Mathematics 1550H – Probability I: Introduction to Probability
Trent University, Summer 2020 (S62)
Assignment #5
Random Walks the Plane

Random decides to take a walk in the Cartesian plane, starting at the origin $(0, 0)$. Random being random, the next step is always decided by simultaneously rolling two fair standard dice, a red one and a blue one. If the blue one comes up 1 or 2, Random moves left one unit; if it comes up 3 or 4 Random doesn't move to either side; if it comes up 5 or 6, Random moves right one unit. At the same time, if the red one comes up 1 or 2, Random moves down by 1; if it comes up 3 or 4, Random doesn't move up or down; if it comes up 5 or 6, Random moves up one unit. For example, if Random is at the point $(3, 1)$ and rolls 1 on the blue die and 4 on the red die, Random will step to the point $(3 - 1, 1 + 0) = (2, 1)$.

Let (X_n, Y_n) denote Random's position after n steps. Of course, $(X_0, Y_0) = (0, 0)$, but after that chance rules.

1. Find the expected values and variances of X_n and Y_n. *[5]*

Solution. This is a bit easier to analyze if you abstract things a little. Let B_k be the random variable that tells us how Random moves horizontally, *i.e.* whether Random moves one unit to the left, doesn't move to either side, or moves one unit to the right, at step k of the walk. Similarly, let R_k be the random variable that tells us how Random moves vertically, *i.e.* whether Random moves one unit down, doesn't move up or downl, or moves one unit up, at step $k \geq 1$ of the walk. Then the various B_k and R_k are all independent of one another (since the die rolls that ultimately define them are) and are identically distributed, with probability distribution function given by $m(-1) = m(0) = m(1) = \dfrac{1}{3}$. It follows that, for all $k \geq 1$,

$$E(B_k) = E(R_k) = (-1) \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} = 0\,.$$

Since

$$E\left(B_k^2\right) = E\left(R_k^2\right) = (-1)^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3} = \frac{2}{3}\,,$$

we also have that $V(B_k) = V(R_k) = E\left(B_k^2\right) - [E(B_k)]^2 = \dfrac{2}{3} - 0^2 = \dfrac{2}{3}$.

Random's position (X_n, Y_n) after $n \geq 1$ steps is determined by the sums $X_n = B_1 + B_2 + \cdots + B_n$ and $Y^n = R_1 + R_2 + \cdots R_n$. Thus, for $n \geq 1$,

$$\begin{aligned} E(X_n) &= E(B_1 + B_2 + \cdots + B_n) = E(B_1) + E(B_2) + \cdots + E(B_n) = n \cdot 0 = 0 \\ \text{and } E(Y_n) &= E(R_1 + R_2 + \cdots + R_n) = E(R_1) + E(R_2) + \cdots + E(R_n) = n \cdot 0 = 0\,, \end{aligned}$$

because expected value is linear. Since the B_k and R_k are independent, their variances also add, so

$$\begin{aligned} V(X_n) &= V(B_1 + B_2 + \cdots + B_n) = V(B_1) + V(B_2) + \cdots + V(B_n) = n \cdot \frac{2}{3} \\ \text{and } V(Y_n) &= V(R_1 + R_2 + \cdots + R_n) = V(R_1) + V(R_2) + \cdots + V(R_n) = n \cdot \frac{2}{3}\,. \quad \square \end{aligned}$$

2. Explain why it is hard to compute the expected value and variance of the distance that Random is from the origin after n steps. Can you suggest reasonable alternatives to the expected value and/or variance of distance that are easier to compute? *[5]*

SOLUTION. After n steps, Random's position is (X_n, Y_n), which is a distance of $\sqrt{X_n^2 + Y_n^2}$ from the origin. While the expected value operator plays nice with addition and multiplication by constants, it does not do so with other operations, such as squaring or taking square roots. This means that to compute the expected value of the distance from the origin after n steps, $E\left(\sqrt{X_n^2 + Y_n^2}\right)$, we would have to fall back on the definition of expected value. Note that, as observed in the solution to **1** above, X_n and Y_n are each sums of n independent and identically distributed random variables – and which are also identically distributed as the ones used to sum the other of X_n or Y_n – X_n and Y_n must have the same probability distribution function, say $p(k)$. Moreover, note that each of X_n and Y_n must be an integer between $-n$ and n inclusive. Then, by definition, the expected value of the distance from the origin is given by:

$$E\left(\sqrt{X_n^2 + Y_n^2}\right) = \sum_{i=-n}^{n} \sum_{j=-n}^{n} \sqrt{i^2 + j^2} \cdot p(i)p(j)$$

Aside from the problem of actually working out the probability distribution function $p(k)$, which is a somewhat modified binomial distribution, this sum is unlikely to easily yield some nice formula in terms of n because of the square roots involved. While one can simplify things a bit by exploiting the facts that $p(-k) = p(k)$ for all k and $\sqrt{(\pm i)^2 + (\pm j)^2} = \sqrt{i^2 + j^2}$ for all i and j, having to compute the sum directly for each n would still be an awful chore.

Are there reasonable alternatives to the expected value (and variance) of distance proper that are easier to handle? There are any number of things one could try.

First, the naive alternative of simply using $\sqrt{[E(X_n)]^2 + [E(Y_n)]^2}$ won't do. Since $E(X_n) = E(Y_n) = 0$ for all $n \geq 0$, we will also have $\sqrt{[E(X_n)]^2 + [E(Y_n)]^2} = \sqrt{0^2 + 0^2} = 0$ for all n. Even a casual inspection of the sum given by its definition tells us that $E\left(\sqrt{X_n^2 + Y_n^2}\right) > 0$ when $n > 0$, so a an laternative that is always 0 is *not* a reasonable one.

A more reasonable idea would be to use "taxicab distance", *i.e.* distance computed while moving only horizontally plus moving only vertically, instead of the usual notion of distance in the plane. This would make the "distance" of the point (x, y) from the origin be $|x| + |y|$, which is a little easier to work with than $\sqrt{x^2 + y^2}$. There are a couple of reasons, besides relative ease of computation, why this is not an unreasonable alternative. For one, Random, by moving -1, 0, or 1 units in each of the x and y directions at each step, is effectively moving like a taxicab navigating unit length city blacks, making the "taxicab distance" a natural fit. For another, the "taxicab distance" of any point from the origin is always no more than $\sqrt{2}$ times its actual distance from the origin (Why?), so "taxicab distance" is not too awful an approximation for actual distance.

Another fairly reasonable idea is to use the square of distance instead of distance, thus eliminating the square root. It turns out that we do not even have to resort to the definition of expected value in this case. Since $E(X_n) = E(Y_n) = 0$ for all $n \geq 0$, we have $V(X_n) = E\left(X_n^2\right)$ and $V(Y_n) = E\left(Y_n^2\right)$ for all $n \geq 0$. Using this in reverse gives us:

$$E\left(X_n^2 + Y_n^2\right) = E\left(X_n^2\right) + E\left(Y_n^2\right) = V(X_n) + V(Y_n) = \frac{2}{3}n + \frac{2}{3}n = \frac{4}{3}n$$

We can then try to approximate the expected value of actual distance, if we need that instead of the square of distance, by the square root of the expected value of the square of distance, *i.e.* use $\sqrt{E\left(X_n^2 + Y_n^2\right)} = \sqrt{\frac{4}{3}n} = \frac{2\sqrt{n}}{\sqrt{3}}$ as a proxy for $E\left(\sqrt{X_n^2 + Y_n^2}\right)$. (Is this a good approximation? It'll take some work with computing $E\left(\sqrt{X_n^2 + Y_n^2}\right)$ the hard way to settle that ...) The variance of the square of distance is also pretty easy to compute, but I'll leave that to you. ■