Unpredictable Walks on the Integers

EJ Infeld *

March 5, 2013

Abstract

In 1997 Benjamini, Premantle and Peres introduced a construction of a walk on $\mathbb{Z}$ that has a faster decaying predictability profile than that of the simple random walk [1], at the rate of $ck^{-\alpha}$ for some $\alpha > 1/2$. A year later, Häggstrom and Mossel improved on this result by constructing two walks with predictability profiles that decay at close to the rate corresponding to the uniform probability distribution [2]. The same year, Hoffman showed in [3] that this new decay rate is in fact optimal.

After introducing the necessary concepts, we will have a look at a few familiar examples and consider their predictability. Then we will follow a construction of a walk introduced in [2] as well as the proof that its predictability profile decays at the postulated rate.

1 Introduction

We have seen many incarnations of a walk over the course of the term. Since we shall limit this discussion to walks on the integers and the nature of the problem is algorithmic, we will define it as follows:

Definition A *(random) walk* on $\mathbb{Z}$ is a sequence of variables $\{S_n\}_{n\in\mathbb{N}}$ such that for $n > 0$, S_n is determined by a (randomised) algorithm A that takes $\{S_0, \ldots, S_{n-1}\}$ (and a random element) as input and returns $S_n = S_{n-1} - 1$ or $S_n = S_{n-1} + 1$. Throughout this discussion we will take $S_0 = 0$.

As an example, consider a simple random walk on $\mathbb{Z}$. At each step n, the input is a history $\{S_0, \ldots, S_{n-1}\}$, and a uniform random variable $U_n \in [0, 1]$, upon which A returns $S_{n-1} + 1$ if

*Department of Mathematics, Dartmouth College, Hanover NH.

$U_n < 1/2$ and $S_{n-1} - 1$ otherwise. Note that in fact for any walk that is a Markov Chain the output of the algorithm A at step n is going to be independent of $\{S_i\}_{i<n-1}$.

So what does it mean for a walk to be unpredictable? Suppose that Alice and Bob are on opposite sides of the fence - Alice wants to construct an algorithm that will generate an unpredictable walk, and Bob will make his best bet on where the walk will go. Bob is going to know Alice's algorithm and can watch the walk for as long as he wants, say n steps. Then he can, based on his knowledge of the algorithm and the observed history, make a bet on where the walk will be at step $n+k$, so k steps into the future. How good is his best bet is the measure of predictability of A.

More formally, suppose we have an algorithm A that generates a walk on $\mathbb{Z}$. Then the measure of predictability of A is the following:

Definition [1] The *predictability profile* of A is:

$$PRE_A(k) = \sup_{i \in \mathbb{Z},\ n \in \mathbb{N},\ \{S_j\}_{j=0}^n} \{\mathbb{P}[S_{n+k} = i : A,\ \{S_j\}_{j=0}^n]\},$$

i.e. the highest probability over all possible values $i \in \mathbb{Z}$, given A and any recorded history, of the walk having that particular value i k steps after the history is no longer recorded.

Notice that if the algorithm A is deterministic, $PRE_A(k) = 1$ for all k always. Also, $PRE_A(k) \geq 1/(k+1)$, since $k+1$ is how many integers at most we can possibly reach on the kth step (remember that the path is bipartite). So the property that we are truly interested in is how fast $PRE_A(k)$ decays in k and in particular, whether its decay is sufficiently close to that of $1/(k+1)$. It's not hard to guess that we would like the corresponding probability distribution to be close to uniform.

This is very much a programmer's problem. Suppose Alice wants to send some information through a network of proxies, and Bob might watch its walk for a while before he decides to try to intercept it, and that he might take time equivalent to k steps to establish the interception attempt. The algorithm is set, the code is open and all Alice can do is make sure Bob, who knows as much about the walk as she does still has little chance of successful interception. This chance is the predictability profile of the walk.

2 Some walks and their predictability profiles

2.1 Simple random walk

Consider a simple random walk. This is a Markov Chain, so the probability of being at any given point is going to be independent of history except for its final entry, and in fact the predictability profile itself will be independent of history entirely - the probability distribution is symmetric regardless of where on $\mathbb{Z}$ we start the walk, and so its maximum value will stay the same. So suppose without loss of generality that the history is $\{S_0 = 0\}$. We are interested in

$$PRE_{SRW}(k) = \sup_{m \in \mathbb{N}} \{S_k = m : S_0 = 0\}.$$

Note that the probability of any individual sequence of moves left or right is the same, and equal to 2^{-k}, so the question comes down to: how many walks terminate at m? The probability of terminating at m is that number times 2^{-k}.

Firstly, note that the path is bipartite. the only values of $m \in \mathbb{Z}$ with nonzero probability $\mathbb{P}[S_k = m]$ are $-k \leq m \leq k$ of the same parity as k. For $m = -k + 2j$, some natural number $j \leq k$, there are $\binom{k}{j}$ possible sequences of left or right moves that terminate at m. Then the highest probability is cumulated around 0, and in fact for k even we expect to be at 0 after k steps with probability about $1/\sqrt{k}$ by the Central Limit Theorem. This is not all that close to $1/k$. Can Alice do better?

2.2 Biased random walk with a random bias

Suppose that at the beginning of the walk we pick $p \in [0, 1]$ uniformly at random and then perform a random walk in which at every stage i, we go right with probability p and left with probability $1 - p$. The probability that we will end up anywhere between $-k$ and k, provided the parity is correct, is now uniformly distributed. We can check this as follows.

The probability that all k values σ_i are equal to 1 is $\int\limits_0^1 p^k dp = 1/(k+1)$. Similarly, the probability that they are all equal to -1 is also $1/(k+1)$. Now suppose that the probability that exactly j values σ_i are -1 is $\binom{k}{j} \int\limits_0^1 p^{k-j}(1-p)^j dp = 1/(k+1)$. Then as long as $j < k$, by integration by parts:

$$\binom{k}{j+1}\int_0^1 p^{k-j-1}(1-p)^{j+1} = \binom{k}{j+1}\left[\frac{p^{k-j}}{k-j}(1-p)^{j+1}\right]_0^1 + \binom{k}{j+1}\int_0^1 \frac{j+1}{k-j}p^{k-j}(1-p)^j dp$$

$$= 0 + \binom{k}{j}\int_0^1 p^{k-j}(1-p)^j dp = \frac{1}{k+1}$$

So what is wrong with this model?

From the history $S_0, \ldots, S_n$ the probability p can be approximated, by simply waiting long enough for $|\{i \leq n : \sigma_i = 1\}|/n$ to be likely close to p. So the longer Bob watches the walk, the less optimal the probability profile becomes. And in fact if we assume Bob knows p, since he can watch the walk for as long as he wants, this model turns out to be at best as unpredictable as the simple random walk.

This might tell us that an ideal walk would keep this random element in the bias, but be constructed in such a way as to hide this bias from the history. Intuition suggests that we then want the bias to be shifting slightly, enough to conceal itself from the history but not enough to become equivalent to the simple random walk.

3 Häggstrom and Mossel construction [2]

The following construction attempts to balance the advantages of a random walk with a random bias with concealing the bias from the history.

Let Alice pick an integer $b > 1$ and a sequence $\{a_j\}_{j=1}^{\infty}$ such that $\sum_{j=1}^{\infty} a_j < 1/2$.

For any j, define a random process $\{p_i^{(j)}\}_{i=1}^{\infty}$ by choosing a value uniformly at random from $[-a_j, a_j]$ and assigning it to b^j consecutive terms, then choosing another value and assigning it to the next b^j terms and so on. For example if $b = 2,\ j = 1$ then $\{p_i^{(j)}\}_{i=1}^{\infty} = \alpha_1, \alpha_1, \alpha_2, \alpha_2, \ldots$ where the α_is, i odd, are chosen independently, uniformly at random from $[-a_j, a_j]$.

Once we have determined $\{p_i^{(j)}\}_{i=1}^{\infty}$ for each j, define $p_i = 1/2 + p_i^{(1)} + p_i^{(2)} + \ldots$. This is the bias for the walk at time i. Then the process is realized by a series of independent uniform random variables $\{U_i\}_{i=1}^{\infty}$, with the step of the walk at time i being σ_i satisfying:

$$\sigma_i = \begin{cases} +1 \text{ if } U_i < p_i \\ -1 \text{ otherwise.} \end{cases}$$

So the walk is $S_n = \sum_{i=1}^{\infty} \sigma_i$, where $\{\sigma_i\}_{i=1}^{\infty}$ are $\{-1, 1\}$-valued independent random variables conditioned on the environment $\{p_i\}_{i=1}^{\infty}$. At each time i, σ_i takes value 1 with probability p_i and -1 with probability $1 - p_i$. We are dealing with a random walk that has varying bias.

Theorem 3.1. *Given b and $\{a_j\}_{j=1}^{\infty}$, there exists a constant $c < \infty$ such that*

$$PRE_S(k) \leq \frac{c}{k a_{\lfloor log_b(k/2) \rfloor}}$$

for all k.

Proof. Assume without loss of generality that $k \geq 2b$, since for any finite set of values of $PRE_S(k)$ we can choose an appropriate c, and denote $m_k = \lfloor log_b(k/2) \rfloor$. Suppose we know $S_0, \ldots, S_n$ and want to predict S_{n+k}.

Consider $p_i^{(m_k)}$, constant on time intervals of length $b^{m_k} = b^{\lfloor log_b(k/2) \rfloor} \geq b^{log_b(k/2)-1} = k/2b$. So the time interval from $n+1$ to $n+k$ necessarily contains a subinterval I of length $k/2b$ on which $p_i^{(m_k)}$ is constant.

Write $\tilde{p}_i^{(m_k)} = p_i - p_i^{(m_k)} = 1/2 + p_i^{(1)} + p_i^{(2)} + \cdots + p_i^{(m-1)} + p_i^{(m+1)} + \ldots$, and note that $\{p_i^{(m_k)}\}_{i=1}^{\infty}$ is independent of $\{\tilde{p}_i^{(m_k)}\}_{i=1}^{\infty}$. Define $X_i = U_i - \tilde{p}_i^{(m_k)}$ and let:

$$Y_i = \begin{cases} 1 \text{ if } X_i < -a_{m_k} \\ 0 \text{ if } X_i \in [-a_{m_k}, a_{m_k}] \\ -1 \text{ if } X_i > a_{m_k}. \end{cases}$$

Note that the relation between these variables and the original variables U_i is as follows:

$X_i < -p_i^{(m_k)} \;\Rightarrow\; U_i < p_i,$

$X_i < -a_{m_k} \;\Rightarrow\; U_i < p_i,$

$X_i > a_{m_k} \;\Rightarrow\; U_i > p_i$, and so:

$$Y_i = 1 \;\Rightarrow\; \sigma_i = 1$$

$$Y_i = -1 \;\Rightarrow\; \sigma_i = -1.$$

Further, each of the Y_i takes value 0 with probability $2a_{m_k}$, so the number of times this happens has a binomial $(b^{m_k}, 2a_{m_k})$ distribution. Given that $Y_i = 0$, what is the probability that $\sigma_i = 1$? We arrive at $U_i \in [-a_{m_k} + \tilde{p}_i^{(m_k)}, a_{m_k} + \tilde{p}_i^{(m_k)}]$, less than p_i with probability:

$$\frac{a_{m_k} + p_i^{(m_k)}}{2a_{m_k}} = \frac{1}{2} + \frac{p_i^{(m_k)}}{2a_{m_k}}.$$

Note that if Bob is given any extra information, his best odds can only improve, since he has the option to ignore that information. We will show that the bound we are attempting to prove can be established for a situation when Bob in fact knows more than we have previously assumed.

Suppose that in addition to $\{\sigma_i\}_{i=1}^n$, Bob knows:

- the values of $\{\sigma_{i\in I_{n+1}^{n+k}\setminus I}\}$ i.e. the movements outside of the subinterval of constant $p_i^{(m_k)}$
- $\{Y_i\}_{i\in I}$.

Then S_{n+k} is necessarily in the interval

$$L = \left[S_n + \sum_{i\in I_{n+1}^{n+k}\setminus I} \sigma_i + \sum_{i\in I} Y_i - \sum_{i\in I} \mathbf{1}_{\{Y_i=0\}} \;,\; S_n + \sum_{i\in I_{n+1}^{n+k}\setminus I} \sigma_i + \sum_{i\in I} Y_i + \sum_{i\in I} \mathbf{1}_{\{Y_i=0\}} \right].$$

And in fact, since the graph is bipartite, S_{n+k} is necessarily an even number of steps from either end of this interval. We then note that on this subset of L of admissible elements, the probability distribution given the extra information is uniform, since the bias $p_I^{(m_k)}$, and X_i are independent, with uniform distribution on $[-a_{m_k}, a_{m_k}]$.

Then:

$$\mathbb{P}[S_{n+k} = x | S_0, \ldots, S_n] \leq \sum_{j=0}^{b^{m_k}} \frac{\mathbb{P}[|\{i \in Y : Y_i = 0\}| = j]}{j+1} \leq \mathbb{P}\left[|\{i \in I :\; Y_i = 0\}| \leq b^{m_k} a_{m_k}\right] + \frac{1}{b^{m_k} a_{m_k}}$$

The term $P\left[|\{i \in I :\; Y_i = 0\}| \leq b^{m_k} a_{m_k}\right]$ decays exponentially in $b^{m_k} a_{m_k}$, so it will not affect the overall rate of decay we are investigating. Therefore:

$$\mathbb{P}[S_{n+k} = x | S_0, \ldots, S_n] \leq \frac{C}{b^{m_k} a_{m_k}} = \frac{2bC}{k a_{m_k}},$$

and so for $c = 2bC$, have $PRE_S(k) \leq \frac{c}{k a_{\lfloor log_b(k/2) \rfloor}}$. $\square$

To sum up, how did we get around the fact that Bob might deduce partial information about the bias from the history? By giving him extra data, we have effectively cut off the part of Alice's algorithm that relied on information he might infer. We distilled the non-deterministic part of the walk to steps that draw their bias afresh after Bob is already reacting. It was shown in [3], that the decay rate achieved above is optimal.

References

[1] I. Benjamini, R. Premantle and Y. Peres, Unpredictable Paths and Percolation (1997)

[2] O. Haggstrom and E. Mossel, Nearest-neighbor walks with low predictability profile and percolation in 2+ϵ dimensions (1998)

[3] C. Hoffman, Unpredictable Nearest Neighbor Processes (1998)