PHY256 Notes on the Central limit theorem, Random Walks and Diffusion

A. C. Quillen

March 24, 2021

Contents

1 Central Limit Theorem

1.1 Probability density functions and histograms

We consider here a random continuous variable x that is described with a probability density function $p(x)$. As the probability should sum to 1 $\int p(x)dx = 1$.

Numerically we can look at a probability distribution by generating a set of the random variables x_i. We can then count the number of x_i values in different bins where each bin covers a small range in x. This gives us a *histogram* (see Figure 1) or a count of the number of values in our set that lie in each bin.

For example, if we randomly generated a set of x values x_1, x_2, x_3.. using a normal or Gaussian probability distribution for x with a mean of 0, the histogram of values should be centered at x and should look like a Gaussian function (see Figure 1). The larger the number of generated values used to generate the histogram, the more likely it is to resemble the Gaussian. We can normalize the histogram and make it look like a probability distribution by dividing by the total number of data samples and taking into account the

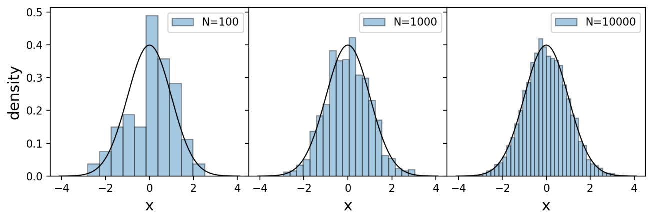

Figure 1: I generated a set of random x values. From these I constructed histograms. The x values are randomly generated so that they should have a normal probability distribution. The normal distribution $\rho(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ is shown with the black lines. The histograms have been normalized so that they integrate to 1. The number of x values generated is labelled in each panel. The more x values generated, the closer the histogram resembles the Gaussian distribution.

bin widths used to create the histogram. Figure 1 shows that with larger numbers of samples, the histogram gets closer to the Gaussian function. This implies that the way the random numbers are generated is consistent with the Gaussian probability distribution. The randomly generated set converges to the probability density function as the number of generated values goes to infinity.

1.2 Mean and variance of a sum of two independent random variables

We start with two independent random variables x, y with normalized probability distribution functions $f(x), g(y)$. Normalized means that the integral of the probability density function sums to 1;

$$\int f(x)dx = 1 \qquad \int g(y)dy = 1$$

The mean of x is also called the expectation value of x,

$$\mu_x = \langle x \rangle = \int dx \; x f(x)$$

and the mean of y

$$\mu_y = \langle g \rangle = \int dy \; y g(y)$$

We can also look at the second moments

$$\langle x^2 \rangle = \int dx \; x^2 f(x)$$

with variance in x of

$$\sigma_x^2 = \langle (x - \mu_x)^2 \rangle = \langle x^2 \rangle - \mu_x^2$$

and likewise for y. Here σ_x is the standard deviation of x and σ_x^2 is its variance.

What is the probability distribution for a random variable z with $z = x + y$?

We first choose x with $f(x)$ and then we must have $y = z - x$ with probability $g(y)$. Summing over all possible x values we find the probability of z is described with

$$p(z) = \int dx \; f(x)g(z - x). \tag{1}$$

We note that this is a convolution!

The distribution of a sum of randomly generated variables should be smoother than the original distributions. This follows as the probability distributions are convolved.

The mean of z is equal to the sum of the mean of x and that of y

$$\begin{aligned}\mu_z &= \int dz \; zp(z) \\ &= \int dz \int dx \; f(x)g(z - x)z\end{aligned}$$

Let $z' = z - x$ so $dz = dz'$ and $z = z' + x$

$$\begin{aligned}\mu_z &= \int dz' \int dx \; f(x)g(z')(z' + x) \\ &= \int dz' \int dx \left[f(x)g(z')z' + f(x)g(z')x\right] \\ &= \mu_y + \mu_x.\end{aligned}$$

Likewise we can show that the variance of z is the sum of that for x and y. Let us make this simpler by setting $\mu_x = \mu_y = \mu_z = 0$ so that $\sigma_x^2 = \langle x^2 \rangle$ and similarly for y. In this case

$$\begin{aligned}\sigma_z^2 &= \int dz \int dx \; f(x)g(z - x)z^2 \\ &= \int dz' \int dx \; f(x)g(z')(z' + x)^2 \\ &= \int dz' \int dx \; \left[f(x)g(z')z'^2 + f(x)g(z')x^2 + f(x)g(z')2xz'\right] \\ &= \langle y^2 \rangle + \langle x^2 \rangle + 2\langle x \rangle \langle y \rangle \\ &= \sigma_x^2 + \sigma_y^2.\end{aligned}$$

In other words

$$\sigma_z = \sqrt{\sigma_x + \sigma_y}. \tag{2}$$

1.3 Consequences of variances adding in quadrature

We list some consequences of equation 2 that variances add in quadrature.

- Errors add in quadrature.
- Integration times in astronomy. With Poisson statistics the standard deviation depends on the number of photons $\sigma = \sqrt{N_p}$. The signal is the number of photons N_p and the error or uncertainty or noise is $\sqrt{N_p}$. The signal to noise is then $S/N = N_p/\sqrt{N_p} = \sqrt{N_p}$.

  The integration time of an astronomical observation t sets the total number of detected photons. If t gives N photons, then twice the integration time gives $2N$ photons. The signal to noise is proportional to $\sqrt{t}$.
- A random walk is a series of steps where each step is drawn from a probability distribution. After N steps, the walker is at a position given by the sum of N randomly variables, all drawn from the same distribution. After N steps the distribution of a bunch of walkers, all starting from the origin, has a standard deviation (that is proportional to) $\sqrt{N}$.
- We can describe diffusion as the continuum limit of many walkers. The total number of steps N taken depends on the duration of the walk t and the time between each step dt with $N = t/dt$. Diffusion of an initially concentrated distribution of walkers gives a number density with width that increases proportional to $t^{1/2}$.

1.4 The central limit theorem

Let $x_1, x_2....$ be independently and identically distributed random variables with mean μ and variance σ^2.

Let

$$z_n = \sum_{i=1}^{n} \sqrt{n}\frac{(x_i - \mu)}{\sigma}$$

The central limit theorem: As $n \to \infty$ the probability distribution of z_n increasingly resembles a normal distribution $N(0,1)$ (a Gaussian with mean 0 and variance 1).

For large n, the sum $X_n = \sum_{i=1}^n x_i$ has distribution $N(n\mu, n\sigma^2)$.

The central limit theorem can be modified to the *multivariate* central limit theorem but with random variables x_i that have different means and variances. Again as n gets larger, the sum $X_n = \sum_{i=1}^n x_i$ has distribution that resembles a Gaussian probability distribution.

The **central limit theorem** implies that sum of random variables will approach a stable distribution that has the same shape as a normal distribution. Additional effort is required to imply that all higher order moments will become similar to those of a normal distribution.

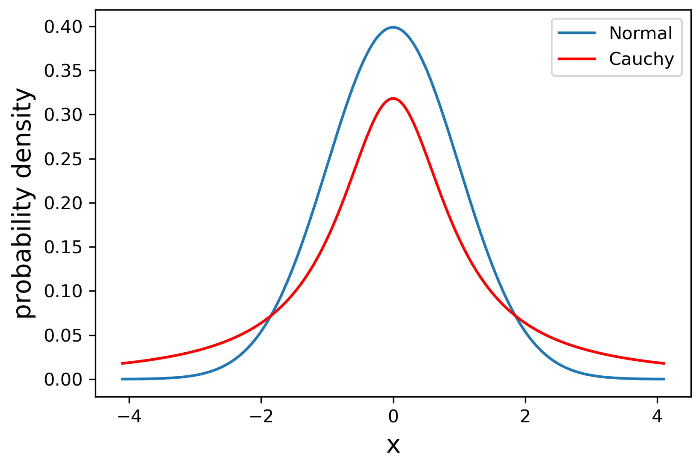

Figure 2: A comparison of the shape of the normal and Cauchy distributions The normal distribution $\rho_N(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. The Cauchy distribution $\rho_C(x) = \frac{1}{\pi}\frac{1}{x^2+1}$

The moments of the distributions of the random variables must exist otherwise the central limit theorem is violated! And when they don't exist (or are infinite) the central limit theorem no longer holds. In this setting we get what is called *anomalous diffusion*, super or/and sub diffusion.

If x, y are random variables described by Gaussian distributions, their sum will also be a Gaussian distribution because the convolution of a Gaussian with a Gaussian is also a Gaussian. The sum of most random variables becomes increasingly closer to a Gaussian. We say that the Gaussian distribution is an *attracting* (or stable) probability distribution.

1.5 Heavy-Tailed distributions

The Cauchy distribution has no moments (they are infinite), but it is a stable (and attracting) distribution. It does have a well defined full width half max (FMHM).

The Cauchy distribution (aka Lorenz distribution)

$$p(x) = \frac{1}{\pi}\frac{1}{x^2+1}$$

is normalized

$$\int_{-\infty}^{\infty} p(x)\, dx = 1$$

so it is a perfectly good probability distribution. However the mean is not well defined

$$\begin{aligned}\int_{-\infty}^{\infty} xp(x)\ dx &= \int_0^\infty \frac{2}{\pi}\frac{x}{x^2+1}dx \\ &= \lim_{x_m\to\infty}\int_0^{x_m}\frac{2}{\pi}\frac{x}{x^2+1}dx \\ &= \lim_{x_m\to\infty}\frac{2}{\pi}\frac{1}{2}\log(x^2+1)\Big]_0^{x_m} \\ &= \lim_{x_m\to\infty}\frac{2}{\pi}\log x_m\end{aligned}$$

So there is no first moment (mean) even though the function is symmetrical about $x = 0$. Likewise the variance

$$\begin{aligned}\int_{-\infty}^{\infty} x^2p(x)\ dx &= \int_0^\infty \frac{2}{\pi}\frac{x^2}{x^2+1}dx \\ &\sim \lim_{x_m\to\infty}\frac{2}{\pi}\int^{x_m} dx \\ &\propto \lim_{x_m\to\infty} x_m\end{aligned}$$

is infinite.

This distribution has tails. But it looks okay if you graph it (see Figure 2)! It has a well defined full width full max.

Consequences of violating the central limit theorem are that tails dominate eventually in a sum of randomly generated variables. The standard deviation of a distribution of random walkers increases without bound.

A random walker obeys something like a Cauchy distribution if every once in a while the walker takes a big step. And every once in a longer time the walker takes an even bigger step. And so-on in a self-similar or power law form. This setting gives you **super-diffusion**. Alternatively you could have your walker every once in a while just stop taking steps for a while. And every once in a longer time stop taking steps for an even longer time. And so on. This gives **sub-diffusion** and is related to a phenomena called intermittency that is sometimes seen in chaotic dynamical systems.

2 Diffusion

Diffusion is the continuum limit of a bunch of random walkers. Instead of describing the location of a single walker with a probability or the positions of a distribution of walkers, we consider the number density of walkers as a function of position and time $\rho(x, t)$ and see how that evolves.

Flux is numbers of walkers going through a surface per unit area on the surface per unit time The flux of walkers depends on the gradient of random walkers. If all walkers start at the same spot, their density distribution spreads out. If there are more walkers in one spot than there are in a neighboring spot, more walkers will move from the denser area than will move from sparser area to the denser area. The flux of walkers depends on the gradient of walkers. .

The rate that the density of walkers changes

$$\frac{\partial \rho}{\partial t}$$

The gradient of walkers

$$\boldsymbol{\nabla} \rho.$$

The flux of walkers we assume depends on the gradient of the density of walkers

$$\mathbf{F} = -A \boldsymbol{\nabla} \rho.$$

The walkers should move from regions of high density to low density, and this gives the minus sign. The coefficient A determine how large a step the walkers take each time step. The change in density depends on the gradient of the flux

$$\frac{\partial \rho}{\partial t} = -\boldsymbol{\nabla} \cdot \mathbf{F}$$

If the coefficient A is independent of position then we can write this as

$$\frac{\partial \rho}{\partial t} = D \nabla^2 \rho \tag{3}$$

where D is a diffusion coefficient that is equal to A. This gives a partial differential equation for the evolution of the density of walkers $\rho(\mathbf{x}, t)$.

Equation 3 is a continuum equation, describing diffusion. It is sometimes called the *heat equation* as heat transfer is often described with the same equation.

Units for diffusion coefficient are L^2/T. The equation is linear in ρ so the units of ρ do not affect the units of the diffusion coefficient.

Viscosity can be considered a diffusion coefficient for velocity.

Turbulence is some times described in terms of a turbulent diffusion coefficient that depends on the sizes and velocities of turbulent eddies.

2.1 The Green's function for the diffusion equation

The diffusion equation

$$\frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial x^2} \tag{4}$$

We can ask what is a solution to this equation that is initially centered at $x = 0$ at $t = 0$? This tells us what happens if we have an initially concentrated amount of die (or randomly walking particles).

Try a normal distribution but with standard deviation $\sigma(t)$.

$$\rho(x,t) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \tag{5}$$

$$\begin{aligned}
\frac{\partial\rho}{\partial t} &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \left[-\frac{\dot\sigma}{\sigma^2} + \frac{1}{\sigma}\frac{2x^2\dot\sigma}{2\sigma^3} \right] \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{\sigma} \left[-\frac{\dot\sigma}{\sigma} + \frac{x^2\dot\sigma}{\sigma^3} \right]
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\rho}{\partial x} &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{x^2}{2\sigma^2}} \left(-\frac{x}{\sigma^2} \right) \\
\frac{\partial^2\rho}{\partial x^2} &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{x^2}{2\sigma^2}} \left[\left(\frac{x}{\sigma^2} \right)^2 - \frac{1}{\sigma^2} \right]
\end{aligned}$$

Inserting partial derivatives back into the diffusion equation (eq. 4) we find

$$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} \frac{1}{\sigma} \left[-\frac{\dot\sigma}{\sigma} + \frac{x^2\dot\sigma}{\sigma^3} \right] = D \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma} e^{-\frac{x^2}{2\sigma^2}} \left[\left(\frac{x}{\sigma^2} \right)^2 - \frac{1}{\sigma^2} \right] \tag{6}$$

Taking the term proportional to x^2,

$$\begin{aligned}
x^2 \frac{\dot\sigma}{\sigma^3} &= D \frac{x^2}{\sigma^4} \\
\dot\sigma\sigma &= D \\
\frac{d}{dt}\left(\frac{\sigma^2}{2} \right) &= D.
\end{aligned}$$

We integrate

$$\begin{aligned}
\frac{d}{dt}\frac{\sigma^2}{2} &= D \\
\sigma^2 &= 2Dt + \text{constant}
\end{aligned}$$

Happily when we take the constant term in equation 6

$$\begin{aligned}
\frac{\dot\sigma}{\sigma} &= \frac{D}{\sigma^2} \\
\dot\sigma\sigma &= D
\end{aligned}$$

we get something consistent with the x^2 term. Altogether our solution (sticking this form for $\sigma(t)$ into equation 5) is

$$\rho(x,t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}}.$$

This describes how fast an initially concentrated (at $x = 0$) set of particles spreads out. The solution is often called a Green's function, which is a solution at later times for an initial distribution that is a delta function. This is the Green's function for the heat (or diffusion) equation!

More complicated initial conditions can be considered with sums (or integrals) of delta function solutions. For more complicated initial conditions we could sum (or integrate) a bunch of Green's functions and get the general solution at later times.

Notice that the standard deviation

$$\sigma(t) = \sqrt{2Dt}. \tag{7}$$

This behavior resembles that of a random walk. Consider a random walk with steps that are taken separated in time by dt and each one has size with standard deviation $\sigma_1 = L$. The number of steps in time t would be $N = t/dt$. After time t the standard deviation of a distribution of walkers begun at a single location would be

$$\sigma(t) = \sigma_1 \sqrt{N} = \sigma_1 \sqrt{\frac{t}{dt}}$$

We compare this to equation 7 and find that the effective diffusion coefficient is

$$D = \frac{\sigma_1^2}{2dt} = \frac{L^2}{2dt}.$$

2.2 Connection between the central limit theorem, random walks, and diffusion

Using variables that have moments, the central limit theorem states that the distribution of their sum will resemble a Gaussian distribution and the standard deviation of the Gaussian will be proportional to $\sqrt{N}$ where N is the number of variables used in the sum. A random walk is the sum of random variables. A ensemble of walkers that starts all in the same spot expands with standard deviation proportional to $\sqrt{N}$ where N is the number of steps. This resembles the Green's function for the heat or diffusion equation. A distribution of walkers will obey the heat/diffusion equation.

What if at each step the walkers move with step size that is given by a heavy tailed distribution? In this case the central limit theorem is violated and the dispersion increases forever. Random walks that are like this are called Levy flights. The associated diffusion obeys different scaling and is called *anomalous diffusion*.