Lecture 1: Introduction to Random Walks and Diffusion

K. Titievsky

Department of Chemical Engineering, MIT

February 4, 2003

Topics

- Syllabus and logistics
- Historical background for random walks
- Statement of the basic random walk problem
- Discrete approach
- Central limit theorem
- Continuous approach
- Connection between the continuous and discrete treatments
- Variations on the basic problem

Objective

To introduce the basic problem of random walks, sketch the essential results of continuum and discrete treatments of the problem and introduce, in the established context, the possible issues and extensions to the problem.

HANDOUT: Historical background from Hughes, *Random Walks and Rangom Environments* Vol. 1, Oxford, 1995, chapter 2.

The Original 'Random Walk' (Pearson 1905; Rayleigh 1880)

Karl Pearson posed the following problem in *Nature* as a model for mosquito infestation: A walker makes steps of a fixed length a and at a random angle θ (in 2-D polar coordinates) after each time interval Δt. Where is the walker after N steps? Lord Rayleigh had already solved the same problem in the context of acoustics in heterogeneous media: In the limit $N \to \infty$? the PDF (probability density function) has the asymptotic form,

$$p(\mathbf{r}, N) \sim \frac{1}{n} \exp(\frac{-\mathbf{r}^2}{2N}) \tag{1}$$

The scaling, $\langle |r| \rangle \sim \sqrt{N}$, is very common for random walks. If the position is simply a sum of a large number independent random displacements, this scaling and the shape of the distribution come naturally out of the *Central Limit Theorem* (CLT).

Discrete approach and CLT

A rough statement of CLT is as follows. Consider a random walk (RW) in one dimension with the following properties:

1. The walk consists of N steps Δx_i.
2. Each step is an *independent identically distributed random variable* (**iid r.v.**), that is:
   - Each step is independent of the prevous steps
   - The length and direction have the same probability density distributions for all steps step.
3. Δx_i has a finite mean and variance:

$$\langle \Delta X_i \rangle = \mu < \infty \tag{2}$$

$$\sigma^2 = \langle \Delta X_i^2 \rangle - \langle \Delta X_i \rangle^2 < \infty \tag{3}$$

The position of the walker after N such steps is

$$X_N = \sum_{i=1}^{N} \Delta X_i \tag{4}$$

We now define the statistic (variable) Z_N, as a normalized version of X_N:

$$Z_N = \frac{X_N - \mu N}{\sigma \sqrt{N}} \tag{5}$$

The CLT states that

$$p(Z_N) \sim \frac{1}{\sqrt{2\pi}} \exp(-\frac{z^2}{2\sigma}) \tag{6}$$

which implies

$$\langle (X_N - \mu N)^2 \rangle \sim \sigma^2 N \tag{7}$$

In other words, the PDF of X_N, when properly normalized, converges to Gaussian (or normal) distribution as $N \to \infty$. The rate of convergence depends on the precise assumptions about the walk. Various changes in these assumptions will be studied in the later parts of the course. In the simple case where the CLT holds, the convergence to a gaussian distribution occurs only in the central region *i.e.*, for $|Z_N| \sim O(1)$, while the tails of the distribution retain non-Gaussian features.

Continuum approach

It turns out that diffusion problems are closely related to the random walk discussed above. We will consider Fickian diffusion. Let $c(x,t)$ be a concentration of some quantity as a function of position and time and $\mathbf{F}(x,t)$ is the flux density (quantity per area per time). Assume also that the quantity of interest is conserved in a sense that it can neither be created nor destroyed, as energy or mass. According to Fick's law (or Fourier's law),

$$\mathbf{F}(x,t) = -D \nabla c \tag{8}$$

Where D is a diffusion coefficient. The conservation equation for a macroscopic volume V bounded by a surface Σ can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\int\limits_V c\ \mathrm{d}V = \oint\limits_\Sigma \mathbf{n}\cdot\mathbf{F}\mathrm{d}a \tag{9}$$

In differential form (using the divergence theorem in the limit $V \to 0$) this becomes:

$$\frac{\partial c}{\partial t} = -\nabla\cdot\mathbf{F} \tag{10}$$

Using Fick's law and assuming that D is constant, we obtain:

$$\frac{\partial c}{\partial t} = D\nabla^2 c \tag{11}$$

The connection to the random walks becomes apparent when this equation is solved for an initial condition $c(x,0) = \delta(0)$, where δ is the Green function (Dirac δ function). In this case, $c(x,t)$ is also Gaussian. For a one dimensional problem, the solution takes on the following form

$$c(x,t) = \frac{1}{\sqrt{4\pi Dt}}\exp\left(-\frac{x^2}{4Dt}\right) \tag{12}$$

Connection betweent the continuum and discrete approaches

The connection is not incidental. In fact it is possible to show that the concentration of simple random walkers approximately obeys the diffusion equation as their number becomes large, at time scales and length scales greatly exceeding those of a single step. Because the walkers are independent, this is equivalent to the $N \to \infty$ limit the of $p(X_N)$ (eq. 4). As a tentative justification for this equivalence, consider a large number M of independent random walkers starting at x = 0 (*i.e.* $x(t=0) = \delta(0)$). After each is allowed to take a sufficiently large number of steps $N = t/\Delta t$, on average we will find $p(X_N)M$ walkers in $(x, x+\mathrm{d}x)$. We have already found that $p(X_N)$ is Gaussian.

A more formal derivation is outlined below. Let
$p(\Delta x)$ = PDF (probability density function) for each step Δx (as defined section on the Discrete approach).
$P_N(x)$ = PDF for X_N.
The independence of the steps can be expressed as

$$P_{N+1}(x) = \int\limits_{-\infty}^{+\infty} p(\Delta x)P_{N-1}(x-\Delta x)\mathrm{d}\delta x \tag{13}$$

This equation states the independence of Nth step from the previous history of the walk. It is sometimes referred to as *Bachelier's equation* since it was analyzed in his 1900 PhD thesis on options pricing.

Since Δx is assumed to have a finite variance σ^2, and X_N will have a variance that scales with N (eq. 7), we expect that $\Delta x \ll x \quad$ for $\quad N \to \infty$.

Thus, Taylor-expanding P_N

$$P_{N+1}(x) \approx \int\limits_{-\infty}^{+\infty} p(\Delta x)(P_N(x) - \Delta x P_N'(x) + \frac{1}{2}\Delta x^2 P_N''(x) + \cdots)\mathrm{d}\Delta x$$

Taking the derivatives of $P_N(x)$ outside the integral we obtain

$$P_{N+1}(x) \approx$$
$$\approx P_N(x)\int p(\Delta x)\mathrm{d}\Delta x - P_N'(x)\int \Delta x p(\Delta x)\mathrm{d}\Delta x + \frac{1}{2}P_N''(x)\int \Delta x^2 p(\Delta x)\mathrm{d}\Delta x + \cdots$$

$$\Downarrow$$

$$P_{N+1}(x) \approx P_N(x) - P_N'(x)\langle \Delta x\rangle + \frac{1}{2}P_N''(x))\langle(\Delta x)^2\rangle + \cdots \tag{14}$$

We now introduce a time as a variable by stating that the steps are taken after every time interval $\Delta t = const$. Thus, from the above equation we obtain:

$$\frac{P_{N+1} - P_N}{\Delta t} + \frac{\langle \Delta x\rangle}{\Delta t}P_N'(x) = \frac{1}{2}\frac{\langle \Delta x^2\rangle}{\Delta t}P_N''(x) \tag{15}$$

For small Δt we obtain approximately

$$\frac{\partial P(x,t)}{\partial t} + v\frac{\partial P(x,t)}{\partial x} = D\frac{\partial^2 P(x,t)}{\partial x^2} \tag{16}$$
$$v = \frac{\langle \Delta x\rangle}{\Delta t} \tag{17}$$
$$D = \frac{1}{2}\frac{\langle \Delta x^2\rangle}{2} \tag{18}$$
$$t = N\Delta t \tag{19}$$

Note, than when $\langle \Delta x\rangle = 0$ we recover the diffusion equation! Otherwise, the equation describes diffusion in a medium flowing with a drift velocity v. This is called an *advection-diffusion* problem, advection referring to the flow term. Thus, we see a clear relation between macroscopic phenomenon of diffusion and the underlying microscopic phenomenon of the random walk. This connection will be studied in detail in future lectures.

Variations on the problem

Many variations on the random walk problem will be discussed in the class, each of which can invalidate the CLT and produce 'anomalous diffusion' with different scaling behavior, not described by the simple diffusion equation in the continuum limit. Some examples are listed below.

- *Non-identical but independent steps.* The CLT can still be recovered with some minor modifications, but anomalous cases arise if the time-averaged variance goes to 0 or ∞.
- *Fat tails.* $\langle \Delta x^m\rangle = \infty$ where $m > 2$ (variance and mean are still finite). CLT can still be recovered, but the convergence is slower and quite different.
- *Levy flights.* $\langle \Delta x^2\rangle = \infty$ The distibution tends to a non-Gaussian distrbiution (Levy stable law), whose width grows faster than $\sqrt{N}$.
- *Correlated steps.* $\langle \Delta x_i \Delta x_j\rangle \neq 0$ for $i \neq j$.

- If the correlations are 'weak', CLT is recovered. Example: Persistent random walk. $\langle \Delta x_i \Delta x_{i+m} \rangle \neq 0$ only for $m > 1, m \leq -1$. Same Gaussian distribution is recovered, with effectively a rescaled step size, called the persistence length.
- If correlations are 'strong', CLT breaks down. Example: self avoiding walk, a very difficult problem.

- *Interactions*. Different random walkers may interact with each other (e.g. attracting, repelling, reacting to die or give birth to new walkers,...) and/or with a heterogeneous environment. This leads to a wide variety of nonlinear diffusion problems, some of which will be described in the last part of the course.

In addition to considering the position distributions of random walks, we will also consider more subtle properties like the time to reach some region for the first time (*first passage*), the fraction of time spent in some region, or the eventual number of returns to some region. For all of these questions, both discrete and continuum methods will typically be used, and the connection between them elucidated.