Physics through Computational Thinking

Random walks: The central limit theorem.

Auditya Sharma and Ambar Jain
Dept. of Physics, IISER Bhopal

Outline

In this module we will cover

1. The central limit theorem. Numerical verification.
2. A general argument for why diffusive motion appears in arbitrary dimensions.

The Central Limit Theorem

The reason for the ubiquitous appearance of the Gaussian distribution has to do with a deep and important theorem called the Central Limit Theorem. It is difficult to prove, but fairly easy to state in a non-rigorous way. Our goal here is to verify it numerically for a number of cases.

Statement (Non-rigorous)

Let $X_1, X_2, \ldots, X_N$ be N independent random variables with means $\mu_1, \mu_2, \ldots, \mu_N$ and standard deviations $\sigma_1, \sigma_2, \ldots, \sigma_N$ respectively. Then the distribution of the sum

$$S = X_1 + X_2 + \ldots + X_N$$

tends to a Gaussian distribution with mean $\mu = \mu_1 + \mu_2 + \ldots + \mu_N$, and variance $\sigma^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_N^2$.

Exercise

- Numerically generate a few sample random walks where the lengths of individual steps are drawn from a uniform distribution in the interval $[-1, 1]$ and visualize them.
- Check if the key result $<m^2> = \alpha N$ is still robust by numerically generating many samples of random walks, and averaging over them.
- Analyticaly compute α. Check by explicit plotting if your numerical data support this expectation.

Solution

```
a = {0};
Do[AppendTo[a, a[[n - 1]] + 1 - 2 RandomReal[]], {n, 2, 500}]
ListPlot[a, Joined → True];

data = Table[{n, Table[Table[1 - 2 RandomReal[], {n}] // Total, {10 000}]^2 // N // Mean}, {n, 100, 1000, 100}];

Show[ListPlot[data], Plot[1/3 x, {x, 0, 1500}]];
```

Exericse

What would be the distribution of the final position you would expect in this case? Test it numerically.

Solution

```
nMax = 1000;
binsize = 10;
histdata = Histogram[Table[Table[1 - 2 RandomReal[], {nMax}] // Total, {10 000}], Automatic, "PDF"];
```

$$\texttt{Show}\left[\texttt{histdata, Plot}\left[\sqrt{\frac{3}{2\,\pi\,\texttt{nMax}}}\;e^{-x^2/(2\,\texttt{nMax}/3)},\ \left\{x,\ -5\sqrt{\texttt{nMax}},\ 5\sqrt{\texttt{nMax}}\right\}\right]\right];$$

Exericse

Carry out the same exercise as above but with individual lengths drawn from a normal distribution of mean zero and standard deviation unity. Play with your code to see what happens if you choose a different mean and standard deviation.

Solution

```
a = {0};
Do[AppendTo[a, a[[n - 1]] + RandomVariate[NormalDistribution[]]], {n, 2, 500}]
ListPlot[a, Joined → True];

data = Table[{n, Table[Table[RandomVariate[NormalDistribution[]], {n}] // Total, {10 000}]^2 // N // Mean},
   {n, 100, 1000, 100}];


Show[ListPlot[data], Plot[x, {x, 0, 1500}]];

nMax = 1000;
binsize = 10;
histdata = Histogram[Table[Table[RandomVariate[NormalDistribution[]], {nMax}] // Total, {10 000}], Automatic, "PDF"];
```

$$\texttt{Show}\left[\texttt{histdata, Plot}\left[\sqrt{\frac{1}{2\,\pi\,\texttt{nMax}}}\;e^{-x^2/(2\,\texttt{nMax})},\ \left\{x,\ -5\sqrt{\texttt{nMax}},\ 5\sqrt{\texttt{nMax}}\right\}\right]\right];$$

A general argument for the scaling of net displacement.

The law of diffusive scaling (Net displacement proportional to the square root of the time traversed) is very general, and seems to hold in a wide variety of contexts. In particular it does not depend on dimensionality. Here is an argument why that happens. Consider a random walk in which each of the steps is a vector. Let us call them $\vec{r_1}$, $\vec{r_2}, \vec{r_3}, \ldots, \vec{r_N}$. Since every direction is assumed to be equally likely the average of each of the vectors must be zero. That is

$$< \vec{r_j} > \; = 0. \tag{1}$$

We can take the typical length of each vector to a. This information appears in the mean of the square of each vector. That is

$$< \vec{r_j}.\vec{r_j} > \; = a^2. \tag{2}$$

The net displacement is then given by

$$\vec{r} = \vec{r_1} + \vec{r_2} + \vec{r_3} + \ldots + \vec{r_N}. \tag{3}$$

The average of the net displacement is also zero.

$$\begin{aligned} < \vec{r} >= \; & < \vec{r_1} > + < \vec{r_2} > + < \vec{r_3} > + \ldots + < \vec{r_N} > \\ = \; & 0 + 0 + 0 + \ldots + 0 = 0. \end{aligned} \tag{4}$$

The average of the square of the net displacement is

$$< \vec{r}.\vec{r} >= \; < \vec{r_1}.\vec{r_1} > + < \vec{r_2}.\vec{r_2} > + < \vec{r_3}.\vec{r_3} > + \ldots + < \vec{r_N}.\vec{r_N} > + \sum_{i,j} < \vec{r_i}.\vec{r_j} > \tag{5}$$

Now comes the key aspect of the argument. The random walk is assumed to be memory-less. So every vector is completely uncorrelated with every other. Mathematically this implies that

$$\begin{aligned} < \vec{r_i}.\vec{r_j} > \; &= \; < \vec{r_i} > . < \vec{r_j} > \\ &= 0. \end{aligned} \tag{6}$$

Therefore

$$< \vec{r}.\vec{r} >= a^2 + a^2 + a^2 + \ldots + a^2 = N\, a^2, \tag{7}$$

which is nothing but the statement of diffusive motion.