Math 118, Spring 2,001

Shlomo Sternberg

January 11, 2001

Contents

Chapter 4

Space and time averages

4.1 histograms and invariant densities

Let us consider a map, $F : [0,1] \to [0,1]$, pick an initial seed, x_0, and compute its iterates, $x_0, x_1, x_2, \ldots, x_m$ under F. We would like to see which parts of the unit interval are visited by these iterates, and how often. For this purpose let us divide the unit interval up into N subintervals of size $1/N$ given by

$$I_k = \left[\frac{k-1}{N}, \frac{k}{N}\right), \; k = 1, \ldots, N-1, \quad I_N = \left[\frac{N-1}{N}, 1\right].$$

We count how many of the iterates $x_0, x_1, \ldots, x_m$ lie in I_k. Call this number n_k. There are $m+1$ iterates (starting with, and counting, x_0) so the numbers

$$p_k = \frac{n_k}{m+1}$$

add up to one:

$$p_1 + \cdots + p_N = 1.$$

We would like to think of these numbers as "probabilities" - the number p_k representing the "probability" that an iterate belongs to I_k. Strictly speaking, we should write $p_k(m)$. In fact, we should write $p_k(m, x_0)$ since the procedure depends on the initial seed, x_0. But the hope is that as m gets large the $p_k(m)$ tend to a limiting value which we denote by p_k, and that this limiting value will be independent of x_0 if x_0 is chosen "generically". We will continue in this vague, intuitive vein a while longer before passing to a precise mathematical formulation. If U is a union of some of the I_k, then we can write

$$p(U) = \sum_{I_k \subset U} p_k$$

and think of $p(U)$ as representing the "probability" that an iterate of x_0 belongs to U. If N is large, so the intervals I_k are small, every open set U can be closely approximated by a union of the I_k's, so we can imagine that the "probabilities", $p(U)$, are defined for all open sets, U. If we buy all of this, then we can write down an equation which has some chance of determining what these "probabilities", $p(U)$, actually are: A point $y = F(x)$ belongs to U if and only if $x \in F^{-1}(U)$. Thus the number of points among the $x_1, dots, x_{m+1}$ which belong to U is the same as the number of points among the $x_0, \ldots, x_m$ which belong to $F^{-1}(U)$. Since our limiting probability is unaffected by this shift from 0 to 1 or from m to $m+1$ we get the equation

$$p(U) = p(F^{-1}(U)). \tag{4.1}$$

To understand this equation, let us put it in a more general context. Suppose that we have a "measure", μ, which assigns a size, $\mu(A)$, to every open set, A. Let F be a continuous transformation. We then define the **push forward** measure, $F_*\mu$ by

$$(F_*\mu)(A) = \mu(F^{-1}(A)). \tag{4.2}$$

Without developing the language of measure theory, which is really necessary for a full understanding, we will try to describe some of the issues involved in the study of equations (4.2) and (4.1) from a more naive viewpoint. Consider, for example, $F = L_\mu, 1 < \mu < 3$. If we start with any initial seed other than $x_0 = 0$, it is clear that the limiting probability is

$$p(I_k) = 1,$$

if the fixed point,$1 - \frac{1}{\mu} \in I_k$ and

$$p(I_k) = 0$$

otherwise. Similarly, if $3 < \mu < 1 + \sqrt{6}$, and we start with any x_0 other than 0 or the fixed point, $1 - \frac{1}{\mu}$ then clearly the limiting probability will be $p(I) = 1$ if both points of period two belong to I, $p(I) = \frac{1}{2}$ if I contains exactly one of the two period two points, and $p(I) = 0$ otherwise. These are all examples of **discrete** measures in the sense that there is a finite (or countable) set of points, $\{z_k\}$, each assigned a positive number, $m(z_k)$ and

$$\mu(I) = \sum_{z_k \in I} m(z_k).$$

We are making the implicit assumption that this series converges for every bounded interval. The **integral** of a function, ϕ, with respect to the discrete measure, μ, denoted by $\langle \phi, \mu \rangle$ or by $\int \phi\mu$ is defined as

$$\int \phi\mu = \sum \phi(x_k) m(x_k).$$

This definition makes sense under the assumption that the series on the right hand side is absolutely convergent. The rule for computing the push forward,

$F_*\mu$ (when defined) is very simple. Indeed, let $\{y_l\}$ be the set of points of the form $y_l = F(x_k)$ for some k, and set

$$n(y_l) = \sum_{F(x_k)=y_l} m(x_k).$$

Notice that there is some problem with this definition if there are infinitely many points x_k which map to the same y_l. Once again we must make some convergence assumption. For example, if the map F is everywhere finite-to-one, there will be no problem. Thus the push forward of a discrete measure is a discrete measure given by the above formula.

At the other extreme, a measure is called **absolutely continuous** (with respect to Lebesgue measure) if there is an integrable function, ρ, called the **density** so that

$$\mu(I) = \int_I \rho(x)dx.$$

For any continuous function, ϕ we define the integral of ϕ with respect to μ as

$$\langle \phi, \mu \rangle = \int \phi\mu = \int \phi(x)\rho(x)dx$$

if the integral is absolutely convergent. Suppose that the map F is piecewise differentiable and in fact satisfies $|F'(x)| \neq 0$ except at a finite number of points. These points are called *critical points* for the map F and their images are called *critical values*. Suppose that A is an interval containing no critical values, and to fix the ideas, suppose that $F^{-1}(A)$ is the union of finitely many intervals, J_l each of which is mapped monotonically (either strictly increasing or decreasing) onto A. The change of variables formula from ordinary calculus says that for any function $g = g(y)$ we have

$$\int_A g(y)dy = \int_{J_k} g(F(x))|F'(x)|dx,$$

where $y = F(x)$. So if we set $g(y) = \rho(x)|1/F'(x)|$ we get

$$\int \rho(x)\frac{1}{|F'(x)|}dy = \int_{J_k} \rho(x)dx = \mu(J_k).$$

Summing over k and using the definition (4.2) we see that $F_*\mu$ has the density

$$\sigma(y) = \sum_{F(x_k)=y} \frac{\rho(x)}{|F'(x)|}. \qquad (4.3)$$

Equation (4.3) is sometimes known as the Perron Frobenius equation, and the transformation $\rho \mapsto \sigma$ as the Perron Frobenius operator.

Getting back to our histogram, if we expect the limit measure to be of the absolutely continuous type, so

$$p(I_k) \approx \rho(x) \times \frac{1}{N}, \quad x \in I_k$$

then we expect that

$$\rho(x) \approx \lim_{m\to\infty} \frac{n_k N}{m+1}, \quad x \in I_k$$

as the formula for the limiting density.

4.2 the histogram of L_4

We wish to prove the following assertions:
(i) *The measure, μ, with density*

$$\sigma(x) = \frac{1}{\pi\sqrt{x(1-x)}} \tag{4.4}$$

is invariant under L_4. In other words it satisfies

$$L_{4*}\mu = \mu.$$

(ii) *Up to a multiplicative constant, (4.4) is the only continuous density invariant under L_4*
(iii) *If we pick the initial seed generically, then the normalized histogram converges to (4.4).*

We give two proofs of (i). The first is a direct verification of the Perron Frobenius formula (4.3) with $y = F(x) = 4x(1-x)$ so $|F'(x)| = |F'(1-x)| = 4|1-2x|$. Notice that the σ given by (4.4) satisfies $\sigma(x) = \sigma(1-x)$ so (4.3) becomes

$$\frac{1}{\pi\sqrt{4x(1-x)(1-4x(1-x))}} = \frac{2}{\pi 4|1-2x|\sqrt{x(1-x)}}.$$

But this follows immediately from the identity

$$1 - 4x(1-x) = (2x-1)^2.$$

For our second proof, consider the tent transformation, T. For any interval, I contained in $[0,1]$, $T^{-1}(I)$ consists of the union of two intervals, each of half the length of I. In other words the ordinary Lebesgue measure is preserved by the tent transformation: $T_*\nu = \nu$ where ν has density $\rho(x) \equiv 1$. Put another way, the function $\rho(x) \equiv 1$ is the solution of the Perron Frobenius equation

$$\rho(Tx) = \frac{\rho(x)}{2} + \frac{\rho(1-x)}{2}. \tag{4.5}$$

It follows immediately from the definitions, that

$$(F \circ G)_*\mu = F_*(G_*\mu),$$

where F and G are two transformations , and μ is a measure. In particular, since $h \circ T = L_4 \circ h$ where

$$h(x) = \sin^2 \frac{\pi x}{2},$$

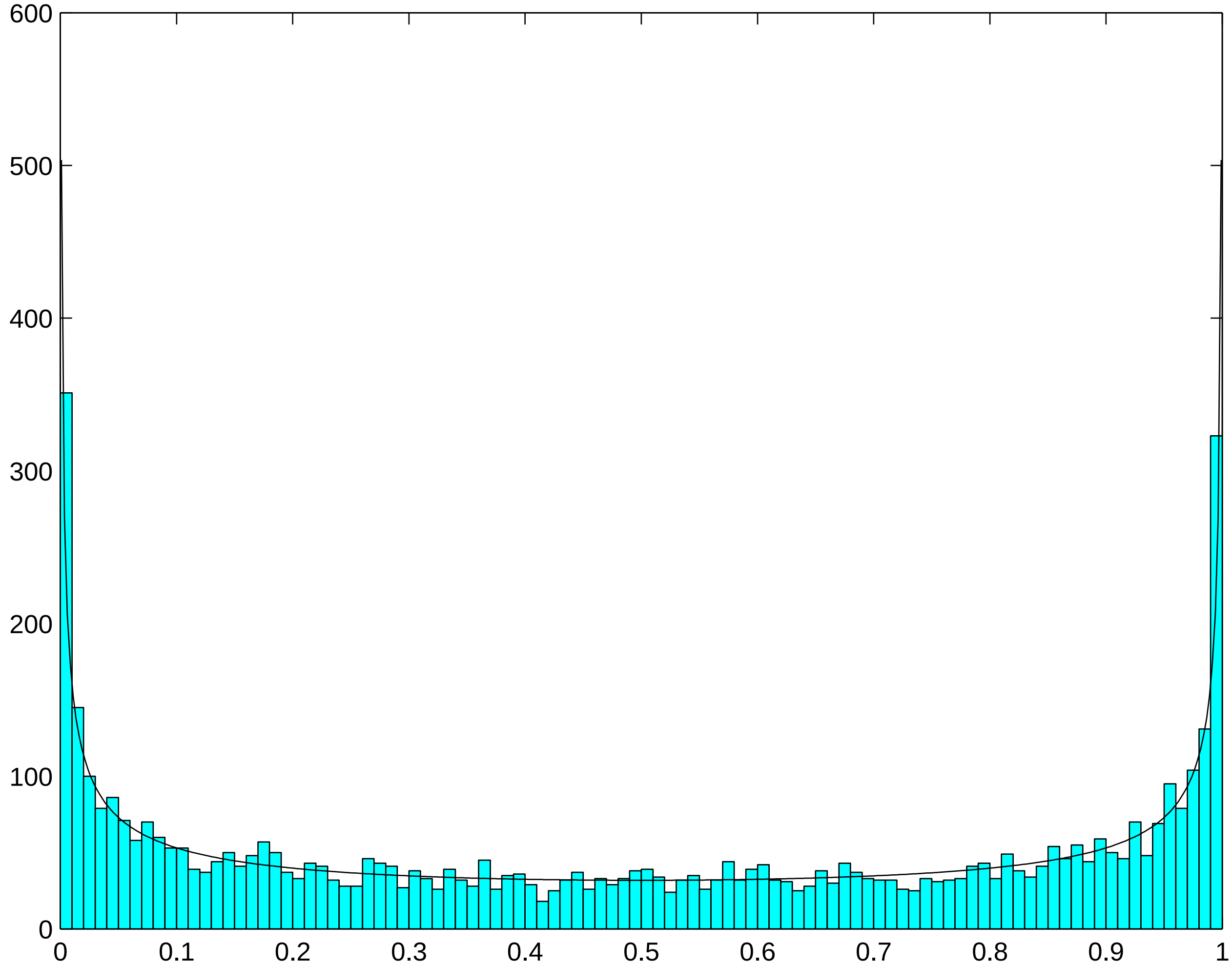

Figure 4.1: The histogram of iterates of L_4 compared with σ.

it follows that if $T_*\nu = \nu$, then $T_*(h_*\nu) = h_*\nu$. So to solve $L_{4*\mu} = \mu$, we must merely compute $h_*\nu$. According to (4.3) this is the measure with density

$$\sigma(y) = \frac{1}{|h'(x)|} = \frac{1}{\pi \sin \frac{\pi x}{2} \cos \frac{\pi x}{2}}.$$

But since $y = \sin^2 \frac{\pi x}{2}$ this becomes

$$\sigma(y) = \frac{1}{\pi\sqrt{y(1-y)}}$$

as desired.

To prove (ii), it is enough to prove the corresponding result for the tent transformation: that $\rho = \text{const.}$ is the only continuous function satisfying (4.5). To see this, let us consider the binary representation of T: Let

$$x = 0.a_1a_2a_3\dots$$

be the binary expansion of x. If $0 \le x < \frac{1}{2}$, so $a_1 = 0$, then $Tx = 2x$ or

$$T(0.0a_2a_3a_4\dots) = 0.a_2a_3a_4\dots.$$

If $x \ge \frac{1}{2}$, so $a_1 = 1$, then

$$T(x) = -2x + 2 = 1 - (2x-1) = 1 - S(x) = 1 - 0.a_2a_3a_4\dots.$$

Introducing the notation

$$\bar{0} = 1, \quad \bar{1} = 0,$$

we have

$$0.a_2a_3a_4\cdots + 0.\bar{a}_1\bar{a}_2\bar{a}_3\cdots = 0.1111\cdots = 1$$

so

$$T(0.1a_2a_3a_4\dots) = 0.\bar{a}_2\bar{a}_3\bar{a}_4\dots.$$

In particular, $T^{-1}(0.a_1a_2a_3\dots)$ consists of the two points

$$0.0a_1a_2a_3\dots \quad \text{and} \quad 0.1\bar{a}_1\bar{a}_2\bar{a}_3\dots.$$

Now let us iterate (4.5) with ρ replaced by f, and show that the only solution is $f = \text{constant}$. Using the notation $x = 0.a_1a_2\cdots = 0.a$ repeated application of (4.5) gives:

$$\begin{aligned} f(x) &= \frac{1}{2}[f(.0a) + f(.1\bar{a})] \\ &= \frac{1}{4}[f(.00a) + f(.01\bar{a}) + f(.10a) + f(.11\bar{a})] \\ &= \frac{1}{8}[f(.000a) + f(.001\bar{a}) + f(.010a) + f(.011\bar{a}) + f(.100a) + \cdots] \\ &\to \int f(t)dt. \end{aligned}$$

But this integral is a constant, independent of x. QED.

The third statement, (iii), about the limiting histogram for "generic" initial seed, x_0, demands a more careful formulation. What do we mean by the phrase "generic"? The precise formulation requires a dose of measure theory, the word "generic" should be taken to mean "outside of a set of measure zero with respect to μ". The usual phrase for this is "for almost all x_0". Then assertion (iii) becomes a special case of the famous Birkhoff ergodic theorem. This theorem asserts that for almost all points, p, the "time average"

$$\lim \frac{1}{n} \sum_{k=0}^{n-1} \phi(L_4^k p)$$

equals the "space average"

$$\int \phi\mu$$

for any integrable function, ϕ. Rather than proving this theorem, we will explain a simpler theorem, von Neumann's mean ergodic theorem, which motivated Birkhoff to prove his theorem.

Let F be a transformation with an invariant measure, μ. By this we mean that $F_*\mu = \mu$. We let H denote the Hilbert space of all square integrable functions with respect to μ, so the scalar product of $f, g \in H$ is given by

$$(f, g) = \int f\bar{g}\mu.$$

The map F induces a transformation $U : H \to H$ by

$$Uf = f \circ F$$

and

$$(Uf, Ug) = \int (f \circ F)(\overline{g \circ F})\mu = \int f\bar{g}\mu = (f, g).$$

In other words, U is an isometry of H. The mean ergodic theorem asserts that the limit of

$$\frac{1}{n} \sum_{0}^{n-1} U^k f$$

exists in the Hilbert space sense, "convergence in mean", rather than the almost everywhere pointwise convergence of the Birkhoff ergodic theorem. Practically by its definition, this limiting element $\hat{f}$ is invariant, i.e. satisfies $U\hat{f} = \hat{f}$. Indeed, applying U to the above sum gives an expression which differs from that sum by only two terms, f and $U^n f$ and dividing by n sends these terms to zero as $n \to \infty$. If, as in our example, we know what the possible invariant elements are, then we know the possible limiting values $\hat{f}$

The mean ergodic theorem can be regarded as a smeared out version of the Birkhoff theorem. Due to inevitable computer error, the mean ergodic theorem may actually be the version that we want.

4.3 The mean ergodic theorem

The purpose of this section is to prove

Theorem 4.3.1 von Neumann's mean ergodic theorem. *Let $U : H \to H$ be an isometry of a Hilbert space, H. Then for any $f \in H$, the limit*

$$\lim \frac{1}{n} \sum U^k f = \hat{f} \tag{4.6}$$

exists in the Hilbert space sense, and the limiting element $\hat{f}$ is invariant, i.e. $U\hat{f} = \hat{f}$.

Proof. The limit, if it exists, is invariant as we have seen. If U were a unitary operator on a finite dimensional Hilbert space, H, then we could diagonalize U, and hence reduce the theorem to the one dimensional case. A unitary operator on a one dimensional space is just multiplication by a complex number of the form $e^{i\alpha}$. If $e^{i\alpha} \neq 1$, then

$$\frac{1}{n}(1 + e^{i\alpha} + \cdots + e^{(n-1)i\alpha}) = \frac{1}{n} \frac{1 - e^{in\alpha}}{1 - e^{i\alpha}} \to 0.$$

On the other hand, if $e^{i\alpha} = 1$, the expression on the left is identically one. This proves the theorem for finite dimensional unitary operators. For an infinite dimensional Hilbert space, we could apply the spectral theorem of Stone (discovered shortly before the proof of the ergodic theorem) and this was von Neumann's original method of proof.

Actually, we can proceed as follows:

Lemma 4.3.1 *The orthogonal complement of the set, D, of all elements of the form $Ug - g$, consists of invariant elements.*

Proof. If f is orthogonal to all elements in D, then,in particular, f is orthogonal to $Uf - f$, so

$$0 = (f, Uf - f)$$

and

$$(Uf, Uf - f) = (Uf, Uf) - (Uf, f) = (f, f) - (Uf, f)$$

since U is an isometry. So

$$(Uf, Uf - f) = (f - Uf, f) = 0.$$

So

$$(Uf - f, Uf - f) = (Uf, Uf - f) - (f, Uf - f) = 0,$$

or

$$Uf - f = 0$$

which says that f is invariant.

So what we have shown, in fact, is

Lemma 4.3.2 *The union of the set D with the set, I, of the invariant functions is dense in H.*

In fact, if f is orthogonal to D, then it must be invariant, and if it is orthogonal to all invariant functions it must be orthogonal to itself, and so must be zero. So $(D \cup I)^{\perp} = 0$, so $D \cup I$ is dense in H.

Now if f is invariant, then clearly the limit(4.6) exists and equals f. If $f = Ug - g$, then the expression on the left in (4.6) telescopes into

$$\frac{1}{n}(U^n g - g)$$

which clearly tends to zero. Hence, as a corollary we obtain

Lemma 4.3.3 *The set of elements for which the limit in (4.6) exists is dense in H.*

Hence the mean ergodic theorem will be proved, once we prove

Lemma 4.3.4 *The set of elements for which the limit in (4.6) exists is closed.*

Proof. If

$$\frac{1}{n}\sum U^k g_i \to \hat{g}_i, \quad \frac{1}{n}\sum U^k g_j \to \hat{g}_j,$$

and

$$\| g_i - g_j \| < \epsilon,$$

then

$$\| \frac{1}{n}\sum U^k g_i - \frac{1}{n}\sum U^k g_j \| < \epsilon,$$

so

$$\| \hat{g}_i - \hat{g}_j \| < \epsilon.$$

So if $\{g_i\}$ is a sequence of elements converging to f, we conclude that $\{\hat{g}_i\}$ converges to some element, call it $\hat{f}$. If we choose i sufficiently large so that $\| g_i - f \| < \epsilon$, then

$$\| \frac{1}{n}\sum U^k f - \hat{f} \| \leq \| \frac{1}{n}\sum U^k (f - g_i) \| + \| \frac{1}{n}\sum U^k g_i - \hat{g}_i \| + \| \hat{g}_i - \hat{f} \| \leq 3\epsilon,$$

proving the lemma and hence proving the mean ergodic theorem.

4.4 the arc sine law

The probability distribution with density

$$\sigma(x) = \frac{1}{\pi\sqrt{x(1-x)}}$$

is called the *arc sine law* in probability theory because, if I is the interval $I = [0, u]$ then

$$\text{Prob } x \in I = \text{Prob } 0 \leq x \leq u = \int_0^u \frac{1}{\pi\sqrt{x(1-x)}} = \frac{2}{\pi} \arcsin \sqrt{u}. \tag{4.7}$$

We have already verified this integration because $I = h(J)$ where

$$h(t) = \sin^2 \frac{\pi t}{2}, \quad J = [0, v], \ h(v) = u,$$

and the probability measure we are studying is the push forward of the uniform distribution. So

$$\text{Prob } h(t) \in I = \text{Prob } t \in J = v.$$

The arc sine law plays a crucial role in the theory of fluctuations in random walks. As a cultural diversion we explain some of the key ideas, following the treatment in Feller [?] very closely.

Suppose that there is an ideal coin tossing game in which each player wins or loses a unit amount with (independent) probability $\frac{1}{2}$ at each throw. Let $S_0 = 0, S_1, S_2, \ldots$ denote the successive cumulative gains (or losses) of the first player. We can think of the values of these cumulative gains as being marked off on a vertical s-axis, and representing the position of a particle which moves up or down with probability $\frac{1}{2}$ at each (discrete) time unit . Let

$$\alpha_{2k,2n}$$

denote the *probability that up to and including time* $2n$*, the last visit to the origin occurred at time* $2k$. Let

$$u_{2\nu} = \binom{2\nu}{\nu} 2^{-2\nu}. \tag{4.8}$$

So $u_{2\nu}$ represents the probability that exactly ν out of the first 2ν steps were in the positive direction, and the rest in the negative direction. In other words, $u_{2\nu}$ is the probability that the particle has returned to the origin at time 2ν. We can find a simple approximation to $u_{2\nu}$ using Stirling's formula for an approximation to the factorial:

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n}$$

where the $\sim$ signifies that the ratio of the two sides tends to one as n tends to infinity. For a proof of Stirling's formula cf. [?]. Then

$$\begin{aligned} u_{2\nu} &= 2^{-2\nu} \frac{(2\nu)!}{(\nu!)^2} \\ &\sim 2^{-2\nu} \frac{\sqrt{2\pi}(2\nu)^{2\nu+\frac{1}{2}} e^{-2\nu}}{2\pi \nu^{2\nu+1} e^{-2\nu}} \\ &= \frac{1}{\sqrt{\pi\nu}}. \end{aligned}$$

The results we wish to prove in this section are

Proposition 4.4.1 *We have*

$$\alpha_{2k,2n} = u_{2k}u_{2n-2k}, \tag{4.9}$$

so we have the asymptotic approximation

$$\alpha_{2k,2n} \sim \frac{1}{\pi\sqrt{k(n-k)}}. \tag{4.10}$$

If we set

$$x_k = \frac{k}{n}$$

then we can write

$$\alpha_{2k,2n} \sim \frac{1}{n}\sigma(x_k). \tag{4.11}$$

Thus, for fixed $0 < x < 1$ *and* n *sufficiently large*

$$\sum_{k<xn} \alpha_{2k,2n} \doteq \frac{2}{\pi}\arcsin\sqrt{x}. \tag{4.12}$$

Proposition 4.4.2 *The probability that in the time interval from* 0 *to* $2n$ *the particle spends* $2k$ *time units on the positive side and* $2n-2k$ *time units on the negative side equals* $\alpha_{2k,2n}$*. In particular, if* $0 < x < 1$ *the probability that the fraction* k/n *of time units spent on the positive be less than* x *tends to* $\frac{2}{\pi}\arcsin\sqrt{x}$ *as* $n \to \infty$*.*

Let us call the value of S_{2n} for any given realization of the random walk, the *terminal point.* Of course, the particle may well have visited this terminal point earlier in the walk, and we can ask when it first reaches its terminal point.

Proposition 4.4.3 *The probability that the first visit to the terminal point occurs at time* $2k$ *is given by* $\alpha_{2k,2n}$*.*

We can also ask for the first time that the particle reaches its maximum value: We say that the *first maximum occurs at time* l if

$$S_0 < S_l, S_1 < S_l, \ldots S_{l-1} < S_l, \qquad S_{l+1} \le S_l, S_{l+2} \le S_l, \ldots S_{2n} \le S_l. \tag{4.13}$$

Proposition 4.4.4 *If* $0 < l < 2n$ *the probability that the first maximum occurs at* $l = 2k$ *or* $l = 2k+1$ *is given by* $\frac{1}{2}\alpha_{2k,2n}$*. For* $l = 0$ *this probability is given by* u_{2n} *and if* $l = 2n$ *it is given by* $\frac{1}{2}u_{2n}$*.*

Before proving these four propositions, let us discuss a few of their implications which some people find counterintuitive. For example, because of the shape of the density, σ, the last proposition implies that the maximal accumulated gain is much more likely to occur very near to the beginning or to the end of a coin tossing game rather than somewhere in the middle. The third proposition implies that the probability that the first visit to the terminal point occurs at time $2k$ is that same as the probability that it occurs at time $2n-2k$

and that very early first visits and very late first visits are much more probable than first visits some time in the middle.

In order to get a better feeling for the assertion of the first two propositions let us tabulate the values of $\frac{2}{\pi}\arcsin\sqrt{x}$ for $0 \leq x \leq \frac{1}{2}$.

x	$\frac{2}{\pi}\arcsin\sqrt{x}$	x	$\frac{2}{\pi}\arcsin\sqrt{x}$
0.05	0.144	0.30	0.369
0.10	0.205	0.35	0.403
0.15	0.253	0.40	0.236
0.20	0.295	0.45	0.468
0.25	0.333	0.50	0.500

This table, in conjunction with Prop. 4.4.1 says that if a great many coin tossing games are conducted every second, day and night for a hundred days, then in about 14.4 percent of the cases,the lead will not change after day five.

The proof of all four propositions hinges on three lemmas. Let us graph (by a polygonal path) the walk of a particle. So a "path" is a broken line segment made up of segments of slope ± 1 joining integral points to integral points in the plane (with the time or $t-$axis horizontal and the $s-$axis vertical). If $A = (a, \alpha)$ is a point, we let $A' = (a, -\alpha)$ denote its image under reflection in the $t-$axis.

Lemma 4.4.1 The reflection principle. *Let $A = (a, \alpha), B = (b, \beta)$ be points in the first quadrant with $b > a \geq 0, \alpha > 0, \beta > 0$. The number of paths from A to B which touch or cross the $t-$ axis equals the number of all paths from A' to B.*

Proof. For any path from A to B which touches the horizontal axis, let t be the abscissa of the first point of contact. Reflect the portion of the path from A to $T = (t, 0)$ relative to the horizontal axis. This reflected portion is a path from A' to T, and continues to give a path from A' to B. This procedure assigns to each path from A to B which touches the axis, a path from A' to B. This assignment is bijective: Any path from A' to B must cross the $t-$axis. Reflect the portion up to the first crossing to get a touching path from A to B. This is the inverse assignment. QED

A path with n steps will join $(0, 0)$ to (n, x) if an only if it has p steps of slope $+1$ and q steps of slope -1 where

$$p + q = n, \quad p - q = x.$$

The number of such paths is the number of ways of picking the positions of the p steps of positive slope and so the number of paths joining $(0, 0)$ to (n, x) is

$$N_{n,x} = \left(\begin{array}{c} p+q \\ p \end{array} \right) = \left(\begin{array}{c} n \\ \frac{n+x}{2} \end{array} \right).$$

It is understood that this formula means that $N_{n,x} = 0$ when there are no paths joining the origin to (n, x).

Lemma 4.4.2 The ballot theorem. *Let n and x be positive integers. There are exactly*

$$\frac{x}{n}N_{n,x}$$

paths which lie strictly above the t axis for $t>0$ and join $(0,0)$ to (n,x).

Proof. There are as many such paths as there are paths joining $(1,1)$ to (n,x) which do *not* touch or cross the $t-$axis. This is the same as the total number of paths which join $(1,1)$ to (n,x) less the number of paths which do touch or cross. By the preceding lemma, the number of paths which do cross is the same as the number of paths joining $1.-1)$ to (n,x) which is $N_{n-1,x+1}$. Thus, with p and q as above, the number of paths which lie strictly above the t axis for $t>0$ and which joint $(0,0)$ to (n,x) is

$$\begin{aligned}
N_{n-1,x-1}-N_{n-1,x+1} &= \binom{p+q-1}{p-1}-\binom{p+q-1}{p} \\
&= \frac{(p+q-1)!}{(p-1)!(q-1)!}\left[\frac{1}{q}-\frac{1}{p}\right] \\
&= \frac{p-q}{p+q}\times\frac{(p+q)!}{p!q!} \\
&= \frac{x}{n}N_{n,x} \quad \text{QED}
\end{aligned}$$

The reason that this lemma is called the Ballot Theorem is that it asserts that if candidate P gets p votes, and candidate Q gets q votes in an election where the probability of each vote is independently $\frac{1}{2}$, then the probability that throughout the counting there are more votes for P than for Q is given by

$$\frac{p-q}{p+q}.$$

Here is our last lemma:

Lemma 4.4.3 *The probability that from time 1 to time $2n$ the particle stays strictly positive is given by $\frac{1}{2}u_{2n}$. In symbols,*

$$\text{Prob}\,\{S_1>0,\ldots,S_{2n}>0\}=\frac{1}{2}u_{2n}. \tag{4.14}$$

So

$$\text{Prob}\,\{S_1\neq 0,\ldots,S_{2n}\neq 0\}=u_{2n}. \tag{4.15}$$

Also

$$\text{Prob}\,\{S_1\geq 0,\ldots,S_{2n}\geq 0\}=u_{2n}. \tag{4.16}$$

Proof. By considering the possible positive values of S_{2n} which can range from

2 to $2n$ we have

$$
\begin{aligned}
\text{Prob}\ \{S_1 > 0, \ldots, S_{2n} > 0\} &= \sum_{r=1}^{n} \text{Prob}\ \{S_1 > 0, \ldots, S_{2n} = 2r\} \\
&= 2^{-2n} \sum_{r=1}^{n} (N_{2n-1,2r-1} - N_{2n-1,2r+1}) \\
&= 2^{-2n} (N_{2n-1,1} - N_{2n-1,3} + N_{2n-1,3} - N_{2n-1,5} + \cdots) \\
&= 2^{-2n} N_{2n-1,1} \\
&= \frac{1}{2} p_{2n-1,1} \\
&= \frac{1}{2} u_{2n}.
\end{aligned}
$$

The passage from the first line to the second is the reflection principle,as in our proof of the Ballot Theorem, from the third to the fourth is because the sum telescopes. The $p_{2n-1,1}$ on the next to the last line is the probability of ending up at $(2n-1, 1)$ starting from $(0, 0)$. The last equality is simply the assertion that to reach zero at time $2n$ we must be at ± 1 at time $2n-1$ (each of these has equal probability, $p_{2n-1,1}$) and for each alternative there is a 50 percent chance of getting to zero on the next step. This proves (4.14). Since a path which never touches the $t-$axis must be always above or always below the $t-$axis, (4.15) follows immediately from (4.14). As for (4.16), observe that a path which is strictly above the axis from time 1 on, must pass through the point $(1, 1)$ and then stay above the horizontal line $s = 1$. The probability of going to the point $(1, 1)$ at the first step is $\frac{1}{2}$, and then the probability of remaining above the new horizontal axis is Prob $\{S_1 \geq 0, \ldots, S_{2n-1} \geq 0\}$. But since $2n-1$ is odd, if $S_{2n-1} \geq 0$ then $S_{2n} \geq 0$. So, by (4.14) we have

$$
\begin{aligned}
\frac{1}{2} u_{2n} &= \text{Prob}\ \{S_1 > 0, \ldots, S_{2n} > 0\} \\
&= \frac{1}{2} \text{Prob}\ \{S_1 \geq 0, \ldots, S_{2n-1} \geq 0\} \\
&= \frac{1}{2} \text{Prob}\ \{S_1 \geq 0, \ldots, S_{2n-1} \geq 0, S_{2n} \geq 0\},
\end{aligned}
$$

completing the proof of the lemma.

We can now turn to the proofs of the propositions.

Proof of Prop.4.4.1 To say that the last visit to the origin occurred at time $2k$ means that

$$S_{2k} = 0$$

and

$$S_j \neq 0, \quad j = 2k+1, \ldots, 2n.$$

By definition, the first $2k$ positions can be chosen in $2^{2k} u_{2k}$ ways to satisfy the first of these conditions. Taking the point $(2k, 0)$ as our new origin, (4.15) says that there are $2^{2n-2k} u_{2n-2k}$ ways of choosing the last $2n - 2k$ steps so as to satisfy the second condition. Multiplying and then dividing the result by 2^{2n} proves Prop.4.4.1.

Proof of Prop.4.4.2. We consider paths of $2n$ steps and let $b_{2k,2n}$ denote the probability that exactly $2k$ sides lie above the $t-$axis. Prop. 4.4.2 asserts that

$$b_{2k,2n} = \alpha_{2k,2n}.$$

For the case $k = n$ we have $\alpha_{2n,2n} = u_0 u_{2n} = u_{2n}$ and $b_{2n,2n}$ is the probability that the path lies entirely above the axis. So our assertion reduces to (4.16) which we have already proved. By symmetry, the probability of the path lying entirely below the the axis is the same as the probability of the path lying entirely above it, so $b_{0,2n} = \alpha_{0,2n}$ as well. So we need prove our assertion for $1 \leq k \leq n-1$. In this situation, a return to the origin must occur. Suppose that the first return to the origin occurs at time $2r$. There are then two possibilities: the entire path from the origin to $(2r, 0)$ is either above the axis or below the axis. If it is above the axis, then $r \leq k \leq n-1$, and the section of path beyond $(2r, 0)$ has $2k - 2r$ edges above the $t-$axis. The number of such paths is

$$\frac{1}{2} 2^{2r} f_{2r} 2^{2n-2r} b_{2k-2r,2n-2r}$$

where f_{2r} denotes the *probability of first return* at time $2r$:

$$f_{2r} = \text{Prob}\, \{S_1 \neq 0, \ldots, S_{2r-1} \neq 0, S_{2r} = 0\}.$$

If the first portion of the path up to $2r$ is spent below the axis, the the remaining path has exactly $2k$ edges above the axis, so $n - r \geq k$ and the number of such paths is

$$\frac{1}{2} 2^{2r} f_{2r} 2^{2n-2r} b_{2k,2n-2r}.$$

So we get the recursion relation

$$b_{2k,2n} = \frac{1}{2} \sum_{r=1}^{k} f_{2r} b_{2k-2r,2n-2r} + \frac{1}{2} \sum_{r=1}^{n-k} f_{2r} b_{2k,2n-2r} \quad 1 \leq k \leq n-1. \tag{4.17}$$

Now we proceed by induction on n. We know that $b_{2k,2n} = u_{2k} u_{2n-2k} = \frac{1}{2}$ when $n = 1$. Assuming the result up through $n-1$, the recursion formula (4.17) becomes

$$b_{2k,2n} = \frac{1}{2} u_{2n-2k} \sum_{r=1}^{k} f_{2r} u_{2k-2r} + \frac{1}{2} u_{2k} \sum_{r=1}^{n-k} f_{2r} u_{2n-2k-2r}. \tag{4.18}$$

But we claim that the probabilities of return and the probabilities of first return are related by

$$u_{2n} = f_2 u_{2n-2} + f_4 u_{2n-4} + \cdots + f_{2n} u_0. \tag{4.19}$$

Indeed, if a return occurs at time $2n$, then there must be a first return at some time $2r \leq 2n$ and then a return in $2n - 2r$ units of time, and the sum in (4.19) is over the possible times of first return. If we substitute (4.19) into the first sum in (4.18) it becomes u_{2k} while substituting (4.19) into the second term yields u_{2n-2k}. Thus (4.18) becomes

$$b_{2k,2n} = u_{2k}u_{2n-2k}$$

which is our desired result.

Proof of Prop.4.4.3. This follows from Prop.4.4.1 because of the symmetry of the whole picture under rotation through 180° and a shift: The probability in the lemma is the probability that $S_{2k} = S_{2n}$ but $S_j \neq S_{2n}$ for $j < 2k$. Reading the path rotated through 180° about the end point, and with the endpoint shifted to the origin, this is clearly the same as the probability that $2n - 2k$ is the last visit to the origin. QED

Proof of Prop 4.4.4. The probability that the maximum is achieved at 0 is the probability that $S_1 \leq 0, \ldots, S_{2n} \leq 0$ which is u_{2n} by (4.16). The probability that the maximum is first obtained at the terminal point, is, after rotation and translation, the same as the probability that $S_1 > 0, \ldots, S_{2n} > 0$ which is $\frac{1}{2}u_{2n}$ by (4.14). If the maximum occurs first at some time l in the middle, we combine these results for the two portions of the path - before and after time ℓ - together with (4.9) to complete the proof. QED

4.5 The Beta distributions.

The arc sine law is the special case $a = b = \frac{1}{2}$ of the Beta distribution with parameters a, b which has probability density proportional to

$$t^{a-1}(1-t)^{b-1}.$$

So long as $a > 0$ and $b > 0$ the integral

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt$$

converges, and was evaluated by Euler to be

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

where Γ is Euler's Gamma function. So the Beta distributions with $A > 0, b > 0$ are given by

$$\frac{1}{B(a,b)}t^{a-1}(1-t)^{b-1}.$$

We characterized the arc sine law ($a = b = \frac{1}{2}$) as being the unique probability density invariant under L_4. The case $a = b = 0$, where the integral does not converge, also has an interesting characterization as an invariant density. Consider transformations of the form

$$t \mapsto \frac{at+b}{ct+d}$$

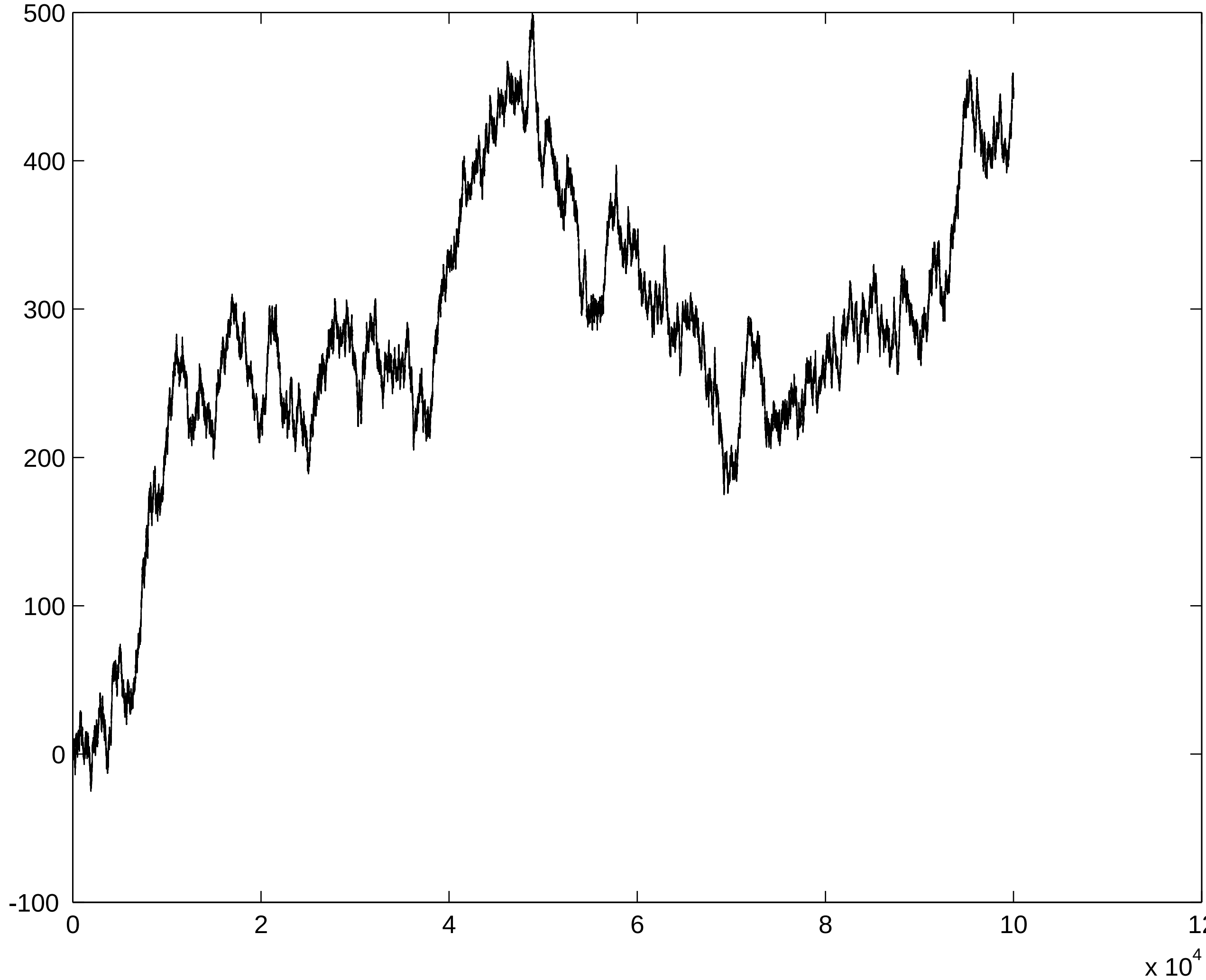

Figure 4.2: A random walk with 100,000 steps. The last zero is at time 3783. For the remaining 96,217 steps the path is positive. According to the arc sine law, with probability 1/5, the particle will spend about 97.6 percent of its time on one side of the origin.

where the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is invertible. Suppose we require that the transformation preserve the origin and the point $t = 1$. Preserving the origin requires that $b = 0$, while preserving the ' point $t = 1$ requires that $a = c + d$. Since $b = 0$ we must have $ad \neq 0$ for the matrix to be invertible. Since multi[lying all the entries of the matrix by the same non-zero scalar does not change the transformation, we may as well assume that $d = 1$, and hence the family transformations we are looking at are

$$\phi_a : \quad t \mapsto \frac{at}{(a-1)t+1}, \qquad a \neq 0.$$

Notice that

$$\phi_a \circ \phi_b = \phi_{ab}.$$

Our claim is that, up to scalar multiple, the density

$$\rho(t) = \frac{1}{t(1-t)}$$

is the unique density such that the measure

$$\rho(t)dt$$

is invariant under all the transformations ϕ_a. Indeed,

$$\phi_a'(t) = \frac{a}{[1-t+at]^2}$$

so the condition of invariance is

$$\frac{a}{[1-t+at]^2}\rho(\phi_a(t)) = \rho(t).$$

Let us normalize ρ by

$$\rho\left(\frac{1}{2}\right) = 4.$$

Then

$$s = \phi_a\left(\frac{1}{2}\right) \Leftrightarrow s = \frac{a}{1+a} \Leftrightarrow a = \frac{s}{1-s}.$$

So taking $t = \frac{1}{2}$ in the condition for invariance and a as above, we get

$$\rho(s) = 4((1-s)/s)[\frac{1}{2} + \frac{1}{2}\frac{s}{1-s}]^2 = \frac{1}{s(1-s)}.$$

This elementary geometrical fact - that $1/t(1-t)$ is the unique density (up to scalar multiple) which is invariant under all the ϕ_a - was given a deep philosophical interpretation by Jaynes, [**?**]:

Suppose we have a possible event which may or may not occur, and we have a population of individuals each of whom has a clear opinion (based on ingrained prejudice, say, from reading the newspapers or watching television) of the probability of the event being true. So Mr. A assigns probability p(A) to the event E being true and (1-p(A)) as the probability of its not being rue, while Mr. B assigns probability P(B) to its being true and (1-p(B)) to its not being true and so on.

Suppose an additional piece of information comes in, which would have a (conditional) probability x of being generated if E were true and y of this information being generated if E were not true. We assume that both x and y are positive, and that every individual thinks rationally in the sense that on the advent of this new information he changes his probability estimate in accordance with Bayes' law, which says that the posterior probability p' is given in terms of the prior probability p by

$$p' = \frac{px}{px + (1-p)y} = \phi_a(p) \quad \text{where} \quad a := \frac{x}{y}.$$

We might say that the population as a whole has been invariantly prejudiced if any such additional evidence does not change the proportion of people within the population whose belief lies in a small interval. Then the density describing this state of knowledge (or rather of ignorance) must be the density

$$\rho(p) = \frac{1}{p(1-p)}.$$

According to this reasoning of Jaynes, we take the above density to describe the prior probability an individual (thought of as a population of subprocessors in his brain) would assign to the probability of an outcome of a given experiment. If a series of experiments then yielded M successes and N failures,Bayes' theorem (in its continuous version) would then yield the posterior distribution of probability assignments as being proportional to

$$p^{M-1}(1-p)^{N-1}$$

the Beta distribution with parameters M, N.