Lecture Notes on Probabilistic Methods in Certain Counting Problems of Ergodic Theory

by

Steven P. Lalley
Purdue University

Technical Report #89-12

Department of Statistics
Purdue University

May 1989

1. Preface

These lectures are an introduction to a "probabilistic" approach to two classes of counting problems — problems concerning the orbit of a point under the action of a discrete group, and problems concerning the periodic orbits of hyperbolic flows. There are other methods, very elegant methods, that have been developed to deal with these problems. Mark Pollicott will talk about some of these in his lectures. The various approaches all have their advantages and disadvantages.

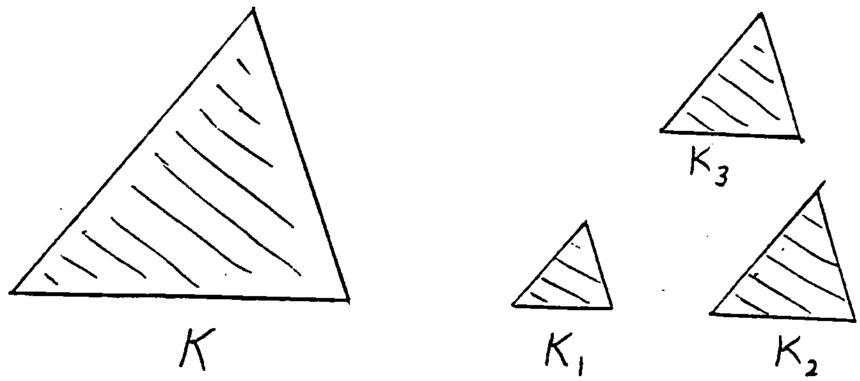
The methods that I will tell you about are based on an analogy with some parts of random walk theory. There is also some heavy machinery involved — Perron-Frobenius theory, Gibbs states, and so on — which, in my view, obscures the main ideas somewhat. So I am going to begin by talking about two simple problems where the machinery isn't needed (secs. 2–6). The solutions will only involve some elementary renewal theory, which you can read about in any basic text on stochastic processes (a resumé is given in sec. 4). I hope these problems won't seem too artificial. If you're patient enough to last through the lectures you will see that these problems are a good starting point. Afterwards (secs. 7–10) I will show how to handle some more difficult problems.

2. Semigroups and Self-Similar Fractals

Let's start by constructing a self-similar fractal in $\mathbb{R}^2$. Take a triangle K and let S_1, S_2, S_3 be similarity transformations of $\mathbb{R}^2$ which shrink K onto pairwise disjoint triangles at the vertices of K, specifically,

$$S_i x = \alpha_i(x - v_i) + v_i$$

where v_1, v_2, v_3 are the vertices of K. The contraction ratios α_i should be such that $0 < \alpha_i < 1$ and $\alpha_i + \alpha_j < 1$ for any pair $i \neq j$. Then S_1, S_2, S_3 map K onto pairwise disjoint triangles K_1, K_2, K_3 contained in K, each similar to K. The self-similar fractal $\wedge$ will be the unique compact set contained in K such that $\wedge = \bigcup_1^3 S_i \wedge$.



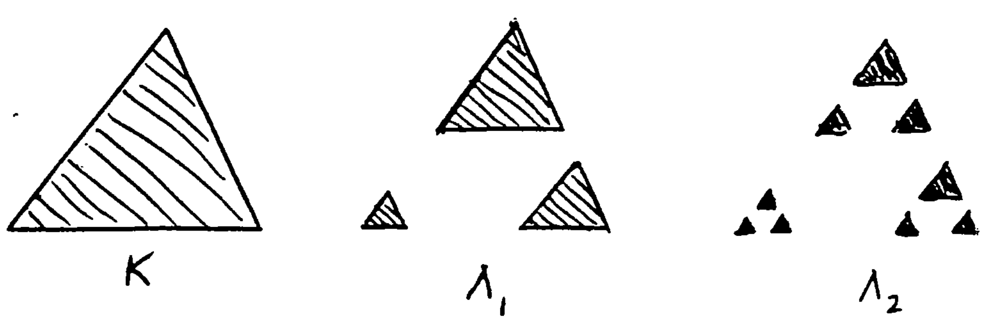
You can obtain Λ as follows. Let $i_1 i_2 \dots i_n$ be a sequence of length n from the index set $\{1, 2, 3\}$, and let Ω_n be the set of all such sequences. Define

$$K_{i_1 i_2 \dots i_n} = S_{i_1} S_{i_2} \dots S_{i_n} K,$$

$$\Lambda_n = \bigcup_{\Omega_n} K_{i_1 i_2 \dots i_n},$$

$$\Lambda = \bigcap_{n=1}^{\infty} \Lambda_n.$$

The intersection $\bigcap \Lambda_n$ is a nonempty compact set, because $K \supset \Lambda_1 \supset \Lambda_2 \supset \dots$ is a nested sequence of compact sets. The set $\Lambda = \bigcap \Lambda_n$ clearly satisfies $\Lambda = \bigcup_1^3 S_i \Lambda$; what's more, each of $S_1\Lambda$, $S_2\Lambda$, and $S_3\Lambda$ is a scaled-down copy of Λ. This is what is meant by "self-similarity".



Take an infinite sequence $i_1 i_2 \dots$ of 1's, 2's, and 3's. Then $K \supset K_{i_1} \supset K_{i_1 i_2} \supset \dots$

and diameter $(K_{i_1 i_2 \dots i_n}) = (\prod_1^n \alpha_i)$ diameter (K). Consequently,

$$k_{i_1 i_2 \dots} = \bigcap_{n=1}^{\infty} K_{i_1 i_2 \dots i_n}$$

consists of a single point. This point is in $\wedge$ because $K_{i_1 i_2 \dots i_n} \subset \wedge_n$. Notice that for distinct sequences $i_1 i_2 \dots$ and $i'_1 i'_2 \dots$ the points $k_{i_1 i_2 \dots}$ and $k_{i'_1 i'_2 \dots}$ are different, because $K_{i_1 i_2 \dots i_n} \cap K_{i'_1 i'_2 \dots i'_n} = \varnothing$ unless $i_j = i'_j \ \forall\, j \leq n$. Thus we have a $1-1$ correspondence between the set $\Omega = \{\text{all sequences } i_1 i_2 \dots\}$ and $\wedge$. The action of S_i on $\wedge$ is just

$$S_i k_{i_1 i_2 \dots} = k_{i i_1 i_2 \dots}.$$

Choose any $x \in K$. For each sequence $i_1 i_2 \dots \in \Omega$ and each $n = 1, 2, \dots$, the point $S_{i_1} S_{i_2} \dots S_{i_n} x \in K_{i_1 i_2 \dots i_n}$. Since the diameters of $K_{i_1 i_2 \dots i_n}$ shrink to zero as $n \to \infty$,

$$\lim_{n \to \infty} S_{i_1} S_{i_2} \dots S_{i_n} x = k_{i_1 i_2 \dots}\ .$$

So $\wedge$ is precisely the set of limit points of the "orbit" $\{S_{i_1} S_{i_2} \dots S_{i_n} x\}$ of x under the action of the semigroup generated by S_1, S_2, S_3. This also shows that $\wedge$ is the unique compact set satisfying $\wedge = \bigcup_{i=1}^{3} S_i \wedge$.

Take $x \in K - (\bigcup_1^3 K_i)$; for each finite sequence $i_1 i_2 \dots i_n \in \bigcup_{n=0}^{\infty} \Omega_n$ ($\Omega_0 = \{$empty sequence e$\}$) define

$$k_{i_1 i_2 \dots i_n} = S_{i_1} S_{i_2} \dots S_{i_n} x \quad (\text{and } k_e = x).$$

None of the points $k_{i_1 i_2 \dots i_n}$ is in $\wedge$, since $x \in K - (\bigcup_1^3 K_i)$, but the set of all possible limit points is $\wedge$. So for any $\epsilon > 0$ only finitely many of the points $k_{i_1 i_2 \dots i_n}$ are farther than ϵ from $\wedge$. Define

$$N(\epsilon) = \text{number of finite sequences } i_1 i_2 \dots i_n$$
$$\text{such that distance } (K_{i_1 i_2 \dots i_n}, \wedge) \geq \epsilon.$$

Problem: *How does $N(\epsilon)$ behave as $\epsilon \downarrow 0$?*

This is motivated by an analogous problem in hyperbolic geometry, the noneuclidean lattice point problem. Consider the orbit of a point x in hyperbolic space under the action of a discrete group of noneuclidean isometries. How many points of this orbit are within (hyperbolic) distance t of x (as $t \to \infty$)? We'll come back to a variant of this later.

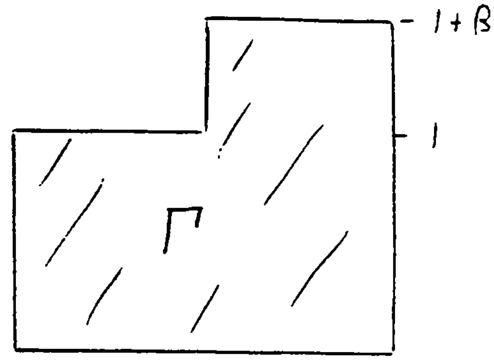
There are a number of other counting functions associated with $\wedge$ that are of interest. For example,

$$N^*(\epsilon) = \text{minimum number of } \epsilon\text{-balls}$$
$$\text{needed to cover } \wedge;$$

this is of interest because the *metric entropy* of $\wedge$ is determined by $N^*(\epsilon)$ (in particular, m.e. $(\wedge) = \lim_{\epsilon \downarrow 0}(\log N^*(\epsilon))/(\log \epsilon)$ provided the limit exists). The method we will use to analyze $N(\epsilon)$ also works for $N^*(\epsilon)$.

3. Periodic Orbits of the Bernoulli Flow

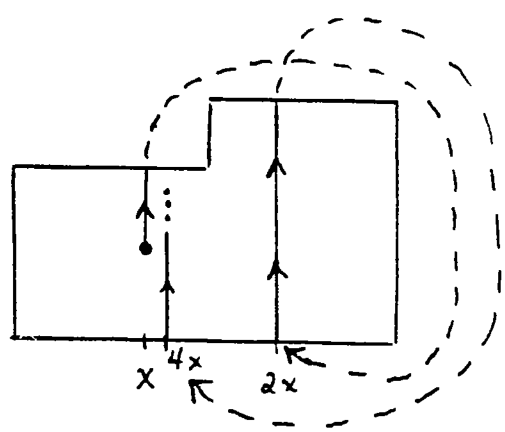
The Bernoulli flow is a contraption much beloved by ergodic theorists, because many chaotic flows are isomorphic to it in a measure-theoretic sense. That isn't important here, though.



Let $\beta > 0$ be a fixed *irrational* number, and define

$$r(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1/2 \\ \beta + 1 & \text{if } \frac{1}{2} < x < 1, \end{cases}$$

$$\Gamma = \{(x,t)\colon\ 0 \le x < 1 \text{ and } 0 \le t \le r(x)\}$$



The Bernoulli flow is the flow on Γ for which the dynamics are as follows. Starting at any $(x,t) \in \Gamma$, travel up the verticle line through (x,t) at unit speed until reaching the "ceiling" $(x, r(x))$; then jump instantaneously to $(2x, 0)$ (where $2x$ means $2x$ mod 1) and proceed at unit speed up the vertical segment through $(2x, 0)$.

Some orbits of this flow are periodic, for example, the orbit through (1/3, 0). In fact, the orbit through $(x, 0)$ is periodic iff x has a periodic binary expansion. If this is the case, and if n is the (smallest) period of the binary expansion, then the (smallest) period of the orbit through $(x, 0)$ is $r(x) + r(2x) + \ldots + r(2^{n-1}x)$. You can see from this that for any $t < \infty$ there are only finitely many periodic orbits with periods $\leq t$. Define

$$\begin{aligned} M(t) = &\text{ number of (distinct) periodic} \\ &\text{ orbits of the Bernoulli flow} \\ &\text{ with (minimal) period } \leq t. \end{aligned}$$

Problem: *How does $M(t)$ behave as $t \to \infty$? How are the periodic orbits distributed in Γ?*

There is some precedent for this problem also. Mathematicians have been interested in the periods of the closed geodesics on a negatively curved, compact manifold for years. For example, they figure prominently in Selberg's trace formula. And it has been known since Margulis' thesis (or longer) that there is a connection between closed geodesics and the noneuclidean lattice-point problem.

4. Some Elementary Renewal Theory

One of the fundamental theorems of probability theory is the *renewal theorem*, which guarantees the convergence to a steady state for large classes of stochastic processes. You can view it either as a Tauberian theorem (see Rudin [12] for a proof based on Wiener's general Tauberian theorem) or as a mixing property of random walk (see Lindvall [9] for a simple and intuitively appealing probabilistic proof). I'll present a very simple proof in the arithmetic case based on the ergodic theorem.

Let $F(dx)$ be a Borel probability measure with support contained in $[0, \infty)$ such that

$$\mu = \int_0^\infty xF(dx) < \infty. \tag{4.1}$$

In all the applications given in these lectures the measure F is supported by a finite set of points. If F is supported by a discrete (additive) subgroup of $\mathbb{R}$, and if $h\mathbb{Z}$ is the smallest such subgroup, call F *h-arithmetic*. Otherwise, call F *nonarithmetic*. Let $z(x)$ be a bounded, real-valued function with isolated discontinuities, satisfying

$$|z(x)| \leq C_1 e^{-C_2|x|} \qquad \forall\, x \in \mathbb{R} \tag{4.2}$$

for some $0 < C_1, C_2 < \infty$. The *renewal equation* is

$$Z(x) = z(x) + \int_0^\infty Z(x-y)F(dy). \tag{4.3}$$

RENEWAL THEOREM: *If F is nonarithmetic and $Z(x)$ satisfies the renewal equation $\forall\ x\ \in\ \mathbb{R}$ and is bounded on finite intervals, then*

$$\lim_{x\to\infty} Z(x) = \int_{-\infty}^\infty z(y)dy/\mu. \tag{4.4}$$

If F is h-arithmetic and $Z(x)$ satisfies the renewal equation $\forall\ x\ \in\ h\mathbb{Z}$ then

$$\lim_{n\to\infty} Z(nh) = \sum_{m\in\mathbb{Z}} z(mh)/\mu. \tag{4.5}$$

Here is an example of how this theorem works in the limit theory of stochastic processes. Suppose you have an electric lamp that burns continuously. Whenever the light bulb fails, you replace it with a new bulb. It is reasonable to assume that the lifetimes $X_1, X_2, \ldots$ of the successive bulbs are independent and satisfy $P\{X_i \in dx\} = F(dx)$ for some F. Suppose you insert a new bulb at time 0, and let R_t be the remaining lifetime of the bulb in the lamp at time t. If $Z_u(t) = P\{R_t > u\}$ then $Z_u(t) = 0$ for $t < 0$ and

$$Z_u(t) = P\{X_1 > u + t\} + \int_0^\infty Z_u(t-y)F(dy),$$

because the system "renews" itself at the instant the first bulb blows. The renewal theorem implies that for any $u > 0$

$$\lim_{t\to\infty} Z_u(t) = \int_0^\infty P\{X_1 > u+t\}dt/\mu,$$

so the distribution of R_t approaches a steady state as $t \to \infty$. (Note: The exponential decay condition (4.2) isn't necessarily satisfied in this example, but don't worry, I haven't stated the renewal theorem in complete generality. If you want the whole story, see Feller, vol. 2, ch. XI.)

The renewal equation can be reformulated in the language of probability as follows. Let $X_1, X_2, \ldots$ be independent random variables each with distribution F, i.e., $P\{X_i \in$

$dx\} = F(dx)$. Write $S_0 = 0$ and $S_n = X_1 + X_2 + \ldots + X_n$. Then the renewal equation (4.3) is equivalent to

$$Z(x) = z(x) + EZ(x - S_1).$$

LEMMA 1: *The unique solution of the renewal equation (4.3) that is bounded on each halfline* $(-\infty, a]$, $a < \infty$, *is*

$$Z(x) = E\sum_{n=0}^{\infty} z(x - S_n). \tag{4.6}$$

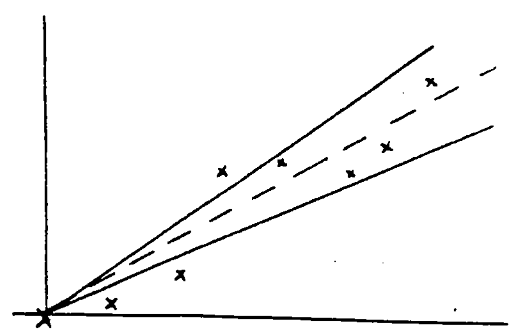
Lemma 1 is a fairly easy consequence of

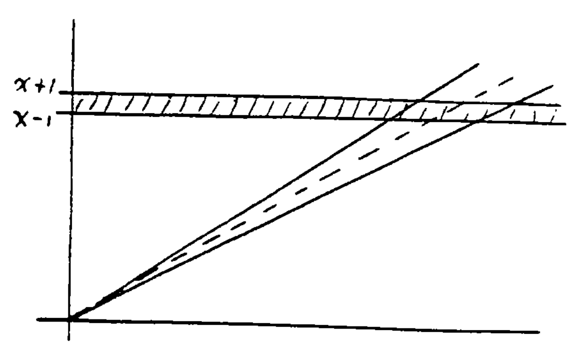
LEMMA 2: *For each* $\epsilon > 0$ *There exists a constant* $C < \infty$ *such that for any finite interval* $[a, b]$ *with* $b - a > \epsilon$,

$$E\sum_{n=0}^{\infty} 1_{[a,b]}(S_n) \leq C(b - a).$$

The function $1_{[a,b]}(x)$ takes the value 1 if $x \in [a, b]$ and is 0 otherwise. Lemma 2 says that the expected number of visits to any interval of length $> \epsilon$ by the sequence $\{S_n\}_{n\geq 0}$ is bounded by a multiple of the length of the interval. You can find proofs of both lemmas in Feller, vol. 2, ch. XI, sec. 1; both are easy.

According to the strong law of large numbers (the ergodic theorem), $S_n/n \to \mu$ a.s. Another way of saying this is that the sequence $(n, S_n)_{n\geq 0}$ will eventually stay between the lines with slopes $\mu + \epsilon$ and $\mu - \epsilon$ through the origin. Now suppose that $z(x)$ has compact support, say $[-1, 1]$. Then with high probability, if x is large, all the non-zero terms in the series (4.6) lie in the range $x/(\mu + \epsilon) \leq n \leq x/(\mu - \epsilon)$. Even if $z(x)$ does not have compact support, but satisfies the exponential decay hypothesis (4.2), the major contribution to the sum (4.6) will come from the terms $x/(\mu + \epsilon) \leq n \leq x/(\mu - \epsilon)$. In other words,





LEMMA 3: *For* $\epsilon > 0$ *define* $Z_\epsilon(x) = E\sum_{n:|n-\frac{x}{\mu}|>\epsilon x} |z(x - S_n)|$. *Then* $\forall\ \epsilon > 0$,

$$\lim_{x\to\infty} Z_\epsilon(x) = 0.$$

EXERCISE: Give a complete proof of this.

Here is a probabilistic proof of the renewal theorem in the arithmetic case. This proof can be adapted to the nonarithmetic case, but there are a few new wrinkles. Other proofs, all analytic in nature, are given in Breiman [4], Feller [5], and Rudin [12].

Asume that F is 1-arithmetic, and, for simplicity, that F is supported by the positive integers and $F(\{1\}) > 0$. Thus, the sequence $S_0, S_1, S_2, \ldots$ has positive jumps, and consequently doesn't visit any point more than once. According to Lemma 1, the solution of the renewal equation is

$$\begin{aligned} Z(m) &= E\sum_{n=0}^{\infty} z(m - S_n), \qquad m \in \mathbf{Z} \\ &= E\sum_{x \in \mathbf{Z}} z(x) \sum_{n=0}^{\infty} 1\{S_n = m - x\} \\ &= \sum_{x \in \mathbf{Z}} z(x) P\{S_n = m - x \text{ for some } n \geq 0\}. \end{aligned}$$

Since $\sum_{x \in \mathbf{Z}} |z(x)| < \infty$, we can prove (4.5) by showing that

$$\lim_{m\to\infty} H(m) = 1/\mu, \text{ where} \tag{4.7}$$
$$H(m) = P\{S_n = m \text{ for some } n \geq 0\}.$$

Suppose we can show that the limit in (4.7) exists. Call it α. Then $\alpha = 1/\mu$, because the existence of the limit implies

$$\begin{aligned} \alpha &= \lim_{M\to\infty} \frac{1}{M} \sum_{m=0}^{M} P\{S_n = m \text{ for some } n \geq 0\} \\ &= \lim_{M\to\infty} \frac{1}{M} \sum_{n=0}^{\infty} P\{S_n \leq M\} \\ &= \lim_{M\to\infty} \frac{1}{M} \sum_{n=1}^{\infty} P\{S_n/n \leq M/n\} \\ &= \lim_{M\to\infty} \frac{1}{M} \sum_{n=1}^{\infty} 1\{\mu \leq M/n\} \\ &= 1/\mu, \end{aligned}$$

by the (weak) law of large numbers.

To prove that the limit in (4.7) exists we introduce a system of random arrows on the integers as follows. Let $\{Y_m\}_{m \in \mathbb{Z}}$ be independent random variables, each with distribution F. For each $m \in \mathbb{Z}$, imagine that there is an arrow from m to $m + Y_m$ as in the figure.

For each starting point $m \in \mathbb{Z}$ the arrows specify an increasing sequence (path) $\gamma^{(m)}$ through the integers, namely $m \to m + Y_m \to m + Y_m + Y_{m+Y_m} \to \cdots$. For each $m \in \mathbb{Z}$ and $k \geq 0$, the probability that $\gamma^{(m)}$ passes through $m + k$ is $H(k)$. Suppose we can show that all of the paths $\gamma^{(m)}$, $m \leq 0$, eventually coalesce, i.e., with probability 1 there exists a (random) integer N such that the intersection of $\gamma^{(0)}$ with $\{N, N+1, N+2, \ldots\}$ is the same as that of $\gamma^{(m)}$ with $\{N, N+1, N+2, \ldots\}$ for every $m < 0$. Then it will follow that for all k, $k' \geq 1$

$$|H(k + k') - H(k)| \leq P\{N \geq k\};$$

since $\lim_{k \to \infty} P\{N \geq k\} = 0$, this will prove that the limit in (4.7) exists.

So to finish the argument we must show that with probability one all of the paths $\gamma^{(m)}$, $m \leq 0$, eventually coalesce. Say that there is a *bottleneck* at $k \in \mathbb{Z}$ if there are no arrows connecting $(-\infty, k-1]$ to $[k+1, \infty)$. If there is a bottleneck at k then all the paths $\gamma^{(m)}$, $m \leq k$, must pass through k. Therefore, we can finish the proof by showing that there is, with probability one, a bottleneck at some $k \geq 0$.

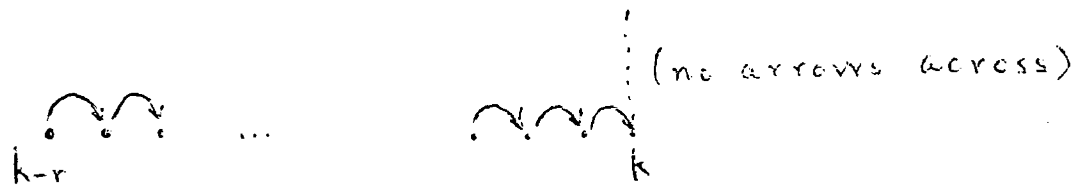
Remember that

$$\mu = \sum_{j=1}^{\infty} j\, F\{j\} = \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} F\{i\} < \infty,$$

so there is an r sufficiently large that

$$\sum_{j=r+2}^{\infty} \sum_{i=j}^{\infty} F\{i\} \leq 1/2.$$

Consequently, for any $k \in \mathbb{Z}$ the probability that there is an arrow connecting $(-\infty, k-r-1]$ to $[k+1, \infty)$ is $\leq 1/2$. Since $F\{1\} > 0$, the probability that $Y_m = 1$ for each $m \in \{k-r, k-r+1, \ldots, k-1\}$ is $(F\{1\})^r > 0$.



Therefore, the probability that k is a bottleneck is at least $(1/2)(F\{1\})^r > 0$. Now let $J_k = 1$ if k is a bottleneck and $J_k = 0$ otherwise. Then $\{J_k\}_{k \geq 0}$ is an ergodic, stationary sequence (since it is derived from the ergodic, stationary sequence $\{Y_m\}_{m \in \mathbb{Z}}$), so by the ergodic theorem

$$\lim_{K \to \infty} \frac{1}{K} \sum_{k=0}^{K} J_k = EJ_0 \geq (1/2)(F\{1\})^r > 0$$

almost surely. Therefore, with probability one, there is a bottleneck at some $k \geq 0$.

PROBLEM 1: Modify this argument to show that the assumption $F\{1\} > 0$ is unnecessary.

*PROBLEM 2: Find a similar proof for the nonarithmetic case.

5. Accumulation of Orbits at Cantor Sets

Let's consider again the counting function $N(\epsilon)$ introduced in sec. 2:

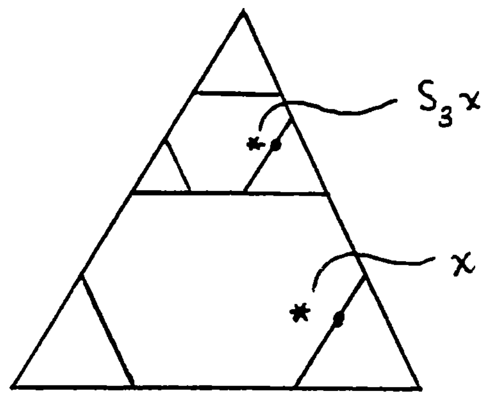
$$\begin{aligned} N(\epsilon) &= \text{number of finite sequences } i_1 i_2 \ldots i_n \\ &\qquad \text{such that distance } (K_{i_1 i_2 \ldots i_n}, \wedge) \geq \epsilon, \end{aligned}$$

where

$$k_{i_1 i_2 \ldots i_n} = S_{i_1} S_{i_2} \ldots S_{i_n} x, \quad x \in K - \bigcup_{i=1}^{3} K_i.$$

Remember that each of S_1, S_2, S_3 is a similarity transformation of $\mathbb{R}^2$, and that S_i multiplies distances by α_i. Since S_i maps $\wedge$ into $\wedge$, you can see (with a little work) that

(5.1) $$\text{distance } (k_{i_1 i_2 \ldots i_n}, \wedge) = \alpha_{i_1} \text{ distance } (k_{i_2 i_3 \ldots i_n}, \wedge)$$



Now consider those points $k_{i_1 i_2 \ldots i_n}$ that are counted in $N(\epsilon)$. Some are in K_1, some in K_2, some in K_3; only one, $k_e = x$, is not in $K_1 \cup K_2 \cup K_3$, and this only if distance $(x, \wedge) \geq \epsilon$. How many are in K_i? Any such $k_{i_1 i_2 \ldots i_n}$ must have $i_1 = i$ and distance $(k_{i_1 \ldots i_n}, \wedge) \geq \epsilon$, so by (5.1) distance $(k_{i_2 i_3 \ldots i_n}, \wedge) \geq \epsilon/\alpha_i$. Conversely, if distance $(k_{i_2 i_3 \ldots i_n}, \wedge) \geq \epsilon/\alpha_i$ then $k_{i i_2 i_3 \ldots i_n}$ is in K_i and its distance to $\wedge$ is $\geq \epsilon$. Therefore,

$$N(\epsilon) = R(\epsilon) + \sum_{i=1}^{3} N(\epsilon/\alpha_i) \qquad \text{where} \tag{5.2}$$

$$R(\epsilon) = \begin{cases} 1 & \text{if distance } (x, \wedge) \geq \epsilon \\ 0 & \text{if distance } (x, \wedge) < \epsilon \end{cases}$$

We can rewrite the functional equation (5.2) as a renewal equation. First, define the similarity dimension δ of $\wedge$ to be the unique (positive) real such that

$$\sum_{i=1}^{3} \alpha_i^{\delta} = 1.$$

Next,

$$Z(t) = e^{-\delta t} N(e^{-t}),$$
$$z(t) = e^{-\delta t} R(e^{-t});$$

then

$$Z(t) = z(t) + \sum_{i=1}^{3} \alpha_i^{\delta} Z(t - \log \alpha_i^{-1}). \tag{5.3}$$

This is a renewal equation. The kernel F is the probability distribution which attaches probabilities $\alpha_1^\delta, \alpha_2^\delta, \alpha_3^\delta$ to the points $\log \alpha_1^{-1}, \log \alpha_2^{-1}, \log \alpha_3^{-1}$. Applying the renewal theorem, we get

$$\lim_{t\to\infty} Z(t) = \int_0^\infty z(x)dx/\mu \qquad \text{(nonarithmetic case)}$$

$$\lim_{n\to\infty} Z(nh) = \sum_{x\in\mathbb{Z}} z(xh)/\mu \qquad \text{(h-arithmetic case)}$$

which is the same as

THEOREM 1: *If the smallest closed subgroup of* $\mathbb{R}$ *containing* $\log \alpha_1^{-1}, \log \alpha_2^{-1}, \log \alpha_3^{-1}$ *is* $h\mathbb{Z}$*, then*

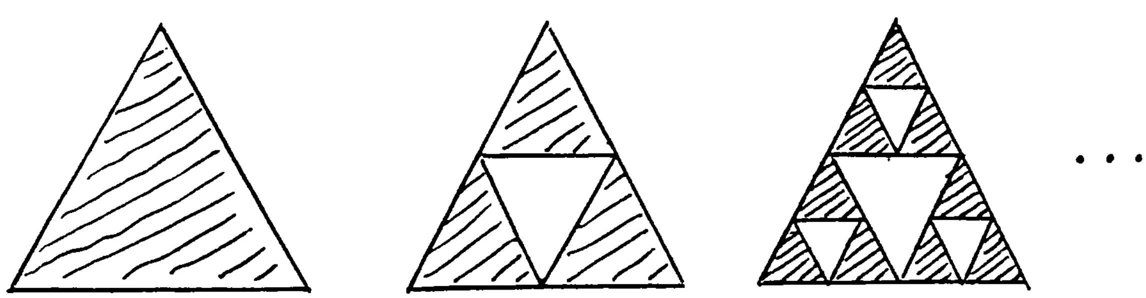
$$N(e^{-nh}) \sim e^{\delta nh} e^{-\delta n_* h}(1-e^{-\delta h})^{-1}(\sum_{i=1}^{3} \alpha_i^{-\delta} \log \alpha_i^{-1})$$

as $n \to \infty$*, where* n_* *is the smallest integer* $\geq h^{-1} \log\,(\text{dist}\ \ (x, \wedge))^{-1}$. *If there is no such subgroup, other than* $\mathbb{R}$ *itself, then*

$$N(\epsilon) \sim \epsilon^{-\delta}(\text{dist}\ (x,\wedge))^{\delta}/\delta \sum_1^3 \alpha_i^{-1} \log \alpha_i^{-1}.$$

PROBLEM 1: In the arithmetic case you can obtain much more precise information about $N(e^{-nh})$ by using generating functions and (5.2). Do so.

PROBLEM 2: Consider the "Sierpinski gasket", obtained by removing middle triangles successively from an equilateral triangle as shown.



Define

$$N^*(\epsilon) = \text{ number of ϵ-balls needed to cover.}$$

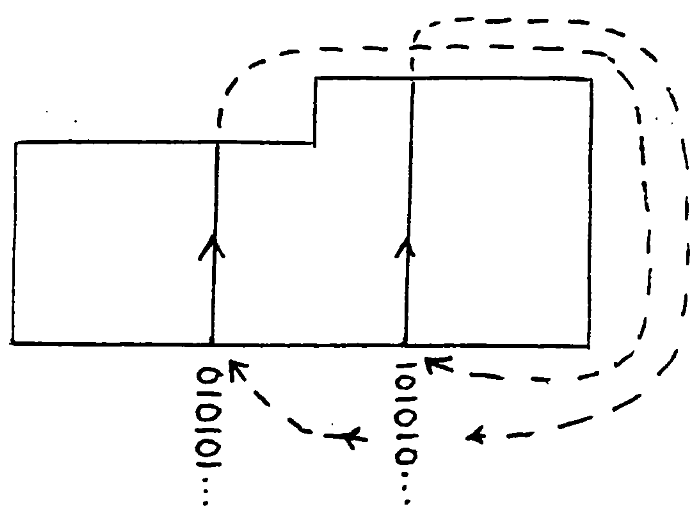
Show that $N^*(2^{-n}) \sim C \cdot 3^n$ for some $0 < C < \infty$.

HINT: See Lalley [6].

6. Counting Periodic Orbits of Bernoulli Flows

Recall that an orbit of the Bernoulli flow is periodic iff it passes through a point $(x,0)$ such that x has a periodic binary expansion. Also, if the least period of the binary expansion $x_1x_2x_3\dots$ of x is n then the period of the orbit through $(x,0)$ is

$$r(x)+r(2x)+r(4x)+\dots+r(2^{n-1}x)=n+\beta\sum_{i=1}^{n}x_i.$$



So it follows that the number of periodic orbits with (minimal) period $\leq t$ is

$$M(t)=\sum_{n=1}^{\infty}\frac{1}{n}\#\{\mathrm{x}=x_1x_2\dots x_n\in\Omega_n:\ n+\beta\sum_{i=1}^{n}x_i\leq t\}$$

where

$$\Omega_n=\{\text{sequences } x_1x_2\dots x_n \text{ of 0's and 1's of length } n \text{ with no period } d<n, d|n\}.$$

Define

$$\tilde{M}(t)=\sum_{n=1}^{\infty}\frac{1}{n}\#\{\mathrm{x}=x_1x_2\dots x_n\in\tilde{\Omega}_n:\ n+\beta\sum_{i=1}^{n}x_i\leq t\},$$

and

$$\tilde{N}(t)=\sum_{n=1}^{\infty}\#\{\mathrm{x}=x_1x_2\dots x_n\in\tilde{\Omega}_n:\ n+\beta\sum_{i=1}^{n}x_i\leq t\},$$

where

$$\tilde{\Omega}_n=\{\text{all sequences } x_1x_2\dots x_n \text{ of 0's and 1's of length } n\}.$$

We will begin by analyzing $\tilde{N}(t)$ and $\tilde{M}(t)$, then return to $M(t)$.

Consider a finite sequence $x_1 x_2 \dots x_n$ satisfying $n + \beta \sum_{i=1}^{n} x_i \leq t$. Either $x_1 = 0$, in which case $(n-1) + \beta \sum_{i=2}^{n} x_i \leq t-1$, or $x_1 = 1$, in which case $(n-1) + \beta \sum_{i=2}^{n} x_i \leq t-1-\beta$. Consequently,

$$\tilde{N}(t) = \tilde{N}(t-1) + \tilde{N}(t-1-\beta) + R(t), \text{ with}$$
$$R(t) = 1\{t \geq 1\} + 1\{t \geq \beta+1\}.$$

Let γ be the unique real number such that $e^{-\gamma} + e^{-\gamma(\beta+1)} = 1$; observe that $\gamma > 0$. Set

$$\tilde{Z}(t) = e^{-\gamma t}\tilde{N}(t),$$
$$\tilde{z}(t) = e^{-\gamma t}R(t).$$

Then

$$\tilde{Z}(t) = \tilde{z}(t) + e^{-\gamma}\tilde{Z}(t-1) + e^{-\gamma(\beta+1)}\tilde{Z}(t-\beta-1),$$

which is once again the renewal equation. The distribution F, given by

$$F\{1\} = e^{-\gamma}, \quad F\{1+\beta\} = e^{-\gamma(\beta+1)} \tag{6.1}$$

is nonarithmetic, because β is irrational. The mean of F is $\mu = e^{-\gamma}(1 + (\beta+1)e^{-\gamma\beta})$, and $\int_0^\infty \tilde{z} = (1/\gamma)$. Thus, by the renewal theorem,

$$\lim_{t\to\infty} \tilde{Z}(t) = \frac{1}{\gamma\mu} \Longrightarrow \tilde{N}(t) \sim \frac{e^{\gamma t}}{\gamma\mu}. \tag{6.2}$$

The idea behind the next step is simple, although the details are a little messy. Remember that $\tilde{Z}(t) = E\sum_{n=0}^{\infty} \tilde{z}(t-S_n)$ (Lemma 1) and that the main contribution to the sum comes from those terms for which $|n - t/\mu| \leq \epsilon t$ (Lemma 3). Now the only difference between the series defining $\tilde{M}(t)$ and $\tilde{N}(t)$ is the factor $1/n$ multiplying each term. If the main contributions to the series come from the range $|n - t/\mu| \leq \epsilon t$ then $1/\{t(\epsilon + \mu^{-1}\} \leq 1/n \leq 1/\{t(-\epsilon + \mu^{-1})\}$. Letting $t \to \infty$, $\epsilon \downarrow 0$, and using (6.2), you get

$$\tilde{M}(t) \sim \frac{e^{\gamma t}}{\gamma t}. \tag{6.3}$$

We'll give a complete justification of (6.3) shortly. First, though, let's consider again the function $M(t)$. The difference between $M(t)$ and $\tilde{M}(t)$ is that M counts only sequences

$x_1x_2\ldots x_n \in \Omega_n$, whereas $\tilde{M}$ counts *all* sequences $x_1x_2\ldots x_n \in \tilde{\Omega}_n$. Hence $M(t) \le \tilde{M}(t)$. Now if $\mathbf{x} = x_1x_2\ldots x_n \in \tilde{\Omega}_n - \Omega_n$ and $n+\beta\sum_1^n x_i \le t$, then $\mathbf{x} = x_1x_2\ldots x_d\, x_1x_2\ldots x_d \ldots x_d$ for $d = n/m$, $m > 1$, and $d+\beta\sum_1^d x_i \le t/m$, so $x_1x_2\ldots x_d$ is counted in $\tilde{M}(t/2)$. Therefore

$$M(t) \le \tilde{M}(t) \le M(t) + \tilde{M}(t/2),$$

and (6.3) implies

THEOREM 2: *$M(t) \sim e^{\gamma t}/\gamma t$ as $t \to \infty$.*

This is a special case of a theorem of Parry [10]. In fact you can obtain Parry's theorem from the renewal theorem — this makes a nice problem.

To prove (6.3) we will show that $\forall\ \epsilon > 0$,

(6.4) $$\lim_{t\to\infty} e^{-\gamma t} N_\epsilon(t) = 0 \text{ and}$$

(6.5) $$\lim_{t\to\infty} t e^{-\gamma t} M_\epsilon(t) = 0$$

where

$$M_\epsilon(t) = \sum_{\epsilon,t} \left(\frac{1}{n}\right) \#\{x_1x_2\ldots x_n \in \tilde{\Omega}_n\colon\ n + \beta\sum_1^n x_i \le t\},$$

$$N_\epsilon(t) = \sum_{\epsilon,t} \#\{x_1x_2\ldots x_n \in \tilde{\Omega}_n\colon\ n + \beta\sum_1^n x_i \le t\},$$

and $\sum_{\epsilon,t}$ denotes the sum over all $n \ge 1$ such that $|n - t/\mu| > \epsilon t$. Once (6.4)–(6.5) are accomplished, since

$$\frac{1}{t(\frac{1}{\mu}+\epsilon)} \le \frac{M(t) - M_\epsilon(t)}{N(t) - N_\epsilon(t)} \le \frac{1}{t(\frac{1}{\mu}-\epsilon)},$$

(6.2) will imply (6.3).

Let's start with (6.4):

$$\begin{aligned} e^{-\gamma t}N_\epsilon(t) &= \sum_{\epsilon,t}\sum_{\mathbf{x}\in\tilde{\Omega}_n} e^{-\gamma t} 1\{n + \beta\sum_1^n x_i \le t\} \\ &= \sum_{\epsilon,t}\sum_{\mathbf{x}\in\tilde{\Omega}_n} \exp\{-\gamma(t - (n+\beta\sum_1^n x_i))\}\exp\{-\gamma(n+\beta\sum_1^n x_i)\}1\{n+\beta\sum_1^n x_i \le t\} \\ &= \sum_{\epsilon,t} Ez(t - S_n) \end{aligned}$$

where $z(x) = e^{-\gamma x}1\{x \geq 0\}$ and $S_n = X_1 + X_2 + \ldots + X_n$ with $X_1, X_2, \ldots$ independent and distributed according to F, defined by (6.1). Explanation: for any $x_1 x_2 \ldots x_n \in \tilde{\Omega}_n$,

$$P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} = \exp\{-\gamma(n + \beta \sum_1^n x_i)\}.$$

Now (6.4) follows directly from Lemma 3.

Finally, consider (6.5):

$$te^{-\gamma t} M_\epsilon(t) = \sum_{\epsilon,t} n^{-1} \sum_{\mathbf{x} \in \tilde{\Omega}_n} te^{-\gamma t} 1\{n + \beta \sum_1^n x_i \leq t\}.$$

Fix $\delta > 0$ small, and write $\sum_{\epsilon,t} = \sum_{\epsilon,t}^* + \sum_{\epsilon,t}^{**}$, where

$$\sum_{\epsilon,t}^* = \text{sum over all } n \text{ included in } \sum_{\epsilon,t} \text{ such that } n \geq \delta t,$$

$$\sum_{\epsilon,t}^{**} = \text{sum over all } n \text{ included in } \sum_{\epsilon,t} \text{ such that } n < \delta t,$$

Then

$$\begin{aligned}
&\sum_{\epsilon,t}^* n^{-1} \sum_{\mathbf{x} \in \tilde{\Omega}_n} te^{-\gamma t} 1\{n + \beta \sum_1^n x_i \leq t\} \\
&\leq \sum_{\epsilon,t}^* \sum_{\mathbf{x} \in \tilde{\Omega}_n} \delta^{-1} e^{-\gamma t} 1\{n + \beta \sum_1^n x_i \leq t\} \\
&\leq \delta^{-1} e^{-\gamma t} N_\epsilon(t) \to 0 \text{ by (6.4).}
\end{aligned}$$

Also

$$\begin{aligned}
&\sum_{\epsilon,t}^{**} n^{-1} \sum_{\mathbf{x} \in \tilde{\Omega}_n} te^{-\gamma t} 1\{n + \beta \sum_1^n x_i \leq t\} \\
&\leq \sum_{\epsilon,t}^{**} \sum_{\mathbf{x} \in \tilde{\Omega}_n} te^{-\gamma t} 1\{n + \beta \sum_1^n x_i \leq t\} \\
&\leq \sum_{n=1}^{\infty} \sum_{\mathbf{x} \in \tilde{\Omega}_n} te^{-\gamma t} 1\{n + \beta \sum_1^n \leq \delta t(1 + \beta)\}
\end{aligned}$$

because if n is included in $\sum_{\epsilon,t}^{**}$ then $n < \delta t$ so $n + \beta \sum_1^n x_i \leq \delta t(1 + \beta)$. The last series equals

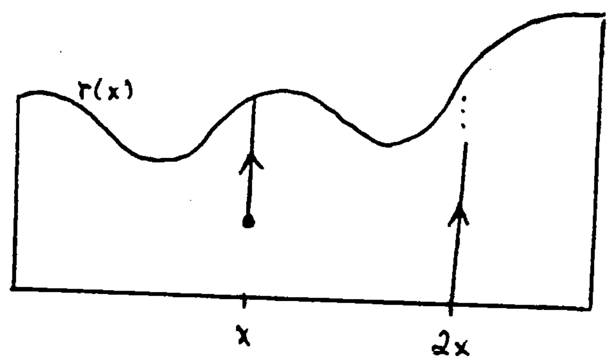
$$te^{-\gamma t} \tilde{N}(\delta t(1 + \beta)) \to 0$$

provided $\delta < (1+\beta)^{-1}$, by (6.2).

7. A More Sophisticated Renewal Theorem

Unfortunately, the usefulness of classical renewal theory in counting problems of the general types considered in secs. 2–3 is very limited. Let's look at yet another problem to see what the difficulty is.

Let $r(x)$ be a C^1, strictly positive function on [0,1] and let $\Gamma_r = \{(x,t)\colon\ 0 \le x < 1$ and $0 \le t \le r(x)\}$. We can define a flow on Γ_r the same way we defined the Bernoulli flow on Γ in sec. 3. Starting at any $(x,t) \in \Gamma_r$, travel up the vertical line segment through (x,t) at unit speed until reaching $(x, r(x))$, then jump instantaneously to $(2x, 0)$ and proceed up the vertical segment through $(2x, 0)$, etc. (As below, $2x$ means $2x$ mod 1.) Call this the *suspension flow* under r.



As for the Bernoulli flow, the periodic orbits of the suspension flow are the orbits through points $(x,0)$ for which x has a periodic binary expansion. If $(x,0)$ is such a point and n is the (smallest) period of the binary expansion then the period of the orbit through $(x,0)$ is

$$r(x) + r(2x) + \cdots + r(2^{n-1}x).$$

For the Bernoulli flow this sum was $n + \beta \sum_1^n x_i$. For the suspension flow, however, $\sum_0^{n-1} r(2^i x)$ cannot be represented as $\sum_1^n f(x_i)$. This is crucial, because if we define

$$\tilde{N}(t) = \sum_{n=1}^{\infty} \#\{\mathrm{x} \in \tilde{\Omega}_n\colon \sum_{j=0}^{n-1} r(2^j x) \le t\}$$

as in sec. 6, then there is no functional equation for $\tilde{N}(t)$ that can be rewritten as a renewal equation. Conditioning on the value of x_1 of the first coordinate in the binary expansion just doesn't get you anywhere, because $r(x)$ isn't a function of only x_1.

So what we need is a "renewal theory" suitable for dealing with counting functions involving sums like $\sum_{j=0}^{n-1} r(\sigma^j x)$, where σ is the forward shift on a sequence space. First, some definitions and notation:

$$\mathcal{A} = \{1, 2, \ldots, \ell\} \quad \text{(the alphabet);} \tag{7.1}$$

$$A(i,j)\text{:} \quad \text{irreducible, aperiodic, 0–1 matrix on } \mathcal{A} \times \mathcal{A}; \tag{7.2}$$

$$\begin{aligned} \Omega = \{&\text{finite or infinite sequences } \mathbf{x} = x_1 x_2 \ldots x_m \text{ or} \\ &\mathbf{x} = x_1 x_2 \ldots \text{such that } A(x_i, x_{i+1}) = 1 \ \forall\ i\}; \end{aligned} \tag{7.3}$$

$$\begin{aligned} \sigma\colon \ \Omega \to \Omega\colon \ &\text{the shift, defined by} \\ &\sigma\colon \ x_1 x_2 \ldots \to x_2 x_3 \ldots \\ &\sigma\colon \ x_1 x_2 \ldots x_m \to x_2 x_3 \ldots x_m \\ &\sigma\colon \ e \to e \quad (e = \text{ empty sequence}); \end{aligned} \tag{7.4}$$

$$\begin{aligned} d = d_\rho\colon \quad &\text{metric on } \Omega \text{ defined by} \\ &d(\mathbf{x}, \mathbf{y}) = \rho^n, n = \min\{i\colon \ x_i \neq y_i\}, 0 < \rho < 1; \\ r, g\colon \ &\Omega \to \mathbb{R} \text{ Lipschitz continuous (some } \rho); \end{aligned} \tag{7.5}$$

$$S_n r(\mathbf{x}) = r(\mathbf{x}) + r(\sigma\mathbf{x}) + \ldots + r(\sigma^{n-1}\mathbf{x}), S_0 r \equiv 0; \tag{7.6}$$

$$N_g(t, \mathbf{x}) = \sum_{n=0}^{\infty} \sum_{\substack{\mathbf{y} \in \Omega: \\ \sigma^n \mathbf{y} = \mathbf{x}}} g(\mathbf{y}) 1\{S_n r(\mathbf{y}) \leq t\}. \tag{7.7}$$

Call $N_g(t, \mathbf{x})$ the *renewal function.* To guarantee that there are only finitely many nonzero terms in the sum defining $N_g(t, x)$, we make the following

ASSUMPTION: $\exists\, n \geq 1$ *such that* $S_n r(\mathbf{x}) > 0$ *for all infinite sequences* $\mathbf{x}$.

As in classical renewal theory there is a difference between the arithmetic and nonarithmetic cases. Say that r is *nonarithmetic* if there is no function $\tilde{r}$ valued in a discrete subgroup of $\mathbb{R}$ such that $r - \tilde{r}$ is cohomologous to zero (i.e., $r - \tilde{r} = h - h \circ \sigma$ for some continuous h).

THEOREM 3: *Assume that r is nonarithmetic and that $S_n r(\mathbf{x}) > 0$ $\forall$ infinite sequences $\mathbf{x}$, for some $n \geq 1$. Then there exist a constant $\delta > 0$, a probability measure $\nu(d\mathbf{x})$ on Ω, and a positive, Lipschitz continuous h: $\Omega \to \mathbb{R}$ such that for every continuous g: $\Omega \to \mathbb{R}$,*

$$N_g(t, \mathbf{x}) \sim e^{\delta t} h(\mathbf{x}) \left\{ \frac{\int g d\nu}{\int r h d\nu} \right\} \text{ as } t \to \infty, \tag{7.8}$$

uniformly for $\mathbf{x} \in \Omega$.

The relation (7.8) not only tells you how many sequences $\mathbf{y}$ there are such that $\sigma^n \mathbf{y} = x$ and $S_n r(\mathbf{y}) \leq t$ for some n; it also tells you how they are distributed in Ω. Fix $\mathbf{x} \in \Omega$, and define $\mu_t^{\mathbf{x}}$ to be the uniform distribution on $\bigcup_{n=1}^{\infty} \{\mathbf{y} \in \Omega\colon \sigma^n \mathbf{y} = \mathbf{x} \text{ and } S_n r(\mathbf{y}) \leq t\}$. Thus for any continuous function g: $\Omega \to \mathbb{R}$,

$$\int g d\mu_t^{\mathbf{x}} = N_g(t, \mathbf{x}) / N_1(t, \mathbf{x}),$$

and according to (7.8), this converges to $\int g d\nu$ as $t \to \infty$. So we have

COROLLARY: *As* $t \to \infty$, $\mu_t^{\mathbf{x}} \xrightarrow{w*} \nu$.

There is also a theorem that explains how $N_g(t; \mathbf{x})$ grows when r is arithmetic. Let's not go into this here (for the whole sordid story, see [7]). In the most interesting applications, r is nonarithmetic. Establishing that r is nonarithmetic is usually difficult, though.

You will probably want to know something about the quantities δ, $h(\mathbf{x})$, and $\nu(d\mathbf{x})$ which occur in (7.8). For this you need a dose of Perron-Frobenius theory.

RUELLE'S P–F THEOREM: *For each $\theta \in \mathbb{R}$ there exist $\lambda_\theta > 0$, a Lipschitz continuous h_θ: $\Omega^* \to (0, \infty)$, and a Borel probability measure $\nu_\theta(dx)$ on Ω^* such that*

$$\lambda_\theta h_\theta(\mathbf{x}) = \sum_{\mathbf{y}:\, \sigma \mathbf{y} = \mathbf{x}} e^{\theta r(\mathbf{y})} h_\theta(\mathbf{y}) \ \forall\, \mathbf{x} \in \Omega^* \tag{7.9}$$

and

$$\lambda_\theta \int g(\mathbf{x})\nu_\theta(d\mathbf{x}) = \int \{ \sum_{\mathbf{y}:\ \sigma\mathbf{y}=\mathbf{x}} e^{\theta r(\mathbf{y})} g(\mathbf{y})\} \nu_\theta(d\mathbf{x}) \tag{7.10}$$

for all continuous functions $g\colon \Omega^* \to \mathbb{R}$.

NOTE: $\Omega^* = \{$*infinite* sequences $x_1x_2\ldots \in \Omega\}$.

If $S_n r > 0$ on Ω^* for some n then λ_θ is a strictly increasing, continuous function of θ such that $\lambda_0 > 1$ and $\lambda_\theta \to 0$ as $\theta \to -\infty$ (proof: see [7]). Define $\delta > 0$ to be the unique real number such that

$$\lambda_{-\delta} = 1; \tag{7.11}$$

also

$$\nu = \nu_{-\delta}, \tag{7.12}$$

$$h = h_{-\delta}. \tag{7.13}$$

PROBLEM: Show that there is a unique continuous extension of h_θ from Ω^* to Ω such that (7.9) holds $\forall\, \mathbf{x} \in \Omega$.

Theorem 3 doesn't look much like a "probabilistic" theorem at all. But this is only because we haven't renormalized the renewal function $N_g(t, \mathbf{x})$. In fact, the theorem says the same thing about the "random walk" $S_n r(\mathbf{x})$ under the probability measure $h(x)\nu(dx)/\int h d\nu$ that the classical renewal theorem says about random walks with positive, independent increments. The novelty is that the increments in $S_n r(\mathbf{x})$ aren't independent. This makes it impossible to deduce Th. 3 from the classical renewal theorem.

If you're not a probabilist you might think all this discussion about whether or not Th. 3 has a probabilistic interpretation is just so much hot air. But it isn't. The reason is that the random walk $S_n r(\mathbf{x})$ behaves a lot like a random walk with independent increments: there is a law of large numbers, a central limit theorem, large deviation theorems, and so on. These have some important ramifications in counting problems. We'll give just one example, an analogue of Lemma 3 in sec. 4.

Define

$$\mu(d\mathbf{x}) = h(\mathbf{x})\nu(d\mathbf{x});$$
$$\overline{g} = \int g d\mu, \quad \forall\, g \in C(\Omega);$$
$$F^{\mathbf{x},t} = \{\mathbf{y}\colon \sigma^n \mathbf{y} = \mathbf{x} \text{ and } S_n r(\mathbf{x}) \leq t, \text{ some } n \geq 0\};$$
$$P^{\mathbf{x},t} = \text{ uniform distribution on } F^{\mathbf{x},t};$$
$$n(\mathbf{y}) = \{\max n\colon \sigma^n \mathbf{y} = \mathbf{x} \text{ and } S_n r(\mathbf{x}) \leq t\}, \ \mathbf{y} \in F^{\mathbf{x},t}.$$

THEOREM 4: *Under the hypotheses of Th. 3, for each $g \in C(\Omega)$ and each $\epsilon > 0$,*

$$P^{\mathbf{x},t}\{\mathbf{y}\colon |\frac{S_{n(\mathbf{y})}g(\mathbf{y})}{n(\mathbf{y})} - \overline{g}| \leq \epsilon\} \to 1 \textit{ and}$$
$$P^{\mathbf{x},t}\{\mathbf{y}\colon |\frac{a}{n(\mathbf{y})} - \overline{r}| \leq \epsilon\} \to 1$$

as $t \to \infty$, uniformly for $\mathbf{x} \in \Omega$.

In other words, "nearly all" $\mathbf{y}$ in $F^{\mathbf{x},t}$ have the property that $\mathbf{y}, \sigma\mathbf{y}, \sigma^2\mathbf{y}, \ldots, \sigma^{n(\mathbf{y})}(\mathbf{y})$ are distributed approximately as μ in Ω.

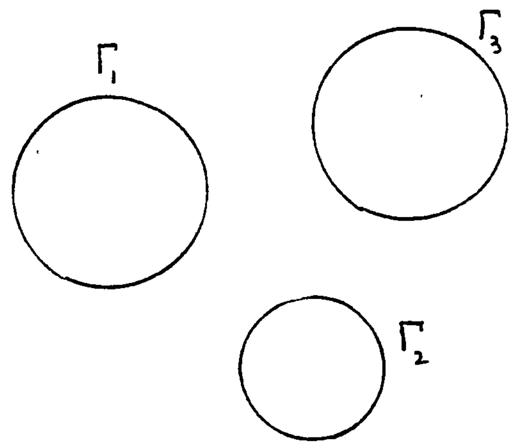
8. Reflection Groups

Now we'll discuss a problem very similar to the noneuclidean lattice point problem, which we mentioned earlier. This concerns reflection groups. Given a circle Γ with, say, center at x_0 and radius r, the reflection T in Γ is

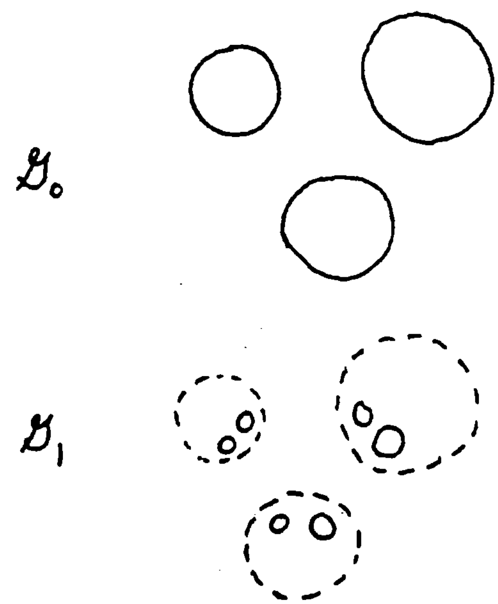
$$Tx = x_0 + (r/|x - x_0|)^2(x - x_0), \ x \in \mathbb{C} \cup \{\infty\}.$$

The points Tx and x are on the same ray emanating from x_0, and distance (x_0, x) and distance (x_0, Tx) multiply to give r^2. So T maps the inside of Γ onto the outside of Γ, and vice versa, and $T \circ T =$ identity.

Take a collection of mutually exterior circles, say three for simplicity, labelled $\Gamma_1, \Gamma_2, \Gamma_3$. Let T_i be the reflection in Γ_i. We are going to be interested in the group G generated by T_1, T_2, T_3. If we were interested in pursuing the lattice-point problem in general, we would consider finitely generated discrete groups of hyperbolic isometries (without parabolic elements), and things would work out in pretty much the same way. But G has a simple structure, the "symbolic dynamics" is very straightforward, and the action of G on $\mathbb{C}$ is easy to visualize, so we'll limit ourselves to G.



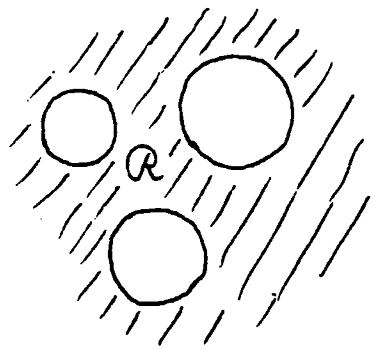
You can begin to get a picture of how G acts on $\mathbb{C}$ by looking at what happens to $\Gamma_1, \Gamma_2, \Gamma_3$. For any $T \in G$, $T\Gamma_i$ is a circle, because each $T \in G$ is a conformal mapping. The circles $\{T\Gamma_i\colon\ T \in G, i = 1,2,3\}$ may be arranged by "generations":



$$\mathcal{G}_0 = \{\Gamma_1, \Gamma_2, \Gamma_3\}$$
$$\mathcal{G}_1 = \{T_i\Gamma_j\colon\ i \neq j\}$$
$$\mathcal{G}_2 = \{T_{i_1}T_{i_2}\Gamma_j\colon\ i_1 \neq i_2, i_2 \neq j\}$$
$$\vdots$$
$$\mathcal{G}_n = \{T_{i_1}\ldots T_{i_n}\Gamma_j\colon\ i_k \neq i_{k+1}, i_n \neq j\}$$

The n^{th} generation circles are mutually exterior, each n^{th} generation circle is contained in an $(n-1)^{th}$ generation circle, and each n^{th} generation circle contains exactly two $(n+1)^{th}$ generation circles (proof by induction). It follows that each $T \in G$ has a unique representation as $T_{i_1}T_{i_2}\ldots T_{i_n}, n \geq 0, i_j \neq i_{j+1}$.

The circles $\{T\Gamma_i\colon\ T \in G,\ i = 1,2,3\}$ chop up $\mathbb{C}$ into countably many disjoint, connected, open sets. If $\mathcal{R}$ is the region exterior to Γ_1, Γ_2, and Γ_3 then the regions $T\mathcal{R}, T \in G$ are precisely the connected comonents of $\mathbb{C} - \bigcup_{i=1}^{3} \bigcup_{G} T\Gamma_i$. If $\xi \in \mathrm{R}$ then the orbit $G\xi$ of ξ has one point in each $T\mathrm{R}$, $T \in G$. Define



$$\wedge = \{\text{accumulation points of } G\xi\}, \tag{8.1}$$

$$M(\epsilon) = \#\{T \in G\colon \text{ distance } (T\xi, \wedge) \geq \epsilon\}. \tag{8.2}$$

PROBLEM: *How does* $M(\epsilon)$ *behave as* $\epsilon \downarrow 0$?

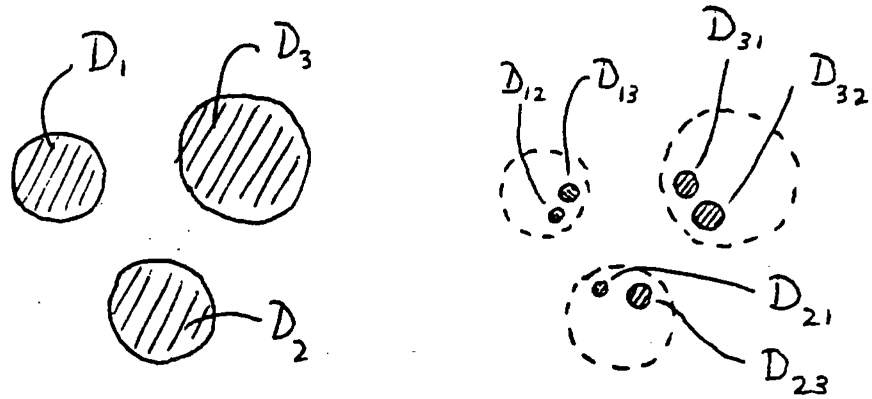
NOTE: distance in (8.2) means Euclidean distance.

This problem should be highly reminiscent of the problem posed in sec. 2 about semigroups of contractive similarity transformations. But there is an important difference. The elements $T \in G$ do *not* scale down distances exactly, as the similarities did. Consequently, you don't get a simple functional equation for $M(\epsilon)$ as we did in sec. 5. Nevertheless, there is "local" self-similarity, because every $T \in G$ is conformal, and this will, in the end, save the day.

As we mentioned earlier, there is a very simple "symbolic dynamics" associated with the action of G on $\mathbb{C}$. We have already seen that there is a $1-1$ correspondence between the set Ω' of *finite* sequences $i_1 i_2 \ldots i_n$ of 1's, 2's, and 3's satisfying $i_j \neq i_{j+1}$ and the group G, given by

$$i_1 i_2 \ldots i_n \to T_{i_1} T_{i_2} \ldots T_{i_n},\ \eta \to \text{ identity}$$

(η is the empty sequence). Let $\Omega^* = \{$infinite sequences $i_1 i_2 i_3 \ldots$ of 1's, 2's, 3's with $i_j \neq i_{j+1}\ \forall j\}$. For each $i_1 i_2 \ldots i_{n+1} \in \Omega'$, $n \geq 0$, define $D_{i_1 i_2 \ldots i_{n+1}}$ to be the closed disc interior to $T_{i_1} T_{i_2} \ldots T_{i_n} \Gamma_{i_{n+1}}$.



For each sequence $i_1 i_2 i_3 \ldots \in \Omega^*$ the discs $D_{i_1 \ldots i_n}$ are nested, i.e., $D_{i_1} \supset D_{i_1 i_2} \supset D_{i_1 i_2 i_3} \supset \ldots$, and therefore have a nonempty intersection. (Recall that the circles $T_{i_1 i_2 \ldots i_n} \Gamma_{i_{n+1}}$ are nested by generations.) Also, $\exists$ constants $C < \infty$ and $0 < \rho < 1$ such that for every $i_1 i_2 i_3 \ldots \in \Omega^*$ and every $n \geq 1$,

(8.3) $$\text{diameter } (D_{i_1 i_2 \ldots i_n}) \leq C \rho^n .$$

This is because each T_i shrinks distances outside of Γ_i. So

$$k_{i_1 i_2 \ldots} \triangleq \bigcap_{n \geq 1} D_{i_1 i_2 \ldots i_n}$$

is a single point. This point $k_{i_1 i_2 \ldots}$ lies in $\wedge$, because it is the limit of the sequence

$$k_{i_1 i_2 \ldots i_n} \triangleq T_{i_1} T_{i_2} \ldots T_{i_n} \xi ,$$

since $k_{i_1 i_2 \ldots i_n} \in D_{i_1 i_2 \ldots i_n}$. Finally, distinct sequences $i_1 i_2 i_3 \ldots$ and $i'_1 i'_2 i'_3 \ldots$ in Ω^* correspond to distinct points $k_{i_1 i_2 \ldots}$ and $k_{i'_1 i'_2 \ldots}$ of $\wedge$, because if $i_1 i_2 \ldots i_n \neq i'_1 i'_2 \ldots i'_n$ then the circles $T_{i_1} T_{i_2} \ldots T_{i_{n-1}} \Gamma_{i_n}$ and $T_{i'_1} T_{i'_2} \ldots T_{i'_{n-1}} \Gamma_{i'_n}$ are mutually exterior, and

$$k_{i_1 i_2 \ldots} \in D_{i_1 i_2 \ldots i_n} = \text{ interior } T_{i_1} T_{i_2} \ldots T_{i_{n-1}} \Gamma_{i_n} ,$$
$$k_{i'_1 i'_2 \ldots} \in D_{i'_1 i'_2 \ldots i'_n} = \text{ interior } T_{i'_1} T_{i'_2} \ldots T_{i'_{n-1}} \Gamma_{i'_n} .$$

So now we have $1-1$ correspondences

$$\Omega' \leftrightarrow G \leftrightarrow G\xi \qquad \text{and}$$
$$\Omega^* \leftrightarrow \wedge$$

given by

$$i_1 i_2 \dots i_n \to T_{i_1} T_{i_2} \dots T_{i_n} \to k_{i_1 i_2 \dots i_n} \qquad \text{and}$$

$$i_1 i_2 i_3 \dots \to k_{i_1 i_2 i_3 \dots}.$$

The counting function $M(\epsilon)$ defined by (8.2) can now be pulled back to the sequence space $\Omega = \Omega' \cup \Omega^*$, where there is some hope of using Th. 3. Define

$$\begin{aligned} r(i_1 i_2 \dots i_n) &= -\log \left\{ \frac{\text{distance } (k_{i_1 i_2 \dots i_n}, \wedge)}{\text{distance } (k_{i_2 i_3 \dots i_n}, \wedge)} \right\}, \\ r(i_1 i_2 \dots) &= -\log |T'_{i_1}(k_{i_2 i_3 \dots})|, \\ r(\eta) &= 0. \end{aligned}$$

Here $|T_i'(z)|$ is the local expansion factor for the mapping T_i at the point z (recall that T_i is conformal, but orientation-reversing). Because of (8.3), r is Lipschitz continuous on Ω, and because $k_{i_2 i_3 \dots}$ is outside Γ_{i_1}, $r(i_1 i_2 \dots) > 0 \ \forall \ i_1 i_2 i_3 \dots \in \Omega^*$. And the most important point:

$$S_n r(i_1 i_2 \dots i_n) = \log \left\{ \frac{\text{distance } (k_{i_1 i_2 \dots i_n}, \wedge)}{\text{distance } (k_\eta, \wedge)} \right\}^{-1},$$

so

$$M(\epsilon) = N_1(-\log(\epsilon/\text{distance } (k_\eta, \wedge)), \eta)$$

where $N_1(t, \mathrm{x})$ is defined by (7.7).

Before we can apply Theorem 3 we must verify that r is a nonarithmetic function. This isn't at all easy. In fact, it is equivalent to the fact that the geodesic flow on the unit tangent bundle of H^3/G^* is topologically mixing, where H^3 is hyperbolic 3-space and G^* is the group of H^3-isometries induced by $\{T_{i_1} T_{i_2} \dots T_{i_n}\colon\ n \text{ even}\}$. See [7], sec. 11, and [13] for details.

Theorem 3 now implies that $M(\epsilon) \sim C\epsilon^{-\delta}$ as $\epsilon \downarrow 0$ for some $\delta > 0$ and $0 < C < \infty$. With some additional work, it can be shown that δ is the Hausdorff dimension of $\wedge$, and that the measure on $\wedge$ induced by ν is the normalized δ-dimensional Hausdorff measure on $\wedge$. (See [3] for a similar calculation.) In summary,

THEOREM 5: *As* $\epsilon \downarrow 0$,

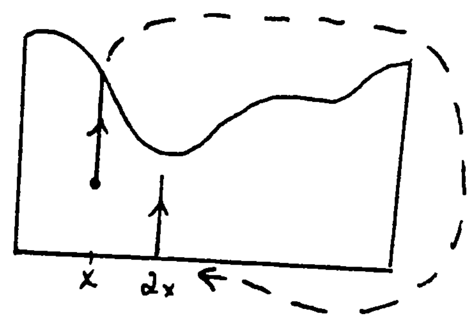
$$M(\epsilon) \sim C\epsilon^{\delta}$$

where δ is the Hausdorff dimension of $\wedge$ and $0 < C < \infty$. If P^ϵ is the uniform distribution on the set of $x \in G\xi$ such that distance $(x, \wedge) \geq \epsilon$, then as $\epsilon \downarrow 0$

$$P^\epsilon \xrightarrow{w^*} \textit{normalized } \delta\textit{-Hausdorff measure on } \wedge\,.$$

9. Periodic Orbits of Suspension Flows

We return now to the suspension flow under under $r(x)$, which we introduced in sec. 7. Remember that this flow works just like the Bernoulli flow — the only difference is the "ceiling" of the region on which the flow takes place. This ceiling is the graph of the function $r(x)$. Assume that r is strictly positive and Lipschitz continuous on $[0, 1]$. Assume also that the suspension flow is topologically mixing.



The suspension flow under r is a more important gadget than you might at first think. This is because a large class of hyperbolic flows, the "Axiom A" flows, can be "represented" by suspension flows over shifts of finite types. I don't want to explain precisely what I mean by "representation" here. You will probably hear something about this in some of the other lectures (or see [14], [1]). Suffice it to say that counting problems for periodic orbits of Axiom A flows can be (and have been) reduced to counting problems for periodic orbits of suspension flows.

So let's think about periodic orbits of the suspension flow under r. We have seen that the periodic orbits of this flow are the orbits that go through points $(x, 0)$ such that x has a periodic binary expansion. If x has a periodic binary expansion with (least) period n then the periodic orbit through $(x, 0)$ has (least) period $S_n r(x) = r(x) + r(\sigma x) + \ldots + r(\sigma^{n-1} x)$ (here σ represents the shift on the binary expansion). Also, this periodic orbit goes through

$(x,0),(\sigma x,0),(\sigma^2 x,0),\ldots,(\sigma^{n-1}x,0)$, but no other points $(y,0)$. Define

$$\mathcal{P}_n = \{\text{sequences } x_1x_2\ldots \text{ of 0's and 1's with least period } n\};$$
$$\overline{\mathcal{P}}_n = \{\text{sequences } x_1x_2\ldots \text{ of 0's and 1's with period } n\};$$
$$M(t) = \sum_{n=1}^{\infty}\frac{1}{n}\sum_{\mathbf{x}\in\mathcal{P}_n} 1\{S_n r(\mathbf{x}) \leq t\};$$
$$\overline{M}(t) = \sum_{n=1}^{\infty}\frac{1}{n}\sum_{\mathbf{x}\in\overline{\mathcal{P}_n}} 1\{S_n r(\mathbf{x}) \leq t\}. \tag{9.1}$$

Then $M(t)$ is the number of periodic orbits of the suspension flow with (least) period $< t$. NOTE: we won't bother to distinguish between sequences $\mathbf{x}$ of 0's and 1's and points x of [0,1).

THEOREM 6: *As* $t \to \infty$,

$$M(t) \sim e^{\delta t}/(\delta t) \tag{9.2}$$

where δ *is the topological entropy of the flow.*

This theorem is essentially the same as the main result of Parry and Pollicott [11], the "Prime Number Theorem" for periodic orbits of an Axiom A flow. Their proof is quite a bit different than the one you're going to see here, and, perhaps, more natural, following one of the standard proofs of the prime number theorem. But the approach taken here leads also, via Theorem 4, to some additional results about the distribution of a "typical" periodic orbit.

To prove Th. 6 it suffices to show that (9.2) holds when $M(t)$ is replaced by $\overline{M}(t)$. You already have seen the argument for this, in sec. 6, so we won't go through it again.

The sum defining $\overline{M}(t)$ looks vaguely similar to that defining $N(t,\mathbf{x})$ in (7.7). There are two differences. First, the factor $\frac{1}{n}$ does not appear in (7.7). The same difficulty arose in sec. 6, and we dealt with it by using Lemma 3; we will be able to deal with it here in much the same way, using Theorem 4 in place of Lemma 3. Second, in (7.7) the sum is over those $\mathbf{y}$ such that $\sigma^n\mathbf{y} = \mathbf{x}$, whereas in (9.1) it is over $\mathbf{y} \in \mathcal{P}_n$. This problem didn't arise in our analysis of the Bernoulli flow (sec. 6). To deal with it we proceed as follows.

Choose a large integer k, and choose infinite sequences $\mathbf{x}^{(i)} = x_1^{(i)} x_2^{(i)} \ldots, i = 1, 2, \ldots, 2^k$ such that the finite sequences $x_1^{(i)} x_2^{(i)} \ldots x_k^{(i)}$, $i = 1, 2, \ldots, 2^k$ are the 2^k distinct sequences of 0's and 1's of length k. Define

$$g_i(\mathbf{x}) = 1\{x_j = x_j^{(i)} \ \forall\, j = 1, 2, \ldots, k\};$$
$$N_n^{(i)}(t) = \sum_{\mathbf{x}:\ \sigma^n \mathbf{x} = \mathbf{x}^{(i)}} g_i(\mathbf{x}) 1\{S_n r(\mathbf{x}) \le t\}.$$

LEMMA 4: *For each $n \ge k$ and $t \ge 0$*

$$\sum_{i=1}^{2^k} N_n^{(i)}(t - \epsilon_k) \le \sum_{\mathbf{x} \in \overline{\mathcal{P}}_n} 1\{S_n r(\mathbf{x}) \le t\} \le \sum_{i=1}^{2^k} N_n^{(i)}(t + \epsilon_k) \tag{9.3}$$

where $\epsilon_k = 2^{-k+1} C$ and C is a Lipschitz constant for r on $[0,1]$, i.e., $|r(x) - r(y)| \le C|x - y| \ \forall\, x, y \in [0,1]$.

Some comments: $N_n^{(i)}(t)$ is precisely the type of sum that occurs in (7.7). The fact that (9.3) may not hold for $n < k$ doesn't matter, because for t sufficiently large the terms $n < k$ in (9.1) add up to *zero*. Observe that $\epsilon_k \to 0$ as $k \to \infty$.

PROOF *of Lemma 4*: For every sequence $\mathbf{x} \in \overline{\mathcal{P}}_n$ there is one and only one $i \in \{1, 2, \ldots, 2^k\}$ such that $g_i(\mathbf{x}) = 1$. Consider the sequence $\tilde{\mathbf{x}}$ given by

$$\tilde{x}_j = x_j \quad \text{for } j \le n + k,$$
$$\tilde{x}_j = x_j^{(i)} \quad \text{for } j > n + k.$$

You have

$$\sigma^n \tilde{\mathbf{x}} = \mathbf{x}^{(i)}, \tag{9.4}$$

$$g_i(\tilde{\mathbf{x}}) = g_i(\mathbf{x}) = 1, \text{ and} \tag{9.5}$$

$$\begin{aligned} |S_n r(\tilde{\mathbf{x}}) - S_n r(\mathbf{x})| &\le \sum_{j=0}^{n-1} |r(\sigma^j \tilde{\mathbf{x}}) - r(\sigma^j \mathbf{x})| \\ &\le C \sum_{j=0}^{\infty} 2^{-k+j} = \epsilon_k. \end{aligned} \tag{9.6}$$

Conversely, if $\tilde{\mathbf{x}}$ is a sequence such that $\sigma^n \tilde{\mathbf{x}} = \mathbf{x}^{(i)}$ and $g_i(\tilde{\mathbf{x}}) = 1$ then you can obtain a sequence $\mathbf{x} \in \overline{\mathcal{P}}_n$ by

$$\begin{aligned} x_j &= \tilde{x}_j \text{ for } j \le n, \\ x_j &= x_{j-n} \text{ for } n < j \le 2n, \\ &\ldots, \end{aligned}$$

and for this sequence $\mathbf{x}$, (9.4)–(9.6) still hold. So there is a $1-1$ correspondence between $\mathbf{x} \in \overline{\mathcal{P}}_n$ and sequences $\tilde{\mathbf{x}}$ satisfying (9.4)–(9.5) for some i, and corresonding $\mathbf{x}$, $\tilde{\mathbf{x}}$ satisfy (9.6). The inequality (9.3) is a direct consequence. □

Theorem 3 only applies when $r(\mathbf{x})$ is a nonarithmetic function. But we assumed that the suspension flow under r is mixing; this implies that $r(\mathbf{x})$ is nonarithmetic.

EXERCISE: A flow T_t on a compact metric space X is topologically mixing provided there is no continuous $f: X \to \mathbb{C}$ such that $f \circ T_t = e^{i\theta t} f \ \forall\, t \ge 0$, some $\theta \in \mathbb{R}$. Use this characterization to show that if the suspension flow under r is topologically mixing, then $r(\mathbf{x})$ is nonarithmetic.

So now Theorem 3 applies, and tells you that

$$\begin{aligned} N^{(i)}(t) &\triangleq \sum_{n=1}^{\infty} N_n^{(i)}(t) \\ &= \sum_{n=1}^{\infty} \sum_{\mathbf{x}:\ \sigma^n \mathbf{x} = \mathbf{x}^{(i)}} g_i(\mathbf{x}) 1\{S_n r(\mathbf{x}) \le t\} \\ &\sim C_i e^{\delta t}, \end{aligned}$$

with

$$C_i = \frac{(\int g_i d\nu) h(\mathbf{x}^{(i)})}{\delta \int (fh) d\nu}.$$

Notice that

$$\begin{aligned} \sum_{i=1}^{2^k} C_i &= \frac{\sum_{i=1}^{k} h(\mathbf{x}^{(i)}) \nu\{\mathbf{x}:\ x_j = x_j^{(i)} \ \forall\, 1 \le j \le k\}}{\delta \int (fh) d\nu} \\ &\approx \int h d\nu / (\delta \int (fh) d\nu) \end{aligned}$$

when k is large. Theorem 4 guarantees that the primary contribution to $N^{(i)}(t)$ comes from those $N_n^{(i)}(t)$ for which

$$|n - t/\overline{f}| < \epsilon t.$$

By letting $\epsilon \downarrow 0$, you get

$$\sum_{i=1}^{2^k} \sum_{n=1}^{\infty} \frac{1}{n} N_n^{(i)}(t) \sim (\sum_{i=1}^{2^k} C_i) e^{\delta t} \overline{f}/t \tag{9.8}$$

EXERCISE: Fill in the details in this last step.

Putting (9.3), (9.7), and (9.8) together gives

$$\begin{aligned}\overline{M}(t) &= \sum_{n=1}^{\infty} \frac{1}{n} \sum_{\mathrm{x} \in \overline{\mathcal{P}}_n} 1\{S_n r(\mathrm{x}) \leq t\} \\ &\sim \left(\frac{e^{\delta t}}{t}\right) \frac{\overline{f} \int h d\nu}{\delta \int f h d\nu} = \frac{e^{\delta t}}{\delta t}\end{aligned}$$

This proves (9.2), except for the identification of δ as the topological entropy of the flow.□

How do periodic orbits distribute themselves on the suspension space $\Gamma_r = \{(x, s)\colon 0 \leq x < 1 \text{ and } 0 \leq s \leq r(x)\}$? Take a continuous function $G\colon \Gamma_r \to \mathbb{R}$, and for a periodic orbit τ define $\tau(G)$ to be the integral of G over (a single period of) τ. The distribution of τ on Γ_r is determined by the integrals $\tau(G)$ for all continuous G.

Define a probability measure Q on Γ_r by

$$Q(d(x,s)) = ds\mu(d\mathrm{x}) / \int r(\mathrm{y}) \mu(d\mathrm{y}) \text{ where}$$

$$\mu(d\mathrm{x}) = h(\mathrm{x}) \nu(d\mathrm{x}) / \int h(\mathrm{y}) \nu(d\mathrm{y})$$

as in sec. 7. Since μ is an invariant probability measure for the shift, Q is invariant for the flow. (EXERCISE: Prove this.) For $\epsilon > 0$ set

$$\begin{aligned}M_\epsilon(t) = {} & \text{number of periodic orbits } \tau \text{ with} \\ & \text{period } \leq t \text{ and } \left|\frac{\tau(G)}{\tau(1)} - \int G dQ\right| > \epsilon.\end{aligned}$$

THEOREM 7: $\lim_{t \to \infty} (M_\epsilon(t)/M(t)) = 0, \ \forall\, \epsilon > 0.$

What this means is that for large t, nearly all periodic orbits with period $\leq t$ are such that the average value of G on the orbit is approximately $\int G dQ$. More simply, most periodic orbits are distributed on Γ_r approximately (in the weak * topology) like Q. There are even more sophisticated results than this — see [8] — but I won't go into these here. I should mention also that Q is the maximum entropy measure for the flow (see [11]).

PROBLEM: Using Theorem 4, prove Th. 7.

You can find the solution in [7], sec. 5.

REFERENCES

[1] Bowen, R. (1973) Symbolic dynamics for hyperbolic flows. *Amer. J. Math.* **95**.

[2] Bowen, R. (1975) Springer *Lecture Notes in Math* **470**.

[3] Bowen, R. (1979) Hausdorff dimension of quasi-circles. *Publ. I.H.E.S.* **50**.

[4] Breiman, L. (19??). *Probability*. Addison-Wesley.

[5] Feller, W. (19??) *An Intro. to Prob. Th. and its Appl.*, vol. 2. Wiley Press.

[6] Lalley, S. (1988) Packing and covering self-similar fractals. *Indiana U.. Math. J.* **37**.

[7] Lalley, S. (1989?) Renewal theorems in symbolic dynamics .... To appear, *Acta Mathematica.*

[8] Lalley, S. (1987) Distribution of periodic orbits .... *Adv. Appl. Math.* **8**.

[9] Lindvall, T. (1977) Probabilistic appraoch to the renewal theorem. *Annals of Prob.* **7**.

[10] Parry, W. (1979) Prime Number Theorem .... *Israel J. Math.*

[11] Parry, W. and Pollicot,, M. (1983) Prime Number Theorem .... *Ann. Math.* **118**.

[12] Rudin, W. (19??) *Functional Analysis.*

[13] Rudolph, D. (1982) in *Ergodic Th. and Dynamical Systems 2.*

[14] Series, C. (1981) Symbolic dynamics for geodesic flows. *Acta Mathematica* **146**.