

From Statistical Physics to Inference with Applications to Data-Driven Modelling in Quantitative Biology: Tutorial 1

S.C.,R.M.,F.Z.

Bayesian inference and single-particle tracking

Characterising the motion of biomolecules (proteins, RNA, ...) or complexes (vesicles, ...) inside the cell is fundamental to the understanding of many biological processes [1]. Optical imaging techniques now allow for the tracking of single particles in real time [2]. The goal of this tutorial is to understand how the diffusion coefficient can be reconstructed from the recordings of the trajectory of a particle, and how the accuracy of the inference is affected by the number of data points (recording length) [3]. The diffusion coefficient depends on the diffusive properties of the environment and on the size of the object. Supposing that the data are obtained in water, from the diffusion coefficient reconstructed from the data the characteristic size of the diffusing object will then be extracted, and a connection with characteristic biological sizes will be made.

Problem

We consider a particle undergoing diffusive motion in the plane, with position $\mathbf{r}(t) = \{x(t), y(t)\}$ at time t . The diffusion coefficient (supposed to be isotropic) is denoted by \mathcal{D} , and we assume that the average velocity vanishes. Measurements give access to the positions x_i, y_i of the particles at times t_i , where i is a positive integer running from 1 to M .

Data:

Several trajectories of the particle can be downloaded from the book webpage, see Tutorial 1 repository. Each file contains a three-column array (t_i, x_i, y_i) , where t_i is the time, x_i and y_i are the measured coordinates of the particle, and i is the measurement index, running from 1 to M . The unit of time is seconds and displacements are in μm .

Questions:

1. Write a script to read the data. Start by the file `dataM1000d2.5.dat`, and plot the trajectories in the (x, y) plane. What are their characteristics? How do they fill the space? Plot the displacement $r_i = \sqrt{x_i^2 + y_i^2}$ as a function of time. Write the random-walk relation between displacement and time in two dimensions, defining the diffusion coefficient \mathcal{D} . Give a rough estimation of the diffusion coefficient from the data.
2. Write down the probability density $p(\{x_i, y_i\}|\mathcal{D}, \{t_i\})$ of the time series $\{x_i, y_i\}_{i=1, \dots, M}$, given \mathcal{D} and the measurement times, and deduce, using Bayes rule, the posterior density of probability for the diffusion coefficient, $P(\mathcal{D}|\{t_i, x_i, y_i\})$.
3. Calculate analytically the most likely value of the diffusion coefficient, its average value, and its variance, assuming a uniform prior on \mathcal{D} .
4. Plot the posterior distribution of \mathcal{D} obtained from the three data files. Compute, for the given datasets, the values of the mean and of the variance of \mathcal{D} , and its most likely value. Compare the results obtained with different number M of measures.
5. Imagine that the data correspond to a spherical object diffusing in water (of viscosity $\eta = 10^{-3} \text{ Pa s}$). Use the Einstein-Stokes relation,

$$\mathcal{D} = \frac{k_B T}{6\pi\eta\ell}, \quad (1)$$

(here ℓ is the radius of the spherical object and η is the viscosity of the medium) to deduce the size of the object. Biological objects going from molecules to bacteria display diffusive motions, and have

characteristic size ranging from nm to μm . For proteins $\ell \approx 1-10$ nm, while for viruses $\ell \approx 20-300$ nm and for bacteria, $\ell \approx 2-5 \mu\text{m}$. Among the molecules or organisms described in the table below, which ones could have a diffusive motion similar to the one displayed by the data?

object	ℓ (nm)
small protein (lysozyme) (100 residues)	1
large protein (1000 residues)	10
influenza viruses	100
small bacteria (e-coli)	2000

6. In many cases the motion of particles is not confined to a plane. Assuming that x_i, y_i are the projections of the three dimensional position of the particle in the plane perpendicular to the imaging device (microscope) how should the procedure above be modified to infer \mathcal{D} ?

References

- [1] H. C. Berg, *Random Walks in Biology*, Princeton University Press (1993).
- [2] N. Ruthardt, D.C. Lamb, and C. Bräuchle, *Single-particle Tracking as a Quantitative Microscopy-based Approach to Unravel Cell Entry Mechanisms of Viruses and Pharmaceutical Nanoparticles*, Molecular Therapy 19 (7), 1199-1211 (2011).
- [3] A. Robson, K. Burrage, M.C. Leake, *Inferring diffusion in single live cells at the single-molecule level*, Phil Trans R Soc B 368: 20120029 (2013).