**Definition 1.** Recall the following definitions: algebra, $\sigma$-algebra, $\sigma$-algebra generated by a collection of subsets, finitely additive, countably additive, measure, measure space.

**Definition 2.** Let $(X, \mathcal{A}, \mu)$ be a measure space. If $\mu(X) = 1$, we say that $(X, \mathcal{A}, \mu)$ is a probability space.

**Definition 3.** Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be measure spaces. Let $T : X_1 \to X_2$.

We say that $T$ is measurable if and only if $B \in \mathcal{A}_2$ implies that $T^{-1}(B) \in \mathcal{A}_1$.

We say that $T$ is measure preserving if and only if $T$ is measurable and $\mu_1(T^{-1}(B)) = \mu_2(B)$ for all $B \in \mathcal{A}_2$.

We say that $T$ is an invertible measure preserving transformation if and only if $T$ is a bijection, and both $T$ and $T^{-1}$ are measure preserving.

**Definition 4.** Let $X$ be a metric space. Let $\mathcal{B}(X)$ denote the smallest $\sigma$-algebra containing each open subset of $X$. We call $\mathcal{B}(X)$ the collection of Borel sets. If $(X, \mathcal{B}(X), \mu)$ is a probability space, we call $\mu$ a Borel probability measure. We will let $M(X)$ denote the set of Borel probability measures.

**Proposition 5.** Let $X,Y$ be metric spaces, and let $f : X \to Y$ be continuous. If $B \in \mathcal{B}(Y)$, then $f^{-1}(B) \in \mathcal{B}(X)$.

**Definition 6.** Let $X$ be a metric space, and let $f : X \to X$ be continuous. Let $\mu \in M(X)$. If $f$ is measure preserving with respect to the probability space $(X, \mathcal{B}(X), \mu)$, then we say that $\mu$ is $f$-invariant. We will let $M(X,f)$ denote the set of $f$-invariant Borel probability measures.

**Definition 7.** Let $X$ be a set, and let $\mathcal{A}$ be a algebra of subsets of $X$. Let $\tau : \mathcal{S} \to [0,\infty]$.

1. We say that $\tau$ is finitely additive if and only if $\tau(\emptyset) = 0$, and $\tau(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} \tau(E_i)$ whenever $\{E_i\}$ is a finite pairwise disjoint collection of elements of $\mathcal{A}$.

2. We say that $\tau$ is countably additive if and only if $\tau(\emptyset) = 0$, and $\tau(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \tau(E_i)$ whenever $\{E_i\}$ is a countable pairwise disjoint collection of elements of $\mathcal{A}$ with $\bigcup_{i=1}^{\infty} E_i \in \mathcal{A}$.

**Proposition 8.** Let $\mathcal{A}$ be an algebra of subsets of $X$, and let $\tau : \mathcal{A} \to [0,\infty]$ be a finitely additive function with $\tau(X) = 1$. Suppose that for every decreasing sequence of sets $\{E_i\}$ with each $E_i \in \mathcal{A}$ and $\bigcap_{n=1}^{\infty} E_n = \emptyset$ we have $\tau(E_n) \to 0$. Then $\tau$ is countably additive.

**Theorem 9.** Let $\mathcal{A}$ be an algebra of subsets of $X$, and let $\tau : \mathcal{A} \to [0,\infty]$ be a countably additive function with $\tau(X) = 1$. Let $\mathcal{B}$ be the $\sigma$-algebra generated by $\mathcal{A}$. Then there is a unique function $\mu : \mathcal{B} \to [0,\infty]$ which extends $\tau$ such that $(X, \mathcal{B}, \mu)$ is a probability space.

**Definition 10.** Let $X$ be a set. A collection $\mathcal{S}$ of subsets of $X$ is called a semi-algebra if and only if

1. $\emptyset \in \mathcal{S}$.
2. If $A, B \in \mathcal{S}$, then $A \cap B \in \mathcal{S}$.
3. If $A \in \mathcal{S}$, then $A^C = \bigcup_{i=1}^{n} E_i$, where $\{E_i\}$ is a finite, pairwise disjoint, collection of elements of $\mathcal{S}$.

**Definition 11.** Let $X$ be a set, and let $\mathcal{S}$ be a semi-algebra of subsets of $X$. Let $\tau : \mathcal{S} \to [0,\infty]$.

1. We say that $\tau$ is finitely additive if and only if $\tau(\emptyset) = 0$, and $\tau(\bigcup_{i=1}^{n} E_i) = \sum_{i=1}^{n} \tau(E_i)$ whenever $\{E_i\}$ is a finite pairwise disjoint collection of elements of $\mathcal{S}$ with $\bigcup_{i=1}^{n} E_i \in \mathcal{S}$.

2. We say that $\tau$ is countably additive if and only if $\tau(\emptyset) = 0$, and $\tau(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \tau(E_i)$ whenever $\{E_i\}$ is a countable pairwise disjoint collection of elements of $\mathcal{S}$ with $\bigcup_{i=1}^{\infty} E_i \in \mathcal{S}$.

**Proposition 12.** Let $\mathcal{S}$ be a semi-algebra of subsets of $X$. The algebra generated by $\mathcal{S}$ consists precisely of those subsets of $X$ which can be expressed as a finite union of pairwise disjoint subsets of $\mathcal{S}$.

**Proposition 13.** Let $X$ be a set, and let $\mathcal{S}$ be a semi-algebra of subsets of $X$. Suppose that $\tau : \mathcal{S} \to [0,\infty]$ is finitely additive. Let $\mathcal{A}$ be the algebra generated by $\mathcal{S}$. Then there is a unique finitely additive function $\tau_1 : \mathcal{A} \to [0,\infty]$ which is an extension of $\tau$. If $\tau$ is countably additive, then so is $\tau_1$. 
Example 14. Let $X = [0, 1]$, and let $\mathcal{S}$ denote the collection of all connected subsets of $X$. Then $\mathcal{S}$ is a semi-algebra of subsets of $X$. There is a unique $\mu : \mathcal{B}(X) \to [0, \infty]$ such that

1. $(X, \mathcal{B}, \mu)$ is a probability space.

2. for any interval $J \subset X$, $\mu(J)$ is the length of $J$.

We will call $\mu$ Lebesgue measure on $\mathcal{B}(X)$.

Example 15. Let $X = S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, and let $\mathcal{S}$ denote the collection of all connected subsets $D$ of $X$ such that the length of $D$ is less than $\pi$. Then $\mathcal{S}$ is a semi-algebra of subsets of $X$. There is a unique $\mu : \mathcal{B}(X) \to [0, \infty]$ such that

1. $(X, \mathcal{B}, \mu)$ is a probability space.

2. for any arc $J \subset X$, $\mu(J)$ is $\frac{1}{2\pi}$ times the length of $J$.

We will call $\mu$ Lebesgue measure on $\mathcal{B}(X)$.

Definition 16. Let $X$ be a set. A collection $\mathcal{M}$ of subsets of $X$ is said to be a monotone class if and only if

1. If $E_1 \subset E_2 \subset E_3 \ldots$ where each $E_i \in \mathcal{M}$, then $\bigcup E_i \in \mathcal{M}$.

2. If $E_1 \supset E_2 \supset E_3 \ldots$ where each $E_i \in \mathcal{M}$, then $\bigcap E_i \in \mathcal{M}$.

Proposition 17. Let $X$ be a set, and let $\mathcal{D}$ be a collection of subsets of $X$. There is a smallest monotone class containing $\mathcal{D}$. We call this the monotone class generated by $\mathcal{D}$.

Proposition 18. Let $X$ be a set, and let $\mathcal{A}$ be an algebra of subsets of $X$. The $\sigma$-algebra generated by $\mathcal{A}$ is the monotone class generated by $\mathcal{A}$.

Proposition 19. Suppose that $(X_1, \mathcal{B}_1, m_1)$ and $(X_2, \mathcal{B}_2, m_2)$ are probability spaces and $T : X_1 \to X_2$. Let $\mathcal{S}_2$ be a semi-algebra which generates $\mathcal{B}_2$. Suppose that for all $A \in \mathcal{S}_2$ we have $T^{-1}(A) \in \mathcal{B}_1$ and $m_1(T^{-1}(A)) = m_2(A)$. Then $T$ is measure preserving.

Example 20. Let $X = [0, 1]$, and let $(X, \mathcal{B}, \mu)$ be the probability space in Example 14. Let $T : X \to X$ be the full tent map given by $T(x) = 2x$ if $x \leq \frac{1}{2}$ and $T(x) = 2 - 2x$ if $x \geq \frac{1}{2}$. Then $T$ is measure preserving.

Example 21. Let $X = S^1$, and let $(X, \mathcal{B}, \mu)$ be the probability space in Example 15. Let $T : X \to X$ be a rotation. Then $T$ is measure preserving.

Example 22. Let $X = S^1$, and let $(X, \mathcal{B}, \mu)$ be the probability space in Example 15. Let $n$ be a positive integer, and let $T : X \to X$ be given by $T(z) = z^n$. Then $T$ is measure preserving.

Example 23. Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be probability spaces. Let $\mathcal{S}$ denote the collection of all sets of the form $D_1 \times D_2$ where $D_1 \in \mathcal{A}_1$ and $D_2 \in \mathcal{A}_2$. Then $\mathcal{S}$ is a semi-algebra of subsets of $X_1 \times X_2$. Let $\mathcal{B}$ be the $\sigma$-algebra generated by $\mathcal{S}$. Then there is a unique function $\mu : \mathcal{B} \to [0, \infty]$ such that

1. $(X_1 \times X_2, \mathcal{B}, \mu)$ is a probability space (which we will call the product space).

2. If $D_1 \times D_2 \in \mathcal{S}$, then $\mu(D_1 \times D_2) = \mu_1(D_1) \cdot \mu_2(D_2)$.

Example 24. Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ the probability space in Examples 15. Let $(X_1 \times X_2, \mathcal{B}, \mu)$ be the product space. Then with respect to this space, it can be shown that the hyperbolic toral automorphism (given in Example 132 last semester) is measure preserving.

Definition 25. Let $(X, \mathcal{B}, \mu)$ be a probability space, and let $T : X \to X$ be measure preserving. Let $B \in \mathcal{B}$. A point $x \in B$ is said to be recurrent with respect to $B$ if and only if there is a positive integer $k$ with $T^k(x) \in B$.

Theorem 26. (sometimes called Poincare Recurrence Theorem) Let $(X, \mathcal{B}, \mu)$ be a probability space, and let $T : X \to X$ be measure preserving. For each $B \in \mathcal{B}$, almost every point of $B$ is recurrent with respect to $B$. 

Theorem 27. (also sometimes called Poincare Recurrence Theorem) Let $(X, B, \mu)$ be a probability space, and let $T : X \to X$ be measure preserving. Let $B \in B$. Let $A$ denote the set of $x \in B$ such that there exist positive integers $n_1 < n_2 < \ldots$ with $T^{n_i}(x) \in B$ for each $i = 1, 2, \ldots$. Then $\mu(A) = \mu(B)$.

Corollary 28. Let $f : X \to X$ be a continuous map of a compact metric space to itself. Suppose that there exists an $f$-invariant Borel probability measure $\mu$. Let $R(f)$ denote the set of recurrent points of $f$ (recall Definition 6 from last semester notes). Then $\mu(R(f)) = 1$.

Corollary 29. Let $f : X \to X$ be a continuous map of a compact metric space to itself. Suppose that there exists an $f$-invariant Borel probability measure $\mu$ with the property that for any nonempty open subset $V$ of $X$, $\mu(V) > 0$. Then the set of recurrent points of $f$ is dense in $X$.

Remark 30. The previous Corollary applies in Examples 20, 21, 22, and 24.

Definition 31. Let $(X, B, \mu)$ be a probability space, and let $T : X \to X$ be measure preserving. Let $B \in B$. We say that $T$ is ergodic if and only if whenever $T^{-1}(B) = B$ for some $B \in B$, we have either $\mu(B) = 0$ or $\mu(B) = 1$.

Definition 32. Let $X$ be a set, and let $A$ and $B$ be subsets of $X$. We use the following notation.

$$A \triangle B = (A - B) \cup (B - A)$$

where $A - B$ denotes the set of $x \in A$ such that $x \notin B$.

Theorem 33. Let $(X, B, \mu)$ be a probability space, and let $T : X \to X$ be measure preserving. The following are equivalent:

1. $T$ is ergodic
2. If $B \in B$ and $\mu(T^{-1}(B) \triangle B) = 0$, then $\mu(B) = 0$ or $\mu(B) = 1$.
3. If $A \in B$ with $\mu(A) > 0$, then $\mu(\bigcup_{n=1}^{\infty} T^{-n}(A)) = 1$.
4. If $A, B \in B$ with $\mu(A) > 0$, $\mu(B) > 0$, then there exists a positive integer $n$ with $\mu(T^{-n}(A) \cap B) > 0$.

Definition 34. Let $X$ be a metric space, and let $\mu$ be a Borel probability measure on $X$. Let $\{V_a : a \in A\}$ denote the collection of all open subsets of $X$ which satisfy $\mu(V_a) = 0$. The support of $\mu$ is defined by

$$\text{supp}(\mu) = X - (\bigcup_{a \in A} V_a).$$

Proposition 35. Let $X$ be a compact metric space, and let $\mu$ be a Borel probability measure on $X$. Then

1. $\text{supp}(\mu)$ is a closed set and $\mu(\text{supp}(\mu)) = 1$.
2. $\text{supp}(\mu)$ is the intersection of all closed subsets $K$ of $X$ with $\mu(K) = 1$.

Definition 36. Let $X$ be a metric space, and let $f : X \to X$ be continuous. Let $\mu \in M(X)$. If $f$ is measure preserving with respect to the probability space $(X, B(X), \mu)$, then we say that $\mu$ is $f$- invariant. We will let $M(X, f)$ denote the set of $f$-invariant Borel probability measures. If $\mu \in M(X, f)$ and $f$ is ergodic with respect to the probability space $(X, B(X), \mu)$, then we say that $\mu$ is ergodic. We will let $E(X, f)$ denote the set of $f$-invariant, ergodic Borel probability measures.

Proposition 37. Let $X$ be a compact metric space, and let $f : X \to X$ be continuous. Suppose that there exists $\mu \in E(X, f)$ with $\text{supp}(\mu) = X$. Then

1. $f$ is strongly topologically transitive
2. almost all points of $X$ have a dense forward orbit.
Definition 38. Let \((X, \mathcal{B}, \mu)\) be a probability space. Recall that the terms measurable and integrable may be defined for real valued or complex valued functions. Let \(p \geq 1\). We let \(L^p_\mathbb{R}(\mu)\) denote the set of real valued measurable functions \(f\) such that \(|f|^p\) is integrable. We let \(L^p_\mathbb{C}(\mu)\) denote the set of complex valued measurable functions \(f\) such that \(|f|^p\) is integrable.

Theorem 39. Let \((X, \mathcal{B}, \mu)\) be a probability space, and let \(T : X \to X\) be measure preserving. Let \(p \geq 1\). The following are equivalent:

1. \(T\) is ergodic
2. If \(f : X \to \mathbb{R}\) is measurable and \((f \circ T)(x) = f(x)\) for all \(x \in X\), then \(f\) is constant almost everywhere.
3. If \(f : X \to \mathbb{R}\) is measurable and \((f \circ T)(x) = f(x)\) for almost all \(x \in X\), then \(f\) is constant almost everywhere.
4. If \(f : X \to \mathbb{R}\) is in \(L^p_\mathbb{R}(\mu)\) and \((f \circ T)(x) = f(x)\) for all \(x \in X\), then \(f\) is constant almost everywhere.
5. If \(f : X \to \mathbb{R}\) is in \(L^p_\mathbb{R}(\mu)\) and \((f \circ T)(x) = f(x)\) for almost all \(x \in X\), then \(f\) is constant almost everywhere.

Theorem 40. The previous Theorem holds for complex valued functions and \(L^p_\mathbb{C}(\mu)\).

Theorem 41. Let \((X, \mathcal{B}, \mu)\) be a probability space. Let \(p \geq 1\). Then \(L^p_\mathbb{C}(\mu)\) and \(L^p_\mathbb{R}(\mu)\) are banach spaces with norm given by:

\[||f||_p = \left(\int |f|^p d\mu\right)^{\frac{1}{p}}.\]

Recall that the elements of these spaces are equivalence classes of functions. In the case \(p = 2\) these spaces are Hilbert spaces.

Theorem 42. Let \((S^1, \mathcal{B}, \mu)\) be the probability space given in Example 15. Then \(L^2_\mathbb{C}(\mu)\) is a separable Hilbert space, and the functions

\[f_k(z) = z^k, \quad k \in \mathbb{Z}\]

form a countable, complete orthonormal system. So for each \(f \in L^2_\mathbb{C}(\mu)\) we may write

\[f(z) = \sum_{n=-\infty}^{\infty} b_n z^n\]

in a unique way. Here the convergence is in \(L^2_\mathbb{C}(\mu)\), and we have

\[||f||_2 = \sqrt{\sum_{n=-\infty}^{\infty} |b_n|^2}.\]

This series is called the Fourier series.

Example 43. Let \((S^1, \mathcal{B}, \mu)\) be the probability space given in Example 15. Any irrational rotation of the circle is ergodic. Any rational rotation of the circle is not ergodic.

Example 44. Let \((S^1, \mathcal{B}, \mu)\) be the probability space given in Example 15. The measure preserving transformation \(T\) given by \(T(z) = z^2\) is ergodic.

Theorem 45. Let \((X, \mathcal{B}, \mu)\) be a probability space. Let \(\mathcal{A}\) be an algebra of subsets of \(X\) which generates \(\mathcal{B}\). Then for every \(B \in \mathcal{B}\) and every \(\epsilon > 0\) there exists \(A \in \mathcal{A}\) with \(\mu(A \Delta B) < \epsilon\).

Theorem 46. Let \((X, \mathcal{B}, \mu)\) be a probability space, and suppose that \(T : X \to X\) is measure preserving. Let \(\mathcal{A}\) be an algebra of subsets of \(X\) which generates \(\mathcal{B}\). Suppose that for each \(A \in \mathcal{A}\), there is a positive integer \(N\) such that if \(B = T^{-N}(A)\), then \(\mu(B \cap A) = \mu(B) \cdot \mu(A)\). Then \(T\) is ergodic.
We will denote \( X = (\Sigma, \mu) \) where the sum is taken over all \( Y \).

Let \( k \) be a positive integer. Fix \( (X, \mathcal{B}, \mu) \) in the previous definition, the two-sided shift on \( X \).

Theorem 48. Let \( X \) and \( \mathcal{B} \) be as in the previous definition. There is a unique measure \( \mu \) such that

1. \((X, \mathcal{B}, \mu)\) is a probability space.

2. For any cylinder set

\[ A = \{(\ldots, x_{-1}, x_0, x_1, \ldots) \in X : -n \leq j \leq n \implies x_j \in A_j\}, \]

we have:

\[ \mu(A) = \prod_{i=-n}^{i=n} \mu_i(A_i). \]

Definition 49. We call the probability space \((X, \mathcal{B}, \mu)\) (as in the previous theorem) the product of the probability spaces \((X_i, \mathcal{B}_i, \mu_i)\) and write

\[ (X, \mathcal{B}, \mu) = \prod_{i=-\infty}^{i=\infty} (X_i, \mathcal{B}_i, \mu_i). \]

Definition 50. Similarly, if we start with probability spaces \((X_i, \mathcal{B}_i, \mu_i)\) for each non-negative integer \( i \) we may form a probability space \((X, \mathcal{B}, \mu)\) which we call the product of the probability spaces \((X_i, \mathcal{B}_i, \mu_i)\) and write

\[ (X, \mathcal{B}, \mu) = \prod_{i=0}^{i=\infty} (X_i, \mathcal{B}_i, \mu_i). \]

Definition 51. Fix a positive integer \( k \), and fix a probability vector \((p_1, \ldots, p_k)\) (so \( p_i \geq 0 \) for all \( i \), and \( p_1 + \cdots + p_k = 1 \)). Let \( Y = \{1, \ldots, k\} \), and let \( \mathcal{C} \) denote the collection of all subsets of \( Y \). Define a measure \( m \) on \( \mathcal{C} \) by \( m(A) = \sum p_i \), where the sum is taken over all \( i = 1, \ldots, k \) such that \( i \in A \). Then \((Y, \mathcal{C}, m)\) is a probability space.

For each integer \( i \), let \((X_i, \mathcal{B}_i, \mu_i) = (Y, \mathcal{C}, m)\). We may form the product

\[ (X, \mathcal{B}, \mu) = \prod_{i=-\infty}^{i=\infty} (X_i, \mathcal{B}_i, \mu_i). \]

We will denote \( X \) by \( \Sigma_k \), and refer to \( \mu \) as the \((p_1, \ldots, p_k)\)-product measure. So we may consider the probability space \((\Sigma_k, \mathcal{B}, \mu)\).

Theorem 52. With respect to the probability space in the previous definition, the two-sided shift on \( k \) symbols is an invertible measure preserving transformation.

Definition 53. Fix a positive integer \( k \), and fix a probability vector \((p_1, \ldots, p_k)\) (so \( p_i \geq 0 \) for all \( i \), and \( p_1 + \cdots + p_k = 1 \)). Let \( Y = \{1, \ldots, k\} \), and let \( \mathcal{C} \) denote the collection of all subsets of \( Y \). Define a measure \( m \) on \( \mathcal{C} \) by \( m(A) = \sum p_i \), where the sum is taken over all \( i = 1, \ldots, k \) such that \( i \in A \). Then \((Y, \mathcal{C}, m)\) is a probability space.

For each non-negative integer \( i \), let \((X_i, \mathcal{B}_i, \mu_i) = (Y, \mathcal{C}, m)\). We may form the product

\[ (X, \mathcal{B}, \mu) = \prod_{i=0}^{i=\infty} (X_i, \mathcal{B}_i, \mu_i). \]

We will denote \( X \) by \( \Sigma_k^+ \), and refer to \( \mu \) as the \((p_1, \ldots, p_k)\)-product measure. So we may consider the probability space \((\Sigma_k^+, \mathcal{B}, \mu)\).
Theorem 54. With respect to the probability space in the previous definition, the one-sided shift on $k$ symbols is a measure preserving transformation.

Lemma 55. Consider the probability space $(\Sigma_k^+, B, \mu)$, and the one-sided shift $T$. Let $A$ be a cylinder set given by

$$A = \{(x_0, x_1, \ldots) \in X : 0 \leq j \leq n \Rightarrow x_j \in A_j\}.$$  

Let $B$ be a cylinder set given by

$$B = \{(x_0, x_1, \ldots) \in X : 0 \leq j \leq n \Rightarrow x_j \in B_j\}.$$  

If $s \geq n + 1$, then

$$\mu(T^{-s}(A) \cap B) = \mu(A) \cdot \mu(B).$$

Lemma 56. Consider the probability space $(\Sigma_k, B, \mu)$, and the two-sided shift $T$. Let $A$ be a cylinder set given by

$$A = \{(\ldots, x_{-1}, x_0, x_1, \ldots) \in X : -n \leq j \leq n \Rightarrow x_j \in A_j\}.$$  

Let $B$ be a cylinder set given by

$$B = \{(\ldots, x_{-1}, x_0, x_1, \ldots) \in X : -n \leq j \leq n \Rightarrow x_j \in B_j\}.$$  

If $s \geq 2n + 1$, then

$$\mu(T^{-s}(A) \cap B) = \mu(A) \cdot \mu(B).$$

Lemma 57. Let $(X, B, \mu)$ be a probability space. Suppose that $A = C_1 \cup \cdots \cup C_n$, a pairwise disjoint union of elements of $B$. Suppose that $B = D_1 \cup \cdots \cup D_n$, a pairwise disjoint union of elements of $B$. Suppose that for all $i, j$ we have $\mu(C_i \cap D_j) = \mu(C_i) \cdot \mu(D_j)$. Then $\mu(A \cap B) = \mu(A) \cdot \mu(B)$.

Theorem 58. With respect to the probability space in Definition 53, the one-sided shift on $k$ symbols is ergodic.

Theorem 59. With respect to the probability space in Definition 51, the two-sided shift on $k$ symbols is ergodic.

Definition 60. Let $(X_1, B_1, \mu_1)$ and $(X_2, B_2, \mu_2)$ be probability spaces, and suppose that $T_1 : X_1 \to X_1$, $T_2 : X_2 \to X_2$ are measure preserving transformations. We say that $T_1$ is isomorphic to $T_2$ if and only if there exist $M_1 \in B_1$, $M_2 \in B_2$ and an invertible measure-preserving transformation $\phi : M_1 \to M_2$ such that

1. $\mu_1(M_1) = \mu_2(M_2) = 1$.
2. $T_1(M_1) \subset M_1$, and $T_2(M_2) \subset M_2$.
3. $\phi(T_1(x)) = T_2(\phi(x))$ for all $x \in M_1$.

Example 61. Let $T_1 : [0, 1] \to [0, 1]$ be given by $T_1(x) = 2x \mod 1$. Let $T_2 : S^1 \to S^1$ be given by $T_2(z) = z^2$. Then $T_1$ is isomorphic to $T_2$.

Proposition 62. Let $(X_1, B_1, \mu_1)$ and $(X_2, B_2, \mu_2)$ be probability spaces, and suppose that $T_1 : X_1 \to X_1$, $T_2 : X_2 \to X_2$ are measure preserving transformations. Suppose that $T_1$ is isomorphic to $T_2$. If $T_1$ is ergodic, then $T_2$ is ergodic.

Theorem 63. (Birkhoff Ergodic Theorem) Suppose that $(X, B, \mu)$ is a probability space, and $T : X \to X$ is measure preserving. Let $f \in L^1(\mu)$. Then the sequence

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

converges almost everywhere to a function $f^* \in L^1(\mu)$. Also, $f^* \circ T = f^*$ almost everywhere and $\int f^* d\mu = \int f d\mu$.

We will return to the proof of this theorem, but we first look at some corollaries.
Corollary 64. (sometimes called the Ergodic Theorem) Suppose that \((X, \mathcal{B}, \mu)\) is a probability space, and \(T : X \to X\) is measure preserving. Then \(T\) is ergodic if and only if for all \(f \in L^1(\mu)\),
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x)) = \int f \, d\mu
\]
for almost all \(x \in X\).

Corollary 65. Suppose that \((X, \mathcal{B}, \mu)\) is a probability space, and \(T : X \to X\) is measure preserving and ergodic. Let \(E \in \mathcal{B}\). For each positive integer \(n\) and each \(x \in X\), let \(K_n(x)\) denote the cardinality of \(\{x, T(x), \ldots, T^{n-1}(x)\} \cap E\). Then for almost all \(x \in X\) we have
\[
\lim_{n \to \infty} \frac{K_n(x)}{n} = \mu(E).
\]

Corollary 66. (Borel’s Theorem on normal numbers) For almost all \(x \in [0,1)\), the frequency of ones in the binary expansion of \(x\) is \(\frac{1}{2}\).

Corollary 67. Suppose that \((X, \mathcal{B}, \mu)\) is a probability space, and \(T : X \to X\) is measure preserving. Then \(T\) is ergodic if and only if for all \(A, B \in \mathcal{B}\),
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) = \mu(A)\mu(B).
\]

Our goal now is to prove the Birkhoff Ergodic Theorem.

Definition 68. Let \(L^0(X, \mathcal{B}, \mu)\) denote the vector space of measurable complex-valued functions. Let \(L^0_{\mathbb{R}}(X, \mathcal{B}, \mu)\) denote the vector space of measurable real-valued functions.

Definition 69. Let \((X_1, \mathcal{B}_1, \mu_1)\) and \((X_2, \mathcal{B}_2, \mu_2)\) be probability spaces, and suppose that \(T_1 : X_1 \to X_2\) is measure preserving. The induced operator \(U^0 : L^0(X_2, \mathcal{B}_2, \mu_2) \to L^0(X_1, \mathcal{B}_1, \mu_1)\) is defined by \((U^0(f))(x) = f(T(x))\) for \(x \in X_1\).

Proposition 70. 1. \(U^0\) is linear.
2. \(U^0(L^0_{\mathbb{R}}(X_2, \mathcal{B}_2, \mu_2)) \subseteq L^0_{\mathbb{R}}(X_1, \mathcal{B}_1, \mu_1)\).
3. \(U^0(f \cdot g) = (U^0(f)) \cdot (U^0(g))\).
4. \(U^0(c) = c\), if \(c\) is a constant function.
5. If \(f \geq 0\), then \(U^0(f) \geq 0\).
6. \(U^0(K_B) = K_{T^{-1}(B)}\) for all \(B \in \mathcal{B}\).

Definitions. \(K_B(x) = 1\) if \(x \in B\), and \(K_B(x) = 0\) if \(x \notin B\).

Proposition 71. Let \((X_1, \mathcal{B}_1, \mu_1)\) and \((X_2, \mathcal{B}_2, \mu_2)\) be probability spaces, and suppose that \(T_1 : X_1 \to X_2\) is measure preserving. If \(F \in L^0(X_2, \mathcal{B}_2, \mu_2)\), then
\[
\int U^0(F) \, d\mu_1 = \int F \, d\mu_2.
\]

Proposition 72. Let \((X_1, \mathcal{B}_1, \mu_1)\) and \((X_2, \mathcal{B}_2, \mu_2)\) be probability spaces, and suppose that \(T_1 : X_1 \to X_2\) is measure preserving. Let \(p \geq 1\). Then \(U^0(L^p(X_2, \mathcal{B}_2, \mu_2)) \subseteq L^p(X_1, \mathcal{B}_1, \mu_1)\), and \(\|U^0(f)\|_p = \|f\|_p\), for all \(f \in L^p(X_2, \mathcal{B}_2, \mu_2)\).

Theorem 73. (Maximal Ergodic Theorem) Suppose that \((X, \mathcal{B}, \mu)\) is a measure space, and let \(U : L^1_{\mathbb{R}}(X, \mathcal{B}, \mu) \to L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)\) be a positive linear operator with \(\|U\| \leq 1\). Let \(N\) be a positive integer, and let \(f \in L^1_{\mathbb{R}}(X, \mathcal{B}, \mu)\). Set \(f_0 = 0\), and for \(n \geq 1\) set \(f_n = f + U(f) + U^2(f) + \cdots + U^{n-1}(f)\). Set \(F_N = \max_{0 \leq n \leq N} f_n\), and set \(A = \{x \in X : F_N(x) > 0\}\). Then \(\int_A f \, d\mu \geq 0\).
Corollary 74. Suppose that \((X, \mathcal{B}, \mu)\) is a probability space, and \(T : X \to X\) is measure preserving. Let \(g \in L^1(\mathbb{R}, \mathcal{B}, \mu)\), and let \(\alpha \in \mathbb{R}\). Set

\[
B_\alpha = \{ x \in X : \sup_{n \geq 1} \frac{1}{n} \sum_{i=0}^{n-1} g(T^i(x)) > \alpha \}.
\]

Suppose that \(A \in \mathcal{B}\) and \(T^{-1}(A) = A\). Then \(\int_{B_\alpha \cap A} g \, d\mu \geq \alpha \cdot \mu(B_\alpha \cap A)\).

We now prove the Birkhoff Ergodic Theorem. We will also prove the following:

Theorem 75. \((L^p \text{ Ergodic Theorem of Von Neumann})\) Suppose that \((X, \mathcal{B}, \mu)\) is a probability space, and \(T : X \to X\) is measure preserving. Let \(1 \leq p < \infty\), and let \(f \in L^p(\mu)\). Then for each \(n \geq 1\), the function \(\frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i\) is in \(L^p(\mu)\). Also the function \(f^*\) given by \(f^*(x) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(T^i(x))\) is in \(L^p(\mu)\). Finally,

\[
\lim_{n \to \infty} ||f^* - \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i||_p = 0.
\]

Definition 76. Suppose that \((X, \mathcal{B}, \mu)\) is a probability space, and \(T : X \to X\) is measure preserving. We say that \(T\) is weak-mixing if and only if for all \(A, B \in \mathcal{B}\),

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} |\mu(T^{-1}(A) \cap B) - \mu(A)\mu(B)| = 0.
\]

We say that \(T\) is strong-mixing if and only if for all \(A, B \in \mathcal{B}\),

\[
\lim_{n \to \infty} \mu(T^{-n}(A) \cap B) = \mu(A)\mu(B).
\]

Theorem 77. Suppose that \((X, \mathcal{B}, \mu)\) is a probability space, and \(T : X \to X\) is measure preserving. If \(T\) is strong-mixing, then \(T\) weak-mixing. If \(T\) is weak-mixing, then \(T\) ergodic.

Definition 78. Let \(n\) be a positive integer, and let \(T : \Sigma_n \to \Sigma_n\) denote the two-sided shift map with symbols \(1, \ldots, n\) as in Definition 51. Let \(P = (p_{i,j})\) be a stochastic \(n \times n\) matrix. This means that each entry of the matrix is nonnegative and each row sum is one. Let \(\vec{p} = (p_1, \ldots, p_n)\) be a probability vector such that \(\vec{p}P = \vec{p}\) (i.e., a left eigenvector corresponding to the eigenvalue one).

We say that a subset \(B\) of \(\Sigma_n\) is an elementary rectangle if and only if there are symbols \(a_1, \ldots, a_k\) and an integer \(q\) such that \(B\) consists of all doubly infinite sequences \((x_i)\) such that \(x_q = a_1, \ldots, x_{q+k-1} = a_k\). We define a function \(\tau\) on the set of elementary rectangles by

\[
\tau(B) = p_{a_1} \cdot P_{a_1,a_2} \cdots P_{a_{k-1},a_k}
\]

where \(B\) is given as above.

Theorem 79. Let \(\Sigma_n\) and \(\mathcal{B}\) be as in Definition 51. There is a unique measure \(\mu\) which extends the function \(\tau\) given in the previous definition, such that \((\Sigma_n, \mathcal{B}, \mu)\) is a probability space. This measure is called the \((\vec{p}, P)\)-Markov measure.

Theorem 80. The two-sided shift map \(T : \Sigma_n \to \Sigma_n\) is an invertible, measure preserving transformation with respect to the probability space \((\Sigma_n, \mathcal{B}, \mu)\) where \(\mu\) is the \((\vec{p}, P)\)-Markov measure. Moreover, if \(A\) is the \(n \times n\) matrix given by \(A_{i,j} = 1\) if \(P_{i,j} > 0\), and \(A_{i,j} = 0\) if \(P_{i,j} = 0\), and \(\Sigma_A\) is the corresponding set of \(A\)-allowable sequences, then \(\mu(\Sigma_A) = 1\).

Proposition 81. If \(P\) is a stochastic matrix then for any positive integer \(k\), \(P^k\) is a stochastic matrix.

Proposition 82. Let \(\mu\) be the \((\vec{p}, P)\)-Markov measure as above. Let \(q\) be an integer, and let \(k\) be a positive integer. Let \(B\) consist of all doubly infinite sequences \((x_i)\) with \(x_q = a_1\) and \(x_{q+k} = a_2\). Then

\[
\mu(B) = p_{a_1}(P^k)_{a_1,a_2}
\]

where \((P^k)_{a_1,a_2}\) denotes the entry of \(P^k\) in row \(a_1\) and column \(a_2\).
Remark 83. When using the \((\vec{p},P)\)-Markov measure as above, we often assume that if \(\vec{p} = (p_1, \ldots, p_n)\) then \(p_i > 0\) for all \(i\).

Theorem 84. Let \(P = (p_{i,j})\) be a stochastic \(n \times n\) matrix, and let \(\vec{p} = (p_1, \ldots, p_n)\) be a probability vector such that \(p_i > 0\) for all \(i\) and \(\vec{p}P = \vec{p}\). Then \(\lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} P^k\) exists.

Proposition 85. Let \(P = (p_{i,j})\) be a stochastic \(n \times n\) matrix, and let \(\vec{p} = (p_1, \ldots, p_n)\) be a probability vector such that \(p_i > 0\) for all \(i\) and \(\vec{p}P = \vec{p}\). Let \(Q = \lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} P^k\).

1. If \(\vec{x}P = \vec{x}\), then \(\vec{x}Q = \vec{x}\).
2. \(Q\) is stochastic.
3. \(PQ = QP = Q\).
4. \(Q^2 = Q\).
5. If \(\vec{x}Q = \vec{x}\), then \(\vec{x}P = \vec{x}\).

Theorem 86. Suppose that \((X, \mathcal{B}, \mu)\) is a probability space, and \(T : X \to X\) is measure preserving. Let \(S\) be a semi-algebra which generates \(\mathcal{B}\). Then \(T\) is ergodic if and only if for all \(A, B \in S\),

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}(A) \cap B) = \mu(A)\mu(B).
\]

Definition 87. Let \(P\) be a non-negative \(n \times n\) matrix. We say that \(P\) is irreducible if and only if for each pair \(i, j\) of positive integers with \(1 \leq i, j \leq n\) there is a positive integer \(k\) with \((P^k)_{i,j} > 0\).

Theorem 88. Let \(P = (p_{i,j})\) be a stochastic \(n \times n\) matrix, and let \(\vec{p} = (p_1, \ldots, p_n)\) be a probability vector such that \(p_i > 0\) for all \(i\) and \(\vec{p}P = \vec{p}\). Let \(Q = \lim_{N \to \infty} \frac{1}{N} \sum_{k=0}^{N-1} P^k\). Let \(T\) denote the shift transformation with the \((\vec{p},P)\)-Markov measure (sometimes called the \((\vec{p},P)\)-Markov shift). The following are equivalent.

1. \(T\) is ergodic.
2. Each row of \(Q\) is \(\vec{p}\).
3. All rows of \(Q\) are identical.
4. Each entry of \(Q\) is positive.
5. \(P\) is irreducible.
6. 1 is a simple eigenvalue of \(P\) (i.e. the eigenspace is one-dimensional).

Definition 89. We will assume that \(X\) is a compact metric space. Also, recall the following. Let \(\mathcal{B}(X)\) denote the smallest \(\sigma\)-algebra containing each open subset of \(X\). We call \(\mathcal{B}(X)\) the collection of Borel sets. If \((X, \mathcal{B}(X), \mu)\) is a probability space, we call \(\mu\) a Borel probability measure. We will let \(M(X)\) denote the set of Borel probability measures.

Theorem 90. Let \(\mu \in M(X)\). For all \(B \in \mathcal{B}(X)\) and all \(\epsilon > 0\), there is an open set \(V_\epsilon\) and a closed set \(C_\epsilon\) with \(C_\epsilon \subset B \subset V_\epsilon\) and \(\mu(V_\epsilon - C_\epsilon) < \epsilon\). (A measure with this property is called regular.)

Corollary 91. Let \(\mu \in M(X)\), and let \(B \in \mathcal{B}(X)\). Then

\[
\mu(B) = \sup \mu(C),
\]

where the supremum is taken over all closed sets \(C\) with \(C \subset B\). Also,

\[
\mu(B) = \inf \mu(V),
\]

where the infimum is taken over all open sets \(V\) with \(B \subset V\).
Definition 92. (and remark) Let $C(X)$ denote the space of (bounded) continuous real valued functions on $X$. Recall that $C(X)$ is a Banach space (a complete normed linear space) with norm given by

$$||f|| = \sup_{x \in X} |f(x)|.$$  

Theorem 93. Let $\mu_1, \mu_2 \in M(X)$. If $\int f d\mu_1 = \int f d\mu_2$ for all $f \in C(X)$, then $\mu_1 = \mu_2$.

Definition 94. (and remark) Let $\mu \in M(X)$. Define a function $J_\mu : C(X) \to \mathbb{R}$ by $J_\mu(f) = \int f d\mu$. Observe that

1. $J_\mu$ is linear.
2. If $f \geq 0$, then $J_\mu(f) \geq 0$.
3. If $f(x) = 1$ for all $x \in X$, then $J_\mu(f) = 1$.
4. The function $\mu \to J_\mu$ is injective.

Definition 95. A linear map $L : C(X) \to \mathbb{R}$ is called a linear functional on $C(X)$. We say that a linear functional $L$ on $C(X)$ is bounded if and only if there exists a positive real number $M$ such that for all $f \in C(X)$ we have

$$|L(f)| \leq M \cdot ||f||.$$  

Proposition 96. Let $\mu \in M(X)$. Then $J_\mu : C(X) \to \mathbb{R}$ is a bounded linear functional on $C(X)$.

Definition 97. (and remark) Let $Y$ be a normed linear space, and let $Y^*$ denote the set of bounded linear functionals on $Y$. We define a norm on $Y^*$ by

$$||L|| = \sup_{x \in Y, x \neq 0} \frac{|L(x)|}{||x||}.$$  

With this norm $Y^*$ is a Banach space. This norm determines a topology on $Y^*$ called the strong topology.

Now given a a collection of $\mathcal{F}$ of linear functionals on $Y^*$ (so $\mathcal{F} \subset (Y^*)^*$), there is a smallest topology on $Y^*$ such that each $f \in \mathcal{F}$ is continuous. If we take $\mathcal{F} = (Y^*)^*$, this determines a topology on $Y^*$ called the weak topology.

Finally, for each $y \in Y$, we obtain an element $\phi_y$ of $(Y^*)^*$ by setting $\phi_y(L) = L(y)$. Let $\mathcal{F} = \{\phi_y : y \in Y\}$. Then $\mathcal{F}$ determines a topology on $Y^*$ called the weak* topology on $Y^*$. In general, the weak* topology is weaker (has fewer open sets) than the weak topology.

Definition 98. Let $L$ be a bounded linear functional on $C(X)$. We say that $L$ is positive if and only if $f \geq 0$ implies that $L(f) \geq 0$. We say that $L$ is normalized if and only if $L(1) = 1$, where the first 1 denotes the constant function 1. Let $NPLF(X)$ denote the subset of $(C(X))^*$ which consists of all bounded, positive, normalized linear functionals on $C(X)$.

Theorem 99. The function $G : M(X) \to NPLF(X)$ defined by $G(\mu) = J_\mu$ is a bijection.

Proof. Surjectivity follows from one form of the Riesz representation theorem.

Definition 100. Let $T_1$ denote the weak* topology on $(C(X))^*$. Let $T_2$ denote the relative topology on $NPLF(X)$ obtained from $T_1$. Let $T_3$ denote the collection of subsets of $M(X)$ of the form $G^{-1}(V)$ where $V \in T_2$. Then $T_3$ is a topology on $M(X)$. We call this topology the weak* topology on $M(X)$.

Theorem 101. Given $f \in C(X)$ we define a function $H_f : M(X) \to \mathbb{R}$ by $H_f(\mu) = \int f d\mu$. The weak* topology on $M(X)$ is the smallest topology on $M(X)$ such that each function $H_f$ is continuous.

Theorem 102. The space $C(X)$ is separable. (There exists a countable, dense subset.)

Corollary 103. There exists a countable, dense subset $F$ of $C(X)$ such that the constant function zero is not in $F$.

Theorem 104. The space $M(X)$ with the weak* topology is metrizable.
Theorem 105. Let $Y$ be a normed linear space, and let $Y^*$ denote the set of bounded linear functionals on $Y$. Then the unit sphere
\[ \{ L \in Y^* : ||L|| \leq 1 \} \]
is compact with the weak* topology. Here the norm is as defined in Definition 97.


Theorem 106. The space $M(X)$ with the weak* topology is compact.

Proposition 107. (and notation) Let $T : X \to X$ be continuous. There is a well-defined function $\hat{T} : M(X) \to M(X)$ given by
\[ (\hat{T}(\mu))(B) = \mu(T^{-1}(B)) \]
for all $B \in \mathcal{B}(X)$.

Proposition 108. Let $\mu \in M(X)$, and let $T : X \to X$ be continuous. Then $\int f d\hat{T}(\mu) = \int (f \circ T) d\mu$ for all $f \in C(X)$.

Proposition 109. Let $\mu \in M(X)$, and let $T : X \to X$ be continuous. The following are equivalent.
1. $T$ is a measure preserving transformation of the probability space $(X, \mathcal{B}(X), \mu)$.
2. $\hat{T}(\mu) = \mu$.
3. $\int (f \circ T) d\mu = \int f d\mu$ for all $f \in C(X)$.

Definition 110. Let $T : X \to X$ be continuous. We let $M(X, T)$ denote the set of all $\mu \in M(X)$ such that $T$ is a measure preserving transformation of the probability space $(X, \mathcal{B}(X), \mu)$. The elements of $M(X, T)$ are called $T$-invariant measures.

Theorem 111. Let $(\mu_n)$ be a sequence in $M(X)$ with the weak* topology, and let $\mu \in M(X)$. Then $\mu_n \to \mu$ if and only if for each $f \in C(X)$,
\[ \int f d\mu_n \to \int f d\mu. \]

Proof. This is a homework problem.

Theorem 112. Let $T : X \to X$ be continuous. Then $M(X, T)$ is a closed subset of $M(X)$ (with the weak* topology). Hence, $M(X, T)$ is compact.

Theorem 113. Let $T : X \to X$ be continuous. Let $(\sigma_n)$ be a sequence in $M(X)$. Define a new sequence $(\mu_n)$ in $M(X)$ by
\[ \mu_n = \frac{1}{n} \sum_{i=0}^{n-1} \hat{T}^i(\sigma_n). \]

Then any subsequential limit of the sequence $(\mu_n)$ is an element of $M(X, T)$. In particular, $M(X, T) \neq \emptyset$.

Proposition 114. Let $\mu_1, \mu_2 \in M(X)$, and let $\lambda \in [0, 1]$. Then there is an element $\mu_3$ of $M(X)$ defined by
\[ \mu_3 = \lambda \mu_1 + (1 - \lambda) \mu_2. \]

Definition 115. Let $K$ be a subset of $M(X)$. We say that $K$ is convex if and only if whenever $\mu_1, \mu_2 \in K$ and $\lambda \in [0, 1]$ we have $(\lambda \mu_1 + (1 - \lambda) \mu_2) \in K$.

Proposition 116. If $T : X \to X$ is continuous, then $M(X, T)$ is a convex subset of $M(X)$.

Definition 117. Let $T : X \to X$ be continuous. An element $\mu$ of $M(X, T)$ is called an extreme point of $M(X, T)$ if and only if whenever $\mu = \lambda \mu_1 + (1 - \lambda) \mu_2$ with $\lambda \in [0, 1]$ and $\mu_1, \mu_2 \in M(X, T)$, we have $\mu = \mu_1$ or $\mu = \mu_2$.

Theorem 118. Let $T : X \to X$ be continuous. Suppose that $\mu \in M(X, T)$, and $\mu$ is an extreme point of $M(X, T)$. Then $T$ is an ergodic measure preserving transformation of the probability space $(X, \mathcal{B}(X), \mu)$. 
**Definition 119.** Let $T : X \to X$ be continuous. Suppose that $\mu \in M(X, T)$, and $T$ is an ergodic measure preserving transformation of the probability space $(X, \mathcal{B}(X), \mu)$. Then we say that $\mu$ is ergodic.

**Definition 120.** Let $(X, \mathcal{B}, \mu)$ and $(X, \mathcal{B}, \mu_1)$ be a measure spaces. We say that $\mu_1$ is absolutely continuous with respect to $\mu$ if and only if $\mu(A) = 0$ implies that $\mu_1(A) = 0$.

**Theorem 121.** (Radon-Nikodym) Let $(X, \mathcal{B}, \mu)$ be a $\sigma$-finite measure spaces. Let $\mu_1$ be a measure which is absolutely continuous with respect to $\mu$. Then there is a nonnegative measurable function $f$ such that for all $E \in \mathcal{B}$ we have

$$\mu_1(E) = \int_E f \, d\mu.$$ 

The function $f$ is unique, in the sense that if $g$ has the same properties, then $g = f$ almost everywhere.

**Theorem 122.** Let $T : X \to X$ be continuous. Suppose that $\mu \in M(X, T)$, and $\mu_1 \in M(X, T)$. Suppose that $\mu_1$ is absolutely continuous with respect to $\mu$. If $\mu$ is ergodic, then $\mu_1 = \mu$.

**Theorem 123.** Let $T : X \to X$ be continuous. Suppose that $\mu \in M(X, T)$. Then $\mu$ is an extreme point of $M(X, T)$ if and only if $\mu$ is ergodic.

**Theorem 124.** (Krein-Milman) Let $X$ be a locally convex topological vector space. Let $K$ be a compact convex subset of $X$. Then $K$ is the closed convex hull of its extreme points. In particular, if $K \neq \emptyset$, then $K$ has extreme points.

**Theorem 125.** Let $Y$ be a normed linear space, and let $Y^*$ denote the set of bounded linear functionals on $Y$. Then $Y^*$ with the weak* topology is a locally convex topological vector space.

**Theorem 126.** Let $T : X \to X$ be continuous. Then there exists a $T$-invariant ergodic measure.

**Definition 127.** Let $(X, \mathcal{B}, \mu_1)$ and $(X, \mathcal{B}, \mu_2)$ be a measure spaces. We say that $\mu_1$ and $\mu_2$ are mutually singular if and only if there exist disjoint sets $A, B \in \mathcal{B}$ with $X = A \cup B$ and

$$\mu_1(A) = 0 = \mu_2(B).$$

**Theorem 128.** (Lebesgue decomposition) Let $(X, \mathcal{B}, \mu_1)$ and $(X, \mathcal{B}, \mu_2)$ be a $\sigma$-finite measure spaces. Then there exists a measure $\mu_3$ on $(X, \mathcal{B})$ which is singular with respect to $\mu_1$ and a measure $\mu_4$ on $(X, \mathcal{B})$ which is absolutely continuous with respect to $\mu_1$ such that $\mu_2 = \mu_3 + \mu_4$. The measures $\mu_3$ and $\mu_4$ are unique.

**Corollary 129.** Let $(X, \mathcal{B}, \mu_1)$ and $(X, \mathcal{B}, \mu_2)$ be probability spaces. Then there exists $p \in [0, 1]$ and probability measures $\mu_5$ and $\mu_6$ on $(X, \mathcal{B})$ such that

1. $\mu_5$ is singular with respect to $\mu_1$.
2. $\mu_6$ is absolutely continuous with respect to $\mu_1$.
3. $\mu_2 = p\mu_5 + (1 - p)\mu_6$.

The measures $\mu_5$ and $\mu_6$ and the number $p$ are unique.

**Remark 130.** Recall that we are assuming in these notes in all items 89 and above (except items 97, 105, 121, 124, 125, 127, 128, and 129) that $X$ is a compact metric space.

**Example.** Let $T : X \to X$ be continuous. Suppose $\mu_1, \mu_2 \in M(X)$ are mutually singular. Then it need not be the case that $\hat{T}(\mu_1)$ and $\hat{T}(\mu_2)$ are mutually singular.

**Proposition 131.** Let $T : X \to X$ be a homeomorphism. Suppose $\mu_1, \mu_2 \in M(X)$ are mutually singular. Then $\hat{T}(\mu_1)$ and $\hat{T}(\mu_2)$ are mutually singular.

**Theorem 132.** Let $T : X \to X$ be a homeomorphism. Suppose $\mu_1, \mu_2 \in M(X, T)$ are ergodic and $\mu_1 \neq \mu_2$. Then $\mu_1$ and $\mu_2$ are mutually singular.
Theorem 133. Let $X$ and $Y$ be compact metric spaces, and let $f : X \to X$ and $g : Y \to Y$ be continuous. Suppose that $f$ and $g$ are topologically conjugate. Then there exists a function $H : M(X,f) \to M(Y,g)$ such that

1. $H$ is a homeomorphism,
2. $H$ is affine,
3. $\mu \in M(X,f)$ is ergodic if and only if $H(\mu) \in M(Y,g)$ is ergodic.

Definition 134. Let $f : X \to X$ be continuous. We say that $f$ is uniquely ergodic if and only if $M(X,f)$ has only one element.

Definition 135. (and Remark) Let $\alpha = (p_1, p_2, \ldots)$ be a sequence of integers where each $p_i \geq 2$. Let $\Delta_\alpha$ denote the set of all sequences $(x_1, x_2, \ldots)$ where $x_i \in \{0, 1, \ldots, p_i - 1\}$ for each $i$. We use the product topology on $\Delta_\alpha$. Observe that $\Delta_\alpha$ is a compact, metrizable space.

Addition in $\Delta_\alpha$ is defined as follows. We set

$$(x_1, x_2, \ldots) + (y_1, y_2, \ldots) = (z_1, z_2, \ldots)$$

where $z_1 = x_1 + y_1 \mod p_1$, $z_2 = x_2 + y_2 + t_1 \mod p_2$, etc. Here $t_1 = 0$ if $x_1 + y_1 < p_1$ and $t_1 = 1$ if $x_1 + y_1 \geq p_1$. So, we carry a one in the second case. Continue adding and carrying in this way for the whole sequence.

With this addition $\Delta_\alpha$ is a topological group known as the $\alpha$-adic adding machine.

We define $f_\alpha : \Delta_\alpha \to \Delta_\alpha$ by

$$f_\alpha(x_1, x_2, \ldots) = (x_1, x_2, \ldots) + (1, 0, 0, \ldots).$$

We will refer to the map $f_\alpha : \Delta_\alpha \to \Delta_\alpha$ as the adding machine map. This map is also sometimes called the odometer map.

Theorem 136. Let $\alpha = (p_1, p_2, \ldots)$ be a sequence of integers with $p_i \geq 2$ for each $i$. Let $j_i = p_1 \cdot p_2 \cdot \ldots \cdot p_i$ for each $i$. Let $f : X \to X$ be a continuous map of a compact metric space $X$. Then $f$ is topologically conjugate to $f_\alpha$ if and only if (1), (2), and (3) hold.

1. For each positive integer $i$, there is a cover $P_i$ of $X$ consisting of $j_i$ pairwise disjoint, nonempty, clopen sets which are cyclically permuted by $f$.
2. For each positive integer $i$, $P_{i+1}$ is a refinement of $P_i$.
3. If $\text{mesh}(P_i)$ denotes the maximum diameter of an element of the cover $P_i$, then $\text{mesh}(P_i) \to 0$ as $i \to \infty$.

Corollary 137. Let $\alpha = (p_1, p_2, \ldots)$ be a sequence of integers with $p_i \geq 2$ for each $i$. Then $\Delta_\alpha$ is a minimal set for $f_\alpha$.

Theorem 138. Let $\alpha = (p_1, p_2, \ldots)$ be a sequence of integers with $p_i \geq 2$ for each $i$. Then the adding machine map $f_\alpha$ is uniquely ergodic.