MATHEMATICS 191, FALL 2004
MATHEMATICAL PROBABILITY
Outline #6 (Random Walks)

Last modified: October 28, 2004

References:

- PRP, sections 3.9 and 3.10
- EP, sections 5.6 through 5.8, 5.20, and 5.22

1. Here are some examples of a simple random walk.

    - In a computer game, the player starts at level and, at every step, either moves up one level with probability p or moves down one level with probability $1-p$.
    - Harvard and Yale play an infinite sequence of football games. At every step, Harvard either moves one game further ahead in the standings with probability p or moves one game less ahead (or further behind) with probability $1-p$.
    - A gambler makes an infinite sequence of \$1 bets at a casino. At every step, her net worth either increases by \$1 with probability p or decreases by \$1 with probability $1-p$.

    Express symbolically, and prove, the following properties of such a process:

    - It is spatially homogeneous.
    - It is temporally homogeneous.
    - It has the Markov property (i.e. no "momentum.")

2. As a preliminary to the mathematics involved in solving the recurrence relation for a random walk, review how to turn the following linear second-order differential equations into algebraic equations by plugging in appropriate trial solutions. For initial conditions, use $y = A$ and $\frac{dy}{dt} = 0$. Physicists will recognize this as the equation for a "damped harmonic oscillator."

    - $$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y = 0$$
      (trial solution $y = e^{at}$)
    - $$\frac{d^2y}{dt^2} + 2\frac{dy}{dt} + y = 0$$
      (trial solutions $y = e^{at}$ and $y = te^{at}$)

3. Suppose that the player starts in level $k \geq 0$, and the game is won if the player reaches level N and lost if the player reaches level 0. Let p_k denote the probability that the player starts in level k and eventually loses the game. Set up and solve a recurrence relation to determine p_k, first in the general case where $p \neq \frac{1}{2}$ and then in the special case where $p = \frac{1}{2}$. The second case is like the second differential equation above.

4. As a preliminary to the mathematics involved in solving the next recurrence relation for a random walk, review how to solve the following inhomogeneous linear differential algebraic equations by finding a "particular solution" to which you add the general solution of the homogeneous equation. Physicists will recognize this as the equation for a "forced harmonic oscillator."

   - $$\frac{d^2y}{dt^2} + 3\frac{dy}{dt} + 2y = e^{-3t}$$
     (trial solution $y = Be^{-3t}$)
   - $$\frac{d^2y}{dt^2} + 2\frac{dy}{dt} + y = e^{-t}$$
     (trial solution $y = Bt^2e^{-t}$)

5. Let D_k denote the expected number of steps after which the player, starting in level $k \geq 0$, wins or loses the game. Set up and solve a recurrence relation for D_k, and solve it for the special case where $p = \frac{1}{2}$. Discuss what happens in the limit $N \to \infty$.

6. Allow "negative levels," and use the reflection principle to count the number of paths $N_n^0(a, b)$ that start at level $a > 0$ and end at level $b > 0$ while passing through level 0. Then state and prove the "ballot theorem," which states that the fraction of paths going from level 0 to level b in n steps that never pass through level 0 again equals $\frac{b}{n}$. As a concrete example, imagine that Harvard and Yale play 7 games, of which Harvard wins 5. You can easily enumerate the 9 of 21 cases in which the series is never tied.

7. State the "principle of reversal", and use it to prove the "hitting time theorem," which states that that the fraction of paths going from level 0 to level b in n steps that never passed through level b before step n also equals $\frac{b}{n}$. Show that the following probabilities are equal:

   - the probability $f_b(n)$ that the player is in level b for the first time at step n, perhaps having spent some time in negative levels.

- the probability $f_b(n)$ that the player is in level b at step n without ever having returned to level 0, but perhaps having spent some time at levels higher than b.

Using the same example as above, you can easily enumerate the 9 of 21 cases in which the Harvard never goes three games ahead until the final game of the series.

8. As an application of the preceding result, imagine that H and Y agree to play a series of independent games, each of which H wins with probability p. The series starts in level 0, and it ends if it ever again return to level zero (becomes tied). Show that the expected number of times μ_b that the series is in level b (H is b games ahead) before the series ends is equal to the probability f_b that the series would eventually reach level b if a tie did not cause it to end, namely $\mu_b = \sum_{n=1}^{\infty} f_b(n)$, and that if $p = \frac{1}{2}$ this sum is equal to 1.

9. Show that the following are all equal:

   - The number of paths from time 0, level 0 to time $2n$, level 0.
   - The number of paths from time 0, level 0 to time $2n$, any level, that never return to level 0 (either always positive or always negative).
   - The number of paths from time 0, level 0 to time $2n$, any level, that remain in non-negative levels.

   For the simple case where $2n = 4$, list all the paths in each of these three cases. State this key lemma also in terms of probabilities for the case where $p = \frac{1}{2}$.

   The proof in PRP (page 81, equation (22), is combinatorial. Attached is an alternative graphical proof that simply sets up a bijection between the first and second types of paths. It is easy to set up a bijection between the second and third type.

10. Probability mass function for last visit to the origin.

    Consider the case where negative levels are allowed (no absorbing barrier), $p = \frac{1}{2}$, and the player starts in level 0. Let u_{2m} denote the probability that the player is again in level 0 after $2m$ steps. Show that, if T_{2n} denotes the last time before step $2n$ that the player was in level 0, then $\mathbb{P}(T_{2n} = 2k) = u_{2k}u_{2n-2k}$.

11. As a partial proof of Stirling's approximation, try approximating $\log(n!) = \log 1 + \log 2 + ... + \log n$ by the integral $\int_{\frac{1}{2}}^{n+\frac{1}{2}} \log x dx$, and show that for large n, $\log(n!)$ is well approximated by $(n + \frac{1}{2}) \log n - n + \log K$, where K is a constant that must be determined by other means. Exponentiate this result to get Stirling's approximation

$$n! \approx K n^n e^{-n} \sqrt{n}$$

Now use this approximation to show that

$$u_{2n} = \frac{(2n)!}{(n!)^2 2^{2n}} \approx \sqrt{\frac{2}{Kn}}$$

In fact, $K = \sqrt{2\pi}$. This is proved on the handwritten attachment, but it is strictly an aside. So

$$u_{2n} = \frac{(2n)!}{(n!)^2 2^{2n}} \approx \sqrt{\frac{1}{n\pi}}$$

Here is a table of values, computed in Mathematica. Even for values of n as small as 3, the error is only a few percent.

$n = 1$ $u_2 = \frac{1}{2} = 0.5$ $\sqrt{\frac{1}{\pi}} = 0.564$

$n = 2$ $u_4 = \frac{3}{8} = 0.375$ $\sqrt{\frac{1}{2\pi}} = 0.398942$

$n = 3$ $u_6 = \frac{5}{16} = 0.3125$ $\sqrt{\frac{1}{3\pi}} = 0.325735$

$n = 4$ $u_8 = \frac{35}{128} = 0.273438$ $\sqrt{\frac{1}{4\pi}} = 0.282095$

$n = 5$ $u_{10} = \frac{63}{256} = 0.246094$ $\sqrt{\frac{1}{5\pi}} = 0.252313$

$n = 6$ $u_{12} = \frac{231}{1024} = 0.225586$ $\sqrt{\frac{1}{6\pi}} = 0.230329$

12. Arc sine law (distribution function) for last visit to the origin.

    Using Stirling's approximation, show that if $0 < x < 1$, then $\mathbb{P}(T_{2n} \leq 2xn)$ is approximately $\frac{2}{\pi} \arcsin \sqrt{x}$.

13. Probability mass function and arc sine law for sojourn times.

    Again consider the case where negative levels are allowed (no absorbing barrier), $p = \frac{1}{2}$, and the player starts in level 0. Show that the probability that the player is in "positive territory" during $2k$ of the first $2n$ steps (this means that the step either begins or ends at a positive level) is also $u_{2k} u_{2n-2k}$. Show that the fraction x of the time that the player spends in "positive territory" during the first $2n$ turns is also approximately $\frac{2}{\pi} \arcsin \sqrt{x}$.

Notes on using Wallis's product to approximate binomial coefficients and get the constant K in Stirling's approximation are attached. These are based on a proof in the book that I used for freshman calculus in 1959-1960, Differential and Integral Calculus by Courant, volume 1, pp. 223-225. I had meant to type these up in TeX but chose to watch the World Series instead. If anyone is learning TeX and wants to practice, this would be a good document to try. Email me your TeX and I will post it on the Web site.

You are not responsible (in this course) for this proof. Physicists should be aware, though that Stirling's approximation is used almost daily in Physics 181, and they might want to learn this proof. It's also just a great proof, one of the best applications of integration by parts that I know.

Also attached is a graphical proof that there is a bijection between paths that start in level 0 and return to level 0 after $2n$ steps and paths that never return to 0 during the first $2n$ steps. This was inspired by a similar proof in Feller's probability textbook. Here is a graph-free version. (I have trouble putting graphs into TeX – can anyone help out?)

Consider a random walk of $2n$ steps, with n steps to the left and n to the right, e.g. for $n = 6$

$RRLLLLLRLRRR$

This path gets as far as 2 steps to the right, 3 to the left. So the maximum excursion is 3. Identify the step that first leads to this maximum excursion, and use it to break the path into three pieces, in the order 3, 1, 2, like this:

$(RRLLLL)(L)(RLRRR)$

Piece 3 is $RRLLLL$
Piece 1 is L
Piece 2 is $RLRRR$

Now modify these pieces and assemble them into a new path as follows to create a new path:

Piece 1' is piece 1, reversed in time so that a left step becomes a right step.
Piece 1' is R
Piece 2' is the same as piece 2: $RLRRR$. Since this returned to 0 from the maximum excursion (which was to the left in our example), it continues in the same direction as piece 1'. It could get back to level 1 if the maximum excursion was achieved multiple times, but never to level 0.
Piece 3' is piece 3, reversed in time. This means that you write the string backwards and replace R by L, L by R.

$RRLLLL \longrightarrow LLLLRR \longrightarrow RRRRLL$

Since piece 3 went in the direction of the maximum excursion (left), the time-reversed version 3' goes the other way.

Concatenate pieces 1', 2', and 3' to get

$(R)(RLRRR)(RRRRLL) = RRLRRRRRRRLL$ This is a path that never returns to the origin.

To invert this process, let k be the final level achieved by a path of $2n$ steps that never returns to the origin. k must be even; in the example, $k = 6$. Now find the place where level $\frac{k}{2}+1$ is first achieved, and split the path where this happens:

$(RRLRRR)(RRRRLL)$. This recovers piece 3'.
Split off the first step to recover piece 1'
$(R)(RLRRR)(RRRRLL)$
Finally, reverse piece 1' to get 1
Piece 1 is (L)
Reverse piece 3' to get 3
Piece 3 is $(RRLLLL)$
Piece 2 is the same as 2', $(RLRRR)$

Assemble the pieces in the order 3, 1, 2 to get $(RRLLLL)(L)(RLRRR)$

Piece 1 goes 1 step opposite to the direction of the original path. In the example, that path went to the right and step 1 goes left. Piece 3' went from level $\frac{k}{2}+1$ to k, which is $\frac{k}{2}-1$ steps, so its reversal 3 goes $\frac{k}{2}-1$ steps in the same direction as 1 (left, in the example) Piece 2 (same as 2')went from level 1 to level $\frac{k}{2}+1$, a total of k steps (to the right in the example).

So 3 followed by 1 goes k steps in one direction, then 2 goes k steps in the other direction to return to level 0.