UNIVERSITY OF
Waterloo

University of Waterloo
Faculty of Mathematics

Centre for Education in
Mathematics and Computing

Intermediate Math Circles
March 31, 2010
Random Walks II

Recall that last week we talked about an eccentric hiker who walks on a trail, and we solved the following problem.

Problem 1. An eccentric hiker walks on a trail, according to the rule described as follows. Before taking each step he tosses a coin, and:

• If the toss results in heads, then he takes one step forward;

• If the toss results in tails, then he takes one step backwards (along the trail).

(a) What is the probability that after taking 20 steps, the hiker finds himself in the same place that he started from?

(b) What is the probability that after taking 20 steps, the hiker finds himself exactly one step ahead of the place that he started from? Exactly 2 steps ahead? Exactly 4 steps ahead?

(c) What is the most likely position of the hiker after he takes 20 steps?

In Problem 1 it was assumed that the trail extends for long enough both in front of and behind the hiker (so there were no restrictions on the hiker doing as many forward and backward steps as he pleases). Upon calculating, we found that the most likely position of the hiker, after 20 steps, is right at the spot where he had started! The probability for the hiker to find himself there after 20 steps turned out to be

$$p = \frac{\binom{20}{10}}{2^{20}} = \frac{184,756}{1,048,576} \simeq 17.62\%.$$

Today we will look at a situation where we change a bit the assumptions on the layout of the trail.

Problem 2. Consider the following modification to the framework of Problem 1: the trail behind the hiker is okay, but he is standing at the edge of a precipice (so even one step ahead will make the hiker fall off of the precipice). Calculate the probability that the hiker gets to take 20 steps without falling off of the precipice.

Problem 3. We continue to look at the situation from Problem 2. For every $n \geq 1$, let P_n denote the number of possibilities for the hiker to take n steps without falling off of the precipice, and where after the n steps he is precisely at the point where he had started from.

(a) Explain why $P_n = 0$ whenever n is an odd positive integer.

(b) Determine the values of P_2, P_4, P_6, P_8. Is there a pattern in these numbers?

(c) Find a general formula for P_n, where n is an even positive integer.

Before starting on the solutions to Problems 2 and 3, we review a very important sequence of numbers which appear in combinatorial enumeration – the *Catalan numbers*.

The sequence of Catalan numbers.
This is a sequence of integers $C_1, C_2, \ldots, C_n, \ldots$ which starts as follows:

$$C_1 = 1, C_2 = 2, C_3 = 5, C_4 = 14, C_5 = 42, \ldots$$

A direct formula for calculating Catalan numbers is

$$C_n = \frac{(2n)!}{n!\,(n+1)!}, \quad \text{for every } n \geq 1.$$

Thus if we want for instance to calculate C_4 by using this formula, we do the following:

$$C_4 = \frac{8!}{4! \cdot 5!} = \frac{8 \cdot 7 \cdot 6 \cdot 5!}{4! \cdot 5!} = \frac{8 \cdot 7 \cdot 6}{4!} = \frac{48 \cdot 7}{24} = 2 \cdot 7 = 14,$$

as stated above.

Another way of calculating the values of Catalan numbers is by an important *recurrence* satisfied by these numbers. This is a formula which allows us to calculate C_n by assumming that we have already calculated the preceding values $C_1, C_2, \ldots, C_{n-1}$. When stating the recurrence formula, it comes in handy to make the convention that we also have a Catalan number C_0, defined by

$$C_0 := 1.$$

Then the recurrence is as follows:
We have:

$$C_2 = C_0C_1 + C_1C_0, \quad C_3 = C_0C_2 + C_1C_1 + C_2C_0, \quad C_4 = C_0C_3 + C_1C_2 + C_2C_1 + C_3C_0,$$

and in general

$$C_n = C_0C_{n-1} + C_1C_{n-2} + \cdots + C_{n-2}C_1 + C_{n-1}C_0, \quad \text{for every } n \geq 2.$$

So for instance, suppose we have already verified that $C_1 = 1, C_2 = 2, C_3 = 5$, $C_4 = 14$. Then in order to calculate C_5 we can do the following:

$$C_5 = C_0C_4 + C_1C_3 + C_2C_2 + C_3C_1 + C_4C_0 = 1 \cdot 14 + 1 \cdot 5 + 2 \cdot 2 + 5 \cdot 1 + 14 \cdot 1 = 42,$$

as stated above.

The recurrence formula is useful because it allows us to *recognize* that a certain sequence of numbers is Catalan. That is, we have the following general fact.

Fact. Let $A_0, A_1, A_2, \ldots, A_n, \ldots$ be a sequence of numbers where $A_0 = A_1 = 1$ and where we have

$$A_2 = A_0A_1 + A_1A_0, \;\; A_3 = A_0A_2 + A_1A_1 + A_2A_0, \;\; A_4 = A_0A_3 + A_1A_2 + A_2A_1 + A_3A_0,$$

and in general

$$A_n = A_0A_{n-1} + A_1A_{n-2} + \cdots + A_{n-2}A_1 + A_{n-1}A_0, \quad \text{for every } n \geq 2.$$

Then $A_n = C_n$ for every $n \geq 0$.

Why does the above fact hold? We have $A_0 = C_0$ and $A_1 = C_1$, and from there we can push through the recurrence formulas for the A_n and the C_n. By doing so we obtain succesively that

$$A_2 = A_0A_1 + A_1A_0 = C_0C_1 + C_1C_0 = C_2,$$

then

$$A_3 = A_0A_2 + A_1A_1 + A_2A_0 = C_0C_2 + C_1C_1 + C_2C_0 = C_3,$$

and so on, until the equality $A_n = C_n$ is eventually obtained for all n. The name used for such an argument is *mathematical induction*. (We will encounter another proof done by mathematical induction later on today, right at the end of the lecture.)

Let us now proceed with the solutions to Problems 2 and 3. It will be convenient to do Problem 3 first; this is because, as it will turn out, the answer to Problem 3(c) can be useful for the solution to Problem 2.

Solution to Problem 3.

Let us recall that last week we encoded the possibilities for how the hiker can do a walk by *words* made with the letters F (for "Forward") and B (for "Backward"). Last week we only looked at words of length 20, but certainly we can look at such words in any length. For example the word

$$BBF\ BFF$$

represents a walk of length 6, where the hiker first takes 2 steps backward, then 1 step forward, then 1 step backward, and ends with 2 steps forward. Note this is an example of walk of the kind that is counted when we calculate P_6 – indeed, at the end of the walk the hiker is back at the starting point, without having fallen off of the precipice.

In order to get an insight of what kind of words we need to focus on, let us start to work on parts (a) and (b) of the problem.

Part (a) is in fact immediate – what happens here is that a walk with an odd number of steps can never return the hiker to the position where he started from. (We also made this observation last week, in the solution to Problem 1(b).) So $P_n = 0$ (no possibilities at all) when n is odd.

Let us then consider the case when $n = 2m$, even. Here we want to look at words of length $2m$ which contain m occurrences of the letter F and m occurrences of B – this is exactly what is needed to make sure that at the end of the walk the hiker is back where he had started from. We don't want to count *all* words with m occurrences of F and m occurrences of B; indeed, the order in which the B's and F's come in the word can really make a difference. To take the simplest possible example: both words BF and FB correspond to walks of length 2 where the hiker ends back at his starting point. But BF represents a "safe" walk, while FB doesn't – the hiker falls off of the precipice right at start!

So then let us calculate, one by one, the particular values P_2, P_4, P_6, P_8.

Calculation of P_2. This was in fact done just above: there are 2 words of length 2 that return the hiker to where he started from, BF and FB. But with FB he falls off, so the only safe possibility is BF. We conclude that $P_2 = 1$.

Calculation of P_4. There is a total of 6 words of length 4 which return the hiker to where he started from, but only 2 of these words (namely $BBFF$ and $BFBF$) avoid the falling off of the precipice. So we find that $P_4 = 2$.

Calculation of P_6. The total number of words of length 6 which can be formed with 3 letters F and 3 letters B (and thus correspond to a 6-step walk where the hiker returns at the starting point) is $\binom{6}{3} = 20$. By examining these 20 words, we find that exactly 5 of them avoid the falling off the edge of the precipice. They are:

$$BFBFBF; \quad BFBBFF; \quad BBFBFF; \quad BBFFBF; \quad BBBFFF.$$

Hence $P_6 = 5$.

Calculation of P_8. The total number of words of length 8 which can be formed with 4 letters F and 4 letters B is $\binom{8}{4} = 70$. It is not so appealing to proceed by examining separately each of these 70 words, so we will try to organize them in a way which makes our calculation more efficient. In the discussion that comes next we will use the term "safe" to refer to words which avoid the falling off the edge of the precipice.

Let us look for instance at words which start with BF. How many safe words of length 8 are there, which start with BF? The answer is that there are 5 such words – indeed, we are just looking at words of the form $BF \star\star\star\star\star\star$, where $\star\star\star\star\star\star$ is one of the words of length 6 that were counted in P_6 (and we saw that $P_6 = 5$).

Now, the words starting with BF are precisely those ones where the hiker has a *first return to his original position* after having taken 2 steps. How about using the "first return to the original position" as a criterion, when we examine various words? More precisely: let us write

$$P_8 = P_8^{(2)} + P_8^{(4)} + P_8^{(6)} + P_8^{(8)},$$

where for every $1 \leq j \leq 4$ we denote

$$P_8^{(2j)} = \begin{cases} \text{number of words with 4 letters } F \text{ and 4 letters } B \\ \text{which avoid falling off the edge of the precipice} \\ \text{and where the hiker returns for the first time at his starting point} \\ \text{after exactly } 2j \text{ steps.} \end{cases}$$

The observation made above about words starting with BF can then be interpreted as saying that

$$P_8^{(2)} = P_6 = 5.$$

We next look at $P_8^{(4)}$. If the hiker must return to his starting point after 4 steps (but not after 2 steps!) then the word must start with $BBFF$. Then the second half of the word has to be one of the patterns counted in P_4 (either $BFBF$ or $BBFF$). We conclude that $P_8^{(4)} = 2$, where the words counted here are $BBFF\,BFBF$ and $BBFF\,BBFF$.

For $P_8^{(6)}$ we make the following observation: what we enumerate here is the set of words of the form

$$B\diamondsuit\diamondsuit\diamondsuit\diamondsuit F\ BF,$$

where “$\diamondsuit\diamondsuit\diamondsuit\diamondsuit$” represents any of the words that were counted in P_4. (Why is that? Think!) This implies that

$$P_8^{(6)} = P_4 = 2.$$

Finally for $P_8^{(8)}$ we observe that we are enumerating the words of the form

$$B\,\diamondsuit\diamondsuit\diamondsuit\diamondsuit\diamondsuit\diamondsuit\,F$$

where “$\diamondsuit\diamondsuit\diamondsuit\diamondsuit\diamondsuit\diamondsuit$” represents any of the words that were counted in P_6; this leads to $P_8^{(8)} = P_6 = 5$.

So altogether we come to

$$\begin{aligned} P_8 &= P_8^{(2)} + P_8^{(4)} + P_8^{(6)} + P_8^{(8)} \\ &= 5 + 2 + 2 + 5 = 14. \end{aligned}$$

This completes the solution to part (b) of the problem.

Solution to part (c). The approach used for calculating P_8 suggests a way of finding a *recurrence* satisfied by the numbers P_n with n even. Specifically, let us fix an even integer $n = 2m$, where say that $m \geq 2$. We write

$$P_{2m} = P_{2m}^{(2)} + P_{2m}^{(4)} + \cdots + P_{2m}^{(2m)},$$

where for every $1 \leq j \leq 2m$ we put

$$P_{2m}^{(2j)} = \begin{cases} \text{number of words with } m \text{ letters } F \text{ and } m \text{ letters } B \\ \text{which avoid the falling off the edge of the precipice} \\ \text{and where the hiker returns for the first time to his starting point} \\ \text{after exactly } 2j \text{ steps.} \end{cases}$$

Observe that $P_{2m}^{(2)} = P_{2m-2}$. This is because the safe words with first return after 2 steps are precisely those of the form

$$BF\,\underbrace{\star\star\cdots\star\star}_{2m-2},$$

where “$\underbrace{\star\star\cdots\star\star}_{2m-2}$” is a word counted in P_{2m-2}. Similarly, we have $P_{2m}^{(2m)} = P_{2m-2}$. Indeed, $P_{2m}^{(2m)}$ counts the safe words where the hiker never returns to his starting point until the very end of the walk; and these are precisely the words of the form

$$B\,\underbrace{\diamondsuit\diamondsuit\cdots\diamondsuit\diamondsuit}_{2m-2}\,F,$$

where "$\underbrace{\diamondsuit\diamondsuit\cdots\diamondsuit\diamondsuit}_{2m-2}$" is a word counted in P_{2m-2}.

Now what can we say about $P_{2m}^{(2j)}$ when j is strictly between 1 and $m-1$? This is a mix of the two cases dicussed in the preceding paragraph: since the first return to starting point is after exactly $2j$ steps, we are enumerating words of the form

$$B\ \underbrace{\diamondsuit\diamondsuit\cdots\diamondsuit\diamondsuit}_{2j-2}\ F\ \underbrace{\star\star\cdots\star\star}_{2m-2j},$$

where "$\underbrace{\diamondsuit\diamondsuit\cdots\diamondsuit\diamondsuit}_{2j-2}$" is a word counted in P_{2j-2}, and where "$\underbrace{\star\star\cdots\star\star}_{2m-2j}$" is a word counted in P_{2m-2j}. This leads to the conclusion that

$$P_{2m}^{(2j)} = P_{2j-2} \cdot P_{2m-2j}, \quad \text{for } 1 < j < m-1.$$

Our recurrence relation for P_{2m} thus takes the form

$$P_{2m} = P_{2m-2} + P_2 \cdot P_{2m-4} + \cdots + P_{2j-2} \cdot P_{2m-2j} + \cdots + P_{2m-4} \cdot P_2 + P_{2m-2}.$$

But from this recurrence we can tell exactly what the numbers P_{2m} are. Indeed, let us introduce a sequence of numbers A_m by putting

$$A_1 = P_2,\ A_2 = P_4, \ldots, A_m = P_{2m}, \ldots$$

and where we also introduce a number A_0, set to be $A_0 = 1$. Then $A_0 = A_1 = 1$, and the recurrence formula found for P_{2m} is transformed into

$$A_m = A_0 \cdot A_{m-1} + A_1 \cdot A_{m-2} + \cdots + A_{j-1} \cdot A_{m-j} + \cdots + A_{m-2} \cdot A_1 + A_{m-1} \cdot A_0.$$

This is exactly the recurrence which determines the Catalan numbers! Therefore, as discussed earlier in the lecture, we must have $A_m = C_m$ for every $m \geq 0$. Since $P_{2m} = A_m$, our final answer to Problem 3(c) is this:

$$P_{2m} = C_m \text{ (the } m\text{-th Catalan number), for every } m \geq 1.$$

[End of solution to Problem 3]

Solution to Problem 2.

We will calculate the required probability p by using the same kind of formula as discussed last week:

$$p = \frac{N}{T},$$

where T is the total number of possibilities for the walk ($T = 2^{20} = 1,048,576$, as we saw while solving Problem 1), and N is the number of walks of 20 steps where the hiker does not fall off the edge of the precipice.

Same as before, we will view the possible walks as words of length 20 made with the letters B and F. We will say that such a word is "safe" when it corresponds to a walk where the hiker does not fall off the edge of the precipice; in the opposite case we will say that the word is "unsafe". With this terminology, the numerator N in the formula for p is

$$N = \text{ number of safe words of length 20, with letters } B \text{ and } F.$$

Now, it is actually more convenient to count unsafe words, rather than safe words. This is because there are various special subsets of unsafe words that can be easily enumerated. For instance all words starting with the letter F are unsafe. How many such words are there? There are 2^{19} of them (first letter is F, then each of the other 19 letters can be picked in 2 ways). Moving to words that start with B we observe that, among them, all the words starting with BFF are unsafe as well. The number of words of this kind is 2^{17}, because the first 3 letters are prescribed (BFF), then each of the remaining 17 letters can be picked in 2 ways.

So let us denote

$$M = \text{ number of unsafe words of length 20, with letters } B \text{ and } F.$$

Then we have

$$M = M_1 + M_3 + M_5 + \cdots + M_{17} + M_{19},$$

where for every $0 \leq k \leq 9$ we denote by M_{2k+1} the number of unsafe words obtained in the following way:

(i) The first $2k$ letters correspond to a safe walk of length $2k$ (no falling off the edge), where the hiker is back at his starting point after the $2k$ steps.

(ii) The $(2k+1)$-th letter of the word is an F. (Ouch!)

(iii) The remaining $20 - (2k+1)$ letters at the end of the word can be anything (each of them can be picked in 2 ways).

By using our answer to Problem 3, we can calculate the value of every M_{2k+1}, $0 \le k \le 9$. Say for instance we want to determine M_9 (so here $k = 4$, $2k + 1 = 9$). M_9 counts words which look like this:

(i) The first 8 letters of the word correspond to a safe walk of length 8 (no falling off of the precipice), where the hiker is back at his starting point after the 8 steps. We have counted this in Problem 3(b) – the number of ways in which it can happen is the Catalan number $C_4 (= 14)$.

(ii) The 9-th letter of the word is an F. (Ouch!)

(iii) The remaining 11 letters of the word can be anything (each of them can be picked in 2 ways). So the number of possibilities for this part of the word is $2^{11} = 2048$. The number of ways in which we can create a word counted in M_9 is then 14×2048 (we create the first 8 letters, then separately from that we create the last 11 letters, and then we assemble everything together with an F on position 9). Hence

$$M_9 = C_4 \times 2^{11} = 14 \times 2048 = 28,672.$$

In general, for every $0 \le k \le 9$ we get (by the same kind of reasoning as shown for M_9):

$$M_{2k+1} = C_k \cdot 2^{19-k},$$

where C_k is the corresponding Catalan number. So then

$$\begin{aligned} M &= M_1 + M_3 + M_5 + \cdots + M_{17} + M_{19} \\ &= C_0 \cdot 2^{19} + C_1 \cdot 2^{17} + C_2 \cdot 2^{15} + \cdots + C_8 \cdot 2^3 + C_9 \cdot 2^1. \end{aligned}$$

Here we need to fill in the list of Catalan numbers up to C_9:

$$\begin{cases} C_0 = 1,\ C_1 = 1,\ C_2 = 2,\ C_3 = 5,\ C_4 = 14,\ C_5 = 42, \\ \quad C_6 = 132,\ C_7 = 429,\ C_8 = 1430,\ C_9 = 4862. \end{cases}$$

Plugging in these numerical values (and also the powers of 2 that are needed) we find that

$$M = 863,820. \quad \text{(Check!)}$$

Now recall that M was the number of "unsafe" words of length 20. What we really wanted was the complementary number N of "safe" words. But this is now readily computed as

$$N = T - M = 1,048,576 - 863,820 = 184,756.$$

Finally, the required probability is

$$p = \frac{N}{T} = \frac{184,756}{1,048,576} \simeq 17.62\%.$$

[End of solution to Problem 2.]

Remark.

Wait a minute – didn't we also get "$\simeq 17.62\%$" as our answer in Problem 1(a)? What kind of coincidence is this? Actually, by looking more carefully, we discover that the number N of safe words calculated so painfully in the solution to Problem 2 is precisely equal to $\binom{20}{10}$! This explains the re-occurence of 17.62%, because in Problem 1(a) we also got to 17.62% from the ratio $\binom{20}{10}/2^{20}$. But the question remains: is it just a coincidence that N came out equal to $\binom{20}{10}$ in the solution to Problem 2, or there is a more general fact behind this?

I will end the lecture with a little calculation indicating that we do have a more general fact in the background (this is certainly not something particular to walks of length 20). Let us recap our derivation for p – it went like this:

$$\begin{aligned} p &= \frac{N}{T} = \frac{T-M}{T} = 1 - \frac{M}{T} \\ &= 1 - \frac{C_0 \cdot 2^{19} + C_1 \cdot 2^{17} + C_2 \cdot 2^{15} + \cdots + C_8 \cdot 2^3 + +C_9 \cdot 2^1}{2^{20}} \\ &= 1 - \Big(\frac{C_0}{2^1} + \frac{C_1}{2^3} + \frac{C_2}{2^5} + \cdots + \frac{C_8}{2^{17}} + \frac{C_9}{2^{19}} \Big). \end{aligned}$$

On the other hand p turned out to be just $\binom{20}{10}/2^{20}$, so what we are dealing with is the following formula:

$$\frac{C_0}{2^1} + \frac{C_1}{2^3} + \frac{C_2}{2^5} + \cdots + \frac{C_8}{2^{17}} + \frac{C_9}{2^{19}} = 1 - \frac{\binom{20}{10}}{2^{20}}.$$

It is tempting to believe this is just a special case of a more general formula, stating that

$$(*) \qquad \frac{C_0}{2^1} + \frac{C_1}{2^3} + \frac{C_2}{2^5} + \cdots + \frac{C_{m-1}}{2^{2m-1}} = 1 - \frac{\binom{2m}{m}}{2^{2m}}$$

for every positive integer m. But how could we prove that something like this holds for every m? Answer: mathematical induction will do the job! With this method we only have to:

(i) check the case $m = 1$, and

(ii) check that we can always "push the m one unit further".

(The precise meaning of (ii) is this: we assume the equality $(*)$ was already proved for some value of m and, by relying on that, we verify the equality also holds for $m+1$.)

Okay, so then let us prove the formula $(*)$ by induction on m. For $m = 1$ the formula claims that

$$\frac{C_0}{2^1} = 1 - \frac{\binom{2}{1}}{2^2}.$$

This is true – both sides of the equation are equal to $\frac{1}{2}$. Then let us suppose that we proved $(*)$ for a certain value of m, and now we want to prove it for $m+1$. We have to verify that

$$(**) \qquad \frac{C_0}{2^1} + \frac{C_1}{2^3} + \frac{C_2}{2^5} + \cdots + \frac{C_{m-1}}{2^{2m-1}} + \frac{C_m}{2^{2m+1}} = 1 - \frac{\binom{2m+2}{m+1}}{2^{2m+2}}.$$

But because we assume $(*)$ was already proved for m, the left-hand side of Equation $(**)$ can be replaced by

$$\left(1 - \frac{\binom{2m}{m}}{2^{2m}}\right) + \frac{C_m}{2^{2m+1}},$$

so we are just left to check that

$$\left(1 - \frac{\binom{2m}{m}}{2^{2m}}\right) + \frac{C_m}{2^{2m+1}} = 1 - \frac{\binom{2m+2}{m+1}}{2^{2m+2}}.$$

This is equivalent (after cancelling a "1", changing signs, and multiplying both sides of the equation by 2^{2m+2}) to

$$(***) \qquad 4 \cdot \binom{2m}{m} - 2 \cdot C_m = \binom{2m+2}{m+1}.$$

Finally, $(***)$ is verified by simply replacing the Catalan number and the binomial coefficients from their definitions in terms of factorials:

$$\begin{aligned}
4 \cdot \binom{2m}{m} - 2 \cdot C_m &= 4 \cdot \frac{(2m)!}{m! \cdot m!} - 2 \cdot \frac{(2m)!}{m! \cdot (m+1)!} \\
&= 2 \cdot \frac{(2m)!}{m! \cdot m!} \cdot \left(2 - \frac{1}{m+1}\right) \\
&= 2 \cdot \frac{(2m)!}{m! \cdot m!} \cdot \frac{2m+1}{m+1}
\end{aligned}$$

$$\begin{aligned}
&= \frac{(2m)!}{m! \cdot m!} \cdot \frac{(2m+1)(2m+2)}{(m+1)(m+1)} \\
&= \frac{(2m+2)!}{(m+1)! \cdot (m+1)!} \\
&= \binom{2m+2}{m+1}.
\end{aligned}$$