Order from Disorder: Probability, Random Walks and Percolation (and ...?)

Jeffrey Schenker

Contents

Lecture 1: Introduction

This course will be an examination of several random constructions that have come up in physics as simplified models to understand phase transitions. (Thinks like melting, boiling, how a magnet becomes magnetized, etc.)

You don't need to know any physics, however, to understand what we will talk about.

Today's lecture is an overview of the three or four topics we will cover in the course:

(1) Random Walks
(2) Self Avoiding Walks
(3) Percolation
(4) Ising model

Random Walks. Imagine standing at a corner in a giant city with streets laid out in a grid. Flip a coin twice to decide if you will travel north, south, east, or west. (Or, flip a four sided die.) Once you get to the next corner. Repeat.

After a long time, your path will look something like the picture that appears on the right hand side of the home page of this website. A moment's thought will show that you can return to where you started. In fact the chances are at least one out of four, since this is the chance that you return after your first step. So we can ask, what are the chances that you ever return? It turns out the answer is one out of one! Every time you start out you will return. How many times will you return? Infinitely many!! (If you are in an infinitely large city and keep walk for infinite time, of course.) These answers are far from obvious, and we would like to understand them.

Furthermore these answers depend crucially on walking on a two dimensional grid. Well, we would get pretty much the same answers on a one dimensional grid (a line), but if we "walk" on a three dimensional grid, our odds of returning to the starting point are less than certain and (consistent with this) the number of times we expect to return is finite. Eventually we escape! (To picture a three dimensional grid, imagine you are in a large many story building with hallways laid out in a rectangular grid and stair cases between floors at every place that two hallways intersect. If you wander randomly in such a building, expect to get lost. But if you never climb a staircase, expect to find your way back to where you started.)

In lecture, we will see that the average of the square of your distance from the starting point after n steps is exactly $\sqrt{n}$ (measured in units of the length of the blocks in the city). That is the "mean squared displacement after steps is." Since this is the mean square distance, we expect the typical distance from the start to be. One of the key things we will look at about the random walk is how to formulate this statement more precisely. This will lead us to the Central Limit Theorem and the "Invariance Principle."

Percolation. Now suppose that some of the streets in the city are blocked. Indeed suppose many of them are blocked with each connection between corners available, or open, only with probability p, where p is a number between 0 and 1. Suppose also that different streets are blocked independently of one another. (What does this mean? We will come back to that.) If p is really small, then it is very hard to get anywhere and we don't expect to be able to find a path that goes very far. On the other hand, if is really close to one then most streets are open and chances are we can go a long distance. Percolation is concerned with more precise answers to the question of "how far can we go?" It turns out, in this example, that there is a sharp transition between going "almost nowhere" and going "as far as you like (with good probability)" when p passes by $\frac{1}{2}$. This is something we will try to understand.

Ising model. This is a more sophisticated model of statistical mechanics, involving "spins" that can be up $+1$ or down -1 at each corner in our grid. A collection of spins is assigned a probability, which is it's chance of occurring, in such a way that spins at neighboring corners tend to agree with one another. There is a parameter, the "temperature," which governs how likely they are to agree. When it is very large they don't care much about agreeing, but when it is small there is a strong penalty for disagreement. Again, there is a transition, but this turns out to be a bit more involved to define and to prove.

CHAPTER 1

Probability

1.1. Lecture 2: Probabilities and Sample Spaces

We start with a quick overview of probability from an operational viewpoint. By "operational," I mean that we are going to skip over philosophical questions of how probabilities are determined and what they mean, focusing instead on how to compute things. Those philosophical questions are quite interesting and are certainly important, but they aren't really mathematics and they would take us too far afield.

Probabilities. A probability will be a number between 0 and 1, and we will associate probabilities to outcomes for some "experiment." A typical experiment would be flipping a coin, say, three times. Then the possible outcomes are:

$HHH, HHT, HTH, THH, HTT, THT, TTH, TTT$

The sample space is the set of all possible outcomes for the experiment. We will typically use Ω to indicate the sample space and ω to denote an outcome. A probability distribution on Ω is a function

$$p : \Omega \to [0,1]$$

such that

$$\sum_{\omega \in \Omega} p(\omega) = 1.$$

For flipping a coin three times, we would typically make the probability $p(\omega) = \frac{1}{8}$ for all outcomes to model an unbiased coin.

Often times we want to compute not the probability of a single outcome but the probability of an event, which is a collection of outcomes sharing some property. More formally and event A is a subset of Ω. For instance, we might play a game in which we toss a coin three times and I win if two or more heads come up, while you win if two or more tails come up. Then the event that I win is simply

$$V = \{HHH, HHT, HTH, THH\},$$

and we would naturally say that the probability of V is

$$P(V) = \sum_{\omega \in V} p(\omega) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}.$$

In general we will define the probability of an event $A \subset \Omega$ to be

$$P(A) = \sum_{\omega \in A} p(\omega).$$

We will call the function P a probability measure on Ω.

Infinite Sample Spaces. Everything that was just said can be carried over to a sample space Ω which is infinite, provided Ω is countable. (Recall that countable means that Ω is in 1-1 correspondence with the natural numbers$\mathbb{N} = \{1, 2, 3, \ldots .\}$)

If the sample space is uncountable, like the real numbers $\mathbb{R}$, then the above definitions really don't make sense. It turns out that the right thing to do is to forget all about the function$p$ and to consider just the function P defined on events. We will come back to this point. However, for the moment we will stick to finite, and sometimes countable, sample spaces.

Properties of probability measures. If Ω is a finite set, the uniform probability measure on Ω is the probability measure P which comes from assigning equal weights to each outcome ω. We have

Proposition 1.1. *Let Ω be a finite sample space and let P be the uniform probability measure on Ω. Then*

$$P(A) = \frac{|A|}{|\Omega|}$$

for all $A \subset \Omega$, where $|A|$ = number of elements of A, also called the cardinality of A.

PROOF. Since $p(\omega)$ is the same for all $\omega \in \Omega$ and $\sum_\omega p(\omega) = 1$, we must have $p(\omega) = 1/|\Omega|$. So,

$$P(A) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|},$$

□

The complement of an event A, denoted A^c, is the event that A does not occur:

$$A^c = \{\omega \,:\, \omega \notin A\}.$$

The intersection of two events A, B, denoted $A \cap B$, is the event that both A and B occur

$$A \cap B = \{\omega \,:\, \omega \in A \text{ and } \omega \in B\},$$

while the union of A and B, denoted $A \cup B$, is the event the one or both of the events A or B occurs

$$A \cup B = \{\omega \,:\, \omega \in A \text{ or } \omega \in B\}.$$

The emptyset $\emptyset = \{\}$ is an event as is the sample space. The probability of the emptyset $P(\emptyset) = 0$ and the probability of Ω is $P(\Omega) = 1$. Two events A_1 and A_2 are disjoint, also called mutually exclusive, if $A_1 \cap A_2 = \emptyset$. A collection of events $A_1, A_2, \ldots A_n$ is called pairwise disjoint, or again mutually exclusive, if $A_i \cap A_j = \emptyset$ for all i, j. We will use the notation $B \setminus A = B \cap A^c$ for the event that B occurs but A does not.

Lemma 1.1. (1) *For mutually exclusive events $A_1, \ldots, A_n$ we have $P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$.*
(2) $P(A^c) = 1 - P(A)$.
(3) *If $A \subset B$ then $P(B) = P(A) + P(B \setminus A)$, so in particular $P(A) \leq P(B)$.*
(4) *(*Inclusion exclusion principle*):* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

PROOF. (1) Use induction on n. Consider first the case $n = 2$. If $A \cap B = \emptyset$ then

$$\sum_{\omega \in A \cup B} p(\omega) = \sum_{\omega \in A} p(\omega) + \sum_{\omega \in B} p(\omega)$$

since an outcome ω can either fall in A or B but not in both. This proves the identity in case $n=2$. Now suppose we have verified the identity for collections of $n-1$ mutually exclusive events, with $n>3$, and suppose given $A_1,\dots,A_n$ that are mutually exclusive. Then

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= P\left(A_n \cup \bigcup_{i=1}^{n-1} A_i\right) \\ &= P(A_n) + P\left(\bigcup_{i=1}^{n-1} A_i\right) \end{aligned}$$

by applying the case$n=2$. By then applying the result for $n-1$ sets we get the identity claimed.

(2) A^c and A are mutually exclusive and $A\cup A^c=\Omega$.

(3) A and $B\setminus A$ are mutually exclusive and $A\cup B\setminus A=B$.

(4) $A\cup B=A\cup(B\setminus A)$ where A and $B\setminus A$ are mutually exclusive. Thus

$$P(A\cup B)=P(A)+P(B\setminus A).$$

Now $B=(B\cap A)\cup(B\setminus A)$ where $B\cap A$ and $B\setminus A$ are mutually exclusive. Thus

$$P(B\setminus A)=P(B)-P(B\cap A)$$

and the result follows. □

The inclusion exclusion principle can be generalized to larger collections of events:

Lemma 1.2. *Given events $A_1,\dots,A_n$ we have*

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i\cap A_j) \\ &\quad + \sum_{i<j<k} P(A_i\cap A_j\cap A_k) \\ &\quad + \cdots + (-1)^n P(A_1\cap\cdots A_n). \end{aligned}$$

1.2. Lecture 3: Conditional Probabilities and Baye's Rule

A common type of question that arises in probability is to compute the likelihood of an event given some piece of information. In the example of last time, flipping a coin three times in a row, we could ask, for instance, "what are the chances of obtaining one or more heads overall given that the first coin came up tails?"

Of course in this case we can determine the answer by brute force. There are four outcomes for which the first coin comes up tails,

$$\{THH, THT, TTH, TTT\},$$

and, of these four, three have at least one head showing. So the answer should be $3/4$, since we are supposing all outcomes are equally likely.

Now what did we do here? We wished to find the probability of the event

$$A=\{HHH, HHT, HTH, THH, HTT, THT, TTH\}$$

given that another event

$$B=\{THH, THT, TTH, TTT\}$$

is known to have occurred. Thus we replaced our sample space with B and considered the event $A \cap B$ that A and B occur. The probability of $A \cap B$ is $P(A \cap B) = \frac{3}{8}$. However, this should be compared only to the fraction of outcomes which lead to B, since B is known to have occurred. The latter fraction is $B = \frac{1}{2}$, giving the answer

$$\frac{P(A \cap B)}{P(B)} = \frac{3}{4}.$$

Conditional Probabilities.

Definition 1.1. If A and B are events in a sample space Ω and $B > 0$ then the *conditional probability of A given B* is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Example 1.1. Here is another example. (Example 1.4.6 in the text.) Consider families with two children. Suppose that boys and girls are equally likely to occur. We go to visit one of these families and a boy answers the door. Given this, what is the probability that the other child is a boy? One might be tempted to give the quick answer $\frac{1}{2}$ since the chances that the other child is a boy or a girl are equal. However, THIS IS WRONG!!!!!!

The easiest way to see what is going on is to work this out with a sample space. What we are asking for is the conditional probability that a family has two boys given that it has at least one boy? The over all sample space is

$$\Omega = \{bb, bg, gb, gg\},$$

where the order indicates the order of children by age, say. (It is very important that our sample space is Ω and not $\{bb, bg, gg\}$ which ignores the order of the children. Try to understand why this is important!) Once a boy opens the door, we know that the event that the family has at least one boy

$$B = \{bb, bg, gb\}$$

is realized. However, the event that there are two boys is

$$A = \{bb\}.$$

Assuming all outcomes are equally likely we see that

$$P(A|B) = \frac{1}{3}$$

so $1/3$ of families with at least one boy have two boys.

You may enjoy reading about an infamous example of this problem here.

Partitions.

Definition 1.2. A collection of events $B_1, \ldots, B_n$ is a *partition* if the events are mutually exclusive and $\bigcup_{i=1}^{n} B_i = \Omega$.

Theorem 1.1. *Let $B_1, \ldots, B_n$ be a partition and assume $P(B_i) > 0$ for all i. For any event A we have*

$$P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

PROOF. Write A as the disjoint union $A = (A \cap B_1) \cup (A \cap B_2) \cup \ldots \cup (A \cap B_n)$. So

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A|B_i)P(B_i).$$

□

Example 1.2. Here is an important example. (Example 1.4.12 in the text.) Suppose we have a population of people, say 100,000 people. Suppose 1% of them are infected with a disease. So there are 1000 sick people. Now there is a test for the disease that is accurate 95% of the time. So roughly 95 out of 100 sick people who take the test give a positive result and roughly 5 out of 100 healthy people who take the test give a positive result. Suppose you are a doctor and give the test to a patient about whom you have no other information and the result comes back positive. What are the chances the person is sick? You may be tempted to answer "95%." You would be in good company as, according to Leonard Mlodinow in his book "The Drunkard's Walk", many physicians would give this answer. However the answer is absolutely wrong! To see this, let us set up two events

$$A = \{\text{the patient is sick}\}$$

and

$$B = \{\text{the patient tests positive}\}.$$

What we want to know is $P(A|B)$. However, the 95% refers to $P(B|A)$, which is not the same thing! Nonetheless we can compute things. Our assumptions about the accuracy of the test give

$$P(B|A) = 0.95\ , \quad P(B|A^c) = 0.05$$

and, since we assumed 1% of the population is sick,

$$P(A) = 0.01.$$

To compute $P(A|B)$ we need to know $P(A \cap B)$ and $P(B)$. We can compute $P(B)$ using the theorem,

$$\begin{aligned} P(B) &= P(B|A)P(A) + P(B|A^c)P(A^c) \\ &= 0.95 * 0.01 + 0.05 * 0.99 = 0.0095 + 0.0495 = 0.059. \end{aligned}$$

To compute $ P(A \cap B)$ we can use $ P(B |A)$ and $ P(A)$

$$P(A \cap B) = P(B|A)P(A) = 0.95 * 0.01 = 0.0095.$$

Thus the probability that the patient is sick is

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.0095}{0.059} = 0.16 = 16\%!$$

This number is probably much smaller than most people realize. The problem here is that the margin of error on the test is actually quite a lot larger than the incidence of the disease.

We can put the pieces from above together to get a general formula

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)},$$

known as Bayes' Rule and which has the following more general form

Theorem 1.2 (Baye's Rule). *Let $B_1,\ldots,B_n$ be a partition with $P(B_i) > 0$ for all i and let A be any event with $P(A) > 0$. Then, for each i,*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}.$$

As above this is proved with the theorem on partitions writing

$$P(A \cap B_i) = P(A|B_i)P(B_i)$$

and

$$P(A) = \sum_{j=1}^{n} P(A|B_j)P(B_j).$$

1.3. Lecture 4: Independent Events; Random Variables

Independence. The last key concept about events we will need is the notion of independent events. The idea is that events A and B are independent if knowing that B occurs does not change the likelihood of A and vice versa. Thus we should have

$$P(A|B) = P(A) \quad \text{and } P(B|A) = P(B)$$

Either of these identities implies that $P(A\cap B) = P(A)P(B)$, so we make the

Definition 1.3. Two events A and B are *independent* if

$$P(A\cap B) = P(A)P(B).$$

More generally events $A_1,\ldots,A_n$ are *independent* if

$$P\left(\bigcap_{j\in J} P(A_j)\right) = \prod_{j\in J} P(A_j)$$

for all sets $J \subset \{1,\ldots,j\}$.

The point of the definition when we have more than two sets is that

$$P(A_k|\bigcap_{j\in h} P(A_j)) = P(A_k)$$

for any index set J and any $k \notin J$. It is not enough to suppose $P(A_j\cap A_k) = P(A_j)P(A_k)$ for all pairs, as the following example shows.

Example 1.3. Flip two fair coins and let

$A_1 =$ first coin comes up heads,

$A_2 =$ second coin comes up heads,

$A_3 =$ both coins come up the same.

Then $P(A_j) = \frac{1}{2}$ for $j = 1,2,3$. Furthermore $P(A_i \cap A_j) = \frac{1}{4}$ for all pairs $i \neq j$. However $A_1 \cap A_2 \subset A_3$ so

$$P(A_1 \cap A_2 \cap A_3) = \frac{1}{4}$$

and $P(A_3 | A_1 \cap A_2) = 1$.

Here are several key facts about independence

Lemma 1.3. (1) *If $A_1, \ldots, A_n$ are independent then so are $A_1, \ldots, A_{n-1}, A_n^c$.*
(2) *If $P(A) = 0$ or 1 then A and B are independent for any event B.*

Remark 1.1. Using (1) we can replace any number of the sets $A_1, \ldots, A_{n-1}$ by their complements.

PROOF. Let $B_i = A_i$ if $1 \leq i < n$ and $B_n = A_n^c$. We need to show

$$P\left(\bigcap_{j \in J} B_j\right) = \prod_{j \in J} P(B_j)$$

for any index set $J \subset \{1, \ldots, n\}$. If $n \notin J$ there is nothing to prove since $B_1 = A_1, \ldots, B_{n-1} = A_{n-1}$ are independent by assumption. On the other hand, if $n \in J$ then

$$\bigcap_{j \in J} B_j = \left(\bigcap_{j \in J \setminus \{n\}} A_j\right) \cap A_n^c.$$

However

$$\bigcap_{j \in J \setminus \{n\}} A_j = \underbrace{\left(\bigcap_{j \in J} A_j\right)}_{E_1} \cup \underbrace{\left(\left(\bigcap_{j \in J \setminus \{n\}} A_j\right) \cap A_n^c\right)}_{E_2},$$

where E_1 and E_2 are mutually exclusive events. Thus

$$P\left(\left(\bigcap_{j \in J \setminus \{n\}} A_j\right) \cap A_n^c\right) = P\left(\bigcap_{j \in J \setminus \{n\}} A_j\right) - P\left(\bigcap_{j \in J} A_j\right),$$

and since $A_1, \ldots, A_n$ are independent

$$\begin{aligned} P\left(\left(\bigcap_{j \in J \setminus \{n\}} A_j\right) \cap A_n^c\right) &= \left(\prod_{j \in J \setminus \{n\}} P(A_j)\right)(1 - P(A_n)) \\ &= \left(\prod_{j \in J \setminus \{n\}} P(A_j)\right) P(A_n^c) \end{aligned}$$

which is what we needed to show.

To prove (2), first note that if $P(A) = 0$ then $P(A \cap B) =$ since $A \cap B \subset A$. Thus

$$P(A \cap B) = 0 = P(A)P(B),$$

so A and B are independent. On the other hand, if $P(A) = 1$ then $P(A^c) = 1 - P(A) = 0$ so A^c and B are independent and it follows from (1) that A and B are independent. $\square$

Random Variables. Imagine we play a game. We flip a coin and every time it comes up heads you give me a dollar if it comes up tails I give you a dollar. If we play this game n times then the amount of money you have one at the end is

$$W = \sum_{j=1}^{n} w_j$$

where w_j denotes your winnings on the $j^t h$ coin toss:

$$w_j = \begin{cases} 1 & \text{if the } j^{\text{th}} \text{ toss comes up heads,} \\ -1 & \text{if the } j^{\text{th}} \text{ toss comes up tails.} \end{cases}$$

The quantities W and w_j depend on the outcome of the n coin tosses and are examples of random variables, where

Definition 1.4. A *random variable* X is a map from a probability space Ω into $\mathbb{R}$.

Thus random variables are just functions on probability spaces.

Definition 1.5. The cumulative distribution function of a random variable X is the increasing function $F_X : \mathbb{R} \to [0,1]$ given by

$$F_X(x) = P(\{\omega | X(\omega) \leq x\}).$$

We will often use the shorthand of suppressing the set quantifier notation writing $P(X \leq x)$ instead of $P(\omega | X(\omega) \leq x)$.

Note that F_X is an increasing function of x and that it is right continuous:

$$F_X(x) = \lim_{y \downarrow x} F_X(y).$$

By contrast

$$\lim_{y \uparrow x} F_x(y) = P(X < x).$$

Thus

$$P(X = x) = F_X(x) - \lim_{y \uparrow x} F_x(y).$$

The function $F_X(x)$ is constant except at the points $x \in \mathbb{R}$ for which there is ω such that $x = X(\omega)$. At each such point $F_X(x)$ has a jump discontinuity, jumping up by $P(X = x)$. (Try plotting $F_{w_1}(x)$ for w_1 as above, assuming that heads and tails are equally likely.)

Definition 1.6. The indcued *probability mass function* for X is

$$p_X(x) = P(X = x).$$

This function defines a probability measure on the uncountable sample space $\mathbb{R}$ since $p_X(x) = 0$ except at finitely many x:

$$P_X(A) := \sum_{x \in A} p_X(x).$$

Indeed with this definition we have $P_X(A) = P(X \in A)$, where $(X \in A)$ is shorthand for the event

$$X^{-1}(A) = \{\omega | X(\omega) \in A\}.$$

1.4. Lecture 5: Expectation Values

Recall that a random variable is a map $X : \Omega \to \mathbb{R}$. The *expectation* of a random variable is the average value:

$$\mathbb{E}(X) = \sum_{\omega \in \Omega} X(\omega) p(\omega).$$

This may also be called the *expected value* or the *average* of X.

Proposition 1.2. *Expectation is linear. That is, if X and Y are random variables and if $c \in \mathbb{R}$ then*

$$\mathbb{E}(X + cY) = \mathbb{E}(X) + c\mathbb{E}(Y).$$

PROOF. Exercise. □

We can express the expectation of X in terms of the induced probability mass function $p_X(x) = P(X = x)$

Proposition 1.3. *For any random variable X,*

$$\mathbb{E}(X) = \sum_{x \in \mathbb{R}} x p_X(x).$$

Remark 1.2. Remember that $p_X(x) = 0$ except for finitely many $x \in \mathbb{R}$, so the sum on the right hand side is well defined.

PROOF. Exercise. □

Another interesting formula expresses the expectation in terms of the cumulative distribution function $F_X(t) = P(X \leq t)$ as follows.

Proposition 1.4. *For any random variable X,*

$$\mathbb{E}(X) = \int_0^\infty (1 - F_X(t))dt - \int_{-\infty}^0 F_X(t)dt.$$

Remark 1.3. Note that $F_X(t) = 1$ for t sufficiently large and $F_X(t) = 0$ for t sufficiently small, so both integrals are over finite intervals. Later on we will encounter random variables on infinite spaces where this may not be true and one must check whether the integrals on the right hand side are convergent.

PROOF. First suppose that X is a non-negative random variable (so $P(X \geq 0) = 1$). Let $0 < x_1 < \cdots < x_N$ be all the positive points with m. In between these points $F_X(x)$ is constant and we have

$$p_X(x_j) = F_X(x_j) - F_X(x_{j-1}).$$

(Draw a graph!) Furthermore, because $X \geq 0$ with probability one we have $p_X(0) = F_X(0)$. Since $F_X(t) = 1$ for $t \geq x_N$ we see that

$$\int_0^\infty (1 - F_X(t))dt = \int_0^{x_N} (1 - F_X(t))dt = \sum_{j=1}^N \int_{x_{j-1}}^{x_j} (1 - F_X(x_{j-1}))dt$$

where we let $x_0 = 0$. Thus

$$\int_0^\infty (1 - F_X(t))dt = x_N - \sum_{j=1}^N (x_j - x_{j-1})F_X(x_{j-1}).$$

We can transform the right hand side using

Fact 1.1 (Summation by parts)**.** Let $a_0, \ldots, a_N$ and $b_0, \ldots, b_N$ be finite sequences of real numbers then

$$\sum_{j=1}^N (a_j - a_{j-1})b_{j-1} = a_N b_N - a_0 b_0 - \sum_{j=1}^N a_j(b_j - b_{j-1}).$$

PROOF. Exercise. □

So,

$$\begin{aligned}\int_0^\infty (1 - F_X(t))dt &= x_N - x_N F_X(x_N) + x_0 F_X(x_0) + \sum_{j=1}^N x_j(F_X(x_j) - F_X(x_{j-1})) \\ &= \sum_{j=1}^N x_j p_X(x_j) = \sum_{x \in \mathbb{R}} x p_X(x) = \mathbb{E}(X).\end{aligned}$$

For a general random variable X we write $X = X_+ - X_-$ where

$$X_+ = \begin{cases} X & X \geq 0 \\ 0 & X < 0 \end{cases}, \qquad X_- = \begin{cases} 0 & X > 0 \\ -X & X \leq 0 \end{cases}.$$

By linearity $\mathbb{E}(X) = \mathbb{E}(X_+) - \mathbb{E}(X_-)$. Now both X_+ and X_- are non-negative random variables, so the representation just derived applies to them. Fix $t > 0$. Then

$$F_{X_+}(t) = P(X_+ \leq t) = P(X \leq t) = F_X(t),$$

and

$$F_{X_-}(t) = P(X_- \leq t) = P(X > -t) = 1 - P(X \leq -t) = 1 - F_X(-t)$$

leading to the claimed formula. □

One particular type of random variable is the indicator function of an event A. This is the function that is 1 on A and 0 on A^c. There are two common notations for this χ_A and $I[A]$ which we will use interchangeably. The expectation of the indicator function is just the probability of A:

$$\mathbb{E}(I[A]) = P(A).$$

So expectation generalizes probability. In a similar way we can generalize the idea of conditional probability.

Definition 1.7. The conditional expectation of a random variable X given and event A is

$$\mathbb{E}(X|A) = \frac{\mathbb{E}(XI[A])}{P(A)}.$$

A common use of this is through an expression like $\mathbb{E}(X|Y = y)$, where X and Y are random variables. For instance we could play the game based on coin flips discussed last time. You can check that your total expected winnings are zero:

$$\mathbb{E}(W) = 0.$$

This follows from the fact that you are as likely to loose a particular sum of money as to win it. However, suppose the first coin comes up heads, so you win a dollar. What are your expected winnings now? That is what is $\mathbb{E}(W|w_1 = 1)$? It is not hard to see that the answer is now 1.

Likewise we can extend the notion of independence to random variables:

Definition 1.8. Random variables $X_1, \ldots, X_n$ are *mutually independent* if for any sets $A_1, \ldots, A_n \subset \mathbb{R}$ the events $\{X_1 \in A_1\}, \ldots, \{X_n \in A_n\}$ are independent.

For example in the coin toss game from last time the random variables $w_1, \ldots w_n$ are independent.

Theorem 1.3. *Suppose $X_1, \ldots, X_n$ are independent. Then*

(1) $\mathbb{E}\left(\prod_{j=1}^{n} X_j\right) = \prod_{j=1}^{n} \mathbb{E}(X_j)$, *and*

(2) $g_1(X_1), \ldots, g_n(X_n)$*are independent given any functions* $g_1, \ldots, g_n : \mathbb{R} \to \mathbb{R}$.

Remark. The notation $g_j(X_j)$ denotes the random variable $g_j \circ X_j : \Omega \to \mathbb{R}$.

PROOF. Let us prove 2 first. Let $A_1, \ldots, A_n$ be subsets of $\mathbb{R}$. Then $g_j(X_j) \in A_j$ if and only if $X_j \in g_j^{-1}(A_j) = \{t : g_j(t) \in A_j\}$. Thus the events $\{g_1(X_1) \in A_1\}, \ldots, \{g_n(X_n) \in A_n\}$ are independent by assumption.

To prove 1, we will use the identity

$$\mathbb{E}\left(\prod_{j=1}^{n} X_j\right) = \sum_{x_1,\ldots,x_n \in \mathbb{R}} x_1 \cdots x_n P\left(\bigcap_{j=1}^{n} \{X_j = x_j\}\right).$$

As above there are only finitely many non-zero terms in the summation. We now use the independence to factor the probability on the right hand side and then factor the summation:

$$\sum_{x_1,\ldots,x_n \in \mathbb{R}} x_1 \cdots x_n P \prod_{j=1}^{n} P(\{X_j = x_j\}) = \prod_{j=1}^{n} \sum_{x_j \in \mathbb{R}} x_j P(\{X_j = x_j\}) = \prod_{j=1}^{n} \mathbb{E}(X_j).$$

The details are left as an exercise. □

If $X_1, \ldots, X_n$ are independent it follows that

$$\mathbb{E}\left(\prod_{j=1}^{n} g_j(X_j)\right) = \prod_{j=1}^{n} \mathbb{E}(g_j(X_j))$$

for any functions $g_1, \ldots, g_n$. This turns out to be a sufficient condition for independence as well:

Theorem 1.4. *Random variables $X_1, \ldots, X_n$ are independent if and only if for any functions $g_1, \ldots, g_n$ from $\mathbb{R} \to \mathbb{R}$ we have*

$$\mathbb{E}\left(\prod_{j=1}^{n} g_j(X_j)\right) = \prod_{j=1}^{n} \mathbb{E}(g_j(X_j))$$

PROOF. We have already seen the forward implication. To prove the backwards note that if $g_1, \ldots, g_n$ are indicator functions of sets $A_1, \ldots, A_n$ then $g_j(X_j) = I[X_j \in A_j]$, so $\mathbb{E}(g_j(X_j)) = P(X_j \in A_j)$ and

$$\mathbb{E}\left(\prod_{j=1}^{n} g_j(X_j)\right) = P\left(\bigcap_{j=1}^{n} \{X_j \in A_j\}\right).$$

Independence follows. □

CHAPTER 2

The Random Walk

2.1. Lecture 6: Random Walk and Stirling's Formula

A common description of Random Walk is as the "drunkard's walk." We imagine an inebriated sailor on his way home on a long straight road. Trouble is he has no idea which way is home so every time he takes a step he stops to think and comes up with a random decision as to which way is home.

Suppose the walker starts at position $x \in \mathbb{Z}$. When he takes a step he moves either to the right by one or to the left by one: $x \mapsto x+1$ or $x-1$. So his position after n steps is

$$X_n = x + S_1 + \cdots + S_n,$$

where $S_1, \ldots, S_n$ are random variables, which we assume to be independent of one another. Furthermore we assume

$$S_j = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases},$$

for each $j = 1, \ldots, n$. This specifies the *joint distribution of the random variables* $S_1, \ldots, S_n$, since if σ_j is a sequence of ± 1 then by independence we have

$$\mathbb{P}\left(\bigcap_{j=1}^{n} S_j = \sigma_j \right) = \prod_{j=1}^{n} \mathbb{P}\left(S_j = \sigma_j\right) = \frac{1}{2^n}.$$

(I have introduced the new notation $\mathbb{P}$ for the probability of an event. This will be useful to distinguish P from p on the board.) So the probability fo any particular outcome is $\frac{1}{2^n}$.

Here is an applet you can play with to explore the sort of trajectories one gets for $1D$ random walks.

What about X_n? Let us start with some easy calculations.

Proposition 2.1. (1) $\mathbb{E}(X_n) = x$, *and*
(2) $\mathbb{E}\left((X_n - x)^2\right) = n$.

PROOF. (1) Follows because $\mathbb{E}(S_j) = 0$ and expectation is linear. To compute the expectation in (2) note that

$$(X_n - x)^2 = \sum_{i=1}^{n} S_i^2 + 2 \sum_{1 \le i < j \le n} S_i S_j.$$

Now $\mathbb{E}\left(S_i^2\right) = 1$, since $S_i^2 = 1$ with probability one. On the other hand, since S_i and S_j are independent, we have

$$\mathbb{E}\left(S_i S_j\right) = \mathbb{E}\left(S_i\right)\mathbb{E}\left(S_j\right) = 0$$

and the result follows. □

The quantity that was computed in (2) is called the *variance* of X_n. More generally we make the

Definition 2.1. Let X be a random variable. The *variance of* X is

$$\mathrm{Var}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right).$$

Proposition 2.2. *If X is a random variable then*

$$\mathrm{Var}(X) = \min_{a \in \mathbb{R}} \mathbb{E}\left((X-a)^2\right) = \mathbb{E}\left(X^2\right) - (\mathbb{E}(X))^2.$$

PROOF. Exercise. □

The variance is a measure of how much a variable deviates from it's mean on average. More specifically it is the "mean square deviation" of the variable.

Let us argue heuristically for a moment. Since the variance of X_n is n we expect $|X_n - x|$ to be roughly of size $\sqrt{n}$. Suppose we want to know

$$\mathbb{P}(X_n = x) = \mathbb{P}(|X_n - x| = 0).$$

We might guess that X_n is more or less uniformly distributed over the sites between $-c\sqrt{n}$ and $c\sqrt{n}$ for some c. Indeed you can check that if a random variable X takes values on the integer sites in $[-a, a]$, with a an integer and with each value occuring with equal probability, then

$$\mathrm{Var}(X) = \frac{2}{2a+1} \frac{(a+1)a(a-1)}{3} \approx \frac{a^2}{3}.$$

So if we ask ourselves "what is the probability that $X_n = 0$?", then a tempting answer is "about $\frac{1}{\sqrt{n}}$."

This heuristic argument is not far from the truth. But we should be suspicious. Mearly knowing the variance tells us, in fact, *very little* about the distribution as the following example shows.

Example 2.1. Let

$$X = \begin{cases} -a & \text{with probability } \varepsilon \\ a & \text{with probability } \varepsilon \\ 0 & \text{with probability } 1 - 2\varepsilon \end{cases}$$

then $\mathbb{E}(X) = 0$ and $\mathrm{Var}(X) = 4a^2\varepsilon$. So, by taking a very very large and ε very very small but with $4a^2\varepsilon = n$ we can produce a random variable which is typically 0 but every once in a while extremely large but nonetheless has variance n.

In fact, our heuristic reasoning missed a key point about X_n, namely $X_n - x$ *is even if n is even and odd if n is odd.* Aside from this, our heuristic argument gives essentially the right result. Now we will proceed to finding a more precise answer to the question: "What is the probability that $X_n = x$?"

In fact it is not hard to compute the probabilities that X_n takes any particular value:

Proposition 2.3. (1) $\mathbb{P}(X_{2n} = x + k) = \begin{cases} 0 & \textit{if } k \textit{ is odd,} \\ \frac{1}{2^{2n}} \binom{2n}{n + \frac{k}{2}} & \textit{if } k \textit{ is even.} \end{cases}$

(2) $\mathbb{P}(X_{2n+1} = x + k) = \begin{cases} \frac{1}{2^{2n+1}} \binom{2n+1}{n + \frac{k-1}{2}} & \textit{if } k \textit{ is odd,} \\ 0 & \textit{if } k \textit{ is even.} \end{cases}$

PROOF. We have essentially done this. And there is a homework problem. □

The trouble with this answer is that it is hard to make sense out of the formulas. They are *precise* but not terribly useful for computing. To proceed we need a good way to estimate factorials.

Stirling's Formula.

Proposition 2.4.

$$\lim_{n\to\infty} \frac{\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}}{n!} = 1$$

Remark 2.1. A common short hand for this statement is $n! \sim \sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n}$.

Let's start by proving something weaker, namely that

$$b_n = \frac{n!}{n^{n+\frac{1}{2}}e^{-n}}$$

has a limit as $n \to \infty$. To see this, we look at

$$\frac{b_n}{b_{n-1}} = \frac{n!}{n^{n+\frac{1}{2}}e^{-n}} \frac{(n-1)^{n-\frac{1}{2}}e^{-n+1}}{(n-1)!} = \left(1-\frac{1}{n}\right)^{n-\frac{1}{2}} e. \tag{2.1.1}$$

If we let $b_0 = 1$ then (2.1.1) does not hold for $n = 1$, instead

$$\frac{b_1}{b_0} = e.$$

Now we can write

$$b_n = \frac{b_n}{b_{n-1}} \frac{b_{n-1}}{b_{n-2}} \cdots \frac{b_1}{b_0}.$$

Thus the existence of the limit $\lim_{n\to\infty} b_n$ follows from the following

Claim 2.1. The infinite product

$$\prod_{n=2}^{\infty} \left[\left(1-\frac{1}{n}\right)^{n-\frac{1}{2}} e \right]$$

converges.

PROOF. Convergence of an infinite product is the same as convegence of the series defining it's logarithm. In this case

$$\sum_{n=2}^{\infty} \left[\left(n-\frac{1}{2}\right) \ln\left(1-\frac{1}{n}\right) + 1 \right].$$

To controll this sum we can apply a comparison test, comparing it with $\sum_{n=2}^{\infty} \frac{1}{n^2}$, which is known to be convergent. To make this comparison we use Taylor's Theorem, which shows that

$$\ln(1-x) = -x - \frac{1}{2}x^2 + R_2(x)$$

where $R_2(x)$ is "big-oh of x^3," that is there is a constant $c < \infty$ such that $|R_2(x)| \le c|x|^3$ provided $|x| \le \frac{1}{2}$, say. (Of course, $|x| \le \frac{1}{2}$ could be replaced by $|x| \le \varepsilon$ for any small ε.) Thus

$$\begin{aligned}
\left(n-\frac{1}{2}\right) \ln\left(1-\frac{1}{n}\right) + 1 &= -\left(n-\frac{1}{2}\right)\left(\frac{1}{n}+\frac{1}{2n^2}\right) + 1 + \left(n-\frac{1}{2}\right) R_2\left(\frac{1}{n}\right) \\
&= -1 + \frac{1}{2n} - \frac{1}{2n} + \frac{1}{4n^2} + 1 + \left(n-\frac{1}{2}\right) R_2\left(\frac{1}{n}\right) \\
&= \frac{1}{4n^2} + \left(n-\frac{1}{2}\right) R_2\left(\frac{1}{n}\right).
\end{aligned}$$

Since $\left|\left(n-\frac{1}{2}\right)R_2\left(\frac{1}{n}\right)\right| \le c\frac{1}{n^2}$ we see that there is a finite constant C such that

$$\left|\left(n-\frac{1}{2}\right)\ln\left(1-\frac{1}{n}\right)+1\right| \le C\frac{1}{n^2}.$$

Thus the series is summable as claimed and thus the product exists. □

Note that this proof does not tell us *what the value of* $\lim b_n$ *is.* Later we will show that it is indeed $\sqrt{2\pi}$. First, however we will apply this result to show what happens to the distribution of X_n when n gets large.

2.2. Lecture 7: Central Limit Theorem for the Random Walk

Probability of finding a random walker at the origin. Given Stirling's formula from last time we can now answer the question: "What is the probability that a random walker in $1D$ returns to his starting position after $2n$ steps?"

Indeed, we have

$$\mathbb{P}(X_{2n}=0) = \binom{2n}{n}2^{-2n} = \frac{(2n)!}{n!n!}2^{-2n}.$$

Applying Stirling's formula in the form we proved last time, namely that the following limit exists

$$\lim_{n\to\infty}\frac{n^{n+\frac{1}{2}}e^{-n}}{n!} = C_0,$$

we see that

$$\begin{aligned}\mathbb{P}(X_{2n}=0) &\sim \frac{C_0 2^{2n+\frac{1}{2}}n^{2n+\frac{1}{2}}e^{-2n}}{\left(C_0 n^{n+\frac{1}{2}}e^{-n}\right)^2}2^{-2n}\\ &= \frac{\sqrt{2}}{C_0}\frac{1}{\sqrt{n}}.\end{aligned}$$

Up to the mysterious constant $\sqrt{2}/C_0$ this is asymptotic relation suggested by the heuristic argument last time.

Similarly, if j is some *fixed* finite number then

$$\begin{aligned}\mathbb{P}(X_{2n}=2j) &= \left(\frac{(2n)!}{(n+j)!(n-j)!}\right)2^{-2n}\\ &\sim \frac{\sqrt{2}}{C_0}\sqrt{\frac{n}{n^2-j^2}}\frac{1}{\left(1+\frac{j}{n}\right)^{n+j}\left(1-\frac{j}{n}\right)^{n-j}}\end{aligned} \tag{2.2.1}$$

Now, if j is fixed and finite then we get

$$\mathbb{P}(X_{2n}=2j) \sim \frac{\sqrt{2}}{C_0\sqrt{n}}\frac{1}{e^j e^{-j}} = \frac{\sqrt{2}}{C_0}\frac{1}{\sqrt{n}},$$

the same result as for $j=0$. Thus if we ask "what is $\mathbb{P}(X_{2n}\in[a,b]\cap\mathbb{Z})$ for some points $a,b\in\mathbb{Z}$, $a\le b$?" the answer is

$$\mathbb{P}(X_{2n}\in[a,b]\cap\mathbb{Z}) = \sum_{j\in[a,b]}\mathbb{P}(X_{2n}=j) \sim \frac{\sqrt{2}}{C_0}\frac{1}{\sqrt{n}}N_{[a,b]} \tag{2.2.2}$$

where $N_{[a,b]}$ = number of even points in $[a,b]$. Whatever a,b are this is a very small number. (It is $\mathrm{O}(1/\sqrt{n})$.) A similar calculation would give

$$\mathbb{P}(X_{2n+1} \in [a,b] \cap \mathbb{Z}) \sim \frac{\sqrt{2}}{C_0} \frac{1}{\sqrt{n}} N'_{[a,b]} \tag{2.2.3}$$

with $N'_{[a,b]}$ = number of odd points in $[a,b]$.

Probability of finding the walker far from the origin. The asymptotic relation (2.2.1) holds even if we increase j as we increase n, i.e. we let $j = j(n)$, provided $|j(n)| \le c|n|$ with c<1 a fixed constant so that both $n + j$ and $n - j$ are large. Rewriting it we have

$$\mathbb{P}(X_{2n} = 2j) \sim \frac{\sqrt{2}}{C_0} \frac{1}{\sqrt{n}} \left(1 - \frac{j^2}{n^2}\right)^{-n-\frac{1}{2}} \frac{\left(1 - \frac{j}{n}\right)^j}{\left(1 + \frac{j}{n}\right)^j}.$$

To under stand this relation we will make use of the limit

$$\left(1 + \frac{x}{n}\right)^n \sim e^x \quad n \to \infty, \tag{2.2.4}$$

which you can prove by taking logarithms an applying Taylor's formula

$$\ln\left(1 + \frac{x}{n}\right)^n = n \ln\left(1 + \frac{x}{n}\right) = x + \mathrm{O}\left(\frac{x^2}{n}\right).$$

(Please read the post on Taylor's Theorem, Big Oh and little oh.) Note that (2.2.4) holds even if $x = x(n)$ grows with n, provided $x^2/n \to 0$. In fact, we have (using the Taylor series for e^x),

$$\left(1 + \frac{x}{n}\right)^n = e^x e^{\mathrm{O}\left(\frac{x^2}{n}\right)} = e^x \left(1 + \mathrm{O}\left(\frac{x^2}{n}\right)\right).$$

Thus, we have

$$\begin{aligned}\mathbb{P}(X_{2n} = 2j) &\sim \frac{\sqrt{2}}{C_0} \frac{1}{\sqrt{n}} \frac{1}{\sqrt{1 - \frac{j^2}{n^2}}} e^{\frac{j^2}{n}} \left(1 + \mathrm{O}\left(\frac{j^4}{n^3}\right)\right) \frac{e^{-\frac{j^2}{n}}\left(1 + \mathrm{O}\left(\frac{j^3}{n^2}\right)\right)}{e^{\frac{j^2}{n}}\left(1 + \mathrm{O}\left(\frac{j^3}{n^2}\right)\right)} \\ &\sim \frac{\sqrt{2}}{C_0} \frac{1}{\sqrt{n}} e^{-\frac{j^2}{n}} \left(1 + \mathrm{O}\left(\frac{j^3}{n^2}\right)\right).\end{aligned}$$

In other words, as long as $j(n)^3/n^2 \to 0$, or $j(n) = \mathrm{o}\left(n^{\frac{2}{3}}\right)$, we have

$$\mathbb{P}(X_{2n} = 2j) \sim \frac{\sqrt{2}}{C_0} e^{-\frac{j^2}{n}}. \tag{2.2.5}$$

Chebyshev's Inequality. Eq. (2.2.5) is mostly interesting for $j = \mathrm{O}\left(\sqrt{\mathrm{n}}\right)$. In fact, using our variance calculation from last time

$$\mathrm{Var}(X_n) = n$$

we can easily show

Proposition 2.5. *For any n and any $\lambda > 0$*

$$\mathbb{P}(|X_n| > \lambda) \le \frac{n}{\lambda^2}.$$

To prove this we will use

Theorem 2.1 (Chebyshev's Inequality)**.** *Let X be a non-negative random variable. Then*

$$\mathbb{P}(X > \lambda) \leq \frac{\mathbb{E}(X)}{\lambda}.$$

PROOF. Notice that

$$\lambda \mathbb{P}(X > \lambda) = \mathbb{E}(\lambda I[X > \lambda]) \leq \mathbb{E}(X I[X > \lambda]) \leq \mathbb{E}(X).$$

□

PROOF. (*Proof of Prop. 2.5*) Apply Chebyshev's inequality to $X = X_n^2$ to get

$$\mathbb{P}(|X_n| > \lambda) = \mathbb{P}\left(X_n^2 > \lambda^2\right) \leq \frac{n}{\lambda^2}.$$

□

Thus we have

$$\mathbb{P}\left(|X_n| > \lambda\sqrt{n}\right) \leq \frac{1}{\lambda^2},$$

so

$$\mathbb{P}\left(|X_n| \leq \lambda\sqrt{n}\right) \geq 1 - \frac{1}{\lambda^2},$$

This shows that X_n is indeed $\mathrm{O}(\sqrt{n})$ with probability close to one, if we take λ large.

Indeed, since we get the asymptotic relation (2.2.5) for $j = \mathrm{o}\left(n^{\frac{2}{3}}\right)$, it is interesting to take $\lambda = n^{\alpha}$ with $0 < \alpha < \frac{1}{6}$, say $\alpha = \frac{1}{12}$. Then we have

$$\mathbb{P}\left(|X_n| > n^{\frac{1}{2}}\right) \leq \frac{1}{n^{2\alpha}}.$$

Central Limit Theorem. The most interesting question to ask is what happens if we let $j \approx r\sqrt{n}$ for some number r. This is interesting because as we have seen X_n is typically of order $\sqrt{n}$ in size. Using (2.2.5) we see that

$$\mathbb{P}(X_{2n} = 2j) \sim \frac{\sqrt{2}}{C_0}\frac{1}{\sqrt{n}}e^{-\frac{j^2}{n}}$$

for $|j| = \mathrm{O}(\sqrt{n})$. Thus for any two real numbers $a < b$

$$\mathbb{P}\left(a\sqrt{2n} \leq X_{2n} \leq b\sqrt{2n}\right) \sim \frac{\sqrt{2}}{C_0} \sum_{a\sqrt{\frac{n}{2}} \leq j \leq b\sqrt{\frac{n}{2}}} \frac{1}{\sqrt{n}}e^{-\frac{j^2}{n}}.$$

Letting $\zeta = \sqrt{\frac{2}{n}}j$, we can write this.

$$\mathbb{P}\left(a\sqrt{2n} \leq X_{2n} \leq b\sqrt{2n}\right) \sim \frac{1}{C_0} \sum_{\zeta \in \sqrt{\frac{2}{n}}\mathbb{Z} \cap [a,b]} \sqrt{\frac{2}{n}} e^{-\frac{1}{2}\zeta^2}.$$

The right hand side looks like a Riemann sum for the integral

$$\frac{1}{C_0}\int_a^b e^{-\frac{1}{2}x^2}dx.$$

Indeed we have the paritition $\sqrt{\frac{2}{n}}\mathbb{Z}\cap[a,b]$ which has mesh size $\sqrt{\frac{2}{n}}$ and the function $e^{-\frac{1}{2}x^2}$ is evaluated at the left end point, say. Thus we have almost proved

Theorem 2.2 (Central Limit Theorem for the $1D$ Random Walk)**.** *Let X_n be a one dimensional nearest neighbor random walk, then for all $a<b$ in $\mathbb{R}$*

$$\lim_{n\to\infty}\mathbb{P}\left(a\sqrt{n}\le X_n\le b\sqrt{n}\right)=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{1}{2}x^2}dx.$$

PROOF. Aside from the identity $C_0=\sqrt{2\pi}$ this was shown above for the subsequence X_{2n}. The proof of the limit for odd n is substantially the same as the derivation given above and is left as an exercise.

The constant C_0 can be computed by noting that

$$1=\mathbb{P}(-\infty<X_{2n}<\infty)=\mathbb{P}\left(|X_{2n}|>\lambda\sqrt{n}\right)+\mathbb{P}\left(|X_{2n}|\le\lambda\sqrt{n}\right).$$

By Chebeyshev's inequality we have $\mathbb{P}(|X_{2n}|>\lambda\sqrt{n})\le\frac{1}{\lambda^2}$ uniformly in n. Thus, after taking limits as $n\to\infty$ we get

$$1=\frac{1}{C_0}\int_{-\lambda}^{\lambda}e^{-\frac{1}{2}x^2}dx+\mathrm{O}\left(\frac{1}{\lambda^2}\right).$$

Taking $\lambda\to\infty$ we see that

$$C_0=\int_{-\infty}^{\infty}e^{-\frac{1}{2}x^2}dx.$$

Thus the proof is completed by the following Lemma □

Lemma 2.1. $\int_{-\infty}^{\infty}e^{-\frac{1}{2}x^2}dx=\sqrt{2\pi}$.

Remark 2.2. This also completes the proof of Stirling's Formula from last time.

PROOF. Let's compute the square of the integral

$$\left(\int_{-\infty}^{\infty}e^{-\frac{1}{2}x^2}dx\right)^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-\frac{1}{2}x^2-\frac{1}{2}y^2}dxdy=\int_0^{\infty}\int_0^{2\pi}e^{-\frac{1}{2}r^2}rdrd\theta=2\pi.$$

□

2.3. Lecture 8: Moment Generating Function and Large Deviations

In Lecture 7 we obtained the central limit theorem which said, in brief, that if X_n is a $1D$ random walk then

$$\lim_{n\to\infty}\mathbb{P}\left(a\sqrt{n}\le X_n\le b\sqrt{n}\right)=\int_a^b\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx. \tag{2.3.1}$$

It follows that

$$\lim_{n\to\infty}\mathbb{E}\left(f\left(\frac{X_n}{\sqrt{n}}\right)\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}f(x)e^{-x^2/2}dx \tag{2.3.2}$$

for just about any funtion f you might write down, a fact that we will explore in this lecture.

Indeed (2.5.1) is just (2.3.2) for

$$f(x)=I[a\le x\le b].$$

By taking linear combinations we see that (2.3.2) holds if f is a *piecewise constant* function supported in a bounded interval:

$$f(x) = \sum_{j=1}^{n} c_j I[a_j \le x \le b_j].$$

Recall that a sequence of functions f_n *converges to* f *uniformly* if

$$\lim_{n\to\infty} \sup_x |f_n(x) - f(x)| = 0.$$

It is a fact from real analysis (see the supplement to Lecture 8) that: *If* $f_n \to f$ *uniformly and* f_n *is integrable on* $[a,b]$ *then* f *is integrable on* $[a,b]$ *and*

$$\int_a^b f(x) = \lim_{n\to\infty} \int_a^b f_n(x).$$

Thus we have

Theorem 2.3. *If there is a sequence of piecewise constant functions* f_k*, all supported on a bounded interval* $[-M,M]$ *and such that* $f_k \to f$ *uniformly then* (2.3.2) *holds for* f.

PROOF. The convergence

$$\lim_{k\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f_k(x) e^{-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-x^2/2} dx$$

is guaranteed by the uniform convergence theorem for integration mentioned above. On the other side, we have

$$\left|\mathbb{E}\left(f\left(X_n/\sqrt{n}\right)\right) - \mathbb{E}\left(f_k\left(X_n/\sqrt{n}\right)\right)\right| \le \mathbb{E}\left(\left|f\left(X_n/\sqrt{n}\right) - f_k\left(X_n/\sqrt{n}\right)\right|\right) \le \sup_x |f_k(x) - f(x)|.$$

Thus given $\varepsilon > 0$ we can choose k so that

$$\left|\mathbb{E}\left(f\left(X_n/\sqrt{n}\right)\right) - \mathbb{E}\left(f_k\left(X_n/\sqrt{n}\right)\right)\right| < \frac{\varepsilon}{3},$$

and

$$\left|\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f_k(x) e^{-x^2/2} dx - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-x^2/2} dx\right| < \frac{\varepsilon}{3}.$$

Fixing k and choosing N large enough that for $n \ge N$ we have

$$\left|\mathbb{E}\left(f_k(X_n/\sqrt{n})\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f_k(x) e^{-x^2/2}\right| < \frac{\varepsilon}{3}$$

gives, upon combining these three inequalities,

$$\left|\mathbb{E}\left(f\left(X_n/\sqrt{n}\right)\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-x^2/2}\right| < \varepsilon$$

for $n \ge N$. As ε and n are arbitrary this gives the desired result. □

Problem 2.1. Show that any compactly supported continuous function f can be obtained as a uniform limit of piecewise constant functions. (Hint: show that f is uniformly continuous and let the piecewise constant functions be constant on intervals of decreasing size as $n \to \infty$.)

Functions without unbounded support. If we look at the proof of the central limit theorem we see that (2.5.1) also holds if $a=-\infty$ or $b=\infty$. In fact, an important part of the proof was the identity

$$1=\mathbb{P}(-\infty<X_n<\infty)\xrightarrow{n\to\infty}\int_{-\infty}^{\infty}\frac{1}{C_0}e^{-x^2/2}dx$$

which allowed us to calculate the constant C_0. To prove (2.5.1) with, say, $a=-\infty$, note that by Chebyshev's inequality

$$\begin{aligned}\mathbb{P}\left(X_n\le b\sqrt{n}\right)&=\mathbb{P}\left(X_n<-\lambda\sqrt{n}\right)+\mathbb{P}\left(\lambda\sqrt{n}\le X_n\le b\sqrt{n}\right)\\&=\mathrm{O}\left(\frac{1}{\lambda^2}\right)+\mathbb{P}\left(\lambda\sqrt{n}\le X_n\le b\sqrt{n}\right).\end{aligned}$$

Thus there is a finite constant $c>0$ such that

$$\limsup_{n\to\infty}\mathbb{P}\left(X_n\le b\sqrt{n}\right)\le\int_{-\lambda}^{b}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx+\frac{c}{\lambda^2}$$

and

$$\liminf_{n\to\infty}\mathbb{P}\left(X_n\le b\sqrt{n}\right)\ge\int_{-\lambda}^{b}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx-\frac{c}{\lambda^2}.$$

Taking the limit $\lambda\to\infty$ we find that

$$\limsup_{n\to\infty}\mathbb{P}\left(X_n\le b\sqrt{n}\right)=\liminf_{n\to\infty}\mathbb{P}\left(X_n\le b\sqrt{n}\right)=\int_{-\infty}^{b}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx,$$

which gives (2.5.1) with $a=-\infty$.

For any C^1 function f with $\lim_{x\to\infty}f(x)=0$ we can write

$$f(x)=-\int_x^{\infty}f'(t)dt=-\int_{-\infty}^{\infty}f'(t)I[x\le t]dt.$$

Thus,

$$\mathbb{E}\left(f\left(\frac{X_n}{\sqrt{n}}\right)\right)=-\int_{-\infty}^{\infty}f'(t)\mathbb{E}\left(I[X_n\le t\sqrt{n}]\right)dt=-\int_{-\infty}^{\infty}f'(t)\mathbb{P}\left(X_n\le t\sqrt{n}\right)dt.\tag{2.3.3}$$

Now (2.3.3) holds for absolutely any C^1 function f such that $\lim_{x\to\infty}f(x)=0$, since

$$\mathbb{P}(X_n<-n)=0.$$

In fact continuity of the derivative is not really necessary here, as long as $f(x)=-\int_x^{\infty}f'(t)dt$. To pass to the limit $n\to\infty$ we must be a little more careful since we need to know that the integral

$$\int_{-\infty}^{\infty}f'(t)\mathbb{P}\left(X_n\le t\sqrt{n}\right)dt\xrightarrow{n\to\infty}\int_{-\infty}^{\infty}f'(t)\int_{-\infty}^{t}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt.$$

In particular the integral on the r.h.s. must make sense, which certainly puts some restrictions on f'. At this point we can easily obtain the following

Corollary 2.1. *Let f be C^1 on $\mathbb{R}$ with*

$$\int_{\mathbb{R}}|f'(t)|dt<\infty.\tag{2.3.4}$$

Then

$$\lim_{n\to\infty}\mathbb{E}\left(f\left(\frac{X_n}{\sqrt{n}}\right)\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}f(x)e^{-x^2/2}dx.$$

PROOF. Since $\int |f'(t)|dt < \infty$ the limit

$$\lim_{x\to\infty} f(x) = f(0) + \lim_{x\to\infty} \int_0^x f'(t)dt$$

exists. Let $\alpha = \lim_{x\to\infty} f(x)$. Then

$$f(x) = \alpha - \int_x^\infty f'(t)dt.$$

By the above argument

$$\mathbb{E}\left(f\left(\frac{X_n}{\sqrt{n}}\right)\right) = \alpha - \int_{-\infty}^\infty f'(t)\mathbb{P}\left(X_n \le t\sqrt{n}\right)dt.$$

We take the limit $n\to\infty$ on the right hand side using Theorem A.3 below with

$$f_n(t) = f'(t)\mathbb{P}\left(X_n \le t\sqrt{n}\right) \quad \text{and} \quad g(t) = |f'(t)|.$$

(Note that $|f_n(x)| \le g(t)$.) This gives

$$\lim_{n\to\infty} \mathbb{E}\left(f\left(\frac{X_n}{\sqrt{n}}\right)\right) = \alpha - \int_{-\infty}^\infty f'(t) \int_{-\infty}^t \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt,$$

from which the corollary follows after integrating by parts. □

Theorem 2.4 (Dominated Convergence)**.** *Let f_n be a sequence of functions and let g be a non-negative function, all on the real line. If*

(1) *f_n and g are* integrable* *on compact intervals,*
(2) *$|f_n(x)| \le g(x)$ for all x,*
(3) *$\int_{-\infty}^\infty g(x)dx < \infty$, and*
(4) *$\lim_{n\to\infty} f_n(x) = f(x)$ exists for every x, and f is* integrable*.

Then

$$\int_{-\infty}^\infty f(x)dx = \lim_{n\to\infty}\int_{-\infty}^\infty f_n(x)dx.$$

PROOF. For a proof, see the supplement to Lecture 8. □

Remark 2.3. *I have stated this theorem assuming the notion of integrabilty with which you are familiar is *Riemann integration*. That is good enough for what we need to do, but makes results like this awkward and somewhat difficult to prove. I you have seen *Lebesgue integration,* then you may know you can get rid of the assumption that f is integrable and that the limit exists everywhere, provided it existed "almost everywhere," meaning off a set of zero length.

Looking back at the proof of Cor. 2.1, we see that we could have used the error bound

$$\mathbb{P}\left(|X_n| \ge \lambda\sqrt{n}\right) \le \frac{1}{\lambda^2}$$

to show that

$$\left|f'(t)\mathbb{P}\left(X_n \le t\sqrt{n}\right)\right| \le \left|f'(t)\right| \begin{cases} 1 & t \ge -1 \\ \frac{1}{t^2} & t \le -1 \end{cases}.$$

Thus we could have replaced the condition (2.3.5) by

$$\int_{-\infty}^{-1} \frac{|f'(t)|}{t^2} dt + \int_{-1}^\infty |f'(t)|dt < \infty,$$

which is weaker. Symmetrizing the assumptions we could prove

Corollary 2.2. *Let f be C^1 on $\mathbb{R}$ with*

$$\int_{\mathbb{R}} \frac{|f'(t)|}{1+t^2} dt < \infty. \tag{2.3.5}$$

Then

$$\lim_{n\to\infty} \mathbb{E}\left(f\left(\frac{X_n}{\sqrt{n}}\right)\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{-x^2/2} dx.$$

PROOF. Exercise! □

The moment generating function. We have missed something here. Indeed, note that

$$\mathbb{E}\left(\left(\frac{X_n}{\sqrt{n}}\right)^2\right) = 1 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx,$$

as can be seen by integrating by parts:

$$\int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx = -\int_{-\infty}^{\infty} x \frac{d}{dx} e^{-x^2/2} dx = \int_{-\infty}^{\infty} e^{-x^2/2} dx = \sqrt{2\pi}.$$

Thus (2.3.2) holds for $f(x) = x^2$ although (2.3.5) fails in this case! The point is that the estimate we have used to controll $\mathbb{P}(|X_n| \geq \lambda\sqrt{n})$ is far from optimal.

Theorem 2.5. *Let X_n be a 1D random walk. Then for each $t \in \mathbb{R}$*

$$\lim_{n\to\infty} \mathbb{E}\left(e^{tX_n/\sqrt{n}}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx - x^2/2} dx = e^{t^2/2}. \tag{2.3.6}$$

Furthermore, the convergence is uniform in t over compact sets, *i.e., given $M < \infty$*

$$\lim_{n\to\infty} \sup_{|t|\leq M} \left|\mathbb{E}\left(e^{tX_n/\sqrt{n}}\right) - e^{t^2/2}\right| = 0. \tag{2.3.7}$$

PROOF. First we prove the second equality:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx-x^2/2} dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(x-t)^2/2} e^{t^2/2} dx = \frac{e^{t^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = e^{t^2/2}.$$

Regarding the limit, recall that $X_n = S_1 + \cdots + S_n$ where $S_1, \ldots, S_n$ are independent and $S_j = \pm$ with probability $1/2$. Thus

$$e^{tX_n/\sqrt{n}} = \prod_{j=1}^{n} e^{tS_j/\sqrt{n}},$$

with the factors on the right being independent random variables. Thus

$$\mathbb{E}\left(e^{tX_n/\sqrt{n}}\right) = \prod_{j=1}^{n} \mathbb{E}\left(e^{tS_j/\sqrt{n}}\right) = \left[\mathbb{E}\left(e^{tS_1/\sqrt{n}}\right)\right]^n,$$

since all S_j have the same distribution.

However, we can compute

$$\mathbb{E}\left(e^{tS_1/\sqrt{n}}\right) = \frac{1}{2}\left(e^{\frac{t}{\sqrt{n}}} + e^{-\frac{t}{\sqrt{n}}}\right) = \cosh\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{t^2}{n} + \mathrm{O}\left(\frac{t^4}{n^2}\right).$$

Thus

$$\mathbb{E}\left(e^{tX_n/\sqrt{n}}\right) = \left(1 + \frac{t^2}{n} + \mathrm{O}\left(\frac{t^4}{n^2}\right)\right)^n \sim e^{t^2}$$

as claimed. If $|t| \le M$ then

$$\left|\mathbb{E}\left(e^{X_n/\sqrt{n}}\right) - e^{t^2}\right| = \mathrm{O}\left(\frac{M^4}{n}\right),$$

which gives the uniform convergence (2.3.7). □

Corollary 2.3. *For each $s > 0$ there is a constant C_s such that*

$$\mathbb{P}\left(|X_n| \ge \lambda\sqrt{n}\right) \le C_s e^{-s\lambda}.$$

PROOF. Note that

$$\mathbb{E}\left(e^{sX_n/\sqrt{n}}\right) = \mathbb{E}\left(e^{-sX_n/\sqrt{n}}\right) = \mathbb{E}\left(\cosh\left(sX_n/\sqrt{n}\right)\right).$$

Thus, applying Chebyshev's inequality the identity (2.3.6) implies

$$\mathbb{P}\left(|X_n| \ge \lambda\sqrt{n}\right) \le e^{-s\lambda}\mathbb{E}\left(\cosh\left(X_n/\sqrt{n}\right)\right) \le C_s e^{-s\lambda}.$$

□

2.4. Lecture 9: Central Limit Theorem Revisited

The central limit theorm answers the question "What does the probability distribution of a random walker look like after n steps?" The answer is "Roughly like the distribution of $\sqrt{n}\times$ a standard normal random variable." Some comments:

(1) A "standard normal random variable" is a random variable on the (uncountable) sample space $\mathbb{R}$ with probability measure

$$\mathbb{P}(a \le X \le b) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx.$$

(2) A more precise statement is that

$$\mathbb{P}\left(a \le \frac{X_n}{\sqrt{n}} \le b\right) \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}}\int_a^b e^{-x^2/2}dx,$$

from which it follows that

$$\mathbb{E}\left(f\left(\frac{X_n}{\sqrt{n}}\right)\right) \xrightarrow{n\to\infty} \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} f(x)e^{-x^2/2}dx \tag{2.4.1}$$

for all reasonable functions f. Such convergence is called *convergence in distribution* and we say that $X_n/\sqrt{n}$ converges in distribution to a standard normal random variable.

(3) You can't see this by looking at an individual random walker. You have to look at an *ensemble.* Suppose we start 1,000 walkers at the origin and let them all go, independently of one another, for a large number of steps – say 10,000 steps. Then we count the number of walkers who end up at particular sites. The central limit theorem tell's us that the histrogram we would draw would look a lot like the graph of the curve $y = e^{-x^2/20{,}000}$. (Here $20{,}0000 = 2\times n$ with $n = 10{,}000$.) This curve is biggest in an interval of length close to 100.

Last time we proved (2.4.1) for the following types of functions

(1) f a compactly supported piecewise continuous function
(2) f a uniform limit of compactly supported piecewise continuous functions, which includes f a compactly supported continuous function

(3) f a C^1 function with

$$\int_{-\infty}^{\infty} \frac{|f'(x)|}{1+x^2} dx < \infty.$$

(4) $f(x) = e^{sx}$ for any s.

Moments and moment generating functions.

Definition 2.2. Given a random variable X the *moments* of X are the averages $\mathbb{E}\left(X^k\right)$ for $k = 1, 2, \ldots$. The *moment generating function* is the

$$\Phi_X(t) = \mathbb{E}\left(e^{tX}\right).$$

Proposition 2.6. *For any random variable X on a finite probability space*

$$\mathbb{E}\left(X^k\right) = \frac{d^k}{dt^k} \Phi_t(X)\bigg|_{t=0}.$$

PROOF. This follows from the identities $\frac{d^k}{dt^k} e^{tx} = x^k e^{tx}$ and $\frac{d^t}{dt^k} \mathbb{E}\left(e^{tX}\right) = \mathbb{E}\left(\frac{d^k}{dt^k} e^{tX}\right)$, both of which are easily seen to hold. ☐

Corollary 2.4. *For each $k \in \mathbb{N}$*

$$\lim_{n\to\infty} \frac{1}{n^{\frac{k}{2}}} \mathbb{E}\left(X_n^k\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k e^{-x^2/2} = \frac{d^k}{(dt)^k} e^{t^2}\bigg|_{t=0}. \tag{2.4.2}$$

PROOF. The theorem appears to follow from an interchange of differentation and limits. However verifying that this interchange is allowed is tricky. It can be done using complex analysis because $\mathbb{E}\left(e^{tX_n/\sqrt{n}}\right) = \phi_n(t)$ is a complex analytic function of t.

To prove (2.4.2) directly we can proceed as follows. Let

$$\phi(x) = \begin{cases} 1 & |x| \le 1 \\ 1 - (2 - |x|) & 1 < |x| \le 2 \\ 0 & 0 \end{cases}.$$

So ϕ is continuous, $0 \le \phi \le 1$, and

$$1 - \phi(x) \le I\left[|x| \ge 1\right]. \tag{2.4.3}$$

Let $\lambda > 0$. Then

$$\lim_{n\to\infty} \frac{1}{n^{k/2}} \mathbb{E}\left(X_n^k \phi\left(\frac{X_n}{\lambda\sqrt{n}}\right)\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^k \phi\left(\frac{x}{\lambda}\right) e^{-x^2/2} dx \tag{2.4.4}$$

since $x^k \phi(x/\lambda)$ is continuous.

Regarding the l.h.s. of (2.4.4), we have by (2.4.3)

$$\frac{1}{n^{k/2}} \left| \mathbb{E}(X_n^k) - \mathbb{E}\left(X_n^k \phi\left(\frac{X_n}{\lambda\sqrt{n}}\right)\right)\right| \le \frac{1}{n^{k/2}} \mathbb{E}\left(|X_n|^k I\left[|X_n| \ge \lambda\sqrt{n}\right]\right).$$

Problem 2.2. Show for each k that there is a constant $C_k < \infty$ such that

$$|x|^k \le C_k \cosh x.$$

Thus

$$\frac{1}{n^{k/2}}\left|\mathbb{E}(X_n^k)-\mathbb{E}\left(X_n^k\phi\left(\frac{X_n}{\lambda\sqrt{n}}\right)\right)\right| \le C_k\mathbb{E}\left(\cosh\left(\frac{X_n}{\sqrt{n}}\right)I\left[|X_n|\ge\lambda\sqrt{n}\right]\right).$$

By the Cauchy Schwarz Inequality (Lemma 2.2 below), we have

$$\frac{1}{n^{k/2}}\left|\mathbb{E}(X_n^k)-\mathbb{E}\left(X_n^k\phi\left(\frac{X_n}{\lambda\sqrt{n}}\right)\right)\right|^2 \le C_k^2\mathbb{E}\left(\cosh^2\left(\frac{X_n}{\sqrt{n}}\right)\right)\mathbb{E}\left(I\left[|X_n|\ge\lambda\sqrt{n}\right]\right).$$

Now, by the central limit theorem

$$\mathbb{E}\left(\cosh^2\left(\frac{X_n}{\sqrt{n}}\right)\right)=1+\frac{1}{4}\mathbb{E}\left(e^{2X_n/\sqrt{n}}\right)+\frac{1}{4}\mathbb{E}\left(e^{-2X_n/\sqrt{n}}\right)\xrightarrow{n\to\infty}1+\frac{1}{2}e^2$$

(see Theorem 7 from last time) and,

$$\mathbb{E}\left(I\left[|X_n|\ge\lambda\sqrt{n}\right]\right)\xrightarrow{n\to\infty}\frac{2}{\sqrt{2\pi}}\int_\lambda^\infty e^{-x^2/2}dx.$$

Thus

$$\limsup_{n\to\infty}\frac{1}{n^{k/2}}\left|\mathbb{E}(X_n^k)-\mathbb{E}\left(X_n^k\phi\left(\frac{X_n}{\lambda\sqrt{n}}\right)\right)\right|\le C\int_\lambda^\infty e^{-x^2/2}dx.$$

Putting the above estimates together we see that, for any $\lambda>0$,

$$\limsup_{n\to\infty}\frac{1}{n^{k/2}}\left|\mathbb{E}(X_n^k)-\frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty x^ke^{-x^2/2}dx\right|\le C\int_\lambda^\infty e^{-x^2/2}dx+\int_{-\infty}^\infty x^k\left(1-\phi\left(\frac{x}{\lambda}\right)\right)e^{-x^2/2}dx.$$

Since the r.h.s. goes to zero as $\lambda\to\infty$ by dominated convergence we see that the lim sup on the l.h.s. is zero. Thus the first equality of (2.4.2) holds.

Problem 2.3. Prove that

$$\frac{d^k}{dt^k}\int_{-\infty}^\infty e^{tx-x^2/2}dx=\int_{-\infty}^\infty x^ke^{tx-x^2/2}dx,$$

and use this to complete the proof by showing that the second equality of (2.4.2) holds.

□

Lemma 2.2 (Cauchy Schwarz inequality)**.** *Let X and Y be random variables on a probability space Ω then*

$$|\mathbb{E}(XY)|\le\sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}.$$

PROOF. Note that

$$|\mathbb{E}(XY)|\le\mathbb{E}(|X||Y|),$$

and

$$2|X||Y|\le X^2+Y^2.$$

The second inequality follows from the fact that $(|X|-|Y|)^2\ge0$. Putting these two inequalities together gives

$$|\mathbb{E}(XY)|\le\frac{\mathbb{E}(X^2)+\mathbb{E}(Y^2)}{2}. \tag{2.4.5}$$

Now (2.4.5) is not the inequality we wanted to show. However it implies what we want very easily. If $\mathbb{E}\left(X^2\right)$ and $\mathbb{E}\left(Y^2\right)$ are both non-zero, then all we have to do is apply (2.4.5) to the variables $\frac{X}{\sqrt{\mathbb{E}(X^2)}}$ and $\frac{Y}{\sqrt{\mathbb{E}(Y^2)}}$ to discover that

$$\frac{|\mathbb{E}(XY)|}{\sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}} \leq 1.$$

Problem 2.4. Complete the proof by showing that if either $\mathbb{E}(X^2) = 0$ or $\mathbb{E}(Y^2) = 0$ then $\mathbb{E}(XY) = 0$. (Hint: in the first case show that $\mathbb{P}(X = 0) = 1$.)

□

2.5. Lecture 10: Recurrence of the Random Walk

As we saw last time, the central limit theorem answers questions about what happens to an ensemble of walkers. For a graphic illustration of this take a look at the following applet:

DiscreteRandomWalk1D.html

You can choose how many walkers to watch simultaneously and see all their trajectories at once, as well as a histrogram of their positions. Play with this and see how a bell curve like shape emerges.

But now we want to ask questions about *one random walk*. For simplicity let us consider a random walk starting at the origin:

$$X_n = S_1 + \cdots + S_n$$

with S_1, ..., S_n independent and all equal to ± 1 with probability $1/2$ in each case.

- What is the probability that the walk ever returns to the origin? In symbols
$$\mathbb{P}(X_n = 0 \text{ for some } n \geq 1) = ?$$
- If the walk returns to the origin, how many times does it return? Namely, can we compute
$$\mathbb{P}(V = k)$$
for $k = 1, \ldots,$ where $V = \#$ of times $X_n = 0$ for $n \geq 1$.

Average number of returns in a long finite interval. You might be worried a little bit about what the above questions mean exactly. To answer them we need to know about the countable family of random variables $X_1, X_2, \ldots,$. We haven't really talked about such families. In particular the probability space we need is not finite, and it is not even countable. We will see below that if we proceed in an intuitive fashion then we can get a sensible answer without worrying about this. Later we will come back and deal a little more precisely with the problems just raised.

Because of this, let's get started with an honest calculation. Let

$$E_n = \text{Event that } X_n = 0.$$

So

$$\mathbb{P}(E_n) = 0, \quad \text{if } n \text{ is odd,}$$

and, if $n = 2k$ is even then

$$\mathbb{P}(E_{2k}) \sim \frac{1}{\sqrt{\pi k}} \tag{2.5.1}$$

for k large, as we saw in Lecture 7.

Now let V_N = number of return visits the random walk makes to 0 up to step N. So,

$$V_n = \sum_{n=1}^{N} I[X_n = 0]$$

and

$$\mathbb{E}(V_n) = \sum_{n=1}^{N} P(E_n) = \sum_{k=1}^{N/2} P(E_{2k}).$$

The asympotic relation (2.5.1) means that

$$\sqrt{\pi k}\,\mathbb{P}(E_{2k}) \xrightarrow{k\to\infty} 1.$$

Thus, given $\varepsilon > 0$, we can find N_ε such that for $2k > N_\varepsilon$ we have

$$\mathbb{P}(E_{2k}) > \frac{1-\varepsilon}{\sqrt{\pi k}}.$$

For our purposes ε need not be particularly small, just less than 1, say $\varepsilon = 1/2$. Then for $N > N_{1/2}$ we have

$$\begin{aligned} \mathbb{E}(V_N) &\geq \sum_{k=1}^{N_{1/2}/2} P(E_{2k}) + \frac{1}{2\sqrt{\pi}} \sum_{k=\frac{N_{1/2}}{2}+1}^{N/2} \frac{1}{\sqrt{k}} \\ &= \mathbb{E}\left(V_{N_{1/2}}\right) + \frac{1}{2\sqrt{\pi}} \sum_{N_{1/2}<2k\leq N} \frac{1}{\sqrt{k}}. \end{aligned}$$

Because $1/\sqrt{x}$ is a decreasing function, we can estimate the sum on the r.h.s. in terms of an integral:

$$\sum_{N_{1/2}<2k\leq N} \frac{1}{\sqrt{k}} \geq \int_{\frac{N_{1/2}}{2}+1}^{\frac{N}{2}} \frac{1}{\sqrt{x}} dx = \sqrt{2}\left[\sqrt{N} - \sqrt{N_{1/2}+2}\right].$$

Thus

$$\mathbb{E}(V_N) \geq \frac{\sqrt{N}}{\sqrt{2\pi}} + \left(\mathbb{E}\left(V_{N_{1/2}}\right) - \frac{1}{\sqrt{2\pi}}\sqrt{N_{1/2}+2}\right).$$

In other words, if N is large, much much larger than $N_{1/2}$, say, then $\mathbb{E}(V_N)$ is at least as big as a constant times $\sqrt{N}$. So we expect the random walk to make around $\sqrt{N}$ visits to zero.

Average number of returns is infinite. The above calculation suggests that if we let the walk run for ever then the expected number of returns to 0 should be infinite. Let us show that now.

To start we write

$$V = \sum_{n=1}^{\infty} I[X_n = 0].$$

So V is a random variable whose value is the number of visits the walk makes to 0. The same argument we just went through shows that

$$\mathbb{E}(V) = \sum_{n=1}^{\infty} \mathbb{P}(E_n) = \infty.$$

Indeed, clearly $V \geq V_N$ — the number of times we visit altogether is not smaller than the number of times we visit up to step N. Thus $\mathbb{E}(V) \geq C\sqrt{N}$ for all N sufficiently large.

Thus we expect the walk to return infinitely often. However, this calculation doesn't really tell us this. It *could* happen that any particular walk returns only a finite number of times, but that rare walks visit many many times in just such a way as to make the expectation infinite.

Example 2.2 (St. Petersburg Paradox)**.** Imagine we play a game. You flip a coin. If it comes up heads you win \$1. If it comes up tails, we play again. Only now if it comes up heads you win \$2. We keep flipping the coin until it comes up heads. When it does the game ends and you win $\$2^n$, where $n =$ number of tails that showed up before the heads that ended the game. What are your expected winnings?

$$\mathbb{E}(\text{Winnings}) = \sum_{n=0}^{\infty} 2^n \mathbb{P}(n \text{ tails and 1 head}) = \sum_{n=0}^{\infty} 2^n \frac{1}{2^{n+1}} = \sum_{n=0}^{\infty} \frac{1}{2} = \infty.$$

According to the standard interpretation of the expectation, you should be willing to part with an infinite sum of money, or at least everything you have, for a chance to play this game. But none of us would be and therein lies the paradox.

Probability of Return. Let's not worry about the paradox, but instead let's worry about whether something similar happens for V. That this is not so can be seen as follows.

Theorem 2.6. $\mathbb{P}(V = \infty) = 1$.

The key point is actually the following

Lemma 2.3. $\mathbb{P}(V \geq 1) = 1$

PROOF. Suppose not. Then $\mathbb{P}(V \geq 1) = q < 1$. Since $\{V = 0\} = \{V \geq 1\}^c$ we have

$$\mathbb{P}(V = 0) = 1 - q.$$

Now we can compute $\mathbb{P}(V = 1)$. Indeed in order to have $V = 1$, we must have a first visit to the origin at some step n, after which we never visit again. Thus

$$\mathbb{P}(V = 1) = \sum_{n=1}^{\infty} \mathbb{P}(\text{never return after step } n |\text{ first return is at step } n) \mathbb{P}(\text{first return is at step } n).$$

Thus follows because the events $\{\text{first return is at step } n\}$ for $n = 1, \ldots$ are mutually exclusive.

Now the key observation is that

$$\mathbb{P}(\text{never return after step } n |\text{ first return is at step } n) = \mathbb{P}(V = 0) = 1 - q.$$

Indeed, consider the random walk at steps $n+1, \ldots$ given that the walk is at 0 at step n (so $X_n = 0$). If we just relable

$$Y_m = X_{n+m}$$

then Y_m looks just like a random walk starting at 0. The key thing here is that in order to find out what X_{n+m} is we only need to know where X_n is and what happens at step $n+1$, $n+2$, $\ldots$, $n+m$. Everything else is irrelevant.

Thus

$$\mathbb{P}(V = 1) = (1-q) \sum_{n=1}^{\infty} \mathbb{P}(\text{first return is at step } n) = (1-q)\mathbb{P}(V \geq 1) = q(1-q),$$

since

$$\bigcup_{n=1}^{\infty} \{\text{first return is at step } n\} = \{V \geq 1\}.$$

Similarly, we have

$$\begin{aligned}\mathbb{P}(V=k) &= \sum_{n=1}^{\infty}\mathbb{P}(\text{return } k\text{times after step } n|\text{ first return is at step } n)\,\mathbb{P}(\text{first return is at step } n)\\ &= \mathbb{P}(V=k-1)\,q.\end{aligned}$$

Thus, by induction,

$$\mathbb{P}(V=k)=(1-q)q^k.$$

Problem 2.5. Show that, for $q\in[0,1)$,

$$\sum_{k=0}^{\infty}(1-q)q^k=1.$$

It follows from the problem that

$$\sum_{k=0}^{\infty}\mathbb{P}(V=k)=1$$

and thus

$$\mathbb{P}(V=\infty)=0.$$

Therefore,

$$\begin{aligned}\mathbb{E}(V)=\sum_{k=0}^{\infty}k\mathbb{P}(V=k) &= (1-q)q\sum_{k=1}^{\infty}kq^{k-1}\\ &= q(1-q)\frac{d}{dq}\sum_{k=0}^{\infty}q^k\\ &= q(1-q)\frac{d}{dq}\frac{1}{1-q}\\ &= \frac{q}{1-q}<\infty,\end{aligned}$$

contradicting the fact that $\mathbb{E}(V)=\infty$ which we showed above. Thus we must have $q=\mathbb{P}(V\geq 1)=1$. □

The Theorem ($\mathbb{P}(V=\infty)=1$) now follows quite easily by a similar argument. More precisely we have, as above,

$$\mathbb{P}(V=1)=\mathbb{P}(V=0)\mathbb{P}(V\geq 1)=(1-q)q,$$

but with $q=1$. Thus $\mathbb{P}(V=1)=0$. Furthermore, as above,

$$\mathbb{P}(V=k)=\mathbb{P}(V=k-1)\,q=\mathbb{P}(V=k-1).$$

So by induction

$$\mathbb{P}(V=k)=\mathbb{P}(V=1)=0$$

for all k. Thus we must have $\mathbb{P}(V=\infty)=1$. *With probability one the random walk returns to zero infinitely often!*

2.6. Lecture 11: Random Walk in 2D and 3D

We now look at the multi-dimensional random walk.

Random Walk in 2D. Let $\mathbf{X}_n$ be the trajectory of a random walk in two dimensions. So,

$$\mathbf{X}_n = \mathbf{S}_1 + \cdots + \mathbf{S}_n$$

where $\mathbf{S}_1$, ..., $\mathbf{S}_n$ are inedependent random vectors with

$$\mathbf{S}_j = \begin{cases} (1,0) & \text{with probability } \frac{1}{4} \\ (0,1) & \text{with probability } \frac{1}{4} \\ (-1,0) & \text{with probability } \frac{1}{4} \\ (0,-1) & \text{with probability } \frac{1}{4} \end{cases}.$$

Then for each $n = 1, 2, \ldots$, $\mathbf{X}_n$ is a random point in

$$\mathbb{Z}^2 = \mathbb{Z} \times \mathbb{Z} = \{(n,m) \ : \ n,m \in \mathbb{Z}\}.$$

Problem 2.6. Show that $\mathbb{E}(X_n) = 0$ and $\mathbb{E}\left(|X_n|^2\right) = n$, where the length of a vector is

$$|(a,b)| = \sqrt{a^2+b^2}.$$

You can watch some trajectories for 2D random walks here. In the applet you can change the width of the square in which you view the walk and also the number of walkers. Try running a simulation with 200 walkers on a square of width 20. Watcg the histograms of the x and y coordinates. See how a bell curve emerges for each. (If you were to let the simulation run long enough the bell curve will get wider and wider and then disappear, once the 200 walkers all got separated from one another.)

Proposition 2.7. *Let $\mathbf{X}_n$ be a 2D random walk. Then*

$$\lim_{n\to\infty} \mathbb{E}\left(e^{\frac{1}{\sqrt{n}}(t,s)\cdot\mathbf{X}_n}\right) = e^{\frac{1}{4}\left(t^2+s^2\right)}. \tag{2.6.1}$$

Remark 2.4. The *dot product* of two vectors is

$$(a,b)\cdot(c,d) = ac+bd.$$

The function

$$\Phi(t,s) = \mathbb{E}\left(e^{(t,s)\cdot\mathbf{X}_n}\right)$$

is the *moment generating function* of $\mathbf{X}_n$. Note that if $\mathbf{X}_n = (x_n, y_n)$ then

$$\left.\frac{\partial^k\partial^\ell}{\partial t^k \partial s^\ell}\Phi(t,s)\right|_{t=s=0} = \mathbb{E}\left(x_n^k y_n^\ell\right).$$

PROOF. First note that, because $\mathbf{S}_1$, ..., $\mathbf{S}_n$ are independent,

$$\begin{aligned}\mathbb{E}\left(e^{(t,s)\cdot\mathbf{X}_n}\right) &= \mathbb{E}\left(\prod_{j=1}^n e^{(t,s)\cdot\mathbf{S}_j}\right) \\ &= \prod_{j=1}^n \mathbb{E}\left(e^{(t,s)\cdot\mathbf{S}_j}\right) \\ &= \left[\tfrac{1}{4}\left(e^t+e^s+e^{-t}+e^{-s}\right)\right]^n \\ &= \left[\tfrac{\cosh t+\cosh s}{2}\right]^n.\end{aligned}$$

Thus

$$\begin{aligned}\mathbb{E}\left(e^{\frac{1}{\sqrt{n}}(t,s)\cdot\mathbf{X}_n}\right) &= \left[1+\tfrac{1}{4n}\left(t^2+s^2\right)+\mathrm{O}\left(\tfrac{t^4+s^4}{n^2}\right)\right]^n \\ &= e^{\frac{1}{4}\left(t^2+s^2\right)}\left(1+\mathrm{O}\left(\tfrac{t^4+s^4}{n}\right)\right).\end{aligned}$$

□

Problem 2.7. Show that

$$e^{\frac{\sigma}{2}t^2} = \frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty} e^{tx-\frac{1}{2\sigma}x^2}\,dx.$$

It follows from (2.6.3) and Problem 2.7 that

$$\lim_{n\to\infty}\mathbb{E}\left(e^{\frac{1}{\sqrt{n}}(t,s)\cdot\mathbf{X}_n}\right) = \frac{1}{\pi}\int_{\mathbb{R}^2} e^{tx+sy}e^{-x^2-y^2}\,dxdy.$$

In words this says that the "moment generating function of $\frac{1}{\sqrt{n}}\mathbf{X}_n$ converges as $n\to\infty$ to the moment generating function of (X,Y) where X and Y are independent normal random variables each with mean zero and variance $\frac{1}{2}$.

Problem 2.8. Show that

$$\frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty} xe^{-\frac{1}{2\sigma}x^2}\,dx = 0 \quad\text{and}\quad \frac{1}{\sqrt{2\pi\sigma}}\int_{-\infty}^{\infty} x^2e^{-\frac{1}{2\sigma}x^2}\,dx = \sigma.$$

Prop. 2.7 is the central limit theorem for a 2D random walk. It implies the following

Theorem 2.7. *Let $\mathbf{X}_n$ be a 2D random walk. Then for any $a<b$, $c<d\in\mathbb{R}$ we have*

$$\mathbb{P}\left(\mathbf{X}_n\in\sqrt{n}[a,b]\times[c,d]\right) \xrightarrow{n\to\infty} \frac{1}{\pi}\int_a^b\int_c^d e^{-x^2-y^2}\,dydx. \tag{2.6.2}$$

We won't give the proof of this theorem. You can give a combinatorial proof along the lines of what we did for a 1D random walk, but it is much more complicated. The better way to see it is to understand that Prop. 2.7 actually already implies (2.6.2). But this requires a bit of work as well.

As a consequence, we find that

$$\mathbb{P}\left(\mathbf{X}_{2n}=0\right) \sim \frac{1}{\pi n}.$$

(As in the 1D case $\mathbb{P}(\mathbf{X}_n=0)=0$ if n is odd.)

Problem 2.9. Show that $\sum_{n=1}^{\infty}\frac{1}{n}$ is divergent.

It follows from the arguments developed last time that:

Theorem 2.8. *Let $\mathbf{X}_n$ be a 2D random walk. Then with probability one $\{n : \mathbf{X}_n=0\}$is infinite. That is, the walk returns to the origin infinitely many times.*

Random Walk in 3D. Now let $\mathbf{X}_n$ be the trajectory of a random walk in *three* dimensions. So,

$$\mathbf{X}_n = \mathbf{S}_1+\cdots+\mathbf{S}_n$$

where $\mathbf{S}_1$, …, $\mathbf{S}_n$ are inedependent random vectors with

$$\mathbf{S}_j = \begin{cases} (1,0,0) & \text{with probability } \frac{1}{6} \\ (0,1,0) & \text{with probability } \frac{1}{6} \\ (0,0,1) & \text{with probability } \frac{1}{6} \\ (-1,0,0) & \text{with probability } \frac{1}{6} \\ (0,-1,0) & \text{with probability } \frac{1}{6} \\ (0,0,-1) & \text{with probability } \frac{1}{6} \end{cases}.$$

So$\mathbf{X}_n$ is a random point in $\mathbb{Z}^3 = \{(n,m,\ell) \ : \ n,m,\ell \in \mathbb{Z}\}$.

Problem 2.10. Show that $\mathbb{E}(X_n) = 0$ and $\mathbb{E}\left(|X_n|^2\right) = n$, where $|(a,b,c)| = \sqrt{a^2+b^2+c^2}$.

You can watch trajectories and histograms for 3D random walks here. Play with the applet. You can rotate the camera view to get an idea of the 3D picture. See the bell curves emerging.

Proposition 2.8. *Let $\mathbf{X}_n$ be a 3D random walk. Then*

$$\lim_{n\to\infty} \mathbb{E}\left(e^{\frac{1}{\sqrt{n}}(t,s,r)\cdot\mathbf{X}_n}\right) = e^{\frac{1}{6}\left(t^2+s^2+r^2\right)}. \tag{2.6.3}$$

Remark 2.5. The *dot product* is $(t,s,r)\cdot(x,y,z) = tx+sy+rz$. As above, the function

$$\Phi(t,s,r) = \mathbb{E}\left(e^{(t,s,r)\cdot\mathbf{X}_n}\right)$$

is the *moment generating function* of $\mathbf{X}_n$.

PROOF. First note that, because $\mathbf{S}_1$, …, $\mathbf{S}_n$ are independent,

$$\begin{aligned} \mathbb{E}\left(e^{(t,s,r)\cdot\mathbf{X}_n}\right) &= \mathbb{E}\left(\prod_{j=1}^{n} e^{(t,s,r)\cdot\mathbf{S}_j}\right) \\ &= \left[\tfrac{1}{6}\left(e^t+e^s+e^r+e^{-t}+e^{-s}+e^{-r}\right)\right]^n \\ &= \left[\tfrac{\cosh t+\cosh s+\cosh r}{6}\right]^n. \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{E}\left(e^{\frac{1}{\sqrt{n}}(t,s,r)\cdot\mathbf{X}_n}\right) &= \left[1+\tfrac{1}{6n}\left(t^2+s^2\right)+\mathrm{O}\left(\tfrac{t^4+s^4+r^4}{n^2}\right)\right]^n \\ &= e^{\frac{1}{6}\left(t^2+s^2+r^2\right)}\left(1+\mathrm{O}\left(\tfrac{t^4+s^4+r^4}{n}\right)\right). \end{aligned}$$

□

It follows from (2.6.3) and Exercise 2.7 that

$$\lim_{n\to\infty} \mathbb{E}\left(e^{\frac{1}{\sqrt{n}}(t,s,r)\cdot\mathbf{X}_n}\right) = \left(\frac{3}{2\pi}\right)^{\frac{3}{2}} \int_{\mathbb{R}^2} e^{tx+sy+rz} e^{-\frac{3}{2}\left(x^2+y^2+z^2\right)} dxdydz.$$

That is, the "moment generating function of $\frac{1}{\sqrt{n}}\mathbf{X}_n$ converges as $n\to\infty$ to the moment generating function of (X,Y,Z) where X, Y, and Z are independent normal random variables each with mean zero and variance $\frac{1}{3}$. As above, this implies

Theorem 2.9. *Let $\mathbf{X}_n$ be a 3D random walk. Then for any $a < b$, $c < d$, $e < f \in \mathbb{R}$ we have*

$$\mathbb{P}\left(\mathbf{X}_n \in \sqrt{n}[a,b] \times [c,d] \times [e,f]\right) \xrightarrow{n\to\infty} \left(\frac{3}{2\pi}\right)^{\frac{3}{2}} \int_a^b \int_c^d \int_e^f e^{-\frac{3}{2}\left(x^2+y^2+z^2\right)} dz dy dx.$$

As a consequence, we find that

$$\mathbb{P}(\mathbf{X}_{2n} = 0) \sim \frac{1}{\pi n^{3/2}}.$$

(As in the 1D case $\mathbb{P}(\mathbf{X}_n = 0) = 0$ if n is odd.)

Problem 2.11. Show that $\sum_{n=1}^{\infty} \frac{1}{n^{3/2}}$ is convergent.

Thus it does not follow that the 3D random walk returns to zero infinitely often. *In fact, it follows that the walk returns only finitely many times!!!!*

Theorem 2.10. *Let $\mathbf{X}_n$ be a 3D random walk. Then with probability one $\{n : \mathbf{X}_n = 0\}$ is finite. That is, the walk returns to the origin only finitely many times.*

To see this we need the following

Lemma 2.4 (Borel-Cantelli)**.** *Let $E_1, \ldots$ be a sequence of events. If*

$$\sum_{j=1}^{\infty} \mathbb{P}\left(E_j\right) < \infty$$

then with probability one only finitely many of the E_j occur.

Remark 2.6. This result doesn't say anything interesting about a finite probability space. The proof we give is clear but a little dishonest because we haven't yet defined what we mean by a probability measure on the spaces where we need to apply the result. That is coming next time.

PROOF. The event A that *infinitely* many E_j occur is

$$A = \bigcap_{j=1}^{\infty} \bigcup_{k=j}^{\infty} E_k. \tag{2.6.4}$$

Indeed by definition

$$A = \left\{\omega : \omega \in E_j \text{ for infinitely many } j\right\}.$$

That is, if $\omega \in A$ then there is a sequence $j_k \to \infty$ such that $\omega \in E_{j_k}$for all k. Since $j_k \to \infty$ we see that $\omega \in A_N$ for each N where

$$A_N = \bigcup_{k=N}^{\infty} E_k.$$

Thus $\omega \in \cap_N A_N$ which is the set on the r.h.s. of (2.6.4). Conversely, if $\omega \in A_N$ then there exists $k_N \geq N$ such that $\omega \in E_k$. If this holds for each N then $k_N \to \infty$, so we have ω in infinitely many E_k.

We see that $A \subset A_N$, so

$$\mathbb{P}(A) \leq \mathbb{P}(A_N) \leq \sum_{k=N}^{\infty} \mathbb{P}(E_k)$$

since the sum on the r.h.s. neglects the overlap between the sets E_k. (See the inclusion exclusion principle in Lecture 2.) However, since $\sum_n \mathbb{P}(E_n)$ converges we have

$$\lim_{N\to\infty} \sum_{k=N}^{\infty} \mathbb{P}(E_k) = 0.$$

Thus $\mathbb{P}(A) = 0$, so $\mathbb{P}(A^c) = 1 - \mathbb{P}(A) = 0$, where A^c is the complementary event to A, namely the event that only finitely many E_j ocur:

$$A^c = \bigcup_{j=1}^{\infty} \bigcap_{k=j}^{\infty} E_k^c.$$

□

PROOF. (Proof of Theorem 2.10) Let E_j =event that $\mathbf{X}_j = 0$. Then

$$\sum_{j=1}^{\infty} \mathbb{P}\left(E_j\right) < \infty$$

since $\mathbb{P}\left(E_j\right) \sim C j^{-3/2}$. By Borel-Cantelli, with probability one only finitely many E_j occur. □

2.7. Lecture 12: A bit of measure theory

In the last lecture we saw that a 3D random walk $\mathbf{X}_n$ has the property that

$$\mathbb{P}\left(\mathbf{X}_n = 0 \text{ for infinitely many } n\right) = 0. \tag{2.7.1}$$

The proof rested on the Borel Cantelli lemma and the fact that

$$\mathbb{P}\left(\mathbf{X}_n = 0\right) \sim \frac{C}{n^{3/2}},$$

so

$$\sum_{n=1}^{\infty} \mathbb{P}\left(\mathbf{X}_n = 0\right) < \infty.$$

The property (2.7.1) for the 3D random walk is called *transience*. This is as opposed to *recurrence* of the random walk in 1D and 2D, for which the corresponding probability is 1.

The step we took to consider the recurrence and transience of random walks marked a shift in our arguments. Previously everything we have done could be formulated in the context of finite probability spaces. But when faced with the very natural question "how many times does the random walk X_n in 1D, say, visit the origin?", we are led to consider the event that it visits infinitely many times. This event

$$I = \bigcap_{N=1}^{\infty} \bigcup_{n\geq N} \{X_n = 0\}$$

does not lie in the sample space of the random walk up to any finite time. Instead we are led to consider the sample space

$$\Omega = \{-1,1\}^{\mathbb{N}}, \tag{2.7.2}$$

which is the set of all sequence $(S_1, S_2, \ldots)$ with $S_j = \pm 1$ for each j.

Remark 2.7. Recall that for sets A and B,

$$A^B = \{f : B \to A\},$$

the set of functions from A to B. A sequence can be thought of as a function with domain $\mathbb{N}$.

Problem 2.12. Show that, for finite sets A, B,

$$\left|A^B\right| = |A|^{|B|},$$

where $|C|$ =number of elements in C.

The sample space (2.7.2) describes all possible steps that a random walk can take up to infinite time. Given a point $\omega = (\sigma_1, \cdots) \in \Omega$ we can express the position of the random walk after the n-th step as

$$X_n = X_n(\omega) = \sigma_1 + \cdots + \sigma_n.$$

That is, we can think of X_n as a map from Ω to $\mathbb{R}$, namely, as a "random variable." Similarly, we can define the random variables S_j with

$$S_j(\omega) = \sigma_j,$$

so that S_j is the j-th step of the walk.

But what is the probability measure on Ω? Recall that for finite spaces we used the notion:

$$\mathbb{P}(A) = \sum_{\omega \in A} p(\omega). \tag{2.7.3}$$

However this does not work here, because the probability of each individual outcome is zero! Indeed given a sequence σ_j of ± 1's, the probability

$$\mathbb{P}\left(S_j = \sigma_j \text{ for all } j\right)$$

should be less than the probability that

$$\mathbb{P}\left(S_j = \sigma_j \text{ for } j = 1, \ldots, n\right) = \frac{1}{2^n} \xrightarrow{n \to \infty} 0.$$

Thus (2.7.3) will not work for us. (This is analogous to the fact that integrals are like sums, but not the same as sums.)

What we will do to proceed is throw out (2.7.3) and work only with the map $\mathbb{P} : \mathscr{F} \to [0,1]$ where $\mathscr{F}$ is a collection of events that we assign probabilities to.

Probability Measures. What probabilities should the map $\mathbb{P}$ and the collection of events $\mathscr{F}$ have?

(1) We want $\Omega \in \mathscr{F}$ and $\emptyset \in \mathscr{F}$ and we want

$$\mathbb{P}(\Omega) = 1 \quad \text{and} \quad \mathbb{P}(\emptyset) = 0.$$

(2) We want $A^c \in \mathscr{F}$ if $A \in \mathscr{F}$, since if we are going to consider an event we should also consider the possibility that the event does not occur. Furthermore, we want

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A).$$

(3) We want $\mathbb{P}(A) \leq \mathbb{P}(B)$ if $A \subset B$ and $A, B \in \mathscr{F}$. This point is basic to our reasoning so far. If a larger collection of outcomes give rise to an event that should increase it's probability.

(4) We certainly want unions and intersections of events to be allowed. That is if $A, B \in \mathscr{F}$ then we want

$$A \cup B \text{ and } A \cap B \in \mathscr{F}.$$

(5) We want inclusion exclusion to hold
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B),$$
whenever $A, B \in \mathscr{F}$. Equivalently, we want
$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$$
if $A, B \in \mathscr{F}$ and $A \cap B = \emptyset$.

(6) In fact, we want something more than 5. In the proof of Borel-Cantelli we need to know that
$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(E_n). \tag{2.7.4}$$
In fact, we expect that if the then
$$\text{If } E_n \text{ are pairwise disjoint, then } \mathbb{P}\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(E_n). \tag{2.7.5}$$
In fact, (2.7.5) is *two* conditions:
(a) If $E_1, \ldots,$ is a sequence in $\mathscr{F}$ then $\bigcup_{n=1}^{\infty} E_n \in \mathscr{F}$.
(b) Eq. (2.7.5) for any sequence $E_1, \ldots$ in $\mathscr{F}$.

Problem 2.13. Show that (2.7.5) implies 3 and 5 by applying it to sequences $A, B, \emptyset, \emptyset, \ldots$.

Problem 2.14. Show that (2.7.4) follows from (2.7.5) by applying (2.7.5) to the sequence
$$\widetilde{E}_j = E_j \setminus \left(\bigcup_{k=1}^{j-1} E_k\right).$$

Definition 2.3. Let Ω be a set, called the sample space. A collection $\mathscr{F}$ of subsets of Ω is a *field* if
(1) $\Omega, \emptyset \in \mathscr{F}$,
(2) $A \in \mathscr{F} \implies A^c \in \mathscr{F}$, and
(3) $A, B \in \mathscr{F} \implies A \cup B \in \mathscr{F}$.

The collection $\mathscr{F}$ is called a σ-*field* if, in place of 3. we assume
3'. If $E_1, E_2, \ldots$ is a sequence in $\mathscr{F}$ then $\bigcup_{n=1}^{\infty} E_n \in \mathscr{F}$.

A field need not be a σ-field. Furthermore, we see that we are naturally interested in σ-fields since we need to take countable unions and countable intersections to get at results like transience or recurrence.

Problem 2.15. Let $\mathscr{F}$ be a σ-field. Show that if $E_1, E_2, \ldots$ is a sequence in $\mathscr{F}$ then $\bigcap_{n=1}^{\infty} E_n \in \mathscr{F}$.

Definition 2.4. Let Ω be a sample space and $\mathscr{F}$ a sigma field on Ω. A *probability measure* on $(\Omega, \mathscr{F})$ is a map $\mathbb{P} : \mathscr{F} \to [0,1]$ such that
(1) $\mathbb{P}(\Omega) = 1$
(2) If $E_1, E_2, \ldots$ is a sequence of pairwise disjoint elements of $\mathscr{F}$ then
$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} E_j\right) = \sum_{j=1}^{\infty} \mathbb{P}(E_j).$$

Given a sigma field $\mathscr{F}$ and a probability measure $\mathbb{P}$ we will call the elements of $\mathscr{F}$ *events.*

Problem 2.16. Let $\mathscr{F}$ be a sigma field and let $\mathbb{P}$ be a probability measure on $\mathscr{F}$. Show that

(1) $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ for every $A \in \mathscr{F}$.
(2) If $E_1, E_2, \ldots$ is a sequence of events then

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} E_n\right) = \lim_{N\to\infty} \mathbb{P}\left(\bigcap_{n=1}^{N} E_n\right).$$

The sigma-Field for the 1D random walk. To use the sample space (2.7.2) to understand random walks, it is not be possible to takethe sigma field $\mathscr{F}$ to be the collection of all subsets of Ω. We will not prove this fact here, as it is technical to see. In any case, we will take a slightly different perspective. Namely, we note that what we *really* care about is understanding the random walk after a finite number of steps and certain events that pertain to limits as $n \to \infty$. We don't really care about arbitrary subsets of $\{-1,1\}^{\mathbb{N}}$, but rather those which are "almost" specified by events that depend on only finitely many of the steps. For the moment this is a bit vague, but hopefully it will seem more precise in a bit.

What kind of events do we want to understand?

(1) Certainly the event $\{S_1 = 1\}$ or $\{S_{2345} = -1\}$ or similar such things, namely

$$\{S_j = 1\} \text{ and } \{S_j = -1\}$$

for each j.
(2) Everything that can be built up from those events by countable unions or intersections or complements.

What is the event $\{S_j = 1\}$? In terms of the sample space $\{-1,1\}^{\mathbb{N}}$ it is the set

$$\{\omega = (\sigma_1, \ldots) \ : \ \sigma_j = 1\}.$$

Namely we can think of it as the cartesian product

$$\{S_j = 1\} = \underbrace{\{-1,1\} \times \cdots \times \{-1,1\}}_{j-1 \text{ factors}} \times \{1\} \times \underbrace{\{-1,1\} \times \{-1,1\} \times \cdots}_{\text{countably many factors}}.$$

More generally we make the following

Definition 2.5. Given $E_1, \ldots, E_N \subset \{-1,1\}$ the *cylinder set over* $E_1 \times \cdots \times E_N$ *is*

$$\mathscr{C}(E_1, \ldots, E_N) := \{\omega = (\sigma_1, \ldots) \ : \ \sigma_j \in E_j \text{ for } j = 1, \ldots, N\}.$$

Remark 2.8. In other words we have

$$\mathscr{C}(E_1, \ldots, E_N) = E_1 \times \cdots \times E_N \times \underbrace{\{-1,1\} \times \{-1,1\} \times \cdots}_{\text{countably many factors}}.$$

Let $\mathscr{F}_0 =$ collection of finite unions of cylinder sets.

Problem 2.17. Show that $\mathscr{F}_0$ is a field, but not a sigma-field.

Anything we event that pertains to the random walker during the first N steps can be expressed in terms of cylinder sets. Thus $\mathscr{F}_0$ contains most of the events in which we are really interested.

Problem 2.18. Try writing $\{X_4 = 0\}$ as a union of cylinder sets. (Hint: you should need 8 sets.)

Definition 2.6. We define the probability of a cylinder set $\mathscr{C}(E_1, \ldots, E_N)$ to be

$$\mathbb{P}(\mathscr{C}(E_1, \ldots, E_N)) = \prod_{n=1}^{N} \frac{|E_n|}{2}. \tag{2.7.6}$$

This definition extends to all of $\mathscr{F}_0$, but does not yet define a probability measure, as we do not have a sigma-field. Recalling the philosophy that we are essentially interested in the walk after a finite number of steps, so events in $\mathscr{F}_0$, we let

$$\mathscr{F} = \text{ the smallest sigma field that contains } \mathscr{F}_0.$$

More precisely,

$$\mathscr{F} = \bigcap \{\mathscr{G} \ : \ \mathscr{G} \text{ is a sigma field and } \mathscr{G} \supset \mathscr{F}_0\}. \tag{2.7.7}$$

For this to make sense we need two things:

(1) The intersection of a collection of sigma fields is a sigma field.
(2) The set on the r.h.s. of (2.7.7) is non empty.

Problem 2.19. Show that the intersection of an arbitrary non-empty collection of sigma fields is a sigma field. (This should be VERY easy. It might make you a little queasy because it is so abstract, but there is very very little to show.)

This takes care of 1. To take care of 2. we note that the power set $2^\Omega = \{A \subset \Omega\}$ is itself a sigma-field.

Theorem 2.11. *There exists a unique probability measure $\mathbb{P}$ on $\mathscr{F}$ defined as above such that* (2.7.6) *holds for all cylinder sets.*

The proof of this result goes beyond what we cover in this course. However, we don't need to know the proof to compute with $\mathbb{P}$. Indeed once we know that $\mathbb{P}$ exists we can compute things like

$$\mathbb{P}(X_n = 0 \text{ for infinitely many } n)$$

by simply noting that the event

$$A = \{X_n = 0 \text{ for infinitely many } n\}$$

can be expressed as

$$A = \bigcap_{N=1}^{\infty} \bigcup_{n \geq N} \{X_n = 0\}.$$

Problem 2.20. Show, for each n, that

$$\{X_n = 0\} \in \mathscr{F}_0.$$

Hint: it is either empty (if n is odd) or it can be expressed in terms of a union $\binom{n}{\frac{n}{2}}$ cylinder sets (if n is even).

Thus $A \in \mathscr{F}$ and

$$\mathbb{P}(A) = \lim_{N \to \infty} \mathbb{P}\left(\bigcup_{n \geq N} \{X_n = 0\} \right)$$

and we can proceed as in the last two lectures.

2.8. Lecture 13: Gambler's Ruin

Let us return to the 1D random walk and ask some more questions. For instance take any site $a \in \mathbb{Z}$. What is the probability that a random walker starting at 0 ever hits a? It should be clear that once the walker lands on a he will return there infinitely often. (To see this, think of the random walk from the first time he hits a onward as a new random walk starting at a. In lecture 10 we showed that this walk will hit it's starting point infinitely often.)

So, what is the probability that the walker ever gets to a? Really the only answer consistent with the facts that

(1) The walk is unbounded, and
(2) The walker returns to the origin infintely often

is that the walk hits a with probability one. But how do we prove this?

Before answering that question, we consider another question which is a little easier to answer. Namely, suppose the walker starts at a point a, with $0 < a < N$ where N is a large number. What is the probability that the walker reaches N before he reaches 0?

First of all we should show that the walker really does reach 0 or N at some point.

Proposition 2.9. *Let X_n be a 1D random walk starting at a, with $0 < a < N$. Then*

$$\mathbb{P}(X_n = 0 \text{ or } X_n = N \text{ for some } n \geq 1) = 1.$$

PROOF. Let $E = \{X_n = 0 \text{ or } X_n = N \text{ for some } n \geq 1\}$. Then E^c is the event

$$E^c = \{0 < X_n < N \text{ for all } n \geq 1\},$$

since X_n cannot leave the set $\{1, \ldots, N-1\}$ without reaching either 0 or N. Note that E^c is the countable intersection:

$$E^c = \bigcap_{n=1}^{\infty} \{0 < X_n < N\}.$$

Thus, for any $n \geq 1$,

$$\mathbb{P}(E^c) \leq \mathbb{P}(0 < X_n < N).$$

However, by the central limit theorem we have

$$\mathbb{P}(0 < X_n < N) \sim \frac{1}{\sqrt{2\pi}} \int_{-\frac{a}{\sqrt{n}}}^{\frac{N-a}{\sqrt{n}}} e^{-\frac{1}{2}x^2} dx \sim \frac{N}{\sqrt{2\pi n}} \to 0$$

as $n \to \infty$. Thus $\mathbb{P}(E^c) = 0$ and so $\mathbb{P}(E) = 1$. □

Based on the propostion, it makes sense to define the following random variable:

$$T = \min\{n \,:\, X_n = 0 \text{ or } X_n = N\}.$$

(This random variable is an example of a "stopping time." I mention this only for those who might have come across this term before.) Now T can in principle be very very large. Indeed there is a small, but positive probability, that the random walk bounces back and forth between a and $a+1$ two million times before wandering off to hit 0 or N. Thus the probability that T is bigger than two million is positive. Indeed, we have

$$T = \sum_{n=1}^{\infty} n I[A_n]$$

where A_n is the event that $T = n$, namely

$$A_n = \{X_n = 0 \text{ or } X_n = N\} \cap \{0 < X_n < N \text{ for all } m = 0, 1, \ldots, n-1\}.$$

(Notice that this event is in the field $\mathscr{F}_0$ that was defined last time.) Thus to determine the value of T we might have to let the walk run for a very very long time. However, if we want to ask "is $T \leq n$?" then we need only follow the walk up to the n-th step.

Using T we may restate our question about the walk as

$$\text{What is } \mathbb{P}(X_T = N | X_0 = a)?$$

(Here we have used the notation $\mathbb{P}(\cdot | X_0 = a)$ to denote, not a conditional probability, but rather the probability that something happens when we start the walk at a.)

Gambler's Ruin. There is a good interpretation of the above question in terms of a game. Suppose you have a dollars that you decide to gamble at a table in a casino. In each round you win a dollar or lose a dollar, with probabililty $\frac{1}{2}$ in each case. You decide to play until you have N dollars or you run out of money. (Maybe the bank at the casino has only $N-a$ dollars so if you win this much you "break the bank.") Thus the question we are asking is "what are your chances of winning?"

To solve the problem let us think of the probability as a function of a,

$$F(a) = \mathbb{P}(X_T = N | X_0 = a).$$

So $F : \{0, \ldots, N\} \to [0,1]$. Furthermore, we have the "boundary conditions"

$$F(0) = 0 \quad \text{and} \quad F(N) = 1, \tag{2.8.1}$$

since if our initial fortune is 0 dollars we can't play and if we start with N dollars then we already have our desired limit. (Or the Casino's bank is already broken.)

To determine F at other points we proceed as follows. Suppose we start with a dollars, where $0 < a < N$. If we play one round of the game then we end up with either $a+1$ or $a-1$ dollars. But now our situation is just as if we walked up to the table with the new sum of money. Thus

$$F(a) = \frac{1}{2}F(a+1) + \frac{1}{2}F(a-1). \tag{2.8.2}$$

We can use (2.9.1) to solve for F as follows. First rewrite the equation as:

$$F(a+1) = 2F(a) - F(a-1) \tag{2.8.3}$$

Start with $a = 1$. Then we learn that

$$F(2) = 2F(1),$$

since $F(0) = 0$ by (2.8.1). We don't know $F(1)$, but don't worry about that. At the next step we learn that

$$F(3) = 2F(2) - F(1) = 4F(1) - F(1) = 3F(1).$$

There is a clear pattern emerging here! Let's compute

$$F(4) = 2F(3) - F(2) = 6F(1) - 2F(1) = 4F(1).$$

All this suggests that

$$F(a) = aF(1)$$

for all a. Indded, it is easy to see that $F(a) = Ca$ satisfies (2.9.1) for any C. But how do we determine $F(1)$? From the other boundary condition! We know that

$$1 = F(N) = NF(1),$$

so

$$F(1) = \frac{1}{N}.$$

Thus the answer to our question seems to be

$$\mathbb{P}(X_T = N | X_0 = a\,0) = \frac{a}{N}.$$

In fact, we didn't really show this yet because we need to show that the solution we obtained is *unique.* That is is follows from the following

Theorem 2.12. *Let α and β be given. There is a* unique *function $F : \{0, \ldots, N\} \to \mathbb{R}$ which satisfies* (2.9.1) *and also $F(0) = \alpha$ and $F(N) = \beta$. The function is*

$$F(a) = \alpha + \frac{(\beta - \alpha)}{N} a. \tag{2.8.4}$$

Problem 2.21. Show that F given by (2.9.2) does indeed satsify (2.9.1) and that $F(0) = \alpha$ and $F(N) = \beta$.

PROOF. Using (2.9.9) we see that for each $j = 1, \ldots, N-1$ we have

$$\begin{pmatrix} F(j+1) \\ F(j) \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F(j) \\ F(j-1) \end{pmatrix}.$$

Thus

$$\begin{pmatrix} F(a) \\ F(a-1) \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}^{a-1} \begin{pmatrix} F(1) \\ F(0) \end{pmatrix}.$$

To raise a matrix to a power, it is very convient to diagonalize it. Actually it is not possible to diagonalize $\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$. (Not all matrices are diagonalizable!) Indeed, if we look at the characteristic polynomial,

$$\det\left(\lambda \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}\right) = \lambda^2 - 2\lambda + 1 = (\lambda - 1)^2.$$

So the only eigenvalue of $\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$ is 1. Thus if the matrix were diagonalizable it would be the identity matrix (which it is not!) However, it does have an eigenvalue. (Every matrix has an eigenvalue!) What is the eigenvector? Looking at (2.9.1) we can guess. Notice that the function $F(a) \equiv 1$ satsifies (2.9.1). Thus $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ should be an eigenvector of $\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$. Indeed,

$$\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The vector $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is orthogonal to $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and we have

$$\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

Thus

$$\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}. \tag{2.8.5}$$

Problem 2.22. Show that

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

From the problem and (A.3.4) it follows that

$$\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix},$$

and that

$$\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}^n = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}^n\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Problem 2.23. Let a,b be real numbers. Show that

$$\begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a+b \\ 0 & 1 \end{pmatrix}.$$

Thus

$$\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}^n = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 2n \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} n+1 & -n \\ n & 1-n \end{pmatrix},$$

so

$$\begin{pmatrix} F(a) \\ F(a-1) \end{pmatrix} = \begin{pmatrix} a & 1-a \\ a-1 & 2-a \end{pmatrix}\begin{pmatrix} F(1) \\ F(0) \end{pmatrix}.$$

We conclude that

$$F(a) = aF(1) + (1-a)F(0) = \alpha + (F(1)-\alpha)\,a.$$

To compute $F(1)$, we plug in $\beta = F(N)$ to get

$$\beta = \alpha + (F(1)-\alpha)N,$$

so

$$F(1) = \alpha + \frac{\beta-\alpha}{N}$$

and the result follows. □

2.9. Lecture 14: Probability to hit a given point

Last time we proved the following theorem:

Theorem 2.13. *Let α and β be given real numbers. There is a* unique *function $F : \{0,\dots,N\} \to \mathbb{R}$ which satisfies*

$$F(a) = \frac{1}{2}F(a+1) + \frac{1}{2}F(a-1). \tag{2.9.1}$$

and also $F(0) = \alpha$ and $F(N) = \beta$. The function is

$$F(a) = \alpha + \frac{(\beta-\alpha)}{N}a. \tag{2.9.2}$$

Our main interest in the theorem was to show that

$$\mathbb{P}(X_T = N \,|X_0 = a) = \frac{a}{N}$$

where X_n is a random walk (in 1D) starting at a and T is the first step n for which $X_n = 0$ or N, i.e.,

$$T = \min\{n \,:\, X_n = 0 \text{ or } X_n = N\}.$$

Last time we gave a perfectly decent proof of this theorem using a little linear algebra. The point is that (2.9.1) shows that if we know F at two neighboring points we can see what F is anywhere else. This was conveniently expressed through the identity

$$\begin{pmatrix} F(j+1) \\ F(j) \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} F(j) \\ F(j-1) \end{pmatrix}, \tag{2.9.3}$$

which follows from (2.9.1). Repeated applications of (2.9.3) show that

$$\begin{pmatrix} F(j+m) \\ F(j+m-1) \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}^m \begin{pmatrix} F(j) \\ F(j-1) \end{pmatrix}. \tag{2.9.4}$$

In our proof of Thm. 2.13 we obtained the identity

$$\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}^m = \begin{pmatrix} m+1 & -m \\ m & 1-m \end{pmatrix}. \tag{2.9.5}$$

The whole situation is quite analogous to the following. Suppose $F : [0,1] \to \mathbb{R}$ is a continuous function that is twice differentiable in $(0,1)$, with continuous derivatives. If

$$F'' = 0$$

then $F(x) = ax+b$, so F is determined everywhere once we pin F and it's derivative down at a point. Indeed you can easily prove the following

Theorem 2.14. *Let α and β be given real numbers. There is a* unique *function $F : [0,1] \to \mathbb{R}$ which satisfies*

(1) $F \in C^2[0,1]$,
(2) $F''(x) = 0$ *for all x,*
(3) $F(0) = \alpha$, *and* $F(1) = \beta$.

The function is

$$F(x) = \alpha + (\beta - \alpha)x.$$

To understand the relation between Thms. 2.13 and 2.14, note that eq. (2.9.1) can be written as

$$\frac{1}{2}\left((F(a+1) - F(a)) - (F(a) - F(a-1))\right) = 0. \tag{2.9.6}$$

The difference $dF(a+1,a) = F(a+1) - F(a)$ can be thought of as the discrete derivative of F along the bond $(a, a+1)$. Thus (2.9.6) is analogous to the statement that the derivative of the derivative vanishes.

Another proof. There is nothing wrong with the proof of Thm. 2.13 we gave last time. However, it did hinge upon a a particular matrix identity (eq. (2.9.5)) in such a way that it might seem the result is an "accident" of sorts. Today we will look at another proof, which uses the relationship of (2.9.1) with the random walk in a decisive way.

It is always interesting to see a number of proofs of a good theorem (even a simple theorem like Thm. 2.13). Each proof gives a particular picture of "why" the result is true. Knowing several truly different proofs enhances ones understanding. In particular, different proofs suggest different types of generalizations.

PROOF. *(Proof of Theorem 2.13)* Consider a random walk X_n that starts at a point $a \in \{0, \ldots, N\}$. Let T be defined as above

$$T = \min\{n \,:\, X_n = 0 \text{ or } X_n = N\}.$$

Now consider the random walk

$$Y_n = X_{n \wedge T},$$

where

$$n \wedge T = \min\{n, T\}.$$

So Y_n is a random walk that looks just like X_n until it reaches the boundary $\{0, N\}$ of the interval $\{0, \ldots, N\}$ when it gets "stuck" to the boundary. Once $Y_n = 0$ or $Y_n = N$ the walk never moves again.

Now let F be a function that satisfies (2.9.1) and let

$$G_n(j) = \mathbb{E}(F(Y_n) | Y_0 = j)$$

Clearly

$$G_0(j) = F(j) \quad \text{for all } j = 0, \ldots, N.$$

I claim that $G_n(j) = F(j)$ for all $n = 1, \ldots$ as well. To prove this it suffices to show

$$G_{n+1}(j) = G_n(j) \tag{2.9.7}$$

for $n = 0, 1, \ldots$.

To show (2.9.7) note that the events $\{Y_n = k\}$, $k = 0, \ldots, N$, form a partition — they are mutually exclusive and their union is the whole sample space. Thus,

$$F(Y_{n+1}) = F(Y_{n+1}) \sum_{k=0}^{N} I[Y_n = k].$$

Therefore

$$\mathbb{E}(F(Y_{n+1}) | Y_0 = j) = \sum_{k=0}^{N} \mathbb{E}(F(Y_{n+1}) | Y_n = k) \mathbb{P}(Y_n = k | Y_0 = j).$$

If $0 < k < N$ then

$$\mathbb{E}(F(Y_{n+1}) | Y_n = k) = \frac{1}{2}(F(k+1) + F(k-1)) = F(k)$$

by (2.9.1). On the other hand, if $k = 0$ or N and $Y_n = k$, then $Y_{n+1} = k$, so

$$\mathbb{E}(F(Y_{n+1}) | Y_n = k) = F(k) \quad k = 0 \text{ or } N.$$

Thus

$$\begin{aligned} G_{n+1}(j) = \mathbb{E}(F(Y_{n+1}) | Y_0 = j) &= \sum_{k=0}^{N} F(k) \mathbb{P}(Y_n = k | Y_0 = j) \\ &= \mathbb{E}(F(Y_n) | Y_0 = j) = G_n(j), \end{aligned}$$

which is (2.9.7).

Thus $F(j) = G_n(j)$ for all n. However, as $n \to \infty$ we have

$$\mathbb{P}(Y_n = k | Y_0 = j) \xrightarrow{n \to \infty} 0$$

for $k \in \{1, \ldots, N\}$, while

$$\mathbb{P}(Y_n = N | Y_0 = j) \xrightarrow{n \to \infty} \mathbb{P}(X_T = N | Y_0 = j)$$

and

$$\mathbb{P}(Y_n = 0 | Y_0 = j) \xrightarrow{n \to \infty} 1 - \mathbb{P}(X_T = N | Y_0 = j).$$

Thus

$$F(j) = G_n(j) = \sum_{k=0}^{N} F(k)\mathbb{P}(Y_n = k | Y_0 = j) \xrightarrow{n\to\infty} F(0) + (F(N) - F(0))\,\mathbb{P}(X_T = N | Y_0 = j),$$

We have shown that any function F which satisfies (2.9.1) with $F(0) = \alpha$ and $F(N) = \beta$ must equal

$$F(j) = \alpha + (\beta - \alpha)\mathbb{P}(X_T = N | Y_0 = j).$$

This shows the uniqueness of the solution. To see that the solution is given also by (2.9.2) we simply need to show that the function defined in (2.9.2) is indeed a solution. This is quite easy, and is left as an exercise □

Probability to hit a point. Let us return to the original question which motivated us last time: what is the probability that a random walk starting at 0 hits a particular point a? Without loss of generality we may take $a > 0$ (since the answer is clearly symmetric under $a \mapsto -a$).

Given $a > 0$, we may ask the more general question: what is the probability that a random walk starting at a point $j \le a$ eventually hits a? To find the answer, let us define

$$F(j) = \mathbb{P}(X_n \text{ never hits } a | X_0 = j)$$

Clearly

$$F(a) = 0. \tag{2.9.8}$$

Furthermore if $X_0 = j < a$ then the random walk first takes a step ending up with $X_1 = j+1$ or $j-1$ each with probability $\frac{1}{2}$. From there on the walk runs independently of this first step. It follows that

$$F(j) = \frac{1}{2}(F(j+1) + F(j-1)), \quad j < a. \tag{2.9.9}$$

That is F satisfies (2.9.1) on the set $\{a, a-1, a-2, \ldots\}$.

Trouble is, (2.9.9) and (2.9.8) do not together pin down F. Indeed

$$F(j) = \alpha(j - a)$$

solves these equations for any number α. However only one of these solutions can work! The point is we also know that $F(j)$, being a probability, is between 0 and 1. In particular, F is a bounded function. This suggests that $F(j) = 0$ for all j. That this is so follows from the following

Theorem 2.15. *Let $F : \mathbb{N} \to \mathbb{R}$ satisfy (2.9.9) for $j \ge 1$. If F is bounded then*

$$F(j) = F(0)$$

for all j.

PROOF. Using the linear algebra approach from the last lecture, we see from (2.9.4) and (2.9.5) that

$$\begin{pmatrix} F(j) \\ F(j-1) \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}^{j-1} \begin{pmatrix} F(1) \\ F(0) \end{pmatrix} = \begin{pmatrix} j & 1-j \\ j-1 & 2-j \end{pmatrix} \begin{pmatrix} F(1) \\ F(0) \end{pmatrix},$$

for $j \ge 1$. Thus

$$F(j) = F(0) + (F(1) - F(0))\,j.$$

Since F is bounded, we must have $F(1) = F(0)$ and thus $F(j) = F(0)$ for all j. □

Returning to the probability that the random walk hits a point $a > 0$, we see that this probability is one as follows. Let

$$F(j) = \mathbb{P}(X_n \text{ never hits } a \,|X_0 = a - j).$$

Then F satisfies the hypotheses of the theorem, so

$$F(j) = F(0) = \mathbb{P}(X_n \text{ never hits } a \,|X_0 = a) = 0$$

In particular

$$\mathbb{P}(X_n \text{ never hits } a \,|X_0 = 0) = F(a) = 0.$$

Thus with probability one a random walk starting at 0 hits the point a.

In fact $X_n = a$ for infinitely many n, since once the random walk gets to a it starts afresh and returns to this point infinitely often. (See lecture 10.) With a little care one can use this to show that with probability a random walk (in 1D) hits every integer infinitely often!

2.10. Lecture 15: Dirichlet Problem for the Discrete Lapalacian

In this lecture we will consider analogues of the Gambler's Ruin problem for higher dimensional random walks. The idea is to look at a finite set A in the lattice $\mathbb{Z}^d$ and ask questions about when and where a random walk that starts at $x \in A$ exits the set. (Here $d = 1, 2, 3$ or a larger integer. Random walks on higher dimensional lattices are defined in much the same way. A central limit theorem holds in those cases as well.)

Discrete Laplacian. If $F : \mathbb{Z}^d \to \mathbb{R}$ is a real valued function we define

$$QF(x) = \frac{1}{2d} \sum_{\substack{|y-x|=1 \\ y \in \mathbb{Z}^d}} F(y)$$

and

$$\mathscr{L}F(x) = QF(x) - F(x) = \frac{1}{2d} \sum_{\substack{|y-x|=1 \\ y \in \mathbb{Z}^d}} (F(y) - F(x)).$$

Q and $\mathscr{L}$ are examples of "linear operators," that is maps on a space of functions that are linear. For example,

$$[\mathscr{L}(F + cG)](x) = \mathscr{L}F(x) + c\mathscr{L}G(x).$$

The operator $\mathscr{L}$ is called the *discrete Laplacian.* (A more proper notation for $\mathscr{L}F$ would be $\mathscr{L}(F)$, which emphasizes that $\mathscr{L}$ is a function on the space of functions $\{F : \mathbb{Z}^d \to \mathbb{R}\}$. Because this notation is awkward when we want to express $\mathscr{L}(F)$ evaluated at a point x – that is $\mathscr{L}(F)(x)$ –, we use the usual convention of linear algebra in which an operator "actx" on the object immediately to its right.)

The operators $\mathscr{L}$ and Q are very useful for understanding the random walk (and vice-a-versa!) The following proposition shows some of the relationship. It's proof is left as an exercise.

Proposition 2.10. *Let X_n be a simple random walk in $\mathbb{Z}^d$. Then*

(1) $\mathscr{L}F(x) = \mathbb{E}(F(X_1) - F(X_0) \,|X_0 = x)$.

(2) $Q^n\delta_0(x) = \mathbb{P}(X_n = x \,|X_0 = x)$, *where* $\delta_0(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$.

Remark 2.9. The notation Q^n indicates Q composed with itself n times: $\underbrace{Q \circ Q \circ \cdots \circ Q}_{n \text{ times}}$.

Dirichlet Problem. Given a finite subset $A \subset \mathbb{Z}^d$, we define the *outer boundary* of A to be the set

$$\partial^+ A := \left\{ x \in \mathbb{Z}^d \ : \ \operatorname{dist}(x,A) = 1 \right\}.$$

So if $A = \{(0,0)\} \subset \mathbb{Z}^2$ then

$$\partial^+ A := \{(1,0),(0,1),(-1,0),(0,-1)\}.$$

The *Dirichlet problem* on A is the following:

Problem 2.24. Given a function $g : \partial^+ A \to \mathbb{R}$ find a function $F : A \cup \partial^+ A \to \mathbb{R}$ such that

(1) $\mathscr{L}F(x) = 0$, $x \in A$, and
(2) $F(x) = g(x)$, $x \in \partial^+ A$.

A function that satisfies $\mathscr{L}F(x) = 0$ is said to be *harmonic* at x. Thus the function we seek in the Dirichlet problem is *harmonic* in A and agrees with $g \in \partial^+ A$. Note that to compute $\mathscr{L}F(x)$ at a point $x \in A$ "at the edge," that is at distance one from $\partial^+ A$, we need to know g.

Example 2.3. Suppose we want to solve the Dirichlet problem on $A = \{(0,0)\} \subset \mathbb{Z}^2$. This is easy. We just define

$$F(0,0) = \frac{1}{4}\left(g(1,0) + g(0,1) + g(-1,0) + g(0,-1)\right).$$

More generally we might try to solve the Dirichlet problem on other regions. If the set A is small then we could succeed by reformulating this as a linear algebra problem. For example if

$$A = \{(0,0),(1,0)\}$$

then, assuming $F(x_1,x_2) = g(x_1,x_2)$ for $(x_1,x_2) \in \partial^+ A$, we have

$$\begin{pmatrix} \mathscr{L}F(0,0) \\ \mathscr{L}F(1,0) \end{pmatrix} = \begin{pmatrix} -1 & \frac{1}{4} \\ \frac{1}{4} & -1 \end{pmatrix} \begin{pmatrix} F(0,0) \\ F(1,0) \end{pmatrix} + \frac{1}{4} \begin{pmatrix} g(-1,0) + g(0,1) + g(0,-1) \\ g(2,0) + g(1,1) + g(1,-1) \end{pmatrix}.$$

Thus, inverting the two by two matrix $\begin{pmatrix} -1 & \frac{1}{4} \\ \frac{1}{4} & -1 \end{pmatrix}$, we can solve $\mathscr{L}F = 0$ in A by taking

$$\begin{aligned} \begin{pmatrix} F(0,0) \\ F(1,0) \end{pmatrix} &= -\frac{4}{3} \begin{pmatrix} -1 & -\frac{1}{4} \\ -\frac{1}{4} & -1 \end{pmatrix} \frac{1}{4} \begin{pmatrix} g(-1,0) + g(0,1) + g(0,-1) \\ g(2,0) + g(1,1) + g(1,-1) \end{pmatrix} \\ &= \frac{1}{3} \begin{pmatrix} 1 & \frac{1}{4} \\ \frac{1}{4} & 1 \end{pmatrix} \begin{pmatrix} g(-1,0) + g(0,1) + g(0,-1) \\ g(2,0) + g(1,1) + g(1,-1) \end{pmatrix}. \end{aligned}$$

However, this procedure gets pretty invovled if the set A gets at all large.

Theorem 2.16. *Let $A \subset \mathbb{Z}^d$ be finite and let $g : \partial^+ A \to \mathbb{R}$ be given. Then there is a unique function $F : A \cup \partial^+ A \to \mathbb{R}$ solving the Dirichlet problem 2.24.*

PROOF. *Existence*: Let X_n be a random walk starting at a point $x \in \overline{A} = A \cup \partial^+ A$. By the central limit theorem

$$\mathbb{P}(X_n \in A) \sim \frac{c}{n^{\frac{d}{2}}} |\overline{A}|$$

where $|\overline{A}|$ is the number of points in $\overline{A}$. (We proved parts of this for $d = 1, 2$ and 3. The same ideas – moment generating functions – carry over to higher dimensions.) Since the right hand side goes to zero as $n \to \infty$ we conclude that

$$\mathbb{P}(X_n \in A \text{ for all } n \in N) = 0.$$

It follows that the following “stopping time”

$$T = \min\left\{n \,:\, X_n \in \partial^+ A\right\}$$

is *finite* with probability one. T is the first time at which the walk visits a site outside of A. In other words it is the “time of the first exit from A.”

Our proposed solution to the Dirichlet problem is:

$$F(x) = \mathbb{E}\left(g(X_T)\,|X_0 = x\right). \tag{2.10.1}$$

It is pretty easy to see that this works. Indeed

(1) If $x \in \partial_+ A$ then $T = 0$ and so $F(x) = g(x)$.
(2) If $x \in A$ then we compute that

$$\begin{aligned}
F(x) &= \sum_{y\in\partial^+ A} g(y)\mathbb{P}\left(X_T = y\,|X_0 = x\right)\\
&= \sum_{y\in\partial^+ A} g(y) \sum_{|e|=1} \mathbb{P}\left(X_T = y\,|X_0 = x+e\right)\mathbb{P}\left(X_1 = x+e\,|X_0 = x\right)\\
&= \sum_{|e|=1} \frac{1}{2d} \sum_{y\in\partial^+ A} g(y)\mathbb{P}\left(X_T = y\,|X_0 = x+e\right)\\
&= \sum_{|e|=1} F(x+e) = QF(x),
\end{aligned}$$

so F is harmonic. (In the second to last step we used the fact that the walk once it gets to the point $x+e$ after the first step may as well be considered as a brand new random walk starting at this point.)

Uniqueness: To see that the solution just constructed is the *only* solution we proceed as in our analysis of the Gambler’s ruin problem. In other words we will suppose given a solution F to the Dirichlet problem 2.24 and show that F is given by the formula (2.10.1).

So suppose F is any solution to Problem 2.24. Let us define

$$G_n(x) = \mathbb{E}\left(F\left(X_{n\wedge T}\right)|X_0 = x\right)$$

where $n \wedge T$ is the minimum of n and T. The walk $X_{n\wedge T}$ is a random walk that gets stuck at the first site it visits in $\partial^+ A$. Namely, $X_{n\wedge T} = X_n$ up to $n = T$ and then equals X_T for all subsequent n. Clearly $G_0(x) = F(x)$. Explicit calculation shows that

$$\begin{aligned}
G_n(x) &= \sum_{y\in\bar{A}} \mathbb{E}\left(F\left(X_{n\wedge T}\right)\left|X_{(n-1)\wedge T} = y\right.\right)\mathbb{P}\left(X_{(n-1)\wedge T} = y\,|X_0 = x\right)\\
&= \sum_{y\in\partial^+ A} F(y)\mathbb{P}\left(X_{(n-1)\wedge T} = y\,|X_0 = x\right) + \sum_{y\in A} \frac{1}{2d} \sum_{|e|=1} F(y+e)\mathbb{P}\left(X_{(n-1)\wedge T} = y\,|X_0 = x\right)\\
&= \sum_{y\in\partial^+ A} F(y)\mathbb{P}\left(X_{(n-1)\wedge T} = y\,|X_0 = x\right) + \sum_{y\in A} F(y)\mathbb{P}\left(X_{(n-1)\wedge T} = y\,|X_0 = x\right)\\
&= G_{n-1}(x),
\end{aligned}$$

where we have used the fact that, since F is harmonic,

$$\frac{1}{2d}\sum_{|e|=1} F(y+e) = F(y).$$

By induction we see that $G_n(x) = G_0(x) = F(x)$ for all $n \in \mathbb{N}$.

Thus

$$F(x) = \lim_{n\to\infty} G_n(x) = \lim_{n\to\infty} \sum_{y\in\overline{A}} F(y)\mathbb{P}(X_{n\wedge T} = y|X_0 = x). \tag{2.10.2}$$

Consider two cases for $y \in \overline{A}$:

(1) If $y \in A$ then

$$\mathbb{P}(X_{n\wedge T} = y|X_0 = x) \le \mathbb{P}(X_m \in A \text{ for all } m \le n|X_0 = x) \le \frac{C}{n^{\frac{d}{2}}} \to 0.$$

(2) If $y \in \partial^+ A$ then the event $\{X_{n\wedge T} = y\}$ is the event $\{n \ge T\} \cap \{X_T = y\}$. Since $T < \infty$ with probability one, we see that

$$\mathbb{P}(X_{n\wedge T} = y|X_0 = x) = \mathbb{P}(n \ge T \text{ and } X_T = y|X_0 = x) \xrightarrow{n\to\infty} \mathbb{P}(X_T = y|X_0 = x).$$

Plugging these limits into (2.10.2) shows that F is indeed given by (2.10.1). □

2.11. Lecture 16: Average Time of Return and the Inhomogeneous Dirichlet Problem

Let us now take a look at the 1D random walk again. Consider the gambler's ruin scenario. However, let us now ask: "How many times will the gambler be able to play the game?" The answer, of course, is a random number. For instance, if the gambler starts with \$1, then he may very well play the game only one time. In fact this happens with probability $\frac{1}{2}$. On the other hand, if the Casino has, say, one million dollars, then he might win exactly one million times and walk away with all of the Casino's money. (We probably shouldn't worry too much about this possibility since the chances are $\frac{1}{2^{10^6}}$ which is very very small.)

So let us begin by asking: "What is the average number of games he plays?" That is, we consider a random walk in 1D starting at some point $x \in \{0,\dots,N\}$ and ask for the value of

$$\tau(x) = \mathbb{E}(T|X_0 = x)$$

where

$$T = \min\{n : X_n = 0 \text{ or } N\}$$

is the random step at which the gambler runs out of money or reaches his goal of N.

First of all, let us note that

$$\tau(0) = \tau(N) = 0, \tag{2.11.1}$$

since $T = 0$ if the walker starts off in the boundary set $\{0,N\}$. Now suppose the walker starts at $x \in \{1,\dots,N\}$. Then $X_1 = x \pm 1$, where each possibility has probability $\frac{1}{2}$. Thus

$$\tau(x) = \frac{1}{2}\mathbb{E}(T|X_1 = x+1 \text{ and } X_0 = x) + \frac{1}{2}\mathbb{E}(T|X_1 = x-1 \text{ and } X_0 = x).$$

Now we can consider the walk $X_2, X_3, \dots$ starting at $X_1 = x \pm 1$ as a new random walk $Y_n = X_{n-1}$ with a new stopping time

$$T_Y = \min\{n : Y_n = 0 \text{ or } N\}.$$

In making this substitution we must take into account that

$$T = T_Y + 1$$

for $X_0 = x \in \{1,\ldots,N\}$. Since the walk Y_j has the same distribution as the walk X_j starting at $X_0 = x \pm 1$ we conclude that

$$\begin{aligned}\tau(x) &= \frac{1}{2}\left(\mathbb{E}\left(T_Y \,|Y_0 = x-1\right) + 1 + \mathbb{E}\left(T_Y \,|Y_0 = x+1\right) + 1\right) \\ &= 1 + \frac{1}{2}\left(\tau(x-1) + \tau(x+1)\right).\end{aligned}$$

Thus

$$\mathscr{L}\tau(x) = -1 \tag{2.11.2}$$

where $\mathscr{L}$ is the "discrete Laplacian" defined last time, which in $1D$ is simply

$$\mathscr{L}F(x) = \frac{1}{2}\left(F(x+1) + F(x-1) - 2F(x)\right).$$

Now we need to solve (2.11.2) for τ satisfying the "boundary conditions" (2.11.1). First off, let us forget about the boundary conditions, and use the intuition that $\mathscr{L}$ is kind of like a second derivative to guess that (2.11.2) might have a quadratic solution. Indeed we can easily check that if $h(x) = x^2$ then

$$\mathscr{L}h(x) = \frac{1}{2}\left((x+1)^2 + (x-1)^2 - 2x^2\right) = 1.$$

Thus

$$\left[\mathscr{L}\left(\tau + h\right)\right](x) = 0$$

so $\tau(x) + h(x) = \tau(x) + x^2$ is a *harmonic* function. What are it's boundary values? Looking at (2.11.1) we see that

$$\tau(0) + h(0) = 0 \quad \text{and} \quad \tau(N) + h(N) = N^2.$$

By Theorem 2 from Lecture 13

$$\tau(x) + x^2 = \left[\tau + h\right](x) = \frac{N^2}{N}x = Nx.$$

Thus we have shown that

$$\tau(x) = x(N-x). \tag{2.11.3}$$

This solution is quite surprising. For instance

$$\tau(1) = N-1.$$

Thus the *average length of the walk starting at* $X_0 = 1$ *is* N. This is quite strange since the walk ends after one step with probability $\frac{1}{2}$. Based on this, we might think that the average length of the walk would be some small finite number, not a quantity that grows with N.

Typical values, median values and mean values. Often times one wants to answer the question: "What is the typical value of X?", where X is a random variable. This is a loosely defined question. We don't really know what typical means. What we really have in mind is some idea for the size of X that holds with good probability.

One way to get an idea of the typical value of X is to look at it's average $\mathbb{E}(X)$, however the value of the average may or may not be typical. Indeed, X need not take values close to it's average. For instance if $X = \pm 10^{23}$ each with probability $\frac{1}{2}$ then $\mathbb{E}(X) = 0$ even though typically (with probability one even!) X is a number close to 10^{23} in magnitude. The problem here is cancellation between positive and negative values. If instead, we consider the positive variable

$Y = X + 10^{23} = 0$ or 2×10^{23} each with probability $\frac{1}{2}$ then $\mathbb{E}(Y) = 10^{23}$ which is a decent measure of the order of magnitude that Y attains with probability at least $\frac{1}{2}$.

However, even if a variable is positive we can run into trouble with the mean. The problem is the "Bill Gates" phenomenon. According to wikipedia Bill Gates' fortune is worth about $50 billion. Hence if Bill Gates were to walk into our classroom, the 20 people in the room would have fortuens, on average, of $2.5 billion each! (We can neglect the total fortune of everyone else in the room since that sum is nowhere near to the significant fiugres in the estimate of $50 billion for Bill Gates.) However, this average is an absolutely lousely estimate for the typical size of the fortunes of everyone in the room except Bill Gates.

Thus an average can be skewed by extremely large rare events. In such situations it is useful to define another concept which gets at the typical size of a variable. One such concept is the idea of *a median*. A *median* for a random variable X is a number $a \in \mathbb{R}$ such that

$$\mathbb{P}(X < a) \leq \frac{1}{2} \quad \text{and} \quad \mathbb{P}(X > a) \leq \frac{1}{2}. \tag{2.11.4}$$

(Equivalently $\mathbb{P}(X \leq a) \geq \frac{1}{2}$ and $\mathbb{P}(X \geq a) \geq \frac{1}{2}$.) A sufficient condition for this is if there is a particular number a such that

$$\mathbb{P}(X \leq a) = \frac{1}{2},$$

but such a number need not exist. However any probability distribution admits a median according to (2.11.4).

If Bill Gates were to walk into our class room, the median fortune in the room would only go up by a very modest amount. (Economists typically look at median income or net worth because it is known that the distribution of incomes has "heavy tails" which skew averages. For example, according to the census bureau the median household income in the U.S. in 2008 was $50,303 although the average houseld income was $68,424 .)

Regarding the Gambler's ruin problem, we see that a *median* value of T is, in fact, one. In fact any number $a \in [1,3)$ is a median for T, since the next possible value of T is 3 so $\mathbb{P}(T \leq a) = \frac{1}{2}$ for any such a. Thus we are justified in saying that T is typically of order one in magnitude. We will be right with probability at least $\frac{1}{2}$.

What is going on with the average of T? We can understand this as follows. A walker starting at $X_0 = 1$ typically will reach the point 0 after a small number of steps. However, according to our solution of the Gambler's ruin problem he will make it to the midpoint of the interval $\{0, \ldots, N\}$ with probability about $\frac{2}{N}$. (Imagine that N is even, or modify the statements accordingly for N odd.) Once he gets to the midpoint, there are even odds that he will exit at 0 or N. How long does it take him to get from the midpoint to 0 or N? About $\left(\frac{N}{2}\right)^2$ steps according to what we saw from the central limit theorem. Thus the average of T should be at least

$$\text{Probability to reach midpoint} \times \text{Typical number of steps from midpoint to boundary} = \frac{2}{N}\left(\frac{N}{2}\right)^2 = \frac{N}{4},$$

which is, indeed, of the order of the solution $\tau(1) = N - 1$ that we found. Thus the average $\tau(x)$ for x close to the boundary points $\{0, N\}$ is dominated by the very rare events of crossing the whole interval in about N^2 steps. Since these occur with probability about $\frac{1}{N}$ they produce a very large contribution to the average when N is large.

Average time of return. We can use our solution (2.11.3) to find the average time for the walker to return to his starting point. Indeed, let X_n be a random walk with $X_0 = 0$ and let

$$\widetilde{T} = \min\{n \geq 1 \, : \, X_n = 0\}.$$

Thus $\widetilde{T}$ is the time of the first return to 0. Considering what the walker does after one step we see that

$$\begin{aligned}\mathbb{E}\left(\widetilde{T}\right) &= \frac{1}{2}\left(\mathbb{E}\left(\widetilde{T}\,|X_1 = 1\right) + \mathbb{E}\left(\widetilde{T}\,|X_1 = -1\right)\right)\\ &= \mathbb{E}\left(\widetilde{T}\,|X_1 = 1\right),\end{aligned}$$

since the distribution of $\widetilde{T}$ given that $X_1 = -1$ should be the same as the distribution given that $X_1 = 1$ by symmetry.

To compute the expectation of $\widetilde{T}$ given that $X_1 = 1$ we can use the following argument. Fix $N \geq 1$ and let

$$T_N = \min\{n \geq 1 \, : \, X_n = 0 \text{ or } X_n = N\}.$$

Clearly

$$\widetilde{T} \geq T_N.$$

However based on our solution for the expected time to leave the interval $\{1, \dots, N\}$ we see that

$$\mathbb{E}\left(T_N \,|X_1 = 1\right) = \tau(1) + 1 = N.$$

Thus

$$\mathbb{E}\left(\widetilde{T}\right) = \mathbb{E}\left(\widetilde{T}\,|X_1 = 1\right) \geq \mathbb{E}\left(T_N\,|X_1 = 1\right) = N.$$

Since N is arbitrary we see that the *average time to return is ∞!*

Again this average is dominated by rare events. The difficulty is not that it typically takes a long time to return to 0. On the contrary with probability $\frac{1}{2}$ the walk returns to 0 in two steps — so a median value for $\widetilde{T}$ is two. Instead, what happens is that

$$\mathbb{P}\left(\widetilde{T} \geq 2N\right) \approx \frac{1}{\sqrt{2N}}$$

so that the average infinite.

Inhomogeneous Dirichlet Problem. We have already seen that there is some give and take between linear equations involving $\mathscr{L}$ and random walk. We have solved the Gambler's ruin problem and found the average time of return using the 1D Dirichlet problem. On the other hand we showed that the Dirichlet problem has solutions using the random walk. The following theorem is another example of this kind:

Theorem 2.17. *Let $A \subset \mathbb{Z}^d$ be a finite set and let X_n be a random walk in $\mathbb{Z}^d$ starting at a point $x \in A \cup \partial^+ A$. Let*

$$T = \min\left\{n \, : \, X_n \in \partial^+ A\right\}$$

and for $y \in A$ let V_y denote the number of times that X_n visits y before $n = T$, that is

$$V_y = \#\{n \leq T \, : \, X_n = y\}.$$

If $h : A \to \mathbb{Z}^d$, then

$$F(x) = \sum_{y \in A} h(y) \mathbb{E}\left(V_y \,|X_0 = x\right) \tag{2.11.5}$$

is the unique solution to

(1) $\mathscr{L}F(x) = -h(x)$ *for* $x \in A$*, and*
(2) $F(x) = 0$ *for* $x \in \partial^+ A$.

Remark 2.10. (1) Using Theorem 5 from last time we see that the unique solution to $\mathscr{L}F(x) = -h(x)$ for $x \in A$ with $F(x) = g(x)$ for $x \in \partial^+ A$ is

$$F(x) = \sum_{y \in A} h(y)\mathbb{E}\left(V_y \,|X_0 = x\right) + \sum_{y \in \partial^+ A} g(y)\mathbb{P}\left(X_T = y\,|X_0 = x\right).$$

(2) Since

$$V_y = \sum_{j=0}^{T-1} I[X_j = y]$$

we have

$$F(x) = \mathbb{E}\left(\sum_{j=0}^{T-1} h(X_j)\,|X_0 = x\right). \tag{2.11.6}$$

(3) The function

$$G_A(x,y) = \mathbb{E}\left(V_y \,|X_0 = x\right)$$

is called the "Green's function" of A.

PROOF. First we should check that (2.11.5) is indeed a solution. To do this we us (2.11.6) to compute that

$$\begin{aligned} F(x) &= h(x) + \sum_{|y-x|=1} \frac{1}{2d}\mathbb{E}\left(\sum_{j=1}^{T-1} h(X_j)\,|X_0 = x \text{ and } X_1 = y\right) \\ &= h(x) + \frac{1}{2d}\sum_{|y-x|=1} \mathbb{E}\left(\sum_{j=0}^{T-1} h(X_j)\,|X_0 = y\right) \\ &= h(x) + \frac{1}{2d}\sum_{|y-x|=1} F(y) \end{aligned}$$

where we have used the fact the random walk from time $n = 1$ onward can be considered as a new random walk starting at point y. Thus $\mathscr{L}F(x) = -h(x)$. On the other hand, for $x \in \partial^+ A$, then $F(x) = 0$ (since $V_y = 0$ for a walk starting at $x \in \partial^+ A$).

To prove uniqueness, we simply reduce to the Dirichlet problem. If F and $\widetilde{F}$ are two possible solutions then

$$\left[\mathscr{L}\left(F - \widetilde{F}\right)\right](x) = 0$$

for $x \in A$ while for $x \in \partial^+ A$ we have

$$\left(F - \widetilde{F}\right)(x) = 0.$$

It follows from Theorem 5 from last time that $F(x) - \widetilde{F}(x) = 0$ for every $x \in A$. □

2.12. Lecture 17:

You may have noticed a difference between the results we have obtained so far for the 1D random walk and those we obtained for the 2D, 3D, etc. random walks. In 1D we have exlicit expressions like

$$\mathbb{P}(X_T = N | X_0 = x) = \frac{x}{N}$$

where

$$T = \min\{n : X_n = 0 \text{ or } N\}.$$

(See Lecture 13.) On the other hand, in higher dimensions, the best we did was to say that the solution to the Dirichlet problem

$$\mathscr{L}F(x) = 0,\ x \in A \quad \text{and} \quad F(x) = g(x),\ x \in \partial^+ A$$

is given in terms of random walk probabilities:

$$F(x) = \sum_{y \in \partial^+ A} g(y) \mathbb{P}(X_{T_A} = y | X_0 = x) = \mathbb{E}(g(X_{T_A}) | X_0 = x),$$

where

$$T_A = \min\left\{n : X_n \in \partial^+ A\right\}.$$

(See Lectures 15 and 16.) *However, we never said how to compute the probabilities*

$$\mathbb{P}(X_{T_A} = y | X_0 = x). \tag{2.12.1}$$

There is a reason for this: in general computing these numbers is a difficult problem. In the next few lectures we will show how to use some linear algebra to get useful expressions for the numbers (2.12.1) for certain special sets A, namely squares or cubes (or "hyper-cubes" in higher dimensions).

To get started, suppose we want to compute instead the numbers

$$p_n(y) := \mathbb{P}(X_{n \wedge T_A} = y | X_0 = x) \tag{2.12.2}$$

for some given site $y \in A$. Since the walk must be at a neighbor of y at step $n-1$ to get to y at step n we have

$$p_n(y) = \frac{1}{2d} \sum_{\substack{|y'-y|=1 \\ y' \in A}} p_{n-1}(y'). \tag{2.12.3}$$

In (2.12.3) the factor $\frac{1}{2d}$ in front of the sum is the probability $\mathbb{P}\left(X_{n \wedge T_A} = y \middle| X_{(n-1) \wedge T_A} = y'\right)$. The sum on the right hand side is restricted to those sites $y' \in A$ since if $y' \in \partial^+ A$ then $\mathbb{P}\left(X_{n \wedge T_A} = y \middle| X_{(n-1) \wedge T_A} = y'\right) = 0$ for $y \in A$ (since if $X_{(n-1) \wedge T_A} = y'$ then $n - 1 \geq T_A$ and the walk is stuck at y').

It is useful to write (2.12.3) as

$$p_n(y) = Q_A p_{n-1}(y)$$

where Q_A is the linear operator on the space of functions $f : A \to \mathbb{R}$ given by

$$Q_A f(x) = \frac{1}{2d} \sum_{\substack{|x'-x|=1 \\ x \in A}} f(x').$$

Then

$$p_n(y) = Q_A^n p_0(y) = Q_A^n \delta_x(y) \tag{2.12.4}$$

where

$$p_0(y) = \delta_x(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases}.$$

Our goal is to understand how to use (2.12.4) to compute the probabilities $p_n(y)$ and also $\mathbb{P}(X_{T_A} = y | X_0 = x)$ for certain sets A. To do this we need to use some linear algebra.

Before presenting the linear algebra we need, notice that we can compute $\mathbb{P}(X_{T_A} = y | X_0 = x)$ from $p_n(y)$. Indeed, in order for the walk to end up at $y \in A$ it must first get to a neighbor y' of y with $y' \in A$. When it reaches y' the probability that it gets to y at the next step is $\frac{1}{2d}$. Thus

$$\mathbb{P}(X_{T_A} = y | X_0 = x) = \sum_{\substack{|y'-y|=1 \\ y' \in A}} \sum_{n=0}^{\infty} \frac{1}{2d} p_n(y').$$

As we will see, this formula is reasonably useful.

Linear Algebra. In linear algebra one usually thinks of a vector $\mathbf{x}$ as a list of numbers, usually written as a "column vector"

$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ \vdots \\ x_n \end{pmatrix}.$$

Here we will take a slightly different perspective, because ultimately we are interested in *functions* defined on a finite set A which may or may not have a reasonable order on it. So we will think of a *vector* as a function from a finite set, say $\{1, \ldots, n\}$, into the real numbers. A column vector can be thought of this way by thinking of $\mathbf{x}$ as the map

$$\mathbf{x}(i) = x_i.$$

Thus we will think of $\mathbb{R}^n$ as the set

$$\mathbb{R}^n = \mathbb{R}^{\{1,\ldots,n\}} = \{f : \{1, \ldots, n\} \to \mathbb{R}\},$$

where a function $f(\cdot)$ is associated with the column vector

$$\begin{pmatrix} f(1) \\ \vdots \\ \vdots \\ f(n) \end{pmatrix}.$$

More generally, if A is any finite set we can consider the vector space

$$\mathbb{R}^A = \{f : A \to \mathbb{R}\}.$$

This space is isomorphic to $\mathbb{R}^{|A|}$ where $|A|$ is the number of elements in A, but to make an isomorphism we need to choose an order for A which may not be a natural thing to do for a particular problem.

The *dot product* or *scalar product* or *inner product* of two vectors $f, g \in \mathbb{R}^A$ is

$$\langle f, g \rangle = \sum_{x \in A} f(x) g(x).$$

A *linear operator* on $\mathbb{R}^A$ can be determined by a function $M : A^2 \to \mathbb{R}$ as follows:

$$Mf(x) = \sum_{y\in A} M(x,y)f(y).$$

If $A = \{1,\ldots,n\}$, you can check that this corresponds to the usual notion of a matrix acting on a vector if we associate M with the matrix

$$\begin{pmatrix} M(1,1) & \cdots & \cdots & M(1,n) \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ M(n,1) & \cdots & \cdots & M(n,n) \end{pmatrix}. \tag{2.12.5}$$

A linear operator $M(\cdot,\cdot)$ is *symmetric* if

$$\langle Mf, g\rangle = \langle f, Mg\rangle$$

for all $f,g \in \mathbb{R}^a$. In other words M is symmetric if

$$\begin{aligned} \sum_{x,y\in A} M(x,y)f(y)g(x) &= \sum_{x\in A}\left(\sum_{y\in A} M(x,y)f(y)\right)g(x) \\ &= \sum_{y\in A} f(y)\left(\sum_{x\in A} M(y,x)g(x)\right) = \sum_{x,y\in A} M(y,x)f(y)g(x) \end{aligned}$$

for all f,g. In order for this to hold it is necessary and sufficient that

$$M(x,y) = M(y,x)$$

for every $x,y \in A$. (To see this for a particular pair x,y take $f = \delta_y$ and $g = \delta_x$.) If $A = \{1,\ldots,n\}$ then M is symmetric if and only if the matrix (2.12.5) is equal to its own transpose.

Recall that a set of vectors $f_1,\ldots,f_m$ is *linearly independent* if whenever $a_1,\ldots,a_m$ are real numbers such that

$$\sum_{j=1}^{m} a_j f_j = 0$$

we have $a_1 = \cdots = a_m = 0$. Two vectors f_1 and f_2 are *orthogonal*, denoted $f_1 \perp f_2$, if

$$\langle f_1, f_2\rangle = 0.$$

Problem 2.25. Let $f_1,\ldots,f_m$ be pairwise orthogonal ($f_i \perp f_j$ for $i \neq j$). Show that $f_1,\ldots,f_m$ is linearly independent.

A set $f_1,\ldots,f_m$ of vectors *spans* $\mathbb{R}^A$ if any vector $f \in \mathbb{R}^A$ can be written as a *linear combination*

$$f = \sum_{j=1}^{m} c_j f_j$$

for some real numbers $c_1,\ldots,c_m$. The set $f_1,\ldots,f_m$ is a *basis* if it is linear independent and spans. It is a basic fact of linear algebra that bases exist and that the number of elements in any basis for $\mathbb{R}^A$ is $|A|$. (The number of elements in a basis is called the dimension of the vector space.)

A basis is an *orthogonal basis* if its elements are pairwise orthogonal. Given any orthogonal basis $f_1,\ldots,f_n$ we can turn it into an *orthonormal basis* by dividing each vector by it's *length*:

$$\frac{1}{\|f_1\|}f_1,\ldots,\frac{1}{\|f_n\|}f_n,$$

where the *length* of a vector f is defined to be

$$\|f\| = \sqrt{\langle f, f\rangle}.$$

It is a basic fact of linear algebra that orthonormal bases for $\mathbb{R}^A$ exist. In fact we can turn any basis into an orthonormal basis using the Gram Schmidt process.

The vector space $\mathbb{R}^A$ is a *metric space* if we define the distance between two vectors f and g to be

$$d(f,g) = \|f - g\|.$$

With this definition the Heine Borel Theorem holds: *A subset $C \subset \mathbb{R}^A$ is compact if and only if C is closed and bounded.* (Here *bounded* means that there is a number R such that $f \in C \implies \|f\| \le R$.)

An *eigenvalue* for a linear operator M is a number $\lambda \in \mathbb{R}$ such that there is a non-zero vector $f_\lambda \in \mathbb{R}^A$ for which

$$M f_\lambda = \lambda f_\lambda.$$

The function f_λ is called an *eigenvector* (or *eigenfunction*) for M.The central result of linear algebra we will use is the following

Theorem 2.18. *Let M be a symmetric linear operator on $\mathbb{R}^A$ then there is an orthonormal basis of eigenfunctions for M.*

PROOF. Let $S = \left\{ f \in \mathbb{R}^A \ : \ \|f\| = 1 \right\}$. Since S is closed and bounded, it is a *compact metric space*. Let $\Phi : S \to \mathbb{R}$ denote the function

$$\Phi(f) = \langle f, Mf\rangle.$$

Problem 2.26. Show that Φ is a continuous function.

Since Φ is continuous, it attains it's maximum at some point. That is there is $f_+ \in S$ such that

$$\Phi(f) \le \Phi(f_+) \quad \text{for all } f \in S.$$

I claim that f_+ is an eigenvector. To see this, fix $g \in \mathbb{R}^A$. For sufficiently small t, $f_+ + tg \neq 0$, so the function

$$\phi(t) = \Phi\left(\frac{1}{\|f_+ + tg\|} (f_+ + tg) \right) = \frac{\langle f_+ + tg, M(f_+ + tg)\rangle}{\langle f_+ + tg, f_+ + tg\rangle}$$

is a real valued map $\phi : (-\varepsilon, \varepsilon) \to \mathbb{R}$ for sufficiently small ε. In fact, ϕ is a ratio of quadratic functions:

$$\phi(t) = \frac{\langle f_+, Mf_+\rangle + 2t\langle g, Mf_+\rangle + t^2 \langle g, Mg\rangle}{1 + 2t\langle g, f_+\rangle + t^2 \|g\|^2},$$

where we have noted that $\|f_+\| = 1$ and used the fact that symmetry of M along with the symmetry of the inner product ($\langle f, g\rangle = \langle g, f\rangle$) to write

$$\langle g, Mf_+\rangle + \langle f_+, Mg\rangle = 2\langle g, Mf_+\rangle.$$

Thus ϕ is differentiable and, since $\phi(0) \ge \phi(t)$ for all $t \in (-\varepsilon, \varepsilon)$,

$$0 = \phi'(0) = 2\langle g, Mf_+\rangle - 2\Phi(f_+)\langle g, f_+\rangle.$$

Hence, if $g \perp f_+$ then $\langle g, Mf_+\rangle = 0$, that is $g \perp Mf_+$. It follows that there is a number λ_+ such that

$$Mf_+ = \lambda_+ f_+,$$

explain

so f_+ is an eigenvector! (One way to see this is to note that we can find an orthonormal basis $f_1, \ldots, f_n$ for $\mathbb{R}^A$ with $f_1 = f_+$. Since this is a basis, we can write

$$Mf_+ = \sum_{j=1}^{n} c_j f_j.$$

The coefficients c_j can be recovered by using the fact that the basis is pairwise orthogonal so that

$$\langle f_j, Mf_+ \rangle = c_j.$$

But we have shown that $\langle f_j, Mf_+ \rangle = 0$ for $j = 2, \ldots, n$. Thus $Mf_+ = c_1 f_1 = \lambda_+ f_+$ where $\lambda_+ = c_1$.)

So far we have worked (a little) hard to show that M has one eigenvector. However, at this point we are basically done! To see this, note that by what we obtained above if $g \perp f_+$ then

$$\langle Mg, f_+ \rangle = \langle g, Mf_+ \rangle = \lambda_+ \langle g, f_+ \rangle = 0,$$

so $Mg \perp f_+$. Let $V = \{g : g \perp f_+\}$. Then we have shown that $g \in V \implies Mg \in V$. Let M_V denote the restriction of M to V that is

$$M_V g = Mg, \quad g \in V.$$

Then M_V is a symmetric linear operator on a finite dimensional space, so by the above argument it has an eigenvector f_2, which by virtue of being in V is perpendicular to $f_1 = f_+$. We can repeat this process to produce $f_1, \ldots, f_n$ which are normalized (length one) eigenvectors, with each constructed to be orthogonal to all previous, so the set is orthonormal. We can continue as long as

$$V_n = \left\{ g \ : \ g \perp f_j \text{ for } j = 1, \ldots, n \right\} \neq \{0\}.$$

We get $V_n = \{0\}$ exactly when $f_1, \ldots, f_n$ is a basis, that is when $n = |A|$. □

Notice that the algorithm used in the proof produces a list of eigenvectors $f_1, \ldots, f_n$ with associated eigenvalues $\lambda_1, \ldots, \lambda_n$ that are *non-increasing*:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

If you look at the algorithm you will see that the eigenvectors are given by

$$\begin{aligned}
\lambda_1 &= \max_{f \in S} \langle f, Mf \rangle, \\
\lambda_2 &= \max_{\substack{f_1, f_2 \in S \\ f_1 \perp f_2}} \min_{j=1,2} \langle f_j, Mf_j \rangle, \\
&\vdots \\
\lambda_m &= \max_{\substack{f_1, \ldots, f_m \in S \\ f_i \perp f_j \ i \neq j}} \min_{j=1,\ldots,m} \langle f_j, Mf_j \rangle, \\
&\ddots
\end{aligned}$$

This is the *min-max* principle, so named because of the combination of minima and maxima that appear. (The principle is often stated in the other order

$$\lambda_n = \min_{f \in S} \langle f, Mf \rangle, \quad \lambda_{n-1} = \min_{\substack{f_1, f_2 \in S \\ f_1 \perp f_2}} \max_{j=1,2} \langle f_j, Mf_j \rangle, \quad \ldots.$$

Thus he name "min-max" instead of "max-min.")

Theorem 2.18 is the "Spectral theorem" for symmetric matrices. The main point of it is that we can understand the action of M as follows. Let $f_1,\dots,f_n$ be an orthonormal basis of eigenvectors, with associated eigenvalues $\lambda_1,\dots,\lambda_n$. For any $f\in\mathbb{R}^A$ we have

$$f=\sum_{j=1}^{n}\langle f_j,f\rangle f_j$$

and thus

$$Mf=\sum_{j=1}^{n}\lambda_j\langle f_j,f\rangle f_j.$$

This, in itself, is not so useful since it is not that hard to evaluate Mf from the definition. However, it is alot more work to evaluate $M^m f$ directly if m is a large power. The spectral theorem makes this *VERY EASY*:

$$M^m f=\sum_{j=1}^{n}\lambda_j^m\langle f_j,f\rangle f_j.$$

Thus is we can find the eigenvectors and eigenvalues for Q_A we will be in good shape to compute the probabilities $p_n(y)$.

2.13. Lecture 18:

After our digression last time to prove the spectral theorem form symmetric operators, we now return to the random walk. Recall that if $A\subset\mathbb{Z}^d$ is a finite set then

$$p_n(y)=\mathbb{P}(X_{T_A\wedge n}=y\,|X_0=x)$$

is given by

$$p_n=Q_A^n\delta_x, \tag{2.13.1}$$

where T_A is the exit time from A,

$$T_A=\min\left\{n\ :\ X_n\in\partial^+A\right\},$$

and Q_A is the linear operator

$$Q_Af(y)=\sum_{\substack{|y'-y|\le 1\\ y'\in A}}f(y').$$

Problem 2.27. Show that Q_A is a symmetric linear operator on the vector space $\mathbb{R}^A$.

Since Q_A is symmetric, it makes sense to try to find it's orthonormal basis of eigenfunctions.

Eigenfunctions in 1D. We start by looking at Q_A in 1D, with

$$A=\{1,\dots,N-1\}.$$

The problem now is to find $f:\{1,\dots,N-1\}\to\mathbb{R}$ such that

$$Q_Af=\lambda f. \tag{2.13.2}$$

Let us think of such an f as a function from the larger domain $\{0,\dots,N\}$ into $\mathbb{R}$ by defining $f(0)=f(N)=0$. Then we see that f satsifies

$$\mathscr{L}f(x)=(\lambda-1)f(x),\quad x\in A, \tag{2.13.3}$$

where $\mathscr{L}$ is the discrete Laplacian

$$\mathscr{L}f(x) = \frac{1}{2}\left(f(x+1)+f(x-1)-2f(x)\right).$$

To understand what functions f might satisfy (2.13.2) and thus (2.13.2) we will use the analogy that $\mathscr{L}$ is like a second derivative. Furthermore, we expect λ in (2.13.2) to satsify $|\lambda| \leq 1$ since Q_A^n is related to *probabilities* through (2.13.1) and so λ^n should not grow. Thus (2.13.3) is analogous to the continuum equation

$$f''(x) = -\omega^2 f(x) \tag{2.13.4}$$

with $\omega^2 \geq 0$. Eq. (2.13.4) is the oscillator equation, which has solutions of the form $f(x) = A\cos(\omega x) + B\sin(\omega x)$. So let us see if (2.13.3) has trig function solutions.

Since we are interested in $f(x)$ with $f(0) = 0$ let us try $f(x) = \sin(\omega x)$. Since we want $f(N) = 0$ we should take $\omega = \frac{\pi j}{N}$ for some $j \in \mathbb{N}$. If $f(x) = \sin\left(\frac{\pi j}{N}x\right)$ we can easily check that

$$\begin{aligned} Q_A f(x) &= \frac{1}{2}\left(\sin\left(\frac{\pi j}{N}(x+1)\right) + \sin\left(\frac{\pi j}{N}(x-1)\right)\right) \\ &= \frac{1}{2}\left(\sin\left(\frac{\pi j}{N}x\right)\cos\left(\frac{\pi j}{N}\right) + \sin\left(\frac{\pi j}{N}\right)\cos\left(\frac{\pi j}{N}x\right) + \sin\left(\frac{\pi j}{N}x\right)\cos\left(\frac{\pi j}{N}\right) - \sin\left(\frac{\pi j}{N}\right)\cos\left(\frac{\pi j}{N}x\right)\right) \\ &= \cos\left(\frac{\pi j}{N}\right)\sin\left(\frac{\pi j}{N}x\right) = \cos\left(\frac{\pi j}{N}\right) f(x). \end{aligned}$$

So we have found an eigenfunction!

In fact, we have found an eigenfunction $f_j(x) = \sin\left(\frac{\pi j}{N}x\right)$, with associated eigenvalue $\cos\left(\frac{\pi j}{N}\right)$, for each natural number $j \geq 1$. This seems a little strange at first, as there should only be $N-1$ eigenfunctions because the vector space $\mathbb{R}^A$ has dimension $N-1$. The resolution to this little paradox is that not all of the eigenfunctions are distinct. In fact, notice that

$$f_N(x) = \sin(\pi x) \equiv 0,$$

since we are considering $x \in \{1,\ldots,N-1\} \subset \mathbb{N}$ and sin vanishes on integer multiples of π. Thus f_N isn't an eigenfunction after all (eigenfunctions need to be non-zero). Similarly, since

$$\sin(n\pi + \phi) = -\sin(n\pi - \phi)$$

for any integer n and any angle ϕ, we have

$$f_{N+j}(x) = \sin\left(\pi x + \frac{\pi j}{N}x\right) = -\sin\left(\pi x - \frac{\pi j}{N}x\right) = -f_{N-j}(x)$$

for all j. Continuing in this way we could obtain the identity

$$f_{aN+j}(x) = \sin\left(a\pi + \frac{\pi j}{N}x\right) = \begin{cases} f_j(x) & \text{if } a \text{ is even,} \\ -f_{N-j}(x) & \text{if } a \text{ is odd.} \end{cases}$$

Thus no new functions are introduced one we get past $j = N-1$.

On the other hand, the eigenvalues $\lambda_j = \cos\left(\frac{2\pi j}{N}\right)$ for $j = 1, \ldots, N-1$ are all distinct. In fact, they are strictly decreasing as j increases from 1 to $N-1$:

$$1 > \lambda_1 > \lambda_2 > \cdots > \lambda_N > -1.$$

It follows that the functions $f_1, \ldots, f_{N-1}$ are all distinct, and in fact that they are orthogonal:

$$0 = \langle f_j, f_k \rangle = \sum_{x=1}^{N-1} \sin\left(\frac{\pi j}{N} x\right) \sin\left(\frac{\pi k}{N} x\right) \quad \text{if } j \neq k,$$

since we have

$$\lambda_j \langle f_j, f_k \rangle = \langle Q_A f_j, f_k \rangle = \langle f_j, Q_A f_k \rangle = \lambda_k \langle f_j, f_k \rangle .$$

Computing the norm of f_j,

$$\|f_j\| = \sqrt{\langle f_j, f_j \rangle}$$

is not so straightforward. In a supplement to this lecture I explain how to use roots of unity and Euler's formula to show that

$$\sum_{x=1}^{N-1} \sin^2\left(\frac{\pi j}{N} x\right) = \frac{N}{2}$$

for each $j = 1, \ldots, N-1$. Thus $\|f_j\| = \sqrt{\frac{N}{2}}$.

Since we have found exactly $N-1$ eigenvectors, we have found all of the eigenvectors of Q_A. To get an orthonormal basis we need to divide each vector by it's length. So let us define

$$h_j = \sqrt{\frac{2}{N}} f_j.$$

Then $h_1, \ldots, h_{N-1}$ is an orthonormal basis, and

$$Q_A^n f(x) = \sum_{j=1}^{N-1} \lambda_j^n \langle h_j, f \rangle h_j(x) = \frac{2}{N} \sum_{j=1}^{N-1} \sum_{y=1}^{N-1} \cos^n\left(\frac{\pi j}{N}\right) \sin\left(\frac{\pi j}{N} x\right) \sin\left(\frac{\pi j}{N} y\right) f(y).$$

Looking back at (2.13.1) we see that

$$\mathbb{P}(X_{n \wedge T_A} = y | X_0 = x) = \frac{2}{N} \sum_{j=1}^{N-1} \cos^n\left(\frac{\pi j}{N}\right) \sin\left(\frac{\pi j}{N} x\right) \sin\left(\frac{\pi j}{N} y\right). \tag{2.13.5}$$

Consider the event $\{T_A \geq n\}$, which is a disjoint union

$$\{T_A \geq n\} = \bigcup_{y=1}^{N-1} \{X_{n \wedge T_A} = y\},$$

since $T_A \geq n$ if and only if $X_{T_A \wedge n} \in A = \{1, \ldots, N-1\}$. We know from the central limit theorem that $\mathbb{P}(T_A \geq n)$ is bounded above by $C/\sqrt{n}$, and hence vanishes in the large n limit. However we can get something stronger out of (2.13.5). We have

$$\mathbb{P}(T_A \geq n | X_0 = x) = \frac{2}{N} \sum_{y=1}^{N-1} \sum_{j=1}^{N-1} \cos^n\left(\frac{\pi j}{N}\right) \sin\left(\frac{\pi j}{N} x\right) \sin\left(\frac{\pi j}{N} y\right).$$

Since

$$\left|\cos\left(\frac{\pi j}{N}\right)\right| \leq \cos\left(\frac{\pi}{N}\right)$$

for each $j = 1, \ldots, N-1$ we see that

$$\mathbb{P}(T_A \geq n | X_0 = x) \leq 2N \left(\cos\left(\frac{\pi}{N}\right)\right)^n$$

and thus that $\mathbb{P}(T_A \geq n | X_0 = x)$ tends to zero exponentially fast as $n \to \infty$.

Suppose we want to consider $\mathbb{P}(X_{T_A} = N \,|X_0 = x)$. Of course, we know this is $\frac{x}{N}$ from our solution of the gambler's ruin problem. However, we can find another interesting formula using (2.13.5). Suppose $T_A = n$ and $X_{T_A} = N$. Then at step $n-1$ the walker must have been at position $N-1$, and at all previous steps the walker was in A. This allows us to write

$$\begin{aligned}
\mathbb{P}(X_{T_A} = N\,|X_0 = x) &= \sum_{n=1}^{\infty} \mathbb{P}(T_A = n \text{ and } X_n = N\,|X_0 = x) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(X_n = N\,|X_{n-1} = N-1)\,\mathbb{P}(X_{T_A \wedge n} = N-1\,|X_0 = x) \\
&= \frac{1}{N} \sum_{n=1}^{\infty} \sum_{j=1}^{N-1} \cos^n\left(\frac{\pi j}{N}\right) \sin\left(\frac{\pi j}{N} x\right) \sin\left(\frac{\pi j (N-1)}{N}\right) \\
&= \frac{1}{N} \sum_{j=1}^{N-1} \frac{\sin\left(\frac{\pi j (N-1)}{N}\right)}{1 - \cos\left(\frac{\pi j}{N}\right)} \sin\left(\frac{\pi j}{N} x\right).
\end{aligned}$$

Of course, this formula looks pretty complicated still, but as we will see in higher dimensions we can use such formulas to extract some information. In the present case the most appealing thing to do is to put this together with our solution to the Gambler's ruin problem to write down the fairly amazing identity:

$$x = \sum_{j=1}^{N-1} \frac{\sin\left(\frac{\pi j (N-1)}{N}\right)}{1 - \cos\left(\frac{\pi j}{N}\right)} \sin\left(\frac{\pi j}{N} x\right)$$

which holds for $x = 0, 1, \ldots, N-1$.

2.14. Lecture 19:

Let us now look at the linear algebra approach to understanding random walks in higher dimensions.

Two dimensions. To begin we consider dimesnion two. Let $A = \{1, \ldots, N-1\}^2 \subset \mathbb{Z}^2$ and suppose the walker starts at a point $\mathbf{x} \in A$. Let

$$T_A = \min\left\{n \,:\, X_n \in \partial^+ A\right\}$$

where $\partial^+ A$ is the outer boundary,

$$\partial^+ A = \{(0,y) \,:\, y = 1, \ldots, N-1\} \cup \{(x,0) \,:\, x = 1, \ldots, N-1\} \cup \{(N,y) \,:\, y = 1, \ldots, N-1\} \cup \{(x,N) \,:\, x = 1, ..$$

Based on what we saw in Lecture 18 we know that

$$\mathbb{P}(X_{n \wedge T_A} = \mathbf{y}\,|X_0 = \mathbf{x}) = Q_A^n \delta_{\mathbf{x}}(\mathbf{y}),$$

where

$$Q_A f(\mathbf{y}) = \frac{1}{4} \sum_{\substack{|\mathbf{y}' - \mathbf{y}| = 1 \\ \mathbf{y}' \in A}} f(\mathbf{y}').$$

Let us try to find the eigenfunctions and eigenvalues of Q_A (which is a symmetric operator). The idea is to guess that there are eigenfunctions of the form

$$f((x_1, x_2)) = g(x_1) h(x_2).$$

This technique is called separation of variables — you may have seen it in a PDE's course. Let's see what needs to happen with g and h for f to be an eigenfunction. Suppose that $g(0) = g(N) = h(0) = h(N) = 0$ so that $f(\mathbf{x}) = 0$ for $\mathbf{x} \in \partial^+ A$. Then we have, for $x \in A$,

$$\begin{aligned} Q_A f((x_1,x_2)) &= \frac{1}{4}\left(g(x_1+1)h(x_2) + g(x_1-1)h(x_2) + g(x_1)h(x_2+1) + g(x_1)h(x_2-1)\right) \\ &= \frac{1}{2}\left(\frac{1}{2}\left(g(x_1+1)+g(x_1-1)\right)h(x_2) + g(x_1)\frac{1}{2}\left(h(x_2+1)+h(x_2-1)\right)\right) \\ &= \frac{1}{2}\left(\left[Q_{1D}g\right](x_1)h(x_2) + g(x_2)\left[Q_{1D}h\right](x_2)\right), \end{aligned}$$

where $Q_{1D}\phi(x) = \frac{1}{2}\left(\phi(x+1)+\phi(x-1)\right)$. To find an eigenfunction $Q_A f = \lambda f$ we need

$$[Q_{1D}g](x_1)h(x_2) + g(x_2)[Q_{1D}h](x_2) = 2\lambda g(x_1)h(x_2), \quad x \in A. \tag{2.14.1}$$

Suppose now that g and h are eigenfunctions of Q_{1D}:

$$Q_{1D}g = \lambda_1 g \quad \text{and} \quad Q_{1D}g = \lambda_2 g.$$

Then (2.14.1) holds with $\lambda = \frac{1}{2}(\lambda_1+\lambda_2)$. Since we already know the eigenfunctions of Q_{1D} from last time, we have just found many eigenfunctions of Q_A.

Let us list the eigenfunctions we have found. The possibilities for g and h are

$$g_j(x_1) = \sqrt{\frac{2}{N}}\sin\left(\frac{\pi j}{N}x_1\right) \quad \text{and} \quad \sqrt{\frac{2}{N}}\sin\left(\frac{\pi j}{N}x_2\right), \quad j = 1,\ldots,N-1.$$

We can take any combination of these, so we define

$$f_{j,k}(x_1,x_2) = g_j(x_1)h_k(x_2) = \frac{2}{N}\sin\left(\frac{\pi j}{N}x_1\right)\sin\left(\frac{\pi k}{N}x_2\right). \tag{2.14.2}$$

Problem 2.28. Show that the functions $f_{j,k}$, $j,k = 1,\ldots,N-1$, are orthonormal in $\mathbb{R}^A$ and that each $f_{j,k}$ is an eigenfunction of Q_A with eigenvalue $\frac{1}{2}\left(\cos\left(\frac{\pi j}{N}\right)+\cos\left(\frac{\pi k}{N}\right)\right)$.

So we have found $(N-1)^2$ orthonormal eigenfunctions. Since the dimension of $\mathbb{R}^A$ is $(N-1)^2$ we must have found all the eigenfunctions!

Remark 2.11. In contrast to the 1D situation, the eigenfunctions are no longer unique. In fact

$$Q_A f_{j,k} = \frac{1}{2}\left(\cos\frac{\pi j}{N} + \cos\frac{\pi k}{N}\right) f_{,jk} = Q_A f_{k,j}$$

so $f_{j,k}$ and $f_{k,j}$ share the same eigenvalue. Of course if $j = k$ then these are the same function, but if $j \neq k$ they are different functions (in fact, orthogonal to one another). This means that for $j \neq k$ there is a *two dimensional* space of eigenfunctions with eigenvalue $\lambda_{j,k} = \frac{1}{2}\left(\cos\frac{\pi j}{N} + \cos\frac{\pi k}{N}\right)$:

$$Q_A\left(af_{j,k} + bf_{k,j}\right) = \lambda_{j,k}\left(af_{j,k} + bf_{k,j}\right)$$

for any $a,b \in \mathbb{R}$. Thus we could replace the pair $f_{j,k}$ and $f_{k,j}$ by, for instance, $\frac{1}{\sqrt{2}}\left(f_{j,k}+f_{k,j}\right)$ and $\frac{1}{\sqrt{2}}\left(f_{j,k}-f_{k,j}\right)$ and still have an orthonormal basis. We will stick with the basis (2.14.2), however.

Thus we have obtained the formula

$$Q_A^n f(x_1,x_2) = \frac{4}{N^2}\sum_{j,k=1}^{N-1}\left[\frac{\cos\frac{\pi j}{N}+\cos\frac{\pi k}{N}}{2}\right]^n \sin\left(\frac{\pi j}{N}x_1\right)\sin\left(\frac{\pi k}{N}x_2\right)\sum_{(y_1,y_2)\in A}\sin\left(\frac{\pi j}{N}y_1\right)\sin\left(\frac{\pi k}{N}y_2\right)f(y_1,y_2).$$

Thus

Theorem 2.19. *Let X_n be a 2D random walk, let $A=\{1,\dots,N-1\}^2$ and let $T_A=\min\{n\ :\ X_n\in\partial^+A\}$. Then*

$$\mathbb{P}\left(X_{n\wedge T_A}=(y_1,y_2)\,|X_0=(x_1,x_2)\right)=\frac{4}{N^2}\sum_{j,k=1}^{N-1}\left[\frac{\cos\frac{\pi j}{N}+\cos\frac{\pi k}{N}}{2}\right]^n\sin\left(\frac{\pi j}{N}y_1\right)\sin\left(\frac{\pi k}{N}y_2\right)\sin\left(\frac{\pi j}{N}x_1\right)\sin\left(\frac{\pi k}{N}x_2\right)$$

As in the 1D case we can use this to compute the exit probability distribution. For instance, since in order for $X_{T_A}=(N,y)$ the walker must arrive at $(N-1,y)$ and hop to the right (which happens with probability $\frac{1}{4}$), we have

$$\begin{aligned}\mathbb{P}\left(X_{T_A}=(N,y)\,|X_0=(x_1,x_2)\right)&=\sum_{n=1}^{\infty}\mathbb{P}\left(T_A=n\text{ and }X_{T_A}=(N,y)\,|X_0=(x_1,x_2)\right)\\&=\sum_{n=1}^{\infty}\frac{1}{4}\mathbb{P}\left(X_{n\wedge T_A}=(N-1,y)\,|X_0=(x_1,x_2)\right)\\&=\frac{1}{N^2}\sum_{j,k}\left[\frac{1}{1-\frac{\cos\frac{\pi j}{N}+\cos\frac{\pi k}{N}}{2}}-1\right]\sin\left(\pi j-\frac{\pi j}{N}\right)\sin\left(\frac{\pi k}{N}y\right)\sin\left(\frac{\pi j}{N}x_1\right)\sin\left(\frac{\pi k}{N}x_2\right)\end{aligned}$$

In the last line we have summed the geometric series $\sum_{n=1}^{\infty}r^n=\frac{1}{1-r}-1$ for $|r|<1$.

Higher dimensions. All of the above carries over to higher dimensions. If we let $A=\{1,\dots,N-1\}^d\subset\mathbb{Z}^d$ and let a walker start at $\mathbf{x}\in A$ then

$$\mathbb{P}\left(X_{n\wedge T_A}=\mathbf{y}\,|X_0=\mathbf{x}\right)=Q_A^n\delta_{\mathbf{x}}(\mathbf{y}),\quad \mathbf{y}\in A,$$

where$T_A=\min\{n\ :\ X_n\in\partial^+A\}$ and

$$Q_Af(\mathbf{x})=\sum_{\substack{\mathbf{x}'\in A\\|\mathbf{x}'-\mathbf{x}|=1}}f(\mathbf{x}').$$

As above the eigenfunctions and eigenvalues of Q_A can be found explicitly:

$$f_{j_1,\dots,j_d}(x_1,\dots,x_d)=\left(\frac{2}{N}\right)^{\frac{d}{2}}\prod_{\alpha=1}^{d}\sin\left(\frac{\pi j_\alpha}{N}x_\alpha\right)\tag{2.14.3}$$

satisfies

$$Q_Af_{j_1,\dots,j_d}=\left(\frac{1}{d}\sum_{\alpha=1}^{d}\cos\left(\frac{\pi j_\alpha}{N}\right)\right)f_{j_1,\dots,j_d}.\tag{2.14.4}$$

Problem 2.29. Show that the function $f_{j_1,\dots,j_d}$ defined in (2.14.3) for $j_\alpha=1,\dots,N-1,\ \alpha=1,\dots,N-1$ are orthonormal and that (2.14.4) holds.

Since (2.14.3) gives exactly $(N-1)^d$ orthonormal eigenvectors, we have found all of them and

$$Q_A^nf(x_1,\dots,x_d)=\frac{2^d}{N^d}\sum_{j_1,\dots,j_d=1}^{N-1}\left[\frac{1}{d}\sum_{\alpha=1}^{d}\cos\frac{\pi j_\alpha}{N}\right]^n\prod_{\alpha=1}^{d}\sin\left(\frac{\pi j_\alpha}{N}x_\alpha\right)\sum_{(y_1,\dots,y_d)\in A}\prod_{\alpha=1}^{d}\sin\left(\frac{\pi j_\alpha}{N}y_\alpha\right)f(y_1,\dots,y_d).$$

Thus

Theorem 2.20. *Let X_n be a random walk in $\mathbb{Z}^d$, let $A=\{1,\ldots,N-1\}^d$ and let $T_A=\min\{n\ :\ X_n\in\partial^+A\}$. Then*

$$\mathbb{P}\left(X_{n\wedge T_A}=(y_1,\ldots,y_d)\,|X_0=(x_1,\ldots,x_d)\right)=\frac{2^d}{N^d}\sum_{j_1,\ldots,j_d=1}^{N-1}\left[\frac{1}{d}\sum_{\alpha=1}^{d}\cos\frac{\pi j_\alpha}{N}\right]^n\prod_{\alpha=1}^{d}\sin\left(\frac{\pi j_\alpha}{N}x_\alpha\right)\sin\left(\frac{\pi j_\alpha}{N}y_\alpha\right).$$

Average number of visits to a site. We can use the formulas derived above to compute interesting formulas for the average number of visits to a site $\mathbf{x}\in\mathbb{Z}^d$. Suppose we start a random walk at site $\mathbf{x}\in A_N=\{1,\ldots,N-1\}^d$. Let

$$V_{\mathbf{y};A_N}=\text{number of visits to }\mathbf{y}\text{ before leaving }A.$$

That is

$$V_{\mathbf{y};A_N}=\sum_{n=0}^{\infty}I\left[X_{n\wedge T_N}=\mathbf{y}\right],$$

where $T_N=T_{A_N}=\min\{n\ :\ X_n\in\partial^+A_N\}$. Thus

$$\mathbb{E}\left(V_{\mathbf{y};A_N}\,|X_0=\mathbf{x}\right)=\sum_{n=0}^{\infty}\mathbb{P}\left(X_{n\wedge T_N}=\mathbf{y}\,|X_0=\mathbf{x}\right),$$

so

$$\mathbb{E}\left(V_{\mathbf{y};A_N}\,|X_0=\mathbf{x}\right)=\frac{2^d}{N^d}\sum_{j_1,\ldots,j_d=1}^{N-1}\frac{1}{1-\frac{1}{d}\sum_{\alpha=1}^{d}\cos\frac{\pi j_\alpha}{N}}\prod_{\alpha=1}^{d}\sin\left(\frac{\pi j_\alpha}{N}x_\alpha\right)\sin\left(\frac{\pi j_\alpha}{N}y_\alpha\right).$$

Average number of visits in 1D. Consider now the 1D case and suppose N is even. If we start X_0 at the site $\frac{N}{2}$ right in the middle of the interval then we have

$$\mathbb{E}\left(V_{\frac{N}{2};A_N}\middle|X_0=\frac{N}{2}\right)=\frac{2}{N}\sum_{j=1}^{N-1}\frac{1}{1-\cos\left(\frac{\pi j}{N}\right)}\sin^2\left(\frac{\pi j}{2}\right).$$

Since $\sin^2(\pi j/2)=1$ if j is odd and 0 otherwise, that is

$$\mathbb{E}\left(V_{\frac{N}{2};A}\middle|X_0=\frac{N}{2}\right)=\frac{2}{N}\sum_{\substack{j=1\\ j\text{ odd}}}^{N-1}\frac{1}{1-\cos\left(\frac{\pi j}{N}\right)}. \tag{2.14.5}$$

The sum on the right hand side looks like a Riemann sum for $\int_0^1\frac{1}{1-\cos(\pi y)}dy$. Indeed, it is precisely the midpoint method sum for the partition $\left\{0,\frac{2}{N},\ldots,\frac{N-2}{N},1\right\}$. (Recall that N is even.) The difficulty here is that the integral $\int_0^1\frac{1}{1-\cos(\pi y)}dy$ is divergent:

Proposition 2.11. $\int_0^1\frac{1}{1-\cos(\pi y)}dy=\infty$.

PROOF. I claim that

$$\cos(y)\geq 1-\frac{y^2}{2},\quad y\in\mathbb{R}. \tag{2.14.6}$$

To see this, note that for $y\geq 0$ we have

$$\cos y-1+\frac{y^2}{2}=\int_0^y(t-\sin t)\,dt=\int_0^y\int_0^t(1-\cos s)\,ds\geq 0.$$

By (2.14.6)

$$\frac{1}{1-\cos\pi y} \geq \frac{2}{\pi^2 y^2}.$$

Thus

$$\int_0^1 \frac{1}{1-\cos\pi y} dy \geq \frac{2}{\pi^2}\int_0^1 \frac{1}{y^2} dy = \infty.$$

□

Therefore it is not immediately clear that the sum on the r.h.s. of (2.14.10) converges to the integral as $N \to \infty$. However, the integral strongly suggests that the sum is divergent in the large N limit.

In fact, we can easily estimate the sum in (2.14.10) by looking at just one term:

$$\frac{2}{N}\sum_{\substack{j=1 \\ j \text{ odd}}}^{N-1} \frac{1}{1-\cos\left(\frac{\pi j}{N}\right)} \geq \frac{4}{\pi^2 N}\sum_{j=1}^{N-1} \frac{N^2}{j^2} \geq \frac{4}{\pi^2} N.$$

It follows that

$$\lim_{N\to\infty} \mathbb{E}\left(V_{\frac{N}{2};A_N} \middle| X_0 = \frac{N}{2}\right) = \int_0^1 \frac{1}{1-\cos\pi y} dy = \infty.$$

Since the random walk probabilities are translation invariant, this is equivalent to

$$\lim_{n\to\infty} \mathbb{E}\left(V_{0;B_n} \middle| X_0 = 0\right) = \int_0^1 \frac{1}{1-\cos\pi y} dy = \infty,$$

where $V_{0;B_n}$ is the number of visits to zero before leaving the interval $B_n = \{-n,\ldots,0,\ldots,n\}$. Using something called the "monotone convergence theorem" this implies that

$$\mathbb{E}\left(V_0 \middle| X_0 = 0\right) = \infty$$

where V_0 is the total number of visits to zero made by the random walk over all time. (We already derived this result in Lecture 10.)

Average number of visits in 2D. Consider now the 2D case, again with N even and again with the initial posisition right in the center of A_N: $X_0 = \left(\frac{N}{2},\frac{N}{2}\right)$. We have

$$\mathbb{E}\left(V_{\left(\frac{N}{2},\frac{N}{2}\right);A_N} \middle| X_0 = \left(\frac{N}{2},\frac{N}{2}\right)\right) = \frac{4}{N^2}\sum_{j,k=1}^{N-1} \frac{1}{1-\frac{1}{2}\left(\cos\left(\frac{\pi j}{N}\right)+\cos\left(\frac{\pi k}{N}\right)\right)} \sin^2\left(\frac{\pi j}{2}\right)\sin^2\left(\frac{\pi k}{2}\right).$$

Again the $\sin^2$ factors force the indices j and k to be odd, so that Since $\sin^2(\pi j/2) = 1$ if j is odd and 0 otherwise, that is

(2.14.7)
$$\mathbb{E}\left(V_{\left(\frac{N}{2},\frac{N}{2}\right);A_N} \middle| X_0 = \left(\frac{N}{2},\frac{N}{2}\right)\right) = \frac{4}{N^2}\sum_{\substack{j,k=1 \\ j,k \text{ odd}}}^{N-1} \frac{1}{1-\frac{1}{2}\left(\cos\left(\frac{\pi j}{N}\right)+\cos\left(\frac{\pi k}{N}\right)\right)}.$$

Again the sum on the right hand side looks like a Riemann sum, this time for the integral

(2.14.8)
$$\int_0^1\int_0^1 \frac{1}{1-\frac{1}{2}(\cos\pi x+\cos\pi y)} dx dy.$$

Proposition 2.12. *The integral in* (2.14.11) *is* ∞.

PROOF. By (2.14.6)

$$\frac{1}{1-\frac{1}{2}(\cos\pi x+\cos\pi y)} \geq \frac{4}{\pi^2(x^2+y^2)}.$$

Since the integrand is positive, we can truncate the region of integration to obtain a lower bound. Thus

$$\begin{aligned}\int_0^1\int_0^1 \frac{1}{1-\frac{1}{2}(\cos\pi x+\cos\pi y)}dxdy &\geq \frac{4}{\pi^2}\iint_{\{x^2+y^2\leq 1\}\cap[0,1]^2}\frac{1}{x^2+y^2}dxdy \\ &= \frac{4}{\pi^2}\int_0^{\frac{\pi}{2}}\int_0^1 \frac{1}{r^2}rdrd\theta \\ &= \frac{2}{\pi}\int_0^1 \frac{1}{r}dr = \infty,\end{aligned}$$

where we have switched to polar coordinates after truncating the domain of integration. □

Again this does not show that the sum in (2.14.10) diverges as $N\to\infty$ but it does suggest divergence. To estimate the sum we must be a little more careful than in the 1D case. We have

$$\frac{4}{N^2}\sum_{\substack{j,k=1\\ j,k\text{ odd}}}^{N-1}\frac{1}{1-\frac{1}{2}\left(\cos\left(\frac{\pi j}{N}\right)+\cos\left(\frac{\pi k}{N}\right)\right)} \geq \frac{16}{\pi^2N^2}\sum_{\substack{j,k=1\\ j,k\text{ odd}}}^{N-1}\frac{N^2}{j^2+k^2} = \frac{16}{\pi^2}\sum_{\substack{j,k=1\\ j,k\text{ odd}}}^{N-1}\frac{1}{j^2+k^2}.$$

Thus, in this case no particular term of the sum diverges. However, we can estimate the sum by comparing to an integral as follows. Note that

$$\frac{1}{j^2+k^2}\geq\frac{1}{x^2+y^2}, \quad j\leq x<j+2,\ k\leq y<k+2.$$

Accounting for the area 4 of the squares $[j,j+2]\times[k,k+2]$ this implies that

$$\sum_{\substack{j,k=1\\ j,k\text{ odd}}}^{N-1}\frac{1}{j^2+k^2}\geq\frac{1}{4}\int_1^{N+1}\int_1^{N+1}\frac{1}{x^2+y^2}dxdy. \tag{2.14.9}$$

Problem 2.30. Transform the integral on the r.h.s. into polar coordinates and truncate the region to prove that

$$\int_1^{N+1}\int_1^{N+1}\frac{1}{x^2+y^2}dxdy\geq C\ln(N+1)$$

for some $C>0$.

Putting it all together we see that

$$\mathbb{E}\left(V_{\left(\frac{N}{2},\frac{N}{2}\right);A_N}\middle|X_0=\left(\frac{N}{2},\frac{N}{2}\right)\right)\geq C\ln N$$

with some constant $C>0$ so that

$$\lim_{N\to\infty}\mathbb{E}\left(V_{\left(\frac{N}{2},\frac{N}{2}\right);A_N}\middle|X_0=\left(\frac{N}{2},\frac{N}{2}\right)\right)=\int_0^1\int_0^1\frac{1}{1-\frac{1}{2}(\cos\pi x+\cos\pi y)}dxdy=\infty.$$

Again this implies, using translation invariance and the monotone convergence theorem,

$$\mathbb{E}(V_0\,|X_0=0)=\infty$$

that where V_0 is the total number of visits to zero made by the random walk over all time. (We derived this result previously in Lecture 11.)

Average number of visits in higher dimensions. Consider $A_N = \{1,\dots,N-1\}^d$ with N even and start a random walk at the center of A_N: $X_0 = \left(\frac{N}{2},\dots,\frac{N}{2}\right)$. Then, accounting for the fact that $\sin^2(\pi j/2) = 1$ if j is odd and 0 otherwise, we see that

$$\mathbb{E}\left(V_{\left(\frac{N}{2},\dots,\frac{N}{2}\right);A_N}\middle| X_0 = \left(\frac{N}{2},\dots,\frac{N}{2}\right)\right) = \frac{2^d}{N^d}\sum_{\substack{j_1,\dots,j_d=1\\ j_\alpha \text{ odd},\ \alpha=1,\dots,d}}^{N-1}\frac{1}{1-\frac{1}{d}\sum_{\alpha=1}^d\cos\left(\frac{\pi j_\alpha}{N}\right)}. \tag{2.14.10}$$

The sum on the right hand side looks like a Riemann sum for the integral

$$\int_0^1\cdots\int_0^1\frac{1}{1-\frac{1}{d}\sum_{\alpha=1}^d\cos\pi x_\alpha}dx_1\cdots dx_d. \tag{2.14.11}$$

Proposition 2.13. *If $d \geq 3$ then the integral in* (2.14.11) *is finite.*

PROOF. First notice that

$$\int\cdots\int_{\left\{\sum_\alpha x_\alpha^2 \geq \delta\right\}\cap[0,1]^d}\frac{1}{1-\frac{1}{d}\sum_{\alpha=1}^d\cos\pi x_\alpha}dx_1\cdots dx_d$$

is finite for any $\delta > 0$ since the integrand is bounded on the domain of integration. Thus it suffices to consider the integral

$$\int\cdots\int_{\left\{\sum_\alpha x_\alpha^2 < \delta\right\}\cap[0,1]^d}\frac{1}{1-\frac{1}{d}\sum_{\alpha=1}^d\cos\pi x_\alpha}dx_1\cdots dx_d.$$

The Taylor expansion of $\cos\pi y$ is

$$\cos\pi y = 1-\frac{\pi^2}{2}y^2+O(y^4).$$

It follows that, given $\varepsilon > 0$ we can find $\delta > 0$ such that

$$\cos\pi y \leq 1-\frac{\pi^2-\varepsilon}{2}y^2$$

if $|y| < \delta$. Fix $\varepsilon < \pi^2$ and choose δ accordingly. If $\sum_\alpha x_\alpha^2 < \delta$ then By (2.14.6)

$$\frac{1}{1-\frac{1}{d}\sum_{\alpha=1}^d\cos\pi x_\alpha} \leq \frac{2d}{(\pi^2-\varepsilon)\sum_\alpha x_\alpha^2}.$$

Thus

$$\int\cdots\int_{\left\{\sum_\alpha x_\alpha^2 < \delta\right\}\cap[0,1]^d}\frac{1}{1-\frac{1}{d}\sum_{\alpha=1}^d\cos\pi x_\alpha}dx_1\cdots dx_d \leq \frac{2d}{\pi^2-\varepsilon}\int\cdots\int_{\left\{\sum_\alpha x_\alpha^2 < \delta\right\}\cap[0,1]^d}\frac{1}{\sum_\alpha x_\alpha^2}dx_1\cdots dx_d.$$

First suppose $d = 3$, then switching to polar coordinates we can evaluate the integral on the r.h.s. as follows

$$\begin{aligned}\iiint_{\left\{x_1^2+x_2^2+x_3^2<\delta\right\}\cap[0,1]^d}\frac{1}{x_1^2+x_2^2+x_3^2}dx_1dx_1dx_3 &= \int_0^\delta\int_0^{\pi/2}\int_0^{\pi/2}\frac{1}{r^2}r^2\sin\theta\, d\phi\, d\theta\, dr\\ &= \frac{\pi}{2}\int_0^\delta dr\\ &= \frac{\pi}{2}\delta,\end{aligned} \tag{2.14.12}$$

This shows that (2.14.11) is finite if $d = 3$.

The final expression obtained in (2.14.12) is $\frac{1}{8}\times$ surface area of the unit sphere $\times \int_0^\delta r^{d-3}dr$. If $d \geq 4$ a similar argument using spherical polar coordinates in higher dimensions shows that

$$\int \cdots \int_{\{\sum_\alpha x_\alpha^2 < \delta\}\cap[0,1]^d} \frac{1}{\sum_\alpha x_\alpha^2} dx_1 \cdots dx_d = \frac{1}{2^d} \times \text{area of a unit hyper sphere in } \mathbb{R}^d \times \int_0^\delta r^{d-3} dr.$$

The main point here is that spherical coordinates in dimension d give a factor $r^{d-1}dr$ in the volume form. This combines with the $1/r^2$ coming from the integrand to produce r^{d-3} which is integrable once $r \geq 3$. Thus (2.14.11) is finite if $d \geq 3$. □

It is also true that (2.14.10) converges to the integral (2.14.11) as $N \to \infty$, although we will not show that. (It requires a bit of work because the integrand is unbounded, so one has to carefully argue about the unbounded piece separately.) Putting it all together we obtain the following

Theorem 2.21. *Consider a random walk X_n in $\mathbb{Z}^d$ and let $V_0 =$ number of visits the walk makes to the origin. Then*

$$\mathbb{E}(V_0 | X_0 = 0) = \int_0^1 \cdots \int_0^1 \frac{1}{1 - \frac{1}{d}\sum_{\alpha=1}^d \cos \pi x_\alpha} dx_1 \cdots dx_d, \tag{2.14.13}$$

where the integral on the r.h.s. is finite if $d \geq 3$ and infinite if $d \leq 2$.

Remark 2.12. It follows that the random walk is transient if $d \geq 3$. It *does not* follow immediately that the random walk is recurrent if $d = 1$ or 2 but one can prove this after some additional work, or by different means as we did in Lecture 10 and Lecture 11.

CHAPTER 3

Percolation

To come....

APPENDIX A

A.1. Review of Riemann Integration

This is a quick review of Riemann integration, mostly aimed at those who have seen this before but might have forgotten some things. If you haven't seen integration done rigorously before, you would probably benefit from reading more details in an analysis book, such as Principles of Mathematical Analysis by Walter Rudin.

A *piecewise constant function* is a function ϕ which is a finite linear combination of characteristic functions of intervals:

$$\phi(x) = \sum_{j=1}^{n} c_j \chi_{I_j}(x), \tag{A.1.1}$$

where, for each $j = 1, \dots, n$, c_j is a real number and

$$I_j = [a_j, b_j] \quad \text{or} \quad [a_j, b_j) \quad \text{or} \quad (a_j, b_j] \quad \text{or} \quad (a_j, b_j)$$

is an interval, with $-\infty \le a_j \le b_j \le \infty$. From the standpoint of integration it makes no difference what the function does at the interval end points, a_j and b_j. We will say that ϕ is *compactly supported* if $\phi(x) = 0$ for $|x|$ sufficiently large. Equivalently all the intervals I_j in (A.1.1) are bounded.

Definition A.1. (1) Let PC = {compactly supported piecewise constant functions} . The *integral* of $\phi \in$ PC over an interval $[a,b]$ is

$$\int_a^b \phi(x)dx := \sum_{j=1}^{n} c_j \, |I_J \cap [a,b)|$$

where the *length of an interval* $I = [c,d]$ or $[c,d)$ or $(c,d]$ or (c,d) is

$$|I| = d - c.$$

(2) A function f is *bounded* on $[a,b]$ if $\sup_{x\in[a,b]} |f(x)| < \infty$. If f is bounded, we define the *upper and lower integrals of f over* $[a,b]$ to be

$$\overline{\int_a^b} f(x)dx = \inf\left\{ \int_a^b \phi(x)dx \;:\; \phi \in PC \text{ and } f(x) \le \phi(x) \text{ for all } x \in [a,b] \right\}$$

and

$$\underline{\int_a^b} f(x)dx = \sup\left\{ \int_a^b \phi(x)dx \;:\; \phi \in PC \text{ and } f(x) \ge \phi(x) \text{ for all } x \in [a,b] \right\}.$$

(3) We say a function f is *Riemann integrable on* $[a,b]$ if f is *bounded* and the upper and lower integrals agree, in which case we define $\int_a^b f(x)dx$ to be their common value:

$$\int_a^b f(x)dx = \overline{\int_a^b} f(x)dx = \underline{\int_a^b} f(x)dx.$$

In calculus you learn that the integral is a limit of Riemann sums as the mesh size of the partition for the sum goes to zero. More precisely, recall that a Riemann sum for $\int_a^b f(x)dx$ is a finite sum of the form

$$S(f,\mathscr{P};\mathscr{S}) = \sum_{j=1}^{n} f(y_j)|x_j - x_{j-1}|,$$

where $\mathscr{P} = \{x_0,\dots,x_n\}$ with $a = x_0 < x_1 < \cdots < x_n = b$ is called a *partition of* $[a,b]$ and the n-tuple $\mathscr{S} = (y_1,\dots,y_n)$, called the *sample set,* is *subordinate* to $\mathscr{P}$, meaning $x_{j-1} \le y_j \le x_j$ for each $j = 1,\dots,n$. The *mesh size of* $\mathscr{P}$, denoted $\Delta\mathscr{P}$, is

$$\Delta\mathscr{P} = \max_j \left|x_j - x_{j-1}\right|.$$

The point of Definition.A.1 is that any Riemann sum can be sandwiched between two integrals of simple functions:

$$\int_a^b \phi^-(x)dx \le S(f,\mathscr{P};y_1,\dots,y_n) \le \int_a^b \phi^+(x)dx,$$

where

$$\phi^-(x) = \sum_{j=1}^{n} \left(\inf_{y\in[x_{j-1},x_j)} f(y)\right) \chi_{[x_{j-1},x_j)}(x) + f(b)\chi_{\{b\}}(x)$$

and

$$\phi^+(x) = \sum_{j=1}^{n} \left(\sup_{y\in[x_{j-1},x_j)} f(y)\right) \chi_{[x_{j-1},x_j)}(x) + f(b)\chi_{\{b\}}(x).$$

This observation is a key point in the proof of the following

Theorem A.1. *Let f be Riemann integrable on $[a,b]$, and let $\mathscr{P}_n$ be a sequence of partitions of $[a,b]$. If $\Delta\mathscr{P}_n \to 0$ and $\mathscr{S}_n$ is any sequence of sample sets subordinate to $\mathscr{P}$ then*

$$\lim_{n\to\infty} S(f;\mathscr{P}_n,\mathscr{S}_n) = \int_a^b f(x)dx.$$

PROOF. Exercise. (Or look it up in an analysis book!) □

Here are some easily verified familiar facts about Riemann integration

Proposition A.1. *Let f and g be Riemann integrable on $[a,b]$ and let $c \in \mathbb{R}$ Then*

(1) *$f + cg$ is Riemann integrable on $[a,b]$ and $\int_a^b (f(x)+cg(x))\,dx = \int_a^b f(x)dx + c\int_a^b g(x)dx$.*
(2) *If $f(x) \le g(x)$ for all x then*

$$\int_a^b f(x)dx \le \int_a^b g(x)dx.$$

(3) *The constant funciton $\phi(x) = c$ is Riemann integrable and $\int_a^b \phi(x)dx = c(b-a)$*
(4) $\min(f(x),c)$ *and* $\max(f(x),c)$ *are Riemann integrable on $[a,b]$.*
(5) *$|f|$ is Riemann integrable on $[a,b]$ and $\left|\int_a^b f(x)dx\right| \le \int_a^b |f(x)|\,dx$.*

Remark A.1. Since $\min(f(x),c) \le f(x) \le \max(f(x),c)$ if follows from 2 and 3 that

$$\int_a^b \min(f(x),c)\,dx \le \int_a^b f(x)dx \le \int_a^b \max(f(x),c)\,dx.$$

Since $|f(x)| = \max(f(x),0) - \min(f(x),0)$, 5 follows from 4 and 1.

PROOF. Exercise. □

A useful fact about Riemann integration is that it interacts nicely with uniform convergence:

Theorem A.2. *Let f_n be a sequence of functions that are Riemann integrable on $[a,b]$. If $f_n \to f$ uniformly on $[a,b]$, that is if*

$$\lim_{n\to\infty} \sup_{x\in[a,b]} |f_n(x) - f(x)| = 0,$$

then f is Riemann integrable on $[a,b]$ and

$$\lim_{n\to\infty} \int_a^b f_n(x)dx = \int_a^b f(x)dx.$$

PROOF. Let $\varepsilon > 0$. Since $f_n \to f$ uniformly there is $N \in \mathbb{N}$ such that $n \ge N$ implies

$$f_n(x) - \frac{\varepsilon}{2(b-a)} \le f(x) \le f_n(x) + \frac{\varepsilon}{2(b-a)} \quad \text{for all } x \in [a,b].$$

Since $f_n + \frac{\varepsilon}{2(b-a)}$ is Riemann integrable on $[a,b]$ we can find $\phi_+ \in \mathrm{PC}$ such that

$$\phi_+(x) \ge f_n(x) + \frac{\varepsilon}{2(b-a)} \ge f(x) \quad \text{for all } x \in [a,b].$$

Since $\phi(x) \ge f(x)$ for all $x \in [a,b]$ we conclude that

$$\overline{\int_a^b} f(x)dx \le \int_a^b \phi(x)dx \le \int_a^b \left(f_n(x) + \frac{\varepsilon}{2(b-a)}\right)dx + \frac{\varepsilon}{2} = \int_a^b f_n(x)dx + \varepsilon,$$

for $n \ge N$. In a similar way, we can see that

$$\underline{\int_a^b} f(x)dx \ge \int_a^b f_n(x)dx - \varepsilon,$$

for $n \ge N$.

We conclude that

(1) $0 \le \overline{\int_a^b} f(x)dx - \underline{\int_a^b} f(x)dx \le 2\varepsilon$ for arbitrary ε so f is Riemann integrable on $[a,b]$,
(2) For every ε there is $N \in \mathbb{N}$ such that if $n \ge N$ then $\left|\int_a^b f_n(x)dx - \int_a^b f(x)dx\right| \le \varepsilon$, i.e. $\int_a^b f_n(x)dx \to \int_a^b f(x)dx$.

□

Corollary A.1. *Suppose there is a sequence $\phi_n \in \mathrm{PC}$ such that $\phi_n \to f$ uniformly on $[a,b]$. Then f is Riemann integrable on $[a,b]$.*

Problem A.1. Let f be continuous on $[a,b]$. Show that there is a sequence $\phi_n \in \mathrm{PC}$ that converges uniformly to f. Thus all continuous functions are Riemann integrable. (Hint: let $\phi_n = \sum_{j=1}^n f(x_j)\chi_{[x_{j-1},x_j)} + f(b)\chi_{\{b\}}$ with $x_j = a + \frac{j}{n}(b-a)$ and use the fact that f is *uniformly continuous.* Why is f uniformly continuous?)

Improper integrals. To consider integrals over the whole real line we take limits.

Definition A.2. If $f:\mathbb{R}\to\mathbb{R}$ is Riemann integrable on each $[a,b]\subset\mathbb{R}$ and if for some $c\in\mathbb{R}$

$$\lim_{b\to\infty}\int_c^b f(x)dx \quad\text{and}\quad \lim_{a\to-\infty}\int_a^c f(x)dx \tag{A.1.2}$$

exist, the we define the integral of f over $\mathbb{R}$ to be

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{b\to\infty}\int_c^b f(x)dx + \lim_{a\to-\infty}\int_a^c f(x)dx. \tag{A.1.3}$$

Proposition A.2. *If* (A.1.2) *holds with $c=c_0\in\mathbb{R}$ then it holds for all values of c and* (A.1.3) *is independent of the choice of c.*

The proof is an easy consequence of

Proposition A.3. *Let $a<c<b$ be real numbers. Then a function f is Riemann integrable on $[a,b]$ if and only if f is Riemann integrable on $[a,c]$ and $[c,b]$. Furthermore*

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

PROOF. Exercise □

Proof of dominated convergence. In Lecture 8 we used a Dominated Convergence Theorem for Riemann integration:

Theorem A.3 (Dominated Convergence)**.** *Let f_n be a sequence of functions and let g be a non-negative function, all on the real line. If*

(1) *f_n and g are Riemann integrable on compact intervals,*
(2) *$|f_n(x)|\le g(x)$ for all x,*
(3) *$\int_{-\infty}^{\infty} g(x)dx<\infty$, and*
(4) *$\lim_{n\to\infty} f_n(x)=f(x)$ exists for every x, and f is Riemann integrable on compact intervals.*

Then

$$\int_{-\infty}^{\infty} f(x)dx = \lim_{n\to\infty}\int_{-\infty}^{\infty} f_n(x)dx.$$

Remark A.2. You won't find this theorem in most analysis books. It is based on a theorem due to Arzela in 1885. The proof given below is based on

- Luxemburg, W. A. J., Arzela's Dominate Convergence Theorem for the Riemann Integral, The American Math Monthly, Vol. 78 (1971), pp. 970-979.

PROOF. First since

$$\left|\int_{-\infty}^{\infty} f_n(x)dx - \int_{-\infty}^{\infty} f(x)dx\right| \le \int_{-\infty}^{\infty} |f_n(x)-f(x)|\,dx,$$

it suffices to prove the result supposing $f\equiv 0$ and $f_n\ge 0$. (Replace f_n by $|f_n-f|$ and replace g by $2g$.) Furthermore, since $\int_{-\infty}^{\infty} g(x)dx<\infty$, given $\varepsilon>0$ we can find $M<\infty$ such that

$$\int_{-\infty}^{-M} g(x)dx + \int_M^{\infty} g(x)dx < \varepsilon.$$

Thus

$$\int_{-\infty}^{-M} |f_n|(x)dx + \int_{M}^{\infty} |f_n|(x)dx < \varepsilon.$$

So it suffices to prove the result for f_n which are zero outside a bounded interval and uniformly bounded.

Thus we assume $f_n \geq 0$ on $[-M,M]$, with $f_n(x) \leq A$ for all x,n, and $f_n(x) \to 0$ pointwise. Consider the functions

$$g_n(x) = \sup_{m \geq n} f_m(x).$$

Clearly $f_n \leq g_n$ and $g_n(x)$ converges monotonically to 0 pointwise. Thus the theorem follows from Lemma A.1 below. □

Lemma A.1. *Let g_n be a sequence of non-negative functions on $[a,b]$ such that*

(1) *g_1 is bounded and $g_{n+1}(x) \leq g_n(x)$ for all $x \in [a,b]$, and*
(2) *$\lim_{n\to\infty} g_n(x) = \inf_{n \geq 0} g_n(x) = 0$ for all $x \in [a,b]$.*

Then,

$$\lim_{n\to\infty} \underline{\int_a^b} g_n(x)dx = \inf_n \underline{\int_a^b} g_n(x) = 0.$$

PROOF. Let $\varepsilon > 0$. By Lemma A.2 below, we can find for each n a continuous function h_n on $[-M,M]$ such that $0 \leq h_n(x) \leq g_n(x)$ and

$$\underline{\int_a^b} g_n(x) \leq \int_a^b h_n(x)dx + \frac{\varepsilon}{2^n}.$$

Let

$$\psi_n(x) = \min_{1\leq k \leq n} h_k(x).$$

Then

(1) $\psi_n(x) \leq h_n(x) \leq g_n(x)$ for each $x \in [a,b]$,
(2) $\psi_n(x)$ decreases monotonically to zero for each $x \in [a,b]$, and
(3) ψ_n is continuous .

It follows from Dini's Theorem (Thm. A.4 below) that $\psi_n \to 0$ uniformly. Thus $\int_a^b \psi_n(x)dx \to 0$. However, for each $i \leq n$

$$h_n = h_i + (h_n - h_i) \leq h_i + \max_{k=i,\dots,n} h_k - h_i \leq h_i + \sum_{j=1}^{n-1} \left(\max_{k=j,\dots,n} h_k - h_j \right).$$

Taking the minimum over $i \leq n$ on the r.h.s. gives

$$h_n \leq \psi_n + \sum_{j=1}^{n-1} \left(\max_{k=j,\dots,n} h_k - h_j \right).$$

Since

$$\max_{k=j,\dots,n} h_k \leq \max_{k=j,\dots,n} g_k = g_j$$

we conclude that

$$\begin{aligned}\underline{\int_a^b} g_n(x)dx \le \int_a^b h_n(x)dx + \frac{\varepsilon}{2^n} &\le \frac{\varepsilon}{2^n} + \int_a^b \psi_n(x)dx + \sum_{j=1}^{n-1}\int_a^b \max_{k=j,\dots,n} h_k(x)dx - \int_a^b h_j(x)dx \\ &\le \frac{\varepsilon}{2^n} + \int_a^b \psi_n(x)dx + \sum_{j=1}^{n-1}\underline{\int_a^b} g_j(x)dx - \int_a^b h_j(x)dx \\ &\le \int_a^b \psi_n(x)dx + \frac{\varepsilon}{2^n} + \sum_{j=1}^{n-1}\frac{\varepsilon}{2^j} \\ &\le \int_a^b \psi_n(x)dx + \varepsilon.\end{aligned}$$

Since $\lim_n \int_a^b \psi_n dx = 0$, we have

$$\limsup_{n\to\infty} \underline{\int_a^b} g_n(x)dx \le \varepsilon$$

and the Lemma follows. □

Lemma A.2. *Let f be a bounded non-negative function on $[a,b]$ and let $\varepsilon > 0$. Then there exists a non-negative continuous function h such that $0 \le h(x) \le f(x)$ on $[a,b]$ and*

$$\underline{\int_a^b} f(x)dx \le \int_a^b h(x)dx + \varepsilon.$$

PROOF. By definition of the lower integral we can find a simple function ϕ such that

$$\int_a^b \phi(x) \le \underline{\int_a^b} f(x)dx \le \int_a^b \phi(x) + \frac{\varepsilon}{2}.$$

Since f is non-negative we may take ϕ to be non-negative, so

$$\phi(x) = \sum_{j=1}^{n} w_j \chi_{I_j} \tag{A.1.4}$$

with $w_j \ge 0$.

Let $\mathscr{I} = \int_a^b \phi(x)dx = \sum_j w_j |I_j|$. If $\varepsilon \ge 2\mathscr{I}$ then the continuous function $h(x) = 0$ will do the job. So suppose $\varepsilon < 2\mathscr{I}$. Each interval I_j in (A.1.4) has endpoints a_j and b_j, midpoint $m_j = \frac{b_j+a_j}{2}$ and length $|I_j| = b_j - a_j$. Let

$$\widetilde{I}_j = \left[m_j - r_j, m_j + r_j\right],$$

with

$$r_j = \left(1 - \frac{\varepsilon}{2\mathscr{I}}\right)\frac{|I_j|}{2},$$

so that $\left|\widetilde{I}_j\right| = \left(1 - \frac{\varepsilon}{2\mathscr{I}}\right)\left|I_j\right|$. Now let

$$\psi_j(x) = \begin{cases} 0 & x \notin I_j \\ 1 & x \in \widetilde{I}_j \\ 1 - \frac{4\mathscr{I}\operatorname{dist}(x,\widetilde{I}_j)}{\varepsilon} & x \in I_j \setminus \widetilde{I}_j. \end{cases}$$

So ψ_j increases linearly from 0 at the end points of I_j to take a value of 1 on $\widetilde{I}_j$. In particular ψ_j is continuous. Thus

$$h(x) = \sum_{j=1}^{n} w_j \psi_j(x)$$

is continuous, $h(x) \leq \phi(x) \leq f(x)$ and

$$\begin{aligned}\int_a^b h(x)dx + \varepsilon \geq \varepsilon + \sum_{j=1}^{n} w_j \left|\widetilde{I}_j\right| &= \varepsilon + \left(1 - \frac{\varepsilon}{2\mathscr{I}}\right) \sum_{j=1}^{n} w_j \left|I_j\right| \\ &= \frac{\varepsilon}{2} + \int_a^b \phi(x)dx \geq \int_a^b f(x)dx\end{aligned}.$$

□

Appendix: Dini's Theorem. Dini's Theorem is a result giving uniform convergence for monotonically convergent sequence of continuous functions on a compact set. It is true in great generality, but the version we need here is the following,

Theorem A.4 (Dini's Theorem)**.** *Let g_n be a sequence of non-negative continuous functions on a compact interval $[a,b]$ that decrease monotonically to zero: $g_n(x) \geq g_{n+1}(x)$ and $\lim_{n\to\infty} g_n(x) = 0$ for all $x \in [a,b]$. Then $g_n \to 0$ uniformly.*

PROOF. Let $\varepsilon > 0$. For each $n \in \mathbb{N}$

$$U_n = \{x \in [a,b] \ : \ g_n(x) < \varepsilon\}.$$

Because g_n is continuous, U_n is a relatively open subset of $[a,b]$. Since g_n decrease we have $U_n \subset U_{n+1}$. Since $\lim_n g_n(x) = 0$ for all $x \in [a,b]$, we have $[a,b] = \cup_n U_n$. Thus, $\{U_n\}$ is an open cover so by compactness of $[a,b]$ we can find a finite subcover: $U_{n_1} \cup \cdots \cup U_{n_k} = [a,b]$. Since the U_n are increasing, we have $U_N = [a,b]$ where $N = \max\left\{n_j : j = 1,\ldots,k\right\}$. Thus for $n \geq N$ we have $U_n = [a,b]$, i.e. $g_n(x) < \varepsilon$ for every $x \in [a,b]$. □

A.2. Taylor's Theorem

Taylor's Theorem. In Lecture 6 we made use of Taylor's Theorem. Here is the statement and a proof.

Theorem A.5 (Taylor's Theorem with integral remainder)**.** *Let f be a C^{n+1} function on an open interval $I \subset \mathbb{R}$. Let $x_0 \in I$. Then*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_n(x), \tag{A.2.1}$$

where

$$R_n(x) = \frac{(x - x_0)^{n+1}}{n!} \int_0^1 f^{(n+1)}(x_0 + t(x - x_0))(1 - t)^n dt. \tag{A.2.2}$$

Remark A.3. 1) A C^{n+1} function is an $n+1$ times differentiable function with continuous $n+1$-st derivative. (It follows that f and it's first n derivatives are also continuous.), 2) $f^{(k)}$ denotes the k^{th} derivative of f, with $f^{(0)} = f$, 3) As usual $0! = 1$.

PROOF. Note that the case $n = 0$ is just the fundamental theorem of calculus, with a change of variables:

$$f(x) = f(x_0) + \int_{x_0}^{x} f^{(1)}(y)dy = f(x_0) + (x - x_0)\int_0^1 f^{(1)}(x_0 + t(x - x_0))dt.$$

We prove the result for $n > 0$ by induction. Suppose (A.2.1) holds for some value of n and let f be a C^{n+2} function. Then f is also C^{n+1}, so (A.2.1) holds with $R_n(x)$ given by (A.2.2). Integrating by parts we see that

$$\begin{aligned} R_n(x) &= \frac{(x-x_0)^{n+1}}{n!}\int_0^1 f^{(n+1)}(x_0 + t(x-x_0))(1-t)^n dt \\ &= -\frac{(x-x_0)^{n+1}}{(n+1)!}\int_0^1 f^{(n+1)}(x_0 + t(x-x_0))\frac{d(1-t)^{n+1}}{dt}dt \\ &= \frac{f^{(n+1)}(x_0)}{(n+1)!} + \frac{(x-x_0)^{n+2}}{(n+1)!}\int_0^1 f^{(n+2)}(x_0 + t(x-x_0))(1-t)^{n+1}dt. \end{aligned}$$

Replacing $R_n(x)$ by the final expression on the r.h.s. gives (A.2.1) with n replaced by $n+1$. The result follows by induction. □

We used:

Theorem A.6 (Integration by Parts)**.** *If f and g are C^1 on an open interval containing $[a,b]$ then*

$$\int_a^b f(x)g^{(1)}(x)dx = f(b)g(b) - f(a)g(a) - \int_a^b f^{(1)}(x)g(x)dx.$$

PROOF. Apply the fundamental theorem of calculus to $\frac{d}{dx}[f(x)g(x)] = f^{(1)}(x)g(x) + f(x)g^{(1)}(x)$. □

The integral remainder term expressed in Taylor's theorem above is a nice exact expression. However, for making estimates it is often useful to use a less precise expression.

Theorem A.7. *Let f be a C^{n+1} function on an open interval $I \subset \mathbb{R}$. Then for any closed interval $K \subset I$ there is a finite constant $c(K)$ such that*

$$\left| f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \right| \le c(K)\,|x-x_0|^{n+1}, \quad x, x_0 \in K. \tag{A.2.3}$$

PROOF. By (A.2.1) and (A.2.2) we have, for any $x, x_0 \in I$

$$\begin{aligned} \left| f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \right| &\le \left[\sup_{t\in[0,1]} |f(x_0 + t(x-x_0))|\right]\int_0^1 (1-t)^n dt \\ &= \left[\sup_{t\in[0,1]} |f(x_0 + t(x-x_0))|\right]\frac{|x-x_0|^{n+1}}{(n+1)!}. \end{aligned}$$

If $x, x_0 \in K$ then the interval $\{x_0 + t(x-x_0) \,:\, t \in [0,1]\} \subset K$, so we have (A.2.3) with

$$c(K) = \frac{1}{(n+1)!}\sup_{y\in K}\left|f^{(n+1)}(y)\right|.$$

□

"Big Oh" and "little oh". Statements like (A.2.3) are very common and it is convenient to have a short hand for them that can be expressed more succinctly. This is accomplished by the "Big Oh" and "little oh" notation. If we write

$$f(t) = g(t) + \mathrm{O}(t) \tag{A.2.4}$$

where t is a variable that we are thinking of as small, what we mean is

> There exists a finite positive constant C such that
> $$|f(t) - g(t)| \leq Ct.$$

Thus (A.2.3) can be written

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \mathrm{O}\left(|x - x_0|^{n+1}\right), \quad x, x_0 \in K.$$

It will be convenient to use this shorthand now and then and it is a good idea for you to get comfortable with it.

Eq. (A.2.5) is read "f of t equals g of t plus big oh of t." As you might guess there is also a "little oh." If we write

$$f(t) = g(t) + \mathrm{o}(t) \tag{A.2.5}$$

where t is a variable that we are thinking of as small, what we mean is

$$\lim_{\substack{t \to 0 \\ t \neq 0}} \frac{f(t) - g(t)}{t} = 0.$$

In other words

> For every $\varepsilon > 0$ there is a $\delta > 0$ such that if $|t| < \delta$ then
> $$|f(t) - g(t)| < \varepsilon t.$$

This notation is used in the same way when applied to a parameter that gets large and also to a discrete parameter. Thus we might write

$$\sqrt{n} = \mathrm{o}(n)$$

or

$$\left(1 + \frac{x}{n}\right)^n = e^x + \mathrm{O}\left(\frac{1}{\mathrm{n}^2}\right).$$

This notation can considerably simplify statements. For instance consider the definition of "f is differentiable at x." Typically we would write

> A function f is differentiable at a point x if there exists a number A such that for every $\varepsilon > 0$ there is $\delta > 0$ so that whenever $|y - x| < \delta$ we have
> $$|f(y) - f(x) - A(y - x)| < \varepsilon |y - x|.$$

With "little oh" we could write instead

> A function f is differentiable at a point x if there exists a number A such that
> $$f(y) = f(x) + A(y - x) + \mathrm{o}(|y - x|).$$

Not only is this shorter, it more clearly gets the point across that we are making a linear approximation to $f(y)$.

To give one more example, let's write down another corollary to Taylor's Theorem

Corollary A.2. *Let f be a C^n function on an open interval $I \subset \mathbb{R}$. Let $x_0 \in I$. Then*

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + \mathrm{o}\left(|x-x_0|^n\right).$$

Remark A.4. The point here is that f may not have an $n+1$-st derivative, so the estimate on the error is much weaker: $\mathrm{o}(|\mathrm{x}-\mathrm{x}_0|^{\mathrm{n}})$ instead of $\mathrm{O}(|x-x_0|^{n+1})$.

PROOF. Apply Taylor's Theorem with n replaced by $n-1$. The remainder term

$$\begin{aligned} R_{n-1}(x) &= \frac{(x-x_0)^n}{(n-1)!}\int_0^1 f^{(n)}(x_0+t(x-x_0))(1-t)^{n-1}dt \\ &= \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n + \frac{(x-x_0)^n}{(n-1)!}\int_0^1 \left[f^{(n)}(x_0+t(x-x_0)) - f^{(n)}(x_0)\right](1-t)^{n-1}dt. \end{aligned}$$

Since $f^{(n)}$ is assumed to be continuous, the term in square brackets is $\mathrm{o}(|x-x_0|)$, giving the result. □

Problem A.2. Why is $(x-x_0)^n \mathrm{o}(|x-x_0|) = \mathrm{o}(|x-x_0|^{n+1})$?

A.3. Root's of Unity and Trig Functions

Euler's formula states that

$$e^{\mathrm{i}\phi} = \cos\phi + \mathrm{i}\sin\phi, \tag{A.3.1}$$

where $\mathrm{i} = \sqrt{-1}$. This identity is easily seen to hold using the power series for the exponential:

$$e^{\mathrm{i}\phi} = \sum_{n=0}^{\infty}\frac{1}{n!}\mathrm{i}^n\phi^n = \underbrace{\sum_{m=0}^{\infty}\frac{(-1)^m}{(2m)!}\phi^{2m}}_{\cos\phi} + \mathrm{i}\underbrace{\sum_{m=0}^{\infty}\frac{(-1)^m}{(2m+1)!}\phi^{2m+1}}_{\sin\phi}.$$

Since $e^z e^w = e^{z+w}$ for any complex numbers z, w, Euler's formula (A.3.1) makes short work of deriving trigonometric identities. For example

$$\cos(2\phi) + \mathrm{i}\sin(2\phi) = e^{\mathrm{i}2\phi} = (\cos\phi + \mathrm{i}\sin\phi)^2 = \cos^2\phi - \sin^2\phi + 2\mathrm{i}\cos\phi\sin\phi,$$

from which we conclude that

$$\cos(2\phi) = \cos^2\phi - \sin^2\phi \quad \text{and} \quad \sin(2\phi) = 2\cos\phi\sin\phi.$$

Given a positive integer n, the complex numbers

$$\omega_j = e^{\mathrm{i}\frac{2\pi j}{n}}, \quad j = 0, \ldots, n-1$$

all satisfy

$$\omega_j^n = e^{\mathrm{i}2\pi j} = 1$$

since $\cos(2\pi j) = 1$ and $\sin(2\pi j) = 0$. Thus the numbers ω_j are roots of the polynomial $z^n - 1$. Since there are n of them, these are all the roots. Thus we have the identity

$$z^n - 1 = \prod_{j=0}^{n-1}\left(z - \omega_j\right). \tag{A.3.2}$$

Expanding the right hand side of (A.3.2) leads to interesting identities. Since

$$\prod_{j=0}^{n-1}\left(z-\omega_j\right)=\sum_{A\subset\{0,\ldots,n-1\}}\left((-1)^{|A|}\prod_{j\in A}\omega_j\right)z^{n-|A|},$$

we conclude from (A.3.2)that

$$\sum_{\substack{A\subset\{0,\ldots,n-1\}\\|A|=m}}\prod_{j\in A}\omega_j=0, \tag{A.3.3}$$

if $m=1,\ldots,n-1$, while

$$\prod_{j=0}^{n-1}\omega_j=(-1)^{n-1}.$$

The last identity could also be seen by noting that

$$\prod_{j=0}^{n-1}\omega_j=e^{\mathrm{i}\frac{2\pi}{n}\sum_{j=0}^{n-1}j}=e^{\mathrm{i}\pi(n-1)}=(-1)^{n-1}$$

since $e^{\mathrm{i}\pi}=-1$. (To derive (A.3.3) from (A.3.2), recall that two polynomials are equal as functions if and only if their coefficients are equal.)

Taking (A.3.3) with $m=1$ gives

$$0=\sum_{j=0}^{n-1}\omega_j,$$

which shows the two identities

$$\sum_{j=0}^{n-1}\cos\left(\frac{2\pi j}{n}\right)=0=\sum_{j=0}^{n-1}\sin\left(\frac{2\pi j}{n}\right).$$

Proposition A.4. *Let n be a positive integer. Then for any $m\in\{1,\ldots,n-1\}$ we have*

$$\sum_{j=0}^{n-1}\cos\left(\frac{2\pi mj}{n}\right)=0=\sum_{j=0}^{n-1}\sin\left(\frac{2\pi mj}{n}\right). \tag{A.3.4}$$

PROOF. We have already proved this for $m=1$. A similar argument shows that (A.3.4) in case the greatest common divisor of m and n is 1. Indeed we have the following

Lemma A.3. *Let $m,n\in\mathbb{N}$ with $m<n$. If $\gcd(m,n)=1$ then for any $j,k=0,\ldots,n-1$ we have*

$$mj\equiv mk \mod n$$

if and only $j=k$.

PROOF. Clearly if $j=k$ then $mj\equiv mk \mod n$. Suppose $mj\equiv mk \mod n$, and without loss assume $j\le k$. Then $mj=an+r$ and $mk=bn+r$ with $a,b,r\in\mathbb{N}$ and $a\le b$. Thus

$$m(k-j)=(b-a)n.$$

Since $\gcd(m,n)=1$ we have $m|(b-a)$. Thus

$$k-j=cn$$

for some $c\in\mathbb{N}$. Since $k-j<n$ this can only hold if $c=0$. Thus $k=j$. □

It follows that

$$e^{\mathrm{i}\frac{2\pi m}{n}j}, \quad j = 0, \ldots, n-1 \tag{A.3.5}$$

are all distinct numbers, since if

$$e^{\mathrm{i}\frac{2\pi m}{n}j} = e^{\mathrm{i}\frac{2\pi m}{n}k}$$

then $2\pi\left(\frac{mj}{n} - \frac{mk}{n}\right)$ is a multiple of 2π, so $mj \equiv mk \mod n$. Since the numbers (A.3.5) are all n^{th} roots of unity, and there are n of them, we may argue as above to conclude

$$\sum_{j=0}^{n-1} e^{\mathrm{i}\frac{2\pi mj}{n}} = 0.$$

(The point is that the list $1, e^{\mathrm{i}\frac{2\pi m}{n}}, \ldots, e^{\mathrm{i}\frac{2\pi m}{n}(n-1)}$ is a rearrangement of the list $1, e^{\mathrm{i}\frac{2\pi}{n}}, \ldots, e^{\mathrm{i}\frac{2\pi}{n}(n-1)}$.)

So, we have established (A.3.4) for $\gcd(m,n) = 1$. However, at this point we are done since we can always reduce the fraction $\frac{m}{n}$. That is, we have

$$\frac{m}{n} = \frac{m'}{n'}$$

with $\gcd(m',n') = 1$. □

Corollary A.3. *Let n be a positive integer. Then for any $m, m' \in \{1, \ldots, n-1\}$ we have*

$$\sum_{j=1}^{n-1} \sin\left(\frac{\pi m}{n}j\right)\sin\left(\frac{\pi m'}{n}j\right) = \begin{cases} \frac{n}{2} & \text{if } m = m', \\ 0 & \text{if } m \neq m'. \end{cases} \tag{A.3.6}$$

Remark A.5. Thus the functions $h_m(x) = \sqrt{\frac{2}{n}}\sin\left(\frac{\pi m}{n}x\right)$, $m = 1, \ldots, n-1$ are orthonormal in $\mathbb{R}^{\{1,\ldots,n-1\}}$. Since there are $n-1$ functions and this is the dimension of the vector space the set $h_1, \ldots, h_{n-1}$ is an orthonormal basis. (See Lecture 18.)

PROOF. First note that since $\sin 0 = 0$ we can include the term corresponding to $j = 0$ on the left hand side of (A.3.6) without changing things. We use the trig identity

$$\sin\phi\sin\theta = \frac{1}{2}\left(\cos(\phi-\theta) - \cos(\phi+\theta)\right) \tag{A.3.7}$$

to rewrite the left hand side of (A.3.6) as

$$\frac{1}{2}\sum_{j=0}^{n-1}\cos\left(\frac{\pi(m-m')}{n}j\right) - \frac{1}{2}\sum_{j=0}^{n-1}\cos\left(\frac{\pi(m+m')}{n}j\right).$$

If $m \neq m'$ then both terms vanish by Proposition A.4. If $m = m'$ then the second term vanishes, but the first term is just $\frac{1}{2}\sum_{j=0}^{n-1} 1 = \frac{n}{2}$. □

Problem A.3. Use Euler's formula (A.3.1) and the identity $e^z e^w = e^{z+w}$ to prove the formula (A.3.7).