Random Walks

Mohan Agrawal *Oct 28, 2021*

1 Random walk derivation

Using the fundamental definition of random walks, we can write

$$Y_n = Y_{n-1} + e_i \qquad e_i \sim \mathcal{N}(0,1) \tag{1}$$

or equivalently

$$Y_n = Y_0 + \sum_{i=1}^{n} e_i \tag{2}$$

As we are free to choose our starting point, we can always set $Y_0 = 0$ in eqn. (2).

Eqn. (2) also allows us to obtain the following useful relations:

$$\langle Y_n \rangle = 0 \tag{3}$$

and since each e_i is independently and identically distributed under the standard normal distribution, $Cov(e_i, e_j)_{i \neq j} = 0$. Therefore, we can use the fact that $Var(x + y) = Var(x) + Var(y)$ and write

$$Var(Y_n) = Var\Big(\sum_{i=1}^{n} e_i\Big) = \langle Y_n^2 \rangle = n \tag{4}$$

The last equality uses the definition of variance $Var(X) = \langle (X - \langle X \rangle)^2 \rangle$. (More on this below.) If the spread of the Normal distribution was σ instead of 1, the above result would instead be $n\sigma^2$.

Let us now consider the following relation

$$\langle (Y_n - Y_{n+\delta})^2 \rangle = |\delta| \tag{5}$$

The RHS of above equation comes from the fact that $Y_{n+\delta}$ is δ more steps further from Y_n. We can show if in the following manner. Using (2) we can write

$$Y_{n+\delta} - Y_n = \sum_{i=1}^{\delta} e_i$$

So that LHS of (5) becomes

$$\Big\langle \sum_{i=1}^{\delta} \sum_{j=1, j\neq i}^{\delta} e_i e_j + \sum_{i=1}^{\delta} e_i^2 \Big\rangle$$

where I have broken up the square of the summation into two parts. The expectation of first term is zero, and the expectation of the second term is just δ, both obtained using the fact that e_i's are i.i.d. Thus, proving equation (5). This method was also used to obtain the last equality in equation (4). Now let us expand the LHS of equation (5) to get

$$\left\langle Y_n^2 + Y_{n+\delta}^2 - 2Y_nY_{n+\delta}\right\rangle = |\delta| \tag{6}$$

If we write out two equations, using $\delta > 0$ and $\delta < 0$, where we expand the second term on LHS – $Y_{n\pm\delta}$, we see that both equations reduce to:

$$\left\langle Y_nY_{n\pm\delta}\right\rangle = n \tag{7}$$

Since equation (8) is nothing but the Covariance of the two terms, we could take one more step and calculate the autocorrelation function of Y_n. Correlation is generally taken as:

$$Corr(x, y) = \frac{Cov(x, y)}{\sigma_x\sigma_y} \tag{8}$$

So that we obtain

$$Corr(Y_n, Y_{n+\delta}) = \sqrt{\frac{n}{n+\delta}}$$

utilizing the fact that $\sigma(Y_n) = \sqrt{n}$. However, we can simplify this further by replacing n with $n - \delta$ (it's just a time-series).

$$Corr(Y_{n-\delta}, Y_n) = \sqrt{\frac{n-\delta}{n}} = \sqrt{1 - \frac{\delta}{n}} \tag{9}$$

After a lot of steps, (enough time has passed), n would be large. This allows us to Taylor expand equation (9) to get the final result:

$$\boxed{Corr(Y_{n-\delta}, Y_n) = 1 - \frac{2\delta}{n}} \tag{10}$$

I would mention here that there is no real need to actually start from equation (5). We could directly calculate the LHS of equation (8) by writing Y's using the fundamental definition of equation (2) and then employing the properties of i.i.d. variables that we used to prove equation (5).

In class, Jon assumed the expectation of the first two terms in equation (6) to be roughly some large number N, instead of substituting the values like I have done. He, therefore, got a result that retained the δ dependence:

$$\left\langle Y_nY_{n+\delta}\right\rangle = \frac{1}{2}(2N - |\delta|) \tag{11}$$

But he did not finish his calculation. I assume that he would have gone on to divide his result by N, and in that case my and his result would match.

The question here is, what exactly are we after – is it equation (8) or equation (11) that we're concerned most about? From equation (8) (which gives the Covariance), it seems that random walk is not stationary at all, since it depends only on the position, and not on the separation between two points. On the other hand the correlation does depend on δ. **In fact, we see that the closer two points are, the more strongly they are correlated.** Since this is what we expect from a random walk, I think it's really equation (11) that we were after.

2 Fourier Transform

Starting with equation (11), with a form like:

$$\langle Y_n Y_{n+\delta} \rangle = c - |\delta| \tag{12}$$

Noting that our correlation function $g(\delta)$ is even, and since Fourier transform assumes any input function (array) to be periodic i.e. $x[N-k] = x[-k]$, we can split the following summation

$$G[k] = \sum_{\delta=0}^{N} g[\delta] e^{-2j\pi k\delta/N} \tag{13}$$

into two parts that are given as:

$$G[k] = \sum_{\delta=0}^{N/2} g[\delta] e^{-2j\pi k\delta/N} + \sum_{\delta=0}^{N/2-1} g[-\delta] e^{2j\pi k\delta/N} - g[0] \tag{14}$$

The last term is because of double counting of $g[0]$ in the summations.

Now, using the even property of $g[\delta]$ from eqn.(12), eqn.(14) becomes simply:

$$G[k] = Nc - \sum_{\delta=0}^{N/2} \delta e^{-2j\pi k\delta/N} - \sum_{\delta=0}^{N/2-1} \delta e^{2j\pi k\delta/N} \tag{15}$$

A closed form expression for above equation can be obtained using software like Mathematica or one could go about solving it using the rules of Arithmetico-Geometrico Progression (AGP).