Today I want to discuss some nice consequences of the Pointwise Ergodic theorem:

**Theorem 1.** Let \((X, \mathcal{F}, \mu)\) be a probability space and \(T : X \rightarrow X\) a measurable, measure-preserving transformation. Also, let \(f \in L^1\) be an integrable function on \(X\). Then the quantity
\[
\lim_{n \to \infty} \frac{f + f \circ T + \ldots + f \circ T^{n-1}}{n}
\]
exists a.e. and defines a \(T\)-invariant function \(f^*\), for which \(\int_X f^* = \int f\). If \(T\) is ergodic, then \(f^*\) is constant a.e..

1 Recurrence

Recall the statement of the Poincare Recurrence Theorem, that if \(X\) is a probability space, \(A \subset X\) has positive measure, and \(T : X \rightarrow X\) is measure-preserving, then for almost every \(x \in A\) we have \(T^n(x) \in A\) for infinitely many \(n\).

A natural question is, how often will \(T^n x\) land in \(A\)? If \(T\) is ergodic, we can get a quick answer. Applying the ergodic theorem to the indicator function \(1_A\), we have
\[
\lim_{n \to \infty} \frac{\sum_{k=0}^{n-1} 1_A(T^k x)}{n} = \int_X 1_A = \mu(A)
\]
for almost every \(x\).

In other words, not only does almost every \(x \in A\) return to \(A\) infinitely often, but in fact almost every \(x \in X\) lands in \(A\) infinitely often, and we can actually get a precise asymptotic for how often it happens.

We can ask a similar problem – given a generic point \(x \in A\), how often do we have to wait before it lands in \(A\) again? In other words, if we let \(n(x) = \inf \{ n | T^n x \in A \text{ and } \forall k < n, T^k x \notin A \}\) be the first return time of \(x\), then how can we quantify the value of \(n(x)\)? One should expect that as \(\mu(A)\) gets smaller, the return times should get larger, and vice versa. Indeed, we have
\[
\int_A n(x) = 1.
\]

Rephrasing the above, we expect the following statement:
Lemma 1. Let \((X, \mathcal{F}, \mu)\) be a probability space and \(T\) a measure-preserving and ergodic transformation. In addition, let \(A \subset X\) is measurable and have positive measure. Then if \(n(x)\) denotes the first return time map we have
\[
\int_A n(x) d\mu = 1.
\]

Proof. Let \(A_k = \{x \in A : n(x) = k\}\) and \(A_k^*\) be the set of elements of \(X \setminus A\) that enter \(A\) for the first time at time \(k\).

As \(A\) has positive measure, then as \(T\) is ergodic we have \(\mu(\bigcup_{n=1}^{\infty} T^{-n}A) = 1\), so almost every element of \(X\) is in one of \(A_k\) or \(A_k^*\). Therefore we can write
\[
1 = \mu(X) = \sum_{i=1}^{n} \mu(A_i) + \mu(A_i^*).
\]

In addition, observe that we have \(T^{-1}A_k^* = A_{k+1} \cup A_{k+1}^*\). As \(T\) is measure-preserving and \(A_{k+1}, A_{k+1}^*\) are disjoint this implies that \(\mu(A_k) = \mu(A_{k+1}) + \mu(A_{k+1}^*)\). Applying this relation repeatedly we get that
\[
\mu(A_n) = \mu(A_m^*) + \sum_{i=n+1}^{m} \mu(A_i).
\]

Taking the limit as \(m \to \infty\) (justified by the fact that the sum is monotonic and bounded by 1), we get
\[
\mu(A_n^*) = \lim_{m \to \infty} \mu(A_m^*) + \sum_{i=n+1}^{\infty} \mu(A_i) = \sum_{i=n+1}^{\infty} \mu(A_i).
\]

Now we compute. Putting together our observations we have
\[
1 = \mu(X)
\]
\[
= \sum_{i=1}^{\infty} \mu(A_i^*) + \mu(A_i)
\]
\[
= \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} \mu(A_k)
\]
\[
= \sum_{k=1}^{\infty} k \mu(A_k)
\]
\[
= \int_A n(x)
\]

In particular, because the induced map \(T_A\) is ergodic, then applying the ergodic theorem we can conclude that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} n(T_A^i(x)) = \frac{1}{\mu(A)}
\]
for almost every \(x \in X\).
2 Characterizations of Ergodicity

There are about 104024010 different characterizations of ergodicity, and the ergodic theorem allows us to give another one. We will later use this to show that the shift map is ergodic.

**Corollary 1.** Let $(X, \mathcal{F}, \mu)$ be a probability space and $T$ a measure preserving transformation. Then $T$ is ergodic iff for all $A, B \in \mathcal{F}$, one has

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

Before proving it, we give a bit of intuition. One expects that an ergodic transformation is precisely one that ‘mixes’ around sets of positive measure. Then essentially the pre-orbit of a set should be equidistributed, so that the ‘events’ $T^{-n}A$ and $B$ should be asymptotically independent in some sense. One can turn this slightly shifted perspective into new (stronger) notions of weak and strong mixing transformations.

**Proof.** First we prove this implies $T$ is ergodic. Let $A$ be a measurable set with $T^{-1}A = A$. Then we have

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

However, from hypothesis, we have

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

Therefore $\mu(A) = \mu(A)^2$, so in particular either $\mu(A) = 0$ or $1$.

Now suppose that $T$ is ergodic. Then we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \int_X 1_{T^{-i}A} \cdot 1_B$$

$$= \lim_{n \to \infty} \int_X 1_B \frac{1}{n} \sum_{i=0}^{n-1} 1_{T^{-i}A}$$

$$= \lim_{n \to \infty} \int_X 1_B \frac{1}{n} \sum_{i=0}^{n-1} 1_A \circ T^i$$

$$= \int_X 1_B \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A \circ T^i \quad \text{by DCT}$$

$$= \int_X 1_B \mu(A) \quad \text{by the Ergodic Theorem}$$

$$= \mu(B)\mu(A)$$

$\square$
In practice, however, one has trouble verifying a statement for arbitrary elements of a \( \sigma \)-algebra — they’re just too big! Luckily, it suffices to check this statement for a generating semi-algebra.

**Proposition 1.** Let \((X, \mathcal{F}, \mu)\) be a probability space, and \(\mathcal{S}\) a generating semi-algebra of \(\mathcal{F}\). Let \(T : X \to X\) be a measure preserving transformation. Then \(T\) is ergodic if and only if for all \(A, B \in \mathcal{S}\) one has

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

**Proof.** If \(T\) is ergodic, then this holds for every pair of elements in \(\mathcal{F}\), so it trivially holds for pairs in \(\mathcal{S}\).

Now suppose that the statement holds for our semi-algebra. It suffices to show that it holds for any pair of elements in the \(\sigma\)-algebra \(\mathcal{F}\). We use the one idea of measure theory: approximate!

Let \(\varepsilon > 0\) be arbitrary and \(A, B \in \mathcal{F}\). Then by lemma 1.2.1 there exists \(A_0\) and \(B_0\) finite disjoint unions of elements of \(\mathcal{S}\) such that

\[
\mu(A \Delta A_0) < \varepsilon, \text{ and } \mu(B \Delta B_0) < \varepsilon.
\]

Then as

\[
(T^{-i}A \cap B) \Delta (T^{-i}A_0 \cap B_0) \subset (T^{-i}A \Delta T^{-i}A_0) \cup (B \Delta B_0).
\]

we know that

\[
|\mu(T^{-i}A \cap B) - \mu(T^{-i}A_0 \cap B_0)| \leq \mu[(T^{-i}A \cap B) \Delta (T^{-i}A_0 \cap B_0)]
\]

\[
\leq \mu(T^{-i}A \Delta T^{-i}A_0) + \mu(B \Delta B_0)
\]

\[
< 2\varepsilon.
\]

Also,

\[
|\mu(A)\mu(B) - \mu(A_0)\mu(B_0)| \leq \mu(A)|\mu(B) - \mu(B_0)| + \mu(B_0)|\mu(A) - \mu(A_0)|
\]

\[
\leq |\mu(B) - \mu(B_0)| + |\mu(A) - \mu(A_0)|
\]

\[
\leq \mu(B \Delta B_0) + \mu(A \Delta A_0)
\]

\[
< 2\varepsilon.
\]

Putting these approximations together, we have

\[
\left(\frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A \cap B) - \mu(A)\mu(B)\right) - \left(\frac{1}{n} \sum_{i=0}^{n-1} \mu(T^{-i}A_0 \cap B_0) - \mu(A_0)\mu(B_0)\right) < 4\varepsilon.
\]

Therefore as this holds for any \(\varepsilon > 0\), and the latter quantity goes to 0 in the limit, we have

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

\(\square\)
3 The Strong Law of Large Numbers

Having given a nice characterization of ergodicity in terms of ‘asymptotic average independence’, we are prepared to show the following:

Proposition 2. Let \((X, \mathcal{F}, \mu)\) be a probability space and \((X^\mathbb{N}, \mathcal{F}^\mathbb{N}, \mu^\mathbb{N})\) be the product space (indexed over the naturals. Then if we let \(T : X^\mathbb{N} \to X^\mathbb{N}\) be the shift map, \(T\) is ergodic.

Proof. The \(\sigma\)-algebra \(\mathcal{F}^\mathbb{N}\) is generated by the semi-algebra \(\mathcal{S}\) of cylinder sets. Let \(A, B \in \mathcal{S}\) be two such cylinder sets. Then if there are only finitely many coordinates in which \(A\) and \(B\) are not all of \(X\). As a result, for some sufficiently high \(i\), the coordinates \(T^{-i}A\) and \(B\) are independent. Since we took the product measure, for such \(i\) we have \(\mu(T^{-i}A \cap B) = \mu(A)\mu(B)\). Therefore in particular we have

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

This actually proves something stronger, that the shift map \(T\) is strongly mixing \(\dagger\).

Now we can prove a strong theorem:

Theorem 2 (Strong Law of Large Numbers). Let \(X_1, X_2, ...\) be a sequence of i.i.d random variables with \(\mathbb{E}|X_1| < \infty\). Then

\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}X_1
\]

almost surely.

Proof. If we consider \(X_1 : (X, \mathcal{F}, \mu) \to \mathbb{R}\), then the pushforward of \(\mu\) under \(X_1\) defines a probability measure \(\mu_*\) on \(\mathbb{R}\). If we let \(\Omega = \mathbb{R}^\mathbb{N}\) be the product space with the product measure \(\mu_*^\mathbb{N}\), and let \(\mathbb{R}^\mathbb{N} \to \mathbb{R}\) be the projection onto the first coordinate, then \(f \in L^1\) because \(\mathbb{E}|X_1| < \infty\).

Now if we let \(T : \Omega \to \Omega\) denote the shift map, then by the ergodic theorem, we have

\[
\lim_{n \to \infty} \frac{X_1 + X_2 + ... + X_n}{n} = \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i = \int_{\Omega} f = \mathbb{E}X_1
\]

almost everywhere. \(\dagger\)

\(\dagger\)Which means that for all \(A, B \in \mathcal{F}\) we have \(\mu(T^{-i}A \cap B) \to \mu(A)\mu(B)\). This isn’t a complete proof that \(T\) is strongly mixing, because we would need to prove a statement similar to proposition 1, which allows us to just verify the statement on cylinder sets.
4 Ergodicity and Measures

Here’s a thing

Theorem 3. Suppose $\mu_1$ and $\mu_2$ are probability measures on $(X, \mathcal{F})$ and $T : X \to X$ is measurable and measure preserving with respect to $\mu_1$ and $\mu_2$.

1. If $T$ is ergodic with respect to $\mu_1$ and $\mu_2$ is absolutely continuous with respect to $\mu_1$. Then $\mu_1 = \mu_2$.

2. If $T$ is ergodic with respect to $\mu_1$ and $\mu_2$, then either $\mu_1 = \mu_2$ or $\mu_1$ and $\mu_2$ are mutually singular.

Proof. 1. Suppose $T$ is ergodic with respect to $\mu_1$ and $\mu_2$ is absolutely continuous with respect to $\mu_1$. For any $A \in \mathcal{F}$ the ergodic theorem tells us that for almost every $x \in X$ one has

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) = \mu_1(A).$$

Let $C_A$ be the set of points $x \in X$ for which this holds, so that $\mu_1(C_A) = 1$ and $\mu_2(C_A) = 1$ by absolute continuity.

Then as $T$ is measure preserving w.r.t $\mu_2$, then for each $n \geq 1$ we get

$$\frac{1}{n} \sum_{i=0}^{n-1} \int_X 1_A(T^i x) d\mu_2 = \mu_2(A).$$

Now by applying the dominated convergence theorem, because this sum is bounded by 1, we get

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \int_X 1_A(T^i x) = \lim_{n \to \infty} \int_X \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) d\mu_2$$

$$= \int_X \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) d\mu_2$$

$$= \int_X \mu_1(A) d\mu_2$$

$$= \mu_1(A),$$

which implies that $\mu_1(A) = \mu_2(A)$. Hence $\mu_1 = \mu_2$.

2. Now suppose that $T$ is ergodic with respect to both measures $\mu_1$ and $\mu_2$. Suppose that they are not the same, then there exists some $A \in \mathcal{F}$ with $\mu_1(A) \neq \mu_2(A)$.

Like before, for $k = 1, 2$ let $C_k$ be the set of points for which $\lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} 1_A(T^i x) = \mu_k(A)$. Then the ergodic theorem tells us that $\mu_k(C_k) = 1$, however $C_1$ and $C_2$ must be disjoint as $\mu_1(A) \neq \mu_2(A)$. 

6
Therefore $\mu_1$ and $\mu_2$ are supported on disjoint sets and so are mutually singular.