APPLICATIONS OF THE ERGODIC THEOREMS

1. Normal numbers.

This is my favourite application of the pointwise ergodic theorem. Let $x \in [0,1]$, and write $x = 0.a_1a_2a_3\dots$ in base k (this each a_i lies in the set $\{0,1,\dots,k-1\}$). For each sequence $b_1 \dots b_j$ of digits, look at those n for which $a_{n+1}\dots a_{n+j} = b_1 \dots b_j$. We say that x is *normal* in base k if

$$\lim_{N\to\infty} \frac{1}{N}\#\{n \leqslant N : a_{n+1}\dots a_{n+j} = b_1\dots b_j\} = k^{-j} \tag{1.1}$$

for all j and for all choices of $b_1,\dots,b_j$, that is to say the number of occurrences of the pattern $b_1\dots b_j$ amongst the base k digits of x is what one expects it to be.

The normality of a number may be interpreted in terms of the $\times k$ maps T_k on the circle $\mathbb{T}$. Indeed $a_{n+1}\dots a_{n+j} = b_1\dots b_j$ if, and only if, $T_k^n x$ lies in the interval

$$I := \frac{b_1}{k} + \dots + \frac{b_j}{k^j} + [0, \frac{1}{k^j}).$$

Since the maps T_k are ergodic (we only proved this for $k = 2$, but the proof for general k is the same) it follows from the pointwise ergodic theorem that

$$\lim_{N\to\infty} \frac{1}{N}\#\{n \leqslant N : a_{n+1}\dots a_{n+j} = b_1\dots b_j\} = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} 1_I(T_k^n x) = \mu(I) = k^{-j}$$

for a.e. x. There are only countable many choices for j and $b_1,\dots,b_j$, so in fact the above holds for all such choices for a.e. x, that is to say almost all $x \in [0,1]$ are normal in base k.

It follows immediately that almost all $x \in [0,1]$ are normal to all bases; such numbers are called *absolutely normal.*

Before writing these notes I was of the belief that no absolutely normal number had been exhibited. There is a sense, however, in which Sierpiński constructed one in 1916. In essence he defined a particular absolutely normal number $x_0 \in [0,1]$ and gave an algorithm for computing it to any desired accuracy. To my knowledge no-one has ever given a simple “closed-form” expression for an absolutely normal number, though it is widely believed that the constants $\sqrt{2}, e, \pi$ and so on are all absolutely normal. It is not hard to construct a number which is normal in some base b; even constructing a number which is normal in two coprime bases b, b' seems difficult.

2. Khintchine’s recurrence theorem.

This result is not an application to number theory, but concerns measure-preserving systems themselves. It only requires von Neumann’s mean ergodic theorem and not the more difficult pointwise ergodic theorem.

Theorem 2.1 (Khintchine's recurrence theorem). *Let (X,μ,T) be a m.p.s. and let $\varepsilon>0$. Let $A\subseteq X$ be any set with $\mu(A)>0$. Then the set*

$$\{n\in\mathbb{N}:\mu(A\cap T^nA)\geqslant\mu(A)^2-\varepsilon\}$$

is syndetic.

Proof. Let $I\subseteq L^2(X)$ be the space of T-invariant functions, and let $\pi:L^2(X)\to I$ be projection . From the basic properties of conditional expectation (see the supplementary notes) it follows that $\pi(1_A\circ T^n)=\pi(1_A)$ for all n. It follows that

$$\begin{aligned}&|\mathbb{E}_{M\leqslant n<M+N}\mu(A\cap T^{-n}A)-\langle 1_A,\pi(1_A)\rangle|\\&\qquad=|\mathbb{E}_{M\leqslant n<M+N}\langle 1_A,1_A\circ T^n-\pi(1_A)\rangle|\\&\qquad=|\langle 1_A,S_N(1_A\circ T^M)-\pi(1_A)\rangle|\\&\qquad=|\langle 1_A,S_N(1_A\circ T^M)-\pi(1_A\circ T^M)\rangle|\\&\qquad=|\langle 1_{T^MA},S_N1_A-\pi(1_A)\rangle|\\&\qquad\leqslant\|S_N1_A-\pi(1_A)\|_2.\end{aligned}$$

The fourth line follows from the third by basic properties of conditional expectation (see the notes) and the fifth from the fourth by change of variables. The last line follows from the previous one by Cauchy-Schwarz.

Now we have $\langle 1_A-\pi(1_A),\pi(1_A)\rangle=\langle 1_A-\pi(1_A),1\rangle=0$. It follows immediately from these facts and the Cauchy-Schwarz inequality that

$$\langle 1_A,\pi(1_A)\rangle=\|\pi(1_A)\|_2^2\geqslant\langle\pi(1_A),1\rangle^2=\langle 1_A,1\rangle^2=\mu(A)^2.$$

It follows from these facts and the von Neumann ergodic theorem that

$$\mathbb{E}_{M\leqslant n<M+N}\mu(A\cap T^{-n}A)\geqslant\mu(A)^2-\varepsilon$$

provided that $N\geqslant N_0(\varepsilon)$ (the bound does not depend on M). In particular there is at least one n in every range $[M,M+N)$ with $\mu(A\cap T^{-n}A)\geqslant\mu(A)^2-\varepsilon$, provided that $N\geqslant N_0(\varepsilon)$. □

3. Continued fractions and the gauss map

Recall that the Gauss map $T:[0,1]\to[0,1]$ is defined by $T(x)=\{1/x\}$ if $x\neq 0$ and $T(0)=0$. The point 0 is only included to make the underlying space X compact, and plays no rôle in what follows.

Lemma 3.1 (Gauss map preserves the Gauss measure). *Let μ be Lebesgue measure on $[0,1]$, and define a new measure ν by $\nu(E):=\frac{1}{\log 2}\int_E\frac{d\mu(x)}{1+x}$. Then (X,T,ν) is a measure-preserving system.*

Proof. It suffices to prove that $\nu(T^{-1}(a,b))=\nu((a,b))$ for all $0<a<b<1$. By the usual limiting arguments one then obtains the required property $\nu(T^{-1}E)=\nu(E)$ for all measurable E. This is just simple calculus, combined with a small amount of thought.

Indeed the inverse image $T^{-1}(a,b)$ is the union $\bigcup_{n=1}^{\infty}\big(\frac{1}{n+b},\frac{1}{n+a}\big)$ and so we have

$$\begin{aligned}\nu(T^{-1}(a,b))\log 2 &= \sum_{n=1}^{\infty}\int_{1/(n+b)}^{1/(n+a)}\frac{dx}{1+x}\\ &= \sum_{n=1}^{\infty}\big(\log\big(\frac{1}{n+a}+1\big)-\log\big(\frac{1}{n+b}+1\big)\big)\\ &= \lim_{N\to\infty}\big(\log\big(\frac{n+a+1}{n+a}\big)-\log\big(\frac{n+b+1}{n+b}\big)\big).\end{aligned}$$

The sum telescopes, so this is

$$\lim_{N\to\infty}(\log(N+a+1)-\log(N+b+1))-\log(a+1)+\log(b+1)=\nu((a,b))\log 2,$$

as required. □

Proposition 3.2. *The Gauss map T is ergodic with respect to the Gauss measure ν.*

Proof. In a sense, this proof is similar in structure to the proof that the $\times 2$ map is ergodic. The details are, however, rather more difficult.

For any choice of integers $k_1,\dots,k_n \geqslant 1$ define the map $\psi_{k_n,\dots,k_1}:(0,1]\to(0,1]$ by

$$\psi_{k_n,\dots,k_1}(x) := \frac{1}{k_n+\frac{1}{k_{n-1}+\dots+\frac{1}{k_1+x}}},$$

that is to say

$$\psi_{k_n,\dots,k_1} = \phi_{k_n}\circ\dots\circ\phi_{k_1}$$

where $\phi_k(x) := \frac{1}{k+x}$. Note that $T^n\circ\psi_{k_n,\dots,k_1}$ is the identity map, and indeed every preimage of x under T^n is of the form $\psi_{k_n,\dots,k_1}(x)$ for some choice of the k_i.

Write $U_{k_n,\dots,k_1} := \psi_{k_n,\dots,k_1}((0,1))$. $U_{k_n,\dots,k_1}$ is easily seen to be an interval; we claim it has length at most $2^{-(n-1)/2}$. To prove this we proceed inductively, assuming the result true for $n-2$ and deducing it for n. For any $0<a<b\leqslant 1$ and for any $k\in\mathbb{N}$ we have

$$\phi_k(a)-\phi_k(b)=\frac{b-a}{(k+a)(k+b)}.$$

This is always less than $b-a$, and is less than $\frac{1}{2}(b-a)$ if $k\geqslant 2$ or if $a,b\geqslant 1/2$. Thus we may certainly proceed inductively if $k_{n-1}\geqslant 2$. But if $k_{n-1}=1$ then $\phi_{k_{n-1}}((a,b)) = (a',b')$ where $a',b'\geqslant 1/2$ and the length of $\phi_{k_n}\circ\phi_{k_{n-1}}((a,b))$ is bounded by $\frac{1}{2}(b-a)$. The claim follows.

If $[a,b]$ and $F:[a,b]\to\mathbb{R}$ is a function which never vanishes, we define the *variation* of $\mathrm{var}_{[a,b]}F$ on $[a,b]$ to be $\sup_{x,y\in[a,b]}|\frac{F(x)}{F(y)}|$. Our next claim is that the variation of the derivative $\psi'_{k_n,\dots,k_1}$ is bounded by an absolute constant, independently of n and the choice of $k_n,\dots,k_1$. This means that the map $\psi_{k_n,\dots,k_1}$ from X to $U_{k_n,\dots,k_1}$ is "almost affine" in a weak sense.

To see this, we use the chain rule to conclude that

$$\mathrm{var}_{[0,1]}\,\psi'_{k_n,\dots,k_1}\leqslant \mathrm{var}_{[0,1]}\,\psi'_{k_{n-1},\dots,k_1}\cdot\mathrm{var}_{\psi_{k_{n-1},\dots,k_1}((0,1])}\,\phi'_{k_n}.$$

By the previous claim the right hand quantity is bounded by $\max_I \operatorname{var}_I \phi'_{k_n}$, the maximum being taken over all intervals I of length at most $2^{-(n-1)/2}$. However

$$\operatorname{var}_{[a,b]} \phi'_k = |\frac{k+b}{k+a}|^2 \leqslant |1+b-a|^2 \leqslant 1+3(b-a).$$

It follows that

$$\operatorname{var}_{[0,1]} \psi'_{k_n,\dots,k_1} \leqslant \prod_{n=1}^{\infty}(1+3\cdot 2^{-(n-1)/2}),$$

an infinite product which is very rapidly convergent.

We call any interval $U_{k_n,\dots,k_1}$ a *standard interval at level* n (compare with the notion of a standard dyadic interval at scale n that we encountered in our discussion of the $\times 2$ map). Let $E \subseteq [0,1]$ be measurable. Let $E' := E \cap U_{k_n,\dots,k_1}$. Then $\psi = \psi_{k_n,\dots,k_1} : E \to E'$ is a bijection, and so by change of variables we have

$$\mu(E') = \int 1_{\psi(E)}(x)\, d\mu(x) = \int 1_E(y)\psi'(y)\, d\mu(y).$$

By the fact that $\operatorname{var}_{[0,1]} \psi'$ is bounded this lies between $M_1\mu(E)$ and $M_2\mu(E)$ where the ration M_2/M_1 is bounded independenetly of E. Noting that when $E = [0,1]$ we have $E' = U_{k_n,\dots,k_1}$, we see that

$$\mu(T^{-n}E \cap U_{k_n,\dots,k_1}) \geqslant c\mu(E)\mu(U_{k_n,\dots,k_1})$$

for some absolute constant $c > 0$ (there is a similar upper bound, but we do not require it). Suppose that E is T-invariant. Then $T^{-n}E = E$, and so we see that

$$\mu(E \cap \tilde{U}) \geqslant c\mu(E)\mu(\tilde{U})$$

for all sets $\tilde{U}$ which are unions of standard intervals at level n. Call such sets $\tilde{U}$ *standard at level* n.

Since T is surjective, every point in $(0,1)$ lies in some standard interval of length n. As the length of such intervals tends to zero with n, we see that for any open set U there is a nested sequence $U_1 \supseteq U_2 \supseteq \dots$, with U_n being a standard set at level n, such that $\bigcap U_n = U$. By the monotone convergence theorem we therefore have

$$\mu(E \cap U) \geqslant c\mu(E)\mu(U)$$

for all open sets U.

By the regularity property of Lebesgue measure there is, for any $\varepsilon > 0$, an open set U with $E^c \subseteq U$ and $\mu(U) \leqslant \mu(E^c) + \varepsilon$. It follows that $\mu(E \cap U) \leqslant \varepsilon$, and therefore $\varepsilon/c \geqslant \mu(E)\mu(U) \geqslant \mu(E)(1-\mu(E))$. Since ε was arbitrary, we must have $\mu(E) = 0$ or $\mu(E) = 1$. Finally, we note that it is a trivial matter to check that this implies that either $\nu(E) = 0$ or $\nu(E) = 1$ (in fact, μ and ν are absolutely continuous with respect to one another). □

Let us derive a corollary about the partial quotients of a "typical" number in $(0,1)$. Recall our observation that if

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

then $a_n = \lfloor 1/T^n x \rfloor$. Writing $f : (0,1] \to \mathbb{R}$ for the function defined by $f(t) := \log(\lfloor 1/t \rfloor)$ it is not hard to check that $f \in L^1([0,1], \nu)$. It follows from the pointwise ergodic theorem that for ν-a.e. x (and hence for μ-a.e. x) we have

$$\frac{1}{N}\sum_{n=1}^{N} \log a_n = S_N f(x) \to \int f \, d\nu = \frac{1}{\log 2}\int_0^1 \frac{\log(\lfloor 1/t \rfloor)}{1+t}\, dt.$$

Splitting the integral into the ranges $(1/(k+1), 1/k)$ and making the substitution $u = 1/t - k$, we see that the right hand side is

$$\frac{1}{\log 2}\sum_{k=1}^{\infty} \log k \int_0^1 \frac{1}{(u+k)(u+k+1)}\, du.$$

The integration is easily accomplished using the obvious partial fraction expansion, and we see that the above is

$$\frac{1}{\log 2}\sum_{k=1}^{\infty} \log k \cdot \log \frac{(k+1)^2}{k(k+2)}.$$

Thus for almost all x the partial quotients satisfy

$$\lim_{N\to\infty} (a_1 \dots a_N)^{1/N} = \prod_{k=1}^{\infty} \left(\frac{(k+1)^2}{k(k+2)} \right)^{\log k/\log 2}.$$

The constant here is called Khintchine's constant, and its numerical value is approximately 2.685452...

More applications of ergodic theory to the study of continued fractions may be found on the second example sheet.