The Birkhoff Ergodic theorem, and an application to continued fractions.

Math 172, Spring 2009 (Gautam Iyer)

Let $(X, \mathcal{M}, \mu)$ be a measure space with $\mu(X) = 1$.

Definition 1. We say $T : X \to X$ is measure preserving if $\mu(T^{-1}(A)) = \mu(A)$ for all $A \in \mathcal{M}$. (Sometimes we say μ is an invariant measure of T)

Note that the definition states that $\mu(T^{-1}A) = \mu(A)$, and not $\mu(TA) = \mu(A)$. Of course, if T is bijective the presence of the inverse isn't relevant. If not, consider the situation where two disjoint sets A_1, A_2 are each mapped bijectively to some set B. Since $T(A_1) = T(A_2) = B$, if you had $\mu(A_1) = \mu(A_2) = \mu(B)$, you certainly won't have $\mu(A_1 \cup A_2) = \mu(B)$ even though $T(A_1 \cup A_2) = B$ (in this case $\mu(A_1 \cup A_2) = 2\mu(B)$). However, since every point in B has two pre-images (one in A_1 and one in A_2), what would be natural instead would be if each pre-image 'counted for half'; namely $\mu(A_1) = \mu(A_2) = \frac{1}{2}\mu(B)$. In this case $\mu(T^{-1}B) = \mu(A_1 \cup A_2) = \mu(B)$, which is our definition.

If you're still not convinced our definition is natural, then next proposition should convince you.

Proposition 2. *If T is measure preserving, $\int_X f \, d\mu = \int_X f \circ T \, d\mu$.*

Proof. As always, it is enough to check this for simple functions. By linearity, we can reduce simple functions to characteristic functions. Now $\chi_A \circ T = \chi_{T^{-1}A}$, hence $\mu(T^{-1}A) = \int \chi_A \circ T$ and $\mu(A) = \int \chi_A$ which are equal by definition of invariance. $\square$

Definition 3. We say $T : X \to X$ is ergodic if whenever $T^{-1}A = A$, $\mu(A) = 0$ or $\mu(A) = 1$.

Intuitively, ergodic maps are maps which 'mix' very well.

Example 4. Let $X = [0, 1]$, μ the Lebesgue measure, and $T(x) = 2x - [2x]$ (i.e. $T(x)$ is the fractional part of $2x$). Then T is measure preserving and ergodic.

Theorem 5 (Birkhoff[1] Ergodic Theorem)**.** *Let $T : X \to X$ be measure preserving and ergodic, and $f \in L^1(X)$. Then almost everywhere*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T^i = \int_X f \, d\mu. \tag{1}$$

Here $T^i = T \circ T \cdots T$ i-times, and T^0 is the identity map. The non-technical way of stating this theorem is that 'space averages are equal to time averages', which can be traced back to an idea of Boltzman: Namely, to compute the average velocity of a gas, one could compute the velocity of each molecule and then divide by the number of molecules. This is exactly the right hand side of (1). Alternately, one could make the assumption that over time, the trajectory of one molecule essentially visits the entire volume occupied by the gas (i.e. the trajectory is 'equi-distributed'). In this case, we could compute the average velocity by computing the average velocity of *this one molecule* over a long period of time. This is exactly the left hand side of (1). Thus the ergodic theorem states that any measure preserving, ergodic map has equi-distributed trajectories.

[1]If we replace the assumption $f \in L^1(X)$ with $f \in L^2(X)$, there is a short 'slick' proof of (1) by Von-Neumann, using elementary Hilbert space techniques. Of course, $\mu(X) = 1$ and $f \in L^2 \implies f \in L^1$, so the Von-Neumann result is a special case of Birkhoff.

The proof of the theorem follows immediately from the following lemma.

Lemma 6. *If $\int_X f < 0$, then $\limsup \frac{1}{n}\sum_0^{n-1} f \circ T^i \leqslant 0$. If $\int_X f > 0$, then $\liminf \frac{1}{n}\sum_0^{n-1} f \circ T^i \geqslant 0$.*

Proof. Define

$$S_n = \frac{1}{n}\sum_0^{n-1} f \circ T^i, \quad F_n = \max_{1\leqslant k\leqslant n} \sum_0^{k-1} f \circ T^i \quad \text{and} \quad A = \{F_n \to +\infty\}.$$

Note $F_{n+1} \geqslant F_n$, thus for all $x \notin A$, there exists M such that $F_n(x) \leqslant M$ for all n. Hence $S_n(x) \leqslant \frac{1}{n}F_n(x) \leqslant \frac{M}{n}$, and so $\limsup S_n(x) \leqslant 0$. It remains to show that $\mu(A) = 0$. Now consider

$$\begin{aligned} F_{n+1} - F_n \circ T &= \max_{1\leqslant k\leqslant n+1} \sum_0^{k-1} f \circ T^i - \max_{2\leqslant k\leqslant n+1} \sum_1^{k-1} f \circ T^i \\ &= f + \max_{1\leqslant k\leqslant n+1} \sum_0^{k-1} f \circ T^i - \max_{2\leqslant k\leqslant n+1} \sum_0^{k-1} f \circ T^i \end{aligned}$$

Now if the maximum in the first term occurs at $k > 1$, then the two maximums are equal and cancel, giving $F_{n+1} - F_n \circ T = f$. The maximum occurring at $k = 0$ is exactly the same as $F_{n+1} = f$, which is equivalent to

$$\max_{2\leqslant k\leqslant n+1} \sum_0^{k-1} f \circ T^i \leqslant f \iff f + F_n \circ T \leqslant f \iff F_n \circ T \leqslant 0.$$

Thus

$$F_{n+1} - F_n \circ T = f - \min\{0, F_n \circ T\} \tag{2}$$

which is the key to the proof. Note that (2) immediately implies that $(F_{n+1} - F_n \circ T)$ is decreasing, and on A decreases to f. By the dominated convergence theorem, $\lim \int_A (F_{n+1} - F_n \circ T) \to \int_A f$.

Next note that A is invariant (i.e. $T^{-1}A = A$). Thus $\int_A F_n \circ T = \int_A F_n$. This gives $0 \leqslant \int_A (F_{n+1} - F_n) = \int_A (F_{n+1} - F_n \circ T) \to \int_A f$, showing $\int_A f \geqslant 0$. Finally since A is invariant, $\mu(A) = 0$ or $\mu(A) = 1$. If $\mu(A) = 1$, then $\int_A f = \int_X f < 0$, which is impossible. Thus $\mu(A) = 0$, finishing the proof. Now replacing f with $-f$, the inequality for $\liminf S_n$ follows. $\square$

Proof of Theorem 5. Let $g = f - \int_X f - \varepsilon$. By the lemma,

$$\left(\limsup \frac{1}{n}\sum_0^{n-1} f \circ T^i\right) - \int_X f - \varepsilon = \limsup \frac{1}{n}\sum_0^{n-1} g \circ T^i \leqslant 0 \quad \text{a.e.}$$

and hence $\limsup \frac{1}{n}\sum_0^{n-1} f \circ T_i \leqslant \int_X f + \varepsilon$ almost everywhere. The reverse inequality for the $\liminf$ follows similarly, finishing the proof. $\square$

The Ergodic theorem has numerous consequences and deep applications. One quick application that follows from Example 4 is a special case of the (strong) law of large numbers!

Problem 7. Let $\Omega = [0,1]$ with the Lebesgue measure. Let $X_n(x)$ be the n^{th} digit in the *binary* expansion of x.

(1) Show that X_n are independent, identically distributed random variables.
(2) Show that $X_{n+1} = X_n \circ T$, where T is defined in Example 4. Now use the ergodic theorem to show $\lim \frac{1}{n}\sum_1^n X_i = \frac{1}{2}$ almost everywhere.

We conclude with a very surprising result about continued fractions. Recall for any $x \in (0,1]$, there exists (unique) integers $a_1(x), a_2(x), \dots$ such that

$$x = \cfrac{1}{a_1(x) + \cfrac{1}{a_2(x) + \cfrac{1}{a_3(x) + \cdots}}}$$

We use the notation $[a_1(x), a_2(x), \dots]$ to denote the expression on the right (called the continued fraction). For any $n \in \mathbb{N}$, we note that $[a_1(x), \dots, a_n(x)]$ is a rational number. We define $p_n(x)$, $q_n(x)$ to be the numerator and denominator of $[a_1(x), \dots, a_n(x)]$ in reduced terms. It is well known that $(\frac{p_n(x)}{q_n(x)}) \to x$, and is in some sense the best approximating sequence to x by rational numbers. Now intuitively, if a sequence of rational numbers converges (well) to some real number, then we expect the denominators to grow exponentially. One question would be to determine the rate at which the denominators grow. The answer is quite surprising.

Theorem 8. *For almost every $x \in (0,1]$,*

$$\lim_{n\to\infty} \frac{\ln q_n(x)}{n} = \frac{\pi^2}{12 \ln 2}.$$

While we prove this for almost all x, it is quite hard to actually produce *one* explicit example of a number x with this property! The proof can be traced down to an ingenious idea of Gauss, and an application of the ergodic theorem.

Proposition 9 (Gauss)**.** *Let $S(x) = \frac{1}{x} - [\frac{1}{x}]$ (i.e. $S(x)$ is the fractional part of $\frac{1}{x}$). Let $d\mu(x) = \frac{1}{(1+x)\ln 2} d\lambda(x)$. Then S is measure preserving, and ergodic!*

Proof. The clever part is guessing what the invariant measure is. Checking invariance and ergodicity although long and technical is not too hard, and we leave it to the interested reader. □

Proof of Theorem 8. The reason the map S comes into play is because if $x = [a_1, a_2, \dots]$, then $\frac{1}{x} = a_1 + [a_2, \dots]$, and hence $S(x) = [a_2, a_3, \dots]$. Thus the map S acts like a shift on the continued fraction expansion of x (which is why we use S instead of T). The remainder of the proof shows how one can write $\frac{\ln q_n(x)}{n}$ in a form where the ergodic theorem applies. Note

$$\frac{p_{n-1}(Sx)}{q_{n-1}(Sx)} = [a_2(x), \dots a_{n-1}(x)] = \frac{1}{[a_1(x), a_2(x), \dots, a_n(x)]} - a_1(x)$$
$$= \frac{q_n(x) - p_n(x)a_1(x)}{p_n(x)},$$

and since the fraction on the right is in reduced terms, we conclude $p_n(x) = q_{n-1}(Sx)$. Thus

$$\frac{1}{q_n(x)} = \frac{p_1(S^{n-1}x)}{q_n(x)} = \frac{p_n(x)}{q_n(x)} \frac{p_{n-1}(Sx)}{q_{n-1}(Sx)} \cdots \frac{p_1(S^{n-1}x)}{q_1(S^{n-1}x)}$$

since $p_1(x) = 1$ for all $x \in (0,1]$. Now taking the logarithm and dividing by n gives

$$\frac{\ln \circ q_n}{n} = \frac{1}{n} \sum_{i=1}^n - \ln \circ \frac{p_i}{q_i} \circ S^{n-i}$$

The key is to realise that in the previous equation $\frac{p_i}{q_i}$ can be *ignored!*

Lemma 10. $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n - \ln \circ \frac{p_i}{q_i} \circ S^{n-i} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n - \ln \circ S^{n-i}$ *pointwise.*

The proof stems from the fact that for any sequence $(b_n) \to b$, the sequence of averages $(\frac{1}{n}\sum_0^{n-1} b_i)$ also converges to b. Now it is easy to check $(\frac{p_i}{q_i}(x)) \to x$ uniformly, at an exponential rate which is the heart of the matter. Of course, a few added technicalities are involved since we have $\frac{p_i}{q_i} \circ S^{n-i}$ instead of $\frac{p_i}{q_i}$, and we leave the details of this to the interested reader.

Now returning to our proof, $\sum_1^n - \ln \circ S^{n-i} = \sum_1^n \ln \circ S^i$. So by the Lemma, Birkhoff, and Proposition 9,

$$\lim_{n\to\infty} \frac{\ln \circ q_n}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^n - \ln \circ S^i = \int_0^1 - \ln(x) \frac{1}{1+x} \frac{dx}{\ln 2},$$

and it only remains to compute the integral on the right. As it turns out, this is one of those integrals you can't do explicitly ⌣̈, so some devious trick is required. Fortunately Gauss has done the hard work for us. Recall $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} \cdots$, then integrating by parts gives

$$\int_0^1 - \ln(x) \frac{1}{1+x} \, dx = \int_0^1 \frac{\ln(1+x)}{x} \, dx = \sum_{n=1}^\infty \int_0^1 (-1)^{n+1} \frac{x^{n-1}}{n} \, dx = \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n^2},$$

and it remains to compute the sum on the right. Note $\sum \frac{1}{n^2} = \sum \frac{1}{(2n)^2} + \frac{1}{(2n-1)^2} \implies \sum \frac{1}{(2n-1)^2} = \frac{3}{4} \sum \frac{1}{n^2}$. Thus $\sum \frac{(-1)^{n+1}}{n^2} = \sum \frac{1}{(2n-1)^2} - \sum \frac{1}{(2n)^2} = (\frac{3}{4} - \frac{1}{4}) \sum \frac{1}{n^2}$, which (by Gauss) is exactly $\frac{\pi^2}{12}$. □