University of Waterloo
Faculty of Mathematics

Centre for Education in
Mathematics and Computing

Intermediate Math Circles
March 24, 2010
Random Walks I

Today and in the next week's meeting we will discuss some facts related to *counting* and *probability*, which arise when we consider "random walk" problems like the one stated next.

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?

Note that in this problem we have assumed the trail extends enough, both in front of and behind the hiker. (So there are no restrictions on the hiker doing as many forward and backward steps as he pleases – this assumption will be changed in Problems 2 and 3.)

Before starting on the solution of Problem 1, we review two fundamental concepts which appear in counting problems.

Review of factorials and of binomial coefficients.

(a) For every positive integer n, the number n *factorial*, denoted as "$n!$", is defined as follows:

$$n! := 1 \cdot 2 \cdots (n-1) \cdot n.$$

So for instance we have $3! = 1 \cdot 2 \cdot 3 = 6$, or $7! = 1 \cdot 2 \cdots 6 \cdot 7 = 5040$. (The actual value of $n!$ grows very fast with n.)

Factorials appear when one counts the number of possible ways of *permuting* a list of distinct symbols. For instance if we have to permute in all possible ways (that is, write in all possible orders) the 3 distinct symbols a, b, c, we get $3! = 6$ possibilities:

$$a - b - c; \quad a - c - b; \quad b - a - c; \quad b - c - a; \quad c - a - b; \quad c - b - a.$$

In general, if instead of 3 distinct letters we have n distinct letters, then the number of possible ways of permuting them is $n!$.

(b) Let k, n be positive integers such that $k \leq n$. We define a new number called *the binomial coefficient n-choose-k* and denoted as "$\binom{n}{k}$", by the following formula:

$$\binom{n}{k} := \frac{n!}{k! \cdot (n-k)!}.$$

For example:

$$\binom{5}{2} = \frac{5!}{2! \cdot 3!} = \frac{120}{2 \cdot 6} = 10.$$

The binomial coefficient $\binom{n}{k}$ appears when one counts in how many ways it is possible to select a subset of k elements out of an (unordered) set of n elements. For instance say we have a set S with 5 elements given as follows:

$$S = \{a, \gamma, \text{green}, \pi, \text{camel}\}.$$

Then the possible ways of choosing a 2-element subset of S are the following:

$$\{a, \gamma\}; \quad \{a, \text{green}\}; \quad \{a, \pi\}; \quad \{a, \text{camel}\};$$

$$\{\gamma, \text{green}\}; \quad \{\gamma, \pi\}; \quad \{\gamma, \text{camel}\}; \{\text{green}, \pi\}; \quad \{\text{green}, \text{camel}\}; \quad \{\pi, \text{camel}\}.$$

There are exactly $\binom{5}{2} = 10$ such possibilities.

Instead of listing all the possibilities, we could have counted them as follows: when constructing a 2-element subset of S we first pick an arbitrary element of S (this can be done in 5 ways), and then we pick an arbitrary second element out of the 4 remaining elements of S. The first element of S can be picked in 5 ways and the second can be picked in 4 ways, which seems to give us $5 \times 4 = 20$ possibilities of constructing the subset. But we have overcounted – every 2-element subset of S has been in fact obtained twice! (For instance the subset $\{\text{green}, \pi\}$ is obtained when we first pick the element "green" and then the element "π", but is also obtained when we first pick "π" and then pick "green".) So the correct number of 2-element subsets of S is $\dfrac{5 \times 4}{2} = 10 = \dbinom{5}{2}$.

A similar argument to the one shown above gives us that, in general, if the set S has n elements, then the number of ways of constructing a k-element subset of S is equal to $\dfrac{n(n-1)\cdots(n-k+1)}{k!}$. This is precisely the binomial coefficient $\dbinom{n}{k}$, because

$$\begin{aligned}\frac{n(n-1)\cdots(n-k+1)}{k!} &= \frac{\big(n(n-1)\cdots(n-k+1)\big)\cdot\big((n-k)(n-k-1)\cdots 2\cdot 1\big)}{k!\cdot\big((n-k)(n-k-1)\cdots 2\cdot 1\big)} \\ &= \frac{n!}{k!\cdot(n-k)!} = \binom{n}{k}.\end{aligned}$$

In order to warm-up on the use of binomial coefficients, let us do a quick (and nice) counting exercise inspired by geometry.

Exercise. Consider a convex polygon with n vertices, where $n \geq 4$. We draw all the diagonals of the polygon, and we assume that no three of them meet at the same point. Let I_n denote the number of intersection points obtained when we draw all these diagonals. What is the formula for I_n?

For instance it is immediate that $I_4 = 1$. Then by drawing a pentagon and its diagonals, one easily counts that $I_5 = 5$. Similarly – draw a hexagon and its diagonals – we find that $I_6 = 15$. Is there a tractable pattern here? And what does this have to do with binomial coefficients?

Here is a hint for this exercise: the answer is that

$$I_n = \binom{n}{k},$$

for a certain special value of k. But what is that special value? (Once you figure out the special value of k, can you find a *proof* for the formula giving I_n?)

Now back to our eccentric hiker. Before proceeding with the solution to Problem 1, we will discuss what it means to "calculate the probability" of the various events mentioned in the statement of the problem. There are many possibilities for how the hiker can do a 20-step walk, and they are all encoded by *words* of length 20 made with the letters F (for "Forward") and B (for "Backward"). For example the word

$$FBFFB\;FBBFF\;BBFFF\;BFFFF$$

describes the walk where the hiker takes first a step forward, then one step backward, then two steps forward, then one step backward, and so on, ending with four steps forward which correspond to the four letters F at the end of the word.

So in how many ways, exactly, can the hiker do his 20-step walk? This is the same as the total number of words of the kind described above. We claim this total number is 2^{20}. Indeed, each of the 20 letters of the word can be chosen in 2 possible ways, and the letters are chosen independently from each other – so combining the possibilities for the various letters we get

$$\underbrace{2 \times 2 \times \cdots \times 2}_{\text{20 times}} \;=\; 2^{20} \;=\; 1,048,576$$

possibilities for the choice of the word.

Finally, let us record the formula for the probability p of something (some well specified event) to happen, in connection to the hiker's walk. The formula is

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

where T is the total number of possibilities for the walk ($T = 2^{20}$, as we saw above), and N is the number of walks when the considered event does actually happen.

With all these things reviewed, let us then solve Problem 1.

Solution to Problem 1.

(a) The event considered here is that the hiker returns to his starting point. This happens if and only if the word corresponding to the walk has 10 letters F and 10 letters B. So $p = \frac{N}{T}$ where $T = 2^{20}$ and N is the number of 20-letter words which have 10 letters F and 10 letters B.

In order to create a 20-letter word with 10 of F and 10 of B, all we have to do is *choose* a set of 10 positions out of the 20 positions of letters in the word. Indeed, once this choice was done we write the letter F on the 10 positions that were chosen, then we write the letter B on the remaining positions, and the word is completely determined! From this observation it follows that

$$N = \binom{20}{10} = 184,756.$$

Thus the answer to this part of the problem is

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

(b) Here we have to calculate the probabilities $p_1 = \dfrac{N_1}{T}$, $p_2 = \dfrac{N_2}{T}$ and $p_4 = \dfrac{N_4}{T}$, where: N_1 is the number of walks ending one step to the right of the starting point;
N_2 is the number of walks ending 2 steps to the right of the starting point;
N_4 is the number of walks ending 4 steps to the right of the starting point.

We claim that $N_1 = 0$ – the hiker's walk can never end exactly one step to the right from where he started! Indeed, suppose we had such a walk (ending 1 step to the right). Let x be the number of forward steps in that walk, and let y denote the number of backward steps. Then $x + y = 20$ (total number of steps) and $x - y = 1$ (because the walk is supposed to end 1 unit to the right of the starting point). This implies that $x = 10.5$ and $y = 9.5$ – impossible, since x and y have to be integers.

So $N_1 = 0$ and it follows that $p_1 = \dfrac{N_1}{T} = 0$.

For N_2 we follows the same strategy as in part (a). The walk ends 2 units to the right of the starting point if and only if there are 11 steps forward and 9 steps backward. So the corresponding 20-letter word must have exactly 11 letters F in it, and the set of positions for these 11 letters F can be chosen in 20-choose-11 ways. It follows that

$$N_2 = \binom{20}{11} = 167,960,$$

hence

$$p_2 = \frac{167,960}{1,048,576} \simeq 16.02\%.$$

Similarly, for N_4 and p_4: we get

$$N_4 = \binom{20}{12} = 125,970, \quad p_4 = \frac{125,970}{1,048,576} \simeq 12.01\%.$$

(c) For every $0 \leq m \leq 20$ let p_m denote the probability that the walk ends m units to the right of the starting point. In parts (a) and (b) we calculated p_0, p_1, p_2 and p_4. Similar calculations show us that $p_m = 0$ whenever m is odd, and that

$$p_m = \frac{\binom{20}{10+j}}{2^{20}} \text{ when } m \text{ is even, } m = 2j.$$

In order to decide which is the largest number among $p_0, p_2, p_4, \ldots, p_{18}, p_{20}$ we must determine which one is the largest among the binomial coefficients

$$\binom{20}{10+j}, \quad 0 \leq j \leq 10.$$

We claim that

$$\binom{20}{10} > \binom{20}{11} > \cdots > \binom{20}{19} > \binom{20}{20}.$$

Indeed, for every $0 \le j \le 9$ the inequality

$$\binom{20}{10+j} > \binom{20}{10+j+1}$$

comes to

$$\frac{20!}{(10+j)!(10-j)!} > \frac{20!}{(11+j)!(9-j)!}.$$

This is equivalent to

$$(10+j)!(10-j)! < (11+j)!(9-j)!$$

(when two fractions have the same numerator, the one with smaller denominator is the larger fraction). The latter equality is equivalent to

$$\frac{(10+j)!(10-j)!}{(11+j)!(9-j)!} < 1.$$

But $\frac{(10+j)!}{(11+j)!} = \frac{1}{11+j}$ and $\frac{(10-j)!}{(9-j)!} = 10 - j$, so we obtain indeed that

$$\frac{(10+j)!(10-j)!}{(11+j)!(9-j)!} = \frac{(10+j)!}{(11+j)!} \cdot \frac{(10-j)!}{(9-j)!} = \frac{10-j}{11+j} \le \frac{10}{11} < 1.$$

The conclusion of the above calculations is that p_0 is the largest among the probabilities p_m with $0 \le m \le 20$. In words: for every $m \ge 1$, the probability p_m that the walk ends m steps to the right of the starting point is smaller than the probability p_0 that the walk ends where it started. By symmetry, the probability for the walk to end m steps to the *left* of the starting point is also equal to p_m, and hence is smaller than p_0 as well.

It follows that among the possible positions of the hiker at the end of the walk, the most likely one is at the very place where he had started. [End of solution to Problem 1]

Once we start thinking about the eccentric hiker, it comes naturally to consider variations of Problem 1, like the ones stated in Problems 2, 3 below (which will be discussed next week).

Problem 2. Consider the framework of Problem 1, with the following modification: the trail behind the hiker is okay, but right ahead of him there is a precipice (so even one step ahead will make the hiker fall off of the precipice). Calculate the probability that the hiker gets to make 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.