"Science Stays True Here"
Journal of Mathematics and Statistical Science (ISSN 2411-2518), Vol.5, 206-211 | Science Signpost Publishing

An Unexpected Random Walk

Thomas J. Osler and Marcus Wright
Mathematics Department, Rowan University, Glassboro, NJ 08028
E-mail: osler@rowan.edu and wright@rowan.edu

1. Introduction

This is an expository paper about a simple random walk. But we examine the walk in a most unusual way that (while known) may not be familiar to the reader. The results are very non intuitive and surprising.

2. A Random Walk in One Dimension

We will examine the uniform random walk in one dimension. The following theorem is known, but not well advertised:

Random walk theorem

If x_i, for $i = 1,2,3,\ldots,N$ is a uniform random variable, taking on the values +1 and -1 only, then the sum $d(N) = \left|x_1 + x_2 + x_{3.}\ldots + x_N\right|$ as $N \to \infty$ has the expected value of

(1) $$E(d(N)) \quad \sqrt{\frac{2}{\pi}}\sqrt{N} \approx 0.798\sqrt{N} .$$

(In the above theorem, the statement $g(N) \quad f(N)$ means that $\lim_{N\to\infty} \frac{g(N)}{f(N)} = 1$ and we say that $g(N)$ is asymptotic to $f(N)$.).

We caution the reader that this result is not at all intuitive. At first glance the reader might think that the expected value is zero. This would be true if the sum under consideration was, (without absolute values),

$x_1 + x_2 + x_{3.} ... + x_N$, this is a classic error of assuming commutativity: Expected value of absolute value ≠ Absolute value of expected value.

3. Expected Random Walk – Simplified Argument

We now give a simple argument that will only crudely suggest the truth of our random walk theorem. Let x_k be a sequence of numbers randomly taking on the values +1 or -1. We wish to study $d(N) = \left| \sum_{k=1}^{N} x_k \right|$ and give some very simple partial explanation for: the expected value of $E(d(N)) \quad \sqrt{\frac{2}{\pi}}\sqrt{N} \approx 0.798\sqrt{N}$.

Let us look at $d(N)^2 = \left(\sum_{k=1}^{N} x_k \right)^2$. This is equal to $d(N)^2 = x_1^2 + x_2^2 + x_3^2 + \cdots + x_N^2 + (other\,terms) =$ $N + (other\,terms)$. Since the "*other terms*" can be both positive and negative, they tend to cancel one another. This leads us to conclude that perhaps the expected value of $E(d(N)) \quad \sqrt{N}$ (or better $E(d(N)) \quad c\sqrt{N}$ for some constant *c*). This completes our simple heuristic argument. A rigorous analysis follows in section 5 and also in [2]. Pages 38 and 39 of [1] give a very nice graphic description of this idea in action.

4. Example of the random walk

Consider tossing a fair coin yielding heads H or tails T, equally likely. You toss the coin 100 times. Suppose you get H 55 times and thus T 45. The difference 55 – 45 is 10. We will say that the *head-tail difference* (HTD) for this toss is 10. The HTD will always be considered non-negative, regardless if the toss yields more heads or more tails. Now reflect on the possible values of the HTD in 100 tosses of the coin. It could be 0 if H = D, (although this is unlikely), or it could even be 100 if either D or T is zero, (again unlikely). So it seems to be more likely that 0 < HTD < 100. In fact on closer reflection, we expect the HTD to be closer to 0 than 100.

Now we ask the question, after tossing the coin 100 times again and again, many times, what is the expected HTD? Our random walk theorem gives us the answer 8! For we can describe the sequence of tosses by a random walk say -1 +1 -1 + … +1, where our first toss is T, the second H, T the third, and finally H the one hundredth. Our random walk theorem tells us that the approximate expected *absolute value* of this sum (the HTD), after 100 tosses is $\sqrt{\frac{2}{\pi}}\sqrt{N} \approx .8\sqrt{100} = 8$. An interesting, conclusion that perhaps now may seem more intuitive!

(Also look at the comment immediately following equation (2) on section 5.)

5. Proving the Random Walk Theorem

Consider a uniform random one-dimensional walk in which we start on the *x* axis at the origin and make *N* random unit steps to the left and right with equal probability. (We prove the case where *N* is even.) The following table shows the probability of terminating at a given location. The numerators are Pascal's triangle. For example, if $N = 4$ steps, the probability that $d(4) = \sum_{k=1}^{4} x_k = -2$ is equal to $\frac{4}{16} = \frac{\binom{4}{3}}{2^4}$, corresponding to three of the four steps being chosen = -1 and one step = 1.

Probability of position after N steps

N	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
0							1						
1						1/2		1/2					
2					1/4		2/4		1/4				
3				1/8		3/8		3/8		1/8			
4			1/16		4/16		6/16		4/16		1/16		
5		1/32		5/32		10/32		10/32		5/32		1/32	
6	1/64		6/64		15/64		20/64		15/64		6/64		1/64

By definition, the expected value of a discrete random variable is the probability-weighted average of all possible values. In other words, each possible value the random variable can assume is multiplied by its probability of occurring, and the resulting products are summed to produce the expected value.

Let $E(d(N))$ denote the expected value of the unsigned distance from the origin after N steps. As an example, consider the case of 4 steps. The calculation would be

$$E(d(4)) = 2\left(0\left(\frac{6}{16}\right) + 2\left(\frac{4}{16}\right) + 4\left(\frac{1}{16}\right)\right) = 2\sum_{k=0}^{2}(2k)\frac{\begin{pmatrix}4\\2-k\end{pmatrix}}{2^4} = \frac{3}{2}.$$

The 2 on the left is needed because the probabilities must be doubled since both the same positive and negative distance count as one. The numbers 0, 2 and 4 are the possible values (moment arms) being multiplied by their corresponding probabilities from the table. It is easy to see that the generalization of this in the case where N is the even number $N = 2n$ is

$$E(d(N)) = E(d(2n)) = 2\sum_{k=0}^{n}(2k)\frac{\binom{2n}{n-k}}{2^{2n}}.$$

We found the sum on the right with Mathematica as $\sum_{k=0}^{n}(2k)\binom{2n}{n-k} = (n+1)\binom{2n}{n-1}.$

But $(n+1)\binom{2n}{n-1} = (n+1)\frac{n}{n+1}\binom{2n}{n} = n\binom{2n}{n} = n\frac{(2n)!}{(n!)^2}.$ We have

(2) $$E(d(N)) = E(d(2n)) = \frac{n \quad (2n)!}{2^{2n-1}(n!)^2}.$$

(It is interesting to note that this **exact** expression gives 7.95892 for N =100 which was studied in the above coin toss example using the asymptotic expression (1).)

The expression (2) gives the exact expected value after a finite number of steps. We finally consider N approaching infinity. We can make the Stirling asymptotic approximation $x! \quad \sqrt{2\pi x}e^{-x}x^{x}$ for large x and get at once $n\frac{(2n)!}{(n!)^2} \quad 2^{2n}\sqrt{\frac{n}{2\pi}}$ and so since $n = N/2$ we finally have

$$E(d(N)) \quad \sqrt{\frac{2}{\pi}}\sqrt{N} \approx 0.798\sqrt{N}.$$

This means that the expected distance from the origin after 1,000,000 random unit steps, each with the same probability of going right or left is about d(1,000,000) = 798. We checked this with Mathematica's random number generator. We calculated d(1,000,000) one thousand times and found the average to be 769 which is an error of less than 4 percent.

References

[1]. Mazur, B. and Stein, W, *Prime Numbers and the Riemann Hypothesis*, Cambridge University Press.

[2]. Weisstein, Eric W. "Random Walk--1-Dimensional." From *MathWorld*--A Wolfram Web Resource. http://mathworld.wolfram.com/RandomWalk1-Dimensional.html