Random walk and binomial distribution

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.

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

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

\[
P = \frac{1}{2}, \quad q = 1 - P
\]

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 step, \( p \neq q \).

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

\[
x = ml \quad \text{where } m \text{ is an integer}
\]

\[\begin{align*}
-\frac{N}{2} & \leq m \leq \frac{N}{2} \\
N & = \text{number (large number)}
\end{align*}\]
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 preferred \( B \))

- Diffusion [End point 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

- Polymers - long molecules (DNA, RNA, proteins, many plastics)
  - made up of small units called monomers
  - Temperature induces 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 \( x = m \ell \) after \( N \) steps?

\[ N \leq m \leq N \]

\( n_r = \) number of steps to the right
\( n_L = \) number of steps to the left

\[
\begin{align*}
\begin{cases}
  m &= n_r - n_L \\
  N &= n_r + n_L
\end{cases} \\
\Rightarrow & \quad m = n_r - (N - n_r) = 2n_r - N
\end{align*}
\]

\( \text{odd} \leftrightarrow \text{odd} \)

Steps are independent on the past

\( p = \) prob. that the step is to the right

\( q = 1 - p = \) " " " " lift

The probability for a specific sequence of \( n_r \) steps to the right and \( n_L \) steps to the lift is

\[
p \cdot p \cdots p \cdot q \cdot q \cdots q = p^{n_r} q^{n_L}
\]

But there are many different ways of taking \( n_r \) steps to the right and \( n_L \) steps to the lift

\[
\frac{N!}{n_r! \cdot n_L!} \quad \text{distinct possibilities}
\]

The probability \( W_N(n,m) \) of taking \( n \) steps to the right and \( m \) steps to the left after \( N \) total steps is then

\[
W_N(n,m) = \frac{N!}{n!m!(N-n)!} \quad \text{Binomial distribution}
\]

\( \Rightarrow \) **Note:**

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

\[
\begin{align*}
\text{RRL} & \\
\text{RLR} & \\
\text{LRR} & \\
\end{align*}
\]

\( 3 \rightarrow \frac{3!}{2!1!} = 3 \)

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

\[
\begin{align*}
\text{RRLL} & \\
\text{RLRL} & \\
\text{RLLR} & \\
\text{LRRR} & \\
\text{LRRL} & \\
\text{LRLR} & \\
\text{LLRR} & \\
\end{align*}
\]

\( 6 \rightarrow \frac{4!}{2!2!} = \frac{4 \cdot 3}{2} = 6 \)

\( \Rightarrow \) **Note:**

The binomial distribution has this name because \( \frac{N!}{n!(N-n)!} \) is the form appearing in \( (p+q)^N \)

\[
(p+q)^N = \sum_{n=0}^{N} \frac{N!}{n!(N-n)!} p^n q^{N-n}
\]
\[ P_n(m) = W_n(nn) \]

\[
\begin{align*}
\sum \begin{array}{l}
\eta_n \eta_m = N \\
\eta_n - \eta_m = m
\end{array}
\end{align*}
\]

\[
\frac{\eta_n}{2} = \frac{N+m}{2} \quad , \quad \frac{\eta_m}{2} = \frac{N-m}{2}
\]

\[
P_n(m) = \frac{N!}{\eta_n! (N-\eta_n)!} p^\eta_n (1-p)^{N-\eta_n} = \frac{N!}{(\frac{N+m}{2})! (\frac{N-m}{2})!} p^{\frac{N+m}{2}} (1-p)^{\frac{N-m}{2}}
\]

**Exercise 1:**
Suppose \( p=q=\frac{1}{2} \), \( N=3 \)

What are the probabilities for \( \eta_n = 0, 1, 2, 3 \)?

\[
\begin{align*}
\eta(0) &= \left(\frac{1}{2}\right)^2 = \frac{1}{4} \\
\eta(1) &= 3 \left(\frac{1}{2}\right)^3 = 3/8 \\
\eta(2) &= \left(\frac{1}{2}\right)^3 = 3/8 \\
\eta(3) &= \left(\frac{1}{2}\right)^2 = \frac{1}{4}
\end{align*}
\]

(totals sum = 1)

**Exercise 2:**
Suppose \( p=q=\frac{1}{2} \), \( N=20 \)

Make a plot illustrating all cases from \( \eta_n = 0 \) to \( \eta_n = 20 \).

Show both \( P_n(m) \) vs \( m \) and \( W_n(nn) \) vs \( \eta_n \).

It should look like

[Graph showing a bell-shaped curve]

Each vertical line indicates the probability for one value of \( \eta_n \).

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

\[ \sum_{n_L=0}^{N} \frac{N!}{n_L!(N-n_L)!} p^{n_L} q^{N-n_L} = (p+q)^N \]

\[ \frac{1}{q} \int_0^q \frac{1}{p} \int_0^p \]

→ Mean number \( \overline{n}_n \) of steps to the right

\[ \overline{n}_n = \sum_{n_L=0}^{N} \frac{N!}{n_L!(N-n_L)!} p^{n_L} q^{N-n_L} \]

1. Let us assume that \( p \) and \( q \) are two arbitrary, independent parameters

\[ n_L p^L = p \frac{\partial}{\partial p} (p^n) \]

\[ \overline{n}_n = \sum_{n_L=0}^{N} \frac{N!}{n_L!(N-n_L)!} p \frac{\partial}{\partial p} (p^n) q^{N-n_L} = \frac{\partial}{\partial p} \left( \sum_{n_L=0}^{N} \frac{N!}{n_L!(N-n_L)!} p^{n_L} q^{N-n_L} \right) \]

1. Now we go back to our specific problem where \( q = 1-p \)

\[ \overline{n}_n = Np \]

\[ \overline{n}_e = Nq \]

\[ \overline{n}_n + \overline{n}_e = N(p+q) = N \]

\[ m = n_n - n_e \Rightarrow \overline{m} = \overline{n}_n - \overline{n}_e = N(p-q) \]

\[ \overline{m} = N(p-q) \]

If \( p = q \) \( \Rightarrow \overline{m} = 0 \)

That is complete symmetry between right and left
Another way to arrive at
\[ \hat{n}_n = Np \]

\[ \hat{n}_n = \sum_{n_1=0}^{N} \frac{n_1 N!}{n_1! (N-n_1)!} p^{n_1} q^{N-n_1} \]

This is zero when \( n_1 = 0 \)

\[ = \sum_{n_1=1}^{N} \frac{n_1 N!}{n_1! (N-n_1)!} p^{n_1} q^{N-n_1} = \sum_{n_1=1}^{N} \frac{N!}{(n_1-1)! (N-n_1)!} p^{n_1} q^{N-n_1} \]

using \( K = n_1 - 1 \) \( \Rightarrow n_1 = K + 1 \)

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

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

\[ = \left( p + q \right)^{N-1} = 1 \]

\[ \hat{n}_n = Np \]
The dispersion

\[ \Delta \mu = \mu - \overline{\mu} \]
\[ (\Delta \mu)^2 = (\mu - \overline{\mu})^2 = \mu^2 - 2\mu \overline{\mu} + \overline{\mu}^2 = \overline{\mu}^2 - \mu^2 \]
\[ \text{variance} \]
\[ \sigma = \sqrt{(\Delta \mu)^2} = \sqrt{\overline{\mu}^2 - \mu^2} \]

\[ (\Delta n_\pi)^2 = \overline{n_\pi}^2 - \overline{n_\pi}^2 \]

Let us already know that \((\overline{n_\pi})^2 = (\overline{n}_p)^2\)

\[ \overline{n_\pi}^2 = \sum_{n=0}^{N} n_\pi \overline{\pi} W_n (n_\pi) = \sum_{n=0}^{N} n_\pi \frac{N!}{n_\pi! (N-n_\pi)!} p_{n_\pi} q^{N-n_\pi} \]

\[ n_\pi \mathbf{p}_{n_\pi} = \frac{p^2 \frac{\partial}{\partial p^2}}{p^2} \mathbf{p}_{n_\pi} + p \frac{\partial}{\partial p} \mathbf{p}_{n_\pi} \]

\[ \mathbf{p}_{n_\pi} = \frac{n_\pi}{n_\pi} (n_\pi-1) \mathbf{p}_{n_\pi-1} \]

\[ \overline{n_\pi}^2 = \sum_{n=0}^{N} \frac{N!}{n_\pi! (N-n_\pi)!} \left( \frac{p^2 \frac{\partial}{\partial p^2}}{p^2} + p \frac{\partial}{\partial p} \right) \mathbf{p}_{n_\pi} q^{N-n_\pi} \]

\[ = \left( \frac{p^2 \frac{\partial}{\partial p^2}}{p^2} + p \frac{\partial}{\partial p} \right) \sum_{n=0}^{N} \frac{N!}{n_\pi! (N-n_\pi)!} \mathbf{p}_{n_\pi} q^{N-n_\pi} \]

\[ = p^2 N (N-1) (p + q)^{N-2} + p N (p + q)^{N-1} \]
\[ \overline{\sigma_i} = p^2 N^2 - p^2 N + p N \]
\[ = \left( \overline{\sigma_i} \right)^2 + Np \left( 1 - p \right) \frac{q}{q} \]
\[ = \left( \overline{\sigma_i} \right)^2 + Np q \]

\[ \overline{(\Delta \sigma_i)^2} = \left( \overline{\sigma_i} \right)^2 + Np q - \left( \overline{\sigma_i} \right)^2 \]

\[ \overline{(\Delta \sigma_i)^2} = Np q \]

\[ \therefore \sigma_m = \sqrt{Np q} \]

So \( \overline{\sigma_i} \) increases like \( N \), but \( \Delta \) width increases only like \( \sqrt{N} \)

\( \rightarrow \) Dispersion of \( m \)

\[ m = \sigma_i - \sigma_i \]
\[ = 2 \sigma_i - N \]

\[ \Delta m = m - \overline{m} = 2 \sigma_i - N - 2 \overline{\sigma_i} + N = 2 (\overline{\sigma_i} - \overline{\sigma_i}) \]

\[ (\Delta m)^2 = 4 (\overline{\sigma_i} - \overline{\sigma_i})^2 = 4 (\Delta \sigma_i)^2 \]

\[ (\Delta m)^2 = 4 (\Delta \sigma_i)^2 = 4 N p q \]

If \( p = q = \frac{1}{2} \)

\[ \begin{cases} (\Delta m)^2 = N \\ \sigma_m = \sqrt{N} \end{cases} \]
Probability distribution for large $N$

To a good approximation, $W(n)$ can be considered as a continuous function

$$L = |W(n+1) - W(n)| \ll W(n)$$

when $N$ is large

$$n = \bar{n} \quad \text{(maximum of } W)$$

$$L \text{ determined from } \frac{dW}{dn} = 0 \quad \text{or equivalently } \frac{d\ln W}{dn} = 0$$

2) Let us expand $\ln W(n)$ in a Taylor's series about $\bar{n}$

[$\ln W$ varies slower than $W$, so the power series expansion emerges faster]

$$n = \bar{n} + \eta$$

$$\ln W(n) = \ln W(\bar{n}) + \frac{d\ln W}{dn} \bigg|_{n=\bar{n}} \eta + \frac{1}{2} \frac{d^2\ln W}{dn^2} \bigg|_{n=\bar{n}} \eta^2 + \ldots$$

0, because we are expanding around the maximum

$$\text{negative} \quad -1B_2$$

$$\ln W(n) \approx \ln W(\bar{n}) - \frac{1}{2} B_2 \eta^2 = \frac{\ln W(\bar{n}) + \ln \left( e^{-\frac{1}{2} B_2 \eta^2} \right)}{\ln \left[ W(\bar{n}) e^{-\frac{1}{2} B_2 \eta^2} \right]}$$

$$W(n) \approx W e^{-\frac{1}{2} B_2 \eta^2}$$

$$\ln W(n) = \ln \left[ \frac{N!}{n! (N-n)!} \right]^{\eta}$$

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

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

if \( n \gg 1 \Rightarrow \ln n! \) is almost a linear function of \( n \)

\[
\frac{d \ln n!}{dn} \approx \frac{\ln (n+1)! - \ln n!}{1} = \ln \left( \frac{(n+1)!}{n!} \right) = \ln (n+1)
\]

\[
\frac{d \ln n!}{dn} \approx \ln n
\]

\[
\Rightarrow \frac{d \ln W(na)}{dna} = -\ln na + \ln (N-na) + \ln p - \ln q
\]

\[
= \ln \left[ \frac{(N-na)(p)}{naq} \right]
\]

Since \( \frac{d \ln W(na)}{DNA} = 0 \) gives \( \tilde{n} \)

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

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

\[
Np - \tilde{n}p = \tilde{n}q
\]

\[
\tilde{n} = \frac{Np}{q}
\]

\[
\Rightarrow \frac{d^2 \ln W(na)}{dna^2} = \frac{1}{na} - \frac{1}{N-na}
\]

\[
\frac{d^2 \ln W(na)}{dna^2} \bigg|_{na = \tilde{n}} = B_2 = -\frac{1}{(Np)} - \frac{1}{\frac{N-Np}{Nq}} = -\frac{(q-p)}{Np \cdot q}
\]

\[
B_2 = -\frac{1}{Np \cdot q}
\]
1. Normalization

\[ \sum_{n=0}^{\infty} W(n) = \int W(n) dn = \int_{-\infty}^{\infty} W(s+n) dn = 1 \]

The integrand makes negligible contribution to the integral whenever \( n \) is large and \( W \) is far from its maximum at \( \bar{n} \).

\[ \tilde{W} \int_{-\infty}^{\infty} e^{-\frac{1}{2} B_2 \eta^2} d\eta = \tilde{W} \sqrt{\frac{\pi}{2 B_2}} = 1 \]

\[ \tilde{W} = \sqrt{\frac{B_2}{2 \pi}} \]

\[ W(n) = \sqrt{\frac{1}{2\pi}} e^{-\frac{1}{2} B_2 (n - \bar{n})^2} \]

\( \bar{n} = N \bar{p} \) \quad \text{and} \quad |B_2| = \frac{1}{N \bar{p} \eta} \]

\[ W(nn) = \frac{1}{\sqrt{2\pi N \bar{p} \eta}} \exp \left[ -\frac{(nn - N \bar{p})^2}{2N \bar{p} \eta} \right] \]

\[ W(nn) = \frac{1}{\sqrt{2\pi \sigma_n^2}} \exp \left[ -\frac{(n - \sqrt{\bar{n}})^2}{2 \sigma_n^2} \right] \]

Gaussian distribution
\[ \int_{-\infty}^{\infty} e^{-Ax^2} \, dx = \sqrt{\frac{\pi}{A}} \]

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

\[ = \left( \int_{-\infty}^{\infty} e^{-A(r^2)} \, dr \, d\theta \right)^{1/2} \]

\[ x = \rho \cos \theta \]
\[ y = \rho \sin \theta \]

\[ dx \, dy = \rho \, d\rho \, d\theta \]

\[ \begin{vmatrix} \partial x/\partial \rho & \partial x/\partial \theta \\ \partial y/\partial \rho & \partial y/\partial \theta \end{vmatrix} = \begin{vmatrix} \cos \theta & -\rho \sin \theta \\ \sin \theta & \rho \cos \theta \end{vmatrix} = \rho \cos^2 \theta + \rho \sin^2 \theta = \rho \]

\[ = \left( 2\pi \int_{0}^{\infty} e^{-\rho^2} \, d\rho \right)^{1/2} = \left( \int_{0}^{\infty} e^{-\rho^2} \, d\rho \right)^{1/2} \]

\[ \rho^2 = \theta \]
\[ 2 \rho \, d\rho = d\theta \]

\[ = \sqrt{\frac{\pi}{A}} \]

\]
\( x = ml = (nn - ne) \ell = 2nn \ell - N\ell \)

\[
\frac{dx}{dm} = 2\ell \frac{dm}{m}
\]

\[
\begin{aligned}
\text{Probability of finding the particle in the range } [x, x + dx] \\
\text{is equal to the probability of finding it between } nn \text{ and } nn + dm
\end{aligned}
\]

\[
\begin{align*}
W(nn) \frac{dm}{m} &= \mathcal{P}(x) \, dx \\
W(nn) \frac{dx}{2\ell} &= \mathcal{P}(x) \, dx
\end{align*}
\]

\[
\mathcal{P}(x) = \frac{1}{\sqrt{2\pi} \sigma^2 4Npq} e^{-\frac{(x - \mu)^2}{2\sigma^2}}
\]

\[
\sigma_m^2 = 4Npq \Rightarrow \sigma^2 = \ell^2 \sigma_m^2
\]

\[
\mathcal{P}(x) = \frac{1}{\sqrt{2\pi} \sigma^2} e^{-\frac{(x - \mu)^2}{2\sigma^2}} \quad \text{standard form of the Gaussian distribution}
\]

\[
\begin{align*}
\mu &\equiv \ell \bar{m} = (p - q) N\ell \\
\sigma^2 &\equiv \ell^2 \sigma_m^2 = 4Npq \ell^2
\end{align*}
\]

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