Lecture 2: Sums of Random Vectors

David Hu (Department of Mathematics)

February 6, 2003

Handouts: Hughes, Chapter 2: Random Walks and Random Flights.

Optional reading: 18.325 Lecture notes Lecture 8 and 9. For much more detail, See Hughes Sec 1.4 and 2.

Definitions

A random flight. This is distinguished from a walk in that during every time step we can jump or fly immediately to a place randomly selected.

Random walk (RW). Here, we pick a place where we want to end up, and go there with constant velocity. The position distribution is the same, but determining the time that it takes to get there is much more subtle. [e.g. if random walk involves very large steps, then flight can differ from a walk. For example in Levy distributions, one takes such big steps that if one waited to walk there, the scaling in time completely changes.]

1 Probability distribution function for the random walk

Starting at the origin $x_0 = 0$, take N steps. Then the position after the random walk, X_N is the sum of random variables where we have Δx_n as the displacements, which in principle can be non-identical (hence subscript n):

$$X_n = \sum_{n=1}^{N} \Delta x_n \tag{1}$$

Let $p_n(X)$ denote the probability density function (PDF) for the random variable X_n. By PDF, we mean the probability of being in some volume Ω is the integral over the volume of that PDF. This is the multidimensional integral

$$P(X \in V) = \in_\Omega \int\int_\Omega p(x)dx \tag{2}$$

The question we ask today is: what is the probability one lands at position X at time N? Note that the analysis is the same whether X is a vector or not. In my analysis, we won't treat discrete cases them separately. For these we will use delta functions.

Now we make some assumptions. Key: independent steps. (There's not much you can say about non-independent variables). If these displacements are independent, then we can write down an exact recursion for the PDF which is basically Bachelier's equation from last time. Fancy versions of this are called Kolmogorov or Rayleigh's equation. If you want to know what the probability of being at position X after N steps, we write down a recursion relation. We integrate over all places we may have been. (We did this last time in 1-D). Say we want to be at position X_{N+1} at time $N+1$. We have to consider all the different places X_N, where we could have been, and we have to take a displacement from these places to $X_{N+1} = x$. Integrating over all possible places we could have been on the previous step,

$$P_{N+1}(x) = \int_{x'} P(x_{N+1} = x | x_N = x') P(x_N = x') dx' \tag{3}$$

Since they steps are independent, we can write them as the product of the probabilities. Independence also allows us to drop the conditional.

$$P_{N+1}(x) = \int_{x'} p_{N+1}(x - x') P_N(x') dx' \tag{4}$$

What remains is just a convolution, which we denote by $*$

$$P_{N+1}(x) = P_N * p_{N+1} \tag{5}$$

So for identical displacements, we have the convolution product

$$P_N(x) = p^{*N}(x) \tag{6}$$

and for non-identical displacements,

$$P_N(x) = p_1 * p_2 * \cdots * p_N(x). \tag{7}$$

The Fourier transform is a very useful tool in the context of discrete random walks. Using Fourier transforms, we can now introduce

$$\hat{P}_N(k) = \int_{-\infty}^{\infty} e^{ikx} P_N(x) dx \tag{8}$$

which can be written as $\hat{P}_N(k) = \hat{p}_1(k)\hat{p}_2(k) \ldots \hat{p}_N(k)$. In the case of identical steps, this simplifies to $\hat{P}_N(k) = (\hat{p}_1(k))^n$.

Using now the inverse transform, the formal solution for the probability distribution of the position of the walker is

$$P_N(x) = \int_{-\infty}^{\infty} e^{-ikx} \prod_{n=1}^{N} \hat{p}_n(k) \frac{dk}{(2\pi)^d} \tag{9}$$

where d is the dimension. This is an exact solution, although it may not be easy to integrate. Hence knowing the characteristic function is equivalent to knowing the probability distribution. Usually one will have to do an long-term asymptotic analysis as $n \to \infty$ to progress further. Notice that this is equivalent to the Central limit theorem. Notation: In probability books, "characteristic function" is used interchangeably for Fourier transform. Its not just an arbitrary that this is the case. For example, its always true for a characteristic function that $\hat{p}(0) = 1$. This says that the expected value of 1 is 1, which properly normalizes your function. The characteristic function must also be positive, a property which the Fourier transform satisfies. The characteristic function for displacements is called the structure function in the random walk literature.

Now we take an aside to derive the convolution theorem for Fourier transforms. The Fourier transform is a very useful tool in the context of discrete random walks.

2 The convolution theorem

The most important property of the Fourier transform though is that is converts convolutions into multiplications:

$$\widehat{f * g}(k) = \hat{f}(k)\hat{g}(k). \tag{10}$$

This follows from writing the integrals after a change of variables. We write

$$\begin{aligned} \widehat{f * g}(k) &= \int_{-\infty}^{\infty} dx e^{-ikx} \int_{-\infty}^{\infty} dy f(x-y) g(y) && (11) \\ &= \int_{-\infty}^{\infty} dy g(y) \int_{-\infty}^{\infty} dx f(x-y) e^{-ikx} && (12) \end{aligned}$$

where we have inverted the order of integration by assuming both integrals are absolutely convergent. If we now let $z = x - y$ we find

$$\widehat{f * g}(k) = \int_{-\infty}^{\infty} dy g(y) e^{-iky} \int_{-\infty}^{\infty} dz f(z) e^{-ikz} = \hat{f}(k)\hat{g}(k). \tag{13}$$

One can arrive at the same result in the case of a discrete Fourier transform:

$$\hat{f}(k)\hat{g}(k) = \left(\sum_y e^{-iky} f(y)\right)\left(\sum_z e^{-ikz} g(z)\right) \tag{14}$$

$$= \sum_x e^{-ikx}\left(\sum_y g(y) f(x-y)\right) \tag{15}$$

$$= \widehat{f * g}(k) \tag{16}$$

where all the terms with an exponent equal to z have been collected on the right hand side of the second line. We also used in both derivations the property $e^{-iky}e^{-ikz} = e^{-ik(y+z)}$ which is the key property here.

3 Examples of random walks

All these examples will involve identical steps. We'll return to non-identical steps later.

3.1 Bernoulli RW

Consider the random walk on the line, with p the probability of stepping the left and q the probability of stepping to the right. Then $p + q = 1$ and

$$p(x) = p\delta(x + a) + q\delta(x - a) \tag{17}$$

If $p = q$, then

$$p(x) = \frac{1}{2}[\delta(x + a) + \delta(x - a)] \tag{18}$$

This is the random walk on the integers. The displacement is a Bernoulli random variable. To generalize the Bernoulli RW, we can walk on a square lattice or even more generally, a hyper-cubic lattice. We'll calculate the walk on the square lattice. Pearson generalized it to take a step in any direction.

3.2 Pearson's RW (d=2)

(isotropic)

$$p(\vec{x}) = \frac{\delta(|\vec{x}| - a)}{2\pi a} \tag{19}$$

We divide by the perimeter to properly normalize it. It has a uniform random angle, but a fixed step length. If we substitute (20) into (9), we get a mess: Bessel functions and Hankle functions. We can calculate up to $n = 3$; afterward we have to rely on long-term asymptotics.

3.3 Rayleigh's RW

(isotropic) Rayleigh looked at this problem in the context of sound waves, since sound waves all have the same amplitude, but have arbitrary phase. So there are random walks in wave-vector space. Again, there is no preferred direction. But now we divide by the surface area of a sphere.

$$p(\vec{x}) = \frac{\delta(|\vec{x}| - a)}{4\pi a^2} \tag{20}$$

Note this is easier to handle than 2-D. The above three walks are isotropic.

3.4 Hypercubic RW

Today's worked example will be the walk on the hypercubic lattice. The distribution function is a sum of delta functions over all the neighbors on the hypercubic lattice. In every direction there are 2 neighbors so,

$$p(\vec{x}) = \frac{1}{2d} \sum_j \delta(\vec{x} - \vec{x_j}) \tag{21}$$

The structure function is the Fourier transform, which gives exponentials

$$\hat{p}(\vec{k}) = \frac{1}{2d} \sum_j e^{i\vec{k}\cdot\vec{x_j}} \tag{22}$$

Since one can move in either of two directions ($\pm a$), we can sum in each direction

$$\hat{p}(\vec{k}) = \frac{1}{d} \sum_j cos(k_j a). \tag{23}$$

where $\vec{k} = (k_1, k_2, \ldots, k_d)$. We will substitute this into (9) to get the probability distribution. (Note we could have used combinatorics for this case) The position probability distribution after N steps may be expressed as

$$\hat{P_N}(\vec{x}) = \int e^{-\vec{k}\cdot\vec{x}} \left(\frac{1}{d} \sum_{j=1}^{d} cos(k_j a) \right)^N \frac{d\vec{k}}{(2\pi)^d}. \tag{24}$$

To actually do the inversion, its easiest to work with cases d =1,2.

3.5 Probability distribution for walk on the integers

Since d = 1, we use scalars rather than vectors.

$$\hat{P_N}(x) = \int e^{-kx} \left(\frac{1}{d} cos(k_j a)\right)^N \frac{dk}{2\pi} = \int e^{-kx} cos^N(k_j a) \frac{dk}{2\pi}. \tag{25}$$

We'll write the cosines as exponentials for this case. [In the end, the probability distribution function will still involve delta functions because we're still on the lattice rather than any position in space.] Using the binomial theorem we get a sum,

$$\hat{P_N}(x) = \int e^{-ikx} \sum_{n=0}^{N} e^{ikan} - e^{-ika(N-n)} \left(\frac{1}{2}\right)^N \frac{dk}{2\pi}. \tag{26}$$

Now Fourier transforming the first exponential into delta function. The Fourier transform of delta function is 1: $1 = \int e^{ikx}\delta(x)dx$. Hence the inverse transform of 1 is a delta function, but rather than evaluated at x, it is evaluated at the value specified: $\delta(x) = \int e^{-ikx} dk$.

$$\hat{P_N}(x) = \sum_{n=0}^{N} \binom{N}{n} \left(\frac{1}{2}\right)^N \left(\frac{1}{2}\right)^{N-n} \delta(x - (2n - N)a). \tag{27}$$

Here n is the number of steps to the right, in the + direction. So if we take all our steps to the right, we appear at position $x = Na$. If we take all our steps to the left, $n = 0$, and we're at $x = -Na$.

3.6 Probability distribution for Rayleigh's RW

Here we will use the binomial theorem twice. We write

$$\hat{P_N}(\vec{k}) = \left(\frac{cos(k_x a) + cos(k_y a)}{2}\right)^N \tag{28}$$

Using the binomial theorem for the first time,

$$\hat{P_N}(\vec{x}) = \sum_{n_x=0}^{N} \binom{N}{n_x} \left(\frac{1}{2}\right)^N cos^{n_x}(k_x a) cos^{N-n_x}(k_y a) \tag{29}$$

Following the same procedure, we write the distribution after N steps

$$\hat{P_N}(\vec{x}) = \frac{1}{(2d)^N} \sum_{n_x=0}^{N} \binom{N}{n_x} \sum_{n_x+}^{n_x} \binom{n_x}{n_{x+}} \delta(x-(2n_{x_+}-n_x)a) \sum_{n_{y+}}^{N-n_x} \binom{N-n_x}{n_{y+}} \delta(y-(2n_{y^+}-(N-n_x)a)) \tag{30}$$

To progress any further, we go back to integrals. This is helpful in getting average properties, for which we will use asymptotics.

3.7 Probability distribution for arbitrary dimension, $N \to \infty$

We will use the precise way of doing this, the saddle point method later. That will give us a uniformly valid approximation that is valid for any value of x, simply for large N. More naively we can use Laplace's method to approximate the probability distribution for typical values of x. (This is not valid for the tails of the distribution.) Now, the structure function is being raised to a large power. We want to find the dominant contribution using Laplace's method. We can write the distribution in a way that's more suggestive:

$$\hat{P_N}(\vec{x}) = \int e^{-\vec{k}\cdot\vec{x}} e^{N\Psi_n(\vec{k})} \frac{d\vec{k}}{(2\pi)^d} \tag{31}$$

where

$$\Psi_N = log\left(\frac{1}{d}\sum_{j=1}^{d} cos(k_j a)\right) \tag{32}$$

The dominant contribution is the region near $\vec{k} = \vec{0}$ as $N \to \infty$. To get the asymptotic expansion, I'm going to use an expansion of cosine and log near $k = 0$.

$$\Psi_N = log\left(1 - \frac{\vec{k}^2 a^2}{2d} + \cdots\right) \sim \frac{\vec{k}^2 a^2}{2d} + \cdots \tag{33}$$

$$\hat{P_N}(\vec{x}) = \int e^{-i\vec{k}\cdot\vec{x}} e^{\frac{\vec{k}^2 a^2}{2d}} \frac{d\vec{k}}{(2\pi)^d} \tag{34}$$

Note this is a Gaussian in higher dimensions. It one of the few integrals we know how to do, and becomes a Gaussian once again. Completing the square,

$$\hat{P_N}(\vec{x}) = \int e^{-\left(\vec{k}a\sqrt{\frac{N}{2d}} + i\frac{\vec{x}}{2a}\sqrt{\frac{2d}{N}}\right)^2} \frac{d\vec{k}}{(2\pi)^d} e^{-\frac{\vec{x}^2 d}{2a^2 N}} \tag{35}$$

Now we eliminate the $O(N^{-1/2})$ term. [Proof: Consider writing this as an integral in 2 dimensions. We can show that $I = \int e^{-(k-k_o)^2} = \int e^{-k^2}$,i.e, we can drop the k_0 (complex). The interested reader can argue this by using contour integration and eliminating parts of the line integral as $N \to \infty$.]

$$\hat{P_N}(\vec{x}) = \int e^{-\frac{a^2 N}{2d}\vec{k}^2} \frac{d\vec{k}}{(2\pi)^d} e^{-\frac{\vec{x}^2 d}{2a^2 N}} \tag{36}$$

$$\hat{P_N}(\vec{x}) = \frac{1}{(2\pi)^d}\left(\frac{2\pi}{a^2 N/(2d)}\right)^{d/2} e^{-\frac{\vec{x}^2 d}{2a^2 N}} \tag{37}$$

$$\hat{P_N}(\vec{x}) = \left(\frac{d}{2\pi a^2 N}\right)^{d/2} e^{-\frac{\vec{x}^2 d}{2a^2 N}} \tag{38}$$

This is a properly normalized d-dimensional Gaussian. When we add random variables, no matter what the dimension, we always get a Gaussian; this phenomenon is an example of the Central Limit Theorem (CLT). The width of the distribution is $\sqrt{n}$.

Prof. Bazant's quote of the day: "It is valid almost everywhere, but not in the mathematical sense. That was a measure theory joke."