# 遍历论

## 目录

# 1 预备知识与遍历定理

## 1.1 测度与积分

### 1.1.1 $\sigma$-代数与测度

**定义 1.1.** 设 $X$ 是一个非空集合，$\mathcal{B} \subset 2^X$，(其中 $2^X$ 表示全体 $X$ 的子集构成的集合)，满足

1. $\emptyset \in \mathcal{B}$；
2. 若 $A \in \mathcal{B}$ 则 $X \backslash A \in \mathcal{B}$；
3. 若 $B_n \in \mathcal{B}$ $(n \geq 1)$，则 $\cup_{n\geq 1} B_n \in \mathcal{B}$.

那么就称 $\mathcal{B}$ 是 $X$ 上的一个 $\sigma$-代数.

**例 1.2.** $\mathcal{B}_0 = \{\emptyset, X\}$ 在集合的包含关系下是最小的 $\sigma$-代数. 类似有 $\mathcal{B}_1 = 2^X$ 是在集合的包含意义下最大的 $\sigma$-代数.

**定义 1.3.** 任给 $\mathcal{C} \subset 2^X$，我们定义 $\bigcap_{\mathcal{B}\supset\mathcal{C}} \mathcal{B}$ 是包含 $\mathcal{C}$ 的最小的 $\sigma$-代数，记作 $\mathcal{B}(\mathcal{C})$.

这里相当于 $\mathcal{B}(\mathcal{C})$ 是 $\mathcal{C}$ 中的元素作可数次交并补运算得到的集族.

**定义 1.4.** 设 $(X, \tau)$ 是一个拓朴空间，其中 $\tau \subset 2^X$ 表示开集构成的集族. 在不引起歧义的情况下，我们记 $\mathcal{B}(\tau) = \mathcal{B}(X)$ 是 $(X, \tau)$ 上的 Borel $\sigma$-代数.

**问题 1.5.** 构造一个 $[0, 1]$ 区间上的非 Borel Lebesgue 可测集.

**例 1.6.** 设 $X = [0, 1]$，$\mathcal{B} = \mathcal{L}$ 是 $[0, 1)$ 中的所有 Lebesgue 可测集，则 $\mathcal{L}$ 为 $X$ 上的 $\sigma$-代数.

**例 1.7.** $(\mathbb{R}, \mathcal{L})$ ($\mathcal{L}$ 是 $\mathbb{R}$ 上的 Lebesgue 测度) 为一个可测空间. 任给 $X \subset \mathbb{R}$ 非空 (未必可测) 那么有 $\mathcal{B} = \{L \cap X : L \in \mathcal{L}\}$ 为 $X$ 上的 $\sigma$-代数.

**定义 1.8.** 若 $\mathcal{B}$ 为 $X$ 上的 $\sigma$-代数，则称 $(X, \mathcal{B})$ 为一个可测空间.

**定义 1.9.** 设 $(X, \mathcal{B})$ 为一个可测空间，$m : \mathcal{B} \to [0, +\infty]$ 满足

1. (空集为零测集) $m(\emptyset) = 0$；
2. (可数可加性质) 若 $A_n \in \mathcal{B}$ $(n \geq 1)$ 且 $A_n$ 互不相交，则

$$m\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n=1}^{\infty} m(A_n).$$

那么就称 $m$ 为 $(X, \mathcal{B})$ 上的一个测度.

**例 1.10.** Lebesgue 测度 $m$ 是 $(\mathbb{R}, \mathcal{L})$ 上的一个测度.

**定义 1.11.** 对于测度空间 $(X, \mathcal{B})$，如果 $0 < m(X) < \infty$，则称 $m$ 是有限测度.

**定义 1.12.** 若 $m(X) = 1$，则称 $m$ 为 $(X, \mathcal{B})$ 的概率测度. 此时 $(X, \mathcal{B})$ 称作概率空间.

**例 1.13.** $([0, 1), \mathcal{L}[0, 1))$ 为一个概率空间，其中 $\mathcal{L}[0, 1)$ 表示集合 $[0, 1)$ 上的 Lebesgue 可测集.

**例 1.14.** 设 $X = [0, 1)$，$p \in X$ 定义 $\delta_p : 2^X \to \{0, 1\}$ 为 $\delta_p(L) = 0$ 若 $p \notin L$，$\delta_p(L) = 1$ 若 $p \in L$. 可以验证 $\delta_p$ 是一个概率测度，这样的测度被称为 $p$ 点处的 Dirac 测度.

**定义 1.15.** 设 $X$ 是一个集合，$\mathcal{C}\subset 2^X$ 满足

1. $\emptyset\in\mathcal{C}$;
2. 若 $A,B\in\mathcal{C}$ 则 $A\cap B\in\mathcal{C}$;
3. 若 $A\in\mathcal{C}$ 则存在有限个 $E_i\in\mathcal{C}$ $i=1,2,...,n$ 互不相交，使得 $X\backslash A=\cup_{i=1}^n E_i$.

那么则称 $\mathcal{C}$ 为 $X$ 上的半代数. 这里特别的如果第 3 条中的 $n=1$ 那么此时 $\mathcal{C}$ 就是代数.

这里半代数的定义和代数的区别就在于补给是有限的交集，不一定是落在这个集族里面. 代数和 $\sigma$-代数的差距在于可数并是否还落在集族里面.

**例 1.16.** $X=[0,1)$，令 $\mathcal{C}=\{[a,b):0\le a\le b<1,$ 这里约定如果$a=b$ 就是空集$\}$. 那么这样我们定义的 $\mathcal{C}$ 是一个半代数而且定义里面的 $n=2$.

**定义 1.17.** 设 $\mathcal{C}\subset 2^X$ 是一个半代数，令 $\mathcal{A}(\mathcal{C})=\cap_{\mathcal{C}\subset\mathcal{A}\text{ 是代数}}\mathcal{A}$ 是包含 $\mathcal{A}$ 的最小的代数.

**定理 1.18.** *设 $\mathcal{C}\subset 2^X$ 是一个半代数，则*

$$\mathcal{A}(\mathcal{C})=\left\{\bigcup_{i=1}^n E_i: E_i\in\mathcal{C}, i=1,2,...,n \text{ 互不相交}\right\}$$

**定义 1.19.** $X$ 的集族 $\mathcal{M}$ 成为单调类，如果

1. 对于单调增加的集合序列的并封闭，即

$$E_1\subset E_2\subset ...,\ E_i\in\mathcal{M}\Rightarrow\bigcup_{i=1}^\infty E_i\in\mathcal{M};$$

2. 对于单调减少的集合序列的交封闭，即

$$F_1\supset F_2\supset ...,\ F_i\in\mathcal{M}\Rightarrow\bigcap_{i=1}^\infty F_i\in\mathcal{M}.$$

**定义 1.20.** 设 $\mathcal{A}$ 是 $X$ 上的代数，则称 $\mathcal{M}(\mathcal{A})=\cap_{\mathcal{A}\subset\mathcal{M}\text{ 是单调类}}\mathcal{M}$ 是由 $\mathcal{A}$ 生成的最小单调类.

**定理 1.21** (代数生成的最小单调类是生成的代数)**.** *设 $\mathcal{A}\subset 2^X$ 是代数，那么 $\mathcal{M}(\mathcal{A})=\mathcal{B}(\mathcal{A})$.*

从上面的定理我们看出，对于集合 $X$ 的半代数 $\mathcal{C}$，通过有限的并集得到所生成的代数 $\mathcal{A}(\mathcal{C})$，通过取最小单调类 $\mathcal{M}(\mathcal{A})$，就可以得到生成的 $\sigma$-代数.

**定理 1.22** (测度的延拓)**.** *设 $X$ 是一个集合. 则*

*1.* 设 $\mathcal{C}\subset 2^X$ 为 $X$ 上的半代数，则若函数 $\tau_0:\mathcal{C}\to[0,1)$ 可数可加并且对空集取值 0，则存在唯一的 $\tau_1$，满足 $\tau_1:\mathcal{A}(\mathcal{C})\to[0,1]$ 可数可加并且 $\tau_1|_{\mathcal{C}}=\tau_0$.

*2.* 设 $\mathcal{A}\subset 2^X$ 为 $X$ 上的半代数，则若函数 $\tau_1:\mathcal{A}\to[0,1)$ 可数可加并且对空集取值 0，则存在唯一的 $\tau_2$，满足 $\tau_2:\mathcal{A}(\mathcal{C})\to[0,1]$ 可数可加并且 $\tau_2|_{\mathcal{A}}=\tau_1$.

**定义 1.23** (概率空间的完备化)**.** 设 $(X,\mathcal{B},m)$ 是概率空间，其的完备化

$$\overline{\mathcal{B}}=\{B\cup C: B\in\mathcal{B},\ \exists Z\in\mathcal{B},\ m(Z)=0\ \text{且}C\subset Z\}.$$

相当于是我们把零测集的子集都定义为测度为 0 的可测集合.

在本节最后，我们定义概率空间的乘积. 先从两个开始，任意有限个可以利用归纳定义. 设 $(X_i,\mathcal{B}_i,m_i)$ $i=1,2$ 为两个概率空间，对于 $X=X_1\times X_2$, 我们定义

$$\mathcal{C}=\{A_1\times A_2 : A_1\in\mathcal{B}_1,\ A_2\in\mathcal{B}_2\}.$$

为半代数. 于是可以生成唯一的 $\sigma$-代数, 另外我们定义 $\tau_0:\mathcal{C}\to[0,1]$, $\tau_0(A_1\times A_2)=m_1(A_1)m_2(A_2)$. 利用定理 1.22 可以唯一延拓为 $\mathcal{B}(\mathbb{C})$ 上的概率测度，于是在没有歧义的情况下，我们令 $\mathcal{B}_1\times\mathcal{B}_2=\mathcal{B}(\mathcal{C})$，有的时候为了避免歧义，记作 $\mathcal{B}_1\otimes\mathcal{B}_2$.

受此启发，我们考虑无穷的情形. 设 $(X_i,\mathcal{B}_i,m_i)$ 为概率空间，$i\in\mathbb{Z}$，考虑集合

$$X=\prod_{i=-\infty}^{+\infty}X_i=\{(x_i)_{i=-\infty}^{+\infty}:x_i\in X_i\}.$$

对于任意的 $n\in\mathbb{Z}$，取 $B_j\in\mathcal{B}_j$ $(-n\le j\le n, j\in\mathbb{Z})$，如下定义集合

$$\prod_{i=-\infty}^{-(n+1)}X_i\times\prod_{i=-n}^{n}B_j\times\prod_{i=n+1}^{+\infty}X_i=\{(x_i)_{i=-\infty}^{+\infty}:x\in B_j\}$$

成为可测矩形. 所有的可测矩形形成一个半代数，类似之前的讨论可以生成唯一的 $\sigma$-代数. 另外我们定义 $\tau_0:S\to[0,1]$ 如下

$$\tau\left(\prod_{i=-\infty}^{-(n+1)}X_i\times\prod_{i=-n}^{n}B_j\times\prod_{i=n+1}^{+\infty}X_i\right)=\prod_{i=-n}^{n}m_j(B_j).$$

可以得到 $\tau_0$ 是可数可加的, 那么可以唯一扩张为 $(X,\mathcal{B})$ 上，此时扩张得到的测度 $m$ 被称为积测度. 称 $(X,\mathcal{B},m)$ 为积概率空间，类似可以定义 $\prod_{i=0}^{+\infty}(X_i,\mathcal{B}_i,m_i)$.

**例 1.24** (Bernoulli 测度)**.** 设 $k$ 是一个正整数，$Y=\{0,1,2,...,k-1\}$. 其上的 $\sigma$-代数为 $2^Y$. 取定一个概率向量数组 $\{p_i\}_{i=0}^{k-1}$ 满足 $\sum_{i=0}^{k-1}p_i=1$ 并且 $p_i>0$ 对 $i=0,1,...,k-1$. 定义 $(Y,2^Y)$ 的测度 $\mu:2^Y\to[0,1]$ 由 $\mu(i)=p_i$ $(i=0,1,2,...,k-1)$ 给出. 由概率空间 $(Y,2^Y,\mu)$ 形成的积概率空间

$$(X,\mathcal{B},m)=\prod_{-\infty}^{+\infty}(Y,2^Y,\mu).$$

之前的可测矩形的定义中 Borel 集就可以取作单点 $a_j$，就形成了基本可测矩形

$$[a_{-n}a_{-n+1}...a_n]=\left\{(x_i)_{-\infty}^{+\infty}\in\prod_{-\infty}^{+\infty}(Y,2^Y,\mu):x_i=a_i,\ i=-n,...,n\right\}.$$

它的测度为 $m([a_{-n}a_{-n+1}...a_n])=\prod_{i=-n}^{n}p_{a_i}$ 对于任意两个整数 $n>\ell$，也称

$$[a_\ell...a_n]=\left\{(x_i)_{-\infty}^{+\infty}\in\prod_{-\infty}^{+\infty}(Y,2^Y,\mu):x_i=a_i,\ i=\ell,...,n\right\}$$

为基本可测矩形，测度为 $m([a_\ell a_{-n+1}...a_n])=\prod_{i=\ell}^{n}p_{a_i}$.

### 1.1.2 可测函数的积分

**定义 1.25.** 设 $(X,\mathcal{B})$ 为可测空间，$f:X\to\overline{\mathbb{R}}=\mathbb{R}\cup\{\pm\infty\}$ 成为可测的，若对于任意的 $c>0$，集合 $\{x\in X:f(x)<c\}=f^{-1}([-\infty,c))$ 是可测的.

很显然，可见 $f$ 可测当且仅当对于 $\mathbb{R}$ 的 Borel 集合的原像是可测的.

**例 1.26.** 设 $(X,\mathcal{B},m)$ 为可测空间，$B\in\mathcal{B}$ 定义特征函数

$$\chi_B(x)=\begin{cases}1 & x\in B,\\ 0 & x\notin B.\end{cases}$$

可知特征函数是可测函数. 设 $A_i(i=1,2,...,n)$ 互不相交，并且 $a_i\in\mathbb{R}$，我们称形如 $f(x)=\sum_{i=1}^n a_i\chi_{A_i}(x)$ 是简单函数. 自然地，简单函数也是可测的.

**命题 1.27** (可测函数的简单性质)**.** *可测函数关于加减乘，取极限 (如果存在的话)，上下极限，取极大，极小都是封闭的.*

**定义 1.28** (可测函数的积分)**.** 设 $(X,\mathcal{B},m)$ 是测度空间，定义积分分为下面几步：

1. 设 $f(x)=\sum_{i=1}^n a_i\chi_{A_i}(x)$ 是简单函数 $A_i$ 互不相交，定义 $\int f\mathrm{d}m=\sum_{i=1}^n a_i m(A_i)$；
2. 设 $f\geq 0$，定义 $\int f\mathrm{d}m=\sup\left\{\int g\mathrm{d}m : 0\leq g\leq f,\ g \text{ 是简单函数}\right\}$ 若 $\int f\mathrm{d}m<+\infty$ 则称 $f$ 可积；
3. 对于一般的 $f$，分为正负部即可，若正负部都可积，则称 $f$ 可积，其中正负部定义为 $f_\pm=\frac{|f|\pm f}{2}$.

**定义 1.29.** 测度空间 $(X,\mathcal{B},m)$ 上的全体可积函数记为 $\mathcal{L}^1(X,\mathcal{B},m)$ 在不引起歧义的时候我们简记为 $\mathcal{L}^1(X)$. 我们可以定义上面的范数 $\|f\|_1=\int|f|\mathrm{d}m$. 如果我们将可测函数几乎处处意义下等同起来，即将几乎处处相等的函数看作相等，那么在商掉这个等价关系下，$\mathcal{L}^1(X)$ 记作 $L^1(X)$. 类似我们可以考虑 $\mathcal{L}^p$ 与 $L^p$ 的情况.

下面我们简单介绍常用的定理. 我们约定在测度空间 $(X,\mathcal{B},m)$ 上进行讨论.

**定理 1.30** (Levi 单调收敛定理)**.** *$0\leq f_1\leq f_2\leq ...$ 可测，并且 $f=\lim_{n\to\infty}f_n$ 则 $\int f\mathrm{d}m=\lim_{n\to\infty}\int f_n\mathrm{d}m$.*

**定理 1.31** (Fatou 引理)**.** *设 $f_n$ 可测，$\int\liminf_{n\to\infty}f_n\mathrm{d}m\leq\liminf_{n\to\infty}\int f_n\mathrm{d}m$.*

**定理 1.32** (Lebesgue 控制收敛定理)**.** *设 $f_n,g$ 可测，并且满足 $|f_n|\leq g$ a.e. 以及 $g\in\mathcal{L}^1$，$f=\lim_{n\to\infty}f_n$ a.e. 那么 $f\in\mathcal{L}^1$ 且 $\int f\mathrm{d}m=\lim_{n\to\infty}\int f\mathrm{d}m$.*

## 1.2 保测映射

### 1.2.1 保测映射的基本概念

**定义 1.33** (保测映射)**.** 设 $(X,\mathcal{B},m)$ 为概率空间，$T:X\to X$ 为映射，我们定义如下

1. (可测性) 若对于任意的 $B\in\mathcal{B}$，$T^{-1}B\in\mathcal{B}$，则称 $T$ 为可测的；
2. (保测性) $T$ 可测，对于任意的 $B\in\mathcal{B}$，$m(T^{-1}B)=m(B)$，此时称 $T$ 保测；
3. (可逆保测性) 若 $T$ 可逆，$T^{-1}$,$T$ 可测，并且 $T$ 保测 (此时 $T^{-1}$ 自然保测) 则称 $T$ 为可逆保测.

**引理 1.34.** *设 $T:(X,\mathcal{B},m)\to(X,\mathcal{B},m)$ 是可测的，则 $T$ 为保测的当且仅当对于任意的 $f\in\mathcal{L}^1(X,\mathcal{B},m)$，满足 $\int f\mathrm{d}m=\int f\circ T\mathrm{d}m$.*

证明. 一方面，对于任意的 $B\in\mathcal{B}$ 我们取函数 $f=\chi_B\in\mathcal{L}^1$ 于是利用条件 $\int f\mathrm{d}m=\int f\circ T\mathrm{d}m$ 得

$$m(B)=\int\chi_B\mathrm{d}m=\int\chi_B\circ T\mathrm{d}m=m(T^{-1}B),$$

其中第三个等式，我们利用了 $\chi_B\circ T=\chi_{T^{-1}B}$ ($\chi_B\circ T(x)=1\Leftrightarrow Tx\in B\Leftrightarrow x\in T^{-1}B\Leftrightarrow\chi_{T^{-1}B}(x)=1$). 这样就证明了 $T$ 是保测映射.

反之假设 $T$ 保测，利用前面的证明，可以验证 $\int f\mathrm{d}m=\int f\circ T\mathrm{d}m$ 对所有的特征函数都成立，进而对所有的简单函数都成立，下面我们证明这对所有的可积函数都成立. 由于可以将函数分为正负部的差，则不妨假设 $f\geq 0$. 取一列简单函数 $\{f_n\}$ 满足 $\lim_{n\to\infty}f_n=f$ a.e，单调收敛. 则可知 $\int f_n\mathrm{d}m=\int f_n\circ T\mathrm{d}m$ 对任意的 $n$ 成立. 可知此时 $f_n\circ T\to f\circ T$ 也是 a.e. 单调收敛 (此处的 a.e. 性质验证需要用到保测性). 于是在等号两侧取极限，就得到对 $f$ 结论也成立. □

**注 1.35.** 对于 $f$ 可测，则利用 $(f\circ T)^{-1}(B)=T^{-1}(f^{-1}(B))$ 可知 $f\circ T$ 也是可测的. 假设 $T$ 仅仅是可测的. 若 $f\in\mathcal{L}^1$ 时候，$f$ 处处有定义，$f\circ T$ 的定义也是明确的，但是如果 $f\in L^1$ 即看作等价类中的元素，那么 $f\circ T$ 的定义就不明确了，这是因为对于 $f=g$ a.e. 也不能保证 $f\circ T=g\circ T$ a.e. 若 $T$ 是保测的，此式成立，不然会有 $T$ 将正测度集映成零测集的情况.

在验证保测映射 $m(T^{-1}B)=m(B)$ 时，对于所有的可测集 $B$ 验证往往比较繁琐，对此有下面的定理，说明只需要在生成可测集 $\sigma$-代数的半代数上验证即可.

**定理 1.36.** 设 $(X,\mathcal{B},m)$ 是一个概率空间，$\mathcal{C}\subset\mathcal{B}$ 是半代数并且满足 $\mathcal{B}(\mathcal{C})=\mathcal{B}$，即 $\mathcal{B}$ 由 $\mathcal{C}$ 生成. 设 $T:(X,\mathcal{B},m)\to(X,\mathcal{B},m)$ 是可测的，若 $m(T^{-1}S)=m(S)$ 对任意的 $S\in\mathcal{C}$ 都成立，那么 $T$ 保测.

证明的想法就是把所有满足 $m(T^{-1}B)=m(B)$ 的集合构成的集族定义出来，证明这一集族在相关运算下封闭，最后由 $\mathcal{C}$ 生成出 $\mathcal{B}$.

证明. 令 $\mathcal{S}=\{B\in\mathcal{B}:m(T^{-1}B)=m(B)\}$. 有假设 $\mathcal{C}\subset\mathcal{S}$. 可以验证

1. 设 $B\in\mathcal{S}$，则 $m(T^{-1}B)=m(B)$ 于是有

   $$m(T^{-1}(X\backslash B))=m(T^{-1}X\backslash T^{-1}B)=m(X\backslash T^{-1}B)=1-m(T^{-1}B)=1-m(B)=m(X\backslash B),$$

   所以 $X\backslash B\in\mathcal{S}$. 这里我们用到了对于概率空间 $m(X)=1$.

2. 对于任意的 $B\in\mathcal{A}(\mathcal{C})$, 则存在有限个 $E_i\in\mathcal{C}, i=1,2,...,n$, 使得 $E_i$ 互不相交并且 $B=\cup_{i=1}^n E_i$. 这时有

   $$m(T^{-1}B)=m\left(T^{-1}\left(\bigcup_{i=1}^n E_i\right)\right)=m\left(\bigcup_{i=1}^n T^{-1}E_i\right)=\sum_{i=1}^n m(T^{-1}E_i)=\sum_{i=1}^n m(E_i),$$

   其中最后一个等号我们用到了 $E_i$ 互不相交的性质.

于是我们已经证明里 $\mathcal{A}(\mathcal{C})\subset\mathcal{S}$. 利用定理 1.21，问题归结为证明 $\mathcal{S}$ 是一个单调类. 假设 $A_n\in\mathcal{S}$ 并且 $A_n$ 单调增加. 于是 $\cup_{k=1}^n A_k=A_n\uparrow\cup_{k\geq 1}A_k$. 所以

$$m\left(\bigcup_{k=1}^n A_k\right)=m\left(T^{-1}\left(\bigcup_{k=1}^n A_k\right)\right)=m\left(\bigcup_{k=1}^n T^{-1}A_k\right).$$

两侧取极限即得 $\cup_{k\geq 1}A_k\in\mathcal{S}$. 这样我们就证明了结论. □

### 1.2.2 保测映射的重要例子

**例 1.37.** 概率空间 $(X,\mathcal{B},m)$ 上的恒等映射 Id 是一个保测映射.

**例 1.38.** $T:([0,1),\mathcal{B}[0,1),m)\to([0,1),\mathcal{B}[0,1),m)$, 其中 $\mathcal{B}[0,1)$ 表示 $[0,1)$ 上的 Borel 集，$m$ 表示 Lebesgue 测度. 定义 $Tx=2x(\mathrm{mod}\,1)$. 可知 $T$ 保测. 证明是显然的，令 $\mathcal{S}=\{[a,b):0\le a\le b\le 1\}$ 为生成 $\mathcal{B}[0,1)$ 的半代数，根据定理 1.36，可知，仅需要验证在半代数 $\mathcal{S}$ 上 $m(T^{-1}B)=m(B)$ 成立即可. 这是因为

$$m(T^{-1}[a,b))=m\left(\left[\frac{a+1}{2},\frac{b+1}{2}\right)\cup\left[\frac{a}{2},\frac{b}{2}\right)\right)=b-a=m([a,b)).$$

**例 1.39** (Gauss 变换)**.** 定义 $G:[0,1)\to[0,1)$ 满足 $Gx=\frac{1}{x}(\mathrm{mod}\,1)$ 若 $x\neq 0$，且 $G0=0$. 考虑 $G:([0,1),\mathcal{B}[0,1),m)\to([0,1),\mathcal{B}[0,1),m)$，其中 $\mathcal{B}[0,1)$ 表示 $[0,1)$ 上的 Borel 集，$m$ 有下面给出

$$m(B)=\frac{1}{\log 2}\int_B\frac{\mathrm{d}x}{1+x},\ B\in\mathcal{B}[0,1).$$

此时 $G$ 是保测的. 可知，对于任意的 $x\in(\frac{1}{n+1},\frac{1}{n}]$，有 $Gx=\frac{1}{x}-n$. 于是对于任意的 $0<a\le b\le 1$

$$m(G^{-1}[a,b))=m\left(\bigcup_{n\ge 1}\left(\frac{1}{n+b},\frac{1}{n+a}\right]\right)=\sum_{n=1}^{\infty}\frac{1}{\log 2}\int_{\frac{1}{n+b}}^{\frac{1}{n+a}}\frac{\mathrm{d}x}{1+x}$$
$$\frac{1}{\log 2}\sum_{n=1}^{\infty}\log\left(\frac{1+\frac{1}{n+a}}{1+\frac{1}{n+b}}\right)=\frac{1}{\log 2}\log\left(\frac{1+b}{1+a}\right)=m([a,b)).$$

关于 $G$ 我们有下面的解释. 对于 $x\in(0,1)$ 可以写作连分数形式

$$x=\frac{1}{a_0+\frac{1}{a_1+\frac{1}{a_2+\cdots}}}.$$

那么取 $x_1=x$，$x_{n+1}=Gx_n$，则有 $\lceil\frac{1}{x_n}\rceil=a_n$，其中 $\lceil\cdot\rceil$ 表示向上取整.

**例 1.40.** 同样考虑 $T:([0,1),\mathcal{B}[0,1),m)\to([0,1),\mathcal{B}[0,1),m)$，其中 $\mathcal{B}[0,1)$ 是 $[0,1)$ 上的 Borel 集，$m$ 为 Lebesgue 测度. $T$ 定义为 $Tx=x+\alpha(\mathrm{mod}\,1)$. 这时候可以类似例 1.38 证明 $T$ 是保测映射.

**定理 1.41** (Haar 测度)**.** *设 $G$ 是一个拓扑群: $G$ 同时是群和拓扑空间并且映射 $G\times G\to G$, $(x,y)\to xy^{-1}$ 是连续的; $\mathcal{B}(G)$ 是 $G$ 上的 Borel 集. 如果 $G$ 是紧致的，则在 $G$ 上存在一个左 Haar 测度 $m:\mathcal{B}(G)\to[0,+\infty)$，即满足下列两条公理的有限测度*

1. *(左乘不变性质) $m(xB)=m(B)$，对于任意的 $B\in\mathcal{B}(G)$，$x\in G$;*
2. *(正则性) 对于任意的 $B\in\mathcal{B}(G)$ 与 $\varepsilon>0$，存在闭集 $U$，开集 $O$ 使得 $C\subset B\subset O$，满足 $m(O\backslash C)<\varepsilon$.*

**注 1.42.** 关于概率测度的正则性，我们有，度量空间上的概率测度都是正则的.

**注 1.43.** 在常熟倍忽略不计的情况下，Haar 测度唯一存在，类似也可以讨论右不变 Haar 测度，当群是交换的时候，两者可以被等同起来.

**例 1.44.** $([0,1),\mathcal{B}[0,1),m)$，其中 $\mathcal{B}[0,1)$ 表示 $[0,1)$ 上的 Borel 集，$m$ 表示 Lebesgue 测度. 那么显然 $[0,1)$ 上可以定义群的加法 $x+y\equiv x+y(\mathrm{mod}\,1)$. 这样由例 1.40 可得 $m$ 是拓扑群 $([0,1),\mathcal{B}[0,1))$ 上的 Haar 测度.

**例 1.45.** 我们可以将例 1.44 中的情形推广到高维情形，即 $[0,1)^d$. 结论完全类似.

**例 1.46.** 设 $m_i$ 为紧致拓扑群 $G_i$ 上的 Haar 测度，$i \geq 1$. 那么我们可以定义乘积空间 $G = \prod_{i=1}^{\infty} G_i$，$m = \prod_{i=1}^{\infty} m_i$，$\prod_{i=1}^{\infty}(G_i, m_i) = (G, m)$. 那么这时候 $m$ 为 $G$ 上的 Haar 测度. 特别地，我们有例子 $G_i = \mathbb{Z}_2 = \{\bar{0}, \bar{1}\}^{\mathbb{Z}_+}$，$\mu_i(\{0\}) = \mu_i(\{1\}) = \frac{1}{2}$.

**例 1.47.** 定义 $\Sigma_2^+ = \{\bar{0}, \bar{1}\}^{\mathbb{Z}_+}$，在其上定义带有进位的加法. 具体来说对于

$$\underline{x} = x_1 x_2 x_3 \ldots,\ \underline{y} = y_1 y_2 y_3 \ldots$$

我们定义 $\underline{z} = \underline{x} + \underline{y}$. 令 $w \in \Sigma_2^+$ 来描述进位

$$w_1 = \begin{cases} 1 & x_1 + y_1 > 1 \\ 0 & x_1 + y_1 \leq 1 \end{cases},\ w_2 = \begin{cases} 1 & x_2 + y_2 + w_1 > 1 \\ 0 & x_2 + y_2 + w_1 \leq 1 \end{cases} \ldots$$

以此类推. 则 $z_1 = x_1 + y_1 - 2w_1$，$z_2 = x_2 + y_2 + w_1 - 2w_2 \ldots$ 可定义 $\underline{z} = z_1 z_2 \ldots$.

**习题 1.48.** 证明 $\Sigma_2^+$ 在上述的加法下是一个紧拓扑群 (此时赋予积拓扑). 找出上面的 Haar 测度.

*证明.* 先验证紧致性. 由于 $\{\bar{0}, \bar{1}\}$ 显然是紧的 (此集合为有限集，子集数量有限)，所以利用 Tychonoff 定理，可数个紧拓扑空间的乘积空间赋予积拓扑也是紧的.

下证是一个群，显然，对于任意的 $\underline{x}, \underline{y} \in \Sigma_2^+$，我们有 $\underline{x} + \underline{y} \in \Sigma_2^+$，并且上述加法满足结合律 (因为对于有限项就是二进制的加法，所以结合律自然存在，而这里的有限项的项数是任意的，所以整体也存在结合律). 于是仅需要验证对于任意的 $\underline{x} \in \Sigma_2^+$，有我们构造 $\underline{y}$ 使得 $\underline{x} + \underline{y} = \underline{0}$. 若 $x_1 = 0$，则取 $y_1 = 0$，反之 $y_1 = 1$. 这时，考虑 $x_2 + w_1$，如果 $x_2 + w_1$ 为 0 或 2 则取 $y_2 = 0$ 反之取 $y_2 = 1$，以此类推，可得 $z_i = 0$ 对任意的 $i \geq 1$ 都成立.

再证明相容性，即要证明在进位加法下，开集被映为开集. 事实上进位加法将基本矩形映射为基本矩形，而基本矩形为 $\Sigma_2^+$ 的拓扑积，于是相容性自然满足.

最后我们验证例 1.46 中的测度即为所需的 Haar 测度. 先验证左不变性，由定理 1.36，仅需要验证在基本矩形上左不变. 这是因为对于任意的基本矩形 $A = [a_1 a_2 \ldots a_n]$ 以及 $\underline{x} = x_1 x_2 \ldots$ 有 $\underline{x} + A$ 仍然是基本矩形，并且长度为 $n$, 记作 $[b_1 b_2 \ldots b_n]$. 于是

$$m(\underline{x} + A) = m([b_1 b_2 \ldots b_n]) = \left(\frac{1}{2}\right)^n = m([a_1 a_2 \ldots a_n]) = m(A).$$

最后验证正则性. 因为 $\{\bar{0}, \bar{1}\}$ 上可以赋予度量 $d$，则可数的积上面也可以赋予度量，根据度量空间上的测度都是正则的即可. □

**命题 1.49.** *设 $G$ 为紧拓扑群，$A : G \to G$ 是连续的满同态，则 $A$ 保持 Haar 测度.*

*证明.* 设 $m$ 为 $G$ 上的 Haar 测度. 要证明对于任意的 $B \in \mathcal{B}(G)$，有 $m(A^{-1}B) = m(B)$ (这里用到了 $A$ 的连续性，所以 $A^{-1}B \in \mathcal{B}$ 有定义). 令 $\mu : \mathcal{B}(G) \to [0, 1]$ 满足 $\mu(B) = m(A^{-1}B)$. 于是有对于任意的 $g \in G$，$B \in \mathcal{B}(G)$，根据 $A$ 是满的，则存在 $h \in G$ 使得 $Ah = g$. 那么

$$\mu(gB) = m(A^{-1}(gB)) = m(A^{-1}gA^{-1}B) = m(hA^{-1}B) = m(A^{-1}B) = \mu(B).$$

并且显然 $\mu$ 也是正则的，所以 $\mu$ 也是 Haar 测度，则由 Haar 测度的唯一性，得 $\mu = m$, 于是 $A$ 保持 Haar 测度 $m$. □

**定义 1.50.** 设 $G$ 是紧拓扑群，我们定义 $G$ 上的连续满同态与平移的复合 (左乘右乘所诱导的映射) 为仿射变换.

从上面的定理证明中，我们事实上用到了一个拉回的概念，我们可以将它推广. 设 $(X,\mathcal{B},m)$ 为概率空间，$(Y,\mathcal{C})$ 中 $\mathcal{C}$ 为拓扑空间 $Y$ 上的 Borel 集，假设 $T:(X,\mathcal{B})\to(Y,\mathcal{C})$ 可测 (开集的原像是可测的)，那么可以通过下面的方式定义 $\mathcal{C}$ 上的开集.

$$\mu:\mathcal{C}\to[0,1]\ (\mu(C)=m(T^{-1}C)).$$

此时 $\mu$ 记作 $\mu=T_*m$. 注意到此时 $T_*$ 是线性的，即对于 $p>0$，$(X,\mathcal{B},m_1)$ 与 $(X,\mathcal{B},m_2)$，有 $T_*(pm_1+(1-p)m_2)=pT_*m_1+(1-p)T_*m_2$.

**例 1.51.** 令 $Y=\{0,1,2,...,d-1\}$，$d\geq 2$，$\mu(\{i\})=p_i$，$\sum_{i=0}^{d-1}p_i=1$. 那么我们可以定义乘积空间 $(\Sigma_d^+,\mathcal{B},m)=((Y,2^Y,\mu))^{\mathbb{Z}_+}$. 这里的测度 $m$ 是之前提到的单边的 Bernoulli 测度. 我们定义左移位映射 $T:\Sigma_d^+\to\Sigma_d^+$ 满足 $T(x_0x_1...)=(x_1x_2...)$. 则首先 $T$ 在积拓扑下是连续的. 其次 $T$ 保持 $m$ 不变. 设 $a_i\in Y$ $i=0,1,2,...,n-1$ 并且 $[a_0a_1...a_{n-1}]$ 是基本矩形是生成 $\mathcal{B}$ 的半代数，那么根据定理 1.36，只需要证明在基本矩形上 $T$ 是保测的即可. 此时注意到

$$T^{-1}([a_0a_1...a_{n-1}])=\bigcup_{a_{-1}\in Y}[a_{-1}a_0...a_{n-1}].$$

于是有

$$m([a_0a_1...a_{n-1}])=p_{a_0}p_{a_1}...p_{a_{n-1}}=\sum_{a_{-1}\in Y}p_{a_{-1}}p_{a_0}p_{a_1}...p_{a_{n-1}}=m\left(T^{-1}([a_0a_1...a_{n-1}])\right).$$

这样就证明成立了. 反之，如果 $m$ 是一个任给的测度，那么，要使得 $T$ 保持测度 $m$，需要满足

$$m([a_0a_1...a_{n-1}])=\sum_{a_{-1}\in Y}([a_{-1}a_0...a_{n-1}]),\ m([a_0a_1...a_{n-1}])=\sum_{a_n\in Y}([a_0...a_{n-1}a_n]).$$

这说明，如果设函数 $P_n:\{0,1,2,...,n\}\to[0,1]$ 满足

$$\sum_{a_0\in Y}P_0(a_0)=1,\ P_n(a_0a_1...a_n)=\sum_{a_{-1}\in Y}P_{n+1}(a_{-1}a_0...a_n)=\sum_{a_{n+1}\in Y}P_{n+1}(a_0a_1...a_{n+1})$$

于是唯一确定一个概率测度 $m$ 满足 $m([a_0a_1...a_n])=P_n(a_0a_1...a_n)$ 并且 $m$ 是 $T$ 不变的. 特别的对于单边的 Bernoulli 测度 $P_n(a_0a_1...a_n)=p_{a_0}p_{a_1}...p_{a_n}$.

**例 1.52** (Markov 链)**.** 令 $p=(p_0,...,p_{d-1})$ 满足 $p_i\geq 0$ 并且 $\sum_{i=0}^{d-1}p_i=1$. 特别地我们称这样的向量是概率向量. 设 $P=(p_{ij})_{1\leq i,j\leq n}$ $p_{ij}\geq 0$ 满足 $\sum_{j=1}^{d-1}p_{ij}=1$ 是随机矩阵. 我们进一步假设 $pP=p$. 构造数列如下

$$P_n:\{0,1,2,...,n\}\to[0,1],\ P_0(a_0)=p_{a_0},\ P_n(a_0a_1...a_n)=p_{a_0}p_{a_0a_1}...p_{a_{n-1}a_n}.$$

于是经过简单验证可得

$$P_n(a_0a_1...a_n)=\sum_{a_{-1}\in Y}P_{n+1}(a_{-1}a_0...a_n)=\sum_{a_{n+1}\in Y}P_{n+1}(a_0a_1...a_{n+1})$$

成立，根据例 1.51 中的分析，就构造了一个左位移下保持的测度. 此时的保测系统 $(X,\mathcal{B},m,T)$ 被称为 Markov 链.

## 1.3 Poincaré 回复定理及其相关结论

### 1.3.1 Poincaré 回复定理的证明

**引理 1.53.** 设 $(X,\mathcal{B},m,T)$ 是保测系统. 对于 $B\in\mathcal{B}$，我们令

$$B_\infty=\bigcap_{n\geq 0}\bigcup_{k\geq n}T^{-k}B=\limsup_{n\to\infty}T^{-n}B.$$

那么 $T^{-1}B_\infty = B_\infty$ 并且 $m(B_\infty) = m(\cup_{k\geq n}T^{-k}B)$ 对于任意的 $n \geq 0$ 都成立.

**注 1.54.** 不特别指出，在本讲义中保测系统都是概率空间上的保持概率测度的系统.

**注 1.55.** $B_\infty$ 的意义就是那些在 $T$ 的作用下，无穷多次进入 $B$ 的点的全体构成的集合.

*证明.* 直接计算得

$$T^{-1}B_\infty = T^{-1}\left(\bigcap_{n\geq 0}\bigcup_{k\geq n}T^{-k}B\right) = \bigcap_{n\geq 0}\bigcup_{k\geq n+1}T^{-k}B = \bigcap_{n\geq 1}\bigcup_{k\geq n}T^{-k}B = B_\infty.$$

其中最后一个等式我们利用了 $\cup_{k\geq n}T^{-k}B$ 是单调减少得这一性质. 更进一步，根据测度的单调收敛性质，得

$$m(B_\infty) = \lim_{n\to\infty} m(\cup_{k\geq n}T^{-k}B).$$

又由于 $T^{-1}(\cup_{n\geq 0}T^{-n}B) = \cup_{n\geq 1}T^{-n}B$，不断重复，再结合上面的式子与 $T$ 的保测性，即可得到 $m(B_\infty) = m(\cup_{k\geq n}T^{-k}B)$ 对于任意的 $n \geq 0$ 成立. □

**定理 1.56** (Poincaré 回复定理)**.** *设 $(X, \mathcal{B}, m, T)$ 是保测系统，$B \in \mathcal{B}$ 满足 $m(B) > 0$ 则 $B$ 中几乎所有的点都无穷次返回 $B$.*

*证明.* 令 $C = B \cap B_\infty$，其中 $B_\infty$ 由引理 1.53 给出. 则注意到注 1.55，我们可知 $C$ 中的点在 $T$ 的作用下无穷次返回. 另外

$$m(C) = m(B \cap B_\infty) = m\left(B \cap \bigcup_{n\geq 0}T^{-n}B\right) = m(B),$$

其中对于第二个等式，我们利用了引理 1.53 关于 $B_\infty$ 测度的结果，即 $B_\infty$ 与 $\cup_{n\geq 0}T^{-n}B$ 测度相等并且前者是后者的子集，对于最后一个等式我们注意到 $B \subset \cup_{n\geq 0}T^{-n}B$. 这样就证明了这个定理. □

### 1.3.2 回复点的定义与刻画

如果保测系统 $(X, \mathcal{B}, m, T)$ 中 $X$ 进一步是紧致度量空间，记为 $(X, d)$ 并且 $T$ 是连续的，则我们由更精细的刻画.

对于 $x \in X$，定义

$$\omega(x) = \{y \in X : \exists 0 < n_1 < n_2 < ... < ...,\ \text{ s.t. } T^{n_i}x \to y\}.$$

这里 $\omega(x)$ 的几何意义可以看作是 $x$ 的 $T$ 映射下的轨道的聚点. 对此我们有下面的定理.

**定理 1.57.** *设 $(X, d)$ 是紧致度量空间，$T : (X, d) \to (X, d)$ 保测连续，$m$ 为 $X$ 上的一个 Borel 概率测度，即可测集就是度量空间中的 Borel 集. 那么有*

1. *$\omega(x) \neq \emptyset$ 并且是闭集，进一步 $T\omega(x) = \omega(x)$.*
2. *$m(\{x \in X : x \in \omega(x)\}) = 1$. 此时这样的点我们成为是回复点.*

**注 1.58.** 这里的回复点，相当于是是自己轨道的据点，如果其作为单点在轨道中出现无穷多次，那么自然是回复点，从而可见回复点是一个更为广泛的性质.

**注 1.59.** 回复点也可以利用 $\liminf_{n\to\infty} d(T^n x, x) = 0$ 这一条件来刻画.

证明. 第一条证明是显然的，在这里略去. 我们主要考察第二条结论的证明. 由于 $(X,d)$ 为紧致度量空间，则有可数的拓扑基，记作 $\{U_n\}_{n\geq 1}$. ($\{U_n\}_{n\geq 1}$ 为拓扑基当且仅当对于任意的 $x\in X$，以及 $x$ 的邻域 $U$，存在 $n$，使得 $x\in U_n\subset U$). 记

$$\widetilde{U}_n=\{x\in U_n:\forall k\geq 1,\ T^kx\notin U_n\}\subset U_n.$$

由 Poincaré 回复定理得 $m(\widetilde{U}_n)=0$. 另一方面我们断言 $X\setminus\cup_{n\geq 1}\widetilde{U}_n=\{x\in X:x\in\omega(x)\}$. 如果这一断言成立，那么可得 $m(\{x\in X:x\in\omega(x)\})=1$ 是显然的.

下面我们来证明这一断言. 对于任意的 $x\in X\setminus\cup_{n\geq 1}\widetilde{U}_n$，以及其邻域 $U$，根据拓扑基的性质，存在 $U_n$，使得 $x\in U_n\subset U$. 由于 $x\notin\widetilde{U}_n$，则存在 $k\geq 1$ 使得 $T^kx\in U_n\subset U$. 由 $U$ 的任意性，就证明了 $x\in\omega(x)$. 反之若 $x\in\omega(x)$，令

$$I_1=\{n:x\in U_n\},\ I_2=\{n:x\notin U_n\}.$$

所以我们可以作下列的讨论.

1. 若 $n\in I_2$，于是 $x\notin\widetilde{U}_n$，这里我们用到了 $\widetilde{U}_n\subset U_n$，
2. 若 $n\in I_1$，由于 $x$ 是回复点，则 $x\notin\widetilde{U}_n$.

根据上面的讨论，得到 $x\in X\setminus\cup_{n\geq 1}\widetilde{U}_n$. □

下面我们给出回复时间的定义，并证明一个相关的定理.

**定理 1.60** (Kač)**.** *设 $(X,\mathcal{B},m,T)$ 是一个保测系统，$E\in\mathcal{B}$ 并且 $m(E)>0$. 定义 $\tau:E\to\mathbb{Z}_+\cup\{+\infty\}$:*

$$\tau(x)=\begin{cases}+\infty & x\in E_0,\\ \min\{k\geq 1:T^kx\in E\} & x\notin E_0,\end{cases}$$

*其中 $E_0:=\{x\in E:T^kx\notin E,\ \forall k\geq 1\}$. 上述的 $\tau$ 根据 Poincaré 回复定理是几乎处处有限的. 于是我们有*

$$\int_E\tau(x)\mathrm{d}m=1-m(E_0^*),$$

*其中 $E_0^*=\{x\notin E:T^kx\notin E,\ \forall k\geq 1\}$.*

**注 1.61.** 上面的 $E_0$ 可以看作是在 $E$ 中不论经过多少次迭代都不会回到 $E$ 中的元素构成的集合，$E_0^*$ 看作不在 $E$ 中的，不论经过多少次迭代都不会进入 $E$ 中的元素构成的集合.

证明. 先证明 $\tau$ 是可测的. 令 $E_n=\tau^{-1}(n)$. 可知

$$\begin{aligned}E_n&=\{x\in E:Tx\notin E,...,T^{n-1}x\notin E,\ T^nx\in E\}\\&=E\cap(X\backslash T^{-1}E)\cap...\cap(X\backslash T^{-(n-1)}E)\cap T^{-n}E.\end{aligned}$$

这个集合显然是可测的，这是因为 $T$ 是可测的. 类似地，我们定义

$$E_n^*=\{x\notin E:Tx\notin E,...,T^{n-1}x\notin E,\ T^nx\in E\}.$$

于是 $\cup_{n\geq 1}E_n=E$，$\cup_{n\geq 0}E_n^*=X\backslash E$. 并且我们可见 $E_i,E_i^*$ 互不相交. 所以可得

$$1=m(X)=m(E_0^*)+\sum_{n=1}^{\infty}(m(E_n)+m(E_n^*)).\quad (*)$$

上面的最后一个等式我们利用了 Poincaré 回复定理，即 $m(E_0)=0$. 我们断言，$T^{-1}(E_n^*)=E_{n+1}\cup E_{n+1}^*$. 这是因为

$$x\in T^{-1}(E_n^*)\Leftrightarrow Tx\in E_n^*\Leftrightarrow Tx\notin E,...,T^n x\notin E,\ T^{n+1}x\in E\quad(**).$$

如果 $x\in E$ 且 $(**)$ 成立，那么 $x\in E_{n+1}$ 反之 $x\in E_{n+1}^*$. 这样就证明了断言. 根据 $T$ 的保测性质，可以得到

$$m(E_n^*)=m(T^{-1}E_n^*)=m(E_{n+1})+m(E_{n+1}^*).$$

重复使用上面的式子，我们有

$$m(E_n^*)=m(E_{n+1})+m(E_{n+2})+m(E_{n+2}^*).$$

以此类推并且注意到 $\lim_{n\to\infty}m(E_n^*)=0$ (根据 $(*)$ 式)，我们可得

$$m(E_n^*)=\sum_{k=n+1}^{\infty}m(E_k).$$

再一次结合 $(*)$ 式，可得

$$1=m(E_0^*)+\sum_{n=1}^{\infty}\sum_{k=n}^{\infty}m(E_k)=m(E_0^*)+\sum_{n=1}^{\infty}\sum_{n=1}^{k}m(E_k)=m(E_0^*)+\int_E\tau(x)\mathrm{d}m.$$

这就证明了我们的结论. □

**习题 1.62.** 设 $(X,\mathcal{B},m,T)$ 是保测系统，$E\in\mathcal{B}$ 满足 $m(E)>0$，$\tau$ 是首次返回时. 设 $B_\infty$ 为引理 1.53 中给出，定义 $R:B_\infty\to B_\infty$ 满足 $Rx=T^{\tau(x)}x$ 证明此时 $R$ 是保测的.

*证明.* □

## 1.4 遍历测度

### 1.4.1 遍历测度的定义与等价刻画

**定义 1.63.** 设 $(X,\mathcal{B},m,T)$ 是保测系统，若满足对于任意的 $B\in\mathcal{B}$ 只要 $T^{-1}B=B$ 就有 $m(B)=1$ 或 0，则称 $m$ 为 $T$ 的遍历测度.

**例 1.64.** 设 $T:(X,\mathcal{B})\to(X,\mathcal{B})$ 可测 ($\mathcal{B}$ 中集合的原像还是 $\mathcal{B}$ 中的集合)，若 $T$ 有不动点 $p$，满足 $Tp=p$，那么关于 $p$ 点的 Dirac 测度 $\delta_p$ (见例 1.14) 为 $T$ 的一个遍历测度. 更一般的，若 $p$ 点为周期点，即 $T^n p=p$ 并且 $T^k\neq p$ 对于任意的 $0<k<n$. 这时候 $p$ 点的轨道可写作

$$\mathrm{Orb}(p)=\{p,Tp,T^2p,...,T^{n-1}p\}.$$

则 $\mu_p=\frac{1}{n}\sum_{i=0}^{n-1}\delta_{T^ip}$ 是一个遍历测度. 证明是显然的，如果 $T^{-1}B=B$ 并且 $m(B)>0$，那么意味着 $T^ip\in B$ 其中 $0\le i\le n-1$. 这样就意味着对于任意的 $0\le k\le n-1$ 有 $T^kp\in B$ 于是 $m(B)=1$.

**注 1.65.** 直观上看，如果 $m$ 不遍历，那么就存在 $B\in\mathcal{B}$，满足 $T^{-1}B=B$ 但是 $0<m(B)<1$. 这时候有 $T^{-1}(X\backslash B)=X\backslash B$ 并且 $0<m(X\backslash B)<1$ 相当于将 $X$ 这个系统分解为两个互不相干的系统了.

下面这个定理给出了遍历的等价刻画.

**定理 1.66.** 设 $(X,\mathcal{B},m,T)$ 保测，则以下四条等价

1. $m$ 遍历；
2. 若 $B \in \mathcal{B}$ 满足 $m(T^{-1}B\Delta B)$ 则 $m(B)=1$ 或 0 其中 $A\Delta B$ 表示集合的对称差；
3. 对于任意的 $A \in \mathcal{B}$，若 $m(A)>0$ 则 $m(\cup_{n\geq 0}T^{-n}A)=1$；
4. 对于任意的 $A,B \in \mathcal{B}$，满足 $m(A),m(B)>0$ 则存在 $n>0$ 使得 $m(T^{-n}A\cap B)>0$.

证明. 1⇒2: 设 $m(T^{-1}B\Delta B)=0$, 对于集合 $B$ 考虑 $B_\infty$ (定义见引理 1.53). 可知 $T^{-1}B_\infty = B_\infty$. 利用遍历性, 可得 $m(B_\infty)=1$ 或 0. 根据引理 1.53 的性质, 可得 $m(B_\infty)=m(\cup_{n\geq 0}T^{-n}B)$. 由此可得, 只要证明 $m(\cup_{n\geq 0}T^{-n}B)=m(B)$ 即可. 根据 $m(T^{-1}B\Delta B)=0$，我们有 $m(T^{-k-1}B\Delta T^{-k}B)=0$ 对于任意的 $k\geq 0$ 成立. 注意到

$$T^{-n}B\Delta B = \Delta_{k=0}^{n-1}(T^{-k-1}B\Delta T^{-k}B) \subset \bigcup_{k=0}^{n-1}(T^{-k-1}B\Delta T^{-k}B)$$

则可得 $m(T^{-n}B\Delta B)=0$ 对于任意的 $n\geq 0$ 成立. 由此即得 $m(\cup_{n\geq 0}T^{-n}B)=m(B)$.

2⇒3：令 $B=\cup_{n\geq 0}T^{-n}A$，可知 $T^{-1}B\subset B$. 利用 $T$ 的保测性，得到 $m(T^{-1}B\Delta B)=0$，利用 2，以及假设 $m(A)>0$，可得 $m(B)=1$.

3⇒4：根据 3，有 $m(\cup_{n\geq 0}T^{-n}A)=1$，于是可知

$$0<m(B)=m\left(B\cap\left(\bigcup_{n\geq 0}T^{-n}A\right)\right)\leq \sum_{n=0}^{\infty}m(T^{-n}A\cap B).$$

于是必然存在某个 $n$ 使得 $m(T^{-n}A\cap B)>0$ 成立.

4⇒1：设 $m(B)>0$ 并且 $T^{-1}B=B$. 那么在 4 中取 $A=X\backslash B$. 若 $m(A)>0$ 则有 $m(T^{-n}A\cap B)=m((X\backslash B)\cap B)=0$ 这与 4 矛盾. 所以 $m(A)=0$ 于是 $m(B)=1$. □

**例 1.67.** 令 $Y=\{0,1,2,...,d-1\}$，$d\geq 2$，$\mu(\{i\})=p_i$，$\sum_{i=0}^{d-1}p_i=1$. 考虑乘积空间 $(\Sigma_d^+,\mathcal{B},m)=((Y,2^Y,\mu))^{\mathbb{Z}_+}$. 这里的测度 $m$ 是单边的 Bernoulli 测度. 左移位映射 $T:\Sigma_d^+\to\Sigma_d^+$ 满足 $T(x_0x_1...)=(x_1x_2...)$. (具体见例 1.51) 那么对于任意的 $A,B\in\mathcal{B}$ 有 $\lim_{n\to\infty}m(T^{-n}A\cap B)=m(A)m(B)$. (满足这样的条件时，测度 $m$ 被称为强混合的显然根据定理 1.66，这一性质强于遍历性).

下面我们给出证明，设 $\mathcal{C}$ 是基本矩形给出的半代数，生成的代数为 $\mathcal{A}$ 再生成 $\sigma$-代数为 $\mathcal{B}$. 我们先验证在半代数 $\mathcal{C}$ 上，强混合性成立. 这是显然的，因为设

$$A=[a_0a_1...a_{k-1}],\ B=[b_0b_1...b_{\ell-1}],$$

有 $T^{-n}A\cap B=\{x\in X: x[0,...,\ell-1]=b_0...b_{\ell-1},\ x[n,...,n+k-1]=a_0...a_{k-1}\}$. 于是自然可得 $m(T^{-n}A\cap B)=m(A)m(B)$，仅需要 $n\geq\ell$ 即可. 于是在代数 $\mathbf{A}$ 上也是成立的. 那么根据下面的引理，在 $\mathcal{B}$ 上也是成立的.

**引理 1.68.** 设 $(X,\mathcal{B},m)$ 为概率空间，$\mathcal{A}$ 是生成 $\sigma$-代数 $\mathcal{B}$ 的代数，即 $\mathcal{B}=\mathcal{B}(\mathcal{A})$，于是，对于任意的 $B\in\mathcal{B}$，任意的 $\varepsilon>0$ 存在 $A\in\mathcal{A}$ 使得 $m(A\Delta B)<\varepsilon$.

这个引理的证明可以从单调类的刻画理解，在此我们略去证明.

### 1.4.2 利用可测函数刻画遍历测度

遍历性同样可以通过可测函数来进行刻画，对此我们有下面的定理.

**定理 1.69.** 设 $(X,\mathcal{B},m,T)$ 为保测系统，则下面的五条等价.

1. $m$ 遍历;
2. 对于任意的 $f \in L^0(X)$ 若 $f \circ T = f$ 则 $f(x) = \text{const}$ $m$-*a.e.* $x$;
3. 对于任意的 $f \in L^0(X)$ 若 $f \circ T = f$ $m$-*a.e.* $x$ 则 $f(x) = \text{const}$ $m$-*a.e.* $x$;
4. 对于任意的 $f \in L^2(X)$ 若 $f \circ T = f$ 则 $f(x) = \text{const}$ $m$-*a.e.* $x$;
5. 对于任意的 $f \in L^2(X)$ 若 $f \circ T = f$ $m$-*a.e.* $x$ 则 $f(x) = \text{const}$ $m$-*a.e.* $x$.

其中 $L^0(X)$ 表示可测函数的全体

证明. 很显然，可以得到 2⇒4，3⇒5，5⇒4 以及 3⇒2. 所以我们仅需要证明 1⇒2 与 4⇒1 即可.

1⇒3：对于任意的 $c \in \mathbb{R}$，我们令 $B = B_c = \{x \in X : f(x) < c\}$. 于是可得

$$T^{-1}B\Delta B = \{x \in X : x \in B,\ Tx \notin B\} \cup \{x \in X : x \notin B,\ Tx \in B\}.$$

于是 $T^{-1}B\Delta B \subset \{x \in X : f(x) \neq f(Tx)\}$. 根据条件,这个右侧的集合是零测集,所以 $m(T^{-1}B\Delta B) = 0$. 再根据遍历性以及定理 1.66 的第 2 条就得到 $m(B) = 0$ 或者 1 这说明 $f$ 在几乎处处的意义下是常数函数. 事实上，令 $a = \operatorname{essinf} f(x)$，$b = \operatorname{esssup} f(x)$ 可知若 $a < b$ 那么对 $c \in (a, b)$ 满足 $m(B_c) = 0$ 或者 1 显然矛盾.

4⇒1：设 $B$ 可测，$T^{-1}B = B$ 令 $f = \chi_B$ 那么根据 4 可得，因为 $\chi_B \circ T = \chi_B$ 所以 $\chi_B$ 几乎处处是常数函数，那么 $m(B) = 0$ 或 1. 这样就证明了遍历性. □

**例 1.70.** 设 $m$ 为 $\mathbb{T} = [0, 1) = \mathbb{R}/\mathbb{Z}$ 上的 Haar 测度，给定 $\alpha \in \mathbb{T}$. 那么 $m$ 为遍历的当且仅当 $\alpha \notin \mathbb{Q}$.

下面我们给出证明，若 $m$ 遍历，那么如果 $\alpha \in \mathbb{Q}$ 则有 $\alpha = \frac{p}{q}$ 成立，并且 $(p, q) = 1$. 令 $f(x) = e^{2\pi i q x} : \mathbb{T} \to \mathbb{R}$ 可知 $f \circ T = f$. 具体来说

$$f(Tx) = e^{2\pi i q\left(x + \frac{p}{q}\right)} = e^{2\pi i q x} = f(x).$$

根据 $m$ 的遍历性，可得 $f$ 为常数函数，这显然是矛盾的.

另一方面，若 $\alpha \notin \mathbb{Q}$ 假设 $f \in L^2$ 并且 $f \circ T = f$. 于是我们可以利用 Fourier 展开，得到

$$f(x) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n x}.$$

另外，注意到 $\int (f \circ T)^2 \mathrm{d}m = \int f^2 \circ T \mathrm{d}m = \int f^2 \mathrm{d}m$，所以 $f \circ T$ 同样是 $L^2$ 可积的，那么可设 $f \circ T$ 的 Fourier 展开

$$f \circ T(x) = \sum_{n \in \mathbb{Z}} b_n e^{2\pi i n x}.$$

利用简单的计算可以求得

$$b_n = \int f \circ T(x) e^{-2\pi i n x} \mathrm{d}m = \int_0^1 f(x + \alpha) e^{-2\pi i n x} \mathrm{d}x = a_n e^{2\pi i n \alpha}.$$

这样根据 $f \circ T = f$ 比较 Fourier 展开的系数以及注意到 $\alpha \notin \mathbb{Q}$ 可得 $a_n = b_n = 0$ 当 $n \geq 1$ 时，这样就证明了结论.

**习题 1.71.** 设 $A = (a_{ij})_{1 \leq i,j \leq 2}$ 其中 $a_{11} = 2$，$a_{12} = a_{21} = a_{22} = 1$. 那么 $T_A : \mathbb{T}^2 \to \mathbb{T}^2$ 是同构，证明 $(\mathbb{T}^2, \mathcal{B}_{\mathbb{T}^2}, m, T_A)$ 这一保测系统中 $m$ 是遍历的，其中 $\mathcal{B}_{\mathbb{T}^2}$ 表示 $\mathbb{T}^2$ 上的 Borel 集合.

## 1.5 遍历定理

### 1.5.1 Von-Neumann 遍历定理

**定理 1.72** (Von-Neumann 遍历定理)**.** 设 $(X,\mathcal{B},m,T)$ 为保测系统，对于任意的 $f\in L^2(X)$ 有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i=\overline{f}\in L^2(X),$$

在 $L^2$ 意义下收敛，并且 $\overline{f}\circ T=\overline{f}$ 并且 $\int\overline{f}\mathrm{d}m=\int f\mathrm{d}m$. 特别地，若 $m$ 遍历，则 $\overline{f}=\int f\mathrm{d}m$.

证明. 令 $U=U_T:L^2(X)\to L^2(X)$ 满足 $f\mapsto f\circ T$. 很显然根据 $T$ 的保测性这是一个等距映射，即 $\|f\|_{L^2}=\|U_T(f)\|_{L^2}$. 我们令

$$\mathcal{S}=\mathcal{S}(T)=\{f\in L^2(X):f\circ T=f\}.$$

很显然，对于任意的 $f\in\mathcal{S}$，结论成立. 再令

$$\mathcal{C}(T)=\{g-g\circ T:g\in L^2(X)\}\subset L^2(X).$$

注意到，对于任意的 $f=g-g\circ T\in\mathcal{C}$ 有

$$\frac{1}{n}\sum_{i=0}^{n-1}f\circ T=\frac{1}{n}\sum_{i=0}^{n-1}(g\circ T^i-g\circ T^{i+1})=\frac{1}{n}(g-g\circ T^n)\to 0,$$

当 $n\to\infty$，其中这里的收敛在 $L^2$ 意义下. 这表明对任意的 $f\in\overline{\mathcal{C}}$ 有 $\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i\to 0$ 在 $L^2$ 意义下，这里的 $\overline{\mathcal{C}}$ 表示 $\mathcal{C}$ 在 $L^2$ 范数下的闭包. 这只需要通过简单的逼近论述就能得到，事实上，对于任意的 $\varepsilon>0$，取 $g-g\circ T\in\mathbf{C}$ 使得 $\|f-(g-g\circ T)\|_{L^2}<\varepsilon$. 于是可得

$$\begin{aligned}\left\|\frac{1}{n}\sum_{i=0}^{n-1}U^i(f)\right\|_{L^2}&\le\left\|\frac{1}{n}\sum_{i=0}^{n-1}(U^i(f)-U^i(g-g\circ T))\right\|_{L^2}+\left\|\frac{1}{n}\sum_{i=0}^{n-1}U^i(g-g\circ T)\right\|_{L^2}\\&\le\varepsilon+\left\|\frac{1}{n}\sum_{i=0}^{n-1}U^i(g-g\circ T)\right\|_{L^2}.\end{aligned}$$

上面的不等式处，我们利用了 $U$ 的等距变换的性质，上面的第二项根据之前的分析，自然是收敛到 0 的. 最后我们证明对于 $L^2(X)$ 有正交分解 $L^2(X)=\mathcal{S}\perp\overline{\mathcal{C}}$ (此处的内积就是 $L^2$ 内积).

对于任意的 $f\in\mathcal{S}$，我们有 $f\circ T=f$，考虑 $g-g\circ T\in\mathcal{C}$ 有

$$(f,g-g\circ T)=(f,g)-(f,g\circ T)=(f,g)-(f\circ T,g\circ T)=0.$$

这说明 $\mathcal{S}\subset\overline{\mathcal{C}}^{\perp}$. 反之，若 $f\in\overline{\mathcal{C}}^{\perp}$，则有

$$\|f-f\circ T\|_{L^2}^2=(f-f\circ T,f-f\circ T)=(f,f)+(f\circ T,f\circ T)-2(f,f\circ T)=2(f,f-f\circ T)=0.$$

这就证明了反向的包含关系. 于是分解关系成立. 所以，对于任意的 $f=g+h$，其中 $g\in\overline{\mathcal{C}}$ 以及 $h\in\mathcal{S}$. 根据之前的分析就能得到

$$\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i=\frac{1}{n}\left(\sum_{i=0}^{n-1}g\circ T^i+\sum_{i=0}^{n-1}h\circ T^i\right)\to h=\overline{f}.$$

另外积分的性质来自于

$$\int\overline{f}\mathrm{d}m=(1,\overline{f})=\lim_{n\to\infty}\left(1,\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i\right)=\int f\mathrm{d}m.$$

这样就完成了 Von-Neumman 遍历定理的证明. □

利用 Von-Neumman 遍历定理可以得到简单的推论.

**推论 1.73.** 设 $(X,\mathcal{B},m,T)$ 是保测系统，那么 $m$ 是遍历的当且仅当对于任意的 $A,B\in\mathcal{B}$ 有

$$\frac{1}{n}\sum_{i=0}^{n-1}m(A\cap T^{-n}B)=m(A)m(B).$$

*证明.* 一方面，假设 $m$ 是遍历的. 设 $A,B\in\mathcal{B}$ 取 $f=\chi_B$，那么 $\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i\to\overline{f}=m(B)$ 在 $L^2$ 意义下收敛，其中最后一个等号我们利用了遍历性. 再取 $g=\chi_A$，我们可以得到

$$\left(g,\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i\right)\to(g,\overline{f})=m(A)m(B).$$

注意到

$$(g,f\circ T^i)=\int\chi_A\chi_B\circ T^i\mathrm{d}m=\int\chi_A\chi_{T^{-i}B}\mathrm{d}m=\int\chi_{A\cap T^{-i}B}\mathrm{d}m=m(A\cap T^{-i}B).$$

于是我们就证明了结论.

另一方面，若 $T^{-1}B=B$ 取 $A=B$ 带入式子中，有 $m(B)=m(B)^2$ 于是 $m(B)=0$ 或 1. □

**注 1.74.** 从这个式子我们也能看出强混合推出遍历.

### 1.5.2 Birkhoff 遍历定理

**定理 1.75** (Birkhoff 遍历定理)**.** *设 $(X,\mathcal{B},m,T)$ 为保测系统，对于任意的 $f\in L^1(X)$ (处处有定义的固定函数，不是等价类) 有*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i=f^*,$$

*在 a.e. 意义下收敛，并且 $f^*\circ T=f^*$ $m$-a.e. 并且 $\int f^*\mathrm{d}m=\int f\mathrm{d}m$. 特别地，若 $m$ 遍历，则 $f^*=\int f\mathrm{d}m$ $m$-a.e. 成立.*

在证明这一定理之前，我们先看一些 Birkhoff 遍历定理的应用.

**例 1.76.** 在 Birkhoff 遍历定理里取 $f=\chi_B$ 那么若 $m$ 是遍历的，则可知，

$$\frac{1}{n}\sum_{i=0}^{n-1}\chi\circ T^i(x)=\frac{1}{n}\sum_{i=0}^{n-1}\chi_{T^{-i}B}(x)\to m(B).$$

这个事实说明

$$\frac{1}{n}\#\left\{i\in\{0,1,2,...,n-1\}:T^ix\in B\right\}\to m(B),$$

刻画了平均逗留时间.

**例 1.77.** 设 $d\in\mathbb{Z}_+$ 并且 $d>1$，$X=\mathbb{T}=[0,1)$ $m$ 为 Lebesgue 测度，则 $X$ 上的几乎所有点都有唯一的 $d$-进展开，即

$$x=\sum_{n=1}^{\infty}\frac{x_n}{d^n}=(0.x_1x_2...)_d,\ x_n\in\{0,1,2,...,d-1\}.$$

那么令 $T=T_d:\mathbb{T}\to\mathbb{T}$ 为 $x\to dx(\mathrm{mod}\,1)$ 那么可知 $m$ 是遍历的. 令 $B=[0]=\{x\in\mathbb{T}:x_1=0\}$，可得 $m(B)=\frac{1}{d}$. 则利用上一个例子的结论，得

$$\frac{1}{n}\#\{i\in\{1,2,...,n\}:x_i=0\}\to\frac{1}{d}.$$

任取 $a_1,a_2,...,a_\ell\in\{0,1,2,...,d-1\}$ 与

$$B=[a_1a_2...a_\ell]=\{x\in\mathbb{T}\}=\{x\in[0,1):x_1...x_\ell=a_1...a_\ell\}.$$

则类似的结论有

$$\frac{1}{n}\#\{i\in\{1,2,...,n\}:x_ix_{i+1}...x_{i+\ell-1}=a_1a_2...a_\ell\}\to\frac{1}{d^\ell}.$$

*Birkhoff 遍历定理的证明.* 因为可将 $f$ 分为正负部的差 $f=f^+-f^-$，所以纸样证明 $f\geq 0$ 的情形. 令

$$\overline{f}(x)=\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i(x),\ \underline{f}(x)=\liminf_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i(x).$$

这两个函数处处有定义并且取值可能是 $+\infty$. 为了证明定理的结论，我们仅需要证明 $\int(\underline{f}-\overline{f})\mathrm{d}m\geq 0$ 即可. 这个还能归结为证明

$$\int\overline{f}\mathrm{d}m\leq\int f\mathrm{d}m\leq\int\underline{f}\mathrm{d}m.$$

我们先证明 $\int\overline{f}\mathrm{d}m\leq\int f\mathrm{d}m$ 成立. 首先，令

$$\overline{f}_M=\begin{cases}\overline{f}(x) & \overline{f}(x)\leq M,\\ M & \overline{f}(x)>M.\end{cases}$$

直接计算，我们有

$$\begin{aligned}\overline{f}\circ T(x)&=\limsup_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i(x)=\limsup_{n\to\infty}\frac{1}{n}\left(\sum_{i=0}^{n}f\circ T^i(x)-f(x)\right)\\&=\limsup_{n\to\infty}\left(\frac{n+1}{n}\frac{1}{n+1}\sum_{i=0}^{n}f\circ T^i(x)-\frac{f(x)}{n}\right)=\overline{f}(x)\end{aligned}$$

对于任意 $x\in X$ 都成立. 于是 $\overline{f}\circ T=\overline{f}$ 进而 $\overline{f}_M\circ T=\overline{f}_M$. 事实上，我们仅需要证明

$$\int\overline{f}_M\mathrm{d}m\leq\int f\mathrm{d}m+\varepsilon,\ \forall M,\varepsilon>0$$

成立. 对于给定的 $\varepsilon>0$，我们令

$$\tau:X\to\mathbb{Z}_+,\ \tau(x)=\min\left\{n\geq 1:\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i(x)\geq\overline{f}_M(x)-\varepsilon\right\}.$$

这里需要验证 $\tau$ 是良定义的，若 $\overline{f}(x)\leq M$，则显然 $\tau(x)$ 有限，因为此时 $\overline{f}_M(x)-\varepsilon=\overline{f}(x)-\varepsilon$. 若 $\overline{f}(x)>M$ 同样的结论也可以被得到. 所以，我们有

$$\tau(x)\overline{f}_M(x)\leq\sum_{i=0}^{\tau(x)-1}f\circ T^i(x)+\tau(x)\varepsilon.$$

利用 $\overline{f}_M$ 的 $T$ 不变性，可得

$$\sum_{i=0}^{\tau(x)-1}\overline{f}_M\circ T^i(x)\leq\sum_{i=0}^{\tau(x)-1}f\circ T^i(x)+\tau(x)\varepsilon.\quad(*)$$

下面我们对 $\tau(x)$ 作截断. 令 $A_n=\{x\in X:\tau(x)>n\}$，于是 $\{A_n\}$ 单调减少，并且极限为空集，所以 $m(A_n)\to 0$. 那么固定 $N\gg 1$，使得 $m(A_N)<\frac{\varepsilon}{M}$. 令 $\widetilde{\tau}:X\to\mathbb{Z}_+$ 是 $\tau$ 的截断，定义为

$$\widetilde{\tau}(x)=\tau(x)\chi_{X\setminus A_N}(x)+\chi_{A_N}(x).$$

另外定义

$$\widetilde{f}(x) = f(x)\chi_{X\setminus A_N}(x) + \max\{f(x), M\}\chi_{A_N}(x).$$

这样，式 $(*)$ 意味着

$$\sum_{i=0}^{\widetilde{\tau}(x)-1} \overline{f}_M \circ T^i(x) \le \sum_{i=0}^{\widetilde{\tau}(x)-1} \widetilde{f} \circ T^i(x) + \widetilde{\tau}(x)\varepsilon. \quad (**)$$

注意到

$$\begin{aligned}\int \widetilde{f}\mathrm{d}m &= \int_{X\setminus A_N} \widetilde{f}\mathrm{d}m + \int_{A_N} \widetilde{f}\mathrm{d}m \le \int_{X\setminus A_N} \widetilde{f}\mathrm{d}m + \int_{A_N} (\widetilde{f} + M)\mathrm{d}m \\ &\le \int f\mathrm{d}m + Mm(A_N) < \int f\mathrm{d}m + \varepsilon\end{aligned}$$

我们又将证明归结为

$$\int \overline{f}_M\mathrm{d}m \le \int \widetilde{f}\mathrm{d}m + \varepsilon,\ \forall M, \varepsilon > 0.$$

我们取 $L \gg 1$，满足 $L > \frac{NM}{\varepsilon}$ 并且定义

$$\widetilde{T}(x) = T^{\widetilde{\tau}(x)}x,\ \widetilde{\tau}_1(x) = \widetilde{\tau}(x),\ \widetilde{\tau}_2(x) = \widetilde{\tau}(\widetilde{T}(x)) + \widetilde{\tau}(x),\ \widetilde{\tau}_n(x) = \sum_{i=0}^{n-1} \widetilde{\tau}(\widetilde{T}^i x).$$

如此，存在 $k$，使得 $\widetilde{\tau}_k(x) \le L < \widetilde{\tau}_{k+1}(x)$. 那么结合 $(**)$ 的计算，可得

$$\begin{aligned}\sum_{i=0}^{L-1} \overline{f}_M \circ T^i(x) &= \sum_{i=0}^{\widetilde{\tau}_1(x)-1} \overline{f}_M \circ T^i(x) + \sum_{i=\widetilde{\tau}_1(x)}^{\widetilde{\tau}_2(x)-1} \overline{f}_M \circ T^i(x) + \dots + \sum_{i=\widetilde{\tau}_{k-1}(x)}^{\widetilde{\tau}_k(x)-1} \overline{f}_M \circ T^i(x) + \sum_{i=\widetilde{\tau}_k(x)}^{L-1} \overline{f}_M \circ T^i(x) \\ &\le \sum_{i=0}^{\widetilde{\tau}_k(x)-1} \widetilde{f} \circ T^i(x) + \widetilde{\tau}_k(x)\varepsilon + NM \le \sum_{i=0}^{L-1} \widetilde{f} \circ T^i(x) + L\varepsilon + NM.\end{aligned}$$

这里我们需要作下面的变形

$$\sum_{i=\widetilde{\tau}_{k-1}(x)}^{\widetilde{\tau}_k(x)-1} \overline{f}_M \circ T^i(x) = \sum_{i=0}^{\widetilde{\tau}(\widetilde{T}^k(x))-1} \overline{f}_M \circ T^i(T^{\widetilde{\tau}_{k-1}(x)}x) = \sum_{i=0}^{\widetilde{\tau}(\widetilde{T}^k(x))-1} \overline{f}_M \circ T^i(\widetilde{T}^{k-1}x).$$

于是对于之前的式子，两边同时除以 $L$，并积分就证明成立. 所以只需要证明 $\int \underline{f}\mathrm{d}m \ge \int f\mathrm{d}m$ 就能完成证明.

首先，我们注意到利用 Fatou 引理 (定理 1.31)，可得

$$\int \underline{f}\mathrm{d}m \le \liminf_{n\to\infty} \frac{1}{n} \int \sum_{i=0}^{n-1} f \circ T^i\mathrm{d}m = \int f\mathrm{d}m.$$

这先验地表明，$\underline{f} \in L^1$. 于是我们不再需要像对 $\overline{f}$ 的操作一样进行截断. 类似之前的证明可得 $\underline{f} = \underline{f} \circ T$. 往证

$$\int \underline{f}\mathrm{d}m \ge \int f\mathrm{d}m - \varepsilon,\ \forall \varepsilon > 0.$$

于是，给定的 $\varepsilon > 0$，令

$$\tau : X \to \mathbb{Z}_+,\ \tau(x) = \min\left\{n \ge 1 : \underline{f}(x) \ge \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i(x) - \varepsilon\right\}.$$

显然 $\tau$ 是良定义的. 所以

$$\tau(x)\underline{f}(x) \ge \sum_{i=0}^{\tau(x)-1} f \circ T^i(x) - \tau(x)\varepsilon.$$

利用 $\underline{f}$ 的 $T$ 不变性，可得

$$\sum_{i=0}^{\tau(x)-1} \underline{f}\circ T^i(x) \geq \sum_{i=0}^{\tau(x)-1} f\circ T^i(x) - \tau(x)\varepsilon.$$

对 $\tau(x)$ 作截断. 令 $A_n = \{x \in X : \tau(x) > n\}$，于是 $\{A_n\}$ 单调减少，并且极限为空集，所以 $m(A_n) \to 0$. 那么固定 $N \gg 1$，使得 $m(A_N) < \varepsilon$. 令 $\widetilde{\tau} : X \to \mathbb{Z}_+$ 是 $\tau$ 的截断，定义为

$$\widetilde{\tau}(x) = \tau(x)\chi_{X\setminus A_N}(x) + \chi_{A_N}(x).$$

所以得到

$$\sum_{i=0}^{\widetilde{\tau}(x)-1} \underline{f}\circ T^i(x) \geq \sum_{i=0}^{\widetilde{\tau}(x)-1} f\circ T^i(x) - \widetilde{\tau}(x)\varepsilon.$$

我们取 $L \gg 1$ 待定，并且定义

$$\widetilde{T}(x) = T^{\widetilde{\tau}(x)}x,\ \widetilde{\tau}_1(x) = \widetilde{\tau}(x),\ \widetilde{\tau}_2(x) = \widetilde{\tau}(\widetilde{T}(x)) + \widetilde{\tau}(x),\ \widetilde{\tau}_n(x) = \sum_{i=0}^{n-1} \widetilde{\tau}(\widetilde{T}^i x).$$

如此，存在 $k$，使得 $\widetilde{\tau}_k(x) \leq L < \widetilde{\tau}_{k+1}(x)$. 于是

$$\begin{aligned}\sum_{i=0}^{L-1} \underline{f}\circ T^i(x) &= \sum_{i=0}^{\widetilde{\tau}_1(x)-1} \underline{f}\circ T^i(x) + \sum_{i=\widetilde{\tau}_1(x)}^{\widetilde{\tau}_2(x)-1} \underline{f}\circ T^i(x) + \dots + \sum_{i=\widetilde{\tau}_{k-1}(x)}^{\widetilde{\tau}_k(x)-1} \underline{f}\circ T^i(x) + \sum_{i=\widetilde{\tau}_k(x)}^{L-1} \underline{f}\circ T^i(x) \\ &\geq \sum_{i=0}^{L-N} f\circ T^i(x) - \widetilde{\tau}_{k+1}(x)\varepsilon \geq \sum_{i=0}^{L-1} f\circ T^i(x) - (L+N)\varepsilon.\end{aligned}$$

两侧同除以 $L$ 并且积分，可得

$$\int \underline{f}\mathrm{d}m \geq \int f\mathrm{d}m - \left(1 + \frac{N}{L}\right)\varepsilon - \frac{N}{L}\int f\mathrm{d}m.$$

只要取 $L > \max\{\frac{N}{\varepsilon}, \frac{N\|f\|_{L^1}}{\varepsilon}\}$ 即可. □

最后，我们介绍 Kingman 次可加遍历定理，这个定理我们不加证明，旨在说明，此定理可以推出 Birkhoff 遍历定理.

**定理 1.78** (Kingman 次可加遍历定理)**.** *设 $(X, \mathcal{B}, m, T)$ 为保测系统 (这里我们不要求一定是概率测度)，我们称 $f_n \in L^0(X)$ 是次可加的，若对于任意的 $n, k \geq 1$ 有 $f_{n+k}(x) \leq f_n(x) + f_k \circ T^n(x)$. 若 $f_n^+ \in L^1(X)$ 那么有 $\frac{1}{n}f_n(x) \to f_*(x)$ $m$-a.e. $x$，其中 $f_*(x)$ 可取值 $-\infty$. 且 $f_* \circ T = f_*$ 以及 $f_*^+ \in L^1(X)$.*

**注 1.79.** 若 $f_n \in L^1(X)$ 那么根据次可加性质，两侧积分，得

$$\int f_{n+k}\mathrm{d}m \leq \int f_n\mathrm{d}m + \int f_k\mathrm{d}m.$$

令 $a_n = \int f\mathrm{d}m$. 则 $\{a_n\}$ 是次可加序列，自然满足关系 $\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\geq 1} \frac{a_n}{n}$.

下面我们利用 Kingman 次可加遍历定理证明 Birkhoff 遍历定理. 令

$$f_n(x) = \sum_{i=0}^{n-1} f\circ T^i(x).$$

于是很容易验证 $f_n$ 满足次可加性质，并且其他的条件也都满足，这样直接利用 Kingman 次可加遍历定理即可. 特别地，我们称满足 $f_{n+k} = f_n + f_k \circ T^n$ 的函数序列为加性上链，这显然是比次可加性质更强的一个性质.

**例 1.80.** 设 $T:\mathbb{R}^d\to\mathbb{R}^d$ 是微分同胚，并且保持 $\mathbb{R}^d$ 上的一个 Borel 测度，那么可得

$$DT^{n+k}(x)=D^k(T^nx)DT^n(x).$$

两侧取算子范数并且取对数，可得

$$\log\|DT^{n+k}(x)\|\log\|DT^k(T^nx)\|+\log\|DT^n(x)\|.$$

于是若 $\log\|DT(x)\|\in L^1(X)$ 则根据 Kingman 次可加遍历定理得到

$$\frac{\log\|DT^n(x)\|}{n}\to\lambda^+(x).$$

称为 $T$ 在 $x$ 点的 Lyapunov 指数.

### 1.5.3 一个来自随机过程的例子：遍历定理的运用

**定理 1.81.** *设 $P$ 是随机矩阵，即 $P=(p_{ij})_{0\le i,j\le d-1}$ 满足 $\sum_{j=0}^{d-1}p_{ij}=1$. 等价于*

$$P\begin{pmatrix}1\\1\\\dots\\1\end{pmatrix}=\begin{pmatrix}1\\1\\\dots\\1\end{pmatrix}.$$

*令 $p=(p_0,p_1,...,p_{d-1})$ 是概率向量，满足 $p_i>0$，$\sum_{i=0}^{d-1}p_i=1$ 并且 $pP=P$. 则下面的极限存在*

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}P^i:=Q.$$

*更进一步，极限矩阵 $Q$ 满足下列性质：*

1. *$Q$ 为随机矩阵；*
2. *$QP=PQ$；*
3. *$Q^2=Q$；*
4. *若 $vP=v$ 则 $vQ=v$.*

*证明.* 设 $(X,\mathcal{B},m,T)$ 为 $(p,P)$ 生成的 Markov 链 (详见例 1.52). 令 $Y=\{0,1,2,...,d-1\}$, $X=Y^{\mathbb{N}}$, $m$ 为 $(p,P)$ 生成的概率，对于任意的 $i\in\{0,1,2,...,d-1\}$，设 $[i]_k=\{x\in X:x_k=i\}$，$T$ 为左移位. 对于任意的 $j\in\{0,1,2,...,d-1\}$ 我们可以定义 $\chi_j=\chi_{[j]_0}$. 利用 Von-Neumann 遍历定理，可得

$$\frac{1}{n}\sum_{k=0}^{n-1}\chi_j\circ T^k\overset{L^2(X)}{\longrightarrow}\chi_j^*,\ (n\to\infty).$$

于是

$$\left(\frac{1}{n}\sum_{k=0}^{n-1}\chi_j\circ T^k,\chi_i\right)\to\left(\chi_i,\chi_j^*\right),\ (n\to\infty).$$

其中，我们有计算

$$\begin{aligned}\int\chi_i(\chi_j\circ T^k)\mathrm{d}m&=\int\chi_{[i]_0\cap(T^{-k}[j]_0)}\mathrm{d}m=m([i]_0\cap(T^{-k}[j]_0))\\&=m([i]_0\cap[j]_k)=\sum_{i_1,\dots,i_{k-1}}p_ip_{ii_1}p_{i_1i_2}\dots p_{i_{k-1}j}=p_ip_{ij}^{(k)},\end{aligned}$$

这里 $P^k=(p_{ij}^{(k)})$. 回到原式，得到

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}p_ip_{ij}^{(k)}\to(\chi_i,\chi_j^*).$$

于是就有

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}p_{ij}^{(k)}\to\frac{1}{p_i}(\chi_i,\chi_j^*):=q_{ij}.$$

令 $Q=(q_{ij})$ 即可. 下面我们证明 1-4 的性质. 对于 1，可知随机矩阵的乘积和凸组合都是随机矩阵，取极限以后自然还是随机矩阵，于是 $Q$ 是随机矩阵. 对于 2，我们注意到

$$P\left(\frac{1}{n}\sum_{i=0}^{n-1}P^i\right)=\left(\frac{1}{n}\sum_{i=0}^{n-1}P^i\right)P$$

两侧取极限即可. 另外，因为取极限有限项不影响，所以可知 $PQ=QP=Q$，根据这一点，可得 $QP^k=PQ^k=Q$，于是

$$Q\left(\frac{1}{n}\sum_{i=0}^{n-1}P^i\right)=Q.$$

两侧取极限 $Q^2=Q$. 另外，对于 $vP=P$，可知

$$v\left(\frac{1}{n}\sum_{i=0}^{n-1}P^i\right)=v,$$

两侧取极限，得 $vQ=v$. □

下面我们对上一定理中提到的保测系统的遍历性作下面的刻画.

**定理 1.82.** 设 $(X,\mathcal{B},m,T)$ 与 $p,P,,Q$ 有上一定理给出，那么下面的结论等价

1. *$m$ 遍历;*
2. *$Q$ 的每行都相同;*
3. *$Q$ 的每行都是正的;*
4. *对于任意的 $i,j$，存在 $k$ 使得 $p_{ij}^{(k)}>0$ (这时候我们称 $P$ 是不可约的);*
5. *1 是 $P$ 的代数与几何重数都为 1 的特征值.*

**注 1.83.** 在证明中，我们为了体现关系，并不局限于完成上述 5 条的循环，尽可能多的给出任意两条之间的证明.

**注 1.84.** 很显然，在第 5 条中，仅需要证明几何重数为 1 即可，这是因为如果代数与几何重数不相等，那么 $P$ 的 Jordan 标准型会出现

$$\begin{pmatrix}1&1\\0&1\end{pmatrix}$$

这样的形式，这显然导致 $\frac{1}{n}\sum_{i=1}^{n-1}P^i$ 的极限不存在.

*证明.* *1 ⇒ 2.* 根据遍历性，可得 $\chi_j^*$ 是常数，那么根据 $Q$ 的定义，有

$$q_{ij}=\frac{1}{p_i}(\chi_i,\chi_j^*)=\frac{\chi_j^*}{p_i}(\chi_i,1)=\chi_j,$$

这里用到了 $(\chi_i,1)=m([i]_0)=p_i$. 于是可得 $q_{ij}=\chi_j^*$，与 $i$ 无关. □

*2 ⇒ 3.* 由于 $q_{ij}$ 与 $i$ 无关，所以可令 $q_{ij}=q_j$. 根据关系 $pP=p$ 以及上一定理中关于 $Q$ 的第 4 条性质，可以推出 $pQ=p$. 于是

$$(p_0,p_1,...,p_{d-1})\begin{pmatrix}q_0...q_{d-1}\\ ...\ ...\ ...\\ q_0...q_{d-1}\end{pmatrix}=q=p.$$

根据 $p_i>0$ 可得 $q_i>0$ 对于任意的 $i$ 成立. □

*3 ⇒ 4.* 根据收敛性可得 $\frac{1}{n}\sum_{k=0}^{n-1}p_{ij}^{(k)}\to q_{ij}>0$. 这直接意味着 4 成立. □

*4 ⇒ 3.* 对于任意的 $\ell,j$ 存在 $i$，使得 $q_{\ell i}>0$ 并且对 $i,j$ 存在 $k$ 使得 $p_{ij}^{(k)}>0$. 根据前一个定理中的证明，可得 $QP^k=Q$，那么就有

$$q_{\ell j}=\sum_{s=0}^{d-1}q_{\ell s}p_{sj}^{(k)}\ge q_{\ell i}p_{ij}^{(k)}>0.$$

根据 $\ell,j$ 的任意性，结论成立. □

*2 ⇒ 1.* 任给基本矩形 $A=[i_0...i_r]_0$ 以及 $B=[j_0...j_s]_0$，其中下标 0 表示从第 0 个位置开始，类似可以有其他的下标，意义以此类推. 那么为证明遍历性，根据定理 1.66，仅需验证

$$\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}m(A\cap T^{-k}B)=m(A)m(B).$$

那么当 $k\ge r+1$ 时，可得

$$T^{-k}B=[j_0...j_s]_k$$

于是

$$\begin{aligned}\lim_{n\to\infty}\frac{1}{n}\sum_{k=0}^{n-1}m(A\cap T^{-k}B)&=\lim_{n\to\infty}\frac{1}{n}\sum_{k=r+1}^{n-1}m(A\cap T^{-k}B)\\ &=\lim_{n\to\infty}\frac{1}{n}\sum_{k=r+1}^{n-1}p_{i_0}p_{i_0i_1}...p_{i_{r-1}i_r}p_{i_rj_0}^{(k-r)}p_{j_0j_1}...p_{j_{s-1}j_s}\\ &=\lim_{n\to\infty}\frac{1}{n}p_{i_0}p_{i_0i_1}...p_{i_{r-1}i_r}\left(\sum_{k=r+1}^{n-1}p_{i_rj_0}^{(k-r)}\right)p_{j_0j_1}...p_{j_{s-1}j_s}\\ &=p_{i_0}p_{i_0i_1}...p_{i_{r-1}i_r}p_{j_0}p_{j_0j_1}...p_{j_{s-1}j_s}=m(A)m(B).\end{aligned}$$

于是这就证明了遍历性. □

*2 ⇒ 5.* 设 $vP=v$, 其中 $v$ 满足归 1 化, 即 $\sum_{i=0}^{d-1}v_i=1$. 再根据前面的证明可知如果 $Q$ 的每一行都相同, 必然可得 $Q$ 的每一行都是 $p$ 构成的行向量, 于是根据上一定理 $Q$ 的第 4 条性质, $vQ=v=p$ 这就说明了 1 的几何重数为 1，再根据前面的注记，就证明了 5. □

*5 ⇒ 2.* 利用 $QP=Q$ 再根据特征值的单性质，就得到 $Q$ 的每行都相等. □

*4 ⇒ 5.* 首先 $p$ 是 $P$ 的关于特征值 1 的特征向量. 假如还有某个特征向量 $v$ 与 $p$ 线性无关，$\lambda$ 使得 $w=v+\lambda p$ 的最小分量恰好为 0. 不失一般性，假设 $w_0=0$ 并且 $w_1>0$. 于是可得

$$w_1=(wP^k)_1=\sum_{j=0}^{d-1}w_jp_{j1}^{(k)}\ge w_1p_{11}^{(k)}.$$

这里我们根据 4. 可取充分大的 $k$ 使得 $p_{11}^{(k)}>0$ 则导出矛盾. □

事实上，我们证明里下图的结论，即完成了定理的证明.

$$1 \Leftrightarrow 2 \Rightarrow 3 \Leftrightarrow 4 \Rightarrow 5 \Leftrightarrow 2.$$

□

## 1.6 保测系统的等价

**定义 1.85** (保测系统的同构)**.** 设 $(X_i, \mathcal{B}_i, m_i, T_i)$ 是两个保测系统，其中 $i = 1, 2$，称 $T_1, T_2$ 是同构的，若存在 $M_i \in \mathcal{B}_i$ 满足 $m_i(M_i) = 1$ 以及可逆的保测变换 (这里的意思是逆变换也是可测的) $\varphi : (X_1, \mathcal{B}_1, m_1, T_1) \to (X_2, \mathcal{B}_2, m_2, T_2)$ 使得 $\varphi \circ T_1|_{M_1} = T_2 \circ \varphi|_{M_2}$ 成立.

**例 1.86.** 设 $T_1 : (X_1, \mathcal{B}_1, m_1) \to (X_1, \mathcal{B}_1, m_1)$ 为 $(\frac{1}{d}, ..., \frac{1}{d})$ $(d \geq 2)$ 的 Bernoulli 系统，$T_2 : (X_2, \mathcal{B}_2, m_2) \to (X_2, \mathcal{B}_2, m_2)$ 为圆周 $\mathbb{T}$ 上的 $\times d$ 映射. 则 $T_1, T_2$ 同构. 证明是容易的，简而言之就是 $d$ 进表示的互相转换. 我们取

$$M_1 = \{x \in X_1 : x \text{ 不终于 } 0 \text{ 或者 } d-1\},\ M_2 = \{x \in \mathbb{T} : x \text{ 具有唯一的 } d \text{ 进展式}\}.$$

下面我们定义 $\varphi : M_1 \to M_2$ 如下

$$(x_1, x_2, x_3, ...) \mapsto \sum_{i=1}^{\infty} \frac{x_i}{d^i}.$$

容易验证这就是一个同构.

设 $(X, \mathcal{B}, m)$ 是一个概率空间，$A, B \in \mathcal{B}$，我们定义等价关系 $A \sim B$ 当且仅当 $m(A \Delta B) = 0$. 因此我们可以定义 $\widetilde{\mathcal{B}} = \mathcal{B}/\sim$. 通过简单的观察，$\sim$ 关系对可数的交并补是相容的，即

$$\widetilde{A}, \widetilde{B} \in \widetilde{B}, \Rightarrow \widetilde{A} \cup \widetilde{B} = \widetilde{A \cup B},\ \widetilde{A}^c = \widetilde{A^c},\ \widetilde{A \cap B} = \widetilde{A \cap B}$$

以及

$$A_n \sim B_n,\ n \geq 1 \Rightarrow \bigcup_{n \geq 1} A_n \sim \bigcup_{n \geq 1} B_n.$$

可见 $\widetilde{\mathcal{B}}$ 也满足类似 $\sigma$-代数的性质. 另外此时可见 $m$ 也可以在等价类意义下定义. 这时候，我们称 $(\widetilde{\mathcal{B}}, \widetilde{m})$ 为 $(X, \mathcal{B}, m)$ 的测度代数.

**定义 1.87.** 设 $(X_i, \mathcal{B}_i, m_i)$，$i = 1, 2$ 为两个概率空间，相应的测度代数为 $(\widetilde{\mathcal{B}}_i, \widetilde{m}_i)$. 设有映射 $\Phi : \widetilde{\mathcal{B}}_2 \to \widetilde{\mathcal{B}}_1$ 满足保持集合运算以及保持测度，那么就称 $\Phi$ 为同态. 进一步，若 $\Phi$ 可逆，且逆可测，则称其为同构.

**例 1.88.** 设 $\varphi : (X_1, \mathcal{B}_1, m_1) \to (X_2, \mathcal{B}_2, m_2)$ 为保测的，即满足 $\varphi^{-1}\mathcal{B}_2 \subset \mathcal{B}_1$ 以及 $m_1(\varphi^{-1}A) = m_2(A)$. 令 $\Phi : \widetilde{\mathcal{B}}_2 \to \widetilde{\mathcal{B}}_1$ 满足 $\widetilde{B}_2 \mapsto \widetilde{C}$，其中 $C = \varphi^{-1}(B_2)$. 由此可得 $\Phi$ 为同态.

**定义 1.89** (保测系统的共轭)**.** 设 $(X_i, \mathcal{B}_i, m_i, T_i)$ 是两个保测系统，若两个测度代数同构，即存在 $\Phi : \widetilde{\mathcal{B}}_2 \to \widetilde{\mathcal{B}}_1$ 同构并且满足 $\widetilde{T}_1^{-1} \circ \Phi = \Phi \circ \widetilde{T}_2^{-1}$. 则称 $T_1$ 与 $T_2$ 共轭.

根据之前同构的定义，可见同构能够推出共轭，只需要将同构中的映射逆过来即可.

**定义 1.90** (保测系统的谱同构)**.** 设 $(X_i, \mathcal{B}_i, m_i, T_i)$ 是两个保测系统. 若存在等距同构 $W : L^2(X_2) \to L^2(X_1)$ 满足 $U_{T_1} \circ W = W \circ U_{T_2}$ 其中 $U_T(f) = f \circ T$，那么就称 $T_1$ 与 $T_2$ 是谱同构的.

一个简单的观察是利用共轭中的 $\Phi$ 定义 $L^2(X_2)$ 到 $L^2(X_1)$ 简单函数的等距同构，再延拓到全体生就得到了一个谱同构，这意味着共轭推出谱同构.

那么对于上面的三种定义，我们有下面的问题.

**问题 1.91.** 什么时候共轭可以推出同构，什么时候谱同构可以推出共轭?

前一个问题有一个结论，即标准概率空间上的共轭可以推出同构. 所谓标准概率空间就是在概率意义下和 $[0,1]$ 上的 Lebesgue 测度同构的概率空间. 另外我们不加证明地给出下面地命题.

**命题 1.92.** *完备可分度量空间上的 Borel 概率测度都是标准的.*

对于后一问题，情况是复杂的，我们需要借助后续知识来解决.

# 2 测度熵

## 2.1 测度熵的定义

### 2.1.1 熵函数的定义与分划的熵

**定义 2.1.** 设 $\varphi:[0,\infty)\to\mathbb{R}$ 满足

$$\varphi(x)=\begin{cases}x\log x & x>0,\\ 0 & x=0.\end{cases}$$

对于上面定义的函数 $\varphi$，我们有下面的性质

1. $\varphi$ 是严格凸函数，即对于任意的 $p=(p_1,p_2,...,p_n)$ 满足 $p_i\geq 0$，$\sum_{i=1}^n p_i=1$，$x_i\geq 0$，有
$$\varphi\left(\sum_{i=1}^n p_ix_i\right)\leq p_i\varphi(x_i).$$
等号成立当且仅当对于 $p_i>0$ 的那些 $x_i$ 都相等.
2. 对于任意的 $0<p<1$ 有 $-\varphi(p)>0$.

记 $\Delta_{n-1}=\{(p_1,p_2,...,p_n)\in\mathbb{R}^n:p_i\geq 0,\ \sum_{i=1}^n p_i=1\}$ 是 $n-1$ 维的标准单形.

**定义 2.2** (熵函数)**.** 令 $H:\Delta_{n-1}\to\mathbb{R}$ 满足

$$H(p_1,p_2,...,p_n)=-\sum_{i=1}^n\varphi(p_i)=-\sum_{i=1}^n p_i\log p_i.$$ [1]

我们称 $H$ 为熵函数.

对于上面定义的熵函数，有下列简单的性质.

1. $H$ 在 $\Delta_{n-1}$ 上连续.
2. $H$ 对坐标有对称性，即 $H(p_1,p_2,...,p_n)=H(p_{\sigma(1)},p_{\sigma(2)},...,p_{\sigma(n)})$，其中 $\sigma\in S_n$ 是 $\{1,2,...,n\}$ 上的置换群中的元素.
3. $H(p_1,p_2,...,p_n)=0$ 当且仅当存在 $i$ 使得 $p_i=1$.

[1] 这里我们直接定义 $0\log 0=0$ 所以不区分 $\varphi$ 与 $x\log x$.

4. $H$ 为严格的凹函数，即对于任意的 $t \in (0,1)$ 有 $H(tp+(1-t)q) \geq tH(p)+(1-t)H(q)$，其中 $p=(p_1,p_2,...,p_n)$，$q=(q_1,q_2,...,q_n)$. 等号成立当且仅当 $p=q$.
5. $H$ 在 $(\frac{1}{n},\frac{1}{n},...,\frac{1}{n})$ 处取得最大值 $\log n$. 事实上根据 $\varphi$ 的严格凸性质,对于任意的 $(p_1,p_2,...,p_n) \in \Delta_{n-1}$，满足
$$\frac{1}{n}H(p_1,p_2,...,p_n) = -\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \leq -\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \leq -\varphi\left(\frac{1}{n}\sum_{i=1}^{n}p_i\right) = -\varphi\left(\frac{1}{n}\right) = \frac{1}{n}\log n.$$
等号成立当且仅当 $p_i=\frac{1}{n}$ 对任意的 $1 \leq i \leq n$ 都成立.

**定义 2.3** (概率空间的分划)**.** 对于概率空间 $(X,\mathcal{B},m)$ 若集族 $\{A_i\}_{i=1}^n$ 满足下列条件

1. $A_i \in \mathcal{B}$;
2. $A_i \cap A_j = \emptyset$ 对于任意的 $i \neq j$ 都成立;
3. $\cup_{i=1}^n A_i = X$,

那么我们称 $\{A_i\}_{i=1}^n$ 是 $X$ 的一个 (有限可测) 的分划. 一般用希腊字母记, 例如 $\xi=\{A_1,A_2,...,A_n\}$.

**定义 2.4** (分划的熵)**.** 对于概率空间 $(X,\mathcal{B},m)$ 上的一个分划 $\xi=\{A_i\}_{i=1}^n$，定义 $\xi$ 的熵为
$$H(\xi)=H(m(A_1),...,m(A_n)).$$

**定义 2.5** (分划的运算与关系)**.** 设 $\xi=\{A_i\}_{i=1}^n$ 与 $\eta=\{B_j\}_{j=1}^m$ 是概率空间 $(X,\mathcal{B},m)$ 上的两个分划，我们记
$$\xi \vee \eta = \{A_i \cap B_j : 1 \leq i \leq n,\ 1 \leq j \leq m\}$$
为分划 $\xi,\eta$ 的交. 若对于任意的 $1 \leq j \leq m$，存在 $1 \leq i \leq n$ 使得 $B_j \subset A_i$ 则称 $\eta$ 是 $\xi$ 的加细.

根据上面的定义，我们很容易得到性质 $\xi \vee \eta \geq \eta,\xi$. 因此可见 $\xi \vee \eta$ 是比 $\eta,\xi$ 都要细的分划中的最粗的那一个.

**定义 2.6** (信息函数)**.** 设 $\xi=\{A_i\}_{i=1}^n$ 是概率空间 $(X,\mathcal{B},m)$ 上的一个分划，那么对应这个分划的信息函数定义为
$$I_\xi(x) = -\sum_{i=1}^{n}\log(m(A_i))\chi_{A_i}(x).$$

很显然，这时候 $\int_X I_\xi(x)\mathrm{d}m = H(\xi)$. 从中可见 $H(\xi)$ 表示的是对于分划 $\xi$ 的平均信息量.

**定理 2.7.** 设 $\xi=\{A_i\}_{i=1}^n$ 与 $\eta=\{B_j\}_{j=1}^m$ 是概率空间 $(X,\mathcal{B},m)$ 上的两个分划，那么 $H(\xi \vee \eta) \leq H(\xi)+H(\eta)$.

*证明.* 根据简单的计算，我们有
$$\begin{aligned}
H(\xi\vee\eta)-H(\eta) &= -\sum_{i=1}^{n}\sum_{j=1}^{m} m(A_i\cap B_j)\log(m(A_i\cap B_j)) + \sum_{j=1}^{m} m(B_j)\log(m(B_j)) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{m} m(A_i\cap B_j)\log\left(\frac{m(A_i\cap B_j)}{m(B_j)}\right) \\
&= -\sum_{j=1}^{m} m(B_j)\left(\sum_{i=1}^{n}\frac{m(A_i\cap B_j)}{m(B_j)}\log\left(\frac{m(A_i\cap B_j)}{m(B_j)}\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j=1}^{m} m(B_j)H(p^j),\ p^j = \left(\frac{m(A_1\cap B_j)}{m(B_j)}, \frac{m(A_2\cap B_j)}{m(B_j)}, ..., \frac{m(A_n\cap B_j)}{m(B_j)}\right) \\
&\le H\left(\sum_{j=1}^{m} m(B_j)p^j\right) = H(\xi).
\end{aligned}
$$

这里的不等式成立的原因来自于 $H$ 是严格凹的函数. □

下面我们来分析上述定理等号成立的条件. 为此，我们先做一些准备工作. 给定 $B_j$ 定义

$$m_j(A) = \frac{m(A\cap B_j)}{m(B_j)},$$

并称其为 $B_j$ 上的条件概率. 记作 $m(\cdot|B_j)$. 从上述计算可得

$$H(\xi\vee\eta) - H(\eta) = \sum_{j=1}^{m} m(B_j)H(m(A_1|B_j), ..., m(A_n|B_j)).$$

于是，我们将 $\sum_{j=1}^{m} m(B_j)H(m(A_1|B_j), ..., m(A_n|B_j))$ 记作 $H(\xi|\eta)$，称为 $\xi$ 在 $\eta$ 下的条件熵. 因此，我们有 $H(\xi\vee\eta) = H(\eta) + H(\xi|\eta)$. 我们很容易得到

$$0 \le H(\xi|\eta) \le H(\xi).$$

另一方面，我们注意到 $H(\xi) \ge 0$ 等号成立条件是分划中有全测集合. 所以 $H(\xi\vee\eta) = H(\eta)$ 当且仅当 $H(\xi|\eta) = 0$，当且仅当对于任意的 $j$，若 $m(B_j) > 0$ 则存在 $i$ 使得 $m(A_i|B_j) = 0$，这意味着在相差零测集的意义下 $B_j \subset A_i$，这是，我们记作 $B_j \overset{\circ}{\subset} A_i$，由此也可以推广加细 $\xi \overset{\circ}{\le} \eta$.

接下来，我们考虑 $H(\xi|\eta) = H(\xi)$ 等号成立的条件，也就是定理 2.7 中等号成立的条件. 根据 $H$ 的严格凹的性质，可得等号成立，当且仅当对于任意的 $j$ 若 $m(B_j) > 0$ 则相应的 $p^j$ 相等，即对于任意的 $j,k$，$m(B_j) > 0, m(B_k) > 0$ 有 $p^j = p^k$，也就是

$$\frac{m(A_i\cap B_j)}{m(B_j)} = \frac{m(A_i\cap B_k)}{m(B_k)} \Leftrightarrow m(B_k)m(A_i\cap B_j) = m(B_k)m(A_i\cap B_k).$$

对 $m(B_k) > 0$ 的 $k$ 求和，就得到 $m(A_i\cap B_j) = m(A_i)m(B_j)$. 总结来说，$H(\xi|\eta) = H(\xi)$ 当且仅当对于任意的 $i,j$ 有 $m(A_i\cap B_j) = m(A_i)m(B_j)$. 这时候我们称 $\xi,\eta$ 是独立的.

利用定理 2.7，我们可以得到下面的定理.

**定理 2.8.** *设 $\xi,\eta,\zeta$ 是概率空间 $(X,\mathcal{B},m)$ 的三个分划，那么*

$$H(\xi\vee\eta|\zeta) = H(\xi|\zeta) + H(\eta|\xi\vee\zeta).$$

*证明.* 根据定理 2.7 (事实上是条件熵的公式) 可得

$$H(\xi\vee\eta\zeta) = H(\xi\vee\zeta) + H(\eta|\xi\vee\zeta) = H(\zeta) + H(\xi|\zeta) + H(\eta|\xi\vee\zeta).$$

另一方面

$$H(\xi\vee\eta|\zeta) = H(\zeta) + H(\xi\vee\eta|\zeta).$$

两式结合即得到结论. □

**推论 2.9.** *若 $\xi \le \eta$ 则 $H(\xi|\zeta) \le H(\eta|\zeta)$.*

*证明.* 若 $\xi \le \eta$ 那么 $\zeta\vee\eta = \eta$. 利用上面的等式

$$H(\eta|\zeta) = H(\xi\vee\eta|\zeta) = H(\xi|\zeta) + H(\eta|\xi\vee\zeta).$$

这样证明了结论成立. □

类似地，我们还可以证明以下关系.

**命题 2.10.** 对于分划 $\xi, \eta, \zeta$，有

1. $H(\xi \vee \zeta|\eta) \leq H(\xi|\eta) + H(\zeta|\eta)$.
2. $H(\xi|\zeta) \leq H(\xi|\eta)$ 若 $\zeta \geq \eta$.

证明. 首先我们断言前者是后者的推论，这是因为根据第二条以及 $\eta \leq \xi \vee \eta$ 可得

$$H(\xi \vee \zeta|\eta) = H(\xi|\eta) + H(\zeta|\xi \vee \eta) \leq H(\xi|\eta) + H(\zeta|\eta).$$

下面我们来证明第二条. 设 $\xi = \{A_i\}_{i=1}^n$，$\zeta = \{C_j\}_{j=1}^m$，$\eta = \{D_k\}_{k=1}^{\ell}$. 因为 $\eta \leq \zeta$，可得 $C_j \cap D_k = \emptyset$ 或者 $D_k$. 进而有

$$\frac{m(A_i \cap C_j)}{m(C_j)} = \sum_k \frac{m(C_j \cap D_k)}{m(C_j)} \cdot \frac{m(A_i \cap D_k)}{m(D_k)}.$$

利用函数 $\varphi(x) = x \log x$ 的凸性质，令

$$\alpha_k = \frac{m(C_j \cap D_k)}{m(C_j)},\ x_k = \frac{m(A_i \cap D_k)}{m(D_k)},$$

可得

$$\begin{aligned}\frac{m(A_i \cap C_j)}{m(C_j)} \log \left(\frac{m(A_i \cap C_j)}{m(C_j)}\right) &= \phi\left(\frac{m(A_i \cap C_j)}{m(C_j)}\right) = \phi\left(\sum_k \alpha_k x_k\right) \leq \sum_k \alpha_k \phi(x_k) \\ &= \sum_k \frac{m(C_j \cap D_k)}{m(C_j)} \phi\left(\frac{m(C_j \cap D_k)}{m(C_j)}\right).\end{aligned}$$

两侧同乘 $m(C_j)$ 得到

$$m(A_i \cap C_j) \log \left(\frac{m(A_i \cap C_j)}{m(C_j)}\right) \leq \sum_k m(C_j \cap D_k) \phi\left(\frac{m(C_j \cap D_k)}{m(C_j)}\right).$$

对 $i, j$ 求和，则有

$$\sum_{i,j} m(A_i \cap C_j) \log \left(\frac{m(A_i \cap C_j)}{m(C_j)}\right) \leq \sum_{i,k} m(A_i \cap D_k) \log \left(\frac{m(A_i \cap D_k)}{m(D_k)}\right),$$

即 $-H(\xi|\zeta) \geq -H(\xi|\eta)$. 这就证明了上述关系. □

对于 $H$ 我们还可以推广到无限求和的情形. 对于概率空间 $(X, \mathcal{B}, m)$ 可以推广到可数的分划 $\xi = \{A_i\}_{i=1}^{\infty}$ 即分划中由可数个两两不相交的集合构成. 此时满足 $\sum_{i=1}^{\infty} m(A_i) = 1$，那么我们可以定义

$$H(\xi) = -\sum_{i=1}^{\infty} m(A_i) \log(m(A_i)).$$

上述求和可能为 $\infty$. 对于不可数的情形，我们考虑下面的例子. 考虑 $[0,1)^2$ 上的 Lebesgue 测度以及分划 $\xi$ 为

$$[0,1)^2 = \bigcup_{x \in [0,1]} I_x = \bigcup_{x \in [0,1)} \{x\} \times [0,1).$$

我们定义对于这个分划的 $H(\xi)$. 首先可得只存在可数个 $\{x_i\}_{i=1}^{\infty}$ 使得 $m(I_{x_i}) > 0$. 这样就转化为可数的情形了. 若 $m(\cup_{i=1}^{\infty} I_{x_i}) < 1$，则定义 $H(\xi) = \infty$. 若 $m(\cup_{i=1}^{\infty} I_{x_i}) = 1$，则定义 $H(\xi) = -\sum_{i=1}^{\infty} m(I_{x_i}) \log(m(I_{x_i}))$.

**例 2.11.** 考虑概率空间 $(X, \mathcal{B}, m)$，自然可以有分划 $X = A_0 \cup A_1$ 满足 $m(A_0) = m(A_1) = \frac{1}{2}$. 我们可以对 $A_1$ 在概率意义下等分为 $A_1 = B_1 \cup B_2 \cup ... \cup B_n$ 满足 $m(B_i) = \frac{1}{2n}$ 其中 $i = 1, 2, ..., n$. 这个分划 $\xi_n$ 的熵为

$$H(\xi_n) = \frac{1}{2} \log 2 + \frac{n}{2n} \log(2n) = \frac{\log 2}{2} + \frac{1}{2} \log(2n) \to \infty.$$

### 2.1.2 测度熵的定义

设 $(X, \mathcal{B}, m, T)$ 为保测系统，设 $\xi = \{A_i\}_{i=1}^n$ 为一分划，可知 $\{T^{-1}A_i\}_{i=1}^n$ 也是分划 (这里用到了 $T$ 的保测性质). 于是可得

$$H(T^{-1}\xi) = -\sum_{i=1}^{n} m(T^{-1}A_i)\log(m(T^{-1}A_i)) = H(\xi).$$

定义 $a_n = H(\xi \vee T^{-1}\xi \vee ... \vee T^{-n+1}\xi) = H(\vee_{i=0}^{n-1}T^{-i}\xi)$.

对于数列 $a_n$ 我们有下列引理.

**引理 2.12.** *$a_{n+m} \leq a_n + a_m$ 对于 $n, m \geq 1$ 成立.*

*证明.* 利用 $H(\xi \vee \eta) \leq H(\xi) + H(\eta)$，可得

$$\begin{aligned}
a_{n+m} &= H(\xi \vee T^{-1}\xi \vee ... \vee T^{-(n+m)+1}\xi) = H\left(\bigvee_{i=0}^{n-1} T^{-i}\xi \vee \bigvee_{i=n}^{n+m-1} T^{-i}\xi\right) \\
&\leq H\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right) + H\left(\bigvee_{i=n}^{n+m-1} T^{-i}\xi\right) = H\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right) + H\left(T^{-n}\left(\bigvee_{i=0}^{m-1} T^{-i}\xi\right)\right) \\
&= H\left(\bigvee_{i=0}^{n-1} T^{-i}\xi\right) + H\left(\bigvee_{i=0}^{m-1} T^{-i}\xi\right) = a_n + a_m.
\end{aligned}$$

上面倒数第二等式我们利用了 $T$ 的保测性. □

在注 1.79 中，已经出现了次可加序列，下面我们总结为下面的引理方便使用.

**引理 2.13.** *若序列 $\{b_n\}$ 满足 $b_{n+m} \leq b_n + b_m$，则 $\lim_{n\to\infty} \frac{b_n}{n} = \inf_{n\geq 1} \frac{b_n}{n}$，这里的 $\inf_{n\geq 1} \frac{b_n}{n}$ 可能为 $-\infty$.*

于是根据上面的分析，$\{a_n\}$ 是次可加，则有 $\lim_{n\to\infty} \frac{a_n}{n}$ 存在并且满足极限在 $[0, \infty)$ 中. 由此，我们可以给出下面的定义.

**定义 2.14.** 对于保测系统 $(X, \mathcal{B}, m, T)$，定义

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

为分划 $\xi$ 在 $T$ 下的测度熵. 我们称 $h_m(T) = \sup_\xi h_m(\xi, T)$ 是 $T$ 的测度熵. 这里上面的 sup 是取遍所有的可测分划而得的.

**注 2.15.** 在不引起歧义的时候，我们可以用 $h(T)$ 与 $h(\xi, T)$ 简记 $h_m(T)$ 与 $h_m(\xi, T)$.

**注 2.16.** 为了简便起见，对于分划 $\xi$，我们用 $\xi_i^j$ 记作 $T^{-i}\xi \vee T^{-i-1}\xi \vee ... \vee T^{-j+1}\xi$. 若 $i = 0$，我们采用 $\xi^j$ 表示 $\xi_0^j$.

## 2.2 测度熵的性质

### 2.2.1 测度熵的基本性质

**定理 2.17.** *设 $(X, \mathcal{B}, m, T)$ 是保测系统，$\xi, \eta$ 是有限可测的分划，则有下列性质*

*性质 1* $h_m(\xi, T) \leq H(\xi)$.

*性质 2* $h_m(\xi \vee \eta, T) \leq h_m(\xi, T) + h_m(\eta, T)$.

*性质 3* 若 $\xi \le \eta$，则 $h_m(\xi, T) \le h_m(\eta, T)$.

*性质 4* $h_m(\xi, T) \le h_m(\eta, T) + H(\xi|\eta)$.

*性质 5* $|h_m(\xi, T) - h_m(\eta, T)| \le H(\xi|\eta) + H(\eta|\xi)$.

*性质 6* 对于任意的 $k \in \mathbb{Z}_+$ 有 $h_m(\xi, T) = h_m(T^{-1}\xi, T) = h_m(\vee_{i=0}^{k-1} T^{-i}\xi, T)$.

*性质 7* 若 $T$ 是可逆的，对于任意的 $k \in \mathbb{Z}_+$，则有 $h_m(\xi, T) = h_m(\vee_{i=0}^{k-1} T^{i}\xi, T) = h(\vee_{i=-k}^{k} T^{-i}\xi, T)$.

*性质 8* 对于任意的 $k \in \mathbb{Z}_{\ge 0}$ 有 $h_m(T^k) = k h_m(T)$，其中约定 $T^0 = \mathrm{Id}$.

*性质 9* 若 $T$ 可逆，则对于任意的 $k \in \mathbb{Z}$，那么 $h_m(T^k) = |k| h_m(T)$.

*性质 10* $\lim_{n\to\infty} \frac{1}{n} H(\xi \vee T^{-1}\xi \vee ... \vee T^{-n+1}\xi) = \lim_{n\to\infty} H(\xi | T^{-1}\xi \vee ... \vee T^{-n+1}\xi)$.

*性质 11* $\frac{1}{n} H(\vee_{i=0}^{n-1} T^{-i}\xi)$ 单调减少收敛到 $h_m(\xi, T)$.

在证明上述定理之前，我们记所有的有限可测分划 $\xi$ 构成的集合为 $\mathcal{P}$，并且由 $\mathcal{P}_n$ 表示所有的由 $n$ 个元素构成的可测分划的集合，即

$$\mathcal{P}_n = \{\xi \in \mathcal{P} : \#\xi = n\}.$$

因为分划可以自然加上零测度的集合，所以我们有

$$\mathcal{P}_1 \subset \mathcal{P}_2 \subset ... \subset \mathcal{P}_n \subset ...$$

*性质 1 的证明.* 根据之前引理 2.12 与 2.13 可得 $h_m(\xi, T) = \lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\ge 1} \frac{a_n}{n} \le a_1 = H(\xi)$，这就证明了此性质. □

*性质 2 的证明.* 根据定理 2.7 有

$$H((\xi \vee \eta) \vee T^{-1}(\xi \vee \eta) \vee ... \vee T^{-n+1}(\xi \vee \eta)) = H(\xi^n \vee \eta^n) \le H(\xi^n) + H(\eta^n).$$

两侧同除以 $n$ 并且取极限，即可得到性质 2 的结论. □

*证明.* 此性质直接来源于 $H$ 保持分划的不等号性质. 事实上 $H(\xi^n) \le H(\eta^n)$，于是两侧同除以 $n$ 取极限即可. □

*性质 4 的证明.* 根据条件熵的性质，以及命题 2.10 可得

$$\begin{aligned} H(\xi^n \vee \eta^n) &= H(\eta^n) + H(\xi^n|\eta^n) = H(\eta^n) + H(\xi \vee T^{-1}\xi \vee ... \vee T^{-n+1}\xi | \eta \vee T^{-1}\eta \vee ... \vee T^{-n+1}\eta) \\ &= H(\eta^n) + H(\xi|\eta^n) + ... + H(T^{-n+1}\xi|\eta^n) \\ &\le H(\eta^n) + H(\xi|\eta) + H(T^{-1}\xi|T^{-1}\eta) + ... + H(T^{-n+1}\xi|T^{-n+1}\eta) \\ &= H(\eta^n) + nH(\xi|\eta). \end{aligned}$$

上面最后一个等式，我们利用了 $T$ 的保测性. 对上式两侧同除以 $n$ 并对 $n$ 取极限即可得到性质 4 的证明，这里我们还用到了 $H(\xi^n \vee \eta^n) = H((\xi \vee \eta)^n)$. □

*性质 5 的证明.* 根据性质 4，性质 5 是显然的. □

性质 5 中，定义 $\rho(\xi, \eta) = H(\xi|\eta) + H(\eta|\xi)$. 并且我们有下列引理.

**引理 2.18.** *$\rho(\xi, \eta)$ 是 $\mathcal{P}$ 上的度量. 这里对于 $\mathcal{P}$ 中的元素，事实熵是在相差一个等价类意义下的，即若 $\xi \stackrel{\circ}{=} \eta$，则将两者视为同一元素. 其中 $\xi \stackrel{\circ}{=} \eta$ 的含义是，对于任意的 $A \in \xi$，存在 $B \in \eta$，$m(A \Delta B) = 0$.*

**注 2.19.** 在上述引理的意义下，可见 $h_m(\xi,T)$ 关于度量 $\rho(\cdot,\cdot)$ 是 Lipschitz 函数.

*引理 2.18 的证明.* 要证明是度量，需要验证三点度量的性质. 首先 $\rho(\cdot,\cdot)$ 具有对称性和正定性是显然的. 因此我们只需要验证三角不等式. 根据 $\rho(\eta,\xi)=H(\eta|\xi)+H(\xi|\eta)$，$\rho(\zeta,\xi)=H(\zeta|\xi)+H(\xi|\zeta)$ 以及 $\rho(\eta,\zeta)=H(\eta|\zeta)+H(\zeta|\eta)$. 因此归结为证明

$$H(\eta|\xi)\le H(\eta|\zeta)+H(\zeta|\xi).$$

事实上根据推论 2.9 和命题 2.10，我们有

$$H(\eta|\xi)\le H(\eta\vee\zeta|\xi)=H(\zeta|\xi)+H(\eta|\xi\vee\zeta)\le H(\zeta|\xi)+H(\eta|\zeta).$$

这样就完成了 $\rho(\cdot,\cdot)$ 是度量的性质的证明. □

*性质 6 的证明.* 根据 $T$ 的保测性，可得

$$h_m(T^{-1}\xi,T)=\lim_{n\to\infty}\frac{H(T^{-1}\xi\vee T^{-2}\xi\vee\ldots\vee T^{-n}\xi)}{n}=\lim_{n\to\infty}\frac{H(\xi\vee T^{-1}\xi\vee\ldots\vee T^{-n+1}\xi)}{n}=h_m(\xi,T).$$

更进一步有

$$H\left(\bigvee_{j=0}^{n-1}T^{-j}\left(\bigvee_{i=0}^{k-1}T^{-i}\xi\right)\right)=H\left(\bigvee_{i=0}^{n+k-2}T^{-i}\xi\right),$$

所以 $h_m(\xi,T)=h_m(\vee_{i=0}^{k-1}T^{-i}\xi,T)$. □

*性质 7 的证明.* 根据性质 6 的证明方法以及观察 $H(\xi)=H(T\xi)=H(T^{-1}\xi)$ 即可. □

*性质 8 的证明.* 首先对于 $k=0$ 时，显然有 $H(\vee_{i=0}^{n-1}\mathrm{Id}^{-i}\xi)=H(\xi)$，于是 $h_m(\mathrm{Id},\xi)=0$ 对于任意的 $\xi\in\mathcal{P}$ 都成立，所以 $h_m(\mathrm{Id})=0$. 对于 $k\ge 1$，根据性质 3，可得

$$\begin{aligned}h(\xi,T^k)\le h\left(\bigvee_{i=0}^{k-1}T^{-i}\xi,T^k\right)&=\lim_{n\to\infty}\frac{H\left(\bigvee_{j=0}^{n-1}T^{-jk}\left(\bigvee_{i=0}^{k-1}\xi\right)\right)}{n}\\&=\lim_{n\to\infty}\frac{1}{nk}H\left(\bigvee_{i=0}^{nk-1}T^{-i}\xi\right)\cdot k=kh_m(\xi,T).\end{aligned}$$

于是对 $\xi\in\mathcal{P}$ 取上确界即可得到 $h_m(T^k)\le kh_m(T)$. 上面的计算还意味着

$$kh_m(\xi,T)=h\left(\bigvee_{i=0}^{k-1}T^{-i}\xi,T^k\right)\le\sup_{\xi\in\mathcal{P}}h(\xi,T^k).$$

于是就得到 $kh(T)\le h(\xi,T^k)$，这样就证明了性质 8. □

*性质 9 的证明.* 根据性质 8，只需要证明 $h(\xi,T^{-1})=h(\xi,T)$ 即可. 这是因为

$$h(\xi,T^{-1})=\lim_{n\to\infty}\frac{1}{n}H(\xi\vee T\xi\vee\ldots\vee T^{n-1})=\lim_{n\to\infty}\frac{1}{n}H(T^{n-1}(\xi\vee T\xi\vee\ldots\vee T^{-n+1}))=h(\xi,T).$$

这样就完成了性质 9 的证明. □

*性质 10 的证明.* 我们断言

$$H(\xi\vee T^{-1}\xi\vee\ldots\vee T^{-n+1}\xi)=\sum_{j=0}^{n-1}H\left(\xi\,\middle|\,\bigvee_{i=1}^{j}T^{-i}\xi\right).$$

在这一断言下，性质 10 是显然的. 下面我们用归纳法证明上述断言. 对于 $n=2$ 时，有

$$H(\xi \vee T^{-1}\xi) = H(\xi) + H(T^{-1}\xi|\xi) = H(T^{-1}\xi) + H(\xi|T^{-1}\xi) = H(\xi) + H(\xi|T^{-1}\xi),$$

其中最后一个等式利用了 $T$ 的保测性. 对于 $n \geq 3$ 时，利用条件熵公式，可得

$$\begin{aligned} H(\xi \vee T^{-1}\xi \vee ... \vee T^{-n+1}\xi) &= H(T^{-1}\xi \vee ... \vee T^{-n+1}\xi) + H(\xi|T^{-1}\xi \vee ... \vee T^{-n+1}\xi) \\ &= H(\xi \vee ... \vee T^{-n+2}\xi) + H(\xi|T^{-1}\xi \vee ... \vee T^{-n+1}\xi). \end{aligned}$$

于是根据归纳假设断言成立. □

*性质 11 的证明*. 我们仅需要证明 $\frac{1}{n}H(\xi^n)$ 的单调减少的性质即可. 归结为证明 $nH(\xi^{n+1}) \leq (n+1)H(\xi^n)$. 另外，根据条件熵公式，可得

$$nH(\xi^{n+1}) = n(H(\xi^n) + H(\xi|\xi_1^n) \leq nH(\xi^n) + nH(\xi|\xi_1^{n-1}).$$

于是我们仅需要证明 $nH(\xi|\xi_1^{n-1}) \leq H(\xi^n)$ 即可. 这是显然的，直接由性质 10 的断言即可得到. □

### 2.2.2 测度熵是同构不变量

下面我们给出一个重要定理，该定理表明，测度熵在保测系统的同构意义下不变.

**定理 2.20.** 设 $(X_i, \mathcal{B}_i, m_i, T_i)$ $i=1,2$ 是两个保测系统，并且 $\varphi : (X_1, \mathcal{B}_1, m_1, T_1) \to (x_2, \mathcal{B}_2, m_2, T_2)$ 是同构映射，那么 $h_{m_1}(T_1) = h_{m_2}(T_2)$.

*证明*. 根据同构的定义我们可以得到 $M_i$ $i=1,2$ (见定义 1.85). 设 $\xi = \{A_i\}_{i=1}^n$ 是 $M_2$ 的有限可测分划，则

$$\varphi^{-1}\xi = (\varphi^{-1}A_1, \varphi^{-1}A_2, ..., \varphi^{-1}A_n)$$

是 $M_1$ 的有限可测分划，由于 $\varphi$ 保测，那么可得

$$m_1(\varphi^{-1}A_i) = m_2(A_i),\ i = 1, 2, ..., n,$$

因此

$$h_{m_2}(T_2, \xi) = h_{m_1}(T_1, \phi^{-1}\xi).$$

故

$$h_{m_2}(T_2) = \sup_{\xi \in \mathcal{P}} h_{m_2}(T_2, \xi) \leq \sup_{\eta \in \mathcal{P}} h_{m_1}(T_1, \eta) = h_{m_1}(T_1).$$

根据对称性可得 $h_{m_2}(T_2) \geq h_{m_1}(T_1)$ 于是就证明了 $h_{m_2}(T_2) = h_{m_1}(T_1)$. □

**注 2.21.** 事实上上述证明中，我们证明了如果 $\varphi$ 是保测的 (不要求可逆)，就能得到 $h_{m_2}(T_2) \geq h_{m_1}(T_1)$.

## 2.3 测度熵的计算

### 2.3.1 Kolmogorov-Sinai 定理

设 $(X, \mathcal{B}, m)$ 是一个概率空间，$\mathcal{C}, \mathcal{D}$ 是 $\mathcal{B}$ 的两个 $\sigma$-代数，如果对于每个 $C \in \mathcal{C}$，存在 $D \in \mathcal{D}$ 使得 $m(C\Delta D) = 0$，又对每个 $D' \in \mathcal{D}$，存在 $C' \in \mathcal{C}$，使得 $m(C'\Delta D') = 0$，则记作 $\mathcal{C} \overset{\circ}{=} \mathcal{D}$.

**定义 2.22.** 设 $T : (X, \mathcal{B}, m) \to (X, \mathcal{B}, m)$ 为概率空间的保测变换，考虑分解 $\eta = \{D_i\}_{i=1}^r \in \mathcal{P}$. 并记 $\vee_{n=0}^{\infty} T^{-n}\eta$ 为包含所有的 $T^{-i}\eta$ $(i = 0, 1, 2, ., , , )$ 的最小 $\sigma$-代数，如果 $\vee_{n=0}^{\infty} T^{-n}\eta \overset{\circ}{=} \mathcal{B}$，则称 $\eta$ 为 $T$ 的生成子.

这里我们需要对 $\vee_{n=0}^{\infty}T^{-n}\eta$ 的含义做一个更加精确的表述. 用 $\mathcal{A}_n$ 记分解 $\vee_{i=0}^{n}T^{-i}\eta$ 之并集之类, 则 $\mathcal{A}_n$ 是子 $\sigma$-代数. 用 $\cup_{n=0}^{\infty}\mathcal{A}_n$ 记这种 $B\in\mathcal{B}$ 的并: 对于某个 $n$, $B\in\mathcal{A}_n$. 则 $\cup_{n=0}^{\infty}\mathcal{A}_n$ 是代数但不是 $\sigma$-代数 (可数并不封闭). 则 $\vee_{n=0}^{\infty}T^{-i}\eta$ 是含代数 $\cup_{n=0}^{\infty}\mathcal{A}_n$ 的最小的 $\sigma$-代数. 基于这一点, 我们给出 $\vee_{n=0}^{\infty}T^{-n}\eta \overset{\circ}{=} \mathcal{B}$ 的等价定义.

**定义 2.23.** 若对任意的 $B\in\mathcal{B}$, $\varepsilon>0$, 存在 $N\in Z_+$ 和集合

$$A_1, A_2, ..., A_k \in \bigvee_{i=0}^{N} T^{-i}\eta,$$

使得 $m(B\Delta\cup_{i=1}^{k}A_i)<\varepsilon$, 则称 $\eta$ 为 $T$ 的生成子.

下面我们给出一个重要的定理.

**定理 2.24** (Kolmogorov-Sinai)**.** *设 $(X,\mathcal{B},m,T)$ 是一个保测系统, $\eta\in\mathcal{P}$ 是有限可测分划, 并且是 $T$ 的生成子, 则 $h_m(T)=h_m(\eta,T)$.*

上述定理说明, 生成子在 $T$ 下的测度熵就等于 $T$ 的熵. 换言之对有限可测分划取上确界, 实际上如果生成子存在, 则最大值就是分划取作生成子的情形.

*Kolmogorov-Sinai 定理的证明.* 我们仅需证明对于任意的 $\alpha\in\mathcal{P}$, 有 $h_m(\alpha,T)\le h_m(\eta,T)$. 于是往证固定 $\alpha=\{A_i\}_{i=1}^{k}\in\mathcal{P}_k\subset\mathcal{P}$, 对于任意的 $\varepsilon>0$, 有

$$h_m(\alpha,T)\le h_m(\eta,T)+\varepsilon.$$

根据定理 2.17 的性质 4 与 6, 可得

$$h_m(\alpha,T)\le h_m(\eta^n,T)+H(\alpha|\eta^n)=h_m(\eta,T)+H(\alpha|\eta^n).$$

于是往证当 $n$ 充分大的时候, $H(\alpha|\eta^n)$ 充分小. 为此, 我们先证明下面的引理.

**引理 2.25.** *对于任意的 $k\in Z_+$, $\varepsilon>0$, 存在 $\delta>0$ 使得对于任意的 $\alpha=\{A_i\}_{i=1}^{k}, \beta=\{B_i\}_{i=1}^{k}\in\mathcal{P}_k$, 只要 $m(A_i\Delta B_i)<\delta$ 对 $i=1,2,...,k$ 都成立, 则有 $H(\alpha|\beta)<\varepsilon$.*

*引理的证明.* 由于 $H_{k^2-1}:\Delta_{k^2-1}\to\mathbb{R}$, 则对于任意的 $\varepsilon>0$ 存在 $\gamma>0$ 使得只要 $|p_i-q_i|<\delta$ 对任意的 $i=1,2,...,k^2$ 都成立, 则 $|H(p)-H(q)|<\varepsilon$, 其中 $p=\{p_i\}_{i=1}^{k^2}, q=\{q_i\}_{i=1}^{k^2}$. 于是取 $\delta=\frac{1}{2}\gamma$, 则, 若对于任意的 $i=1,2,...,k^2$, $m(A_i\Delta B_i)<\frac{1}{2}\gamma$, 就有

$$H(\alpha|\beta)=H(\alpha\vee\beta)-H(\beta)=H(\alpha\vee\beta)-H(\beta\vee\beta)<\varepsilon.$$

上述不等式中, 我们用到了 $\alpha\vee\beta=\{A_i\cap B_j\}$, $\beta\vee\beta=\{B_i\cap B_j\}$ 以及

$$m((A_i\cap B_j)\Delta(B_i\cap B_j))\le m(A_i\Delta B_i)<\frac{1}{2}\gamma.$$

这里我们取 $p=\{m(A_i\cap B_j)\}, q=\{B_i\cap B_j\}$. □

下面我们用此引理证明这一定理. 对于最开始的 $\varepsilon>0$ 利用上述引理, 得到对应的 $\delta>0$. 由于 $\vee_{n=0}^{\infty}T^{-n}\eta=\mathcal{B}$ 所以存在 $N\in\mathbb{Z}_+$, 使得 $\{B_i\}_{i=1}^{k}$(不一定是分划) $\subset\vee_{n=0}^{N-1}T^{-n}\eta$ 并且 $m(A_i\Delta B_i)<\gamma$ ($\gamma$ 待定) 满足对

$$C_1=B_1,\ C_2=B_2\backslash B_1, ..., C_{k-1}=B_{k-1}\backslash(B_1\cup B_2\cup...\cup B_{k-2}),\ C_k=C\backslash(C_1\cup...\cup C_{k-1})$$

有 $m(C_i\Delta A_i)<\delta\ i=1,2,...,k$. 这里只需要 $\gamma$ 充分小总是成立的. 此时 $\xi=\{C_i\}_{i=1}^k$ 是 $\vee_{n=0}^{\infty}T^{-n}\eta$ 中的元素构成的分划. 于是

$$\begin{aligned}h_m(\alpha,T)&\le h_m(\xi,T)+H(\alpha|\xi)\le h_m(\xi,T)+\varepsilon\\&\le h_m\left(\bigvee_{n=0}^{N-1}T^{-n}\eta,T\right)+\varepsilon=h_m(\eta,T)+\varepsilon.\end{aligned}$$

根据 $\varepsilon>0$ 的任意性就完成了结论的证明. □

由此可得, 若 $\eta\in\mathcal{P}_r$, $h_m(T)=h_m(\eta,T)\le H(\eta)\le\log r$ 则 $r\ge e^{h_m(T)}$. 这样就给出了生成子的分划中元素个数的一个估计. 更精确的, 我们给出下面的定理并不加以证明. 当然在这之前, 我们需要对可逆的保测变换的生成子加以定义.

**定义 2.26.** 对于可逆的保测系统 $(X,\mathcal{B},m,T)$, 若 $\eta$ 为有限可测的分划, 满足 $\vee_{n=-\infty}^{\infty}T^n\eta\overset{\circ}{=}\mathcal{B}$, 则称 $\eta$ 是 $T$ 的生成子. 这里 $\vee_{n=-\infty}^{\infty}T^n\eta$ 是包含所有的 $T^i\eta\ i\in\mathbb{Z}$ 的最小 $\sigma$-代数

**定理 2.27** (Krieger)**.** *若 $(X,\mathcal{B},m,T)$ 可逆并且遍历的, 则存在生成子 $\{D_1,D_2,...,D_r\}$ 满足 $e^{h_m(T)}\le r\le e^{h_m(T)}+1$.*

**注 2.28.** 这里注意到可逆的保测变换的生成子的定义与不可逆的时候有所不同, 这是因为如果保测系统 $(X,\mathcal{B},m,T)$ 是可逆的, 并且分划 $\eta=\{D_i\}_{i=1}^r$ 满足 $\vee_{n=0}^{\infty}T^{-n}\eta\overset{\circ}{=}\mathcal{B}$ 则有 $h_m(T)=0$.

最后我们不加证明的给出一个定理, 此定理表明测度熵的值实际上是分划对 $\mathcal{B}$ 的一个穷竭的过程.

**定理 2.29.** *设 $(X,\mathcal{B},m,T)$ 是保测系统, $\alpha_n\in\mathcal{P}$ 并且 $\alpha_n\uparrow\mathcal{B}$, 那么 $h_m(T)=\lim_{n\to\infty}h_m(\alpha_n,T)$. 其中 $\alpha_n\uparrow\mathcal{P}$ 的含义是 $\vee_{k=0}^{\infty}T^{-k}\alpha_n$ 单调增加收敛到 $\mathcal{B}$.*

### 2.3.2 熵计算的例子

**命题 2.30.** *设 $(X,\mathcal{B},m,T)$ 是单边的 $p$-Bernoulli 系统, 其中 $p=(p_1,p_2,...,p_d)$. 那么 $h_m(T)=-\sum_{i=1}^d p_i\log p_i$.*

*证明.* 记 $[i]_0=A_i$ 表示只规定 0 处为 $i$ 的基本矩形, 令 $\alpha=\{A_1,A_2,...,A_d\}$. 我们断言 $\alpha$ 是 $T$ 的生成子. 这是因为

$$\bigvee_{i=0}^{n-1}T^{-i}\alpha=\{[i_0i_1...i_{n-1}]\}$$

恰好是规定了前 $n$ 个位置的基本矩形的全体, 这些基本矩形正好是是生成 $\mathcal{B}$ 的半代数. 于是断言成立, 那么根据 Kolmogorov-Sinai 定理, 可得

$$\begin{aligned}h_m(T)=h_m(\alpha,T)&=\lim_{n\to\infty}\frac{1}{n}H\left(\bigvee_{i=0}^{n-1}T^{-i}\alpha\right)=-\sum_{\underline{i}\in\{1,2,...,d\}^n}m(A_i)\log m(A_i)\\&=-\sum_{i_0i_1...i_{n-1}}p_{i_0}p_{i_1}...p_{i_{n-1}}\log(p_{i_0}p_{i_1}...p_{i_{n-1}})=-\left(\sum_{i=1}^d p_i\log p_i\right).\end{aligned}$$

这就完成了证明. □

**注 2.31.** 可见对于 Bernoulli 系统, $(\frac{1}{n},...,\frac{1}{n})$ 对应的熵是 $\log n$. 再根据熵是同构不变量, 则单位圆周熵的 $\times 2$ 与 $\times 3$ 映射是不同够的, 这是因为根据例 1.86, 单位圆周上的 $\times d$ 映射可以和 Bernoulli 系统同构. 但是在一组基上定义, 可证明这两者是谱同构的, 这说明, 同构的概念是严格强于谱同构的.

**命题 2.32.** *设 $(X,\mathcal{B},m,T)$ 是关于 $(p,P)$ 的 Markov 链，则其熵为*

$$h_m(T)=-\sum_{i,j}p_ip_j\log p_{ij}.$$

*证明.* 类似 Bernoulli 系统的证明，$\alpha=\{A_1,A_2,...,A_d\}$ 为 $T$ 的生成子. 于是有

$$\begin{aligned}H\left(\bigvee_{i=0}^{n-1}T^{-1}\alpha\right)&=-\sum_{i_0i_1...i_{n-1}}m([i_0i_1...i_{n-1}])\log(m([i_0i_1...i_{n-1}]))\\&=-\sum_{i_0i_1...i_{n-1}}p_{i_0}p_{i_0i_1}...p_{i_{n-2}i_{n-1}}(\log p_{i_0}+\log p_{i_0i_1}+...+\log p_{i_{n-2}i_{n-1}})\\&=-\sum_{i_0}p_{i_0}\log p_{i_0}-\sum_{i_0i_1}p_{i_0}p_{i_0i_1}\log p_{i_0i_1}-...-\sum_{i_{n-2}i_{n-1}}p_{i_{n-2}}p_{i_{n-2}i_{n-1}}\log p_{i_{n-2}i_{n-1}}.\end{aligned}$$

两侧同除以 $n$ 取极限就得到了结论成立. □

# 3 Shannon-McMillan-Breiman 定理

## 3.1 简介与引入

**定理 3.1** (Shannon-McMillan-Breiman 定理)**.** *设 $(X,\mathcal{B},m,T)$ 为遍历系统，$\alpha\in\mathcal{P}$ 是有限可测分划，则*

$$-\lim_{n\to\infty}\frac{\log(m(\alpha^n(x)))}{n}=h_m(\alpha,T),$$

*其中 $\alpha^n=\alpha\vee T^{-1}\alpha\vee...\vee T^{-n+1}\alpha$，$\alpha^n(x)$ 表示分划 $\alpha^n$ 中含有 $x$ 的那些元素的全体. 特别地，如果 $\alpha$ 是生成子，那么*

$$-\lim_{n\to\infty}\frac{\log(m(\alpha^n(x)))}{n}=h_m(T).$$

上述定理说明，对于遍历系统，对于 $n$ 充分大的时候，有 $m(\alpha^n(x))\sim e^{-nh}$，其中 $h=h_m(\alpha,T)$.

在具体考虑这个定理的证明之前，我们先考虑一个例子，即 Bernoulli 系统. 类似命题 2.30 中的设定，我们取 $\alpha=\{A_1,A_2,...,A_d\}$ 恰好是生成子. 此时对于某一个 $x$，可见

$$\alpha^n(x)=[x_0x_1...x_{n-1}]_0,$$

上面 $[x_0x_1...x_{n-1}]_0$ 即表示为从第 0 个坐标开始，之后的 $n$ 个位置上面正好是 $x_0,x_1,...,x_{n-1}$. 于是可得

$$m(\alpha^n(x))=p_{x_0}p_{x_1}...p_{x_{n-1}}.$$

所以

$$-\frac{1}{n}\log(m(\alpha^n(x)))=-\frac{1}{n}\sum_{i=0}^{n-1}\log p_{x_i}.$$

我们令 $f(x)=\log p_{x_0}$ 并且注意到 $T$ 是左推移映射，所以 $\log p_{x_i}=f\circ T^i(x)$. 于是，根据遍历性以及命题 2.30 的关于 $T$ 的熵性质，则验证了 Shannon-McMillan-Breiman 定理的正确性.

## 3.2 条件期望，条件测度和条件熵

**定义 3.2** (测度的绝对连续性)**.** 设 $m,\mu$ 为可测空间 $(X,\mathcal{B})$ 上的两个概率测度 (不是概率测度也可以推广，简便起见不加赘述)，称 $\mu$ 是关于 $m$ 绝对连续的，若对于任意的 $A\in\mathcal{B}$，若 $m(A)=0$，则有 $\mu(A)=0$，记作 $\mu\ll m$.

于是我们不加证明地给出著名的 Radon-Nikodym 定理，这一定理刻画了绝对连续的性质.

**定理 3.3** (Radon-Nikodym 定理)**.** *设 $m, \mu$ 为可测空间 $(X, \mathcal{B})$ 上的两个概率测度，则 $\mu \ll m$ 当且仅当存在 $f \in L^1(X, \mathcal{B}, m)$ 并且满足*

$$f \geq 0, \ \int_X f \mathrm{d}m = 1,$$

*使得对于任意的 $A \in \mathcal{B}$ 有 $\mu(A) = \int_A f \mathrm{d}m$. 另外此时的 $f$ 在 $m$-a.e. 的意义下是唯一的，并且我们称 $\mu$ 关于 $m$ 的 Radon-Nikodym 导数是 $f$，简记为 $\frac{\mathrm{d}\mu}{\mathrm{d}m} = f$.*

对于上述定理，一侧的证明是简单，即若存在 $f$，满足 $f \geq 0$ 以及 $\int_X f \mathrm{d}m = 1$ 则根据 $\mu(A) = \int_X f \mathrm{d}m$ 定义出来的测度 $\mu$ 是关于 $m$ 绝对连续的. 另一方面的证明就比较复杂了，需要用到 Riesz 表示定理.

**例 3.4** ($[0,1]$ 区间上的 Lebesgue 测度)**.** 设 $m$ 是 $[0,1]$ 上的 Lebesgue 测度, $\mu$ 是上面的另一个概率测度，于是我们可以令 $F(x) = \mu([0,x])$. 则 $\mu \ll m$ 当且仅当 $\mu(A) = \int_A f \mathrm{d}m$，其中 $0 \leq f \in L^1(X, \mathcal{B})$ 并且 $\int_X f \mathrm{d}m = 1$. 所以 $F(x) = \int_0^x f(t) \mathrm{d}t$. 那么根据实分析中的经典结论，可得 $F(x)$ 是绝对连续的函数并且 $F'(x) = f(x)$ 在 Lebesgue 测度意义下几乎处处成立.

上述例子中，我们可见实分析中绝对连续的函数的定义与我们这里的绝对连续是相容的. 事实上，将 $\mu \ll m$ 的定义改写，我们可见 $\mu \ll m$ 当且仅当

$$\forall \varepsilon > 0, \ \exists \delta > 0, \ \text{s.t.} \ \forall A \in \mathcal{B}, \ \text{if} \ m(A) < \delta, \ \text{then} \ \mu(A) < \varepsilon.$$

**例 3.5** (拉回测度的导数)**.** 设 $m$ 为 $[0,1]$ 上的 Lebesgue 测度，$f : [0,1] \to [0,1]$ 是保向的微分同胚，令 $\mu = f_* m$ 是测度 $m$ 的拉回映射，即 $\mu(A) = m(f^{-1}(A)) = \int_{f^{-1}(A)} 1 \mathrm{d}x = \int_A (f^{-1})'(y) \mathrm{d}y$. 于是有 $\frac{\mathrm{d}\mu}{\mathrm{d}m} = (f^{-1})'$.

在定义条件期望之前，我们先给出一个基本的定理，这一定理是跳进期望定义的基础.

**定理 3.6.** *设 $(X, \mathcal{B}, m)$ 是概率空间，$\mathcal{C} \subset \mathcal{B}$ 是 $\mathcal{B}$ 的 $\sigma$-子代数，那么对于任意的 $f \in L^1(X, \mathcal{B}, m)$，存在唯一的 $g \in X \to \mathbb{R}$ 使得*

1. *$g \in L^1(X, \mathcal{C}, m)$;*
2. *对于任意的 $C \in \mathcal{C}$ 有 $\int_C f \mathrm{d}m = \int_C g \mathrm{d}m$.*

**注 3.7.** 上面的唯一性来自于下面的观察, 对于概率空间 $(X, \mathcal{B}, m)$, 以及可积函数 $f, h \in L^1(X, \mathcal{B}, m)$, $f = h$ $m$-a.e. 当且仅当 $\int_B f \mathrm{d}m = \int_B h \mathrm{d}m$ 对于任意的 $B \in \mathcal{B}$ 都成立.

*证明.* 我们在 $(X, \mathcal{C})$ 上定义两个测度，$m = m|_{\mathcal{C}}$ 以及

$$\mu(C) = \frac{\int_C f \mathrm{d}m}{\int_X f \mathrm{d}m}, \ C \in \mathcal{C}.$$

于是很容易验证 $\mu(X) = 1$ 并且 $\mu \ll m$ 在 $\mathcal{C}$ 上成立. 因此根据 Radon-Nikodym 定理, 可得, 存在唯一的 $\widetilde{g} \in L^1(X, \mathcal{C}, m)$ 使得 $\mu(C) = \int_C \widetilde{g} \mathrm{d}m$ 成立 (这里因为 $f$ 可能并不一定是正函数, 所以需要分为正负部两部分来考虑, 为简便起见我们这里假设 $f \geq 0$ 且几乎处处为 0). 那么我们令 $g = (\int_X f \mathrm{d}m) \widetilde{g}$ 即可. □

根据上面的定理，我们可以给出条件期望的定义.

**定义 3.8** (条件期望)**.** 设 $(X,\mathcal{B},m)$ 为概率空间，$f\in L^1(X,\mathcal{B},m)$，$\mathcal{C}\subset\mathcal{B}$ 是 $\mathcal{B}$ 的 $\sigma$-子代数，根据定理 3.6，确定的唯一的 $g\in L^1(X,\mathcal{C},m)$ 被称为 $f$ 在 $\mathcal{C}$ 下的条件期望，记作 $g=E(f|\mathcal{C})$. 根据定理 3.6，$E(f|\mathcal{C})$ 由下列两条性质唯一确定.

1. $E(f|\mathcal{C})\in L^1(X,\mathcal{C},m)$.
2. 对于任意的 $C\in\mathcal{C}$，有 $\int_C f\mathrm{d}m=\int_C E(f|\mathcal{C})\mathrm{d}m$.

**例 3.9** (有限代数上的条件期望)**.** 设 $(X,\mathcal{B},m)$ 是概率空间，$\alpha\in\mathcal{P}$ 是有限可测分划，记 $\alpha=\{A_1,A_2,...,A_k\}$. 令 $\mathcal{A}(\alpha)$ 为 $\alpha$ 生成的代数 (因为 $\alpha$ 是有限的所以生成的实际上是 $\sigma$-代数). 于是，对于任意的 $f\in L^1(X,\mathcal{B},m)$ 我们可知

$$E(f|\mathcal{A}(\alpha))=\sum_{i=1}^k a_i\chi_{A_i}(x).$$

其中 $a_i=\frac{1}{m(A_i)}\int_{A_i}f\mathrm{d}m$. 这是因为对于任意的 $t\in\mathbb{R}$，我们要求 $\{E(f|\mathcal{A}(\alpha))<t\}$ 是 $\mathcal{C}$ 中的集合，于是，只能写成上述简单函数的形式，而前面的系数 $\{a_i\}_{i=1}^k$ 可以根据积分的性质得到. 特别地，如果 $\mathcal{C}=\{\emptyset,X\}$，则知 $E(f|\mathcal{C})=\int f\mathrm{d}m$，这和我们概率论里面的期望定义是吻合的.

下面我们来考察这里的条件期望和我们之前顶柜的条件概率有什么样的关系. 在上述例子中，取 $f=\chi_B$ 于是

$$E(\chi_B|\mathcal{A}(\alpha))(x)=\sum_{i=1}^k\frac{m(A_i\cap B)}{m(A_i)}\chi_{A_i}(x)=\sum_{i=1}^k m(B|A_i)\chi_{A_i}(x).$$

于是可知 $E(\chi_B|\alpha)=m(B|\alpha(x))$，这里 $\alpha(x)$ 就是我们之前提到的包含 $x$ 的 $\alpha$ 中的元素. 我们简记为 $E(B|\alpha)=m(B|\alpha(x))$. 这和我们之前定义的条件测度非常接近. 下面，我们给出条件测度的推广.

**定义 3.10.** 设 $(X,\mathcal{B},m)$ 为概率空间，$\mathcal{C}\subset\mathcal{B}$ 是 $\mathcal{B}$ 的 $\sigma$-子代数，我们定义 $m(\cdot|\mathcal{C})(\cdot):(\mathcal{B},X)\to\mathbb{R}$ 是 $\sigma$-子代数 $\mathcal{C}$ 下的条件测度 (给定一个 $x\in X$ 就是一个测度)，满足 $m(B|\mathcal{C})(x)=E(\chi_B|\mathcal{C})(x):=E(B|\mathcal{C})(x)$.

下面我们介绍一些关于条件期望和条件概率的基本性质，总结为定理如下.

**定理 3.11.** *设 $(X,\mathcal{B},m)$ 为概率空间，$\mathcal{C}\subset\mathcal{B}$ 为 $\sigma$-子代数，则*

1. *$E(\cdot|\mathcal{C}):L^1(X,\mathcal{B},m)\to L^1(X,\mathcal{B},m)$ 是线性的.*
2. *若 $f\geq 0$ 并且 $f\in L^1(X,\mathcal{B},m)$ 则 $E(f|\mathcal{C})\geq 0$.*
3. *对于任意的 $f\in L^1(X,\mathcal{B},m)$ 以及 $g\in L^\infty(X,\mathcal{B},m)$，有 $E(fg|\mathcal{C})=gE(f|\mathcal{C})$.*
4. *设 $\varphi:\mathbb{R}\to\mathbb{R}$ 连续凸，$f,\varphi\circ f\in L^1(X,\mathcal{B},m)$ 则*
$$\varphi\circ E(f|\mathcal{C})\leq E(\varphi\circ f|\mathcal{C}).$$
5. *对于任意的 $f\in L^1(X,\mathcal{B},m)$，有 $|E(f|\mathcal{C})|\leq E(|f||\mathcal{C})$.*
6. *设 $\mathcal{C}_1\subset\mathcal{C}_2\subset\mathcal{B}$ 则 $E(E(f|\mathcal{C}_2)|\mathcal{C}_1)=E(f|\mathcal{C}_1)$.*
7. *对于任意的 $f\in L^2(X,\mathcal{B},m)$，则 $E(f|\mathcal{C})\in L^2(X,\mathcal{B},m)$ 是 $f$ 在 $L^2(X,\mathcal{C},m)$ 上的正交投影 (事实上根据这一条，以及 $L^2(X,\mathcal{B},m)$ 在 $L^1(X,\mathcal{B},m)$ 中的稠密性，也可以定义条件期望).*

因为此定理中的结论都是比较容易证明的，在此，我们略去证明. 仅给出一些必要的解释. 首先 1，2 是显然的，对于 3，可以先考虑 $f$ 是特征函数的情况，进而考虑其是简单函数的情况，最后用单调收敛定理即可证明. 4 是 Jessen 不等式的直接推论，5 可以在 4 中取 $\varphi = |\cdot|$ 即绝对值函数. 6，7 直接根绝定义验证即可.

从 Von-Neumann 遍历定理和 Birkhoff 定理的证明以及上述第 7 条性质中，我们可以得到下述定理.

**定理 3.12** (Birkhoff 定理)**.** 设 $(X, \mathcal{B}, m, T)$ 为保测系统，$f \in L^1(X, \mathcal{B}, m)$ 那么对于 $n \to \infty$，

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \to E(f|\mathcal{S}),\ m\text{-}a.e.$$

并且 $L^1(X, \mathcal{B}, m)$ 意义下也收敛. 这里 $\mathcal{S}$ 表示所有的 $\mathcal{B}$ 中满足 $T^{-1}A = A$ 的集合构成的 $\sigma$-子代数.

证明. 根据 Birkhoff 遍历定理，可得 $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i \to f^*$ 在几乎处处意义下以及 $L^1$ 意义下收敛. 很显然 $f^* \circ T = f^*$. 这表明 $f^* \in L^1(X, \mathcal{S}, m)$.[2] 要证明 $f^* = E(f|\mathcal{S})$，就需要证明对于任意的 $A \in \mathcal{S}$ 有 $\int_A f^* \mathrm{d}m = \int_A f \mathrm{d}m$. 这是因为

$$\int_A f \mathrm{d}m = \int_X f\chi_A \mathrm{d}m = \int_X (f\chi_A)^* \mathrm{d}m = \int_X f^*\chi_A \mathrm{d}m.$$

其中第二个等式处，我们对函数 $f\chi_A$ 用了一次 Birkhoff 遍历定理，对于第三个等式，我们用了

$$\begin{aligned}\chi_A(x) f^*(x) &= \lim_{n\to\infty} \chi_A(x)\left(\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i(x)\right) = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=0}^{n-1}\chi_A(x) f \circ T^i(x)\right)\\ &= \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=0}^{n-1}\chi_A \circ T^i(x) f \circ T^i(x)\right) = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{i=0}^{n-1}(\chi_A f) \circ T^i(x)\right).\end{aligned}$$

以及 $T^{-1}A = A$. 主义上面的极限是几乎处处意义与 $L^1$ 意义下的. □

## 3.3 Shannon-McMillan-Breiman 定理的证明

在正式证明之前，我们回顾信息函数的定义，并加以推广.

**定义 3.13.** 设 $(X, \mathcal{B}, m)$ 是概率空间，$\alpha = \{A_1, A_2, ..., A_k\} \in \mathcal{P}$ 是有限可测分划. 定义函数

$$I(\alpha)(x) = -\sum_{i=1}^{k}\chi_{A_i}(x)\log(m(A_i))$$

是 $\alpha$ 的信息函数，并且 $\int_X I(\alpha)\mathrm{d}m = H(\alpha)$. 设 $\mathcal{C}$ 是 $\sigma$-子代数，定义条件信息函数 $I(\alpha|\mathcal{C})$ 为

$$I(\alpha|\mathcal{C}) = -\sum_{i=1}^{k}\chi_{A_i}(x)\log(m(A_i|\mathcal{C})(x)) = -\sum_{i=1}^{k}\chi_{A_i}(x)\log(E(A_i|\mathcal{C})(x)).$$

**习题 3.14.** 设 $(X, \mathcal{B}, m)$ 是概率空间，并且 $\alpha, \beta \in \mathcal{P}$ 是有限可测的分划，那么

$$H(\alpha|\beta) = \int I(\alpha|\mathcal{A}(\beta))\mathrm{d}m.$$

根据上面这个习题的结论，我们可以做下面的定义.

[2]这里实际上我们并没有区分 $\mathcal{S}$ 与 $\widetilde{\mathcal{S}} = \{A \in \mathcal{B} : m(T^{-1}A\Delta A) = 0\}$. 可见 $E(f|\mathcal{S}) = E(f|\widetilde{\mathcal{S}})$ 以及 $L^1(X, \mathcal{S}, m) = L^1(X, \widetilde{\mathcal{S}}, m)$.

**定义 3.15.** 设 $(X,\mathcal{B},m)$ 是概率空间，并且 $\alpha\in\mathcal{P}$ 是有限可测的分划，令 $\mathcal{C}\subset\mathcal{B}$ 是 $\sigma$-子代数，我们定义条件熵 $H(\alpha|\mathcal{C})$ 如下

$$H(\alpha|\mathcal{C})=\int I(\alpha|\mathcal{C})\mathrm{d}m.$$

上面的定义中,我们需要验证右侧的可积性质,这是容易得到的,事实上假设 $\alpha=\{A_1,A_2,...,A_k\}$,利用定理 3.11 的第 4 条 (取 $\varphi(x)=x\log x$) 并且注意到函数的凹性，得到

$$\begin{aligned}\int I(\alpha|\mathcal{C})\mathrm{d}m&=\int E(I(\alpha|\mathcal{C})|\mathcal{C})\mathrm{d}m=-\int E(\chi_{A_i}|\mathcal{C})\log E(\chi_{A_i}|\mathcal{C})\mathrm{d}m\\&\le-\sum_{i=1}^{k}\varphi\left(\int E(\chi_{A_i}|\mathcal{C})\mathrm{d}m\right)=-\sum_{i=1}^{k}m(A_i)\log(m(A_i))=H(\alpha).\end{aligned}$$

设说明我们的定义是良定义的.

关于条件信息函数，我们有下面的结论.

**引理 3.16.** 设 $(X,\mathcal{B},m)$ 是概率空间，并且 $\alpha,\beta\in\mathcal{P}$ 是有限可测的分划，令 $\mathcal{C}\subset\mathcal{B}$ 是 $\sigma$-子代数. 那么

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

回忆我们之前在定理 2.8 中得到的式子，上述引理中将之前关于熵函数的恒等式，推广到了信息函数上去，并且对于某一个分划推广到了 $\sigma$-子代数.

*证明.* 根据定义可得

$$\begin{aligned}I(\alpha|\mathcal{C})&=-\sum_{A\in\alpha}\chi_A\log E(\chi_A|\mathcal{C})=-\sum_{A\in\alpha,B\in\beta}\chi_A\chi_B\log E(\chi_A|\mathcal{C}),\\I(\alpha\vee\beta|\mathcal{C})&=-\sum_{A\in\alpha,B\in\beta}\chi_{A\cap B}\log E(\chi_{A\cap B}|\mathcal{C})\\I(\beta|\alpha\vee\mathcal{C})&=-\sum_{B\in\beta}\chi_B\log E(\chi_B|\alpha\vee\mathcal{C}).\end{aligned}$$

于是可得

$$I(\alpha\vee\beta|\mathcal{C})-I(\alpha|\mathcal{C})=\sum_{B\in\beta}\chi_B\sum_{A\in\alpha}\chi_A(\log E(\chi_A|\mathcal{C})-\log E(\chi_{A\cap B}|\mathcal{C})).$$

于是要证明结论，往证

$$\log E(\chi_B|\alpha\vee\mathcal{C})=\sum_{A}\chi_A\log\left(\frac{E(\chi_{A\cap B}|\mathcal{C})}{E(\chi_A|\mathcal{C})}\right)$$

在几乎处处的意义下成立. 在每一个 $A\in\alpha$ 上考虑，归结为证明

$$E(\chi_B|\alpha\vee\mathcal{C})=\sum_{A}\frac{\chi_A E(\chi_{A\cap B}|\mathcal{C})}{E(\chi_A|\mathcal{C})}.$$

首先要证明两侧的相等,因为两侧的函数都是关于 $\alpha\vee\mathcal{C}$ 是可测的于是只要证明在上面的任意集合的积分相等即可. 特别地,由于我们注意到 $\alpha\vee\mathcal{C}$ 中的元素可以表示为 $\cup_\alpha(A\cap C_A)$,其中 $A\in\alpha,C_A\in\mathcal{C}$,所以只需验证一个 $A\cap C$ 上面的积分相等即可. 于是上式左侧有

$$\mathrm{LHS}=\int_{A\cap C}E(\chi_B|\alpha\vee\mathcal{C})\mathrm{d}m=\int_{A\cap C}\chi_B\mathrm{d}m=m(A\cap B\cap C).$$

另一方面，右侧为

$$
\begin{aligned}
\mathrm{RHS} &= \int_{A\cap C} \chi_A \frac{\chi_A E(\chi_{A\cap B}|\mathcal{C})}{E(\chi_A|\mathcal{C})} \mathrm{d}m = \int_C \frac{\chi_A E(\chi_{A\cap B}|\mathcal{C})}{E(\chi_A|\mathcal{C})} \mathrm{d}m \\
&= \int_C E\left(\frac{\chi_A E(\chi_{A\cap B}|\mathcal{C})}{E(\chi_A|\mathcal{C})}\right) \mathrm{d}m = \int_C E(\chi_A|\mathcal{C}) \frac{E(\chi_{A\cap B}|\mathcal{C})}{E(\chi_A|\mathcal{C})} \mathrm{d}m = m(A\cap B\cap C).
\end{aligned}
$$

于是结论就证明成立了. 其中第三个等式我们用到了条件期望的性质. □

下面我们不加证明地叙述著名地鞅定理.

**定理 3.17** (鞅定理). 设 $(X, \mathcal{B}, m)$ 是概率空间，$\{\mathcal{C}_n\}$ 是单调增加的 $\sigma$-子代数，并且即 $\infty = \vee_{n\geq 1}\mathcal{C}_n$. 那么任给 $f \in L^1(X, \mathcal{B}, m)$ 可得

$$
E(f|\mathcal{C}_n) \xrightarrow[a.e.]{L^1} E(f|\mathcal{C}_\infty).
$$

下面我们介绍钟开莱不等式.

**定理 3.18** (钟开莱不等式). 设 $(X, \mathcal{B}, m, T)$ 是保测习题，$\alpha \in \mathcal{P}$ 是有限可测分划，则

$$
\sup_{n\geq} I(\alpha|\alpha_1^n) \in L^1(X, \mathcal{B}, m).
$$

**注 3.19.** 首先这里我们为了简便起见，我们将 $\alpha_i^j$ 与 $\alpha^n$ 的定义改成下面所示 (和之前的定义略有不同但没有本质区别)：

$$
\alpha_i^j = T^{-i}\alpha \vee ... \vee T^{-j}\alpha,\ \alpha^n = \alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha.
$$

特别地，我们记 $\alpha_1^\infty = \vee_{n\geq 1}\alpha_1^n$.

**注 3.20.** 事实上在钟开莱不等式中我们实际上证明更强的结论，$\sup_{n\geq} I(\alpha|\alpha_1^n) \in L^1(X, \alpha_1^\infty, m)$ 并且 $\int \sup_{n\geq} I(\alpha|\alpha_1^n)\mathrm{d}m \leq 1 + H(\alpha)$. 这也就是这个定理为什么叫做不等式的原因.

**推论 3.21.** 当 $n \to \infty$ 时，有

$$
I(\alpha|\alpha_1^n) = -\sum_{A\in\alpha} \chi_A \log E(\chi_A|\alpha_1^n) \xrightarrow[a.e.]{L^1} I(\alpha|\alpha_1^\infty) = -\sum_{A\in\alpha} \log E(\chi_A|\alpha_1^\infty).
$$

证明. 首先根据 $A$ 是有限可测分划以及鞅定理，可得上述式子中的 a.e. 收敛，再根据钟开莱不等式以及控制收敛定理，可得此收敛在 $L^1$ 意义下也成立. □

**注 3.22.** 在上述推论中，若两侧一积分，根据信息函数和关于分划的熵的关系，可得 $h_m(\alpha, T) = \lim_{n\to\infty} H(\alpha|\alpha_1^n) = H(\alpha|\alpha_1^\infty)$. 这正是定理 2.17 中的性质 10 的结论.

下面我们给出钟开莱不等式的证明.

证明. 利用积分和分布函数的关系，即 $\int f\mathrm{d}m = \int_0^\infty m(\{x : f(x) > t\})\mathrm{d}t$，可得

$$
\int f\mathrm{d}m = \sum_{A\in\alpha} \int \chi_A f\mathrm{d}m = \sum_{A\in\alpha} \int_0^\infty m(\{x : \chi_A f(x) > t\})\mathrm{d}t.
$$

对于任意的 $t > 0$，我们记

$$
\{x : \chi_A f(x) > t\} = A \cap \{x : f(x) > t\} = A \cap B(t) = \coprod_{n\geq 1} A \cap B_n(t),
$$

其中

$$
B_n(t) = \{x : E(\chi_A|\alpha_1^1)(x) \geq e^{-t}, ..., E(\chi_A|\alpha_1^{n-1})(x) \geq e^{-t},\ E(\chi_A|\alpha_1^n) < e^{-t}\}.
$$

于是，对于 $A\cap B(t)$ 的测度，我们等式

$$m(A\cap B(t))=\sum_{n\geq 1}m(A\cap B_n(t))=\sum_{n\geq 1}\int\chi_A\chi_{B_n(t)}\mathrm{d}m$$
$$=\sum_{n\geq 1}\int_{B_n(t)}\chi_A\mathrm{d}m=\sum_{n\geq 1}\int_{B_n(t)}E(\chi_A|\alpha_1^n)\mathrm{d}m.$$

上面最后一个等式，我们用了 $B_n(t)\in\alpha_1^n$ 以及条件期望的性质. 那么，根据 $B_n(t)$ 的定义，可以得到下面的估计.

$$m(A\cap B(t))\leq\sum_{n\geq 1}e^{-t}m(B_n(t))\leq e^{-t}.$$

最后一个不等式来自概率空间的假设. 所以，我们有

$$\int f\mathrm{d}m=\sum_{A\in\alpha}\int_0^\infty m(A\cap B(t))\mathrm{d}t\leq\sum_{A\in\alpha}\int_0^\infty\min\{e^{-t},m(A)\}\mathrm{d}t$$
$$\leq\sum_{A\in\alpha}(m(A)-m(A)\log m(A))=1+H(\alpha).$$

这就证明了钟开莱不等式. □

**注 3.23.** 在上面钟开莱不等式的证明中，我们也可以推广到可数的分划 $\alpha$，只要 $H(\alpha)$ 是有限的，那么钟开莱不等式一样成立.

在正式证明 Shannon-McMillan-Breiman 定理之前，我们还需要最后一个工具，即 Breiman 定理.

**定理 3.24** (Breiman)**.** *设 $(X,\mathcal{B},m,T)$ 是保测系统，$f_n,f\in L^1(X,\mathcal{B},m)$ 满足*

1. $f_n\to f$ *a.e.*
2. $\sup_{n\geq 1}|f_n|\in L^1(X,\mathcal{B},m)$.

*则 $\frac{1}{n}\sum_{i=0}^{n-1}f_{n-i}\circ T^i\xrightarrow[a.e.]{L^1}E(f|\mathcal{S})$ 成立. 其中 $\mathcal{S}=\{A\in\mathcal{B}:T^{-1}A=A\}$.*

**注 3.25.** 上述 Breiman 定理可以看作是 Birkhoff 遍历定理的一个推广的形式.

*证明.* 根据 Birkhoff 遍历定理，可得 $\frac{1}{n}\sum_{i=0}^{n-1}f\circ T^i\xrightarrow[a.e.]{L^1}E(f|\mathcal{S})$，于是我们仅需要证明

$$\frac{1}{n}\sum_{i=0}^{n-1}|f_{n-i}-f|\circ T^i\xrightarrow[a.e.]{L^1}0,\ n\to\infty$$

即可. 记 $F=\sup_{n\geq 1}|f_n|$ 以及 $F_k=\sup_{i>k}|f_i-f|$. 根据假设，可得 $0\leq F_k\leq 2F\in L^1(X,\mathcal{B},m)$. 根据 $f_i\to f$ a.e. 成立，可知 $F_k\xrightarrow[a.e.]{L^1}0$ 当 $k\to\infty$ 成立. 这里的 $L^1$ 收敛，我们利用了控制收敛定理. 所以，利用 Cauchy 极限的方法 (算数平均的极限和原本的极限相同)，很容易得到

$$\left\|\frac{1}{n}\sum_{i=0}^{n-1}|f_{n-i}-f|\circ T^i\right\|_{L^1(X,\mathcal{B},m)}=\left\|\frac{1}{n}\sum_{i=0}^{n-1}|f_{n-i}-f|\right\|_{L^1(X,\mathcal{B},m)}\to 0.$$

接下来我们来考察 a.e. 收敛. 给定 $k>1$，我们注意到

$$\frac{1}{n}\sum_{i=0}^{n-1}|f_{n-i}-f|\circ T^i=\frac{1}{n}\left(\sum_{i=0}^{n-k-1}|f_{n-i}-f|\circ T^i\right)+\frac{1}{n}\sum_{i=n-k}^{n-1}|f_{n-i}-f|\circ T^i$$
$$\leq\frac{1}{n}\left(\sum_{i=0}^{n-k-1}F\circ T^i+\sum_{i=1}^{k}|f_i-f|\circ T^{n-i}\right).$$

对于任意的 $1 \le i \le k$，我们有 $\frac{1}{n}|f_i - f| \circ T^{n-i} \to 0$ a.e. 成立. 所以，两侧同时取上极限，在 a.e. 的意义下，有

$$\limsup_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} |f_{n-i} - f| \circ T^i \le F_k^* = E(F_k|\mathcal{S}).$$

再由于 $\{F_k\}$ 是单调减少的函数序列，于是 $\{F_k^*\}$ 同样是单调减少的序列. 课设 $\lim_{k\to\infty} F_k^* = \varphi$. 所以要完成定理的证明，仅需要验证 $\int \varphi \mathrm{d}m = 0$ 即可. 这是因为

$$\int \varphi \mathrm{d}m = \lim_{k\to\infty} \int F_k^* \mathrm{d}m = \lim_{k\to\infty} \int F_k \mathrm{d}m = 0.$$

上面的第一个等式我们用了积分的单调收敛定理，第二个等式用了 Birkhoff 遍历性定理. □

下面我们来整理 Shannon-McMillan-Breiman 定理. 我们首先给出此定理的非遍历情形，然后再对遍历情形作简单的说明.

**定理 3.26** (非遍历情形下的 Shannon-McMillan-Breiman 定理)**.** 设 $(X, \mathcal{B}, m, T)$ 是保测系统，$\alpha \in \mathcal{P}$. 令 $f = I(\alpha|\alpha_1^\infty)$，则

$$-\frac{\log(m(\alpha^n(x)))}{n} \xrightarrow[a.e.]{L^1} E(f|\mathcal{S})(x),\ n \to \infty,$$

其中 $\mathcal{S} = \{A \in \mathcal{B} : T^{-1}A = A\}$.

证明. 可知 $I(\alpha^n) = -logm(\alpha^n(x)) = -\sum_{A\in\alpha} \chi_A \log m(A)$. 于是又

$$\begin{aligned} I(\alpha^{n+1}) &= I(\alpha \vee T^{-1}\alpha \vee ... \vee T^{-n}\alpha) = I(\alpha \vee T^{-1}\alpha_0^{n-1}\alpha) = I(T^{-1}\alpha_0^{n-1}) + I(\alpha|T^{-1}\alpha_0^{n-1}) \\ &= I(\alpha_0^{n-1}) \circ T + I(\alpha|\alpha_1^n) = ... = I(\alpha) \circ T^n + I(\alpha|T^{-1}\alpha) \circ T^{n-1} + ... + I(\alpha|\alpha_1^n). \end{aligned}$$

于是我们令 $f_i = I(\alpha|\alpha_1^i)$ 则有

$$I(\alpha^{n+1}) = \sum_{i=0}^{n-1} f_{n-i} \circ T^i + I(\alpha) \circ T^n.$$

两侧同时除以 $n+1$，利用钟开莱不等式，可知 $\{f_i\}$ 满足 Breiman 定理的条件，再利用 Breiman 定理，即可证明上述定理. □

**注 3.27.** 如果在上述定理中，我们假设遍历性的条件，那么

$$-\frac{\log(m(\alpha^n(x)))}{n} \xrightarrow[a.e.]{L^1} h_m(\alpha, T),\ n \to \infty.$$

这是因为

$$-\frac{\log(m(\alpha^n(x)))}{n} = \frac{I(\alpha^n)}{n} \to \int E(f|\mathcal{S})\mathrm{d}m = \int f\mathrm{d}m = \int I(\alpha|\alpha_1^\infty)\mathrm{d}m = H(\alpha|\alpha_1^\infty) = h_m(\alpha, T).$$

# 4 拓扑熵

## 4.1 拓扑熵的开覆盖定义

设 $X$ 为紧致的 Hausdorff 空间 (不一定有度量)，$T : X \to X$ 是连续的映射，则称 $(X, T)$ 为拓扑动力系统. 设 $\alpha \subset 2^X$ 是 $X$ 集合的子族，满足

1. 对于任意的 $A \in \alpha$，$A$ 为开集；
2. $\cup_{A\in\alpha} A = X$.

则称 $\alpha$ 为 $A$ 的开覆盖. 设 $\alpha \subset 2^X$ 是 $X$ 的一个开覆盖，记

$$N(\alpha) = \min\{\#\beta : \beta\text{为}\alpha\text{的有限子覆盖}\}.$$

这里的子覆盖的意思是指对于任意的 $B \in \beta$，有 $B \in \alpha$，并且 $\cup_{B\in\beta} B = X$. 换言之 $\beta$ 为可以覆盖 $X$ 的 $\alpha$ 的子族.

设 $\alpha, \beta$ 都是 $X$ 的开覆盖，我们定义 $\alpha, \beta$ 的交如下

$$\alpha \vee \beta = \{A \cap B : A \in \alpha, B \in \beta\}.$$

可见这里的定义和分划的交类似. 很显然 $\alpha \vee \beta$ 也是 $X$ 的开覆盖. 称 $\alpha \leq \beta$ ($\beta$ 细于 $\alpha$)，若对于任意的 $B \in \beta$ 存在 $A \in \alpha$ 满足 $B \subset A$.

**例 4.1.** 若 $\beta$ 为 $\alpha$ 的有限子覆盖，则 $\alpha \leq \beta$.

很显然，对于拓扑动力系统 $(X, T)$，$\alpha$ 是开覆盖，那么 $T^{-1}\alpha$ 也是开覆盖. 这是因为

$$\bigcup_{A\in\alpha} T^{-1}A = T^{-1}\left(\bigcup_{A\in\alpha} A\right) = T^{-1}(X) = X.$$

定义 $H(\alpha) = \log N(\alpha)$. 那么我们有下列性质.

**注 4.2.** 因为我们假设 $X$ 是紧致的拓扑空间，所以任何开覆盖都有有限的子覆盖，于是 $H(\cdot)$ 是良定义的.

**定理 4.3** ($H(\cdot)$ 的性质)**.** *设 $(X, T)$ 是拓扑动力系统，$\alpha, \beta$ 是两个开覆盖.*

1. *$H(\alpha) \geq 0$ 且 $H(\alpha) = 0$ 当且仅当 $X \in \alpha$.*
2. *若 $\alpha \leq \beta$ 则 $H(\alpha) \leq H(\beta)$.*
3. *$H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$.*
4. *$H(T^{-1}\alpha) \leq H(\alpha)$ 并且当 $T$ 是满射的时候等号成立.*

证明. 性质 1 中不等式的证明是显然的，棍鱼等式，我们可知 $H(\alpha) = 0$ 当且仅当 $N(\alpha) = 1$，此条件成立的充要条件是 $X \in \alpha$. 对于性质 2, 我们取 $\beta$ 的子覆盖 $\gamma$ 使得 $\#\gamma = N(\beta)$，即 $\gamma = \{B_i\}_{i=1}^{N(\beta)}$. 根据 $\alpha \leq \beta$ 可知存在 $A_i \in \alpha$ $(i = 1, 2, ..., N(\beta))$ 使得 $B_i \subset A_i$. 这意味着 $\cup_{i=1}^{N(\beta)} A_i = X$. 从而可知 $N(\alpha) \leq N(\beta)$. 于是 $H(\alpha) \leq H(\beta)$. 在性质 3 中，设 $\{A_i\}_{i=1}^{N(\alpha)}$ 与 $\{B_j\}_{j=1}^{N(\beta)}$ 是对应的极小开覆盖，那么 $\{A_i \cap B_j\}_{i,j}$ 也是 $\alpha\beta\beta$ 的一个子覆盖，很显然 $\#\{A_i \cap B_j\}_{i,j} \leq N(\alpha)N(\beta)$. 两侧同时取对数，则得到 $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$. 最后我们完成性质 4 的证明. 设 $\{T^{-1}A_i\}_{i=1}^{N(T^{-1}\alpha)}$ 是 $T^{-1}\alpha$ 的有限子覆盖，那么 $\cup_{i=1}^{N(T^{-1}\alpha)} T^{-1}A_i = X$. 于是可得

$$T\left(\bigcup_{i=1}^{N(T^{-1}\alpha)} T^{-1}A_i\right) = T(X) = X = \bigcup_{i=1}^{N(T^{-1}\alpha)} T(T^{-1}\alpha) \subset \bigcup_{i=1}^{N(T^{-1}\alpha)} A_i.$$

根据上面的式子，可得 $N(\alpha) \geq N(T^{-1}\alpha)$. 可见上述关系中集合的包含关系是等号当 $T$ 是满射的时候成立，这也表明 $N(T^{-1}\alpha) = N(\alpha)$. 于是可得 $H(\cdot)$ 的关系. □

**引理 4.4.** *设 $(X, T)$ 为拓扑动力系统，$\alpha$ 为 $X$ 的开覆盖，则对于任意的 $n \in \mathbb{Z}_+$ 与 $\alpha^n = \alpha \vee T^{-1}\alpha \vee ... \vee T^{-n+1}\alpha$，有 $a_{n+m} \leq a_n + a_m$. 特别地，$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\geq 1} \frac{a_n}{n}$ 成立.*

证明. 很显然，我们有

$$\begin{aligned}a_{n+m} &= H(\alpha^{n+m}) = H(\alpha \vee ... \vee T^{-n+1}\alpha \vee T^{n}\alpha \vee ... \vee T^{-n-m+1}\alpha)\\ &\leq H(\alpha^{n} \vee T^{-n}\alpha^{m}) \leq H(\alpha^{n}) + H(T^{-1}\alpha^{m}) \leq H(\alpha^{n}) + H(\alpha^{m}) = a_{n} + a_{m}.\end{aligned}$$

这就完成了证明. □

**定义 4.5.** 设 $(X,T)$ 是拓扑动力系统，$\alpha$ 是开覆盖，令 $h(\alpha,T) = \lim_{n\to\infty} \frac{1}{n} H(\alpha^{n})$ 为关于开覆盖 $\alpha$ 的熵，$h(T) = \sup_{\alpha} h(\alpha,T)$ 为 $T$ 的拓扑熵.

**注 4.6.** 很显然 $h(T) = \sup_{\alpha} h(\alpha,T) = \sup_{\alpha:\#\alpha<\infty} h(\alpha,T)$. 这是因为我们假设 $x$ 是紧的，所以任意的开覆盖都有有限子覆盖，而取子覆盖后 $H(\cdot)$ 的取值反而变大，这样就表明我们求关于 $T$ 的熵只要在有限的开覆盖上面去上确界即可.

**例 4.7.** 若 $T = \mathrm{Id}$ 是恒等映射，那么 $T^{-1}\alpha = \alpha$，并且 $\alpha^{n} = \alpha$. 于是 $h(\alpha,T) = 0$ 对任意的开覆盖 $\alpha$ 都成立，于是 $h(T) = 0$.

**定理 4.8.** 设 $(X,T)$ 是拓扑动力系统，并且 $T$ 是可逆的，则 $h(T) = h(T^{-1})$.

证明. 由于 $X$ 是紧 Hausdorff 的，于是 $T^{-1}$ 也是连续的，所以

$$\begin{aligned}h(\alpha,T) &= \lim_{n\to\infty} \frac{1}{n} H(\alpha^{n}) = \lim_{n\to\infty} \frac{1}{n} H(T^{n-1}(\alpha \vee ... \vee T^{-n+1}))\\ &= \lim_{n\to\infty} \frac{1}{n} H(\alpha \vee ... \vee T^{n-1}\alpha) = h(\alpha, T^{-1}).\end{aligned}$$

这样就证明了结论. □

**定义 4.9.** 设 $(X_i,T_i)$ $i=1,2$ 为两个拓扑动力系统，若有连续的满射 $\varphi : X_1 \to X_2$ 使得 $\varphi \circ T_1 = T_2 \circ \varphi$ 则称 $\varphi$ 为 $T_1$ 为 $T_2$ 的扩充，$T_2$ 为 $T_1$ 的一个因子，此时 $\varphi$ 称为 $T_1$ 与 $T_2$ 之间的半共轭 ($T_1$ 与 $T_2$ 的地位不等价). 当 $\varphi$ 为同胚的时候，则称为共轭.

**定理 4.10.** 设 $(X_i,T_i)$ $i=1,2$ 为两个拓扑动力系统，$(X_1,T_1)$ 为 $(X_2,T_2)$ 的一个扩充，则 $h(T_1) \geq h(T_2)$. 特别地，如果 $\varphi : X_1 \to X_2$ 是共轭，那么 $h(T_1) = h(T_2)$.

证明. 设 $\varphi$ 是 $T_1$ 与 $T_2$ 之间的半共轭，任给 $\alpha$ 为 $X_2$ 的开覆盖，我们断言 $h(\alpha,T_2) = h(\varphi^{-1}\alpha, T_1)$. 根据 $\varphi \circ T_1 = T_2 \circ \varphi$，可得 $\varphi^{-1} \circ T_2^{-n} = T_1^{-n} \circ \varphi^{-1}$，有

$$\begin{aligned}(\varphi^{-1}\alpha)_{T_1}^{n} &= \varphi^{-1}\alpha \vee T_1^{-1}\varphi^{-1}\alpha \vee ... \vee T_1^{-n+1}\varphi^{-1}\alpha\\ &= \varphi^{-1}\alpha \vee \varphi^{-1}T_2^{-1}\alpha \vee ... \vee \varphi^{-1}T_2^{-n+1}\alpha\\ &= \varphi^{-1}(\alpha \vee T_2^{-1}\alpha \vee ... \vee T_2^{-n+1}\alpha) = \varphi^{-1}(\alpha_{T_2}^{n}).\end{aligned}$$

由此可得 $N((\varphi^{-1}\alpha)_{T_1}^{n}) = N(\varphi^{-1}(\alpha_{T_2}^{n})) = N(\alpha_{T_2}^{n})$. 这里的第二个等式来自于 $\varphi$ 是满射 (利用类似定理 4.3 最后一条性质的证明). 这样我们就证明了 $h(\alpha,T_2) = h(\varphi^{-1}\alpha,T_1)$. 那么就有

$$h(T_2) = \sup_{\alpha} h(\alpha,T_2) = \sup_{\alpha} h(\varphi^{-1}\alpha,T_1) \leq \sup_{\beta} h(\beta,T_1) = h(T_1).$$

这样我们就完成了证明. □

**注 4.11.** 上述定理证明拓扑熵在共轭下是不变量.

## 4.2 拓扑熵的等价定义

### 4.2.1 两种方法定义的拓扑熵

之后，不加以强调，我们都假设 $(X,d)$ 是紧致的度量空间，$T:X\to X$ 是连续映射. 此时，我们和之前一样，用 $(X,T)$ 表示为拓扑动力系统. 对于任意的 $n\geq 1$，我们定义

$$d_n(x,y)=\max\{d(T^ix,T^iy),0\leq i\leq n-1\}.$$

从上面的定义可见 $d_1(x,y)=d(x,y)$，并且

$$d_1(x,y)\leq d_2(x,y)\leq ...\leq d_n(x,y).$$

对于任意的 $\varepsilon>0$，记 $B_n(x,\varepsilon)=\{y\in X:d_n(x,y)<\varepsilon\}$. 我们称 $B_n(x,\varepsilon)$ 为 $(n,\varepsilon)$ Bowen 球. 根据定义我们可以得到

$$\begin{aligned}B_n(x,\varepsilon)&=\{y\in X:d(x,y)<\varepsilon,\ d(Tx,Ty)<\varepsilon,...,d(T^{n-1}x,T^{n-1}y)<\varepsilon\}\\&=\bigcap_{i=0}^{n-1}\{y\in X:d(T^ix,T^iy)<\varepsilon\}=\bigcap_{i=0}^{n-1}T^{-i}\{z\in X:d(z,T^ix)<\varepsilon\}=\bigcap_{i=0}^{n-1}T^{-i}(B(T^ix,\varepsilon)).\end{aligned}$$

**定义 4.12** (生成集)**.** 设 $(X,T)$ 是拓扑动力系统，设 $n\geq 1$，$\varepsilon>0$. $F\subset X$ 被称为 $T$ 的一个 $(n,\varepsilon)$ 生成集，若 $\cup_{x\in F}\overline{B}_n(x,\varepsilon)=X$. 其中 $\overline{B}_n(x,\varepsilon)=\{y\in X:d_n(x,y)\leq\varepsilon\}$，即那些使得 $d(x,y)\leq\varepsilon$, $d(Tx,Ty)\leq\varepsilon,...,d(T^{n-1}x,T^{n-1}y)\leq\varepsilon$ 的 $X$ 中的点 $y$.

**注 4.13.** 一个值得注意的事实是在一般的度量空间中 (不一定是欧式度量),$\overline{B}_n(x,\varepsilon)$ 不一定是 $B_n(x,\varepsilon)$ 的闭包. 至于反例，可以直接考虑离散度量.

**定义 4.14.** 记 $r_n(\varepsilon)=\min\{\#F:F$ 为$(n,\varepsilon)$ 生成集$\}$ (注意因为 $X$ 是紧的，$T$ 是连续的，所以 $r_n(\varepsilon)$ 是良定义的). 很显然，$r_n(\varepsilon)$ 关于 $n\geq 1$ 单调增加，关于 $\varepsilon\in(0,1)$ 单调减少. 于是我们定义

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{\log r_n(\varepsilon)}{n}=h_r(T)$$

是由生成集定义的拓扑熵.

**定义 4.15** (分离集)**.** 设 $(X,T)$ 是拓扑动力系统 $E\subset X$ 称为 $(n,\varepsilon)$ 为分离集,若对于任意的 $x,y\in E$，满足 $x\neq y$ 则有 $d_n(x,y)>\varepsilon$. 即存在 $i\in\{0,1,2,...,n-1\}$ 使得 $d(T^ix,T^iy)>\varepsilon$. 很显然根据 $X$ 是紧的，$T$ 是连续的，所以分离集一定是有限. 于是 $s_n(\varepsilon)=\max\{\#E:E$ 为$(n,\varepsilon)$ 分离集$\}<\infty$ 是良定义的. 于是我们称

$$\lim_{\varepsilon\to 0}\limsup_{n\to\infty}\frac{\log s_n(\varepsilon)}{n}=h_s(T)$$

为由分离集定义的拓扑熵. 在上面的定义中，我们同样用到了 $s_n(\varepsilon)$ 关于 $n$ 单调增加，关于 $\varepsilon$ 单调减少的性质.

### 4.2.2 两种拓扑熵定义的等价性

在这一部分，我们希望证明由生成集与分离集定义的两种拓扑熵是等价的.

**引理 4.16.** *对于任意的* $n\geq 1$ *以及* $\varepsilon>0$，*有* $r_n(\varepsilon)\leq s_n(\varepsilon)\leq r_n\left(\frac{\varepsilon}{2}\right)$.

*证明.* 设 $E$ 是 $(n,\varepsilon)$ 分离集，并且满足 $s_n(\varepsilon)=\#E$. 我们断言 $E$ 为 $(n,\varepsilon)$ 生成集. 这是因为，对于任意的 $x\in X$，根据 $\#E$ 的最大性，可知，若 $x\notin E$ 则 $x\cup E$ 不是 $(n,\varepsilon)$ 分离集. 从而存在 $y\in E$ 使得 $d_n(x,y)\leq\varepsilon$. 这样就证明了断言，并且根据此断言，我们可得 $r_n(\varepsilon)\leq s_n(\varepsilon)$.

另一方面，任取 $F$ 为 $(n,\frac{\varepsilon}{2})$ 生成集，则对于任意的 $x\in E$，存在 $y\in F$ 使得 $d_n(x,y)\le\frac{\varepsilon}{2}$. 这样就给出了一个映射 $\varphi:E\to F$. 我们断言，这样的映射是一个单射. 如果断言成立，那么结论就成立了. 设 $x_1,x_2\in E$，满足 $\varphi(x_1)=\varphi(x_2)=y$ 则 $d_n(x_1,y)\le\frac{1}{2}\varepsilon$ 并且 $d_n(x_2,y)\le\frac{1}{2}\varepsilon$. 于是由三角不等式可得 $d(x_1,x_2)\le\varepsilon$. 由于 $E$ 是分离集，则 $x_1=x_2$ 这就证明了 $\varphi$ 是单射. □

下面我们简单回顾 Lebesgue 数的定理.

**定理 4.17** (Lebesgue 数)**.** *设 $(X,d)$ 为紧致的度量空间，$\alpha$ 是 $X$ 的一个开覆盖，则存在 $\delta>0$ 使得对于任意的子集 $B$，只要满足 $\operatorname{diam}(B)\le\delta$ 就存在 $A\in\alpha$ 使得 $B\subset A$.*

上述定理中这样的 $\delta$ 就被称为是覆盖 $\alpha$ 的 Lebesgue 数. 接下来的命题给出了覆盖数和 Lebesgue 数的关系.

**命题 4.18.** *设 $(X,T)$ 是拓扑动力系统以及 $\alpha$ 是一个开覆盖. $\delta$ 是覆盖 $\alpha$ 的一个 Lebesgue 数. 则对于任意的 $n\ge1$，有 $N(\alpha_0^{n-1})\le r_n(\frac{\delta}{2})$，其中 $\alpha_0^{n-1}=\alpha\vee...\vee T^{-n+1}\alpha$.*

**注 4.19.** $n=1$ 时，结论是显然的，这是因为，设 $F$ 为 $(1,\frac{\delta}{2})$ 生成集，则 $\cup_{x\in F}\overline{B}(x,\frac{\delta}{2})=X$. 可知 $\operatorname{diam}(\overline{B}(x,\frac{\delta}{2}))\le\delta$. 于是对于任意的 $x\in F$，利用 $\delta$ 是 Lebesgue 数的条件，可得，存在 $A_x\in\alpha$ 使得 $\overline{B}(x,\frac{\delta}{2})\subset A_x$ 于是 $N(\alpha)\le r_n(\frac{\delta}{2})$. 最后这个不等式利用了 $F$ 是 $\alpha$ 的一个加细.

*命题 4.18 的证明.* 设 $F$ 为 $(n,\frac{\delta}{2})$ 的生成集具有最小基数，即 $\#F=r_n(\frac{\delta}{2})$ 于是 $\cup_{x\in F}\overline{B}_n(x,\frac{\delta}{2})=X$. 将 $\overline{B}_n(x,\frac{\delta}{2})$ 按照定义写开，有

$$\begin{aligned}\overline{B}_n(x,\frac{\delta}{2})&=\left\{y\in X:d(x,y)\le\frac{\delta}{2},\ d(Tx,Ty)\le\frac{\delta}{2},...,d(T^{n-1}x,T^{n-1}y)\le\frac{\delta}{2}\right\}\\&=\overline{B}\left(x,\frac{\delta}{2}\right)\cap T^{-1}\left(\overline{B}\left(x,\frac{\delta}{2}\right)\right)\cap...\cap T^{-n+1}\left(\overline{B}\left(x,\frac{\delta}{2}\right)\right).\end{aligned}$$

因为 $\alpha_0^{n-1}$ 中的元素 (代表元) 可以写作

$$A_0\cap T^{-1}A_1\cap...\cap T^{-n+1}A_{n-1}$$

的形式，其中 $A_i\in\alpha$. 由于 $\delta$ 是 Lebesgue 数，对于球 $\overline{B}(x,\frac{\delta}{2})$，我们可以取出 $A_0,A_1,...,A_{n-1}$ 使得

$$\overline{B}\left(x,\frac{\delta}{2}\right)\subset A_0,\ \overline{B}\left(Tx,\frac{\delta}{2}\right)\subset T^{-1}A_1,...,\overline{B}(T^{n-1}x,\frac{\delta}{2})\subset A_{n-1}.$$

于是

$$\overline{B}\left(x,\frac{\delta}{2}\right)\subset A_0,\ T^{-1}\left(\overline{B}\left(x,\frac{\delta}{2}\right)\right)\subset T^{-1}A_1,...,T^{-n+1}\left(\overline{B}(x,\frac{\delta}{2})\right)\subset T^{-n+1}A_{n-1}.$$

由此可见 $\overline{B}\left(x,\frac{\delta}{2}\right)\subset A_0\cap T^{-1}A_1\cap...\cap T^{-n+1}A_{n-1}$. 从而可见 $F$ 是 $a_0^{n-1}$ 的一个加细，那么就有结论成立. □

更进一步我们可以得到下面的估计.

**命题 4.20.** *设 $\varepsilon>0$，$\alpha$ 是 $X$ 的开覆盖，并且 $\operatorname{diam}\alpha<\varepsilon$ (表示此开覆盖中的每个元素的直径都小于 $\varepsilon$). 则 $r_n(\varepsilon)\le s_n(\varepsilon)\le N(\alpha_0^{n-1})$.*

*证明.* 事实上第一个不等号来自引理 4.16，于是我们只需要证明第二个不等式就可以了. 取 $(n,\varepsilon)$ 分离集 $E$，并且满足 $\#E=s_n(\varepsilon)$. 往证 $\alpha_0^{n-1}$ 中的任意元素至多只含有 $E$ 中的一个元素，如果此结论证明成立，再根据覆盖的性质，第二个不等式很容易就证明成立. 若不然，有不同的两点 $x,y\in E\cap(A_0\cap T^{-1}A_1\cap T^{-2}A_2\cap...\cap T^{-n+1}A_{n-1})$. 由于 $E$ 是分离集，并且 $\operatorname{diam}\alpha<\varepsilon$ 这就导致 $x=y$，矛盾，于是这样我们就证明了结论. □

**引理 4.21.** 设 $\{\alpha_{(n)}\}_{n\geq 1}$ 是一列开覆盖，并且满足 $\operatorname{diam}\alpha_{(n)}\to 0$. 则 $h(T)=\lim_{n\to\infty}h(\alpha_{(n)},T)$.

*证明.* 根据拓扑熵的定义，很显然 $h(T)=\sup_\alpha h(\alpha,T)\geq\limsup_{n\to\infty}h(\alpha_{(n)},T)$. 往证

$$\liminf_{n\to\infty}h(\alpha_{(n)},T)\geq h(T)$$

成立. 只要说明对于任意的开覆盖，若 $n$ 充分大，就有 $h(\alpha_{(n)},T)\geq h(\alpha,T)$ 即可. 首先根据定理 4.17，可以取 $\alpha$ 的 Lebesgue 数 $\delta$. 因为 $\operatorname{diam}\alpha_{(n)}\to 0$ 所以当 $n$ 充分大的时候 $\operatorname{diam}\alpha_{(n)}<\delta$. 此时很显然 $\alpha_{(n)}\geq\alpha$ 成立. 从而就有 $N(\alpha_{(n)})\geq N(\alpha)$，两侧同取交，简单计算即得到结论成立. □

下面我们给出最重要的定理证明，即用生成集和分离集定义的拓扑熵和原先的拓扑熵都是相等的.

**定理 4.22.** 设 $(X,T)$ 是有度量的拓扑动力系统，则 $h(T)=h_r(T)=h_s(T)$.

*证明.* 设 $\alpha_\varepsilon$ 是一个直径 $<\varepsilon$ 的一个开覆盖，设 $0<\delta(\varepsilon)<\varepsilon$ 为 $\alpha_\varepsilon$ 的一个 Lebesgue 数 (存在性由定理 4.17 保证)，于是根据命题 4.18 和 4.20 可得

$$N((\alpha_\varepsilon)_0^{n-1})\leq r_n\left(\frac{\delta}{2}\right)\leq s_n\left(\frac{\delta}{2}\right)\leq N\left((\alpha_{\frac{\delta}{2}})_0^{n-1}\right).$$

令 $\varepsilon\to 0$，很显然可得 $\delta(\varepsilon)\to 0$. 再根据命题 4.21 的结论，得

$$h(T)\leq h_r(T)\leq h_s(T)\leq h(T).$$

这样就完成了证明. □

## 4.3 拓扑熵的计算

### 4.3.1 一些计算拓扑熵的理论

**定义 4.23.** 设 $T:(X,d)\to(X,d)$ 是拓扑动力系统，若有常数 $e>0$ 使得对于任意的 $x\neq y$ 都存在 $n\geq 0$ 使得 $d(T^nx,T^ny)\geq e$ 则称 $T$ 是正向可扩的，换言之不断迭代以后总有一次距离会大于 $e$.

若存在 $L>1$ 以及 $\delta_0>0$ 使得对于任意的 $x,y$ 不等，并且 $d(x,y)<\delta$ 有 $d(Tx,Ty)\geq Ld(x,y)$ 则称 $T$ 是扩张.

满足上面可扩定义中的常数 $e$ 被我们称为 $T$ 的一个可扩常数.

**注 4.24.** 扩张性质可以推出可扩性质，这是因为可以将 $e$ 取成 $\frac{1}{2}\delta_0$ 就可以了.

下面我们给出一个定理，用以刻画可扩的性质.

**定理 4.25.** *设 $T:(X,d)\to(X,d)$ 为拓扑动力系统，则下面几条等价.*

1. *$T$ 为可扩的.*
2. *存在 $e>0$ 使得，只要 $d(T^nx,T^ny)\leq e$ 对于任意的 $n$ 成立，就有 $x=y$.*
3. *$T$ 有 (有限) 生成子，即存在开覆盖 $\alpha=\{A_1,A_2,...,A_d\}$ 使得对于任意的 $i\in\{1,2,...,d\}^{\mathbb{N}}$，有 $\#\cap_{n\geq 0}T^{-n}\overline{A}_{i_n}\leq 1$.*
4. *$T$ 有弱生成子，即存在开覆盖 $\beta=\{B_1,...,B_\ell\}$ 使得对于任意的 $i\in\{1,2,...,d\}^{\mathbb{N}}$，有 $\#\cap_{n\geq 0}T^{-n}B_{i_n}\leq 1$.*

*证明.* 很显然第一条和第二条是等价的，并且第三条可以推出第四条，于是我们仅需要证明第四条可以推出第三条以及第三条和第一条等价就可.

*4⇒3 的证明*. 设 $\beta=\{B_1,B_2,...,B_\ell\}$ 是一个弱生成子，我们取 $\delta>0$ 为 $\beta$ 的一个 Lebesgue 数. 那么任取一个有限开覆盖 $\alpha=\{A_1,...,A_d\}$，满足 $\operatorname{diam}\alpha\le\delta$ 由 $\operatorname{diam}A_i=\operatorname{diam}\overline{\mathbf{A}}_i\le\delta$ 可知，存在 $\sigma(i)\in\{1,2,...,\ell\}^{\mathbb{N}}$ 使得 $\overline{A}_i\subset B_{\sigma(i)}$. 于是我们就有

$$\bigcap_{n\ge0}T^{-n}\overline{A}_{i_n}\subset\bigcap_{n\ge0}T^{-n}B_{\sigma(i_n)}.$$

这样就证明了 $\alpha$ 是有限的生成子. □

*1⇒3 的证明*. 任取 $X$ 的直径小于 $e$ 的有限开覆盖 $\alpha=\{A_i\}_{i=1}^d$. 若 $x,y\in\cap_{n\ge0}T^{-n}\overline{A}_{i_n}$ 则根据第三条性质，右侧中的元素至多有一个，这就表明 $x=y$. 于是第一条就成立了. □

*3⇒1 的证明*. 考虑有限开覆盖 $\alpha=\{A_i\}_{i=1}^d$，并且这个有限开覆盖由 Lebesgue 数 $\delta>0$. 我们断言 $\delta$ 为一个可扩常数，这是因为对于任意的 $x,y\in X$，若 $d(T^nx,T^y)\le\delta$ 对于任意的 $n$ 都成立，那么利用 Lebesgue 数的性质，存在 $i_n\in\{1,2,...,d\}$ 使得 $\{T^nx,T^y\}\subset A_{i_n}$. 于是就有 $x,y\in\cap_{n\ge0}T^{-n}A_{i_n}$. 再根据第三条性质，右侧这个集合中最多包含一个元素，就有 $x=y$. □

由此，我们就证明了上述四条性质是等价的. □

接着，根据上面的性质，我们可以得到下面的定理.

**定理 4.26.** *设 $T:(X,d)\to(X,d)$ 是可扩的，$\alpha$ 是一个生成子，那么 $h(\alpha,T)=h(T)$.*

*证明*. 根据定理 4.21，只要证明 $\operatorname{diam}\alpha_0^{n-1}\to0$ 即可. 反设结论不成立，则存在 $\varepsilon_0$ 以及在抽子列意义下（简便起见不改变记号）$\operatorname{diam}\alpha_0^{n-1}\ge\varepsilon_0$. 于是存在 $A_i\in\alpha$ $i=1,2,...,n-1$ 使得 $\operatorname{diam}(\cap_{i=0}^{n-1}T^{-i}A_i)\ge\varepsilon_0$. 于是就有 $x_n,y_n\in\cap_{i=0}^{n-1}T^{-i}A_i$ 使得 $d(x_n,y_n)\ge\varepsilon_0$，必要时候抽子列，可知存在 $x_n\to x$ 以及 $y_n\to y$ 满足 $d(x,y)\ge\varepsilon_0$ 以及 $T^nx,T^ny\in A_{i_n}$，这表明 $x,y\in\cap_{n\ge0}T^{-n}\overline{A}_{i_n}$. 根据生成子的性质，结论成立. □

### 4.3.2 一些具体计算的例子

**例 4.27.** 设 $\Sigma_d=\{0,1,2,...,d-1\}^{\mathbb{N}}$ 其中 $d\ge2$. 并且在 $\{0,1,2,...,d-1\}$ 上面赋予离散拓扑，即每一个单点都是开集，那么可以在 $\Sigma_d$ 上面诱导一个乘积拓扑，在此拓扑下，可以定义左移位 $\sigma:\Sigma_d\to\Sigma_d$ $(\sigma(x_0x_1...)=x_1x_2...)$. 很显然，左移位是连续的. 当然在 $\Sigma_d$ 上，我们可以定义与乘积拓扑相容的度量 $d$，满足 $d(x,y)=\sum_{n=0}^{\infty}\frac{\delta(x_n,y_n)}{2^n}$ 其中 $\delta(x_n,y_n)$ 仅在 $x_n=y_n$ 时取 1 其他情况取作 0. 在乘积拓扑下，我们可知 $x$ 与 $y$ 接近，即存在充分大的 $N$ 使得 $x[0,N]=y[0,N]$. 根据 Tychonoff 定理，这个空间是紧的. 于是 $\sigma$ 左移位的拓扑熵是 $h(\sigma)=\log d$.

*证明*. 令 $A_k=[k]_0=\{x\in\Sigma_d:x_0=k\}$，$k=0,1,2,...,d-1$ 是 $\Sigma_d$ 上既开又闭得子集 (这是因为 $A_0\cup A_1\cup...\cup A_{d-1}=\Sigma_d$). 于是可见 $\alpha=\{A_k\}_{k=0}^{d-1}$ 既是开覆盖，又是一个分划. 可知 $\frac{1}{2}$ 是可扩常数，此时我们断言 $\alpha$ 是一个 $\sigma$ 的生成子. 这是因为对于任意的 $i\in\Sigma_d$,

$$\#\bigcap_{n\ge0}\sigma^{-n}\overline{A}_{i_n}=\#\bigcap_{n\ge0}\sigma^{-n}A_{i_n}=\#\{i\}=1.$$

所以我们有

$$h(\alpha,\sigma)=h(\sigma)=\lim_{n\to\infty}\frac{\log N(\alpha_0^{n-1})}{n},$$

其中

$$\alpha_0^{n-1}=\alpha\vee\sigma^{-1}\alpha\vee...\vee\sigma^{-n+1}\alpha=\{A_{\underline{j}}:\underline{j}\in\{0,1,2,...,d-1\}^n\}.$$

(表示把前面 $n$ 个元素定住) 于是可知 $N(\alpha_0^{n-1})=d^n$. 这是因为其中的每个元素都是互不相交的，所以一个都不能少. 据此，很容易得到 $h(\sigma)=\log d$. □

**注 4.28.** 这里，如果 $d=2$，我们可知 $\Sigma_2$ 与 Cantor 集是同胚的，对应我们可以定义映射 $x\mapsto \sum_{n=0}^{\infty}\frac{2x_n}{3^{n+1}}$. 即做 3 进制展开，不出现 1.

根据上面的例子，我们考虑 $T:X\to X$ 是可扩系统，并且 $\{A_i\}_{i=0}^{d-1}$ 是一个 $T$ 的开覆盖生成子，即 $\#\cap_{n\geq 0}T^{-n}A_{i_n}\leq 1$ 对于任意的 $i\in\Sigma_d$ 都成立. 这样我们可以定义一个集合

$$Y=\left\{i\in\Sigma_d:\bigcap_{n\geq 0}T^{-n}A_{i_n}\neq\emptyset\right\}.$$

由此，我们可以定义一个映射，将任意的 $i\in Y$，映为 $\cap_{n\geq 0}T^{-n}A_{i_n}$. 这样相当于做了一个编码.

**例 4.29.** 设 $X\subset\Sigma_d$ 是一个闭子集，则将 $\sigma X\subset X$. $\sigma|_X:X\to X$ 称为符号系统. 记

$$\theta_n(X)=\#\{\underline{i}\in\{1,2,...,d-1\}^n:\exists x\in X,\ \text{s.t.}\ x[0,n-1]=\underline{i}\}.$$

我们有 $h(\sigma|_X)=\lim_{n\to\infty}\frac{\log\theta_n(X)}{n}$. 这里的 $\lim_{n\to\infty}\frac{\log a_n}{n}$ 可以看作是 $a_n$ 的指数增长率.

*证明.* 令 $\alpha=\{A_k\}_{k=1}^{d-1}$ 如上例所示，则 $\alpha_X=\{A_k\cap X\}_{k=1}^{d-1}$ 为 $\sigma|_X$ 的生成子，于是我们有

$$(\alpha_X)_0^{n-1}=\alpha_X\vee(\sigma|_X)^{-1}\alpha_X\vee...\vee(\sigma|_X)^{-n+1}\alpha_X.$$

所以有 $N((\alpha_X)_0^{n-1})=(\alpha_X)_0^{n-1}$ 中非空的元素个数. 再根据观察

$$(A_{i_0}\cap X)\cap((\sigma|_X)^{-1}(A_{i_1}\cap X))\cap...\cap((\sigma|_X)^{-n+1}(A_{i_{n-1}}\cap X))$$

非空当且仅当存在 $x\in X$ 使得 $x[0,n-1]=i_0i_1...i_{n-1}$. 于是 $N((\alpha_X)_0^{n-1})=\theta_n(X)$. 据此，我们证明了结论. □

**例 4.30.** 设 $A_{d\times d}$ 为 $0-1$ 矩阵，即任意的 $a_{ij}\in\{0,1\}$. 定义

$$\Sigma_A=\{x\in\Sigma_d:a_{x_ix_{i+1}}=1,\ \forall i\geq 0\},\ x_i\in\{0,1,2,...,d-1\}$$

为拓扑 Markov 链. 一个特殊的例子，我们取 $A$ 为 $2\times 2$ 矩阵并且满足 $a_{22}=0$ 其他都是 1. 那么 $\Sigma_A$ 表示的就是 $\Sigma_2$ 中不出现"11" 的那些元素. 容易看到，任意一个矩阵 $A$，就构成了一个邮箱图 $G$ (类似 Markov 链中的表述). 另一方面 $\Sigma_A$ 为 $\Sigma_d$ 中的左平移不变的闭子集，我们有 $h(\sigma|_{\Sigma_A})=\log r(A)$. 其中 $r(A)$ 表示 $A$ 的谱半径，即最大的特征值的模长.

*证明.* 设 $\underline{i}\in\{0,1,2,...,d-1\}^n$ 在 $\Sigma_A$ 中出现，使得 $x[0,n-1]=\underline{i}$. 那么我们有 $a_{i_0i_1}=...=a_{i_{n-2}i_{n-1}}=1$ 也就是 $a_{i_0i_1}a_{i_1i_2}...a_{i_{n-2}i_{n-1}}=1$. (如果加上条件，对于任意的 $i$，存在 $j$ 使得 $a_{ij}=1$ 那么反之也对). 于是可得

$$\theta_n(\Sigma_A)=\sum_{\underline{i}\in\{0,1,2,...,d-1\}}a_{i_0i_1}a_{i_1i_2}...a_{i_{n-2}i_{n-1}}=\sum_{i,j}(A^{n-1})_{ij}.$$

我们记 $\|B\|=\sum_{i,j}b_{ij}$ 是矩阵的范数，于是根据上一个例子中的结论

$$h(\sigma|_{\Sigma_A})=\lim_{n\to\infty}\frac{\log\|A^{n-1}\|}{n}=\lim_{n\to\infty}\log(\|A^n\|^{\frac{1}{n}})=\log(r(A)).$$

上面我们还用到了在有限维空间中所有的范数都是差了常数，这个在取极限的过程中自然消失了. □

**注 4.31.** 特别地，我们不加证明地叙述 Perron-Frobenius 定理：设 $A$ 是 $0-1$ 矩阵，那么 $r(A)$ 为 $A$ 的最大特征值.

**例 4.32.** 设 $(X,d)$ 为紧度量空间 $T:X\to X$ 是 Lipschitz 的，即存在 $L>0$ 使得 $d(Tx,Ty)\le Ld(x,y)$ 对于任意的 $x,y\in X$ 都成立. 则 $h(T)\le\log(\max\{1,L\})\overline{\dim}_B X$. 这里的 $\overline{\dim}_B X$ 的含义，我们需要做一定的解释. 对于任意的 $\delta>0$，记 $N_\delta(X)$ 为覆盖 $X$ 的 $\delta$ 球的最小数，那么我们定义

$$\overline{\dim}_B X=\limsup_{\delta\to0^+}\frac{\log N_\delta(X)}{-\log\delta}.$$

一个很简单的特例是对于 $T:\mathbb{R}\to\mathbb{R}$ 满足 $T:x\mapsto 2x$. 则 $T\times T\times T$ 诱导了 $\mathbb{T}^3$ 到自身的映射. 这里的拓扑熵为 $3\log2$，$\log2$ 是单个 $\mathbb{T}$ 的拓扑熵，3 是维数.

证明. 当 $L\le1$ 时候，我们有，对于任意的 $x\in X$，$\varepsilon>0$ 满足 $T(\overline{B}_\varepsilon(x))\subset\overline{B}_\varepsilon(Tx)$. 设 $\{x_i\}_{i=1}^{r_1(\varepsilon)}$ 为 $X$ 的 $(1,\varepsilon)$ 生成集，则我们断言 $\{x_i\}_{i=1}^{r_1(\varepsilon)}$ 是 $X$ 的 $(n,\varepsilon)$ 生成集. 这是因为，对于任意的 $y\in X$，存在 $x_i$ 使得

$$d(y,x_i)\le\varepsilon,\ d(Ty,Tx_i)\le\varepsilon,...,d(T^{n-1}y,T^{n-1}x_i)\le\varepsilon$$

因此 $r_n(\varepsilon)=r_1(\varepsilon)$ 于是可得

$$h(T)=\lim_{\varepsilon\to0}\limsup_{n\to\infty}\frac{\log r_n(\varepsilon)}{\mathbf{n}}=0$$

于是就证明了此情形时候结论成立. 下面我们假设 $L>1$. 对于任意的 $\varepsilon>0$，$n>0$ 令

$$p=\overline{\dim}_B X=\limsup_{\delta\to0}\frac{\log N_\delta(x)}{-\log\delta}=\limsup_{\delta\to0}\frac{\log r_1(\delta)}{-\log\delta}.$$

上面我们用到了 $N_\delta(X)$ 和 $r_1(\delta)$ 相等的性质. 对于任意的 $\eta>0$，存在 $\delta_0>0$ 对于任意的 $\delta\in(0,\delta_0]$ 时，根据上极限的性质

$$\frac{\log N_\delta(X)}{-\log\delta}\le p+\eta.$$

也就是说此时 $N_\delta(X)\le\frac{1}{\delta^{p+\eta}}$. 那么我们令 $\delta=\frac{\varepsilon}{L^n}$. 那么若 $\{x_i\}_{i=1}^{N_\delta(X)}$ 为 $(1,\delta)$ 生成集，那么 $\{x_i\}_{i=1}^{N_\delta(X)}$ 也是 $(n,\varepsilon)$ 生成集. 所以 $r_n(\varepsilon)\le N_\delta(X)\le\frac{1}{\delta^{p+\eta}}=\frac{L^{n(p+\eta)}}{\varepsilon^{p+\eta}}$. 于是根据熵的计算公式，可得

$$h(T)=\lim_{\varepsilon\to0}\limsup_{n\to\infty}\frac{\log r_n(\delta)}{n}\le\lim_{\varepsilon\to0}\limsup_{n\to\infty}\frac{n(p+\eta)\log L-(p+\eta)\log\varepsilon}{n}=(p+\eta)\log L.$$

根据 $\eta$ 的任意性，结论成立. □

对于 $T:\mathbb{T}\to\mathbb{T}$ 满足 $x\mapsto 2x$ 那么 $h(T)=\log2$. 这说明上面这个不等式的等号是能取到的.

**例 4.33.** 考虑矩阵

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

和它诱导的环面双曲自同构 $T:\mathbb{T}^2\to\mathbb{T}^2$，即 $T(x_1,x_2)=(2x_1+x_2,x_1+x_2)\pmod 1$. 矩阵 $A$ 有特征值

$$\lambda_1=\frac{3+\sqrt5}{2},\quad\lambda_2=\frac{3-\sqrt5}{2}.$$

下面我们证明 $h(T)=\log\frac{3+\sqrt5}{2}=\log\lambda_1$.

固定 $\varepsilon>0$. 我们用有限多个球

$$B((x_1^i,x_2^i),\varepsilon)=\{(z_1,z_2)\in\mathbb{T}^2:\|(x_1^i,x_2^i)-(z_1,z_2)\|\le\varepsilon\},\quad i=1,2,...,k$$

覆盖 $\mathbb{T}^2$. 这些球的中心分别在 $(x_1^1,x_2^1),...,(x_1^k,x_2^k)$，半径为 $\varepsilon$. 我们可设定单位圆周 $S^1$ 的弧长是 1 且 $\mathbb{T}^2=S^1\times S^1$，进而限制 $k\le\frac{4}{\varepsilon^2}$. 在每个点 $(x_1^i,x_2^i)$ 邻近我们可以将这些球表示为

$$\mathrm{Box}(x_1^i,x_2^i)=\{(x_1^i,x_2^i)+\alpha v_1+\beta v_2:-\varepsilon\le\alpha,\beta\le\varepsilon\},$$

其中 $v_1, v_2$ 是 $A$ 相应于 $\lambda_1, \lambda_2$ 的单位特征向量. 对每个 $n \ge 1$ 和 $1 \le i \le k$, 我们可以考虑 $\mathrm{Box}(x_1^i, x_2^i)$ 的有限子集

$$R(x^i) = R(x_1^i, x_2^i) = \left\{(x_1^i, x_2^i) + \frac{j\varepsilon}{\lambda_1^n} v_1 : j = -[\lambda_1^n], ..., [\lambda_1^n]\right\},$$

这里 $[a]$ 表示不超过 $a$ 的最大整数. 此集合的基数是 $2[\lambda_1^n]+1$. 令 $R = \cup_{i=1}^k R(x^i)$, 则 $R$ 的基数不超过 $k(2[\lambda_1^n]+1)$. 我们有以下两个断言:
**断言 1**: $R$ 是一个 $(n, 2\varepsilon)$ 生成集.

*断言 1 的证明*. 对任意 $(z_1, z_2) \in \mathbb{R}^2/\mathbb{Z}^2$, 我们可以找到某个 $i = 1, 2, ..., k$, 使得 $\|(x_1^i, x_2^i) - (z_1, z_2)\| < \varepsilon$. 特别地, 这意味着 $(z_1, z_2) \in \mathrm{Box}(x_1^i, x_2^i)$. 于是, 我们可以把 $(z_1, z_2)$ 写成 $(z_1, z_2) = (x_1^i, x_2^i) + \alpha v_1 + \beta v_2$ 的形式, 这里对应的 $\alpha, \beta$ 取值于 $[-\varepsilon, \varepsilon]$. 若选 $-[\lambda_1^n] \le j \le [\lambda_1^n]$, 满足 $|\alpha - \frac{j\varepsilon}{\lambda_1^n}| \le \frac{\varepsilon}{2\lambda_1^n}$, 那么可以在 $R(x_1^i, x_2^i)$ 选出点 $(\omega_1, \omega_2) = (x_1^i, x_2^i) + \frac{j\varepsilon}{\lambda_1^n} v_1$. 对任意 $0 \le r \le n-1$, 我们有

$$T^r(z_1, z_2) = T^r(\omega_1, \omega_2) + \left(\alpha - \frac{j\varepsilon}{\lambda_1^n}\right) A^r v_1 + \beta A^r v_2 = T^r(\omega_1, \omega_2) + \left(\alpha - \frac{j\varepsilon}{\lambda_1^n}\right) \lambda_1^r v_1 + \beta \lambda_2^r v_2$$

于是有

$$\|T^r(z_1, z_2) - T^r(\omega_1, \omega_2)\| \le \left|\left(\alpha - \frac{j\varepsilon}{\lambda_1^n}\right)\right| \lambda_1^r \|v_1\| + |\beta| \lambda_2^r \|v_2\| \le \frac{\varepsilon \lambda_1^r}{2\lambda_1^n} + \varepsilon \le \frac{\varepsilon}{2} + \varepsilon < 2\varepsilon.$$

断言得证. □

由断言 1 我们得到

$$h(T) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log r_n(2\varepsilon) \le \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log k(2[\lambda_1^n]+1) = \lim_{\varepsilon \to 0} \log \lambda_1 = \log \lambda_1.$$

为了得到反向的不等式, 我们固定一个点 $(x_1, x_2) \in \mathbb{T}^2$. 对 $\varepsilon > 0, n \ge 1$, 我们考虑集合

$$S = \left\{(x_1, x_2) + \frac{j\varepsilon}{\lambda_1^n} v_1 : j = -[\lambda_1^n], ..., [\lambda_1^n]\right\}.$$

**断言 2**: $S$ 是一个 $(n+1, \varepsilon)$ 分离集.

*证明*. $S$ 中任何两个不同的点都可以写为下面的形式:

$$(u_1, u_2) = (x_1, x_2) + \frac{i\varepsilon}{\lambda_1^n} v_1, \quad (\omega_1, \omega_2) = (x_1, x_2) + \frac{j\varepsilon}{\lambda_1^n} v_1,$$

其中 $-[\lambda_1^n] \le i, j \le [\lambda_1^n]$. 对于 $0 \le r \le n$, 我们有

$$\|T^r(u_1, u_2) - T^r(\omega_1, \omega_2)\| = \left\|\frac{(j-i)\varepsilon}{\lambda_1^n} T^r(v_1)\right\| = |j-i| \frac{\varepsilon}{\lambda_1^{n-r}}.$$

总会存在某个 $0 \le r \le n$, 使得 $\|T^r(u_1, u_2) - T^r(\omega_1, \omega_2)\| > \varepsilon$. 断言得证. □

$(n+1, \varepsilon)$ 分离集 $S$ 的基数为 $2[\lambda_1^n]+1$, 故 $2[\lambda_1^n]+1 \le s_{n+1}(\varepsilon)$. 于是有

$$h(T) = \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log s_{n+1}(\varepsilon) \ge \lim_{\varepsilon \to 0} \limsup_{n \to \infty} \frac{1}{n} \log(2[\lambda_1^n]+1) = \lim_{\varepsilon \to 0} \log \lambda_1 = \log \lambda_1.$$

上面的分析可知 $h(T) = \log \lambda_1 = \log\left(\frac{3+\sqrt{5}}{2}\right)$.

**例 4.34.** $T : \mathbb{T} \to \mathbb{T}$ 式同胚, 则 $h(T) = 0$.

证明. 取 $\mathbb{T}$ 上的 $N$ 等分点 $A=\{x_i\}_{i=1}^N$ 使得 $\frac{1}{N}<\varepsilon$. 则有 $r_1(\varepsilon)\le N$. 下面我们证明 $r_n(\varepsilon)\le nN$. 根据这一点以及拓扑熵的定义，则 $h(T)=0$. 我们用归纳法证明，首先 $n=1$ 的时候结论成立. 假设 $n$ 时，结论成立，那么此时取对应的 $(n,\varepsilon)$ 生成集 $F_n$，我们取 $F_{n+1}=F_n\cup T^{-n}A$. 下面我们证明 $F_{n+1}$ 是 $(n+1,\varepsilon)$ 生成集. 对应任意的 $y\in\mathbb{T}$，存在 $x\in F_n$ 使得 $d_n(x,y)\le\varepsilon$ (这是根据 $F_n$ 是 $(n,\varepsilon)$ 生成集的性质). 于是我们考虑 $d(T^nx,T^ny)$. 有两种情况，若 $d(T^nx,T^ny)\le\varepsilon$ 则 $d_{n+1}(x,y)\le\varepsilon$. 若 $d(T^nx,T^ny)>\varepsilon$ 我们可以在最前面取 $\varepsilon$ 充分小，使得 $\mathbb{T}$ 上连接距离小于 $\varepsilon$ 的两点的劣弧被 $T$ 映射成劣弧 (用到了 $T$ 的一致连续性)，那么就存在 $z\in A$ 使得 $d(z,T^ny)<\varepsilon$ 从而有对于任意的 $0\le k\le n$ 满足 $d(T^kT^{-n}z,T^ky)<\varepsilon$. 这样就完成了证明. □

# 5 变分原理

设 $(X,d)$ 是紧度量空间，对于 $T:(X,d)\to(X,d)$，$\mu$ 是 $T$ 的不变测度，我们期待结论 $h(T)=\sup_\mu h_\mu(T)$. 这样的结论是受 $\Sigma_d$ 上的左平移的启发，在这个模型中，拓扑熵是 $\log d$. 如果我们在 Bernoulli 模型中取概率向量为 $(p_0,p_1,...,p_{d-1})$ 诱导的测度，那么其测度熵为 $-\sum_{i=0}^{d-1}p_i\log p_i$. 其最大值正好是 $\log d$.

## 5.1 度量空间上的测度

设 $(X,d)$ 为紧度量空间，$\mathcal{B}(X)$ 是 $X$ 的 Borel $\sigma$-代数，$\mathcal{B}(X)$ 上的概率称为 Borel 测度，记 $M(X)$ 为 $X$ 上的所有 Borel 概率构成的集合，对于 $\mu,m\in\mathcal{B}(X)$，记号 $\mu=m$ 的意思是指它们作为 $\mathcal{B}(X)\to\mathbb{R}_+$ 上的函数的时候是相等的. 我们还记 $C(X)$ 是 $X$ 上的连续函数的全体.

**引理 5.1.** 设 $\mu,m\in M(X)$，则 $\mu=m$ 当且仅当 $\int f\mathrm{d}m=\int f\mathrm{d}\mu$ 对于任意的 $f\in C(X)$ 都成立.

证明. 一方面，若 $\mu=m$，那么 $\int f\mathrm{d}m=\int f\mathrm{d}\mu$ 对于任意的 $f\in C(X)$ 是显然的，只是因为我们可以用简单函数来逼近连续函数. 反之，我们需要先证明一个引理如下.

**引理 5.2.** $m\in M(X)$，则对于任意的 $B\in\mathcal{B}(X)$ 有

$$m(B)=\inf\{m(U):B\subset U,U\text{为开集}\}=\sup\{m(C):C\subset B,C\text{为闭集}\}.$$

这个引理实际上就是之前的注 1.42. 因此，我们只要验证对于任意的闭集 $C$，有 $m(C)=\mu(C)$ 即可. 令 $U_n=\{x\in X:d(x,C)<\frac{1}{n}\}$，那么有 $C=\cap_{n\ge1}U_n$. 我们取 $f_n:X\to\mathbb{R}$ 使得

$$f_n(x)=\frac{d(x,U_n^c)}{d(x,U_n^c)+d(x,C)}.$$

很显然 $f_n(x)\to\chi_C(x)$ 并且一直有界，于是根据有界收敛定理，可得

$$m(C)=\int\chi_C(x)\mathrm{d}m=\lim_{n\to\infty}\int f_n\mathrm{d}m=\lim_{n\to\infty}\int f_n\mathrm{d}\mu=\mu(C).$$

这样我们就完成了证明. □

**定义 5.3.** 设 $\mu\in M(X)$，定义 $J=J_\mu:C(X)\to\mathbb{R}$ 为泛函，满足 $J_\mu(f)=\int f\mathrm{d}\mu$.

对于上面定义的泛函，我们很容易得到一些简单的性质. $J_\mu$ 为线性的，$J_\mu(1)=1$，$J_\mu(f)\ge0$ 对于 $f\ge0$ 成立，即 $J_\mu$ 是正的. $J_\mu$ 是有界的，此时我们在 $C(X)$ 上面定义 $\|f\|=\sup_{x\in X}|f(x)|$，并且 $C(X)$ 为 Banach 空间.

**定理 5.4** (Riesz 表示定理)**.** *设 $X$ 为紧度量空间，$J : C(X) \to \mathbb{R}$ 为满足上述四条性质的线性泛函，那么存在唯一的测度 $\mu \in M(X)$，使得 $J = J_\mu$ 并且 $J(1) = 1$.*

上述定理证明中的唯一性是显然的，但是存在性证明非常复杂，在这里，我们略去证明.

**定义 5.5** ($M(X)$ 上的弱 * 拓扑)**.** $M(X)$ 上的弱 * 拓扑是使得对于任意的 $f \in C(X)$ $J_\mu(f) \to \int f\mathrm{d}\mu$ 都连续的最小拓扑.

上述最小的拓扑的含义是指，对于任意的 $c \in \mathbb{R}$，$\{\mu \in M(X) : \int f\mathrm{d}m > c\}$ 与 $\{\mu \in M(X) : \int f\mathrm{d}m < c\}$ 都是开集，事实上，上述的集合可以被视为是该拓扑的拓扑子基.

更近一步，对于任意的 $f_1, f_2, ..., f_k \in C(X)$，$\varepsilon > 0$ 以及 $m \in M(X)$ 则

$$V_m(f_1, f_2, ..., f_k, \varepsilon) = \left\{\mu \in M(X) : \left|\int f_i\mathrm{d}m - \int f_i\mathrm{d}\mu\right| < \varepsilon,\ \forall 1 \le i \le k\right\}$$

为 $m$ 的一个邻域基. 事实上，对于任意的 $m$ 的邻域 $U$，存在 $f_1, f_2, ..., f_k$，$\varepsilon > 0$ 使得

$$V_m(f_1, f_2, ..., f_k, \varepsilon) \subset U.$$

**定理 5.6.** *设 $\mu_n, m \in M(X)$ 则 $\mu_n \rightharpoonup^* m$ 当且仅当对于任意的 $f \in C(X)$ 满足*

$$\lim_{n\to\infty} \int f\mathrm{d}\mu_n = \int f\mathrm{d}m.$$

*证明.* 若 $\mu_n \rightharpoonup^* m$ 则考虑 $J_f : M(X) \to \mathbb{R}$ 连续，则 $J_f(\mu_n) \to J_f(m)$ 于是就成立了. 另一方面，反设对于任意的 $f \in C(X)$ 有 $J_f(\mu_n) \to J_f(m)$. 我们有，要证明，对于任意的 $m$ 的邻域 $U$，存在 $N$，对于任意的 $n > N$ 有 $\mu_n \in U$. 对于 $U$，我们可以找到 $f_1, f_2, ..., f_k$ 有 $V_m(f_1, f_2, ..., f_k) \subset U$. 那么根据假设结论就成立了. □

**定理 5.7.** *设 $(X, d)$ 为紧度量空间，$M(X)$ 上的弱 * 拓扑可度量化，也即 $M(X)$ 存在距离 $d$，使得 $d$ 诱导上述的弱 * 拓扑.*

*证明.* 任给 $C(X)$ 的稠密子集 $\{f_k\}_{k\ge1}$. 对于任意的 $\mu, m \in M(X)$ 我们定义

$$d(\mu, m) = \sum_{k=1}^{\infty} \frac{\left|\int f_i\mathrm{d}m - \int f_i\mathrm{d}\mu\right|}{2^k\|f_k\|}.$$

我们只要验证三条度量的性质就可，显然对称性和三角不等式都是显然的，于是我们只需要证明正定性即可，即证明 $d(\mu, m) = 0$ 可以推出 $\mu = m$. 这说明对于任意的 $f_n$ 有 $\int f_n\mathrm{d}\mu = \int f_n\mathrm{d}m$. 对于任意的 $f \in C(X)$ 以及 $\varepsilon > 0$，我们可以选择 $f_n$ 使得 $\|f_n - f\| < \varepsilon$. 这样就能够证明 $\int f\mathrm{d}\mu = \int f\mathrm{d}m$. 再用连续函数逼近特征函数即可证明 $\mu = m$. □

**注 5.8.** 设 $x_n \to x$，则我们注意到 $\delta_{x_n} \to \delta_x$. 但是值得注意的是，对于 $A \subset X$，$\delta_{x_n}(A) \to \delta_x(A)$ 并不是总成立的.

根据上面的注记，我们有下面的刻画.

**定理 5.9.** *设 $\mu_n \rightharpoonup^* \mu$，$\mu_n, \mu \in M(X)$ 则有下面的性质成立.*

1. *若 $C$ 为闭集，则 $\limsup_{n\to\infty} \mu_n(C) \le \mu(C)$.*
2. *若 $U$ 为开集，则 $\liminf_{n\to\infty} \mu_n(U) \ge \mu(U)$.*
3. *若 $A \in \mathcal{B}(X)$ 并且 $\mu(\partial A) = 0$ 则 $\lim_{n\to\infty} \mu_n(A) = \mu(A)$.*

*证明.* 我们分开证明上述三条.

*1 的证明.* 对于任意的 $\varepsilon > 0$ 取 $U$ 为开集，满足 $C \subset U$，使得 $\mu(U \backslash C) < \varepsilon$. 取 $f \in C(X)$ 满足 $f(x) \in [0,1]$ 并且 $f|_C \equiv 1$ 以及 $f|_{U^c} \equiv 0$. 所以可得

$$\mu_n(C) = \int \chi_C \mathrm{d}\mu_n \leq \int f \mathrm{d}\mu_n.$$

两侧取上极限即可得到结论. □

*2 的证明.* 根据 1，可得 $\liminf_{n\to\infty} \mu_n(U) = \liminf(1-\mu(U^c)) = 1 - \limsup_{n\to\infty} \mu_n(U^c)$. 于是就证明了 2 的结论. □

*3 的证明.* 利用前两条结论，可得

$$\liminf_{n\to\infty} \mu_n(A) \geq \liminf_{n\to\infty} \mu_n(\mathring{A}) \geq \mu(\overline{A}) \geq \limsup_{n\to\infty} \mu_n(\overline{A}) \geq \limsup_{n\to\infty} \mu_n(A).$$

于是就证明了结论. □

□

下面我们给出在弱 * 拓扑下，$M(X)$ 的紧性.

**定理 5.10** (Alaoglu)**.** *$M(X)$ 在弱 * 拓扑下是紧的.*

此证明是基本的，利用对角线原理，这里我们略去证明.

## 5.2 不变测度与遍历分解

设 $(X,d)$ 是紧度量空间，$T : (X,d) \to (X,d)$ 连续，$M(X)$ 为 $X$ 上的概率测度全体，上面赋予弱 * 拓扑，则 $M(X) \subset (C(X))^*$. 那么对于任意 $\mu, m \in M(X)$，我们可以定义 $p\mu + (1-p)m$ 如下:

$$(p\mu + (1-p)m)(A) = p\mu(A) + (1-p)m(A),$$

其中 $A \in \mathcal{B}(X)$.

回到最开始的定义，我们假设 $T : X \to X$ 是连续的，实际上是假设 $T^{-1}(\mathcal{B}(X)) \subset \mathcal{B}(X)$，即 Borel 集的原像是 Borel 集合，于是我们可以有下面的定义.

**定义 5.11.** 定义 $T_* : M(X) \to M(X)$ 满足 $\mu \mapsto T_*\mu = \mu \circ T^{-1}$，即 $T_*\mu(B) = \mu(T^{-1}B)$ 对于任意的 $B \in \mathcal{B}(X)$ 成立.

根据上面的顶柜我们很容易得到下面的性质.

1. 对于任意的 $f \in C(X)$，$T_*\mu(f) = \int f \mathrm{d}T_*\mu = \int f \circ T \mathrm{d}\mu$. (证明遵循现在特征函数，简单函数到连续函数这一套流程.) 据此，可以定义 $U_T : C(X) \to C(X)$ 满足 $\langle T_*\mu, f\rangle = \langle \mu, U_T f\rangle$，这里 $U_T f = f \circ T^{-1}$.
2. $T_* : M(X) \to M(X)$ 是连续的线性映射. (这里的线性性质是显然的，连续性只需要利用测度弱 * 收敛的定义即可直接验证.)
3. 定义 $M(X,T) = \{\mu \in M(X) : T_*\mu = \mu\}$ 为关于 $T$ 的不变测度，那么 $M(X,T)$ 是一个紧的并且凸的. (这两条性质来自于上一条 $T_*$ 的连续性和线性性.)

之后我们会研究 $M(X,T)$ 的一些性质，所以我们先给出一个重要定义.

**定义 5.12.** $\mu \in M(X,T)$ 被称为顶点，若 $\mu = p\mu_1 + (1-p)\mu_2$，$0<p<1$ 并且 $\mu_1,\mu_2 \in M(T)$ 能推出 $\mu=\mu_1=\mu_2$.

定义 $E(X,T) \subset M(X,T)$ 表示 $T$ 的遍历测度的全体. 下面我们给出一个关于 $E(X,T)$ 的刻画.

**定理 5.13.** 设 $\mu \in M(X,T)$. 那么 $\mu \in E(X,T)$ 当且仅当 $\mu$ 是 $M(X,T)$ 的顶点.

证明. 一方面若 $\mu$ 是顶点，那么若 $\mu \notin E(X,T)$ 则有存在 $B \in \mathcal{B}(X)$ 使得 $T^{-1}B = B$ 并且 $0 < \mu(B) < 1$. 那么我们可以定义 $\mu_1(\cdot) = \mu(\cdot|B) = \frac{\mu((\cdot)\cap B)}{\mu(B)}$ 以及 $\mu_2(\cdot) = \mu(\cdot|B^c) = \frac{\mu((\cdot)\cap B^c)}{\mu(B^c)}$. 所以可得

$$\mu(B)\mu_1 + \mu(B^c)\mu_2 = \mu.$$

这样就与顶点矛盾，因为这里 $\mu_1,\mu_2$ 显然不同. 另一方面若 $\mu \in E(X,T)$，设 $\mu_1,\mu_2 \in M(X,T)$ 以及 $p \in (0,1)$ 满足 $\mu = p\mu_1 + (1-p)\mu_2$. 那么 $\mu(A) = 0$ 意味着 $\mu_1(A) = \mu_2(A) = 0$. 于是有 $\mu_1,\mu_2 \ll \mu$. 根据 Radon-Nikodym 定理，可得，存在 $f = \frac{\mathrm{d}\mu_1}{\mathrm{d}\mu}$. 我们只要证明 $f = 1$ 即可. 令 $E = \{x \in X : f(x) < 1\} \in \mathcal{B}(X)$，可知由于 $\mu_1 \in M(X,T)$，可得 $\mu_1(E) = \mu_1(T^{-1}E)$，那么有

$$\mu_1(E \backslash T^{-1}E) = \mu_1(T^{-1}E \backslash E).$$

于是根据 Radon-Nikodym 导数的性质，可得

$$\int_{E\backslash T^{-1}E} f\mathrm{d}\mu = \int_{T^{-1}E\backslash E} f\mathrm{d}\mu.$$

若 $\mu(E\backslash T^{-1}E) > 0$，则有上式左侧严格小于 $\mu(E\backslash T^{-1}E)$ 右侧大于等于 $\mu(T^{-1}E\backslash E)$，而根据 $\mu \in M(X,T)$ 就得到了矛盾，所以 $\mu(E\backslash T^{-1}E) =$，于是 $\mu(E\Delta T^{-1}E) = 0$. 再根据 $\mu$ 的遍历性，可得 $\mu(E) = 0$ 或 1. 若 $\mu(E) = 0$ 证明就自动成立了，若 $\mu(E) = 1$，有

$$1 = \mu_1(X) = \int_X f\mathrm{d}\mu = \int_E f\mathrm{d}\mu < \mu(E) = 1$$

矛盾，所以就完成了证明. □

**推论 5.14.** $E(X,T)$ 是 $M(X,T)$ 的 $G_\delta$ 集.

证明. 令 $F_n = \{\mu \in M(X,T) : \exists \mu_1,\mu_2 \in M(X,T), \mu = \frac{\mu_1+\mu_2}{2},\ d(\mu_1,\mu_2) \geq \frac{1}{n}\}$. 那么可得 $M(X,T)\backslash E(X,T) = \cup_{n\geq 1} F_n$. 由于 $F_n$ 都是闭集，那么就能得到 $E(X,T)$ 是 $G_\delta$ 集，即可数个开集的交. □

下面我们证明一个重要定理，该定理表明 $M(X,T)$ 是非空的.

**定理 5.15** (Krylov-Bogliuloov)**.** $M(X,T) \neq \emptyset$. 特别地，对于任给的 $\mu_n \in M(X)$，则 $\{\frac{1}{n}\sum_{i=0}^{n} T_*^i \mu_i\}$ 的任一极限点都在 $M(X,T)$ 中.

证明. 设当 $k \to \infty$ 时，$\frac{1}{n_k}\sum_{i=0}^{n_k-1} T_*^i\mu_i \rightharpoonup *\mu \in M(X)$. 往证对于任意的 $f \in C(X)$ 有 $\int f\circ T\mathrm{d}\mu = \int f\mathrm{d}\mu$. 那么可得

$$\left|\int f\circ T\mathrm{d}\mu - \int f\mathrm{d}\mu\right| = \lim_{k\to\infty}\left|\int f\circ T\mathrm{d}m_k - \int f\mathrm{d}m_k\right|,$$

其中 $m_k = \frac{1}{n_k}\sum_{i=0}^{n_k-1} T_*^i\mu_i$. 把 $m_k$ 的定义带入后，会发现上式能够错位相消，进而得到

$$\left|\int f\circ T\mathrm{d}\mu - \int f\mathrm{d}\mu\right| \leq \frac{2\|f\|}{n_k}.$$

于是就证明了结论成立. □

对于任意的 $x \in X$ 有 $\frac{1}{n}\sum_{i=0}^{n-1} T_*^i \delta_x = \mu_{n,x}$ 为经验测度，很显然对于任意的 $f \in C(X)$ 有

$$\mu_{n,x}(f) = \frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i x.$$

若 $x$ 为 $T$ 的 $n$ 周期点，那么 $\frac{1}{n}\sum_{i=0}^{n-1} \delta_{T^i x}$ 是 $T$ 的不变测度，事实上是 $T$ 的遍历测度.

我们回忆 Birkhoff 遍历定理如下.

**定理 5.16.** 设 $T : (X, \mathcal{B}, \mu) \to (X, \mathcal{B}, \mu)$ 是保测系统，$\mu$ 为遍历的，则对于任意的 $f \in L^1$ 有对 *a.e.* $x$ 使得 $\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i x \to \int f \mathrm{d}\mu$.

下面我们给出一个关于连续函数的刻画.

**定理 5.17.** 设 $T : (X, d) \to (X, d)$ 是拓扑动力系统，并且 $\mu \in E(X, T)$ 则存在 $Q_\mu \in \mathcal{B}(X)$ 使得 $\mu(Q_\mu) = 1$ 并且对于任意的 $x \in Q_\mu$ 满足 $f \in C(X)$ 都有

$$\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i x \to \int f \mathrm{d}\mu,$$

对 $n \to \infty$ 成立.

上述定理相当于是找到了一个与函数无关的集合，使得在这个集合上面，连续函数的平均收敛到 $f$ 上面.

*证明.* 任取 $C(X)$ 的稠密子集 $\{f_k\}_{k \geq 1}$ 并且令

$$Q_\mu^k = \left\{ x \in X : \frac{1}{n}\sum_{i=0}^{n-1} f_k \circ T^i x \to \int f_k \mathrm{d}\mu \right\} \in \mathcal{B}(X).$$

很显然根据上面的 Birkhoff 遍历定理，可得 $\mu(Q_\mu^k) = 1$ 令 $Q_\mu = \cap_{k \geq 1} Q_\mu^k$，于是有 $\mu(Q_\mu) = 1$. 那么这里的 $Q_\mu$ 即为所求. 事实上，对于任意的 $x \in Q_\mu$ 以及 $f \in C(X)$，

$$\begin{aligned}\left|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i x - \int f \mathrm{d}\mu\right| \leq& \left|\frac{1}{n}\sum_{i=0}^{n-1} f \circ T^i x - \frac{1}{n}\sum_{i=0}^{n-1} f_k \circ T^i x\right| \\ &+ \left|\frac{1}{n}\sum_{i=0}^{n-1} f_k \circ T^i x - \int f_k \mathrm{d}\mu\right| + \left|\int f \mathrm{d}\mu - \int f_k \mathrm{d}\mu\right|.\end{aligned}$$

我们先取 $f_k$ 充分接近 $f$，再取 $n$ 充分大，即可证明结论. □

**定义 5.18.** 我们称上面得到的 $Q_\mu$ 为 $\mu$ 的通有点.