# Bayesian approach : Estimation of the Diffusion constant

**Simulating brownian motion:**

We have considered the problem of estimating the mean and the best estimate for the diffusion constant of a Gaussian process with a measure of reliability using Bayes theorem for the given experimental data sets.

**Single particle trajectory :**

We generated the motion of a single particle in 1 dimension by generating a vector of random displacements for a continuous time stochastic process and is shown in Figure 1. N is the number of samples to be generated and the cumulative sum represents the trajectory of the particle in 1D.

**Example : Diffusion Equation**

$$\partial_t P = D \partial_x^2 P \tag{1}$$

The solution of the above equation is given by $P(xt|x't') = \frac{1}{\sqrt{(2\pi D(t-t'))}} \exp(-\frac{(x-x'^2)}{D(t-t')})$ where D is the diffusion constant. The mean squared displacement of the particle is given by $(x-x'^2) = \int dx P(xt|x't')(x-x'^2)$. According to theory, the mean squared displacement of the particle is proportional to the time interval which is given by $< (x-x'^2) >= 2D((t-t')$, where x(t) = position, D = diffusion coefficient, and $\tau$ be the time lag. To accurately model a real particle, the distribution of random displacements should match the experimental conditions. The displacement should increase in proportion to the square root of time(t). However, the mean squared displacement for the real data for our study does not match the theory and we noticed drift in the data sets. The commands and the results of the simulated data from the experiments are given below :

# 1 Bayesian inference: Parameter estimation

The given data $x = x_1, x_2...x_n$ generated from brownian motion are collected at times $t_1 < t_2 < t_n$. The probability of the brownian path is given by

$$P(x_n t_n....x_1 t_1|D) = P(x_n t_n|x_{n-1} t_{n-1})...P(x_1 t_1)P(D) = (\frac{1}{2\pi D(t_n - t_{n-1})})^{\frac{N-1}{2}} \exp(-\sum_{i=1}^{N-1} \frac{(x_n - x_{n-1})^2}{2D(t_n - t_{n-1})})$$

The poterior probability distribution function for the diffusion constant is given as follows using Bayes theorem.

$$P(D|xt) = \frac{P(xt|D)P(D)}{P(xt)} \tag{2}$$

$P(D|xt$ : Posterior probability, $P(xt|D)$ : likelihood, P(D) be the Jefferey's prior. The denominator is the normalisation constant and $xt$ is the data set from the experiments.

The posterior probability for the above sample path is given by

$$P(D|xt) = \frac{(\frac{1}{\sqrt{(2\pi D \nabla_t)}})^{N-1} \exp(-\sum_{i=1}^{N-1} \frac{(\delta x_i)^2}{2D\nabla_t}) P(D)}{P(\{xt\})} \tag{3}$$

The poterior pdf gives our inference about the value of the parameters for the given data. The best estimate of the quantity of interest ( D in our problem) is given by the maximum posterior pdf about its maximum value $D_0$ and the measure of reliability or the error bar of this best estimate depends how the width of the posterior pdf about $D_0$ looks like. Taking logarithm of Eq.(3) is given by

$$L = \log(P(D|xt)) \tag{4}$$

Expand L about $D = D_0$, we get

$$L = L_0 + \frac{dP}{dt}|D_0(D - D_0) + \frac{1}{2}\frac{d^2 P}{dt^2}|D_0(D - D_0)^2 + ... \tag{5}$$

$$\log(P(D|xt) = -\frac{N-1}{2}\log(2\pi\nabla_t) - \log D - \frac{\delta x^2}{2D\nabla_t} \tag{6}$$

The best estimate of $D_0$ is given by the condition $\frac{dP}{dt}|D_0 = 0$ and it is estimated as by taking derivative of Eq.(6)

$$D_0 = \frac{\delta x^2}{(N+1)t} \tag{7}$$

To find the error bar the second derivative of Eq.(6) is taken and it is evaluated at $D = D_0$. It is calculated as

$$\frac{d^2 P}{dt^2}|D_0 = -\frac{1}{D_0^2}\left[N - 2 + \frac{\delta x^2}{D_0 t}\right] \tag{8}$$

The measure of reliability is found using $\sigma = (-\frac{d^2 P}{dt^2}|D_0)^{-0.5}$ and for the diffusion process of our experimental data the error bar is estimated as $\frac{D}{\sqrt{((N-2+\frac{\delta x^2}{D_0\nabla_t})}}$ The optimal value of D is calculated using $D = D_0 \pm \sigma(D) = \frac{\delta x^2}{(N+1)t} \pm \frac{D_0}{\sqrt{((N-2+\frac{\delta x^2}{D_0\nabla_t})}}$ $D_0$ is the best estimate of the value of D and $\sigma$ is the error bar. Taking exponential of Eq.(6), we get

$$(P(D|xt) = A\exp(\frac{1}{2}\frac{d^2 P}{dt^2}|D_0(D - D_0)^2) \tag{9}$$

$$P(D_0 - \sigma < D < D_0 + \sigma|xt = \int_{D_0-\sigma}^{D_0+\sigma P(D|xt} P(D)dD \tag{10}$$

. A is the normalisation constant and the posterior pdf is approximated by the Gaussian distribution which tells us about the positions of its centre and is we assign a simple uniform pdf for the prior,

$$P(D) = \frac{1}{(D_{max} - D_{min})} \tag{11}$$