The Random Walk For Dummies

Richard A. Monte

Abstract. We look at the principles governing the one-dimensional discrete random walk. First we review five basic concepts of probability theory. Then we consider the Bernoulli process and the Catalan numbers in greater depth. Finally we determine the probability that, if a drunk is found again at the bar, then this is his first return visit.

1. Introduction. The random walk has been a topic of interest in many disciplines, but it has been of particular interest in probability theory. Indeed, every student of probability theory has heard of the random walk, especially in the form of a drunk leaving a bar and wandering aimlessly up and down the street.

Among other issues, the following four have been investigated in one, two, and three dimensions:

- The expected position x of the drunk after n steps.
- The maximum position x that the drunk has reached after n steps.
- The expected time of the drunk's last visit to 0.
- The probability that the drunk hasn't stumbled upon his own path after n steps.

These issues are discussed, for example, in Rota's book on probability [**3**].

The one-dimensional discrete case is most widely known because it is simple, yet illustrates many interesting and important features. In this paper, we treat this case of the first issue above. Thus we answer a basic question: What is the probability that the drunk is at a certain distance x, from the bar, after n steps?

In Section 2, we review five basic topics: the sample space, the binomial coefficient, random variables, the Bernoulli process, and Catalan numbers. In Section 3, we find the probability distribution for the position of the drunk after n steps. In Section 4, we proceed to calculate the probability that the drunk arrives at the bar for the first time after n steps. Finally, in Section 5, we determine the conditional probability that the drunk's first return occurs on the nth step given that he is indeed at the bar then; the formula is surprisingly simple.

2. The Basics. In this section, we review five concepts of probability theory, which we will use to study the random walk of the drunk.

First, a *sample space* Ω is defined to be the set of all possible outcomes of an experiment. Consider the coin toss as an example. A single toss has two possible outcomes: heads H, or tails T; thus $\Omega = \{H, T\}$. If we toss the coin twice, then there are 2^2 different possibilities; now $\Omega = \{TT, TH, HT, HH\}$.

Second, the *binomial coefficient* $\binom{n}{k}$ is defined to be the number of k-combinations of an n-element set. In other words, it is the number of different ways to pick k elements out of n. The binomial coefficient is given by the formula,

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

This formula is proved in [**1**, p. 61].

Third, a *random variable* is defined to be a function X that assigns to each element c, in the sample space of an experiment, one and only one real number $X(c)$. The sample space Ω of X is the set of real numbers x such that $x = X(c)$ for some c in the sample space of the experiment. This definition is found in [**2**, p. 28].

Fourth, a *Bernoulli process* is defined to be a sequence of random variables X_1, X_2, X_3, ... Each X_i records the outcome of an experiment modeled by the toss of a coin. Let p equal the probability of getting a head, and q the probability of getting a tail. Since these are the only possible outcomes, $p + q = 1$. Let X_i be 1 if the nth trial yields a head, and be -1 if a tail. Thus, the sample space Ω of a Bernoulli process is the set of all possible sequences of 1 and -1.

Fifth and finally, the nth *Catalan number* c_n counts certain arrangements of $2n$ parentheses. A *well-formed arrangement* is a list of $2n$ parentheses where each open parenthesis can be paired with a corresponding closed parenthesis to its right. The nth Catalan number is the number of such well-formed arrangements. For example, when $n = 3$, the possible well-formed arrangements are

$$()()(),\ ()(()),\ (())(),\ (()()),\ ((())).$$

According to [**1**, p. 253], the nth Catalan number c_n is given by the formula,

$$c_n = \frac{1}{n+1}\binom{2n}{n}.$$

Thus, for example, the first six Catalan numbers are $1, 1, 2, 5, 14$, and 42.

3. How It All Ties In. Suppose a drunk leaves a bar and walks aimlessly up and down the street, totally disoriented. We model the street as a line with the bar at the origin, and assume that the drunk takes unit steps, so we may record his position with an integer. Thus, for example, if he takes 5 steps to the left, he will be at position -5.

We now calculate the probability that the drunk is at position x after n steps. Assume that the drunk's walk can be modeled by a Bernoulli process. Say that the ith step is represented by the random variable X_i. A value of '-1' indicates a step to the left; a value of '1', a step to the right, so that

$$G_n = X_1 + X_2 + \ldots + X_n$$

gives the position of the drunk after n steps. We want to know the distribution of the random variable G_n.

Denote the value of G_n by x. Denote the number of steps taken to the right by r, the number to the left by l. Then

$$x = r - l \text{ and } n = r + l.$$

It follows that

$$r = \tfrac{1}{2}(x+n) \text{ and } l = \tfrac{1}{2}(n-x).$$

Now, there are $\binom{n}{l}$ ways that l given steps can occur among n total steps. This is also the number of ways of arriving at the point x, and each way has probability $p^r q^l$. Note that n and x must have the same parity because $n - x = 2l$. We can therefore conclude that the probability distribution at point x is given by the formula,

$$P(G_n = x) = \begin{cases} \binom{n}{l} p^r q^l, & \text{if } x = n \bmod 2; \\ 0, & \text{otherwise.} \end{cases}$$

Consider for example the case $p = q = \frac{1}{2}$; this is the case of the *symmetric random walk*. The probability $P(G_n = x)$ of being at position x after n steps is given in Table 3-1. This table is simply a pascal triangle interspersed with 0s.

Table 3-1
The Symmetric Random Walk

$n \setminus x$	-5	-4	-3	-2	-1	0	1	2	3	4	5
0						1					
1					$\frac{1}{2}$	0	$\frac{1}{2}$				
2				$\frac{1}{4}$	0	$\frac{2}{4}$	0	$\frac{1}{4}$			
3			$\frac{1}{8}$	0	$\frac{3}{8}$	0	$\frac{3}{8}$	0	$\frac{1}{8}$		
4		$\frac{1}{16}$	0	$\frac{4}{16}$	0	$\frac{6}{16}$	0	$\frac{4}{16}$	0	$\frac{1}{16}$	
5	$\frac{1}{32}$	0	$\frac{5}{32}$	0	$\frac{10}{32}$	0	$\frac{10}{32}$	0	$\frac{5}{32}$	0	$\frac{1}{32}$

The probability distribution in this particular case is given by

$$P(G_n = x) = \binom{n}{l} \left(\frac{1}{2}\right)^n = \binom{n}{l} \Big/ 2^n.$$

Since every path is equally likely in the symmetric random walk, the probability can be interpreted purely combinatorially. The probability is the number of different ways of arriving at x divided by the size of the sample space. The sample space is the set of all the possible paths of length n. Since the drunk has two choices at each point, and he takes a total of n steps, the total number of possibilities is 2^n.

4. Taking a Step Further. From now on, we assume that the random walk is symmetric; that is, $p = \frac{1}{2}$. What is the probability that the drunk's first return is at the $2n$th step? Here is where Catalan numbers enter. Observe that there is a one-to-one correspondence between paths and arrangements of parentheses. First, let an open parenthesis represent a step to the left, and a closed parenthesis a step to the right. For now, we consider only the case where the drunk's first step is towards the left, since the case to the right is clearly symmetric to it.

From Section 2, recall the definition of the Catalan number c_n: it is the number of $2n$ well-formed arrangements of parentheses. Note that, for well-formed arrangements, the number of open parentheses is always at least that of closed parentheses, regardless of what first k parenthesis we pick. Because of the correspondence between paths and arrangements, the Catalan numbers count the number of paths that the drunk can take, which start and end at the bar, without ever crossing to the right side of the bar.

We now adapt our correspondence to the problem of the drunk's first return. The drunk takes his first step to the left. On the next $2n-2$ steps, we insist that the drunk's path corresponds to a well-formed arrangement so that at step $2n-1$ he is at position

-1 without ever crossing to the right of -1. Then the final step brings the drunk back into the bar. This counts the total number of different paths that the drunk can take given the condition that the first return be at the $2n$th step.

Summarizing, we have the following sequence of steps:

- **Step 1.** The drunk takes one step to the left.
- **The next 2n-2 steps.** The drunk follows a path, which corresponds to a $2n-2$ well-formed arrangement, that restricts the drunk to be either at point -1 or to the left of it, and that forces him to be at -1 after the $(2n-1)$st step.
- **Step 2n.** The drunk's first return visit occurs when he takes one step to the right and into the bar.

Hence, the number of paths the drunk can take is simply the Catalan number c_{n-1}, since the well-ordered restriction applies only to the middle $2n-2$ steps.

We know that the nth Catalan number is equal to

$$\frac{1}{n+1}\binom{2n}{n}.$$

So the $(n-1)$st Catalan number is simply

$$\frac{1}{n}\binom{2n-2}{n-1}.$$

Allowing the drunk to take his first step to the right will double the total number of paths that the drunk can take. So the total is

$$2\frac{1}{n}\binom{2n-2}{n-1}.$$

To compute the probability of first return, we need only divide this quantity by the size of the total sample space. The latter was seen at the end of Section 3 to be 2^{2n} since there are $2n$ steps. Therefore, the probability that the drunk reaches the bar for the first time after $2n$ steps is

$$2\frac{1}{n}\binom{2n-2}{n-1} \Big/ 2^{2n}.$$

So, for example, consider the probability that the first return is at Step $2n = 6$; this probablity is given by

$$2\frac{1}{3}\binom{4}{2}\Big/2^{6} = \frac{1}{16}.$$

5. Conditional Probability and the First Return. We now consider a variant of the problem studied in the previous section: Given that the drunk is at the bar at Step $2n$, what is the probability that this is his first return visit? This problem is like the previous one. The only difference is that the sample space has been reduced. Instead of considering all of the possible paths, we now consider the total number of paths given that, in the end, the drunk will be at the bar, which he may or may not have passed by earlier.

To find the probability that the drunk will be at the bar after $2n$ steps, we use the conclusions from Section 3 to obtain

$$P(G_{2n} = 0) = \binom{2n}{n} p^{\frac{1}{2}(2n)} q^{\frac{1}{2}(2n)}.$$

Since $p = q = \frac{1}{2}$, this expression becomes

$$\binom{2n}{n} \Big/ 2^{2n}.$$

The numerator in this expression represents the number of paths the drunk can take provided he is at the bar at the $2n$th step. This is sample space we need for the conditional probability.

From Section 4, recall that, if the drunk's first return visit is on the $2n$th step, then the number of paths that the drunk can take is

$$2\frac{1}{n}\binom{2n-2}{n-1}.$$

On the other hand, in Section 3, we found the number of paths that the drunk can take that put him back at the bar on Step $2n$; this number is simply

$$\binom{2n}{n}.$$

Finally we divide these two numbers, getting

$$2\frac{1}{n}\binom{2n-2}{n-1} \Big/ \binom{2n}{n} = \frac{2}{n}\frac{(2n-2)!}{(n-1)!(n-1)!} \Big/ \frac{(2n)!}{n!n!},$$

which simplifies to

$$\frac{2}{n} \Big/ \frac{2n(2n-1)}{n^2} = \frac{2n}{2n(2n-1)} = \frac{1}{2n-1}.$$

The final simplicity is amazing! The formula is just

$$\frac{1}{\#(\text{steps}) - 1},$$

and is unexpectedly simple!

REFERENCES

[**1**] Brualdi, R., "Introductory Combinatorics," Prentice-Hall, Third Edition, 1999.

[**2**] Hogg, Robert V., "Introduction to Mathematical Statistics," Fifth Edition, 1995.

[**3**] Rota, G. C., "Probability Theory," Preliminary Edition, 1998.

This page will be blank.