Random walk and binomial distribution

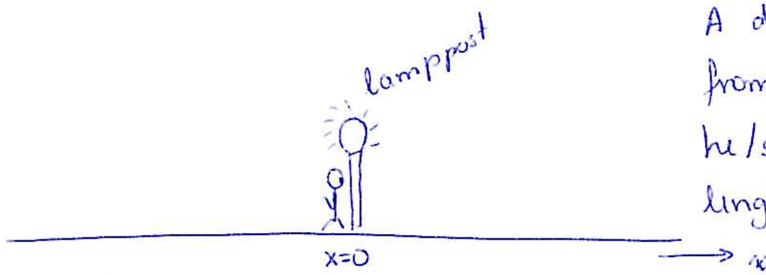
[Fundamentals of Statistical and Thermal Physics
F. Reif (Ch. 1)]

Any process involves noise, randomness. Absolute precision does not exist.

How can we take this into account when solving problems with the computer?

We will start discussing the random walk problem

1D



A drunk person starts out from a lamppost. Each step he/she takes is of equal length ℓ. The direction of each step is completely independent of the previous one.

Each time he/she takes a step, the probability of its being to the right or to the left is respectively

p $\quad$ $q = 1 - p$

In the simplest case, where he/she would throw a coin, $p = q$.

But if there is some sort of drift, such as an inclined street $p \neq q$

After N steps, what is the probability for he/she to be at

N (large number)

$$\boxed{x = m\ell}$$ where m is an integer

$-N \leq m \leq N$

Random walk = sum of random variables

There are several applications to physical and biological sciences, even economics, of the random walk problem, especially when extended to 2D and 3D

- magnetization of an ensemble of spins - 1/2 (with no preferable $\vec{B}$)

- diffusion [endpoint of the random walk has a probability distribution that obeys a simple continuum law, the diffusion equation]
  - molecule in a gas collides with others. How far will it be after N collisions?
  - perfume molecule
  - photon diffusion in the Sun
  - energy transfer
  - motion of microorganisms on surfaces

- shape of polymers → long molecules (DNA, RNA, proteins, many plastics) made up of small units called monomers
  Temperature introduces fluctuations in the angle between adjacent monomers
  But they cannot intersect each other
  [self-avoiding random walks]

- stock market

→ Random walks: scale invariance
  form jagged, fractal patterns

(FIGURE)

In $1D$

what is the probability $P_N(m)$ of finding the particle at position

$$\overline{\underline{x = m\ell}} \qquad\qquad N \leq m \leq N$$

after N steps?

n_r = number of steps to the right

n_ℓ = number of steps to the left

$$\left\{ \begin{array}{l} \overline{\underline{m = n_r - n_\ell}} \\ \overline{\underline{N = n_r + n_\ell}} \end{array} \right\} \quad m = n_r - (N - n_r) = \overline{\underline{2n_r - N}}$$

$$\begin{array}{ccc} \downarrow & & \downarrow \\ \text{even} & \Longleftarrow & \text{even} \\ \text{odd} & \Longleftarrow & \text{odd} \end{array}$$

Steps are independent on the past

p = prob. that the step is to the right

$q = 1 - p$ = " " " left

The probability for a specific sequence of n_r steps to the right and n_ℓ steps to the left is

$$\underbrace{p \cdot p \cdots p}_{n_r} \; \underbrace{q\, q \cdots q}_{n_\ell} = p^{n_r} q^{n_\ell}$$

But there are many different ways of taking n_r steps to the right and n_ℓ steps to the left

$$\frac{N!}{n_r!\, n_\ell!} \quad \text{distinct possibilities}$$

The probability $W_N(n_r)$ of taking n_r steps to the right and n_ℓ steps to the left after N total steps is then

$$\boxed{W_N(n_r) = \frac{N!}{n_r!\, n_\ell!}\, p^{n_r} q^{n_\ell}} \leftarrow \underline{\text{Binomial distribution}}$$

⟹) <u>NOTE</u>:

i) Suppose we take 3 steps: 2 to the right and 1 to the left

$$\left.\begin{matrix} RRL \\ RLR \\ LRR \end{matrix}\right\} 3 \longrightarrow \frac{3!}{2!\,1!} = 3$$

ii) Suppose we take 4 steps: 2 to the right and 2 to the left

$$\left.\begin{matrix} RRLL \\ RLRL \\ RLLR \\ LRRL \\ LRLR \\ LLRR \end{matrix}\right\} 6 \longrightarrow \frac{4!}{2!\,2!} = \frac{4 \cdot 3}{2} = 6$$

⟹) <u>NOTE</u>:

The binomial distribution has this name because $\frac{N!}{n_r!\, n_\ell!}$ is the term appearing in $(p+q)^N$

$$(p+q)^N = \sum_{n=0}^{N} \frac{N!}{n!\,(N-n)!}\, p^n q^{N-n}$$

$$P_N(m) = W_N(n_r)$$

$$\left\{\begin{array}{l} n_r + n_l = N \\ n_r - n_l = m \end{array}\right\} \quad n_r = \frac{N+m}{2}, \quad n_l = \frac{N-m}{2}$$

$$P_N(m) = \frac{N!}{n_r!\,(N-n_r)!}\, p^{n_r} (1-p)^{N-n_r} = \frac{N!}{\left(\frac{N+m}{2}\right)!\left(\frac{N-m}{2}\right)!}\, p^{\frac{N+m}{2}} (1-p)^{\frac{N-m}{2}}$$

Exercise 1:

Suppose $p = q = 1/2$, $N = 3$

what are the probabilities for $n_r = 0, 1, 2, 3$?

(mathematica)

$$\left\{\begin{array}{l} W(0) = (1/2)^3 = 1/8 \\ W(1) = 3\,(1/2)^3 = 3/8 \\ W(2) = 3\,(1/2)^3 = 3/8 \\ W(3) = (1/2)^3 = 1/8 \end{array}\right.$$

(total sum = 1)

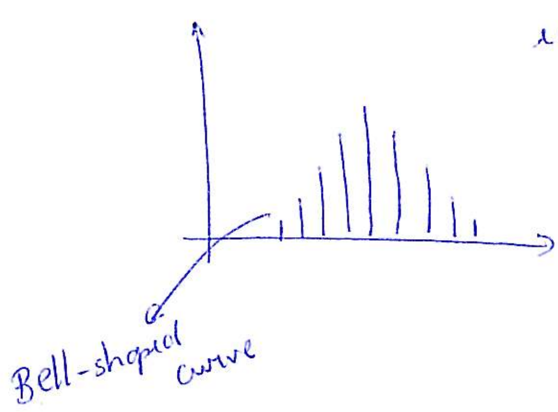
Exercise 2:

Suppose $p = q = 1/2$, $N = 20$

make a plot illustrating all cases from $n_r = 0$ to $n_r = 20$.

Show both $P_N(m)$ vs m and $W_N(n_r)$ vs n_r

It should look like



each vertical line indicates the probability for one value of n_r

Conclusion: after N random steps, the probability of the particle being a distance of N steps from the origin is very small, while the probability for being close to the origin is large

→) Normalization condition

is $\sum_{n_r=0}^{N} W_N(n_r) \overset{?}{=} 1$?

$$\sum_{n_r=0}^{N} \frac{N!}{n_r!\,(N-n_r)!}\, p^{n_r}\, q^{N-n_r} = (p+q)^N \underset{q=1-p}{=} 1^N = 1 \quad ✓$$

→) Mean number $\overline{n_r}$ of steps to the right

$$\overline{n_r} = \sum_{n_r=0}^{N} n_r\, W_N(n_r) = \sum_{n_r=0}^{N} \frac{N!}{n_r!\,(N-n_r)!}\, n_r\, p^{n_r}\, q^{N-n_r}$$

(!) let us assume that p and q are two arbitrary, independent parameters

$$n_r\, p^{n_r} = p \frac{\partial}{\partial p}\left(p^{n_r}\right)$$

(!)

$$\overline{n_r} = \sum \frac{N!}{n_r!\,(N-n_r)!}\, p \frac{\partial}{\partial p}\left(p^{n_r}\right) q^{N-n_r} = p \frac{\partial}{\partial p}\left(\sum_{n_r=0}^{N} \frac{N!}{n_r!\,(N-n_r)!}\, p^{n_r}\, q^{N-n_r}\right)$$

$$= p \frac{\partial}{\partial p}(p+q)^N = p\,N\,(p+q)^{N-1}$$

(!) now we go back to our specific problem where $q = 1-p$

$$\boxed{\overline{n_r} = Np} \qquad \overline{n_\ell} = Nq$$

$$\overline{n_r} + \overline{n_\ell} = N(p+q) = N$$

$$m = n_r - n_\ell \Rightarrow \overline{m} = \overline{n_r} - \overline{n_\ell} = N(p-q) \qquad \boxed{\overline{m} = N(p-q)}$$

if $p = q \Rightarrow \overline{m} = 0$

there is complete symmetry between right and left

Another way to arrive at

$$\overline{n_R = Np}$$

$$\overline{n_R} = \sum_{n_R=0}^{N} n_R \frac{N!}{n_R!\,(N-n_R)!} p^{n_R} q^{N-n_R}$$

↳ this is zero when $n_R = 0$

so

$$= \sum_{n_R=1}^{N} n_R \frac{N!}{n_R!\,(N-n_R)!} p^{n_R} q^{N-n_R} = \sum_{n_R=1}^{N} \frac{N!}{(n_R-1)!\,(N-n_R)!} p^{n_R} q^{N-n_R}$$

using $\overline{K = n_R - 1} \rightarrow n_R = K + 1$

$$= \sum_{K=0}^{N-1} \frac{N!}{K!\,[(N-1)-K]!} p^{K+1} q^{(N-1)-K}$$

$$= N p \underbrace{\sum_{K=0}^{N-1} \frac{(N-1)!}{K!\,[(N-1)-K]!} p^{K} q^{(N-1)-K}}_{(p+q)^{N-1} = 1}$$

$$\overline{n_R} = Np$$

→) The dispersion

NOTE : $\Delta u = u - \bar{u}$

$$\overline{(\Delta u)^2} = \overline{(u-\bar{u})^2} = \overline{u^2 - 2u\bar{u} + \bar{u}^2} = \overline{u^2} - \bar{u}^2$$

(variance)

dispersion: $\sigma = \sqrt{\overline{(\Delta u)^2}} = \sqrt{\overline{u^2} - \bar{u}^2}$

$$\overline{(\Delta n_r)^2} = \overline{n_r^2} - \overline{n_r}^2$$

↳ we already know that $(\overline{n_r})^2 = (Np)^2$

$$\overline{n_r^2} = \sum_{n=0}^{N} n_r^2 \, W_N(n_r) = \sum_{n=0}^{N} n_r^2 \frac{N!}{n_r!\,(N-n_r)!} p^{n_r} q^{N-n_r}$$

$$n_r^2 p^{n_r} = \underbrace{p^2 \frac{\partial^2}{\partial p^2} p^{n_r}}_{\substack{p^2 \, n_r (n_r - 1) p^{n_r - 1} \\ n_r^2 p^{n_r} - n_r p^{n_r}}} + p \frac{\partial}{\partial p} p^{n_r}$$

$$\overline{n_r^2} = \sum_{n=0}^{N} \frac{N!}{n_r!\,(N-n_r)!} \left(p^2 \frac{\partial^2}{\partial p^2} + p \frac{\partial}{\partial p} \right) p^{n_r} q^{N-n_r}$$

$$= \left(p^2 \frac{\partial^2}{\partial p^2} + p \frac{\partial}{\partial p} \right) \underbrace{\sum_{n=0}^{N} \frac{N!}{n_r!\,(N-n_r)!} p^{n_r} q^{N-n_r}}_{(p+q)^N}$$

$$= p^2 N(N-1)(p+q)^{N-2} + pN(p+q)^{N-1}$$

$$\overline{n_r^2} = p^2 N^2 - p^2 N + pN$$

$$= (\overline{n_r})^2 + Np(\underbrace{1-p}_{q})$$

$$= (\overline{n_r})^2 + Npq$$

$$\overline{(\Delta n_r)^2} = \cancel{(\overline{n_r})^2} + Npq - \cancel{(\overline{n_r})^2}$$

$$\overline{(\Delta n_r)^2} = \boxed{Npq} \implies \boxed{\sigma_{n_r} = \sqrt{Npq}}$$

So $\overline{n_r}$ increases like N, but the width increases only like $\sqrt{N}$

→) Dispersion of m

$$m = n_r - n_l = 2n_r - N$$

$$\Delta m = m - \overline{m} = 2n_r - N - 2\overline{n_r} + N = 2(n_r - \overline{n_r})$$

$$(\Delta m)^2 = 4(n_r - \overline{n_r})^2 = 4(\Delta n_r)^2$$

$$\overline{(\Delta m)^2} = 4\overline{(\Delta n_r)^2} = \boxed{4Npq}$$

if $p = q = 1/2 \implies \begin{cases} \overline{(\Delta m)^2} = N \\ \boxed{\sigma_m = \sqrt{N}} \end{cases}$

Probability distribution for large N

To a good approximation, W can be considered as a continuous function ↳ $|W(n_1+1) - W(n_1)| \ll W(n_1)$ when N is large

$n_1 = \tilde{n}$ (maximum of W)

↳ determined from $\frac{dW}{dn_1} = 0$ or equivalently $\frac{d \ln W}{dn_1} = 0$

∘) Let us expand $\ln W(n_1)$ in a Taylor's series about $\tilde{n}$

[ln W varies slower than W, so the power series expansion converges faster]

$n_1 = \tilde{n}_1 + \eta$

$$\ln W(n_1) = \ln W(\tilde{n}) + \underbrace{\cancel{\left.\frac{d \ln W}{dn_1}\right|_{n_1=\tilde{n}}}}_{\text{0, because we are expanding around the maximum}} \eta + \frac{1}{2} \underbrace{\left.\frac{d^2 \ln W}{dn_1^2}\right|_{n_1=\tilde{n}}}_{\text{negative } -|B_2|} \eta^2 + \dots$$

$$\ln W(n_1) \simeq \ln W(\tilde{n}) - \frac{|B_2|}{2}\eta^2 = \underbrace{\ln W(\tilde{n}) + \ln\left(e^{-\frac{|B_2|}{2}\eta^2}\right)}_{\ln\left[W(\tilde{n})\, e^{-\frac{|B_2|}{2}\eta^2}\right]}$$

$$W(n_1) \simeq \tilde{W} e^{-\frac{1}{2}|B_2|\eta^2}$$

∘) $\ln W(n_1) = \ln \left[\frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1} \right]$

$$\ln W(n_r) = \ln N! - \ln n_r! - \ln (N-n_r)! + n_r \ln p + (N-n_r) \ln q$$

if $n \gg 1 \Rightarrow \ln n!$ is almost a continuous function of n

Hence

$$\frac{d \ln n!}{dn} \simeq \frac{\ln (n+1)! - \ln (n!)}{1} = \ln \left(\frac{(n+1)!}{n!} \right) = \ln (n+1)$$

$$\frac{d \ln n!}{dn} \simeq \ln n$$

$$\rightarrow \frac{d \ln W(n_r)}{dn_r} = -\ln n_r + \ln (N-n_r) + \ln p - \ln q$$

$$= \ln \left[\frac{(N-n_r)}{n_r} \frac{p}{q} \right]$$

Since $\frac{d \ln W(n_r)}{dn_r} = 0$ gives $\tilde{n}$

$$\Rightarrow \ln \left[\frac{(N-\tilde{n})}{\tilde{n}} \frac{p}{q} \right] = 0$$

$$(N-\tilde{n}) p = \tilde{n} q$$

$$Np - \tilde{n} p = \tilde{n} - \tilde{n} p$$

$$\boxed{\tilde{n} = Np}$$

$$\rightarrow \frac{d^2 \ln W(n_r)}{dn_r^2} = -\frac{1}{n_r} - \frac{1}{N-n_r}$$

$$\left. \frac{d^2 \ln W(n_r)}{dn_r^2} \right|_{n_r=\tilde{n}} = B_2 = -\frac{1}{(Np)} - \frac{1}{\underbrace{N-Np}_{Nq}} = \frac{-(q+p)}{Npq}$$

$$\boxed{B_2 = -\frac{1}{Npq}}$$

→) Normalization

$$\sum_{n_\pi=0}^{\infty} W(n_\pi) \simeq \int W(n_\pi)\, dn_\pi = \int_{-\infty}^{\infty} W(\tilde{n}+\eta)\, d\eta = 1$$

integrand makes negligible contribution to the integral whenever $|\eta|$ is large and W is far from its maximum at $\tilde{n}$

$$\tilde{W} \int_{-\infty}^{\infty} e^{-\frac{1}{2}|B_2|\eta^2} d\eta = \tilde{W} \sqrt{\frac{\pi\, 2}{|B_2|}} = 1$$

see next page how to solve this integral

$$W(n_\pi) = \sqrt{\frac{|B_2|}{2\pi}}\; e^{-\frac{1}{2}|B_2|(n_\pi-\tilde{n})^2}$$

$\tilde{n} = Np \quad ; \quad |B_2| = \dfrac{1}{Npq}$

$$W(n_\pi) = \frac{1}{\sqrt{2\pi Npq}} \exp\left[-\frac{(n_\pi - Np)^2}{2Npq}\right]$$

$$\boxed{W(n_\pi) = \frac{1}{\sqrt{2\pi\, \sigma_{n_\pi}^2}} \exp\left[-\frac{(n_\pi - \overline{n_\pi})^2}{2\,\sigma_{n_\pi}^2}\right]}$$

Gaussian distribution

Showing that

$$\int_{-\infty}^{\infty} e^{-Ax^2}\,dx = \sqrt{\frac{\pi}{A}}$$

$$= \left(\int_{-\infty}^{\infty} e^{-Ax^2}\,dx \int_{-\infty}^{\infty} e^{-Ay^2}\,dy \right)^{1/2} = \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-A(x^2+y^2)}\,dx\,dy \right)^{1/2}$$

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \qquad dx\,dy = r\,dr\,d\theta \leftarrow \begin{vmatrix} \partial x/\partial r & \partial x/\partial\theta \\ \partial y/\partial r & \partial y/\partial\theta \end{vmatrix} =$$

$$= \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r\cos^2\theta + r\sin^2\theta = r$$

$$= \left(\int_0^{2\pi} d\theta \int_0^{\infty} r\,dr\, e^{-Ar^2} \right)^{1/2} = \left(2\pi \,\frac{1}{2} \int_0^{\infty} dq\, e^{-Aq} \right)^{1/2} = \left(\pi \left. \frac{e^{-Aq}}{(-A)} \right|_{0}^{\infty} \right)^{1/2}$$

$r^2 = q$

$2r\,dr = dq$

$$= \sqrt{\frac{\pi}{A}} \;//$$

→) In terms of the actual displacement (x)

$$x = ml = (n_R - n_L)\,l = 2n_R\,l - Nl$$

$$\underline{dx = 2l\,dn_R}$$

{ Probability of finding the particle in the range $[x, x+dx]$ is equal to the probability of finding it between n_R and $n_R + dn_R$

$$\left.\begin{aligned} W(n_R)\,dn_R &= P(x)\,dx \\ W(n_R)\,\frac{dx}{2l} &= P(x)\,dx \end{aligned}\right\} \Rightarrow P(x) = \frac{W(n_R)}{2l}$$

$$P(x) = \frac{1}{\sqrt{2\pi\, l^2 4Npq}}\, e^{-\frac{(2l n_R - lN - 2l\tilde{n} + lN)^2}{2(l^2 4)Npq}}$$

$$\sigma_m^2 = 4Npq \longrightarrow \boxed{\sigma^2 = l^2\sigma_m^2}$$

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

← standard form of the Gaussian distribution

$$\left\{\begin{aligned} \mu &\equiv l\bar{m} = (p-q)Nl && \text{(average)} \\ \sigma^2 &\equiv l^2\sigma_m^2 = 4Npq\, l^2 && \text{(variance)} \end{aligned}\right.$$

Central limit theorem: the sum of many independent random variables has a probability distribution that converges to a Gaussian