*Electronic Proceedings of Undergraduate Mathematics Day*
*2004, No. 1, http://www.udayton.edu/~mathdept/EPUMD/EPMUD.htm*

University of Dayton
Kenneth C. Schraut Memorial Lecture
November 1, 2003

# *How Not to Get Lost While on a Random Walk*

by
Robert Edward Lewand '66

---

It is a high honor to be invited to give a talk in memory of Dr. K. C. Schraut, for he played a *very* significant role in my life. True, he was only one of the many great teachers I had. Much of what I know about teaching I learned from Mr. John Baker, and it was Brother Joe Stander who introduced me to the beauty of abstract algebra. But Doc's influence extended beyond the classroom. ***Far*** beyond the classroom. He had strong opinions on matters ranging from who would be appropriate roommates for the majors in his department to what graduate schools would best suit the talents and personalities of the graduating seniors. He was extremely proud (and justifiably so) of the fact that among all non-Ph.D. granting colleges and universities in the country, U.D. at that time had the fourth highest number of its mathematics graduates going on to get Ph.D.s. And he was the person most responsible for achieving this recognition.

Dinners at his home, souvenirs from his trip to Germany, his annual Christmas letters, the many squash games – these and many other recollections all add up to the unforgettable experience that was my education here at Dayton. I am greatly in his debt and I regard this talk as homage to a wonderful man.

I am humbled to be included among those who have previously given this lecture: Joe Diestel ('64 – at Kent State), MacArthur Prize Winner Richard Schoen ('72 – at Stanford), and my good friend Paul Campbell ('67 – at Beloit). A very elite group of mathematicians.

Before starting the talk, I would also like to mention how pleased I am that the tradition of strong mathematics education at U.D. is still very much in evidence. I personally know some of the faculty here and I am convinced that the care for and careful attention to students that so characterized the administration of K. C. Schraut continues today.

---

## *Introduction*

I have selected for my talk a topic that is far afield of my primary area of interest, namely cryptology. The subject of random walks is an appropriate area to observe the sometimes counter-intuitive fact that patterns can emerge from an apparently disorderly process. Because of this phenomenon, a random walk, typically encountered as a rather straightforward process in a probability course, has a wide range of applicability to fields as diverse as the stock market, particle physics and medicine; any field where chance plays a role.

Putting aside the modifier "random" for the time being, a *walk* may be thought of as a mapping from a set T (usually considered to be discrete units of time, e.g., seconds) to a set S, the state space (usually thought to be integral coordinates of a position in either one, two or three-dimensional space).

For example, if we choose T to be the set T = {0, 1, 2, 3, 4,...}, and S the set S = {.., -3, -2, -1, 0, 1, 2, 3,…}, then a walk in one dimension would be a function X: T $\rightarrow$ S where X(t) = the position in the state space of an object at time t. So we could have X(0) = 3, X(1) = 7, X(2) = -8, etc.

A two-dimensional walk would be a function X: T $\rightarrow$ SxS, etc. An example would be the relation $X(t) = ((-1)^t t^2, t+1)$ in which case X(0) = (0,l), X(1) = (-1,2), X(2) = (4,3). This last example is completely deterministic, i.e., there really is no randomness to this walk at all. For any value $t \in T$, X(t) is unambiguously determined.

To establish randomness in a walk, it is necessary to introduce probability into the description of the function X. We should insist that the walk can take any of several directions, each with a certain probability.

In this talk we will consider three independent vignettes from this theory.

---

### *Prerequisites*

Except for your very first mathematics class in pre-school, every mathematics lesson you have experienced has had a set of prerequisites. This talk is no exception, however the list of prerequisites is minimal. You need to know about factorials and about binomial coefficients. Here is a brief overview.

Definition: For ***n*** a positive integer, ***n*** ***factorial*** is denoted n! and is defined by

$$n! = (n)(n-1)(n-2)\ldots(2)(1)$$

Examples: 4! = 24, 5! = 120, 2! = 2, 1! = 1, 12! = 479,001,600. As a special case, 0! = 1.

Definition: For ***n*** a non-negative integer and for ***k*** an integer $0 \le k \le n$, the binomial coefficient $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.

Examples: $\binom{3}{0}=\frac{3!}{0!3!}=1$; $\binom{3}{1}=\frac{3!}{1!2!}=3$; $\binom{3}{2}=\frac{3!}{2!1!}=3$; $\binom{3}{3}=\frac{3!}{3!0!}=3$.

Observe that since $(x+y)^3 = x^3+3x^2y+3xy^2+y^3$, these so-called binomial coefficients are well named.

Also note the symmetry that exists in the binomial coefficients: $\binom{n}{k}=\binom{n}{n-k}$.

There is another interpretation of the expression $\binom{n}{k}$, namely $\binom{n}{k}$= number of subsets of size $k$ that can be extracted from a set of size $n$.

Examples: Let X = {a, b, c} (n=3). Here are the subsets of size 2: {a, b}, {a, c}, {b,c}. Note there are three of them.

What if X = {a, b, c, d} and we wanted to list all subsets of size 2? They would be {a,b}, {a,c}, {a, d}, {b, c}, {b, d}, {c, d}. There are six of them. Note that $\binom{4}{2}=\frac{4!}{2!2!}=6$.

Suppose now that you have a set with n elements. You count the number of subsets with 0 elements, then the number of subsets with 1 element, then the number of subsets with 2 elements, etc., finally the number of subsets with n elements. By the time you have added all these up, you will have the total number of subsets of the set, which happens to be $2^n$. This leads us to the following important identity that we shall label

Prerequisite #1: $$\sum_{k=0}^{n}\binom{n}{k}=2^n.$$

That's the first prerequisite for understanding this talk. There are three others.

Prerequisite #2: $$\sum_{k=0}^{n}\binom{n}{k}^2=\binom{2n}{n}$$

Proof: If you were choosing a committee of n people from a collection of 2n people (n men and n women), you could do it in this many ways:
$\sum_{k=0}^{n}\binom{n}{k}\binom{n}{n-k}=\sum_{k=0}^{n}\binom{n}{k}\binom{n}{k}=\sum_{k=0}^{n}\binom{n}{k}^2$, because you would be selecting k men and n-k women for all values of k between 0 and n. But the expression for selecting a group of n from a group of 2n is also given by $\binom{2n}{n}$. So the lemma is proved.

Prerequisite #3: $\sum_{k=0}^{n}\binom{2n}{k}\binom{2n-k}{k}\binom{2n-2k}{n-k}\binom{n-k}{n-k}=\binom{2n}{n}^2$.

Proof: $\sum_{k=0}^{n}\binom{2n}{k}\binom{2n-k}{k}\binom{2n-2k}{n-k}\binom{n-k}{n-k}=\sum_{k=0}^{n}\frac{(2n)!}{k!(n-k)!k!(n-k)!}=$

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

The fourth and final prerequisite for understanding all that is to follow is Stirling's Formula for approximating the size of n!. This will go without proof:

Stirling's Formula: $n! \approx \sqrt{2p}\, n^{n+\frac{1}{2}}e^{-n}$.

We saw above, for example, that 12! = 479,001,600 ; Stirling's approximation would be (rounded off to the nearest integer): 475,687,486. This approximation is within 0.7%.

Summarizing what we have so far:

Prerequisite #1: $\sum_{k=0}^{n}\binom{n}{k}=2^n$.

Prerequisite #2: $\sum_{k=0}^{n}\binom{n}{k}^2=\binom{2n}{n}$.

Prerequisite #3: $\sum_{k=0}^{n}\binom{2n}{k}\binom{2n-k}{k}\binom{2n-2k}{n-k}\binom{n-k}{n-k}=\binom{2n}{n}^2$.

Stirling's Formula: $n! \approx \sqrt{2p}\, n^{n+\frac{1}{2}}e^{-n}$.

That was the hardest part of the talk. Having established these identities, the rest of the talk will be simple; so if you have tuned out at this point, here's the perfect opportunity to jump back into the game. Let us now consider our three vignettes.

---

### *Vignette #1*

Let's first consider the one dimensional case where X(t+1) = X(t) + 1 or X(t) - 1, i.e., the only possible motions are one unit to the right and one unit to the left, and just for symmetry's sake, X(0) = 0.

Suppose p = probability of a jump to the right (X(t+1) = X(t) + 1),
q = probability of a jump to the left (X(t+1) = X(t) -1).

We are assuming that X(t+1) $\neq$ X(t), and so q = 1 - p.

Now imagine a toad, positioned on a long straight highway in some city who jumps each second either to the right (with probability p), or to the left (with probability q).
Note that the ratio X(t)/t represents the average displacement from the origin (0) over time t.

For p = 0.5, does it not seem natural that $\lim_{t\to\infty} \frac{X(t)}{t} = 0$? After all, after, say 1000 jumps, one would expect about 500 jumps to the right and 500 to the left. And so X(t) should be pretty close to 0. The larger t, the more likely this scenario.

Note that this situation is just like flipping a coin with probability 0.5 of heads and probability 0.5 of tails.
What if we change p from 0.5 to 0.75? Then what would you expect $\lim_{t\to\infty} \frac{X(t)}{t}$ to be?
Well, after, say 1000 jumps, one would expect about 750 to the right and 250 to the left. This would result in X(t) being approximately 500 and X(t)/t being approximately 0.5.

In general, since the expected number of jumps to the right after time t is pt, and the expected number to the left is qt, then the expected value for X(t) is pt - qt, and so X(t)/t = p - q in the limiting case. So, for example, if p = 0.9 and q = 0.1, then X(t)/t $\rightarrow$ 0.8.

---

## *Vignette #2*

Let's now change this scenario slightly and ask a different but related question.

Suppose the toad is at the origin and that bliss awaits him a certain number of moves to the right, while disaster awaits him that same number of moves to the left. This situation is usually referred to as "the gambler's ruin," where bliss is defined as winning a certain amount of money and ruin is defined as going broke. The p and q are determined by the game the gambler is playing and the origin is interpreted as the money with which the gambler begins the gaming, his initial purse.

For example, if one is playing roulette and betting on numbers {00, 0, 1, .. , 36}, then p = 1/38 and q = 37/38. A player may begin with $100 and stop when he/she either (a) has nothing left or (b) has a total of $500. Bliss = $500; Disaster = $0.

If instead of wagering on numbers, one wages on black or red, then p = 18/38 = 9/19.

Let's define a function h: S $\rightarrow$ [0,1], the happiness function, as h(x) = the probability of achieving bliss, i.e., reaching state N, from position x.

Then a little thought suggests that h(x) = p h(x+1) + q h(x-1); for if you are at state x then you will either advance to state x+1 (with probability p) at which point your probability of achieving bliss is h(x+1), or else you will retreat to state x-1 (with probability q) in which case your probability of achieving bliss is h(x-1).

For example, if p = 0.9, then h(x) = 0.9h(x+1) + 0.1h(x-1). This equation is known as a "difference equation" and is a topic ordinarily covered in a discrete mathematics course. This difference equation happens to have a "closed form solution," that is, a solution that does not define h in terms of itself. The purpose of this talk is not to demonstrate how to find closed form solutions of difference equations, but let's at least indicate in two different ways why the solution to this particular equation, in the special case when p = q = 0.5 is a linear equation, i.e., one of the form h(x) = mx + b.

<u>Method #1:</u> Check out the solution. If h(x) = mx + b, then we get

ph(x+1) + qh(x-1) =
p(m(x+1) + b) + q(m(x-1) + b) =
pmx + pm + pb + qmx - qm + qb =
(p+q)(mx) + (p+q)b + m(p-q) =
(1)mx + (1)b + 0 =
mx + b =
h(x).

<u>Method #2:</u> If h(x) = 0.5 h(x+1) + 0.5 h(x-1), then 2 h(x) = h(x+1) + h(x-1), or h(x+1) - h(x) = h(x) - h(x-1). So there is a constant h increment for each x increment. This is precisely the nature of a straight line.

Now we can be even more specific. Let's assume that ruin occurs at x = 0 and that bliss occurs at x = N. This translates into the two boundary conditions, h(0) = 0 and h(N) = 1.

Substituting into h(x) = mx + b, we get

0 = h(0) = b, and
1 = h(N) = mN + b.

So b = 0 and m = 1/N. Therefore h(x) = x/N.

<u>Example:</u> If N = 10, then h(x) = x/10. In particular h(5) = 0.5, h(7) = 0.7, h(1) = 0.1. Consider the implications of this for gambling at a game with p = 0.5, initial purse $5 and goal $10. Goal $100. Goal $500.

Purse $5 means x = 5; Goal $10 means N = 10. Therefore h(5) = 5/10 = ½.

Purse \$5, Goal \$100 means h(5) = 5/100 = 1/20.
Purse \$5, Goal \$500 means h(5) = 5/500 = 1/100.

So that takes care of the case when p = q = 0.5. But what about a closed form expression for h(x) in the case when p and q are not both 0.5?

Let's now address that problem. Well, again, I won't show you how to solve the equation h(x) = p h(x+1) + q h(x-1) when p ≠ q, but I will tell you the solution and then we can verify it. Here it is: $h(x) = C_1 + C_2\left(\frac{q}{p}\right)^x$.

Verification:

$$ph(x+1)+qh(x-1) = p(C_1 + C_2\left(\frac{q}{p}\right)^{x+1}) + q(C_1 + C_2\left(\frac{q}{p}\right)^{x-1}) =$$

$$C_1(p+q) + C_2\left(\frac{q^{x+1}}{p^x} + \frac{q^x}{p^{x-1}}\right) = C_1 + C_2\left(\frac{q^{x+1}+q^x p}{p^x}\right) = C_1 + C_2\frac{q^x(q+p)}{p^x} =$$

$$C_1 + C_2\left(\frac{q}{p}\right)^x = h(x).$$

Now let's throw in the two boundary conditions, h(0)=0 and h(N)=1 to conclude $0 = C_1 + C_2$, (i.e., $C_1 = -C_2$), and $1 = C_1 + C_2\left(\frac{q}{p}\right)^N$.

These equations imply $1 = -C_2 + C_2\left(\frac{q}{p}\right)^N = C_2\left(\left(\frac{q}{p}\right)^N - 1\right)$, or

$C_2 = \frac{1}{\left(\frac{q}{p}\right)^N - 1}$ and $C_1 = -\frac{1}{\left(\frac{q}{p}\right)^N - 1}$. Therefore

$$h(x) = -\frac{1}{\left(\frac{q}{p}\right)^N - 1} + \frac{\left(\frac{q}{p}\right)^x}{\left(\frac{q}{p}\right)^N - 1} = \frac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^N - 1}.$$

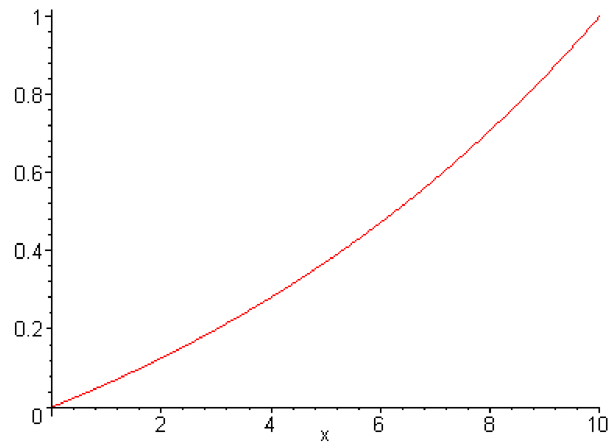
You can see why this formula does not hold when p = q = 0.5.

As an example, if p = 9/19 and N = 10, values of h(x) would be as they appear in this table:

| *x* | *h(x)* |
|---|---|
| 0 | 0 |
| 1 | .0594 |
| 2 | .1256 |
| 3 | .1990 |
| 4 | .2806 |
| 5 | .3713 |
| 6 | .4720 |
| 7 | .5839 |
| 8 | .7082 |
| 9 | .8465 |
| 10 | 1 |

Notice that x would have to be at 7 before there is a better than even chance to reach bliss.

Here's what the graph of h(x) looks like in this case:




Clearly h is no longer a linear function.

---

***Vignette #3***

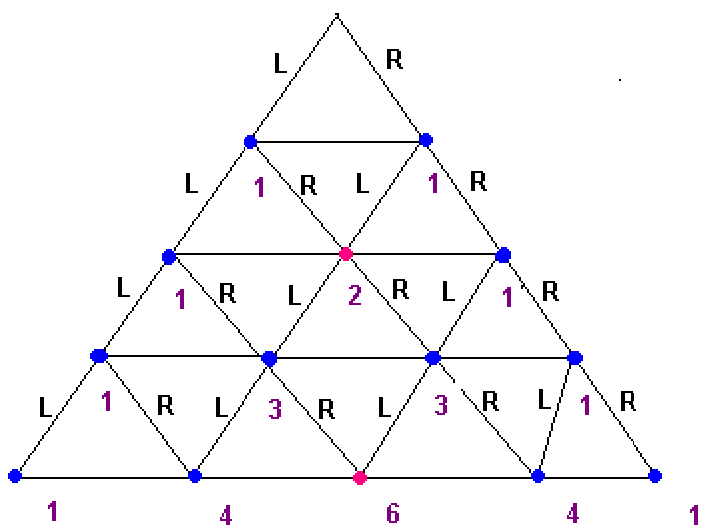
The third vignette addresses the question: What is the probability that the toad will ever return to the starting position assuming **equal** probability for a jump in any direction?

In the one-dimensional case, at each X(t), the frog has two options: leap left or leap right.

X(1) therefore is a result of a leap left or leap right.
X(2) is a result of a leap left or right from X(1) (or, from X(0), X(2) would be the result of leaps either

- left, left
- left, right
- right, left, or
- right, right).

This situation for four ticks of the clock is depicted in the following diagram:




Each horizontal line represents one tick of the clock and each dot represents a state X(t). The red dots correspond to the origin. (Note that in order to be situated at the origin, an even number of moves is required.) The digits beside the states indicate the number of different paths leading to that particular state.

What we have here is Pascal's Triangle. Note that the sum of the digits at horizontal level n is $2^n$.

Since an even number of moves is required to return to the origin, the number of ways this can be done at level 2n is $\binom{2n}{n}$, for the frog would need to select n left leaps and n right leaps in any order. Since order is irrelevant, this is a combination problem.

Therefore the probability of returning to the origin after 2n moves is $\frac{\binom{2n}{n}}{2^{2n}}$.

Referring to this expression as $u_n$, $u_n$ = P(returning to the origin after 2n moves).

Note $u_1 = 0.5$ (i.e., 2/4), $u_2 = 0.375$ (6/16), $u_3 = 0.3125$ (20/64), etc. Compare with the diagram above.

By Stirling's Formula, ( $n! \approx \sqrt{2p}\, n^{n+\frac{1}{2}} e^{-n}$ ), $u_n \approx \dfrac{1}{\sqrt{pn}}$ because

$$u_n = \frac{\binom{2n}{n}}{2^{2n}} = \frac{(2n)!}{n!n!2^{2n}} = \frac{\sqrt{2p}\,(2n)^{2n+\frac{1}{2}} e^{-2n}}{(\sqrt{2p}\, n^{n+\frac{1}{2}} e^{-n})^2\, 2^{2n}} = \frac{2^{2n}\sqrt{2} n^{2n}\sqrt{n}}{\sqrt{2p}\, n^{2n+1} 2^{2n}} = \frac{1}{\sqrt{p\,n}}.$$

Although $u_n \to 0$, $\sum_{n=1}^{\infty} u_n \approx \sum_{n=1}^{\infty} \frac{1}{\sqrt{p\,n}} = \frac{1}{\sqrt{p}} \sum_{n=1}^{\infty} \frac{1}{\sqrt{n}}$ diverges. This can be interpreted as meaning that the frog will almost definitely return to the origin after some number of moves!! George Polya proved an AMAZING result in 1921: The probability of returning home (or any other specified state for that matter) = 1. The frog will necessarily return home!

Well, what about the ***two-dimensional*** case?

In this scenario, at each X(t), the frog has 4 possibilities (leap up, leap down, leap right, leap left). As in the one-dimensional case, an even number of leaps is a necessary (though not sufficient) condition for returning to the origin. But now, at level n, rather than $2^{2n}$ possible paths, there are $4^{2n}$ possibilities. Think, for example, of n = 1 (i.e.. two moves). The 16 possibilities are UU, UD, UR, UL, DU, DD, DR, DL, LU, LD, LR, LL, RU, RD, RR, RL (where U is up, D is down, R is right, and L is left). Note that 4 of these sequences get the frog back to the origin.

The question arises: what is the probability of returning to the origin at the $(2n)^{th}$ move? To answer this question, let's assume that of the 2n moves, k of them are leaps up. (Note $0 \le k \le n$.). Then the frog would need k leaps down. This leaves (2n - 2k) moves, half of which must be left leaps and the other half right leaps. The number of ways in which this can be accomplished (without regard to order) is:

$\binom{2n}{k}\binom{2n-k}{k}\binom{2n-2k}{n-k}\binom{n-k}{n-k}$ or, more simply $\binom{2n}{k}\binom{2n-k}{k}\binom{2n-2k}{n-k}$.

Expanding, we get

$$\binom{2n}{k}\binom{2n-k}{k}\binom{2n-2k}{n-k} = \frac{(2n)!}{k!(2n-k)!}\frac{(2n-k)!}{k!(2n-2k)!}\frac{(2n-2k)!}{(n-k)!(n-k)!} = \frac{(2n)!}{(k!)^2((n-k)!)^2}.$$

Examples:

| **n** | **k** | $\binom{2n}{k}\binom{2n-k}{k}\binom{2n-2k}{n-k}$ |
|---|---|---|
| 2 | 0 | 6 (LLRR, LRLR, LRRL, RRLL, RLRL, RLLR) |
| 2 | 1 | 24 (UDLR, ULRD, UDRL, URLD, DULR, LURD, DURL, RULD, DLUR, …) |
| 2 | 2 | 6 (UUDD, UDUD, UDDU, DDUU, DUDU, DUUD) |

Summing up these expressions for k = 0 to k = n and using Prerequisite #3, ($\sum_{k=0}^{n}\binom{2n}{k}\binom{2n-k}{k}\binom{2n-2k}{n-k}\binom{n-k}{n-k}=\binom{2n}{n}^2$), we get the number of ways of returning to the origin after 2n moves is equal to $\binom{2n}{n}^2$, and therefore the probability ($v_n$) of returning to the origin in 2n moves is $\frac{\binom{2n}{n}^2}{4^{2n}}$. This expression is the square of $u_n$ in the one-dimensional case. As we concluded in the case of one dimension that $u_n \approx \frac{1}{\sqrt{pn}}$, so in the two-dimensional case, $v_n \approx \frac{1}{pn}$.

Again we get $v_n \to 0$, but $\sum_{n=1}^{\infty} v_n \approx \sum_{n=1}^{\infty}\frac{1}{pn} = \frac{1}{p}\sum_{n=1}^{\infty}\frac{1}{n}$ diverges. So again, it's very likely the frog will return to the origin. George Polya proved an AMAZING result in 1921: The probability of returning home (or any other specified state for that matter) = 1. The frog will necessarily return home!

Now, for the three-dimensional case, the mathematics gets a bit more complex; for now at each turn the frog has 6 possible moves: Up, Down, Left, Right, Back, Forth. But it happens that after a similar analysis, $w_n \to 0$ and $\sum_{n=1}^{\infty} w_n$ ***converges.*** Implication: It is unlikely that the frog will return to the origin after any even number of moves. Polya proved that in 3-dimensions, the probability of returning home is less than 1. This result is known as *Polya's Dichotomy*.

Since then it has been learned that the probability of returning home in various dimensions is as follows:[1]

[1] http://mathworld.wolfram.com/PolyasRandomWalkConstants.html

| Dimension | Probability(Returning Home) |
|---|---|
| 1 | 1 |
| 2 | 1 |
| 3 | ≈ 0.340537 |
| 4 | ≈ 0.193206 |
| 5 | ≈ 0.135178 |
| 6 | ≈ 0.104715 |
| 7 | ≈ 0.085845 |
| 8 | ≈ 0.072913 |

The exact expression for the probability of returning home in 3-dimensions is

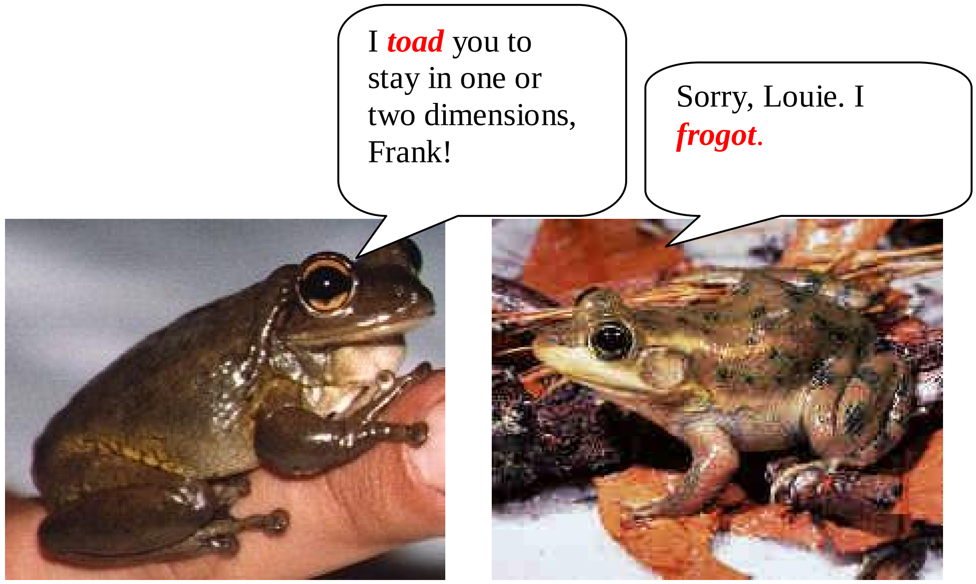
$$1 - \frac{(2p)^3}{3\int_{-p}^{p}\int_{-p}^{p}\int_{-p}^{p} \frac{dx\,dy\,dz}{3-\cos(x)-\cos(y)-\cos(z)}}$$

---

***Summary***

Here's what we've learned about the toad:

- In the one-dimension situation where the probability of a jump to the right is p and the probability of a jump to the left is q ( = 1-p ), then X(t)/t → p - q = 2p - 1. If p = q = 0.5, then the toad is bound to return to the origin in 2n moves for some n.
- In the one-dimension case, if p = q, then the likelihood the toad will achieve bliss (attain position N) before oblivion (attain position 0) from position x is x/N. If p ≠ q, then the probability is $h(x) = \frac{\left(\frac{q}{p}\right)^x - 1}{\left(\frac{q}{p}\right)^N - 1.}$.
- In two dimensions if the toad starts at the origin and moves left, right, up, or down with equal likelihood, then he is bound to return to his starting point after a certain (even) number of moves.
- In three dimensions, if the toad moves left, right, up, down, back, or forth with equal likelihood, chances are only about 34% that he will ever find his way home.




So take Louie's advice: If you don't want to get lost while taking a random walk, confine yourself to one or two dimensions! Thank you for your kind attention and warm reception.

---