

## 1.1: Statistical Properties of Random Walks

### One-Dimensional Random Walk

Consider the mechanical system depicted in Fig. 1.1.1, a version of which is often sold in novelty shops. A ball is released from the top, which cascades consecutively through  $N$  levels. The details of each ball's motion are governed by Newton's laws of motion. However, to predict where any given ball will end up in the bottom row is difficult, because the ball's trajectory depends sensitively on its initial conditions, and may even be influenced by random vibrations of the entire apparatus. We therefore abandon all hope of integrating the equations of motion and treat the system statistically. That is, we assume, at each level, that the ball moves to the right with probability  $p$  and to the left with probability  $q = 1 - p$ . If there is no bias in the system, then  $p = q = \frac{1}{2}$ . The position  $X_N$  after  $N$  steps may be written

$$X = \sum_{j=1}^N \sigma_j, \quad (1.1.1)$$

where  $\sigma_j = +1$  if the ball moves to the right at level  $j$ , and  $\sigma_j = -1$  if the ball moves to the left at level  $j$ . At each level, the probability for these two outcomes is given by

$$P_\sigma = p \delta_{\sigma,+1} + q \delta_{\sigma,-1} = \begin{cases} p & \text{if } \sigma = +1 \\ q & \text{if } \sigma = -1 \end{cases}. \quad (1.1.2)$$

This is a normalized discrete probability distribution of the type discussed in section 4 below. The multivariate distribution for all the steps is then

$$P(\sigma_1, \dots, \sigma_N) = \prod_{j=1}^N P(\sigma_j). \quad (1.1.3)$$

Our system is equivalent to a one-dimensional *random walk*. Imagine an inebriated pedestrian on a sidewalk taking steps to the right and left at random. After  $N$  steps, the pedestrian's location is  $X$ .

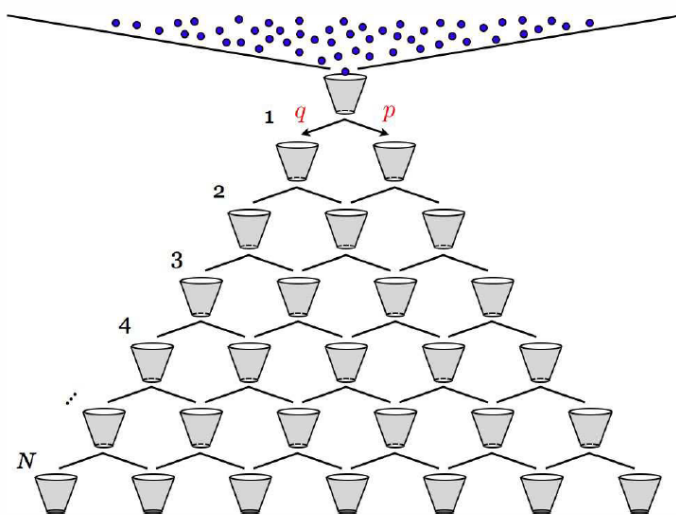


Figure 1.1.1: The falling ball system, which mimics a one-dimensional random walk.

Now let's compute the average of  $X$ :

$$\langle X \rangle = \left\langle \sum_{j=1}^N \sigma_j \right\rangle = N \langle \sigma \rangle = N \sum_{\sigma=\pm 1} \sigma P(\sigma) = N(p - q) = N(2p - 1). \quad (1.1.4)$$

This could be identified as an *equation of state* for our system, as it relates a measurable quantity  $X$  to the number of steps  $N$  and the local bias  $p$ . Next, let's compute the average of  $X^2$ :

$$\langle X^2 \rangle = \sum_{j=1}^N \sum_{j'=1}^N \langle \sigma_j \sigma_{j'} \rangle = N^2 (p-q)^2 + 4Npq. \quad (1.1.5)$$

Here we have used

$$\langle \sigma_j \sigma_{j'} \rangle = \delta_{jj'} + (1 - \delta_{jj'}) (p-q)^2 = \begin{cases} 1 & \text{if } j = j' \\ (p-q)^2 & \text{if } j \neq j' \end{cases}. \quad (1.1.6)$$

Note that  $\langle X^2 \rangle \geq \langle X \rangle^2$ , which must be so because

$$\text{Var}(X) = \langle (\Delta X)^2 \rangle \equiv \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2. \quad (1.1.7)$$

This is called the *variance* of  $X$ . We have  $\text{Var}(X) = 4Npq$ . The *root mean square* deviation,  $\Delta X_{rms}$ , is the square root of the variance:  $\Delta X_{rms} = \sqrt{\text{Var}(X)}$ . Note that the mean value of  $X$  is linearly proportional to  $N^1$ , but the RMS fluctuations  $\Delta X_{rms}$  are proportional to  $N^{1/2}$ . In the limit  $N \rightarrow \infty$  then, the ratio  $\Delta X_{rms} / \langle X \rangle$  vanishes as  $N^{-1/2}$ . This is a consequence of the central limit theorem (see §4.2 below), and we shall meet up with it again on several occasions. We can do even better. We can find the complete probability distribution for  $X$ . It is given by

$$P_{N,X} = \frac{N!}{N_{\text{right}}! N_{\text{left}}!} p^{N_{\text{right}}} q^{N_{\text{left}}},$$

where  $N_{\text{right}}$  are the numbers of steps taken to the right/left, with  $N = N_{\text{right}} + N_{\text{left}}$ , and  $X = N_{\text{right}} - N_{\text{left}}$ . There are many independent ways to take  $N_{\text{right}}$  steps to the right. For example, our first  $N_{\text{right}}$  steps could all be to the right, and the remaining  $N_{\text{left}} = N - N_{\text{right}}$  steps would then all be to the left. Or our final  $N_{\text{right}}$  steps could all be to the right. For each of these independent possibilities, the probability is  $p^{N_{\text{right}}} q^{N_{\text{left}}}$ . How many possibilities are there? Elementary combinatorics tells us this number is

$$\frac{N!}{N_{\text{right}}! N_{\text{left}}!}.$$

Note that  $N_{\text{right}} = \frac{N+X}{2}$ , so we can replace  $N_{\text{right}} = \frac{N+X}{2}$ . Thus,

$$P_{N,X} = \frac{N!}{\left(\frac{N+X}{2}\right)! \left(\frac{N-X}{2}\right)!} p^{(N+X)/2} q^{(N-X)/2}. \quad (1.1.8)$$

## Thermodynamic Limit

Consider the limit  $N \rightarrow \infty$  but with  $x \equiv X/N$  finite. This is analogous to what is called the *thermodynamic limit* in statistical mechanics. Since  $N$  is large,  $x$  may be considered a continuous variable. We evaluate  $\ln P_{N,X}$  using Stirling's asymptotic expansion

$$\ln N! \simeq N \ln N - N + \mathcal{O}(\ln N). \quad (1.1.9)$$

We then have

$$\begin{aligned} \ln P_{N,X} &\simeq N \ln N - N - \frac{1}{2} N(1+x) \ln \left[ \frac{1}{2} N(1+x) \right] + \frac{1}{2} N(1+x) \\ &\quad - \frac{1}{2} N(1-x) \ln \left[ \frac{1}{2} N(1-x) \right] + \frac{1}{2} N(1-x) \ln p + \frac{1}{2} N(1-x) \ln q \\ &= -N \left[ \left( \frac{1+x}{2} \right) \ln \left( \frac{1+x}{2} \right) + \left( \frac{1-x}{2} \right) \ln \left( \frac{1-x}{2} \right) \right] + N \left[ \left( \frac{1+x}{2} \right) \ln p + \left( \frac{1-x}{2} \right) \ln q \right]. \end{aligned}$$

Notice that the terms proportional to  $N \ln N$  have all cancelled, leaving us with a quantity which is linear in  $N$ . We may therefore write  $\ln P_{N,X} = -N f(x) + \mathcal{O}(\ln N)$ , where

$$f(x) = \left[ \left( \frac{1+x}{2} \right) \ln \left( \frac{1+x}{2} \right) + \left( \frac{1-x}{2} \right) \ln \left( \frac{1-x}{2} \right) \right] - \left[ \left( \frac{1+x}{2} \right) \ln p + \left( \frac{1-x}{2} \right) \ln q \right]. \quad (1.1.10)$$

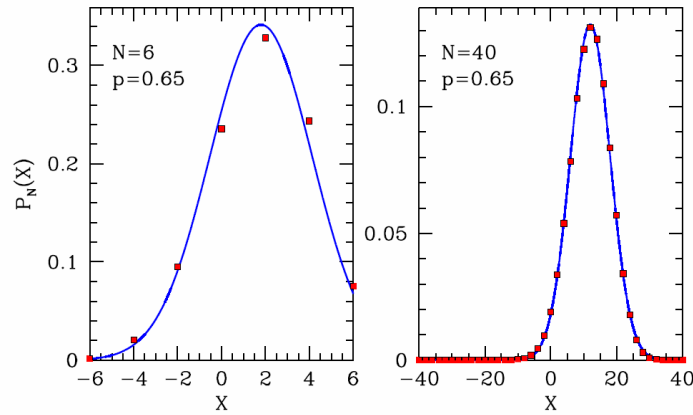


Figure 1.1.2: Comparison of exact distribution of Equation 1.1.8 (red squares) with the Gaussian distribution of Equation 1.1.16 (blue line).

We have just shown that in the large  $N$  limit we may write

$$P_{N,X} = \mathcal{C} e^{-Nf(X/N)}, \quad (1.1.11)$$

where  $\mathcal{C}$  is a normalization constant<sup>2</sup>. Since  $N$  is by assumption large, the function  $P_{N,X}$  is dominated by the minimum (or minima) of  $f(x)$ , where the probability is maximized. To find the minimum of  $f(x)$ , we set  $f'(x) = 0$ , where

$$f'(x) = \frac{1}{2} \ln \left( \frac{q}{p} \cdot \frac{1+x}{1-x} \right). \quad (1.1.12)$$

Setting  $f'(x) = 0$ , we obtain

$$\frac{1+x}{1-x} = \frac{p}{q} \quad \Rightarrow \quad \bar{x} = p - q. \quad (1.1.13)$$

We also have

$$f''(x) = \frac{1}{1-x^2}, \quad (1.1.14)$$

so invoking Taylor's theorem,

$$f(x) = f(\bar{x}) + \frac{1}{2} f''(\bar{x}) (x - \bar{x})^2 + \dots \quad (1.1.15)$$

Putting it all together, we have

$$P_{N,X} \approx \mathcal{C} \exp \left[ -\frac{N(x - \bar{x})^2}{8pq} \right] = \mathcal{C} \exp \left[ -\frac{(X - \bar{X})^2}{8Npq} \right], \quad (1.1.16)$$

where  $\bar{X} = \langle X \rangle = N(p - q) = N\bar{x}$ . The constant  $\mathcal{C}$  is determined by the normalization condition,

$$\sum_{X=-\infty}^{\infty} P_{N,X} \approx \frac{1}{2} \int_{-\infty}^{\infty} dX \mathcal{C} \exp \left[ -\frac{(X - \bar{X})^2}{8Npq} \right] = \sqrt{2\pi Npq} \mathcal{C}, \quad (1.1.17)$$

and thus  $\mathcal{C} = 1/\sqrt{2\pi Npq}$ . Why don't we go beyond second order in the Taylor expansion of  $f(x)$ ? We will find out in §4.2 below.

## Entropy and energy

The function  $f(x)$  can be written as a sum of two contributions,  $f(x) = e(x) - s(x)$ , where

$$\begin{aligned} s(x) &= -\left(\frac{1+x}{2}\right) \ln \left(\frac{1+x}{2}\right) - \left(\frac{1-x}{2}\right) \ln \left(\frac{1-x}{2}\right) \\ e(x) &= -\frac{1}{2} \ln(pq) - \frac{1}{2} x \ln(p/q). \end{aligned}$$

The function  $S(N, x) \equiv N s(x)$  is analogous to the *statistical entropy* of our system<sup>3</sup>. We have

$$S(N, x) = N s(x) = \ln \left( \frac{N!}{N_1! N_2! \dots N_R!} \right) = \ln \left( \frac{N!}{N! 2^{N(1+x)}} \right) .$$

Thus, the *statistical entropy* is the *logarithm of the number of ways the system can be configured so as to yield the same value of  $X$*  (at fixed  $N$ ). The second contribution to  $f(x)$  is the energy term. We write

$$E(N, x) = N e(x) = -\frac{1}{2} N \ln(pq) - \frac{1}{2} N x \ln(p/q) . \quad (1.1.18)$$

The energy term biases the probability  $P_{N, X} = \exp(S - E)$  so that *low energy configurations are more probable than high energy configurations*. For our system, we see that when  $p < q$  ( $p < \frac{1}{2}$ ), the energy is minimized by taking  $x$  as small as possible (meaning as negative as possible). The smallest possible allowed value of  $x = X/N$  is  $x = -1$ . Conversely, when  $p > q$  ( $p > \frac{1}{2}$ ), the energy is minimized by taking  $x$  as large as possible, which means  $x = 1$ . The average value of  $x$ , as we have computed explicitly, is  $\bar{x} = p - q = 2p - 1$ , which falls somewhere in between these two extremes.

In actual thermodynamic systems, entropy and energy are not dimensionless. What we have called  $S$  here is really  $S/k_B$ , which is the entropy in units of Boltzmann's constant. And what we have called  $E$  here is really  $E/k_B T$ , which is energy in units of Boltzmann's constant times temperature.

---

This page titled [1.1: Statistical Properties of Random Walks](#) is shared under a [CC BY-NC-SA](#) license and was authored, remixed, and/or curated by [Daniel Arovas](#).