THE UNIVERSITY OF UTAH
RESEARCH EXPERIENCE FOR UNDERGRADUATES
SUMMER 2002
LECTURE NOTES

Davar Khoshnevisan
Department of Mathematics
The University of Utah
`davar@math.utah.edu`
`http://www.math.utah.edu/~davar`

Contents

FOREWORD

These notes constitute sketches from my July and August lectures on random walks for the Summer REU Program at the Department of Mathematics, The University of Utah. In addition to my eleven lectures on the subject of random walks and their applications, we had a lecture by Prof. Nelson Beebe (U of U) on "*Random Number Generation and Network Security,*" as well as several lectures by the participants on their research.

Participants. There were three research groups comprised of undergraduate researchers, and three graduate students. The research groups presented one or two lectures each, all based on their summer research. It is expected that some or all of this will be ready in a report-format by the middle of the Fall 2002 semester. Also in attendance were John Schweitzer and Matthew Taylor.

The participating graduate students presented two lectures each on a topic of their choice. The participants and their lecture titles are listed below:

Graduate Students. Lars Louder (U of U) Title: *Random walks and electrical networks*; Sarah Geneser (U of U) Title: *Matlab tutorials*; Robert Thorpe (U of U) Title: *Three Games of chance*.

Undergraduate Students. The attending undergradautes' lectures/teams were:

Team 1. Micah Allred (BYU) Amanda Ellis (U of U). Title: *The Mathematics of Finance, Numerical Solutions to Stochastic Differential Equations, and Simulating the Wright–Fischer Model for Gene Frequencies.*

Team 2. Rex Butler (U of U). Title: *The Linear and Nonlinear Voter Models, Random Cellular Automata, and Interacting Particle Systems.*

Team 3. Ron McKay and Song Du (U of U). Title: *Numerical Computation of Hitting Probabilities of Brownian Motion.*

Davar Khoshnevisan
Salt Lake City, UT, August 2002

LECTURE 1: THE SIMPLE WALK

The *simple random walk* is a mathematical motion for one-dimensional molecular motion, and is defined as follows: At time $n = 0$, the particle's position is $S_0 = 0$. Then you toss a fair coin to go left or right with probability $\frac{1}{2}$ each. Let S_1 denote the position of the particle at time 1 obtained in this way. Now repeat the process, making sure that everytime you toss a coin, it is tossed independently of the coin preceding it. This gives you a random (or stochastic) process $S := \{S_n\}_{n\geq 1}$.

You can think of the process S as a random "dynamical system." It is a dynamical system roughly because you apply the same procedure at time n to determine the value at time $n+1$; it is random since this procedure involves random tosses of coins.

§1. A COMBINATORIAL INTERPRETATION

Suppose you want to know the probability that the random process S has "done something before time n." For instance, what is the probability that some time before time n, the random walk passed the point k. (In symbols, $P\{\max_{1\leq j\leq n} S_j \geq k\} =?$) Or, what is the probability that you never hit zero before time n (In symbols, $P\{\min_{1\leq j\leq n} S_j > 0\} =?$)

Combinatorics (or counting) give us one way to make such calculations. Let us say that $\pi_0, \pi_1, \pi_2, \ldots, \pi_n$ is a *path of length n* if $\pi_0 = 0$, and for all $1 \leq i \leq n$, $|\pi_{i+1} - \pi_i| = 1$. Note that each realization of the random walk by time n gives a path of length n.

(1.1) Observation. *There are 2^n paths of length n. Moreover, if $\pi_0, \ldots, \pi_n$ is any given path of length n, then*

$$P\{S_1 = \pi_1, \ldots, S_n = \pi_n\} = 2^{-n}.$$

In other words, all paths are equally likely to be the random walk path. This is an easy exercise.

§2. A PROBABILISTIC INTERPRETATION

For $i = 1, 2, \cdots$ define $X_i := S_i - S_{i-1}$. The values $X_1, X_2, \ldots$ are the displacement values at times $1, 2, \cdots$. In other words, if the coin at time j told us to go to the right, then $X_j = +1$, else $X_j = -1$. Since the coins were independent, the X_i's are independent random variables. Finally, they all have the same distribution which is given by $P\{X = -1\} = P\{X = +1\} = \frac{1}{2}$. Finally, note that $S_n = X_1 + \cdots + X_n$.

Notation. *Any process of the form $T_n = Y_1 + \cdots + Y_n$, where the Y_i's are independent and identically distributed, is called a random walk. In particular, the simple walk is a random walk.*

§3. PRELIMINARY CALCULATIONS

Let us compute a few momonts to get a feeling for the behavior of the simple walk S. First,

$$E\{S_n\} = E\{X_1\} + \cdots + E\{X_n\}.$$

But the X_i's are have the same distribution, and so they all have the same expectation, which is $E\{X\} = 1 \times P\{X = 1\} + (-1) \times P\{X = -1\} = 1 \times \frac{1}{2} + (-1) \times \frac{1}{2} = 0$. Therefore, we have

(3.1) Expected Value. *For each n, $E\{S_n\} = 0$.*

Suppose you are playing a fair game many times in succession. Everytime you play, the probability of winning a dollar is the same as that of losing (i.e., $= \frac{1}{2}$), and you play the game independently each time. Then, S_n is the fortune (if > 0 and loss if ≤ 0) that you have amassed by time n. The above tells us that you expect to come out even in a fair game. Not a surprise. But there are fluctuations and the expected fluctuation is the standard deviation, i.e., the square root of the variance.

(3.2) Variance. *For each n, $\mathrm{Var}(S_n) = n$.*

Proof: In order to make this computation, recall that for any random variable Y, $\mathrm{Var}(Y) = E(Y^2) - |E\{Y\}|^2$. Therefore, $\mathrm{Var}(S_n) = E\{S_n^2\}$. We compute this as follows: First note that

$$S_n^2 = (X_1 + \cdots + X_n)^2 = \sum_{j=1}^{n} X_j^2 + \sum\sum_{i \neq j} X_i X_j.$$

When $i \neq j$, X_i and X_j are independent, so $E\{X_i X_j\} = E\{X_i\}E\{X_j\}$, which is 0. Therefore, $E\{S_n^2\} = \sum_{j=1}^{n} E\{X_j^2\} = nE\{X^2\}$. But $E\{X^2\} = 1^2 \times P\{X + 1\} + (-1)^2 \times P\{X = -1\} = 1$, which shows us that the variance of S_n is indeed n. ♣

On the other hand, we could get an even better idea of the size of S_n by computing higher moments. Note that $E\{S_n^4\} = E\{|S_n - E(S_n)|^4\}$.

(3.3) Fourth Moment. *For each n, $E\{S_n^4\} = 3n^2 - 2n$.*

Proof: We proceed as before and expand S_n^4:

$$\begin{aligned} S_n^4 = & \sum_{i=1}^{n} X_i^4 + \binom{4}{2} \cdot \frac{1}{2} \sum\sum_{i \neq j} X_i^2 X_j^2 \\ & + \binom{4}{3} \sum\sum_{i \neq j} X_i X_j^3 + \frac{4!}{1! \cdot 1! \cdot 2!} \cdot \frac{1}{2} \sum\sum\sum_{i \neq j \neq k} X_i X_j X_k^2 \\ & + \frac{4!}{1! \cdot 1! \cdot 1! 1!} \sum\sum\sum\sum_{i \neq j \neq k \neq l} X_i X_j X_k X_l. \end{aligned}$$

By the independence of the X's, and since their means are 0, after we take expectations, only the first two terms contribute, i.e.,

$$E\{S_n^4\} = nE\{X^4\} + \frac{4!}{2! \cdot 2!} \frac{n(n-1)}{2} \left(E\{X^2\}\right)^2 = nE\{X^4\} + 3n(n-1) \left(E\{X^2\}\right)^2.$$

But we have already seen that $E\{X^2\} = 1$, and one computes just as easily that $E\{X^4\} = 1$. The calculation of the fourth moment follows. ♣

§4. CHEBYSHEV'S AND MARKOV'S INEQUALITIES

The Markov, and more generally, the Chebyshev inequality are inequalities that state that for random variables that have sufficiently many moments are large with very little probability.

(4.1) Markov's Inequality. *Suppose X is a nonnegative random variable. Then for all $\lambda > 0$,*

$$P\{X \geq \lambda\} \leq \frac{E\{X\}}{\lambda}.$$

Proof: For any number (random not) $X \geq 0$, we have $X \geq X\mathbf{1}_{\{X\geq\lambda\}} \geq \lambda\mathbf{1}_{\{X\geq\lambda\}}$, where $\mathbf{1}_A$ is the indicator of the event A, i.e.,

$$\mathbf{1}_A = \begin{cases} 1, & \text{if } A \text{ happens,} \\ 0, & \text{if } A^c \text{ happens.} \end{cases} \tag{4.2}$$

Therefore, we take expectations to deduce that

$$E\{X\} \geq \lambda E\left(\mathbf{1}_{\{X\geq\lambda\}}\right) = \lambda P\{X \geq \lambda\}, \tag{4.3}$$

since for any random event A, $E(\mathbf{1}_A) = 1 \times P\{A\} + 0 \times P\{A^c\} = P\{A\}$. Divide (4.3) by $\lambda > 0$ to get Markov's inequality. ♣

Markov's inequality states that if $X \geq 0$ has a finite mean, then the probability that X is large is very small. If X has more moments, this probability is even smaller in sense.

(4.4) Chebyshev's Inequality. *Suppose X is a random variable that has a finite variance, and let $\mu := E\{X\}$ denote its means. Then for all $\lambda > 0$,*

$$P\{|X-\mu| \geq \lambda\} \leq \frac{\mathrm{Var}(X)}{\lambda^2}.$$

Proof: Let $Y := |X-\mu|^2$ and note that $P\{|X-\mu| \geq \lambda\} = P\{Y \geq \lambda^2\}$. Since $E\{Y^2\} = \mathrm{Var}(X)$, apply Markov's inequality to finish. ♣

There are higher-moment versions of Chebyshev's inequality. Here is one. I will omit the proof, since it is the same as that of (4.4).

(4.5) Chebyshev's Inequality for Fourth Moments. *Suppose X is a random variable that has a finite fourth moment, and suppose $E\{X\} = 0$. Then for all $\lambda > 0$,*

$$P\{|X| \geq \lambda\} \leq \frac{E\{X^4\}}{\lambda^4}.$$

LECTURE 2: THE SIMPLE WALK IN DIMENSION ONE

Laws of large numbers are a class of results that state that, in one way or another, averaging many independent random quantities yields their expectation as long as you average enough things.

For example, suppose you wanted to know the average output α of a machine. If you could simulate the output of this machine on your computer, it would be natural to run several simulations, average the outputs, and declare that as an "estimate" for α. The following shows that this procedure actually works. You may need to refer to §2 of Lecture 1 for further motivation.

(0.1) Kolmogorov's Strong Law of Large Numbers. *Suppose $X_1, X_2, \ldots$ are independent and identically distributed. If $S_n := X_1 + \cdots + X_n$ denote the corresponding random walk, and if $\mu := E\{X_1\}$ exists, then*

$$P\left\{ \lim_{n\to\infty} \frac{S_n}{n} = \mu \right\} = 1.$$

In the unbiased case where $\mu = 0$, this shows that the asymptotic value of the walk is much smaller than n. In fact, in most of these cases, the asymptotic value is of order $\sqrt{n}$.

(0.2) The Central Limit Theorem. *Suppose $X_1, X_2, \ldots$ are independent and identically distributed. If $S_n := X_1 + \cdots + X_n$ denote the corresponding random walk, and if $E\{X_1\} = 0$ and $0 < \sigma^2 := \mathrm{Var}(X_1) < +\infty$, then for any real number x,*

$$\lim_{n\to\infty} P\left\{ \frac{S_n}{\sqrt{n}} \le x \right\} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} e^{-y^2/2\sigma^2}\, dy.$$

In the physics literature, this type of $\sqrt{n}$-growth is referred to as "diffusive."

§1. THE STRONG LAW FOR THE SIMPLE WALK

Once again, S_n is now the simple walk (on the integer lattice). While the general form of the Kolmogorov strong law is a rather difficult result, for the simple walk, things are not so bad as we shall see.

Here is a start: Let us apply Chebyshev's inequality from (4.4) of Lecture 1 to see that for any $\varepsilon > 0$,

$$P\{|S_n| \ge n\varepsilon\} \le \frac{\mathrm{Var}(S_n)}{n^2\varepsilon^2} = \frac{1}{n\varepsilon^2}. \tag{1.1}$$

We are using two more facts from Lecture 1. Namely, that the expectation of S_n is zero (3.1, Lecture 1) and its variance is n (3.2, Lecture 1). This shows that for *any* $\varepsilon > 0$ (however small),

$$\lim_{n\to\infty} P\left\{ \left|\frac{S_n}{n}\right| \ge \varepsilon \right\} = 0.$$

This is not quite as strong as the strong law, but it has the right flavor. We will enhance this calculation to get the strong law.

Proof of The Strong Law For the Simple Walk: We can improve (1.1) by using higher moments than the second moment (i.e., the variance). Namely, let us use the Chebyshev inequality for fourth moments (4.5, Lecture 1) and the fact that $E\{S_n^4\} = 3n^2 - 2n \le 3n^2$ (3.3, Lecture 1) to obtain the following: For all $\varepsilon > 0$,

$$P\{|S_n| \ge n\varepsilon\} \le \frac{E\{S_n^4\}}{\varepsilon^4 n^4} \le \frac{3}{\varepsilon^2 n^2}. \tag{1.2}$$

So in fact the abov probability goes to zero faster than the rate of $(n\varepsilon^2)^{-1}$ stated in (1.1). Now let $\mathcal{N}$ denote the number of times the random walk is at least $n\varepsilon$ units away from the origin. That is,

$$\mathcal{N} := \sum_{n=1}^{\infty} \mathbf{1}_{\{|S_n| \ge n\varepsilon\}},$$

where $\mathbf{1}_A$ is the indicator of the event A; cf. (4.2, Lecture 1). Since $E\{\mathbf{1}_A\} = P\{A\}$, $E\{\mathcal{N}\} = \sum_{n=1}^{\infty} P\{|S_n| \ge n\varepsilon\}$. In particular, by (1.2) above, and using the fact that $1, \frac{1}{4}, \frac{1}{9}, \ldots, \frac{1}{n^2}, \cdots$ is a summable sequence, we see that $E\{\mathcal{N}\} < +\infty$. This means that $\mathcal{N}$ is finite with probability one. In other words, we have shown that with probability one, for any $\varepsilon > 0$, there exists a random time $\mathcal{N}$ past which $|S_n| \le n\varepsilon$. This is the same as saying that with probability one, $S_n/n \to 0$. ♣

§2. RETURNS TO THE ORIGIN

What we have done is to show that S_n is much smaller than n as $n \to \infty$. One rough explanation for this is that S_n is fluctuating as $n \to \infty$; so much so that it has little time to go very far from the origin. This is one of the reasons that the movement of the simple walk has proven to be an important model for "one-dimensional molecular motion." (The more realistic three-dimensional setting will be covered soon.)

One way in which we can study the said fluctuation phenomenon more precisely, is by considering the notion of *recurrence*. In the context of nonrandom dynamical systems, this notion is due to the work of H. Poincaré.

Remember that S_0 is zero. That means that the random walk always starts at the origin. So it makes sense to consider N_n which is the number of returns to the origin by time n; i.e.,

$$N_n := \sum_{j=1}^{n} \mathbf{1}_{\{S_j = 0\}}, \qquad n = 1, 2, 3, \ldots.$$

(2.1) The Expected Number of Returns. *As $n \to \infty$, $E\{N_n\} \sim \sqrt{2n/\pi}$, where $a_n \sim b_n$ means that $a_n/b_n \to 1$ as $n \to \infty$.*

Proof: Note that

$$E\{N_n\} = E\left[\sum_{j=1}^{n} \mathbf{1}_{\{S_j=0\}}\right] = \sum_{j=1}^{n} P\{S_j = 0\}.$$

So it suffices to estimate $P\{S_j = 0\}$ for $j \to \infty$. First, we note that if j is an odd number $S_j \neq 0$. So it suffices to estimate $E\{N_n\}$ for n even. Moreover, if n is even,

$$E\{N_n\} = \sum_{j=1}^{n/2} P\{S_{2j} = 0\}.$$

Here is where combinatorics come in: Thanks to (1.1, Lecture 1), the probability that $S_{2j} = 0$ is equal to 2^{-2j} times the number of paths of length $2j$ such that at time j the path is at 0. Any such path $\pi_0, \ldots, \pi_{2j}$ hits 0 at time j if and only if it has gone to the right exactly j times, and gone to the left exactly j times. There are $\binom{2j}{j}$-many ways for choosing where these rights and lefts are, so

$$P\{S_{2j} = 0\} = 2^{-2j}\binom{2j}{j}.$$

This and the preceding display, together show

$$E\{N_n\} = \sum_{j=1}^{n/2} 2^{-2j}\binom{2j}{j}. \tag{2.2}$$

But $\binom{2j}{j} = (2j)!/(j!)^2$, and this can be estimated by

(2.3) Stirling's Formula. *As $k \to \infty$, $k! \sim \sqrt{2\pi} k^{k+\frac{1}{2}} e^k$.*

We use this to see that

$$\begin{aligned} E\{N_n\} &\sim \sum_{j=1}^{n/2} 2^{-2j} \frac{\sqrt{2\pi}(2j)^{2j+\frac{1}{2}} e^{-2j}}{\left(\sqrt{2\pi} j^{j+\frac{1}{2}} e^{-j}\right)^2} = \sum_{j=1}^{n/2} \frac{1}{\sqrt{2\pi}} \frac{2^{\frac{1}{2}}}{j^{\frac{1}{2}}} = \sqrt{\frac{1}{\pi}} \sum_{j=1}^{n} \frac{1}{\sqrt{j}} \\ &= \sqrt{n} \cdot \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\sqrt{j/n}}. \end{aligned} \tag{2.4}$$

But $\frac{1}{n}\sum_{j=1}^{nT} f(j/n) \to \int_0^T f(x)\,dx$ if f is continuous; in fact this is the Riemann-sum approximation of the calculus of real functions. Apply this with $f(x) := 1/\sqrt{x}$ to see that $(1/n)\sum_{j=1}^{n} 1/\sqrt{j/n} \sim \sqrt{n} \cdot \int_0^{1/2} 1/\sqrt{x}\,dx = \sqrt{2n}$. Together with (2.4), this completes our asymptotic evaluation of $E\{N_n\}$. ♣

§3. THE REFLECTION PRINCIPLE

Here is another application of the combinatorial way of thinking. This is a deep result from the 1887 work of D. André:

(3.1) The Reflection Principle. *For any* $\lambda, n = 1, 2, \ldots$,

$$P\left\{\max_{1\le j\le n} S_j \ge \lambda\right\} = 2P\{S_n \ge \lambda\}.$$

Proof: The combinatorial representation of the simple walk (1.1, Lecture 1) tells us that the above is *equivalent* to showing that

$$\begin{aligned}\#\{\text{paths that go over } \lambda \text{ before time } n\} \\ = 2 \times \#\{\text{paths that are go over } \lambda \text{ at time } n\}.\end{aligned} \tag{3.2}$$

There are two types of paths that go over λ before time n: The first are those that are over λ at time n, i.e., those paths for which $\pi_n \ge \lambda$ (*Type 1*). The second (*Type 2*) are those that go over λ some time before time n and then go below it so that at time n, $\pi_n < \lambda$. If you think about it for a moment, you will see that (3.2) is really stating that the number of paths of *Type 2* is equal to the number of paths of *Type 1*. But this is clear from a picture; for example, see the picture at

`http://www.math.utah.edu/~davar/REU-2002/notes/lec2.html`.

Namely, any path of *Type 2*, can be reflected about the line $y = \lambda$ at the first time it hits λ. This gives a paths of *Type 1*. Conversely, any paths of *Type 1* can be reflected to give a path of *Type 2*. This shows that there are as many paths of each type, and we are done. ♣

§4. APPENDIX: STIRLING'S FORMULA

It would be a shame for you not to see why Stirling's formula (2.3 above) is true; so I have added this section to explain it, although we did not discuss this section's material in our meeting.

Consider $\ln(k!) = \sum_{i=2}^{k} \ln(i)$. By the integral test of calculus,

$$\int_1^k \ln(x)\,dx \le \ln(k!) \le \int_1^{k+1} \ln(x)\,dx.$$

But $\int_1^T \ln(x)\,dx = T\ln(T) - 1$. Therefore,

$$k\ln(k) - 1 \le \ln(k!) \le (k+1)\ln(k+1) - 1. \tag{4.1}$$

Now, recall Taylor's expansions for $\ln(1+y)$:

$$\ln(1+y) = 1 + y - \frac{y^2}{2} + \cdots. \tag{4.2}$$

We don't apply this to $\ln(k+1)$ but rather note that $\ln(k+1) = \ln(k) + \ln((k+1)/k)) = \ln(k) + \ln(1 + \frac{1}{k})$. Apply (4.2) with $y = \frac{1}{k}$ to deduce that

$$\ln(k+1) = \ln(k) + \frac{1}{k} - \frac{1}{2k^2} + \cdots.$$

Put this back in to (4.1) to get

$$\begin{aligned} k\ln(k) \le \ln(k!) &\le (k+1)\left[\ln(k) + \frac{1}{k} - \frac{1}{2k^2} + \cdots\right] \\ &\le (k+1)\left[\ln(k) + \frac{1}{k}\right] \\ &= k\ln(k) + \ln(k) + 1 + \frac{1}{k}. \end{aligned}$$

Since the exponential of $k \ln k$ is k^k, we can exponentiate the above inequalities to obtain

$$k^k \le k! \le k^{k+1} \times e^{1+\frac{1}{k}} \sim ek^{k+1}.$$

Stirling's formula is a much sharper version of these bounds. (For instance note that both sides are off by $k^{\frac{1}{2}}$ to the leading order.)

LECTURE 3: THE SIMPLE WALK IN HIGH DIMENSIONS

Let us continue our discussion on the simple walks to higher dimensions. To do so, it helps to introduce a more abstract walk first (and briefly).

§1. THE SIMPLE WALK ON A GRAPH

(1.1) Graphs. A *graph* is a collection of points (or *vertices*) and a set of neighboring relations (*edges*) between these vertices. An example of a graph is $\mathbf{Z}^1$—the one-dimensional integer lattice—which can be thought of as a graph: The vertices are $0, \pm 1, \pm 2, \ldots$ and there is an edge between two vertices a and b if and only if $|a-b| = 1$. In particular, every vertex has two neighbors.

An obvious generalization to this is $\mathbf{Z}^d$, which is the d-dimensional integer lattice. This can be thought of as a graph with vertices of type $(z_1, \ldots, z_d)$ where the z_i's are integers, and there is an edge between $z = (z_1, \ldots, z_d)$ and $w = (w_1, \ldots, w_d)$ if and only if $\sum_{i=1}^d |w_i - z_i| = 1$ (check this!) So every vertex has $(2d)$ neigbors on this graph. (Can you compute this from the formal definition that I have written?)

A third example of an interesting graph is a *binary tree*. Here, you start with one vertex; it then branches into two; each of these branches into two, and so on. Check that at the nth level of this construction, there are 2^n vertices. The edges are the natural ones: Two vertices are neighbors (i.e., have an edge in common) if and only if one of them branched off into the other. You should check that evey vertex except for the first one (the *root*) has three neighbors, whereas the root has two neighbors.

As a fourth and final example, consider the *complete graph on n vertices*. Here, the graph is made up of a finite number (n) of vertices, and everyone is the neighbor of everyone else.

(1.2) The Simple Walk. The simple walk on a graph is the random process that starts someplace in the graph (call it the origin if you want), and then moves to one of the nearest neighboring vertices with equal probability. (**Warning:** This makes sense only if the graph has no vertices with infinitely many neighbors, of course.) And the walk proceeds this way, everytime going to a nearestneighbor independently of all his/her other moves, and always, all neighbors are equally likely.

§2. THE SIMPLE WALK ON $\mathbf{Z}^2$

Returning to $S_1, S_2, \ldots$ being the simple random walk on the planar integer lattice $\mathbf{Z}^2$, we ask, "*how many times is the walk expected to return to its origin?*" We have already seen in (2.1, Lecture 2) that the one-dimensional walk returns to the origin about $\sqrt{n}$-times in the first n steps, as $n \to \infty$. One should expect fewer returns for the planar walk, since there is "more space." Here is the precise result.

(2.1) Expected Number of Returns. *If N_n denotes the number of times the simple walk returns to the original before time n, then for n even,*

$$E\{N_n\} = \sum_{j=1}^{n/2} 4^{-2j} \binom{2j}{j}^2.$$

In particular, for some constant c, $E\{N_n\} \sim c\log(n)$.

A Semi-Proof: I gave a geometric proof of this in the lecture; the idea was that if you rotate the xy-plane, you rotate the simple walk S_n on to the simple walk $\widetilde{S}_n$ which is a simple walk on the lattice in which the neighbors of the origin $(0,0)$ are the 4 points,

$$\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right), \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right).$$

Since we have only turned the plane, $S_n = 0$ if and only if $\widetilde{S}_n = 0$, so these two events have the same probability, but $P\{\widetilde{S}_n = 0\} \sim c/\sqrt{n}$ (cf. the Stirling-formula approximation in (2.4, Lecture 2)). So, $P\{\widetilde{S}_n = 0\} \sim C^2/n$. On the other hand, just as in the one-dimensional case, $E\{N_n\} = \sum_{j=1}^n P\{S_j = 0\}$, so that $E\{N_n\} \sim \sum_{j=1}^n C^2/j$. Let us see how this sum behaves:

$$\begin{aligned} E\{N_n\} &\sim C^2 \sum_{j=1}^n \frac{1}{j} = C^2 \frac{1}{n} \sum_{j=1}^n \frac{1}{(j/n)} \\ &\sim C^2 \int_{1/n}^1 \ln(x)\, dx, \end{aligned}$$

by a Riemann-Sum approximation. (How did the lower limit of the integral become $(1/n)$?) As $n \to \infty$, this behaves like $c\log(n)$—check!

When done carefully, as we did in the lecture, the exact calculation follows also. ♣

§3. THE SIMPLE WALK ON $\mathbf{Z}^d$, $\mathbf{d} \geq 3$

In higher dimensions the rotation trick fails, but our intuition that the coordinates of S_n are almost independent simple walks is in a sense correct and can be made precise. This leads to $P\{S_n = 0\} \sim (C/\sqrt{n})^d = cn^{-d/2}$. On the other hand, since $d \geq 3$, this sums and we have

(3.1) The Simple Walk in $\mathbf{d} \geq 3$ is transient. *We have $E\{N_\infty\} < +\infty$. Therefore, the expected number of times to hit any point is finite. Therefore, after a finite (but random) number of steps, S_n will leave any finite neighborhood of the origin, and this is the property that the word "transient" is referring to.*

§4. THE SELF-AVOIDING WALK Certain models of polymer chemistry lead to the *self–avoiding walk*, which is defined as follows: First consider all paths of length n in your favorite infinite lattice, say $\mathbf{Z}^d$. On the latter, there are $(2d)^n$ such paths, but many of them self-intersect, i.e., there are distinct $i, j \leq n$ such that $\pi_i = \pi_j$. Let χ_n denote the total number of self-avoiding paths of length n, and from these χ_n self-avoiding paths, choose one at random. This is the *self-avoiding walk* of length n.

(4.1) Bounds on χ_n. *I claim that for every n, $d^n \leq \chi_n \leq (2d)^n$.*

Actually much better bounds are possible (say when $d = 2$), but this is good enough.

Proof: To get the upper bound of $(2d)^n$ note that every self-avoiding path is a path, and so $\chi_n \leq$ the number of *all* paths of length n, which is $(2d)^n$. The lower bound is not much more difficult. When $d = 2$, note that every path that only goes "up" or to the "right" is self-avoiding. There are clearly 2^n such paths. Note that paths of this type (i.e., the "up-right" paths) are those that move in the direction of either vector $(1, 0)$ or $(0, 1)$.

When $d = 3$, the analogue of "up-right" paths are those that move in the direction of $(1, 0, 0)$, $(0, 1, 0)$, $(0, 0, 1)$. There are 3^n such paths. In general, only choose the directions that keep you going "up" in the positive quadrant, and note that these paths are (i) self-avoiding; and (ii) there are d^n many of them. ♣

(4.2) The Connectivity Constant $\mathbf{C(d)}$. *There exists a constant $d \leq C(d) \leq 2d$, such that*

$$\lim_{n\to\infty} \frac{\chi_n}{n} = C(d).$$

This $C(d)$ is called the connectivity constant.

(4.3) Remarks.

- **a.** Such a result holds on many infinite graphs that are "self-similar."
- **b.** In rough terms, the above states that χ_n behaves (roughly again!) like $(C(d))^n$ for large value of n.

Proof: Note that on every self-avoiding path of length $n+m$, certainly the first n steps are self-avoiding, and the next m steps are also self-avoiding. Therefore,

$$\chi_{n+m} \leq \chi_n \cdot \chi_m.$$

In words, the sequence $\chi_1, \chi_2, \ldots$ is *submultiplicative*. This is equivalent to the *subadditivity* of $\log(\chi_n)$'s, i.e.,

$$\log(\chi_{n+m}) \leq \log(\chi_n) + \log(\chi_m).$$

Therefore, by the subbadditivity lemma below, $\log(\chi_n)/n$ has a limit. Note that this limit is between d and $(2d)$ by (4.1). ♣

(4.4) The Subadditivity Lemma. *Any sequence $a_1, a_2, \ldots$ that is subadditive (i.e., $a_{n+m} \leq a_n + a_m$) satisfies*

$$\lim_{k\to\infty} \frac{a_k}{k} = \min_{n\geq 1} \left(\frac{a_n}{n}\right).$$

In particular, the above limit is always $\leq a_1$ which is finite. However, this limit could be $-\infty$!

(4.5) Limits. I will prove this shortly. However, we need to be careful when dealing with limits, especially since the entire point of this exercise is to show that the limit exists. So let us start with some preliminaries: For any sequence $x_1, x_2, \ldots$

$$\limsup_{k\to\infty} x_k := \min_{n\geq 1} \max_{j\geq n} x_j, \qquad \text{and} \qquad \liminf_{k\to\infty} x_k := \max_{n\geq 1} \min_{j\geq n} x_j.$$

In other words, the $\limsup$ is the largest possible accumulation point of the x_j's and the $\liminf$ is the smallest. It should be obvious that for any sequence $x_1, x_2 \ldots$, we always have $\liminf_j x_j \leq \limsup_j x_j$. When the two are equal, this value is the limit $\lim_j x_j$, and this is the only case in which the limit exists.

> *Exercise 1.* For our first example, consider the sequence $x_j := 1/j$ ($j = 1, 2 \cdots$). Then you should check that $\liminf_{j\to\infty} x_j = \limsup_{j\to\infty} = 0$. More generally, check that for any sequence $x_1, x_2, \ldots$, $\lim_{j\to\infty} x_j$ exists if and only if $\liminf_{j\to\infty} x_j = \limsup_{j\to\infty} x_j$.
>
> *Exercise 2.* Show that the sequence $x_j := (-1)^j/j$ ($j = 1, 2, \ldots$) has no limit. Do this by explicitly computing $\liminf_j x_j$ and $\limsup_j x_j$.
>
> *Exercise 3.* A point a is defined to be an `accumulation point` for the sequence $x_1, x_2, \ldots$ if there exists a subsequence $n(k) \to \infty$, such that $x_{n(k)} \to a$. Show that $\limsup_j x_j$ and $\liminf_j x_j$ are always accumulation points of (x_j).
>
> *Exercise 4.* Show that the sequence of Exercise 2 only has 2 accumulation points. Construct a sequence $x_1, x_2, \ldots$ that has k accumulation points for any predescribed integer k. Can you construct a sequence $x_1, x_2, \ldots$ that has infinitely many accumulation points?

Now we are ready for

(4.6) Proof of (4.4). Since $a_k/k \geq \min_n(a_n/n)$ for any k, it follows that

$$\liminf_{k\to\infty} \frac{a_k}{k} \geq \min_n \left(\frac{a_n}{n}\right).$$

It suffices to show that $\limsup_{k\to\infty}(a_k/k) \leq \min_n(a_n/n)$. (For then, the $\limsup$ and the $\liminf$ agree.) We do this in a few easy stages: Thanks to subbadditivity, $a_k \leq a_{k-1} + a_1$. But the same inequality shows that $a_{k-1} \leq a_{k-2} + a_1$, so that by iterating this we get

$$\begin{aligned} a_k &\leq a_{k-1} + a_1 \\ &\leq a_{k-2} + a_1 + a_1 = a_{k-2} + 2a_1 \\ &\leq a_{k-3} + 3a_1 \\ &\;\;\vdots \\ &\leq k a_1. \end{aligned}$$

Therefore, $\limsup_k(a_k/k) \leq a_1$. Next, we show that this $\limsup$ is also $\leq (a_2/2)$. "By induction," this argument boosts itself up to show that for any n, $\limsup_k(a_k/k) \leq (a_n/n)$, which is what we want to show but in disguise.

To finish, I will show that

$$\limsup_{k\to\infty} \frac{a_k}{k} \leq \frac{a_2}{2}. \tag{4.7}$$

I will then leave the "induction" part up to you as a nice exercise.

By subbaditivity, for all $k > 2$, $a_k \le a_{k-2} + a_2$. Applying it again, subadditivity yields $a_k \le a_{k-4} + 2a_2$ for all $k > 4$ and so on. In general, we see that for all $k > 2j$,

$$a_k \le a_{k-2j} + ja_2. \tag{4.8}$$

Now, if k is even, choose $j = (k/2) - 1$ to see that (a) $k > 2j$; and so (b) $a_k \le (k/2)a_2$. If k is odd, choose $j = (k-1)/2$ to see that (c) $k > 2j$; and so (d) $a_k \le a_1 + \frac{k-1}{2}a_2$. So regardless of whether or not k is even, we always have

$$a_k \le \left(\frac{k}{2}\right) a_2 + |a_1| + |a_2|.$$

(why?) Divide by k and let $k \to \infty$ to deduce (4.7). ♣

(4.9) Exercise on the Connectivity Constant. Improve (4.1) by showing that in all dimensions, $\chi_n \le (2d)\cdot(2d-1)^{n-1}$. Conclude from this and from (4.2) the following slightly better bound on the connectivity constant: $d \le C(d) \le (2d-1)$, e.g., $2 \le C(2) \le 3$.
(`Hint.` For step 1, you have $(2d)$ choices, but then you cannot go back to where you were.)

LECTURE 4: OTHER RELATED MODELS

Having discussed some of the basics for simple walks, I will start talking about various different related model, to help you choose a research topic as soon as possible.

§1. MORE ON THE SELF-AVOIDING WALK

Recall the self-avoiding walk from §4 of Lecture 3. A number of interesting questions present themselves that you may wish to think about:

♠ `Open Problem 1.` One can show, by appealing to subadditivity again, that the self-avoiding walk on any regular lattice $\mathbf{L}$ satisfies

$$\lim_{n\to\infty} \frac{\log \chi_n}{n} = c(\mathbf{L}),$$

exists, and $c(\mathbf{L})$ is the *connectivity constant* of the lattice $\mathbf{L}$. In particular, $c(\mathbf{Z}^d)$ is nothing but the constant $C(d)$ of Lecture 3, §4. Two possible starting points for your research along these lines:

- ♡ Find a numerical method for computing $c(\mathbf{L})$ for some interesting lattices $\mathbf{L}$ such as $\mathbf{Z}^2$, $\mathbf{Z}^3$, or the hexagonal lattice.
- ♡ Can you numerically verify the conjecture that for the honeycomb lattice, the connectivity constant is $\sqrt{\sqrt{2}+2}$?

♠ `Open Problem 2.` It is conjectured that, quite generally, there exists a *universal constant* α such that on any lattice $\mathbf{L} \subseteq \mathbf{Z}^d$, χ_n grows like a constant times $n^\alpha (c(\mathbf{L}))^n$. The point is that α is supposed to be independent of the choice of the lattice but can (and ought to) depend on the ambient dimension d. Assuming this hypothesis, can you find a numerical approximation for α when $d = 2$? Physicists conjecture that when $d = 2$, $\alpha = \frac{43}{32}$.

- ♡ For any $x \in \mathbf{R}^d$, let $|x|^2 := x_1^2 + \cdots + x_d^2$ denote the square of the distance between x and the origin. If X_n denotes the position of the randomly selected self-avoiding path of length n in $\mathbf{Z}^d$, what is $A_n := E\{|X_n|^2\}$? It is conjectured that A_n should grow like a constant times n^β for some constant β. Can you numerically estimate β? When $d = 2$, β is conjectured to be $\frac{3}{4}$. This would suggest that self-avoiding walks grow faster than "diffusions," which is why this type of growth is called "super-diffusive." (Another phrase that refers to this property is "anamolous diffusion.")
- ♡ Can you find numerical ways to estimate $E\{|X_n|^p\}$ for some other values of $p > 2$ as well? It is conjectured that $E\{|X_n|^p\}$ should behave like the square of $E\{|X_n|^{p/2}\}$ as $n \to \infty$.

§2. DIFFUSION-LIMITED AGGREGATION (DLA)

Diffusion-limited aggregation (or DLA) is a mathematical model devised by Witten and Sanders to model crystal growth. The aim is to grow a random set in $\mathbf{Z}^d$, in successive time-steps, in order to obtain a sequence of growing random sets $A(0) \subseteq A(1) \subseteq \cdots$.

To start with, set $A(0) = \{0\}$ so this means that at "time" 0, the "crystal" is a point. Then start a random walk from infinitely (or nearly) far away from $A(0)$ and wait until this random walk hits a neighbor of $A(0)$. This defines the crystal at time 1; namely, let $A(1)$ be the set comprised of the origin together with this last value of the stopped random walk. Having created $A(n)$, create $A(n+1)$ by, once again, starting an independent random walk infinitely far awar from $A(n)$ and waiting until it hits a neighbor of $A(n)$. Add that point to $A(n)$ to create $A(n+1)$ and so on.

Although there are many predictions and conjectures, very few facts are rigorously known to be true. Here are some suggestions for interesting problems that you can try to learn about by simulation analysis. You may be able to come up with others. (Don't forget library and web research for further gaining inspiration and motivation.)

♠ `Open Problem 1.` One of the big open problems in this area is to decide whether or not $A(n)$ grows by growing long spindly arms. (The conjecture is that it does; this should make physical sense to you.) Can you decide if this is so? To what extent does the "shape" of $A(n)$ depend on the geometry of the lattice on which the random walks are being run?

♠ `Open Problem 2.` Since the notion of "shape" is not usually easy to grasp, one can ask simpler questions that are still quite interesting. For instance, how long are the arms of the DLA? (This is the title of a 1987 paper of Harry Kesten by the way.) In 1987, H. Kesten proved that for DLA on $\mathbf{Z}^d$, if $r_n := \max\{|x| : x \in A(n)\}$, then with probability one, r_n grows more slowly than $n^{2/3}$ if $d = 2$ and more slowly than $n^{2/d}$ if $d = 3$. While these results have been improved by the subsequent works of H. Kesten as well as those of G. Lawler, the known facts are very far from what is expected to be the true growth rate of $A(n)$. Can you decide what this rate is? Let me be more concrete. Suppose r_n grows like a constant times n^β for some exponent β. Can you find a simulation prediction for β?

§3. INTERNAL DIFFUSION-LIMITE AGGREGATION (IDLA)

In 1991, Diaconis and Fulton formulated a means by which subsets of certain commutative rings could be multiplied together. This uses a random process that is (like) a random walk on that commutative ring. When the said ring is $\mathbf{Z}^d$, their "random walk" becomes the following random process known as the internal diffusion-limit aggregation (IDLA for short):

Let $A(0) = \{0\}$; having defined $A(0), \ldots, A(n)$, we now construct $A(n+1)$ by running a random walk, independently of all else, until the random walk hits a point that is not in $A(n)$. When that happens, stop the walk and add the newly-visited point to $A(n)$ thereby creating $A(n+1)$.

This is a simpler process than the DLA, but it is far from being a simple object. Here is a fact that was shown to be true by M. Bramson, D. Griffeath, and G. Lawler (1992):

(3.1) Asymptotic Shape of the IDLA. *Let B_d denote the ball of radius 1 in $\mathbf{Z}^d$, and let ω_d denote its volume (e.g., $\omega_1 = 2$ and $\omega_2 = \pi$.) Then, as $n \to \infty$, the following happens with probability one:*

$$\left(\frac{\omega_d}{n}\right)^{\frac{1}{d}} A(n) \Rightarrow B_d, \tag{3.2}$$

where by $\Rightarrow$ I mean that for any $\varepsilon > 0$, the left-hand side is eventually contained in the ε-enlargement of B_d (i.e., ball of radius $(1+\varepsilon)$ for any ε), and eventually contains the ε-reduction of B_d (i.e., the ball of radius $(1-\varepsilon)$).

In other words, for large values of n (i.e., in large time), the IDLA set $A(n)$ looks more and more like the centered ball of radius $(n/\omega_d)^{1/d}$. For instance, when $d = 2$, this is the centered ball of radius $\sqrt{n/\pi}$.

- ♠ `Open Problem 1.` What happens in other lattices? For instance, what about the hexagonal or the triangular lattice? What if the lattice is inhomogeneous? (This is due to Matthew Taylor.)
- ♠ `Open Problem 2.` Continuing with the above, what if you have a lattice that is random? For instance, suppose you run a random walk on the infinite cluster of an independent percolation process (see §5 below). Then what behavior should you expect to see?
- ♠ `Open Problem 3.` One may think that $A(n)$ really looks filled in and like a ball. However, in her Ph. D. thesis, D. Eberz has proven that with probability one, there exist infinitely many n's such that $A(n)$ "has holes" in it. A good research problem would be to explore the fluctuations; i.e., to explore how different $A(n)$ is from the ball. As a concrete way to state this, consider the number of points that are (i) in $A(n)$ and not in $(n/\omega_d)^{1/d}B_d$; or are (ii) in $(n/\omega_d)^{1/d}B_d$ but not in $A(n)$. How many of them are there for large values of n? To be even more concrete, hypothesize that this number grows like a constant times n^γ. Can you estimate γ by simulation analysis?

§4. BOND PERCOLATION

For any number $0 < p < 1$, and for any lattice $\mathbf{L}$, we can define bond percolation on $\mathbf{L}$ as follows: Each edge of $\mathbf{L}$ is open with probability p and closed with probability $(1-p)$, and all edges are open/closed independently from one another. We can then say that *percolation* occurs if with positive probability, one can find some random open path that connects a given point of $\mathbf{L}$ (call it the origin) to infinity (i.e., if there is an infinite self-avoiding path emenating from the origin, all of whose edges are open.)

Let $\theta(p)$ denote the probability of percolation on a given lattice. That is, $\theta(p)$ is the probability that there is an infinite open connected path starting from the origin.

(4.1) The Critical Probability. *There exists a critical probability p_c such that whenever $p > p_c$, $\theta(p) > 0$, but when $p < p_c$, $\theta(p) = 0$.*

This follows from showing that $\theta(p)$ increases as p goes up; although it is true, this is not a trivial fact. Here is how you prove it:

Proof: On each edge e in the lattice, set down independent edge-weights X_e such that $P\{X_e \leq x\} = x$ for all $x \in [0,1]$. In other words, X_e is uniformly distributed on $[0,1]$. Now every time $X_e \leq p$, call that edge open, otherwise it is closed. This procedure produces the percolation process with parameter p simultaneously for all p, since $P\{e \text{ is open }\} = P\{X_e \leq p\} = p$. Moreover, if $X_e \leq p$, then for any $p' > p$, $X_e \leq p'$ also. Therefore, the percolation cluster for p is contained in the percolation cluster for p'. In particular, if there is percolation at level p, there is certainly percolation at level p'. This is another way to state that $\theta(p) \leq \theta(p')$. To finish, define p_c to be the smallest value of p such that $\theta(p) > 0$. This is well-define since θ is increasing (draw a picture!) ♣

LECTURE 5: THE CRITICAL PERCOLATION PROBABILITY FOR BOND PERCOLATION

Recall that, in percolation, each edge in $\mathbf{Z}^d$ is open or closed with probability p or $(1-p)$, and the status of all edges are independent from one another. In (4.1, Lecture 4) we showed that there exists a *critical probability* p_c (sometimes written as $p_c(\mathbf{Z}^d)$ to emphasize the lattice in question), such that for all $p > p_c$, there is percolation (i.e., with positive probability, there exists an infinite connected open path from the origin), and for $p < p_c$, there is no percolation. However, this statement is completely vacuous if the numercial value of p_c were trivial in the sense that p_c were 0 or 1. In this lecture, we will show that this is not the case. In fact, we will show that in all dimensions $d \geq 2$,

$$\textbf{(0.1)} \qquad \frac{1}{C(d)} \leq p_c(\mathbf{Z}^d) \leq 1 - \frac{1}{C(d)},$$

where $C(d)$ is the connectivity constant of $\mathbf{Z}^d$; see (§4.2, lecture 3).

(0.2) Concrete Bounds on $p_c(\mathbf{Z}^d)$. Since that $d \leq C(d) \leq (2d)$ (§4.2, lecture 3), then it follows from (0.1) above that $\frac{1}{2d} \leq p_c(\mathbf{Z}^d) \leq 1 - \frac{1}{2d}$. This can be easily improved upon, since by §4.9 of lecture 4, $C(d) \leq (2d-1)$, so that $\frac{1}{2d-1} \leq p_c(\mathbf{Z}^d) \leq 1 - \frac{1}{2d-1}$. in particular, $p_c(\mathbf{Z}^d)$ is strictly between 0 and 1, which is the desired claim. ♣

(0.3) The Planar Case. The planar case deserves special mention: The previous bounds show that $p_c(\mathbf{Z}^2)$ is between $\frac{1}{3}$ and $\frac{2}{3}$. In fact, it has been shown that

- **a.** $p_c(\mathbf{Z}^2) = \frac{1}{2}$ (Harris and Kesten);
- **b.** If $p = p_c(\mathbf{Z}^2)$, then there is no percolation (Bezuidenhout and Grimmett). ♣

§1. THE LOWER BOUND IN (0.1).

We first verify the lower bound of (0.1) on p_c. Note that showing $p_c \geq \frac{1}{C(d)}$ amounts to showing that whenever $p < \frac{1}{C(d)}$, then $P\{\text{percolation}\} = 0$.

First note that the chance that any self-avoiding path π of length n is open is p^n. Therefore,

$$\textbf{(1.1)} \qquad \begin{aligned} E\{\#\text{ of self-avoiding paths of length } n\} &= E\left[\sum_\pi \mathbf{1}\{\pi \text{ is open}\}\right] \\ &= \sum_\pi P\{\pi \text{ is open}\} = \sum_\pi p^n, \end{aligned}$$

where $\sum_\pi$ denotes the summation over all self-avoiding paths of length n, and $\mathbf{1}\{\cdots\} := \mathbf{1}_{\{\cdots\}}$ is the indicator of $\{\cdots\}$. Since there are χ_n many self-avoiding paths of length n,

$$\textbf{(1.2)} \qquad E\{\#\text{ of self-avoiding paths of length } n\} \leq \chi_n p^n.$$

But $\chi_n \approx \{C(d)\}^n$, where

$$\text{(1.3)} \qquad a_n \approx b_n \qquad \text{mean} \qquad \lim_{n\to\infty} \frac{\log a_n}{\log b_n} = 1.$$

This means that as soon as $p < \frac{1}{C(d)}$, then

$$\text{(1.4)} \qquad E\{\#\text{ of self-avoiding paths of length } n\} \to 0, \qquad (n \to \infty).$$

(Why? Be sure that you understand this!) But for any n,

$$\text{(1.5)} \qquad \begin{aligned} P\{\text{percolation}\} &\le P\{\#\text{ of self-avoiding paths of length } n \ge 1\} \\ &\le E\{\#\text{ of self-avoiding paths of length } n\}, \end{aligned}$$

thanks to Markov's inequality (§4.1, lecture 2). Since $P\{\text{percolation}\}$ is independent of n, (1.3) shows that it must be zero as long as $p < \frac{1}{C(d)}$. This shows that $p_c \ge C(d)$, which is the desired result. ♣

§2. THE UPPER BOUND IN (0.1).

Now we want to prove the second inequality in (0.1). That is, we wish to show that if $p > 1 - \frac{1}{C(d)}$, then $P\{\text{percolation}\} > 0$. This is trickier to do, since we have to produce an open path or an algorithm for producing such a path, and this is a tall order. Instead, let us prove the (logically equivalent) converse to the bound that we are trying to prove. Namely, we show that if $P\{\text{percolation}\} = 0$, then $p \le 1 - \frac{1}{C(d)}$. For this, we need to briefly study a notion of duality for percolation, and one for graphs. From now on, we will only work with $\mathbf{Z}^2$; once you understand this case, you can extend the argument to get the upper bound in (0.1) for any $d \ge 2$.

(2.1) The Dual Lattice. Briefly speaking, the dual lattice $\widetilde{\mathbf{Z}^2}$ of $\mathbf{Z}^2$ is the lattice

$$\text{(2.2)} \qquad \widetilde{\mathbf{Z}^2} := \mathbf{Z}^2 + \left(\frac{1}{2}, \frac{1}{2}\right).$$

At this point, some of you may (and should) be asking yourselves, "*What does it mean to sum a set and a point?*" In general, $A + x$ is short-hand for the set $\{y + x;\, y \in A\}$. That is, $A + x$ is A shifted by x. Consequently, the dual lattice $\widetilde{\mathbf{Z}^2}$ is the lattice $\mathbf{Z}^2$ shifted by $(0.5, 0.5)$. Pictorially speaking, the dual lattice $\widetilde{\mathbf{Z}^2}$ looks just like $\mathbf{Z}^2$, except that its origin is the point $(0.5, 0.5)$ instead of $(0, 0)$; i.e., its origin has been shifted by $(0.5, 0.5)$. You should plot $\widetilde{\mathbf{Z}^2}$ to see what is going on here.

(2.3) Dual Percolation. Each edge $\mathbf{e}$ in $\mathbf{Z}^2$ intersects a unique edge in $\widetilde{\mathbf{Z}^2}$ halfway in the middle. We can call this latter edge the *dual edge* to $\mathbf{e}$. Whenever an edge in $\mathbf{Z}^2$ is open, its dual is *declared* close, and conversely, if an edge in $\mathbf{Z}^2$ is closed, we declare its dual edge in $\widetilde{\mathbf{Z}^2}$ open. Clearly, this process creates a percolation process on the dual lattice $\widetilde{\mathbf{Z}^2}$, but the edge-probabilities are now $(1-p)$ instead of p. Now if there is no percolation on $\mathbf{Z}^2$, this means that on $\widetilde{\mathbf{Z}^2}$, there must exist an open "circuit" surrounding the origin. For a picture of this, see

`http://www.math.utah.edu/~davar/REU-2002/notes/lec5.html`

The probability that any given circuit, surrounding the origin, of length n is dual-open is $(1-p)^n$. So,

$$E\left[\#\text{ of open circuits in }\widetilde{\mathbf{Z}^2}\text{ of length }n\right] \le C_n(1-p)^n, \tag{2.4}$$

where C_n denotes the number of circuits—in $\widetilde{\mathbf{Z}^2}$—of length n that surround the origin. Thus, we have shown that

$$P\left\{\text{no perocolation in }\mathbf{Z}^2\right\} \le C_n(1-p)^n. \tag{2.5}$$

We want to show that is p is large enough, the above goes to zero as $n \to \infty$. To do so, we need a bound for C_n.

(2.6) Bounding C_n. It is easier to count the number of circuits of length n in $\mathbf{Z}^2$ (not the dual) that surround the origin. This number is also C_n (why?). But for a path $\pi := \pi_0, \ldots, \pi_n$ to be a circuit of length n about $(0,0)$, it must be that any $(n-1)$ steps in π form a self-avoiding path, and that π must go through one of the points $(1,0), (1,\pm 1), (1,\pm 2), \ldots, (1, \pm\lfloor \frac{n}{2} \rfloor)$. (There are at most $(n+1)$ of these points.) Therefore, $C_n \le (n+1)\chi_{n-1}$ (why?) Recalling (1.3) above, and since $\chi_{n-1} \approx \{C(d)\}^{n-1}$, this and (2.5) show that whenever $p > 1 - \frac{1}{C(d)}$, then there can be no percolation, which is the desired result. ♣

LECTURE 6: STARTING SIMULATION

§1. THE ABC'S OF RANDOM NUMBER GENERATION

(1.1) Computing Background. I will start the lectures on simulation by first assuming that you have access to (i) a language (such as C or better still C_{++}); or (ii) an environment (such as `Matlab`.) If you do not know how to use any programming, you need to get a crash-course, and your T.A.'s (in particular, Sarah and Robert) will help you along if you seek their help. At this point, you should make sure that you (i) have a computer account; and (ii) know how to log in, check mail, and run a program that you know how to run.

(1.2) Generating a Uniformly Distributed Random variable. All of simulation starts with the question, "*How do I choose a random number uniformly between* 0 *and* 1*?*" This is an intricate question, and you will have a detailed lecture on this topic from Dr. Nelson Beebe later this week or the next. These days, any self-respecting programming language or environment has a routine for this task (typically something like `rand`, `rnd`, or some other variant therefrom). Today, we will use such random number generators to generate a few other random variables of interest; we will also apply these methods to simulate random walks.

(1.3) Generating a ± 1 Random Variable. Our first task is to generate a random variable that takes the values ± 1 with probability $\frac{1}{2}$ each. Obviously, we need to do this in order to simulate the one-dimensional simple walk.

The key observation here is that if U is uniformly distributed on $[0,1]$, then it follows that $P\{U \leq \frac{1}{2}\} = \frac{1}{2}$. So, if we defined

$$X := \begin{cases} +1, & \text{if } U \leq \frac{1}{2}, \\ -1, & \text{if } U > \frac{1}{2}, \end{cases} \tag{1.4}$$

then $P\{X = +1\} = P\{U \leq \frac{1}{2}\} = \frac{1}{2}$ and $P\{X = -1\} = P\{U \geq \frac{1}{2}\} = \frac{1}{2}$. That is, we have found a way to generate a random variable X that is ± 1 with probability $\frac{1}{2}$ each. This leads to the following.

(1.5) Algorithm for Generating ± 1-Random Variables

1. Generate U uniformly on $[0,1]$
2. If $U \leq \frac{1}{2}$, let $X := +1$, else let $X := -1$

(1.6) Exercises. Try the following:

- **(a)** *Write a program that generates* 100 *independent random variables, each of which is* ± 1 *with probability* $\frac{1}{2}$ *each.*
- **(b)** *Count how many of your generated variables are* ± 1*, and justify the statement that, "with high probability, about half of the generated variables should be* ± 1*."*
- **(c)** *Come up with another way to construct* ± 1 *random variables based on uniforms; a variant of (1.5) is acceptable.*

(1.7) The Inverse Transform Method. We now want to generate other kinds of "discrete random variables," and we will do so by elaborating on the method of (1.5). Here is the algorithm for generating a random variable X such that $P\{X = x_j\} = p_j$ $j = 0, 1, \ldots$ for any predescribed set of numbers $x_0, x_1, \ldots$, and probabilities $p_0, p_1, \ldots$. Of course, the latter means that $p_0, p_1, \ldots$ are numbers with values in between 0 and 1, such that $p_0 + p_1 + \cdots = 1$.

(1.8) Algorithm for Generating Discrete Random Variables.

1. Generate U uniformly on $[0,1]$
2. Define

$$X := \begin{cases} x_0, & \texttt{if } U < p_0\,\texttt{,} \\ x_1, & \texttt{if } p_0 \le U < p_0 + p_1\,\texttt{,} \\ x_2, & \texttt{if } p_0 + p_1 \le U < p_0 + p_1 + p_2\,\texttt{,} \\ \vdots & \vdots \end{cases}$$

(1.9) Exercise. Prove that the probability that the outcome of the above simulation is x_j is indeed p_j. By specifying $x_0, x_1, \ldots$ and $p_0, p_1, \ldots$ carefully, show that this "inverse transform method" generalizes Algorithm (1.5).

(1.10) Exercise. In this exercise, we perform numerical integration using what is sometimes called *Monte Carlo simulations.*

(a) (Generating random vectors) *Suppose that $U_1, \ldots, U_d$ are independent random variables, all uniformly distributed on $[0,1]$, and consider the random vector $\mathbf{U} = (U_1, \ldots, U_d)$. Prove that for any d-dimesional hypercube $A \subseteq [0,1]^d$, $P\{\mathbf{U} \in A\}$ = the volume of A. In other words, show that $\mathbf{U}$ is uniformly distributed on the d-dimensional hypercube $[0,1]^d$.*

(b) *Let $\mathbf{U}_1, \ldots, \mathbf{U}_n$ be n independent random vectors, all distributed uniformly on the d-dimensional hypercube $[0,1]^d$. Show that for any integrable function f with d variables, the following holds with probability one:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} f(\mathbf{U}_\ell) = \int_0^1 \cdots \int_0^1 f(x_1, \ldots, x_d)\, dx_1 \cdots dx_d. \tag{1.11}$$

(c) *Use this to find a numerical approximation to the following integrals:*

i. $\int_0^1 e^{-x^2}\, dx$.

ii. $\int_0^1 \int_0^1 y^x\, dx\, dy$.

§2. SHORT-CUTS: GENERATING BINOMIALS

(2.1) The Binomial Distribution. A random variable is said to have the *binomial distribution* with parameters n and p if

$$P\{X = j\} = \binom{n}{j} p^j (1-p)^{n-j}, \qquad j = 0, 1, \ldots, n. \tag{2.2}$$

Here n is a positive integer, and p is a real number between 0 and 1.

(2.3) Example. For example, suppose n independent success/failure trials are performed; in each trial, $P\{\text{success}\} = p$. Then, if we let X denote the total number of successes, this is a random variable whose distribution is binomial with parameters n and p. ♣

(2.4) Example. Suppose $\xi_1, \ldots, \xi_n$ are independent with $P\{\xi = 1\} = p$ and $P\{\xi = 0\} = 1 - p$. Then, $X := \xi_1 + \cdots + \xi_n$ is binomial.

Proof: Let $\xi_i = 1$ if the ith trial succeeds and $\xi_i = 0$ otherwise. Then X is the total number of successes in n independent success/failure trials where in each trial, $P\{\text{success}\} = p$. ♣

(2.5) Example. If S_n denotes the simple walk on the integers, then $S_n = X_1 + \cdots + X_n$, where the X's are independent and every one of the, equals ± 1 with probability $\frac{1}{2}$ each. On the other hand, $Y_i := \frac{1}{2}(X_i + 1)$ is also an independent sequence and equals ± 1 with probability $\frac{1}{2}$ each (why?) Since $X_i = 2Y_i - 1$,

$$S_n = 2 \sum_{i=1}^{n} Y_i - n. \tag{2.6}$$

Therefore, the distribution of the simple walk at a fixed time n is the same as that of $2 \times \text{binomial}(n, p) - n$.

(2.7) A Short-Cut. Suppose we were to generate a binomial(n, p) random variable. A natural way to do this is the inverse transform method of (1.7) and (1.8). Here, $x_0 = 0, x_1 = 1, \ldots, x_n = n$, and p_j is the expression in (2.2). The key here is the following short cut formula that allows us to find p_{j+1} from p_j without too much difficulty:

$$\begin{aligned}
p_{j+1} &= \binom{n}{j+1} p^{j+1} (1-p)^{n-j-1} \\
&= \frac{p}{p-1} \times \frac{n!}{(j+1)! \times (n-j-1)!} \times p^j (1-p)^{n-j} \\
&= \frac{p}{p-1} \times \frac{n-j}{j+1} \times \binom{n}{j} p^j (1-p)^{n-j} \\
&= \frac{p}{p-1} \times \frac{n-j}{j+1} \times p_j.
\end{aligned} \tag{2.8}$$

So we can use this to get an algorithm for quickly generating binomials.

(2.9) Algorithm for Generating Binomials.

1. Generate U uniformly on $[0,1]$.
2. Let $\texttt{Prob} := (1-p)^n$ and $\texttt{Sum} := \texttt{Prob}$.
3. For $j = 0, \ldots, n$, do:
   i. If $U < \texttt{Sum}$, then let $X = j$ and stop.
   ii. Else, define

$$\texttt{Prob} := \frac{\texttt{Prob}}{1-\texttt{Prob}} \times \frac{n-j}{j+1} \times \texttt{Prob}, \qquad \text{and} \qquad \texttt{Sum} := \texttt{Prob} + \texttt{Sum}.$$

You should check that this really generates a binomial. ♣

(2.10) Algorithm for Generating the One-Dimensional Simple Walk. Check that the following generates and plots a $1-d$ simple walk.

1. (Initialization) Set $\texttt{W} := 0$ and plot $(0,0)$.
2. For $j = 0, \ldots, n$, do:
   i. Generate $X = \pm 1$ with prob. $\frac{1}{2}$ each.
   (See (1.5) for this subroutine.)
   ii. Let $\texttt{W} := \texttt{W} + X$ and plot $(j, \texttt{W})$.

If you are using a nice plotting routine like the one in `Matlab`, try filling in between the points to see the path of the walk.

(2.11) Exercise. Generate 2-dimensional simple walks that run for (a) $n = 100$ time units; (b) $n = 1000$ time units.

LECTURE 7: FRACTAL PERCOLATION

§1. FRACTAL PERCOLATION

(1.1) Mandelbrot's Fractal Percolation. Consider the square $S := [0,1] \times [0,1]$. That is, S is the set of all points (x, y) such that $0 \leq x \leq 1$ and $0 \leq y \leq 1$. We will divide S into four equal-sized squares,

$$\begin{aligned} S_1 &:= \left[0, \frac{1}{2}\right] \times \left[0, \frac{1}{2}\right], \; S_2 := \left[0, \frac{1}{2}\right] \times \left[\frac{1}{2}, 1\right], \\ S_3 &:= \left[\frac{1}{2}, 1\right] \times \left[0, \frac{1}{2}\right], \; S_4 := \left[\frac{1}{2}, 1\right] \times \left[\frac{1}{2}, 1\right], \end{aligned} \tag{1.2}$$

For each square, you toss an independent coin; with probability $p \in (0,1)$, you keep that square, and with probability $(1-p)$ you jettison it. So now you have a random number of kept squares (some random number between 0 and 4.) Split each into four equal-sized squares, and toss an independent p-coin for each to see if you want to keep, and repeat. Fact: If p is sufficiently large, and if you continue *ad infinitum*, then with positive probability you end up with a nonempty random set that Mandelbrot calls a "random curdle," and these days is referred to as fractal percolation.

(1.3) Hard Question. Use simulation to find the critical probability p_c past which you can get fractal percolation. ♣

§2. FRACTALS AND MINKOWSKI (BOX) DIMENSION

(2.1) The Tertiary Cantor Set. Georg Cantor invented the following strange set that is nowhere dense, has length zero, and yet is uncountable. It is the archetype of what is nowadays is called a fractal.

Start with the interval $I = [0,1]$; split it into three equal parts, and jettison the middle-third to get two intervals $I_1 := [0, \frac{1}{3}]$, and $I_2 := [\frac{2}{3}, 1]$. Take the remaining two intervals, split them in threes, and jettison the middle-third interval, and repeat. After the nth stage of this construction, you will get a set C_n that is made up of 2^n intervals of length 3^{-n}. In particular, the length of C_n is $(2/3)^n$, which goes to zero. It is not hard to see that $C := \cap_n C_n \neq \emptyset$, although it has length zero. A little more work shows that it is nowhere dense.

(2.2) The Minkowski Dimension. Note that in the nth stage of the construction of the tertiary Cantor set of (2.1), we have in pricniple 3^n intervals of length 3^{-n}, but we only keep 2^n of them. Therefore, the total number of intervals of length 3^{-n} that cover the tertiary Cantor set should be 2^n. In general, let N_k denote the total number of the intervals (in higher dimensions, cubes) of length k^{-1} that cover the portion of your fractal in $[0,1]$, and define the *Minkowski* or *box dimension* of your fractal to be the number α such that $N_k \approx k^\alpha$, if such a number exists. (Recall that $a_k \approx b_k$ means that as $k \to \infty$, $\log(a_k) \div \log(b_k) \to 1$.)

(2.3) Example. Consider the tertiary Cantor set of (2.1), and check that $N_{3^{-n}} = 2^n$. Formally let $k = 3^n$ and convince yourself that as $k \to \infty$, $N_k \approx k^\alpha$ where $\alpha = \log(2)/\log(3)$. That is, the tertiary Cantor set is a "**fract**al" of "**fract**ional dimension" $\log(2)/\log(3)$ which is about equal to 0.63. ♣

(2.4) Projects for Extensions. You can try constructing other Cantor-type fractals by either (i) splitting into intervals of other sizes than $\frac{1}{3}$; (ii) retaining/jettisoning intervals by a different algorithm; or (iii) constructing higher-dimensional fractals. For instance, try starting with the square $[0,1] \times [0,1]$; split it into 9 equal-sized squares; retain all but the middle one, and repeat. ♣

(2.5) Projects for Fractal Percolation. Now go back to fractal percolation, and ask:

- ♢ What is the critical probability p_c, such that whenever $p > p_c$, you can end up with a nonempty random fractal, and when $p < p_c$, the entire construction ends at some random stage since everything has been jettisoned? The answer to this is known by theoretical considerations.
- ♢ When $p > p_c$, can you find the box dimension of the resulting random fractal? The answer to this is known by theoretical considerations.
- ♢ When $p > p_c$, can you estimate the probability that there exists a left-to-right path on the resulting random fractal? The answer to this is unknown.

(2.6) Relation to Percolation on Trees. The act of splitting each square into four equal-sized ones can be represented by a rooted tree in which each vertex splits into four vertices in the next level of the tree. Now go through the edges of this tree, and with probability p keep an edge, and with probability $(1-p)$ discard it. Question: Is there an infinite kept path starting from the root? You should make sure that you understand the following assertion: This is *exactly* the same mathematical question as, "Is there fractal percolation?"

(2.7) Relation to Branching Processes. Consider the following model for geneology of a gene: You start with one "grandmother gene." Upon death (or mutation or whatever else is the case), this gene splits into a random number of "offspring," where the offspring distribution is: With probability p^4 there are 4 offpsring; with probability $\binom{4}{1}p^3(1-p)$ there are 3 offspring; with probability $\binom{4}{2}p^2(1-p)^2$ there are 2 offspring; and with probability $(1-p)^4$ there are no offpsring. How large should p be in order for this gene population to survive forever? Make sure that you understand that this is the same mathematical problem as the one in (2.6), which is itself the same as asking whether or not one has fractal percolation.

LECTURE 8: BROWNIAN MOTION

§1. A LITTLE HISTORY

(1.1) Robert Brown (1828). In 1828, an English botanist by the name of Robert Brown discovered that if you grains of pollen suspended in water, then each individual grain seems to undergo a rather erratic movement. He also posed the problem of describing this movement that has come to be known as "molecular motion" or "diffusion." This work was largely ignored by the scientific community for some time.

(1.2) Louis Bachelier (1900). Independently from Brown's work, in his 1900 Ph.D. thesis at the University of Paris, and under the guidance of the great French mathematician H. Poincaré, Louis Bachelier worked out a theory for the fluctuations of the stock market that involved the development of a stochastic process that is now called "Brownian motion." [See L. Bachelier (1900). Théorie de la spéculation, *Annales de l'Ecole Normale Superiure*, *Ser.* 3, **17**, 21–86. See also the English translation: L. Bachelier (1964). *The Random Character of Stock Market Prices*, P. H. Cootner editor, MIT Press, Cambridge.] Unfortunately, Bachelier's work went largely unnoticed for nearly a century, since his arguments contained flaws nearly all of which are now known to be minor. However, amongst the accomplishments of Bachelier's thesis were his discovery of two deep and fundamental facts about the Brownian motion: One, that it has a Markovian character (in words, given the position at time t, you do not need the prior positions to predict or simulate the future behavior of the process); and two, that it has the reflection property: If $W(s)$ denotes the position of the Brownian motion at time s, then the maximal displacement by time t (i.e., $\max_{s\le t} W(s)$) has the same distribution as the absolute displacement at time t (i.e., $|W(t)|$). The latter has a simple distribution and this leads to Bachelier's wonderful calculation:

$$P\left\{\max_{0\le s\le t} W(s) \le \lambda\right\} = \sqrt{\frac{2}{\pi t}} \int_0^\lambda e^{-x^2/2t}\, dx. \tag{1.3}$$

(1.4) Albert Einstein (1905). In 1905 Albert Einstein came to the problem of Brownian motion independently (and unaware of) of Bachelier's work; his motivation was to answer Brown's question by proposing a mathematical model for molecular motion. [See A. Einstein (1956). Investigations on the theory of the Brownian movement, New York.] In particular, he used the connections between the Brownian motion and the diffusion equation to get the sharpest estimates of that time for the Avagadro's number and hence the diameter of a hydrogen atom. With hindsight, we now know that Bachelier went much further in his analysis than Einstein. However, it is easier to describe Einstein's prediction for what the Brownian motion should be. Actually proving that such an object exists and developing a calculus for it required a tremendous mathematical development to which I will come shortly. However, let me mention in passing that a good number of physicists continued Einstein's analysis and applications of Brownian motion in physics; some of the names that you should know about are Smoluchowski, Fokker and Planck, Uhlenbeck, and many others.

(1.5) Einstein's Predicates. Einstein predicted that the one-dimensional Brownian motion is a random function of time written as $W(t)$ for "time" $t \geq 0$, such that:

(a) At time 0, the random movement starts at the origin; i.e., $W(0) = 0$.

(b) At any given time $t > 0$, the position $W(t)$ of the particle has the normal distribution with mean 0 and variance t.

(c) If $t > s > 0$, then the displacement from time s to time t is independent of the past until time s; i.e., $W(t) - W(s)$ is independent of all the values $W(r)$; $r \leq s$.

(d) The displacement is time-homogeneous; i.e., the distribution of $W(t) - W(s)$ is the same as the distribution of $W(t-s)$ which is in turn normal with mean 0 and variance $t - s$.

(e) The random function W is continuous.

(1.6) Norbert Wiener (1923). In 1923, Norbert Wiener (a professor at MIT and a child prodigy) proved the existence of Brownian motion and set down a firm mathematical foundation for its further development and analysis. Wiener used the recently-developed mathematics of É. Borel and H. Steinhaus (the subject is called measure theory), and cleverly combined it with a nice idea from a different mathematical discpline (harmonic analysis) to show the in fact the following random series converges with probability one to an object that satisfies (nearly) all of Einstein's predicates: For all $0 \leq t \leq 1$,

$$W(t) = \frac{1}{\sqrt{\frac{\pi}{2}}} t X_0 + \frac{1}{\sqrt{\pi}} \sum_{j=1}^{\infty} \left[\frac{\sin(\pi j t)}{j} X_j, \frac{\cos(\pi j t)}{j} X_{-j} \right], \tag{1.7}$$

where $X_0, X_{\pm 1}, X_{\pm 2}, \ldots$ are independent standard normal random variables. [See the last two chapters of R. E. A. C. Paley and N. Wiener (1934). *Fourier Transforms in the Complex Plane*, New York.]

(1.8) Paul Lévy (1939). Finally, the classical development of Brownian motion was complete in a 1939 work of Paul Lévy who proved the following remarkable fact: If you replace the normal distribution by any other distribution in Einstein's predicate (cf. 1.5), then either there is no stochastic process that satisfies the properties (a)–(d), or (e) fails to hold! Lévy's work was closely related to the concurrent and independent work of A. I. Khintchine in Russia, and is nowadays called *The Lévy–Khintchine Formula.*

(1.9) Kiyosi Itô (1942/1946). The work of Paul Lévy started the modern age of random processes, and at its center, the theory of Brownian motion. The modern literature on this is truly vast. But all probabilists would (or should) agree that a center-piece of the classical literature is the 1942/1946 work of K. Itô who derived a calculus—and thereby a theory of stochastic differential equations—that is completely different from the ordinary nonstochastic theory. This theory is nowadays at the very heart of the applications of probability theory to mathematical finance. [See K. Itô (1942). On stochastic processes. 1. *Japanese J. Math.*, **18**, 261–301; K. Itô (1946). On a stochastic integral equation, *Proc. Jap. Aca.*, **22**, 32–25.]

(1.10) Monroe Donsker (1951). For us, the final important step in the analysis of Brownian motion was the 1951 work of Donsker who was a Professor of mathematics at The

New York University. [See M. Donsker (1951). An invariance principle for certain probability limit theorems, *Memoires of the American Math. Society*, **6**, and M. Donsker (1952). Justififcation and extension of Doob's heuristic approach to the Kolmogorov–Smirnov theorem, *The Annals of Math. Stat.*, **23**, 277–281.] Amongst other things, Donsker verified a 1949 conjecture of the great American mathematician J. L. Doob by showing that once you run them for a long time, all mean-zero variance-one random walks look like Brownian motion! [The said conjecture appears in J. L. Doob (1949). Heuristic approach to the Kolmogorov–Smirnov statistic, *The Annals of Math. Stat.*, **20**, 393–403].

§2. BROWNIAN MOTION

(2.1) Donsker's Theorem. As I mentioned in (1.10), Donsker's theorem states that once you run them for a long time, all mean-zero variance-one random walks look like Brownian motion. Here is a slightly more careful description: *Let $X_1, X_2, \ldots$ denote independent, identically distributed random variables with mean zero and variance one. The random walk is then the random sequence $S_n := X_1 + \cdots + X_n$, and for all n large, the random graph of $S_1/\sqrt{n}, S_2/\sqrt{n}, \ldots S_n/\sqrt{n}$ (linearly interpolate inbetween the values as Matlab does automatically), is close to the graph of Brownian motion run until time one.*

(2.2) Algorithm for Running a Brownian Motion W. Choose a large value of n and a starting value x, and perform the following. It uses Donsker's theorem above, and wwill plot the path of a one-dimensional Brownian motion run until time 1.

```
– For i=1 to n;
    ◦ W(i) = x;
– end;        % Initialize the Brownian motion to
                have all values equal to the starting point x.
– Plot (0,W(1))        % This plots the starting point.
– For i=2 to n;   % When i=1, W(i)=x already.
    ◦ Generate a random variable Z := ±1 with probability 1/2.
    ◦ Set W(i) = Z/√n + W(i − 1);
    ◦ Plot (i/n,W(i));
– end;
```

(2.3) Application: Bachelier's Reflection Principle. Recall the reflection principle of D. Andre from (3.1, Lecture 2): If S_n is the simple walk, then

$$P\left\{\max_{1\le k\le n} S_n \ge \lambda\right\} = 2P\left\{S_n \ge \lambda\right\}. \tag{2.4}$$

But the same number of these simple-walk paths are over λ as they are under $-\lambda$. Thus,

$$P\left\{\max_{1\le k\le n} S_n \ge \lambda\right\} = P\left\{|S_n| \ge \lambda\right\}. \tag{2.5}$$

What this says is that the distribution of the maximum displacement of the walk is the same as the distribution of the absolute displacement. Replace λ by $\sqrt{n}\lambda$, let $n \to \infty$, and appeal to Donsker's theorem to deduce the following: *If W denotes Brownian motion, then for all $\lambda > 0$,*

$$P\left\{\max_{0\leq s\leq 1} W(s) \geq \lambda\right\} = P\left\{|W(s)| \geq \lambda\right\}. \tag{2.6}$$

But the distribution of $W(s)$ is a normal with mean zero and variance s. From this, one readily obtains Bachelier's reflection principle (cf. equation 1.3) with $t = 1$. The general case $t > 0$ is handled similarly. ♣

LECTURE 9: BROWNIAN MOTION AND DIFFUSION

§1. MONTE-CARLO SIMULATION

(1.1) The St.-Petersbourg Paradox. Let W denote Brownian motion, and for any number $\lambda > 0$ (say $\lambda = 1$ for the sake of concreteness) define

$$T_\lambda := \min\{s \geq 0 : \; W(s) = \lambda\}, \tag{1.2}$$

which is the first time Brownian motion attains the level λ. For this random variable, one has the property that $P\{T_\lambda < \infty\} = 1$, and yet $E\{T_\lambda\} = +\infty$. That is, although Brownian motion will eventually reach λ, it is never expected to! This is a variant of the St.-Petersbourg paradox of gambling. That $P\{T_\lambda < +\infty\} = 1$ is not hard to see on a simulation. But how does one verify that $E\{T_\lambda\} = +\infty$? One starts with a formula from measure theory:

$$E\{T_\lambda\} = \int_0^\infty P\{T_\lambda > x\}\, dx. \tag{1.3}$$

So if we could show that as $x \to \infty$, $P\{T_\lambda > x\} \sim Cx^{-1/2}$ (say), it would follow that for large n, $\int_n^\infty P\{T_\lambda > x\}\, dx \sim C \int_n^\infty x^{-1/2}\, dx = +\infty$. This is indeed the case:

(1.5) Theorem. *For each $\lambda > 0$, there exists some uninteresting constant C such that as $x \to \infty$, $P\{T_\lambda > x\} \sim Cx^{-1/2}$.*

(1.6) Simulation Verification. How does one verify this theorem by simulation methods? Note that $P\{T_\lambda > x\}$ is an expectation and hence can be simulated by Monte–Carlo simulation (cf. Lecture 6, Exercise 1.10). Indeed, $P\{T_\lambda > x\} = E\{\mathbf{1}_{\{T_\lambda > x\}}\}$. So, one can Monte-Carlo-simulate this by generating a large number (N) of independent Brownian motions $W^1, \ldots, W^N$, each until the first time they hit λ. Let $T_\lambda^1, \ldots, T_\lambda^N$ denote their respective hitting times to λ and note that $T_\lambda^1, \ldots, T_\lambda^N$ are independent and identically distributed. Thus, by Kolmogorov's strong law of large numbers (Theorem 0.1, Lecture 2),

$$\lim_{N\to\infty} \frac{1}{N} \sum_{\ell=1}^N \mathbf{1}_{T_\lambda^\ell > x\}} = P\{T_\lambda > x\}. \tag{1.7}$$

In words, generate N independent Brownian motions and see how many of them take up at least x units of time to reach λ. If N is large, then this should be close to $P\{T_\lambda x\}$. Now conjecture that for some $\alpha > 0$, $P\{T_\lambda > x\} \approx x^{-\alpha}$. If so, then $\log P\{T_\lambda > n\} \sim -\alpha \log n$ for large n, and this means that if you plot the log-plot of the function $P\{T_\lambda > x\}$, you will see a constant function; the constant is $-\alpha$ and, thanks to Theorem (1.5), it should be equal to $-\frac{1}{2}$. ♣

(1.8) Random Walk Projects. Suppose S denotes the two-dimensional simple walk; fix some nice set $A \subset \mathbf{R}^2$, and let $T_A := \min\{n \geq 0 : \; S_n \in A\}$ be the first time that you hit that set. For instance, A could be a single point $\{(1, 1)\}$ (say), a curve, etc. Can you find the extent to which the geometry of A affects the rate of decay of $P\{T_A > x\}$ as $x \to \infty$? For instance, can you detect a discernable difference between the two cases $A := \{(1, 1)\}$ and $A :=$ the square $\{(x, y) : |x| = 1,\ |y| = 1\}$? (There is a huge difference.)

(1.9) Brownian Motion Projects. Continuing with our discussion of (1.8), we may ask, "what if the set A changes with time?" The most interesting case is if A is replaced by $\sqrt{n}A$. To be conrete, consider the example of (1.6) but replace λ by $\sqrt{x}\lambda$; let us also write n for x to remind ourselves that it is an integer. So, to summarize: Let $S_n :=$ the simple walk on $\mathbf{Z}$, and more generally consider $P\{\mathcal{T}_{\sqrt{n}\lambda} > nx\}$, where $\mathcal{T}_m$ denotes the first time the random walk hits m for any m. Then, by Donsker's theorem (Theorem 2.1, Lecture 8),

$$\lim_{n\to\infty} P\{\mathcal{T}_{\sqrt{n}\lambda} > nx\} = P\{T_\lambda > x\}, \tag{1.10}$$

and recall that T_λ is the first time Brownian motion hits λ. If instead of $\sqrt{n}\lambda$ you write $n^\alpha\lambda$ for $\alpha \neq \frac{1}{2}$, then nothing interesting happens. Either the probabilities are too small, or they converge to positive constants.

§2. ITÔ DIFFUSIONS

(2.1) A Model in Discrete Time. Suppose that you want to model the random-walk-like movement of a particle in space, but now the space is inhomogeneous, so that in some parts, the walk moves rapidly, and in others very slowly. (Think of a random walk in space that is in part filled with air and in part with oil.) We will restrict "space" to one-dimensions since it is easier to imagine what is going on.

One way to proceed is to construct independent molecular fluctuations, $X_1, X_2, \ldots$. These are—as before—equal to ± 1 with probability $\frac{1}{2}$ each, and are independent random variables. Our "diffusion" (or random walk in inhomogeneous media) will be denoted by the process $Y_0.Y_1, \ldots$, where Y_0 is wherever the process starts (say at the origin.) So, $Y_0 := 0$, and having constructed $Y_0, \ldots, Y_k$, define $Y_{k+1} := Y_k + a(Y_k)X_{k+1}$, where the function a tells us how much to alter the usual fluctuations of the ordinary walk (based on X's), depending on where the diffusion Y is at time k. For instance, by sure that you understand that if the function $a(x) := 2$ for all x, then the diffusion Y is just a simple walk times 2; i.e., a simple walk that fluctuates twice as wildly. We can add a drift term to this diffusion as well to model the effect of a push. That is, $Y_{k+1} = a(Y_k)X_{k+1} + b(Y_k)$.

(2.2) Itô Diffusions in Continuous Time. Just as Brownian motion was obtained as limits of random walks, we can proceed to construct continuous-time diffusions by discrete-time approximations. Here is the simulation algorithm; it will construct an Itô diffusion in continuous time whose fluctuation are guided by some function a and whose drift is by some function b:

$$Y(0) := 0, \qquad Y\left(\frac{k+1}{n}\right) := Y\left(\frac{k}{n}\right) + a\left(Y\left(\frac{k}{n}\right)\right) \cdot \frac{X_{k+1}}{\sqrt{n}} + b\left(Y\left(\frac{k}{n}\right)\right) \cdot \frac{1}{n}. \tag{2.3}$$

The $1/\sqrt{n}$ term is just central limit theorem scaling as in Donsker's theorem for Brownian motion. Indeed, if $a(x) := 1$ and $b(x) := 0$, the process Y is Brownian motion. Another way to write this is to bring the term $Y(k/n)$ to the left-hand side to convince yourselves that Y "solves" the following "stochastic differential equation:"

$$dY(t) = a(Y(t))dW(t) + b(Y(t))dt, \tag{2.4}$$

where W is Brownian motion.

(2.5) Warning. The above stochastic differential equation has very different properties (as well as a different meaning) than ordinary differential equations of the calculus of real functions. For instance, Paley, Wiener, and Zygmund proved that with probability one, the Brownian motion W is nowhere differentiable, so that $dW(t)$ is not the usual "differential." [See R. E. A. C. Paley, N. Wiener, and A. Zygmund (1933). Notes on random functions, *Math. Zeit.*, **37**, 647–668.] The difference is best seen when trying to understand Itô's formula that is next.

(2.6) Itô's Formula. If you consider two differentiable functions f and g, then by the chain rule of the calculus of real functions,

$$(f(g))' = f'(g) \times g'. \tag{2.7}$$

If g is the random function W instead (i.e., Brownian motion), it is nowhere differentiable (cf. 2.6 above), and hence $W'(s)$ does not exist at any s. Itô's formula tells us what happens to chain rule in this case: For a twice continuously differentiable function f,

$$f(W(t)) = f(W(0)) + \int_0^t f'(W(s))\,dW(s) + \frac{1}{2}\int_0^t f''(W(s))\,ds, \tag{2.8}$$

where the "stochastic integral" $\int f'dW$ needs to be defined. It can be shown to satisfy the following natural approximation, but the choice of the so-called left-point rule is absolutely essential now:

$$\int_0^t g(W(s))\,dW(s) = \lim_{n\to\infty}\sum_{j=0}^{n} g\left(W\left(\frac{jt}{n}\right)\right) \times \left\{W\left(\frac{(j+1)t}{n}\right) - W\left(\frac{jt}{n}\right)\right\}, \tag{2.9}$$

where "limit" needs to be understood in some carefully stated sense. What is important about this approximation is that it shows quite clearly that the stochastic integral will have mean zero always! Indeed, note that $g(W(jt/n))$ and $\{W((j+1)t/n) - W(jt/n)$ are independent thanks to Einstein's predicate (1.5c, Lecture 8). Now elementary probability theory tells us that whenever ξ and ζ are independent random variables, then $E\{\xi\zeta\} = E\{\xi\}E\{\zeta\}$. Since $E\{W(t) - W(s)\} = 0$, this shows that stochastic integrals are always mean-zero processes; i.e.,

$$E\left\{\int_0^t g(W(s))\,dW(s)\right\} = 0. \tag{2.10}$$

(2.11) Itô's Formula for the Diffusion Y. The diffusion Y also has an Itô formula; it is the following more complicated one:

$$f(Y(t)) = f(Y(0)) + \int_0^t f'(Y(s))\, dY(s) + \frac{1}{2}\int_0^t f''(Y(s))\, [a(Y(s)]^2\ ds. \tag{2.12}$$

Plug the value of $dY(s)$ from (2.4) and we obtain the Itô formula,

$$\begin{aligned} f(Y(t)) =& f(Y(0)) + \int_0^t f'(Y(s))a(Y(s))\, dW(s) + \int_0^t f'(Y(s))b(Y(s))\, ds \\ &+ \frac{1}{2}\int_0^t f''(Y(s))\, [a(Y(s)]^2\ ds. \end{aligned} \tag{2.13}$$

The point is that this formulation has a stochastic integral in terms of dW which we have already seen is mean-zero.

(2.14) Existence of the Diffusion. Unfortunately, the simulation algorithm of (2.3) will produce something that yields nonsense unless the functions a and b are "nice." By this I mean that the Itô equation (2.4) will have solutions only if a and b are nice. One such condition is that a' and b' exist and are bounded functions. Under this condition, with probability one, (2.4) can be shown to have a unique solution process Y.

LECTURE 10: ITÔ's FORMULA AND THE WRIGHT–FISCHER MODEL

§1. ITÔ's FORMULA

(1.1) An Itô Formula. Suppose Y solves the stochastic differential equation,

$$dY(t) = a(Y(t))dW(t) + b(Y(t))dt, \tag{1.2}$$

and recall from (2.11) of Lecture 10 that for any nice function f,

$$f(Y(t)) = f(Y(0)) + \int_0^t f'(Y(s))a(Y(s))\,dW(s) + \frac{1}{2}\int_0^t f''(Y(s))\left[a(Y(s))\right]^2 ds. \tag{1.3}$$

From this, and a few lines, one can show the following.

(1.4) Probabilistic Interpretation of a and b. As $h \downarrow 0$,

$$E\left\{\frac{Y(t+h)-Y(t)}{h}\,\middle|\, Y(t)=x\right\} \to b(Y(t))$$
$$E\left\{\frac{\left[Y(t+h)-Y(t)\right]^2}{h}\,\middle|\, Y(t)=x\right\} \to a(Y(t)).$$

This gives further credance to our intuition that $a(x)$ determines the strength of the fluctuation if Y enters the value x, and $b(x)$ determines the drift (or push) if Y enters $b(x)$.

§2. THE WRIGHT–FISCHER GENE FREQUENCY MODEL

(2.1) A Haploid Model. The haploid model is the simplest model for asexual gene reproduction; here, there are no genetic effects due to genetic mutation or selection for a specific gene.

Let $2N$ denote a fixed population size comprised of two types of individuals (more aptly, genes): Type A and Type B. If the parent consists of i type-A individuals (and hence $2N-i$ type-B), then in the next generation, each gene becomes type-A with probability $\frac{i}{2N}$ and type-B with the remaining probability $1-\frac{i}{2N}$. All genes follow this prescription independently, and this works to construct a random process that evolves from generation to generation.

Let $X_n :=$ the number of type-A individuals in generation n. Then, given that we have simulated the process until time $(n-1)$ and observed $X_{n-1}=j$, we have:

$$P\{X_n = j \mid X_{n-1}=i\} = \binom{2N}{j}\left(\frac{i}{2N}\right)^j\left(1-\frac{i}{2N}\right)^{2N-j}, \qquad \forall j=0,\ldots,2N. \tag{2.2}$$

A question arises that is the genetics' analogue of the maze-problem from Robert Thorn's talk:

(2.3) Question. What is the probability that starting with i type-A individuals for some $i = 0, \ldots, 2N$, X_n is eventually equal to 0? Can you answer this by simulation when N is large? ♣

(2.4) A Diffusion-Approximation. Consider the entire random process $\frac{X_k}{2N}$ where $k = 1, \ldots, 2N$, and N is fixed but large. Then, one can show that when N is large, this process looks like the solution to the following stochastic differential equation (called Feller's equation) run until time one:

$$d(Y(t)) = Y(t)\big\{1 - Y(t)\big\}dW(t). \tag{2.5}$$

Thinking of this SDE as we did in (2.3, Lecture 10), you should convince yourself that when the solution Y hits 0 or 1, it sticks there forever.

(2.6) An Argument to Convince you of (2.5). This is not a rigorous argument, but its intuitively convincing: Based on the conditional-binomial formula (2.2) above, and a few calculations involving the means and variances of binomials, we have the following: As $h \to 0$, and for each $0 \le t \le 1$,

$$\begin{aligned} E\left\{ \frac{X_{2N(t+\frac{1}{N})} - X_{2Nt}}{2N} \,\middle|\, X_{2Nt} = i \right\} &= 0 \to 0 \\ E\left\{ \frac{\left[X_{2N(t+\frac{1}{N})} - X_{2Nt}\right]^2}{2N} \,\middle|\, X_{2Nt} = i \right\} &= \frac{1}{2N}\left(\frac{i}{2N}\right)\left(1 - \frac{i}{N}\right). \end{aligned} \tag{2.7}$$

So let $h = \frac{1}{2N}$ and consider the process $Y_N(t) := \frac{1}{2N} X_{\lfloor 2Nt \rfloor}$ to "see" that Y_N should look like Y in light of (1.4). ♣

(2.9) Simulation Project. Simulate the Wright–Fischer haploid model, as well as Feller's diffusion, and "compare." You should think hard about what this means, since we are talking about different random processes. ♣

LECTURE 11: PROBABILISTIC SOLUTION TO ELLIPTIC PDE'S

§1. ANOTHER ITÔ's FORMULA

We now explore some of the many connections between Brownian motion and second-order partial differential equations (PDE's). To start, we need a variant of Itô's formula. This one is an Itô-type development for a function $f(x,t)$ of space-time (x,t); the "space variable" is $x \in \mathbf{R}^d$, and the "time variable" is $t \geq 0$.

Throughout, W denotes d-dimensional Brownian motion.

(1.1) Another Itô's Formula. For any $T \geq t \geq 0$,

$$\begin{aligned} f(W(t), T-t) &= f(W(0), t) + \sum_{j=1}^{d} \int_0^T \frac{\partial}{\partial x_j} f(W(s), T-s)\, dW_j(s) \\ &\quad + \sum_{j=1}^{d} \int_0^T \frac{1}{2}\Delta f(W(s), T-s)\, ds + \int_0^T \frac{\partial}{\partial t} f(W(s), T-s)\, ds, \end{aligned} \tag{1.2}$$

where $\Delta f(x,t) := \sum_{j=1}^d \frac{\partial^2}{\partial x_j^2} f(x,t)$ is the *Laplacian* of f in the "space variable" $x \in \mathbf{R}^d$.

§2. THE HEAT EQUATION

The heat equation is the equation that governs the flow of heat in a nice medium. If $u(x,t)$ denotes the amount of heat at place $x \in \mathbf{R}^d$ at time t, then it states that u is "the continuous solution" to the following:

$$\begin{aligned} \frac{\partial}{\partial t} u(x,t) &= \frac{1}{2}\Delta u(x,t), \quad t \geq 0, x \in \mathbf{R}^d, \\ u(x,0) &= f(x), \qquad x \in \mathbf{R}^d, \end{aligned} \tag{2.1}$$

where f is the function that tells us the initial amount of heat introduced at each point $x \in \mathbf{R}^d$ in space, and u tells us how this heat propagates (i.e., cooling). The number $\frac{1}{2}$ is chosen for the sake of convenience and can be replaced by any other number c; in general, this is the so-called *thermal conductivity* of the medium that is being heated, and can be obtained by a change of variables of type $v(x,t) := u(\sqrt{c}x, t)$. Indeed, note that $\frac{\partial}{\partial t} v(x,t) = \frac{\partial}{\partial t} u(ax,t)$ and $\frac{\partial^2}{\partial x_j^2} v(x,t) = c\frac{\partial^2}{\partial x_j^2} u(ax,t)$. So that v solves

$$\begin{aligned} \frac{\partial}{\partial t} v(x,t) &= c\Delta v(x,t), \quad t \geq 0, x \in \mathbf{R}^d, \\ v(x,0) &= f(x/\sqrt{c}), \qquad x \in \mathbf{R}^d. \end{aligned} \tag{2.2}$$

So we might as well study (2.1) when the thermal conductivity is $\frac{1}{2}$.

(2.3) The Probabilistic Solution. *The solution to (2.2) can be written as follows, where W denotes d-dimensional Brownian motion: $u(x,T) = E_x\{f(W(T))\}$, where E_x denotes the expectation relative to Brownian motion started at $x \in \mathbf{R}^d$.*

(2.4) Itô's Formula Once More. We can deduce (2.3) from (1.2) with $T := t$ as follows:

$$\begin{aligned} u(W(T),0) &= u(W(0),T) + \int_0^T -\frac{\partial}{\partial t} u(T-s, W(s))\,ds \\ &\quad + \int_0^T \frac{1}{2}\Delta u(T-s, W(s))\,ds + \text{stoch. integral.} \end{aligned} \tag{2.5}$$

All that we care about is that the expected value of the stochastic integral is zero; cf. the simulation approximation (2.9, Lecture 9) to convince yourselves of this. Moreover, the other two integrals are equal to $\int_0^T \left(\frac{1}{2}\Delta u - \frac{\partial}{\partial t} u\right) = 0$, since u solves the heat equation (2.1). So, we can take the expectation of (2.5) conditional on $W(0) = x$ (i.e., start your Brownian motion at $x \in \mathbf{R}^d$) to get $E_x\{u(W(T),0)\} = u(x,T)$. Since $u(y,0) = f(y)$ for all y, this proves (2.4). ♣

(2.6) Project. How would you simulate $E_x\{f(W(T))\}$? (Hint: Kolmogorov's strong law of large number (0.1, Lecture 2).)

(2.7) THE DIRICHLET PROBLEM. If you put a unit of charge in the middle of a sphere, it charges the outer shell of the sphere and the charge distribution is uniform. More generally, if D is a nice domain in $\mathbf{R}^d$ (the analogue of the sphere), and if f is the charge distribution on the boundary (or shell) ∂D of D, then we have a charge distribution $u(x)$ at x that is given by the Dirichlet problem:

$$\begin{aligned} \Delta u(x) = 0 \quad & x \in D \quad \text{i.e., no-flux inside} \\ u = f, \quad & \text{on } \partial D. \end{aligned} \tag{2.8}$$

The probabilistic solution, using Brownian motion, is $u(x) := E_x\{f(W(\tau_D))\}$, where W denotes Brownian motion started at x and in d dimensions, and τ_D is the first time W leaves D. How would you simulate this?