\mathrm{H}(\mathcal{A}|\mathcal{B}_m) \to 0$ just repeat Proof 2 of Theorem 1.8.6. The convergence in measure $\mu$ of $\mu_{\mathcal{B}_n(x)}(A \cap \mathcal{B}_n(x)))$ to $\mu_{\varepsilon(x)}(A \cap \varepsilon(x)))$ writes as $E(1\!\!1_A|\mathcal{B}_n) \to 1\!\!1_A$, which has just been assumed.

**Corollary 1.8.7"'.** If $X$ is a compact metric space and $\mathcal{F}$ the $\sigma$-algebra of Borel sets (generated by open sets), then if $\sup_{B\in\mathcal{B}_m}\big(\mathrm{diam}(B)\big) \to 0$ as $m \to \infty$, then $\mathrm{h}(T) = \lim_{m\to\infty} \mathrm{H}(T, \mathcal{B}_m)$.

**Proof.** It is sufficient to check $E(1\!\!1_A|\mathcal{B}_m) \to 1\!\!1_A$ in measure. First note that for every $\delta > 0$ there exist an open set $U$ and closed set $K$ such that $K \subset A \subset U$ and $\mu(U \setminus K) \le \delta$. This property is called *regularity* of our measure $\mu$ and is true for every finite measure on the $\sigma$-algebra of Borel sets for a metric space (compactness is not needed here). It can be proved by Caratheodory's argument, compare Proof of Th.1.1.1. Namely we construct the outer measure with the help of open sets, as in the sketch of the proof of theorem 1.7.2 (where we used $\mathcal{G}_0$) and notice that since each closed set is an intersection of a decreasing sequence of open sets we will have the same outer measure if in the construction of outer measure we use the algebra generated by open sets. Now we can refer to Theorem 1.7.2.

Next, due to compactness of $X$, hence $K$, for $m$ large enough the set $A' := \bigcup\{B \in \mathcal{B}_m : B \cap K \neq \emptyset\}$ contains $K$ and is contained in $U$, hence $\mu(A \div A') \le \delta$. This implies that

$$\int_X |E(1\!\!1_A|\mathcal{B}_m) - 1\!\!1_A|\, d\mu =$$

$$\int_{X\setminus(A\cup A')} E(1\!\!1_A|\mathcal{B}_m)\, d\mu + \int_{A\div A'} |E(1\!\!1_A|\mathcal{B}_m) - 1\!\!1_A|\, d\mu + \int_{A\cap A'} 1 - E(1\!\!1_A|\mathcal{B}_m)\, d\mu \le$$

$$\frac{\delta}{\mu(X\setminus A')}\mu(X\setminus(A\cup A')) + \delta + \Big(1 - \frac{\mu(A\cap A')}{\mu(A')}\Big)\mu(A\cap A') \le 3\delta.$$

Hence $\mu\{x : |E(1\!\!1_A|\mathcal{B}_m) - 1\!\!1_A| \ge \sqrt{3\delta}\} \le \sqrt{3\delta}$. ♣

For a simpler proof, omitting Theorem 1.8.6, see Exercise 1.9'.

We end this Section with the *theorem on decomposition into ergodic components* and the adequate entropy formula. Compare this with Choquet representation theorem: Th. 2.1.11, and Th. 2.1.13.

Let $T$ be a measure preserving endomorphism of a Lebesgue space. A measurable partition $\mathcal{A}$ is said to be $T$-invariant if $T(A) \subset A$ for almost every $A \in \mathcal{A}$. The induced map $T_A = T|_A : A \to A$ is a measurable endomorphism of the Lebesgue space $(A, \mathcal{F}_A, \mu_A)$. One calls $T_A$ a *component* of $T$.

**Theorem 1.8.8.** (a) There exists a smallest $T$-invariant measurable partition $\mathcal{A}$ (mod 0) (called the *ergodic decomposition*). Almost all of its components are ergodic.

(b) $h(T) = \int_{X/\mathcal{A}} h(T_A)\, d\mu_{\mathcal{A}}(A)$.

**Proof.** We shall not prove here the part (a). Let us mention only that the ergodic decomposition partition corresponds (see Sec.6) to the completion of $\mathcal{I}$, the $\sigma$-subalgebra of $\mathcal{F}$ consisting of $T$ invariant sets in $\mathcal{F}$ (compare Theorem 1.2.2).

To prove the part (b) notice that for every $T$-invariant measurable partition $\mathcal{A}$, for every finite partition $\xi$ and almost every $A \in \mathcal{A}$, writing $\xi_A$ for the partition $\{s \cap A : s \in \xi\}$, we obtain

$$\mathrm{h}(T_A, \xi_A) = H(\xi_A | \xi_A^-) = \int_A I_{\mu_A}(\xi_A|\xi_A^-)\, d\mu_A.$$

Notice next that the latter information function is equal a.e. to $I_\mu(\xi|\xi^- \vee \mathcal{A})$ restricted to $A$. Hence

$$\int_{X/\mathcal{A}} h(T_A)\, d\mu_{\mathcal{A}}(A) = \int_{X/\mathcal{A}} d\mu_{\mathcal{A}} \int_A I_{\mu_A}(\xi_A|\xi_A^-)\, d\mu_A =$$

$$\int_X I_\mu(\xi|\xi^- \vee \mathcal{A})\, d\mu = \mathrm{H}(\xi|\xi^- \vee \mathcal{A}) = \mathrm{h}(T, \xi)$$

The latter equality follows from an approximation of $\mathcal{A}$ by finite $T$-invariant partitions $\eta \nearrow \mathcal{A}$ and from

$$\mathrm{H}(\xi|\xi^- \vee \eta) = \mathrm{H}(\xi \vee \eta|\xi^- \vee \eta^-) = \lim_{n\to\infty} \frac{1}{n}\mathrm{H}((\xi \vee \eta)^n) =$$

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{H}(\xi^n \vee \eta) = \lim_{n\to\infty} \frac{1}{n}\mathrm{H}(\xi^n) = H(T, \xi).$$

Let now $\xi_n$ be a sequence of finite partitions such that $\xi_n \nearrow \varepsilon$. Then $\mathrm{h}(T, \xi_n) \nearrow \mathrm{h}(T)$ and $\mathrm{h}(T_A, (\xi_n)_A) \nearrow \mathrm{h}(T_A)$. So $\mathrm{h}(T, \xi_n) = \int_{X/\mathcal{A}} h(T_A, \xi_n)\, d\mu_{\mathcal{A}}(A)$ and Lebesgue monotone convergence theorem prove (b) ♣

## §1.9 COUNTABLE TO ONE MAPS, JACOBIAN AND ENTROPY OF ENDOMORPHISMS .

We start with a formulation of a deep theorem by Rohlin:

**Theorem 1.9.1.** Suppose that $\mathcal{A}$ and $\mathcal{B}$ are two measurable partitions of a Lebesgue space $(X, \mathcal{F}, \mu)$ such that $\mathcal{A}|_B$ is countable (mod 0 with respect to $\mu_B$) for almost every $B \in \mathcal{B}$. Then there exists a countable partition $\gamma = \{\gamma_1, \gamma_2, ...\}$ of $X$ (mod 0) such that such that each $\gamma_j \in \gamma$ intersects almost every $B$ at not more than one point, which is then an atom of $\mu_B$, in particular

$$\mathcal{A} \vee \mathcal{B} = \gamma \vee \mathcal{B} \pmod 0.$$

Furthermore, if $\mathrm{H}(\mathcal{A}|\mathcal{B}) < \infty$, then $\gamma$ can be chosen so that

$$\mathrm{H}(\gamma) < H(\mathcal{A}|\mathcal{B}) + 3\sqrt{H(\mathcal{A}|\mathcal{B})} < \infty.$$

**Definition 1.9.2.** Let $(X, \mathcal{F}, \mu)$ be a Lebesgue space. Let $T : X \to X$ be a measurable endomorphism. We say that $T$ is *essentially countable to one* if the measures $\mu_A$ of a canonical system of conditional measures for the partition $\mathcal{A} := T^{-1}(\varepsilon)$ are purely atomic (mod 0 with respect to $\mu_A$), for almost all $A$. We say that $T$ is *countable to one* if we can omit the phrase "mod 0 with respect to $\mu_A$" above.

**Lemma 1.9.3.** If $T$ is essentially countable to one and preserves $\mu$ then there exists a measurable $Y \subset X$ of full measure such that $T(Y) \subset Y$ and

1. $T^{-1}(x) \cap Y$ for a.e. $x \in Y$ is countable, moreover it consists only of atoms of the conditional measure $\mu_{T^{-1}(x)}$;
2. $T(B)$ is measurable if $B \subset Y$ is measurable;
3. $T|_Y$ is forward quasi-invariant, namely $\mu(B) = 0$ for $B \subset Y$ implies $\mu(T(B)) = 0$.

**Proof.** Let $Y'$ be the union of atoms mentioned in Definition 1.9.2.. We can write, due to Theorem 1.9.1, $Y' = \bigcup_j \gamma_j$, so $Y'$ is measurable. Set $Y = \bigcap_{n=0}^{\infty} T^{-n}(Y')$. Denote the partition $T^{-1}(\varepsilon)$ in $Y$ by $\zeta$. Property 1. follows from the construction. To prove 2. we use the fact that $(Y/\zeta, \mathcal{F}_\zeta, \mu_\zeta)$ is a Lebesgue space and the factor map $T_\zeta : Y_\zeta \to X$ is an automorphism (Th.1.6.5). So, for measurable $B \subset Y$, the set

$$\{A \in \zeta : \mu_A(B \cap A) \neq 0\} = \{A \in \zeta : A \cap B \neq \emptyset\} \tag{1.9.1}$$

is measurable by (1.6.3) and therefore its image under $T_\zeta$, equal to $T(B)$, is measurable. If $\mu(B) = 0$, then the set in (1.9.1) has measure $\mu_\zeta$ equal to 0, hence as $T_\zeta$ is isomorphism we obtain that $T(B)$ is measurable, of measure 0. ♣

The key property in the above proof is the equality (1.9.1). Without assuming that $\mu_A$ are purely atomic there could existed $B$ of measure 0 with $C := \{A \in \zeta : \mu_A(B \cap A) \neq 0\}$ not measurable in $\mathcal{F}_\zeta$.

To have such a situation just consider a non-measurable $C \subset Y/\zeta$. Consider the disjoint union $D := C \cup Y$ and denote the embedded $C$ by $C'$. Finally, defining measure on $D$, put $\mu(C') = 0$ and $\mu$ on the embedded $Y$. Define $T(c') = T(C)$ for $C \ni c$ and $c'$ being the image of $c$ under the abovementioned embedding. Thus $C'$ is measurable, of measure 0, whereas $T(C')$ is not measurable because $C$ is not measurable and $T_\zeta$ is isomorphism.

**Definition 1.9.4.** Let $(X, \mathcal{F}, \mu)$ and $(X', \mathcal{F}', \mu')$ be probability measure spaces. Let $T : X \to X'$ be a measurable homomorphism. We say that a real, nonnegative, measurable function $J$ is a *weak Jacobian* if there exists $E$ of measure 0 such that for every measurable $A \subset X \setminus E$ on which $T$ is injective, the set $T(A)$ is measurable and $\mu(T(A)) = \int_A J \, d\mu$. We say $J$ is *strong Jacobian* if the above holds without assuming $A \subset X \setminus E$.

Notice that if $T$ is forward quasi-invariant, namely $(\mu(A) = 0) \Rightarrow (\mu'(T(A)) = 0)$, then automatically weak Jacobian is strong Jacobian.

**Proposition 1.9.5.** Let $(X, \mathcal{F}, \mu)$ be Lebesgue space and $T : X \to X$ be a measurable, essentially countable to one, endomorphism. Then there exists a weak Jacobian $J$. It is unique (mod 0). For $T$ restricted to $Y$ (from Lemma 1.9.3.) $J$ is strong Jacobian.

**Proof.** Consider the partition $\gamma = \{\gamma_1, \gamma_2, ...\}$ given by Theorem 1.9.1. Then for each $j$ the map $T|_{\gamma_j \cap Y}$ is injective. Moreover by Lemma 1.9.3 $T|_{\gamma_j \cap Y}$ maps measurable sets onto measurable sets and is forward quasi-invariant. Therefore $J$ exists on each $\gamma_j \cap Y$ by Radon-Nikodym theorem.

By the presentation of each $A \subset Y$ as $\bigcup_{j=1}^{\infty} A \cap \gamma_j$ the function $J$ satisfies the assertion of the Proposition. The uniqueness follows from the uniqueness of Jacobian in Radon-Nikodym theorem on each $\gamma_j \cap Y$.

**Theorem 1.9.6.** Let $(X, \mathcal{F}, \nu)$ be a Lebesgue space. Let $T : X \to X$ be a $\nu$ preserving endomorphism, essentially countable to one. Then its Jacobian, strong on $Y$ defined in Lemma 1.9.3, weak on $X$, has logarithm equal to $I(\varepsilon|T^{-1}(\varepsilon))$. ($I$ stands for the information function, see Sections 1.4 and 1.8)

**Proof.** Consider already $T$ restricted to $Y$. Let $Z \subset Y$ be an arbitrary measurable set such that $T$ is 1–to–1 on it. For each $y \in Y$ denote by $A(y)$ the element of $\zeta = T^{-1}(\varepsilon)$ containing $y$. We obtain

$$\begin{aligned}
\nu\big(T(Z)\big) = \nu\Big(T^{-1}\big(T(Z)\big)\Big) &= \int_{T^{-1}\big(T(Z)\big)} \left(\int_{A(y)} 1\!\!1 \, d\nu_{A(y)}\right) d\nu(y) = \\
&= \int_{T^{-1}\big(T(Z)\big)} \left(\int_{A(y)} 1\!\!1_Z(y)/\nu_{A(y)}\{y\}) \, d\nu_{A(y)}\right) d\nu(y) = \\
&= \int_{T^{-1}\big(T(Z)\big)} 1\!\!1_Z(y)/\nu_{A(y)}\{y\}) \, d\nu(y) = \int_Z 1/\nu_{A(y)}\{y\}) \, d\nu(y)
\end{aligned}$$

♣

Theorem 1.9.6 gives rise to the so called Rohlin formula:

**Theorem 1.9.7.** Let $(X, \mathcal{F}, \mu)$ be a Lebesgue space. Let $T : X \to X$ be a $\mu$ preserving endomorphism, essentially countable to one. Suppose that on each component $A$ of the ergodic decomposition (cf. Th.1.8.8) the restriction $T_A$ has a countable generator of finite entropy. Then for the Jacobian $J$

$$h_\mu(T) = \mathrm{H}(\varepsilon|T^{-1}(\varepsilon)) = \int I(\varepsilon|T^{-1}(\varepsilon)) \, d\mu = \int \log J \, d\mu.$$

**Proof.** The third equality follows from Theorem 1.9.6, the second one is the definition of the conditional entropy, see Sec. 8. To prove the first equality we can assume, due to Theorem 1.8.8, that $T$ is ergodic. Then, for $\zeta$, a countable generator of finite entropy, with the use of Theorems 1.8.5 and 1.8.6, we obtain

$$\mathrm{H}(\varepsilon|T^{-1}(\varepsilon) = \mathrm{H}(\varepsilon|\zeta^-) = \lim_{n\to\infty} \mathrm{H}(\zeta^n|\zeta^-) = \mathrm{H}(\zeta|\zeta^-) = \mathrm{h}(T, \zeta) = \mathrm{h}(T). \clubsuit$$

**Remark.** The existence of a countable generator is a general, not very difficult, fact, namely the following holds:

**Theorem 1.9.8.** Let $(X, \mathcal{F}, \mu)$ be Lebesgue space. Let $T : X \to X$ be a $\mu$-preserving aperiodic endomorphism, essentially countable to one. Then there exists a *countable generator*, namely a countable partition $\zeta$ such that $\zeta^- = \varepsilon$ (mod 0).

Aperiodic means there exists no $B$ of positive measure and a positive integer $n$ so that $T^n|_B$ =id. For the proof see [Rohlin, 1967, Sec.10.12-13] or [Parry]. To construct $\zeta$ one uses the partition $\gamma$ found for $\varepsilon$ and $T^{-1}(\varepsilon)$ according to Theorem 1.9.1 and so-called *Rohlin towers*.
The existence of a generator with finite entropy is in fact equivalent to $H(\varepsilon|\varepsilon^-) = \mathrm{h}(T) < \infty$. The proof of the implication to the right is contained in Proof of Th.1.9.7. The reverse implication, the construction of the partition, is not easy, it uses in particular the estimate in Th.1.9.1.
The existence of a generator with finite entropy is a strong property. It may fail even for *exact* endomorphisms, see Sec.10 and Exercise 13. Neither its existence implies exactness, Exercise 13. To the contrary, for automorphisms, two-sided generators, even finite, always exist, provided the map is aperiodic.

## §1.10. MIXING PROPERTIES.

In this section we examine briefly some mixing properties of a measure preserving endomorphism, stronger than ergodicity. A measure preserving endomorphism is said to be *weakly mixing* if and only if for every two measurable sets $A$ and $B$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} |\mu(T^{-j}(B) \cap A) - \mu(A)\mu(B)| = 0$$

To see that a weakly mixing transformation is ergodic, suppose that $T^{-1}(B) = B$. Then $T^{-k}(B) = B$ for all $k \geq 0$ and consequently for every $n$, $\frac{1}{n}\sum_{j=0}^{n-1} |\mu(T^{-j}(B) \cap A) - \mu(A)\mu(B)| = |\mu(B) - \mu(B)^2|$. Thus $\mu(B) - \mu(B)^2 = 0$ and therefore $\mu(B) = 0$ or 1.
A measure preserving endomorphism is said to be *mixing* if and only if for every two measurable sets $A$ and $B$

$$\lim_{n\to\infty} \mu(T^{-n}(A) \cap B) - \mu(A)\mu(B) = 0$$

Clearly, every mixing transformation is weakly mixing. The property equivalent to the mixing property is the following: for every square integrable functions $f, g$

$$\lim_{n\to\infty} \int f(g \circ T^n)\, d\mu = \int f\, d\mu \int g\, d\mu.$$

Indeed, the former property follows from the latter one if we substitute the indicator functions $\mathbb{1}_A, \mathbb{1}_B$ in place of $f, g$. To prove the opposite implication notice that with the help of Hölder inequality it is sufficient to restrict to simple functions $f = \sum_i a_i \mathbb{1}_{A_i}, g = \sum_j a_j \mathbb{1}_{A_j}$ for finite partitions $(A_i)$ and $(B_j)$. Then

$$\left| \int f(g \circ T^n)\, d\mu - \int f\, d\mu \int g\, d\mu \right| = \left| \sum_{i,j} a_i b_j (\mu(A_i \cap T^{-n}(B_j)) - \mu(A_i)\mu(B_j)) \right| \to 0$$

because every summand converges to 0 as $n \to \infty$.
In the sequel we will deal also with stronger mixing properties. An endomorphism is called *$K$-mixing* if for every measurable set $A$ and finite partition $\mathcal{A}$

$$\lim_{n\to\infty} \sup_{B \in \mathcal{F}(\mathcal{A}_n^\infty)} |\mu(A \cap B) - \mu(A)\mu(B)| = 0,$$

Recall that $\mathcal{F}(\mathcal{A}_n^\infty)$ for $n \geq 0$ means the complete $\sigma$-algebra assigned to the partition $\mathcal{A}_n^\infty = \bigvee_{j=n}^\infty T^{-j}(\mathcal{A})$. The following theorem provides us with alternative definitions of the $K$-mixing property in case $T$ is an automorphism.

**Theorem 1.10.1.** If $T : X \to X$ is a measure-preserving automorphism of a Lebesgue space, then the following conditions are equivalent:
(a) $T$ is $K$-mixing.
(b) For every finite partition $\mathcal{A}$ $\text{Tail}(\mathcal{A}) := \bigwedge_{n=0}^\infty \bigvee_{k=n}^\infty T^{-k}(\mathcal{A})$ is equal to the trivial partition $\nu = \{X\}$
(c) For every finite partition $\mathcal{A} \neq \nu$, $\mathrm{h}_\mu(T, \mathcal{A}) > 0$ ($T$ has *completely positive entropy*)
(d) There exists a forward invariant exhausting measurable partition $\alpha$ (i.e. satisfying $T^{-1}(\alpha) \leq \alpha$, $T^n(\alpha) \nearrow \varepsilon$, see Def. 1.7.4) such that $T^{-n}(\alpha) \searrow \nu$.

The property $\text{Tail}(\mathcal{A}) = \nu$ is a version of the 0-1 Law. An automorphism satisfying (d) is usually called: *$K$-automorphism*. The symbol $K$ comes from the name: Kolmogorov. Each partition satisfying the properties of $\alpha$ in (d) is called $K$-partition.

**Remark**: the properties (a)-(c) make sense for endomorphisms and they are equivalent (proofs are the same as for automorphisms). Moreover they hold for an endomorphis iff they hold for its natural extension.

**Proof.** (a piece of) To show the reader what is the Theorem about let us prove at least some implications:

(a)⇒(b) Let $A \in \mathcal{F}(\text{Tail}(\mathcal{A}))$ for a finite partition $\mathcal{A}$. Then $A \in \mathcal{F}(\bigvee_{k=n}^\infty T^{-k}(\mathcal{A}))$ for every $n$. Hence, by $K$-mixing, $\mu(A \cap A) - \mu(A)\mu(A) = 0$ and therefore $\mu(A) = 0$ or 1.

(b)⇒(c) Suppose $\mathrm{h}(T, \mathcal{A}) = 0$ for a finite partition $\mathcal{A}$. Then $\mathrm{H}(\mathcal{A}|\mathcal{A}^-) = 0$, hence $I(\mathcal{A}|\mathcal{A}^-) = 0$ a.s. (see Sec.8), hence $\mathcal{A} \leq \mathcal{A}^-$. Hence

$$\bigvee_{k=0}^\infty T^{-k}(\mathcal{A}) = \bigvee_{k=1}^\infty T^{-k}(\mathcal{A}) \text{ and } \bigvee_{k=m}^\infty T^{-k}(\mathcal{A}) = \bigvee_{k=n}^\infty T^{-k}(\mathcal{A})$$

for every $m, n \geq 0$. So $\bigwedge_{n=0}^{\infty} \bigvee_{k=n}^{\infty} T^{-k}(\mathcal{A}) = \bigvee_{k=0}^{\infty} T^{-k}(\mathcal{A})$. The latter partition is Tail$\mathcal{A}$, so it is equal to $\nu$ by (b). But it is finer than $\mathcal{A}$, hence $\mathcal{A} = \nu$. So each finite partition different from $\nu$, the trivial one, has positive entropy.

(c)⇒(d) (in case there exists a finite two-sided generator $\mathcal{B}$, i.e. $\bigvee_{n=-\infty}^{\infty} T^n(\mathcal{B}) = \varepsilon$) $\xi = \bigvee T_{n=0}^{\infty} T^{-n}(\mathcal{B})$ is a $K$-generator. ♣

Let us finish the Section with the following useful:

**Definition 1.10.2.** A measure preserving endomorphism is said to be *exact* if

$$\bigwedge_{n=0}^{\infty} T^{-n}(\varepsilon) = \nu,$$

where $\varepsilon$ is the partition into points and $\nu$ is the trivial partition $\{X\}$.

**Exercise**: Prove that exactness is equivalent to the property: $\mu_e(T^n(A)) \to 1$ for every $A$ of positive measure ($\mu_e$ is outer measure), or to the property: $\mu(T^n(A)) \to 1$ provided $\mu(A) > 0$ and the sets $T^n(A)$ are measurable.

The property *exact* implies the natural extension is a $K$-automorphism (in Theorem 1.10.1(d) set for $\alpha$ the lift of $\varepsilon$). The converse is of course false. Non-one atom space automorphisms are not exact. Observe however that if $T$ is an automorphism and $\alpha$ is a measurable partition satisfying (d), then the factor map $T/\alpha$ on $X/\alpha$ is exact. Exercise: prove that $T$ is the natural extension of $T/\alpha$. Remind finally (Sec. 9) that even for exact endomorphisms $\mathrm{h}(\varepsilon|T^{-1}(\varepsilon))$ can be strictly less than $\mathrm{h}(T)$.

## §1.11. PROBABILITY LAWS AND BERNOULLI PROPERTY.

For $(X, \mathcal{F}, \mu)$ a probability space (or whenever it is needed: a Lebesgue space). Let $f$ and $g$ be real square-integrable functions on $X$. For every positive integer $n$ the $n$-th *correlation* of the pair $f, g$, is the number

$$C_n(f,g) := \int f \cdot (g \circ T^n)\, d\mu - \int f\, d\mu \int g\, d\mu.$$

provided the above integrals exist. Notice that due to the $T$-invariance of $\mu$ we can also write

$$C_n(f,g) = \int (f - Ef)\big((g - Eg) \circ T^n\big)\, d\mu,$$

where we write $Ef = \int f\, d\mu$ and $Eg = \int g\, d\mu$

Let $g : X \to R$ be a square-integrable function. The limit

$$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \int \Big(\sum_{j=0}^{n-1} g \circ T^j - nEg\Big)^2\, d\mu \tag{1.11.1}$$

is called *asymptotic variance* or *dispersion*, provided it exists.

Write $g_0 = g - Eg$. Then we can rewrite the above as
$\sigma^2 = \lim_{n\to\infty} \frac{1}{n} \int \left( \sum_{j=0}^{n-1} g_0 \circ T^j \right)^2 d\mu$.

Another useful expression for the asymptotic variance is the following

$$\sigma^2(g) = \int g_0^2 \, d\mu + 2 \sum_{j=1}^{\infty} \int g_0 \cdot (g_0 \circ T^j) \, d\mu. \tag{1.11.2}$$

The convergence of the series of correlations $C_n(g,g)$ in (1.11.2) easily implies that $\sigma^2(g)$ from this formula is equal to $\sigma^2$ defined in (1.11.1), compare the computation in the proof of Theorem 1.11.3 later on.

We say that the *law of iterated logarithm*, LIL, is satisfied for $g$ if $\sigma^2(g)$ exists (i.e. the above series converges) and

$$\limsup_{n\to\infty} \frac{\sum_{j=0}^{n-1} g \circ T^j - nEg}{\sqrt{n \log\log n}} = \sqrt{2\sigma^2} \qquad \mu - \text{ almost surely }. \tag{1.11.3}$$

($\mu$ almost surely, a.s., means $\mu$ almost everywhere, a.e. This is the probability theory language.)
We say that the *central limit theorem*, CLT, holds, if

$$\mu\left(\left\{ x \in X : \frac{\sum_{j=0}^{n-1} g \circ T^j - nEg}{\sqrt{n}} < r \right\}\right) \to \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{r} e^{-t^2/2\sigma^2} \, dt. \tag{1.11.4}$$

For $\sigma \neq 0$ the convergence is for all $r$, for $\sigma^2 = 0$ the convergence holds for $r \neq 0$ and on the right hand side one sets 0 for $r < 0$ and 1 for $r > 0$.

The LIL and CLT for $\sigma^2 \neq 0$ are often, and this is the case in Theorem 1.11.1 below, a consequence of the *almost sure invariance principle*, ASIP, which says that the sequence of random variables $g$, $g \circ T$, $g \circ T^2$, *centered at the expectation value* i.e. provided $Eg = 0$, is "approximated with the rate $n^{1/2-\varepsilon}$" for an $\varepsilon > 0$, depending on $\delta$ in Theorem 1.11.1 below, by a martingale difference sequence and a respective Brownian motion.

**Theorem 1.11.1.** Let $(X, \mathcal{F}, \mu)$ be a probability space and $T$ an endomorphism preserving $\mu$. Let $\mathcal{G} \subset \mathcal{F}$ be a $\sigma$-algebra. Write $\mathcal{G}_m^n := \bigvee_{j=m}^{n} T^{-j}(\mathcal{G})$ (notation from §.1.6) for $m \le n \le \infty$ and suppose that the following property called *$\phi$-mixing* holds:
There exists a sequence $\phi(n), n = 0, 1, ..$ of positive numbers satisfying

$$\sum_{n=1}^{\infty} \phi^{1/2}(n) < \infty \tag{1.11.5}$$

such that for every $A \in \mathcal{G}_0^m$ and $B \in \mathcal{G}_n^\infty$, $0 \le m \le n$ we have

$$|\mu(A \cap B) - \mu(A)\mu(B)| \le \phi(n-m)\mu(A). \tag{1.11.6}$$

Finally consider a $\mathcal{G}_0^\infty$ measurable function $g : X \to I\!R$ such that

$$\int |g|^{2+\delta}\, d\mu < \infty \quad \text{for some } \delta > 0,$$

and that for all $n \geq 1$

$$\left(\int |h - E(h|\mathcal{G}_0^n)|^{2+\delta}\right)^{-(2+\delta)} \leq Kn^{-s}, \qquad K > 0, s > 0 \text{ large enough.} \tag{1.11.7}$$

(A concrete formula for $s$ can be given, depending on $\delta$.)

Then $g$ satisfies CLT and LIL.

LIL for $\sigma^2 \neq 0$ is a special case, for $\psi(n) = \sqrt{2\log\log n}$, of the following: for every real positive non-decreasing function $\psi$ one has, provided $\int g\, d\mu = 0$,

$$\mu\{x \in X : \sum_{j=0}^{n} g(T^j(x)) > \psi(n)\sqrt{\sigma^2 n} \text{ for infinitely many } n\} = 0 \text{ or } 1$$

according as $\int_1^\infty \frac{\psi(t)}{t} \exp(-\frac{1}{2}\psi^2(t))\, dt$ converges or diverges.

As we already remarked, this Theorem, for $\sigma^2 \neq 0$, is a consequence of ASIP and the similar conclusions for the standard Brownian motion. We do not give the proofs here. For ASIP and further references see [Philipp, Stout, Ch.4,7]. Let us discuss only the existence of $\sigma^2$. It follows from the following consequence of (1.11.6): For $\alpha, \beta$ square integrable real functions on $X$, $\alpha$ measurable in $\mathcal{G}_0^m$ and $\beta$ measurable in $\mathcal{G}_n^\infty$ we have

$$\left|\int \alpha\beta\, d\mu - E\alpha\, E\beta\right| \leq 2(\phi(n-m))^{1/2}\|\alpha\|_2\|\beta\|_2. \tag{1.11.8}$$

The proof of this inequality is not difficult, but tricky, with the use of Hölder inequality, see [Ibragimov] or [Billingsley, 1968]. It is sufficient to work with the functions $\alpha = \sum_i a_i 1\!\!1_{A_i}, \beta = \sum_j a_j 1\!\!1_{A_j}$ for finite partitions $(A_i)$ and $(B_j)$, as with mixing in Sec. 10. Note that if instead of (1.11.6) we have stronger:

$$|\mu(A\cap B) - \mu(A)\mu(B)| \leq \phi(n-m)\mu(A)\mu(B), \tag{1.11.9}$$

as will happen in Ch.3, then we very easily obtain in (1.11.7) the estimate by $\phi(n-m)\|\alpha\|_1\|\beta\|_1$, by the computation the same as for mixing in Sec.10.

We may assume that $g$ is centered at the expectation value. Write $g = k_n + r_n = E(g|\mathcal{G}_0^{[n/2]}) + (g - E(g|\mathcal{G}_0^{[n/2]})$. We have

$$\left|\int g(g\circ T^n)\, d\mu\right| \leq$$

$$\left|\int k_n(k_n\circ T^n)\,d\mu\right| + \left|\int k_n(r_n\circ T^n)\,d\mu\right| + \left|\int r_n(k_n\circ T^n)\,d\mu\right| + \left|\int r_n(r_n\circ T^n)\,d\mu\right| \leq$$

$$2(\phi(n-[n/2]))^{1/2}\|k_n\|_2^2 + 2\|k_n\|_2\|r_n\|_2 + \|r_n\|_2^2 \leq$$

$$2(\phi(n-[n/2]))^{1/2}\|k_n\|_2^2 + 2K[n/2]^{-s}\|k_n\|_2 + K[n/2]^{-2s},$$

the first summand estimated according to (1.11.8). For $s>1$ we obtain convergence of the series of correlations.

Let us go back to the discussion of the $\phi$-mixing. If $\mathcal{G}$ is associated to a finite partition that is a generator, $\phi$-mixing with $\phi(n)\to 0$ as $n\to\infty$ implies $K$-mixing (see Sec.10). Indeed $B$ is the same in both definitions, whereas $A$ in $K$-mixing can be approximated by sets belonging to $\mathcal{G}_0^m$. We leave details to the reader.

Intuitively both notions mean that any event $B$ in remote future weakly depends on the present state $A$, i.e. $|\mu(B)-\mu(B|A)|$ is small.

In applications $\mathcal{G}$ will be usually associated to a finite or countable partition.

In Theorems 1.11.1, the case $\sigma^2=0$ is easy. It relies on Theorem 1.11.3 below. Let us first introduce the following fundamental

**Definition 1.11.2.** Two functions $f,g: X\to \mathbb{R}$ (or $\mathbb{C}$) are said to be *cohomologous* in a space $\mathcal{K}$ of real (or complex) -valued functions on $X$ (or $f$ is called cohomologous to $g$), if there exists $h\in\mathcal{K}$ such that

$$f-g=h\circ T-h. \tag{1.11.10}$$

If $f,g$ are defined mod 0, then (1.11.10) is understood a.s.. This formula is called a *cohomology equation.*

**Theorem 1.11.3.** Let $f$ be a square integrable function on a probability space $(X,\mathcal{F},\mu)$, centered at the expectation value. Assume that $\sum_{n=0}^{\infty} n|\int f\cdot(f\circ T^n)\,d\mu|<\infty$. Then the following three conditions are equivalent:

(a) $\sigma^2(f)=0$;

(b) All the sums $S_n=S_nf=\sum_{j=0}^{n-1} f\circ T^j$ have the norm in $L^2$ (the space square integrable functions) bounded by the same constant;

(c) $f$ is cohomologous to 0 in the space $\mathcal{H}=L^2$.

**Proof.** (c)$\Rightarrow$(a) follows immediately from (1.11.1) after substituting $f=h\circ T-h$. Let us prove (a)$\Rightarrow$(b): Write $C_j$ for the correlation $\int f\cdot(f\circ T^j)\,d\mu, j=0,1,...$. Then

$$\int |S_n|^2\,d\mu = nC_0^2 + 2\sum_{j=1}^{n}(n-j)C_j$$

$$= n\Big(C_0^2+2\sum_{j=1}^{\infty}C_j\Big) - 2n\sum_{j=n+1}^{\infty}C_j - 2\sum_{j=1}^{n} j\cdot C_j = n\sigma^2 - I_n - II_n.$$

Since $I_n \to 0$ and $II_n$ stays bounded as $n \to \infty$ and $\sigma^2 = 0$, we deduce that all the sums $S_n$ are uniformly bounded in $L^2$.

(b)⇒(c): $f = h \circ T - h$ for any $h$ a limit in weak*-topology of the bounded sequence $\frac{1}{n} S_n$. We leave the easy computation to the reader. (This computation will be provided in detail in the similar situation of Bogolyubov-Krylov Theorem, in 2.1.14.). ♣

Now Theorem 1.11.1 for $\sigma^2 = 0$ follow from (c), which gives $\sum_{j=0}^{n-1} f \circ T^j = h \circ T^n - h$, with the use of Borel-Cantelli lemma.

**Remark..** Th.1.11.1 in the two-sided case: where $g$ depens on $\mathcal{G}_j = T^j(\mathcal{G})$ for $j = ..., -1, 0, 1, ...$ for an automorphism $T$, also holds. In 1.11.8 one should replace $\mathcal{G}_0^n$ by $\mathcal{G}_{-n}^n$

Given two finite partitions $\mathcal{A}$ and $\mathcal{B}$ of a probability space and $\varepsilon \geq 0$ we say that $\mathcal{B}$ is *$\varepsilon$-independent* of $\mathcal{A}$ if there is a subfamily $\mathcal{A}' \subset \mathcal{A}$ such that $\mu(\bigcup \mathcal{A}') > 1 - \varepsilon$ and for every $A \in \mathcal{A}'$

$$\sum_{B \in \mathcal{B}} \left| \frac{\mu(A \cap B)}{\mu(A)} - \mu(B) \right| \leq \varepsilon. \tag{1.11.11}$$

Given an ergodic measure preserving endomorphism $T : X \to X$ of a Lebesgue space, a finite partition $\mathcal{A}$ is called *weakly Bernoulli* (abbr. WB) if for every $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that the partition $\bigvee_{j=n}^{s} T^{-j}(\mathcal{A})$ is $\varepsilon$-independent of the partition $\bigvee_{j=0}^{m} T^{-j}(\mathcal{A})$ for every $0 \leq m \leq n \leq s$ such that $n - m \geq N$.
Of course in the definition of $\varepsilon$-independence we can consider any measurable (maybe uncountable) partition $\mathcal{A}$ and write conditional measures $\mu_A(B)$ in (1.11.11). Then for $T$ an automorphism we can replace in the definition of WB $\bigvee_{j=n}^{s} T^{-j}(\mathcal{A})$ by $\bigvee_{j=0}^{s-n} T^{-j}(\mathcal{A})$ and $\bigvee_{j=0}^{m} T^{-j}(\mathcal{A})$ by $\bigvee_{j=-n}^{m-n} T^{-j}(\mathcal{A})$ and set $n = \infty, n - m \geq N$. WB in this formulation becomes one more version of weak dependence of present (and future) from remote past. If $\varepsilon = 0$ and $N = 1$ then all partitions $T^{-j}(\mathcal{A})$ are mutually independent (recall that $\mathcal{A}, \mathcal{B}$ are called independent if $\mu(A \cap B) = \mu(A)\mu(B)$ for every $A \in \mathcal{A}, B \in \mathcal{B}$.). We say then that $\mathcal{A}$ is Bernoulli. If $\mathcal{A}$ is a generator (two-sided generator), then clearly $T$ on $(X, \mathcal{F}, \mu)$ is isomorphic to one-sided (two-sided) Bernoulli shift of $\sharp\mathcal{A}$ symbols, see Chapter 0, Examples 0.8. The following famous theorem of Friedman and Ornstein holds:

**Theorem 1.11.4.** If $\mathcal{A}$ is a weakly Bernoulli finite two-sided generating partition of $X$ for an automorphism $T$, then $T$ is isomorphic to a two-sided Bernoulli shift.

Of course the standard Bernoulli partition (in particular the number of its states) in the above Bernoulli shift can be different from the image under the isomorphism of the WB partition.

Bernoulli shift above is unique in the sense that each two two-sided Bernoulli shifts of the same entropy are isomorphic [O].

Note that $\phi$-mixing in the sense (1.11.9), with $\phi(n) \to 0$, for $\mathcal{G}$ associated to a finite partition $\mathcal{A}$, implies weak Bernoulli.

Central Limit Theorem is a much weaker property than LIL. We want to end this Section with a useful abstract theorem that allows us to deduce CLT for $g$ without specifying $\mathcal{G}$. This Theorem similarly as Theorem 1.11.1 can be proved with the use of an approximation by a martingale difference sequence.

**Theorem 1.11.5.** Let $(X, \mathcal{F}, \mu)$ be a probability space and $T : X \to X$ an automorphism preserving $\mu$. Let $\mathcal{F}_0 \subset \mathcal{F}$ be a $\sigma$-algebra such that $T^{-1}(\mathcal{F}_0) \subset F_0$. Denote $\mathcal{F}_n = T^{-n}(\mathcal{F}_0)$ for all integer $n = ..., -1, 0, 1, ...$ Let $g$ be a real square integrable function. If

$$\sum_{n \geq 0} \|E(g|\mathcal{F}_n)\|_2 + \|g - E(g|\mathcal{F}_{-n})\|_2 < \infty,$$

then $g$ satisfies CLT.

**Exercises.**

0. Prove that for any two $\sigma$-algebras $\mathcal{F} \ni \mathcal{F}'$ and $\phi$ an $\mathcal{F}$-measurable function, the conditional expectation value operator $L^p(X, \mathcal{F}, \mu) \ni \phi \to E(\phi|\mathcal{F}')$ has norm 1 in $L^p$, for every $1 \leq p \leq \infty$. (Hint: Prove that $E((\vartheta \circ |\phi|)|\mathcal{F}') \geq \vartheta \circ E((|\phi|)|\mathcal{F}')$ for convex $\vartheta$, in particular for $t \mapsto t^p$.)

1. Let $T$ be an ergodic automorphism of a probability non-atomic measure space and $\mathcal{A}$ its partition into orbits $\{T^n(x), n = ..., -1, 0, 1...\}$. Prove that $\mathcal{A}$ is not measurable.

Suppose we do not assume ergodicity of $T$. What is the largest measurable partition, smaller than the partition into orbits? (Hint: Th.1.8.8.)

2. Prove that the following partitions of measure spaces are not measurable:

(a) Let $T : S^1 \to S^1$ be a mapping of the unit circle with Haar measure defined by $T(z) = e^{2\pi i\alpha} z$ for an irrational $\alpha$. $\mathcal{P}$ is the partition into orbits;

(b) $T$ is the automorphism of the 2-dimensional torus $\mathbb{R}^2/\boldsymbol{Z}^2$, given by a hyperbolic integer matrix of determinant 1. Let $\mathcal{P}$ be the partition into stable, or unstable, lines (i.e. straight lines parallel to an eigenvector of the matrix);

(c) Let $T : S^1 \to S^1$ be defined by $T(z) = z^2$. Let $\mathcal{P}$ be the partition into *grand orbits*, i.e. equivalence classes of the relation $x \sim y$ iff $\exists m, n \geq 0$ such that $T^m(x) = T^n(y)$.

3. Prove that every Lebesgue space is isomorphic to the unit interval equipped with the Lebesgue measure together with countably many atoms.

4. Prove that every separable complete metric space with a measure on the $\sigma$-algebra containing all open sets, minimal among complete measures, is Lebesgue space.

Hint: [Rohlin 1949, 2.7].

5. Let $(X, \mathcal{F}, \mu)$ be a Lebesgue space. Then $Y \subset X, \mu_e(Y) > 0$ is measurable iff $(Y, \mathcal{F}_Y, \mu_Y)$ is Lebesgue, where $\mu_e$ is the outer measure, $\mathcal{F}_Y = \{A \cap Y : A \in \mathcal{F}\}$ and $\mu_Y(A) = \frac{\mu_e(A \cap Y)}{\mu_e(Y)}$.

Hint: If $\mathbf{B}$=$(B_n)$ is a basis for $(X, \mathcal{F}, \mu)$, then $B'_n = B_n \cap Y)$ is a basis for $(Y, \mathcal{F}_Y, \mu_Y)$. Add to $Y$ one point for each sequence $(B'_n)^{\varepsilon_n}$ whose intersection is missing in $Y$ and in the space $\tilde{Y}$ obtained in such a way generate complete measure space $(\tilde{Y}, \tilde{\mathcal{F}}, \tilde{\mu})$ from the extension $\tilde{\mathbf{B}}$ of the basis $(B'_n)$. Borel sets with respect to $\mathbf{B}$ in $X$ correspond to Borel sets with respect to $\tilde{\mathbf{B}}$ and sets of $\mu$ measure 0 correspond to sets of $\tilde{\mu}$ measure 0. So measurability of $Y$ implies $\tilde{\mu}(\tilde{Y} \setminus Y) = 0$.

6. Prove Th.1.6.3.

Hint: In the case both spaces are unit intervals with standard Lebesgue measure, consider all intervals $J'$ with rational ends. $J = T^{-1}(J')$ is contained in a Borel set $B_J$ with $\mu(B_J \setminus J) = 0$. Remove from $X$ a Borel set of measure 0 containing $\bigcup_J (B_J \setminus J)$. Then $T$ becomes a Borel map, hence it is a Baire function, hence due to the injectivity it maps Borel sets to Borel sets.

7. (a) Consider the unit square $[0,1] \times [0,1]$ equipped with Lebesgue measure. For each $x \in [0,1]$ let $\mathcal{A}_x$ be the partition into points $(x', y)$ for $x' \neq x$ and the interval $\{x\} \times [0,1]$. What is $\bigwedge_x \mathcal{A}_x$ ? Let $\mathcal{B}_x$ be the partition into the intervals $\{x'\} \times [0,1]$ for $x' \neq x$ and the points $\{(x,y) : y \in [0,1]\}$. What is $\bigwedge_x \mathcal{B}_x$ ?

(b) Find two measurable partitions $\mathcal{A}, \mathcal{A}'$ of a Lebesgue space such that their set-theoretic intersection (i.e. the largest partition such that $\mathcal{A}, \mathcal{A}'$ are finer than it) is not measurable.

8. Find an example of $T : X \to X$ an endomorphism of a probability space $(X, \mathcal{F}, \mu)$, injective and onto, such that for the system ... $\xrightarrow{T} X \xrightarrow{T} X$, natural extension does not exist.

Hint: Set $X$ the unit circle and $T$ irrational rotation. Let $A$ be a set consisting of exactly one point in each $T$-orbit. Set $B = \bigcup_{j \geq 0} T^j(A)$. Notice that $B$ is not Lebesgue measurable and that the outer measure of $B$ is 1 (use *unique ergodicity* of $T$, i.e. that (1.2.1a) holds for every $x$)

Let $\mathcal{F}$ be the $\sigma$-algebra consisting of all the sets $C = B \cap D$ for $D$ Lebesgue measurable, set $\mu(C) = \mathrm{Leb}(D)$, and of $C \subset X \setminus B$, set then $\mu(C) = 0$. Note that $\bigcap_{n \geq 0} T^n(B) = \emptyset$ and in the set-theoretic inverse limit the set $\pi_{-n}^{-1}(B) = \pi_0^{-1}(T^n(B))$ would be of measure 1 for every $n \geq 0$.

9. (a) Prove that in a Lebesgue space $d(\mathcal{A}, \mathcal{B}) := H(\mathcal{A}|\mathcal{B}) + H(\mathcal{B}|\mathcal{A})$ is a metric in the space $Z$ of countable partitions (mod 0) of finite entropy. Prove that the metric space $(Z, d)$ is separable and complete.

(b) Prove that if $T$ is an endomorphism of the Lebesgue space, then the function $\mathcal{A} \to \mathrm{h}(T, \mathcal{A})$ is continuous for $\mathcal{A} \in Z$ with respect to the above metric $d$.

Hint: $|\mathrm{h}(T, \mathcal{A}) - \mathrm{h}(T, \mathcal{B})| \leq \max\{\mathrm{H}(\mathcal{A}|\mathcal{B}), \mathrm{H}(\mathcal{B}|\mathcal{A})\}$. Compare Proof of Th.1.4.5.

9'. (a) Let $d_0(\mathcal{A},\mathcal{B}) := \sum_i \mu(A_i \div B_i)$ for partitions of a probability space into $r$ measurable sets $\mathcal{A} = \{A_i, i = 1, ..., r\}$ and $\mathcal{B} = \{B_i, i = 1, ..., r\}$. Prove that for every $r$ and every $d > 0$ there exists $d_0 > 0$ such that if $\mathcal{A},\mathcal{B}$ are partitions into $r$ sets and $d_0(\mathcal{A},\mathcal{B}) < d_0$, then $d(\mathcal{A},\mathcal{B}) < d$

(b) Using (a) give a simple proof of Corollary 1.8.7"'. (Hint: Given an arbitrary finite $\mathcal{A}$ construct $\mathcal{B} \le \mathcal{B}_m$ so that $d_0(\mathcal{A},\mathcal{B})$ be small for $m$ large. Next use (a) and Theorem 1.4.4.d).

10. Prove that there exists a finite generator for every $T$, a continuous positively expansive map of a compact metric space (see the definition of *positively expansive* in Ch.2, Sec.5).

11. Compute the entropy $\mathrm{h}(T)$ for Markov shifts.

12. Prove that the entropy $\mathrm{h}(T)$ defined either as supremum of $\mathrm{H}(T,\mathcal{A})$ over finite partitions, or over countable partitions of finite entropy, or as $\sup \mathrm{H}(\xi|\xi^-)$ over all measurable partitions $\xi$ that are forward invariant (i.e. $T^{-1}(\xi) \le \xi$) is the same.

13. Let $T$ be an endomorphism of the 2-dimensional torus $I\!R^2/\boldsymbol{Z}^2$, given by an integer matrix of determinant larger than 1 and with eigenvalues $\lambda_1, \lambda_2$ such that $|\lambda_1| < 1$ and $|\lambda_2| > 1$.

Let $S$ be the endomorphism of $I\!R^2/\boldsymbol{Z}^2$ being the cartesian product of $S_1(x) = 2x$ (mod 1) on the circle $I\!R/\boldsymbol{Z}$ and of $S_2(y) = y + \alpha$ (mod 1), the rotation by an irrational angle $\alpha$. Which of the maps $T, S$ is exact? Which has a countable generator of finite entropy?

(Answer: $T$ does not have the generator, but it is exact. The latter holds because for each small parallelepiped $P$ spanned by the eigendirections there exists $n$ such that $T^n(P)$ covers the torus with multiplicity bounded by a constant not depending on $P$. See ??? $S$ is not exact, but it is ergodic and has a generator.)

14. Prove that if the definition of partition $\mathcal{A}$ $\varepsilon$-independent of partition $\mathcal{B}$ is replaced by $\sum_{A\in\mathcal{A},B\in\mathcal{B}} |\mu(A\cap B) - \mu(A)\mu(B)|$, then the definition of weakly Bernoulli is equivalent to the old one. (Note that now the expression is symmetric with respect to $\mathcal{A},\mathcal{B}$.)

**Bibliographical notes:**

For the Martingale Convergence Theorem see for example [Doob], [Billingsley, 1979], [Petersen] or [Stroock]. Its standard proof goes via a *maximal function.* We followed this way in Proof of Shannon, McMillan, Breiman Theorem in Sec.5, L.1.5.1, where we relied on [Petersen] and [Parry]. Remark 1.1.2 is taken from [Neveu,Ch.4.3], see for example [Hoover] for a more advanced theory.

Standard proofs of Birkhoff's Ergodic Theorem also use the idea of maximal function. This concerns in particular the extremaly simple proof in Sec. 2, which has been taken from [KH].

For the material of Sec.6 and related exercises see [Rohlin, 1949]. It is also written in an elegant and a very concise way in [Cornfeld, Fomin, Sinai].

The consideration in Sec.7 leading to the extension of the compatible family $\tilde{\mu}_{\Pi,n}$ to $\tilde{\mu}_{\Pi}$ is known as Kolmogoroff Theorem on the existence of stochastic process. First, one verifies $\sigma$-additivity of a measure on an algebra, next uses the Extension Theorem 1.7.2. Our proof of $\sigma$-additivity of $\tilde{\mu}$ on $\tilde{X}$ via Lusin theorem is also a variant of Kolmogoroff's proof. The proofs of $\sigma$-additivity on algebras depend unfortunately on topologocal concepts. Halmos wrote [Halmos, p. 212]: "this peculiar and somewhat undesirable circumstance appears to be unavoidable" Indeed the $\sigma$-additivity may be not true, see [Halmos, p.214]. Our example of non-existence of natural extension, Exercise 8, is in the spirit of Halmos' example. There might be troubles even with extending a measure from cylinders in product of two measure spaces, see [MR] for a counterexample. On the other hand product measures extend to generated $\sigma$-algebras without any additional assumptions [Halmos], [Billingsley, 1979].

For Th.1.8.1: the existence of a countable $\gamma$ such that $\mathcal{A} \vee \mathcal{B} = \gamma \vee \mathcal{B}$, see [Rohlin, 1949]; for the estimate that follows, see for example [Rohlin 1967] or [Parry]. The simple proof of Th.1.8.6 via convergence in measure has been taken from [Rohlin 1967] and [Wa]. Proof of Th.1.8.8 (b) is taken from [Rohlin, 1967, sec.8.10-11 and 9.8].

For Th.1.9.6 see [Parry, L.10.5]; our proof is different. For the construction of generator and 2-sided generator see again [Rohlin 1967],[Parry] or [CFS]. The same are references to the theory of measurable invariant partitions: exhausting and extremal, and to Pinsker partition, which we omitted because we do not need these notions further in the book, but which are fundamental to understand deeper the measure-theoretic entropy theory. Finally we encourage the reader to become acquainted with spectral theory in relation to mixing properties [CFS].

Th. 1.11.1 can be found in [Philipp, Stout]. See also [PUZ, Part 1.]. For (1.11.8) see [Ibragimov, 1.1.2] or [Billingsley, 1968]. For Th.1.11.3 see [Leonov], [Ibragimov, 1.5.2] or [PUZ, Part 1., L.1]. Theorem 1.11.5 can be found in [Gordin]

**References.**


[Billingsley, 1965] P. Billingsley: Ergodic Theory and Information. Wiley, New York, 1965.

[Billingsley, 1979] P. Billingsley: Probability and Mesure. Wiley, New York, 1979.

[CFS] I.P.Cornfeld, S.V.Fomin, Ya.G.Sinai: Ergodic Theory. Springer-Verlag, New York, 1982.

[Carathéodory] C. Carathéodory: Vorlesungen über reelle Funktionen. Leipzig-Berlin, 1927.

[Doob] J.L. Doob: Stochastic Processes, Wiley, New York, 1953

[Gordin]

[Hoover] D.N. Hoover: Convergence in distribution ond Skorokhod Convergence for the general theory of processes. Probab. Th. Rel. Fields 89 (1991), 239-259.

[Ibragimov] I.A. Ibragimov: Some limit theorems for stationary processes. Theor. Prob. Appl. 7 (1962), 349-382.

[KH] A. Katok, B. Hasselblatt: Introduction to the Modern Theory of Dynamical Systems. Cambridge University Press, 1995

[Kolmogoroff] A.N. Kolmogoroff: Gundbegriffe der Wahrscheinlichkeitsrechnung. Berlin 1933.

[Lang] S. Lang: Algebra. Addison-Wesley, 1970.

[Leonov] V.P. Leonov: On the dispersion of time-dependent means of stationary stochastic process. Theor. Prob. Appl. 6 (1961), 93-101.

[MR] E. Marczewski, C. Ryll-Nardzewski:

[Neveu] Jacques Neveu: Bases Mathématiques du Calcul des Probabilités. Masson et Cie, Paris, 1964.

[Parry] W.Parry: Entropy and Generators in Ergodic Theory. W.A.Benjamin, Inc., New York, 1969.

[Petersen] K. Petersen: Ergodic Theory. Cambridge University Press, 1983.

[Philipp, Stout] W.Philipp, W.Stout: Almost sure invariance principles for partial sums of weakly dependent random variables. Mem.AMS 161 (1975).

[PUZ] F. Przytycki, M. Urbański, A. Zdunik: Harmonic, Gibbs and Hausdorff measures for holomorphic maps. Part 1. Anals of Math. 130 (1989), 1-40. Part 2. Studia Math. 97 (1991), 189-225.

[Rohlin, 1949] V.A.Rohlin: On the fundamental ideas of measure theory. Mat.Sb. 67.1 (1949), 107-150. Amer. Math. Soc. Trans. (1)10 (1962), 1-54.

[Rohlin, 1967] V.A.Rohlin: Lectures on the entropy theory of transformations with invariant measure. Uspehi Mat. Nauk 22.5 (1967), 3-56. Russian Math. Surveys, 22.5 (1967), 1-52.

[Stroock] D.W.Stroock: Probability Theory, an Analytic View. Cambridge University Press, 1993.

[Wa] - P. Walters: An Itroduction to Ergodic Theory. Springer-Verlag, New York, 1982.