The Annals of Probability
1998, Vol. 26, No. 4, 1781–1787

UNPREDICTABLE NEAREST NEIGHBOR PROCESSES

BY CHRISTOPHER HOFFMAN

University of Maryland

Benjamini, Pemantle and Peres constructed nearest neighbor processes which have predictability profiles that decay faster than that of the simple random walk. Häggström and Mossel found processes with even faster decaying predictability profiles. We prove that the rate of decay achieved by Häggström and Mossel is optimal.

1. Introduction.

Benjamini, Pemantle and Peres [1] constructed a class of nearest neighbor processes on the integers which are in a certain sense more random than simple random walk. That is, given the past of the process, it is more difficult to predict what the value of the process will be at a time in the distant future than it is to predict where a simple random walk will be at that time. To make this precise we define the predictability profile of a process.

DEFINITION 1.1. For a sequence of random variables $S = \{S_i\}_{i\geq 0}$ taking values in $\mathbb{Z}$, we define its *predictability profile* $\{\mathrm{PRE}_S(j)\}_{j\geq 1}$ by

$$\mathrm{PRE}_S(j) = \sup \mathrm{P}[S_{i+j} = z | S_0, \ldots, S_i],$$

where the supremum is over all $z \in \mathbb{Z}$, all $i \geq 0$ and all histories $S_0, \ldots, S_i$.

Thus $\mathrm{PRE}_S(j)$ is the maximal chance of guessing S correctly j steps into the future, given the past of S. The predictability profile of simple random walk on $\mathbb{Z}$ is asymptotic to $Cj^{-1/2}$ for some $C > 0$.

For any $\alpha < 1$, Benjamini, Pemantle and Peres constructed a nearest neighbor process S_α taking values in $\mathbb{Z}$ such that $\mathrm{PRE}_{S_\alpha}(j) < C_\alpha j^{-\alpha}$. This result was extended by the following theorem of Häggström and Mossel.

THEOREM 1.1 (Häggström and Mossel [4]). *For any decreasing positive sequence* $\{f(j)\}_{j=1}^{\infty}$ *such that*

$$\sum_{j=1}^{\infty} \frac{f(j)}{j} < \infty,$$

there exists a constant $C < \infty$ *and a nearest neighbor walk* $S = \{S_i\}_{i\geq 0}$ *on* $\mathbb{Z}$ *such that*

$$\mathrm{PRE}_S(j) \leq \frac{C}{jf(j)}$$

for all $j \geq 1$.

Received July 1997.
AMS 1991 *subject classification*. Primary 60G07.
Key words and phrases. Predictability profile, nearest neighbor processes.

The processes constructed in [1] were used to give a new proof of Grimmet, Kesten and Zhang's theorem [1] that simple random walk on the unique infinite percolation cluster on $\mathbb{Z}^3$ is transient. The processes constructed in [4] were used to show the transience of simple random walk on infinite percolation clusters on certain subgraphs of $\mathbb{Z}^3$. These include the "$2+\varepsilon$ dimensional" wedge graphs

$$\{(x, y, z) \,|\, |z| \leq |x|^{\varepsilon}\}$$

for any $\varepsilon > 0$. They also included

$$\mathscr{W}_{\alpha} = \{(x, y, z) \,|\, |z| \leq (\log |x|)^{\alpha}\}$$

for any $\alpha > 2$. Simple random walk is transient on the graphs $\mathscr{W}_{\alpha}$ if and only if $\alpha > 1$ (see [3]). Thus for $1 < \alpha \leq 2$, simple random walk on the graph $\mathscr{W}_{\alpha}$ is transient, but it is unknown whether simple random walk on percolation clusters on the graph $\mathscr{W}_{\alpha}$ is transient or recurrent. This paper does not resolve that question. It does show that the method of [1] cannot be extended to prove that percolation clusters on these graphs are transient.

Benjamini, Pemantle and Peres also obtained a bound for how fast the predictability profile of a nearest neighbor process on $\mathbb{Z}$ can decay. It is clear the predictability profile of a nearest neighbor process on the integers can decay no more rapidly than $1/j + 1$. Benjamini, Pemantle and Peres showed that for nearest neighbor processes the predictability profile cannot decay at a rate $O(1/j)$. They did this by showing that for any process there exists another process with the same predictability profile so that the sequence of random variables $\{S_i - S_{i-1}\}$ is stationary and ergodic. The ergodic theorem shows that the predictability profile of such a process cannot decay at a rate $O(1/j)$.

In this paper we sharpen the result in [1] and prove that the condition given by Häggström and Mossel is sharp.

THEOREM 1.2. *For any positive sequence $\{f(j)\}_{j=1}^{\infty}$ such that $\sum_{n=1}^{\infty} f(k^n) = \infty$ for some integer integer $k > 1$, there exists no nearest neighbor process $S = \{S_i\}_{i \geq 0}$ on $\mathbb{Z}$ such that*

$$\mathrm{PRE}_S(j) \leq \frac{1}{jf(j)}$$

for all $j \geq 1$. In particular the theorem holds for any decreasing positive sequence $f(j)$ such that $\sum_{j=1}^{\infty}(f(j)/j) = \infty$.

By taking $f(j) = c/\log j$ we can see that it is impossible for a nearest neighbor process to have a predictability profile that decays on the order of $O(\log j/j)$. On the other hand, by Theorem 1.1, there are nearest neighbor processes that have predictability profiles which decay on the order of $O((\log j)^{1+\varepsilon}/j)$ for any $\varepsilon > 0$. Similarly, it is impossible for a nearest neighbor process to have a predictability profile that decays like $O((\log j)(\log\log j)/j)$ but possible to have one that decays like $O((\log j)(\log\log j)^{1+\varepsilon}/j)$.

2. Proof. First we sketch the idea of the proof. We will force a contradiction by showing that certain predictability profiles force $\mathrm{P}(S_{i+n}-S_i<-n)>0$ for large n. Choose some large integer k. The predictability profile and the fact that S is a nearest neighbor process gives us an upper bound for $\mathrm{P}(S_{i+k}-S_i>tk|S_0,\ldots,S_i)$ for each $t\in[-1,1]$, which is uniform over the choice of i and past $S_0,\ldots,S_i$. Consider the distribution which assigns probability $\mathrm{PRE}(k)$ to $1,1-1/k,\ldots,1-(\lfloor 1/\mathrm{PRE}(k)\rfloor-1)/k$ and probability $1-\mathrm{PRE}(k)\lfloor 1/\mathrm{PRE}(k)\rfloor$ to $1-\lfloor 1/\mathrm{PRE}(k)\rfloor/k$, where $\lfloor x\rfloor$ is the greatest integer less than or equal to x. Call this distribution D_1 and its mean m_1. The distribution D_1 has the property that

$$\mathrm{P}_{D_1}(x>t)\geq\mathrm{P}(S_{i+k}-S_i>tk|S_0,\ldots,S_i)$$

for every $t\in[-1,1]$, integer i and past $S_0,\ldots,S_i$. This inequality is a simple consequence of the predictability profile.

Now for each t we want to find uniform bounds on $\mathrm{P}(S_{i+k^2}-S_i>tk^2|S_0,\ldots,S_i)$. To do this we take the distribution D_1, convolve it with itself k times and then rescale it so that distribution is supported in the interval $[-1,1]$. We call the rescaled convolution $\tilde{D}_2$. Because of the inequalities that D_1 satisfies, the distribution $\tilde{D}_2$ provides upper bounds for $\mathrm{P}(S_{i+k^2}-S_i>tk^2|S_0,\ldots,S_i)$, for any t, i, and $S_0,\ldots,S_i$.

Now we employ a flattening procedure to the rescaled convolution $\tilde{D}_2$, to make a distribution D_2. The distribution D_2 will provide even better estimates for $\mathrm{P}(S_{i+k^2}-S_i>tk^2|S_0,\ldots,S_i)$ than we got with $\tilde{D}_2$. The flattening procedure is done as follows. Take the largest t such that

$$\mathrm{P}_{\tilde{D}_2}(x=t)>\mathrm{PRE}(k^2).$$

Create a new distribution that decreases the probability assigned to t to $\mathrm{PRE}(k^2)$ and increases the probability assigned to $t-1/k^2$ by the same amount as the probability assigned to t was decreased by. This new distribution still provides bounds for

$$\mathrm{P}(S_{i+k^2}-S_i>tk^2|S_0,\ldots,S_i).$$

Then we repeat this procedure until we have a distribution that assigns to every t probability at most $\mathrm{PRE}(k^2)$. Call this distribution D_2 and its mean m_2.

Next we take the distribution D_2, convolve it with itself k times, and rescale it to form $\tilde{D}_3$. Then we perform a flattening procedure like the one described above to create D_3 which provides bounds for $\mathrm{P}(S_{i+k^3}-S_i>tk^3|S_0,\ldots,S_i)$. This procedure is repeated forming distributions D_n. At each stage we give upper bounds on $\sigma^2(D_n)$ and m_n, the mean of the distribution D_n. We continue until we force the means of distributions less than -1. This implies that $\mathrm{P}(S_{k^n}-S_0<k^n)>0$, which contradicts the fact that $\{S\}_{i\geq 0}$ is a nearest neighbor process.

Vaguely speaking, this process succeeds in forcing the distributions to be supported on smaller and smaller t for the following reason. The rescaled convolutions take mass from the edges of the distribution and concentrate it near its mean. Then we use the predictability profile to flatten the convolution,

moving the extra mass leftward. So mass from the right-hand side of the distribution gets moved to the middle by the convolution and to the left by the flattening, while the mass at the left gets moved to the center by the convolution and then back to the left by the flattening. The net effect is that under each iteration of this procedure the mean of the distribution has been decreased. It turns out that the amount that the mean of the nth distribution is less than the mean of the $(n-1)$st distribution is proportional to $f(k^n)$. Thus if $\sum f(k^n) = \infty$, the distance we can move the means of the distributions is unbounded.

To make this precise we assume that $\mathrm{PRE}_S(j) = 1/jf(j)$. This causes no loss of generality because, if this is not the case, then we can define $g(j) = 1/\mathrm{PRE}_S(j)j$. Then $g(j) \geq f(j)$ so $\sum g(k^j) = \infty$ for some $k > 1$ and the theorem still applies.

Now pick a k such that $\sum f(k^j) = \infty$. For technical reasons we need to choose a subsequence n_i and do the flattening procedure along the sequence D_{n_i}. Pick an l so that $k^l > 20{,}000$. If $\sum f(k^n) = \infty$ then there exists a $q < l$ so that $\sum_j f(k^{q+jl}) = \infty$. For notational reasons we assume $q = 1$. This causes no loss of generality. Set $n_1 = 1$. Given n_{i-1}, pick n_i to be the smallest integer greater than n_{i-1} so that $n_i \in 1 + l\mathbb{Z}$ and $f(k^{n_{i-1}})/f(k^{n_i}) \leq 2^{(n_i - n_{i-1})/l}$.

LEMMA 2.1. *If $\sum f(k^n) = \infty$ then $\sum_i f(k^{n_i}) = \infty$.*

PROOF. By the way that the n_i were chosen, if there exists some $n \in 1 + l\mathbb{Z}$ such that $n_i < n < n_{i+1}$, then $f(k^n) \leq f(k^{n_i})2^{(n_i - n)/l}$. So

$$\sum_{n \in 1 + l\mathbb{Z},\, n_i < n < n_{i+1}} f(k^n) \leq f(k^{n_i}) \sum_{j=1}^{\infty} 2^{-j} \leq f(k^{n_i}).$$

By the assumption made in the previous paragraph, we have that $\sum_j f(k^{1+lj}) = \infty$. The preceding argument implies that if $\sum_j f(k^{1+lj}) = \infty$ then $\sum_i f(k^{n_i}) = \infty$. Thus the lemma is true. □

Again, we define D_1 to be the distribution that assigns probability $\mathrm{PRE}(k)$ to $1, 1 - 1/k, \ldots, 1 - (\lfloor 1/\mathrm{PRE}(k) \rfloor - 1)/k$ and probability $1 - \mathrm{PRE}(k)\lfloor 1/\mathrm{PRE}(k) \rfloor$ to $1 - \lfloor 1/\mathrm{PRE}(k) \rfloor / k$. For all n we set $\tilde{D}_n$ to be the rescaled version of $(D_{n-1})^{*k}$, where $(D_{n-1})^{*k}$ is the convolution of k copies of D_{n-1}. We only do the flattening procedure for the n_i. So for $n \notin \cup n_i$ the distribution $D_n = \tilde{D}_n$. For $n \in \cup n_i$, the distribution D_n is obtained by applying the flattening procedure described above to $\tilde{D}_n$.

First we repeat the properties that our distributions D_n have.

LEMMA 2.2. *The distribution D_n satisfies the inequality*

$$\mathrm{P}_{D_n}(x > t) \geq \mathrm{P}\big(S_{i+k^n} - S_i > tk^n | S_0, \ldots, S_i\big),$$

for every $t \in [-1, 1]$, integer i, and past $S_0, \ldots, S_i$.

PROOF. The proof is by induction. It is easy to check that each D_n, $\tilde{D}_n$ and intermediate distribution created in the flattening procedure satisfies the appropriate collection of inequalities. We leave it to the reader to check the details. □

The next step in the proof is bounding the variance of the distributions D_{n_i}.

LEMMA 2.3. *Now* $\sigma^2(D_{n_i}) \le 2f(k^{n_i})^2$ *for all* i. *This implies that* $4\sigma(\tilde{D}_{n_i}) < (1/8)f(k^{n_i})$ *for all* i.

PROOF. The proof is by induction. It is easy to check that $D_{n_1} = D_1$ satisfies this condition. This can be proved in the same way that line one is proved below. For the inductive step we bound the variance by bounding

$$\sum (x - m_{n_{i-1}})^2 \mathsf{P}_{D_{n_i}}(x).$$

This is done by estimating the contribution of the portion of D_{n_i} that is greater than or equal $m_{n_{i-1}}$, the part less than $m_{n_{i-1}}$ but greater than or equal to $m_{n_{i-1}} - 1.25f(k^{n_i})$ and the part that is less than $m_{n_{i-1}} - 1.25f(k^{n_i})$.

The flattening procedure only moves mass leftward. So mass that started out in the first region goes to points that are less than $m_{n_{i-1}}$ or to points greater than $m_{n_{i-1}}$ but still closer to $m_{n_{i-1}}$. Thus the contribution of each piece of mass to the sum above is less than or equal to its contribution to

$$\sum_{x \ge m_{n_{i-1}}} (x - m_{n_{i-1}})^2 \mathsf{P}_{\tilde{D}_{n_i}}(x).$$

So this sum provides a bound on the contribution of the first region.

A bound over the second region is obtained by assuming that every possible x in D_{n_i} has probability $\mathrm{PRE}(k^{n_i})$. This bound gives us

$$\mathrm{PRE}(k^{n_i}) \sum_{i=0}^{\lfloor 1.25/\mathrm{PRE}(k^{n_i}) \rfloor} \left(\frac{i}{k^{n_i}}\right)^2 < \left(\frac{1}{k^{n_i}\mathrm{PRE}(k^{n_i})}\right)^2 = f(k^{n_i})^2. \tag{2.1}$$

The third region is many standard deviations of $\tilde{D}_{n_i}$ away from $m_{n_{i-1}}$ so it has little mass under $\tilde{D}_{n_i}$. In fact it is so far away that the flattening procedure is unlikely to have much effect on the mass in this region. Thus the contribution of this region is likely to be only a little more than

$$\sum_{x \le m_{n_i} - \frac{9}{8} f(k^{n_i})} (x - m_{n_{i-1}})^2 \mathsf{P}_{\tilde{D}_{n_i}}(x).$$

We will show that each piece of mass gets moved by the flattening procedure only a small distance compared to the distance it was from $m_{n_{i-1}}$. So the contribution of each piece to the above sum is more than half of its contribution to $\sum (x - m_{n_{i-1}})^2 \mathsf{P}_{D_{n_i}}(x)$.

To make the bound of the third part rigorous, notice that the flattening procedure can move a piece of mass no more than $f(k^{n_i})$ to the left. (This can happen only if the distribution before flattening was supported on one point.)

So all the mass in $\tilde{D}_{n_i}$ which was in the interval $[m_{n_{i-1}} - \frac{1}{8}f(k^{n_i}), 1]$ must wind up in the interval $[m_{n_{i-1}} - \frac{9}{8}f(k^{n_i}), 1]$ after flattening. By the inductive hypothesis

$$4\sigma(\tilde{D}_{n_i}) = 4\sigma(D_{n_{i-1}})/k^{n_i - n_{i-1}} < 8f(k^{n_{i-1}})/k^{n_i - n_{i-1}}.$$

The way we chose the n_i implies that $4\sigma(\tilde{D}_{n_i}) < \frac{1}{8}f(k^{n_i})$. So a Chebyshev argument shows that there can be at most (1/16)th of the mass that is in the interval $[-1, m_{n_{i-1}} - \frac{1}{8}f(k^{n_i})]$ before flattening. Thus there can be at most (1/16)th of the mass that is in the interval $[-1, m_{n_{i-1}} - \frac{9}{8}f(k^{n_i})]$ after flattening.

Since at least (15/16)ths of the mass is in the interval $[m_{n_{i-1}} - \frac{9}{8}f(k^{n_i}), 1]$ after flattening, every piece of mass that starts out in the interval $[-1, m_{n_{i-1}} - \frac{9}{8}f(k^{n_i})]$ before flattening can be moved a distance at most $(1/16)f(k^{n_i})$. Thus the mass in the interval $[-1, m_{n_{i-1}} - 1.25f(k^{n_i})]$ in D_{n_i} was all in the interval $[-1, m_{n_{i-1}} - \frac{9}{8}f(k^{n_i})]$ in the distribution $\tilde{D}_{n_i}$. Thus the flattening procedure moved each piece of the mass in $\tilde{D}_{n_i}$ in this interval by a distance that is a small fraction of its distance from $m_{n_{i-1}}$. So a bound on the contribution of this portion is

$$2 \sum_{x \le m_{n_i} - \frac{9}{8}f(k^{n_i})} (x - m_{n_{i-1}})^2 \mathrm{P}_{\tilde{D}_{n_i}}(x).$$

Adding up these three bounds gives us

$$\begin{aligned}
\sigma^2(D_{n_i}) &\le \sum_{x \ge m_{n_{i-1}}} \left((x - m_{n_{i-1}})^2 \mathrm{P}_{\tilde{D}_{n_i}}(x)\right) + f(k^{n_i})^2 \\
&\quad + 2 \sum_{x \le m_{n_i} - \frac{9}{8}f(k^{n_i})} (x - m_{n_{i-1}})^2 \mathrm{P}_{\tilde{D}_{n_i}}(x) \\
&\le 2\sigma^2(\tilde{D}_{n_i}) + f(k^{n_i})^2 \\
&\le \frac{4f(k^{n_{i-1}})^2}{k^{n_i - n_{i-1}}} + f(k^{n_i})^2 \\
&\le 2f(k^{n_i})^2.
\end{aligned}$$

The next to last step is due to the inductive hypothesis and the definition of $\tilde{D}_{n_i}$. The last step is possible because of the way we chose the n_i. □

The previous lemma makes it easy to calculate how far the flattening procedure moves the mean of the distribution.

LEMMA 2.4. *Here* $m_{n_{i-1}} - m_{n_i} > (1/16)f(k^{n_i})$.

PROOF. Recall from the previous proof that $4\sigma(\tilde{D}_{n_i}) < \frac{1}{8}f(k^{n_i})$. This and a Chebyshev estimate show that at least (15/16)ths of the mass of $\tilde{D}_{n_i}$ is contained in the interval

$$\left[m_{n_{i-1}} - \tfrac{1}{8}f(k^{n_i}), m_{n_{i-1}} + \tfrac{1}{8}f(k^{n_i})\right].$$

Now we calculate an upper bound on how much mass D_{n_i} can put in that same interval. From this information we get a lower bound on how much the flattening procedure moved the mean of the distribution.

$$\mathrm{P}_{D_n}\left(|x - m_{n_{i-1}}| \le \tfrac{1}{8} f(k^{n_i})\right) \le \tfrac{1}{4} f(k^{n_i})(k^{n_i})\mathrm{PRE}(k^{n_i}) \le \tfrac{1}{4}.$$

So 15/16 of the mass of $\tilde{D}_{n_i}$ was in this region before the flattening procedure and at most 1/4 is left after the flattening is done. Thus at least half of the mass must be moved out of this interval. So the flattening must move at least 1/4 of the mass from this interval to spots at least $\frac{1}{4} f(k^{n_i})$ to the left of the interval. This implies that $m_{n_{i-1}} - m_{n_i} > (1/16) f(k^{n_i})$. □

PROOF OF THEOREM 1.2. Lemmas 2.4 and 2.1 imply that $m_{n_i} \searrow -\infty$. So for n large, Lemma 2.2 implies that

$$\mathrm{P}(S_{i+k^n} - S_i < -k^n) > 0.$$

This is a contradiction with the fact that S is a nearest neighbor process. □

The same proof can be used under more general hypotheses. The proof goes through almost verbatim for any process on the integers with bounded step size. More generally if the process has bounded step size in one direction and the mean of how far it can move in the other direction given any past is uniformly bounded, then the proof also applies. The proof can also be used to give bounds on the constants achievable in Theorem 1.1.

Acknowledgments. The author thanks Yuval Peres for suggesting the problem and Yuval Peres and Elchanan Mossell for a careful reading of the paper.

REFERENCES

[1] BENJAMINI, I., PEMANTLE, R. and PERES, Y. (1997). Unpredictable paths and percolation. *Ann. Probab.* To appear.

[2] GRIMMETT, G. R., KESTEN, H. and ZHANG, Y. (1993). Random walk on the infinite cluster of the percolation model. *Probab. Theory Related Fields* 96 33–44.

[3] LYONS, T. (1983) A simple criterion for transience of a reversible Markov chain. *Ann. Probab.* 11 393–402.

[4] HÄGGSTRÖM, O. and MOSSEL, E. (1998). Nearest-neighbor walks with low predictability profile and percolation in $2 + \varepsilon$ dimensions. *Ann. Probab.* 26 1212–1231.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742
E-MAIL: hoffman@math.umd.edu