Lecture 14: Non-identically Distributed Steps

Ryan J. Larsen

Division of Engineering and Applied Sciences, Harvard University

April 1, 2003

We have already discussed two things that can cause the Central Limit Theorem (CLT) to break down; steps with infinite variance, and correlations between steps. Today we introduce a third possibility: non-identically distributed steps.

1 The Continuum Limit

Consider, as before, the sum, X_N, of random steps

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

where the steps, Δx_n, are independent but not identically distributed.

We will begin by (formally) taking the continuum limit, assuming it exists, taking

$$< \Delta x_n >= 0 \tag{2}$$

$$Var\left(\Delta x_n\right) = {\sigma_n}^2. \tag{3}$$

As before, we first write Bachelier's equation

$$P_{N+1}\left(x\right) = \int P_N\left(x-y\right) * p_{N+1}\left(y\right) dy. \tag{4}$$

We assume that the entire distribution has a length scale much larger than than the size of the individual steps, expand the distribution about x, and divide by the time step, τ_N, to obtain

$$\frac{P_{N+1} - P_N}{\tau_N} = \frac{{\sigma_N}^2}{2\tau_N} {P_N}''\left(x\right). \tag{5}$$

We then replace P_N with the slowly varying density function, ρ, such that $\rho\left(x,t\right) = P_N\left(x\right)$ as $N \to \infty$ where $t = \sum_{n=1}^{N} \tau_N$.

We thus obtain

$$\frac{\partial \rho}{\partial t} = D(t) \frac{\partial^2 \rho}{\partial x^2} \tag{6}$$

where $D(t) = \frac{\sigma_N{}^2}{2\tau_N}$.

Notice that this is diffusion equation where the diffusion coefficient varies with time. From looking at this expression, it appears that all combinations of σ_N and τ_N which give the same $D(t_N)$ are equivalent, but we will see later that this is not the case.

We can find the exact solution to the Green function, (where $\rho(x,0) = \delta(x)$) by introducing

$$u = \int_0^t D(t')\, dt' \tag{7}$$

which converts the diffusion equation to

$$\frac{\partial \rho}{\partial u} = \frac{\partial^2 \rho}{\partial x^2} \tag{8}$$

which we can solve to obtain

$$\rho(x,t) = \frac{\exp \frac{-x^2}{4u(t)}}{\sqrt{4\pi u(t)}} \tag{9}$$

This suggests replacing X_N with the scaled variable

$$Z_N = \frac{X_N}{\sqrt{2u(t)}} \tag{10}$$

which can change with time. Note that in the case of constant diffusion over time, we obtain $< X_N{}^2 > \sim 2Dt$ as before.

The diffusion equation will have higher-order terms if $p(x)$ has higher-order cumulants. Assuming all higher order cumulants exist we define, as before, $c_{n,l}$ to be the l^{th} cumulant of the PDF associated with Δx_n. Since cumulants are additive, we have the cumulants for X_N given by $C_{N,l} = \sum_{n=1}^{N} c_{n,l}$. Therefore, we have the distribution scaling as

$$Z_N = \frac{X_N}{\sqrt{C_{N,2}}} \tag{11}$$

and the higher cumulants scaling as

$$\lambda_{N,l} = \frac{C_{N,l}}{C_{N,2}{}^{l/2}}. \tag{12}$$

where $\lambda_{N,l}$ for $l > 2$ must go to zero for a standard normal distribution to be present.

As before, if $p(x)$ is the same for all steps, then $c_{n,l} = c_l$ implies $C_{N,l} = Nc_l$, which leads to the scalings

$$Z_N = \frac{X_N}{\sigma\sqrt{N}} \tag{13}$$

and

$$\lambda_{N,l} = \frac{Nc_l}{\left(\sigma\sqrt{N}\right)^l} \to 0 \ for \ l > 2. \tag{14}$$

2 CLT for Non-identical Variables

Imagine there is an average variance $\bar{\sigma}^2 = \frac{1}{N}\sum_{n=1}^{N} {\sigma_n}^2 \neq 0, \infty$. In this case there is another Berry-Esseen which states that if $< \Delta x_n >= 0$, $< {\Delta x_n}^2 >= {\sigma_n}^2$, and $< {\Delta x_n}^3 >= r_n < \infty$, then we have a standard normal distribution that scales according to

$$Z_N = \frac{X_N}{\bar{\sigma}\sqrt{N}} \tag{15}$$

Furthermore, if $\Phi_N(z)$ is the CDF of Z_N we have

$$|\Phi_N(z) - \Phi(z)| \leq \frac{6\bar{r}}{\bar{\sigma}^{3/2}\sqrt{N}} \tag{16}$$

where $\bar{r} = \frac{1}{N}\sum_{n=1}^{N} r_n$. This is true even in the tails of the distribution. Notice how how this is the same as the Berry-Esseen Theorem introduced earlier if we replace σ_n by σ, r_n by r, and the 6 by 3.

3 An Interesting class of examples: $\Delta X_n = a_n \epsilon_n$

Consider a random walk given by $\Delta X_n = a_n \epsilon_n$ where the the ϵ_n are iid steps whose amplitudes are modified by a_n. We will take $< \epsilon_n >= 0$ and $< \epsilon_n^2 >= 1$.

An example of this type of random walk would be the diffusion of an impurity atom in a crystal lattice. In this case the diffusion constant, D, is proportional to $exp\frac{-E}{kT}$ where E is the energy barrier to diffusion and kT is, as usual, the product of the Boltzman constant and temperature. If the temperature varies with time, then the diffusion constant will also vary with time.

Another example may be found in financial time series. In this case a_n would represent human factors that effect the market, such as political and economic events. The time τ_n between steps is a constant set by transaction costs and the liquidity of the market. J.P. Bouchard has investigated the case where the a_n are random variables with long-range correlations.

4 Example 1: Power-law decay/growth

Consider the case were

$$a_n = \frac{1}{n^\alpha}. \tag{17}$$

When $\alpha < 1/2$ the CLT holds and the distribution is given by the gaussian.

For these random walks the second cumunlant is given by

$$C_{N,2} = \sum_{n=1}^{N} a_n^2 = \sum_{n=1}^{N} \frac{1}{n^{2\alpha}} \tag{18}$$

$$C_{N,2} = \begin{cases} \zeta(2\alpha) & \text{when } \alpha > 1/2, \\ logN & \text{when } \alpha = 1/2, \\ N^{-2\alpha+1} & \text{when } \alpha < 1/2. \end{cases} \tag{19}$$

As usual, $\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}$.

An example of $\alpha = 3$ occurs in sigle molecule spectroscopy and can be found in Barkai and Silbey, Chem. Phys. Lett., vol. 310, 287, (1999).

5 Example 2: Exponential decay/growth

Consider the case where $a_n = a^n$. An example of this case would be inelastic diffusion. For example, consider a ball bouncing on a rough surface where each bounce is inelastic with a coefficient of restitution of r. Although we may expect the shape of the PDFs for each bounce to be the same, we expect the length of each step to be reduced by a factor of $\sqrt{r}$. A common feature of these random walks is that the PDF of all the steps $\sum \Delta X_n$ is sensitive to the shape of the PDF associated with the ϵ_n. This is a breakdown of the CLT which predicts that the total PDF will be a Gaussian regardless of the PDFs associated with each step. We will consider two types of ϵ_n distributions: standard normal and bernoulli walks.

5.1 Standard normal

In this case the typical length scale is given by

$$\sigma_N^2 = \sum_{n=1}^{N} a_n^2 = \sum_{n=1}^{N} a^{2n} = \frac{a^2\left(1 - a^{2N}\right)}{1 - a^2} \tag{20}$$

which gives us

In all of these cases the shape of the distribution is a Gaussian. However, the shapes will be different if the PDFs of the ϵ_n steps are not Gaussian.

5.2 Bernoulli Walk

Consider the case where $\epsilon_n = \pm 1$, each with probability of $1/2$. This case was studied in the 1930s by Erdos and by Kersner and Winter.

For this lecture we will only describe the results quantitativly; the analytic details will be in the next lecture.

Consider the case where $a = 1/3$. After the the first step of length 1, imaginge that the rest of the steps are in a single direction. Thus the total lenght the walker can move after the first step is $\sum_{n=1}^{\infty} = \frac{1}{3}\frac{1}{1-1/3} = \frac{1}{2}$. So the final position of the walker must lie between -1.5 and -0.5 or between 0.5 and 1.5. The walker can never occupy the middle third (-0.5 to 0.5) of the total range (-1.5 to 1.5). This pattern continues for all subsequent steps leading to a fractal PDF with all walks separated by zones where there is no probability. The PDF is uniformly distributed on the *fractal* Cantor middle-thirds set!

From this argument, it is clear that for all $a < 1/2$, the final PDF is uniform on a fractal set because after any step, the walker cannot return to the same position (thus excluding some region from the support of the PDF).

Consider the borderline case when $a = 1/2$. Now the sum of all remaing steps after the first step is 1, so is is possible for the walker to barely make it back to its beginning location. This happens, i.e. the walker can just barely return, recursively at each step. The resulting PDF is a *continous* uniform distribution in the range -1 to 1.

We will prove this directly using Fourier transforms in the next lecture. Moreover, by exploiting relations betweent he PDF for different values of a, we will show that when $a = 1/\sqrt{2}$, the distribution is piecewise linear (one derivative); when $a = 1/\sqrt[3]{2}$ the distribution is is piecewise parabolic (two derivatives); and so on until $a \to 1$ which returns the Gaussian which contains all derivatives continuous. Of course, the last result is just the Central Limit Theorem in action, once sufficient self-averaging kicks in.

These values for a are special cases where exact solutions are possible. At intermediate values of a one often recovers complicated singular distributions. One interesting case is the Golden Mean, $a = (\sqrt{5}-1)/2$, which satisfies, $a^2 + a = 1$. This means that a walker can return to a site in exactly two steps. This property of recombinance in the 'space-time tree' leads to a very complicated singular (non-differentiable) distribution.