Estimating the diffusion constant from a trajectory

Suppose we make a number of observations of the position of an object undergoing Brownian motion. How can we best estimate the diffusion constant D from those observations? How good will our estimate be? We will work in one dimension for the sake of simplicity.

Warm-up: a single observation

Even a single observation, where we measure the position $X(t_1)$ of the diffusing object at time t_1, is enough to make a very crude estimate of the diffusion constant D. Using the relation

$$\langle X_1^2 \rangle \equiv \langle X(t_1)^2 \rangle = 2D\,t_1 \tag{1}$$

it is clear that the only good estimator for D is the following (we will always use hats to denote estimators):

$$\hat{D_1} \equiv \frac{X_1^2}{2\,t_1} \tag{2}$$

Just how crude is this estimate for D, which is based on a single observation of the diffuser? Its variance is

$$\mathrm{Var}\,\hat{D_1} = \langle \hat{D_1}^2 \rangle - \langle \hat{D_1} \rangle^2 \tag{3}$$

To evaluate the first term, we use the fact that the fourth moment of a Gaussian variable X is related simply to the second moment ("Wick's Theorem"):

$$\langle X^4 \rangle = 3\langle X^2 \rangle^2 \tag{4}$$

Therefore

$$\mathrm{Var}\,\hat{D_1} = \frac{3 \cdot (2D\,t_1)^2}{(2\,t_1)^2} - D^2 = 2D^2 \tag{5}$$

So, not such a good estimate: our standard relative error will be 141%! In addition to finding the variance of the estimator, we can also compute its complete distribution. If we divide our estimator by D, we get a quantity that is the square of a normalized, centered Gaussian. That means that $\hat{D_1}/D$ follows a χ^2 distribution with a single degree of freedom. Fortunately, there is an analytic expression for this distribution. From it, we get

$$\mathrm{Prob}\left[\frac{\hat{D_1}}{D} = x\right] = \frac{1}{\sqrt{2} \cdot \Gamma(\frac{1}{2})} \cdot x^{-1/2} e^{-x/2} \tag{6}$$

This probability distribution is plotted in Figure 1. Note that it diverges for small values!

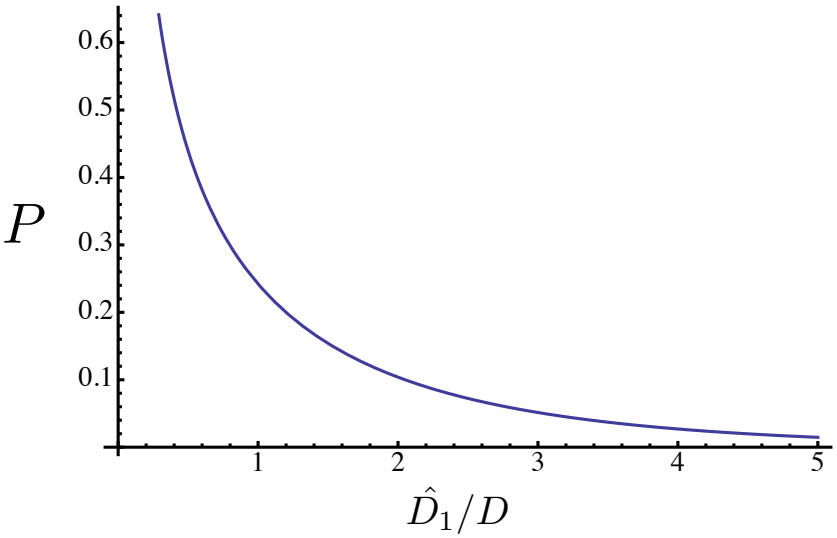

Figure 1: Probability density for the estimator $\hat{D}_1$.

The first nontrivial case: two observations

If the diffuser's position is measured at two times $t_2 > t_1$, giving results X_1 and X_2, how do we now estimate the diffusion constant? There are at least three terms that we can play with:

$$\frac{X_1^2}{2\,t_1}, \qquad \frac{(X_2 - X_1)^2}{2\,(t_2 - t_1)}, \qquad \frac{X_2^2}{2\,t_2} \tag{7}$$

Each of these is an unbiased estimator for D. We will consider the estimator $\hat{D}_2$ to be a weighted average of these terms, with coefficients chosen later so as to minimize the variance (thus giving the 'best' estimator).

$$\hat{D}_2 \equiv \alpha \cdot \frac{X_1^2}{2\,t_1} + \beta \cdot \frac{(X_2 - X_1)^2}{2\,(t_2 - t_1)} + (1 - \alpha - \beta) \cdot \frac{X_2^2}{2\,t_2} \tag{8}$$

The variance of $\hat{D}_2$ can be calculated. It's not worth writing out all the steps here; the result is a nasty expression quadratic in α and β, which can then be minimized with respect to these parameters. One finds that the minimal variance is achieved when $\alpha = \beta = 1/2$. That is, the best estimator for D

using two observations is

$$\hat{D_2} \equiv \frac{1}{2}\left[\frac{X_1^2}{2\,t_1} + \frac{(X_2 - X_1)^2}{2\,(t_2 - t_1)}\right] \tag{9}$$

Its variance is

$$\mathrm{Var}\,\hat{D_2} = \frac{3}{4}\cdot D \tag{10}$$

There are two interesting things to note. First of all, the optimal estimator involves the squared displacement from time 0 to t_1 and the displacement from t_1 to t_2, but not the total displacement from 0 to t_2. It seems that this information is redundant, and would not improve our estimate at all. A second interesting point is that the quality of our estimator (as measured by its variance) does not depend at all on the times at which we measure the position of the diffuser. We could take t_1 to be extremely small and t_2 to be very large, or we could take $t_2 = 2 \cdot t_1$. It doesn't matter at all. If you go through the calculation of the variance, you find that the times t_1 and t_2 cancel out of the result. This is a reflection of the scale-invariance of true Brownian motion, I think. For a real diffusing particle, the choice of observation times would probably not matter, so long as the time intervals were larger than the characteristic decay time of the velocity autocorrelation function.

The case of n observations

Now we move on to the general case. It's pretty easy, now, to guess what the optimal estimator for the diffusion constant should be. But let's derive it in a different way, from the maximum likelihood criterion. That is, we will choose $\hat{D_n}$ so that, given some observed values $x_1, x_2, \ldots x_n$, the likelihood of these observations is maximized (assuming Brownian motion with diffusion constant $\hat{D_n}$). That likelihood of the observations, assuming diffusion constant D, is

$$\mathrm{Prob}_D(X_1 = x_1, X_2 = x_2, \ldots) = \prod_{k=1}^{n} \frac{\exp\left[-(x_k - x_{k-1})^2/4D(t_k - t_{k-1})\right]}{\sqrt{4\pi D(t_k - t_{k-1})}} \tag{11}$$

Here we have defined $t_0 = 0$ for convenience. Maximizing the logarithm of the likelihood is equivalent (and much easier!) than maximizing the likelihood.

$$0 = \frac{\partial \log \mathrm{Prob}_D}{\partial D} = \sum_{k=1}^{n} \frac{(x_k - x_{k-1})^2}{4D^2(t_k - t_{k-1})} - \frac{n}{2D} \tag{12}$$

We define the estimator $\hat{D}_n$ to be the value of D that maximizes the likelihood:

$$\hat{D}_n = \frac{1}{n} \sum_{k=1}^{n} \frac{(x_k - x_{k-1})^2}{2\,(t_k - t_{k-1})} \tag{13}$$

To analyze the quality of this estimator, let's relate it to a random variable with a known distribution. Specifically, we note that $n\hat{D}_n/D$ is a sum of n centered Gaussians normalized to have unit variance. Such a random variable is distributed according to a χ^2 distribution with parameter (number of degrees of freedom) n:

$$\frac{n\hat{D}_n}{D} = \sum_{k=1}^{n} \frac{(x_k - x_{k-1})^2}{2\,D(t_k - t_{k-1})} \sim \chi^2_n \tag{14}$$

This distribution, which takes on positive real values only, has the probability density function (according to Wikipedia)

$$f(x) = \frac{x^{n/2-1} e^{-x/2}}{2^{n/2}\Gamma(n/2)} \tag{15}$$

In our problem, x denotes the possible value of $n\hat{D}_n/D$. This implies (via the usual confusing change-of-variables thing with probability density functions) that $\hat{D}_n/D$ is distributed according to

$$\text{Prob}\left[\frac{\hat{D}_n}{D} = x\right] = nf(nx) = \frac{n(nx)^{n/2-1} e^{-nx/2}}{2^{n/2}\Gamma(n/2)} \tag{16}$$

Several of these distributions are plotted in Figure 2. It can be shown that

$$\text{Var}\,\hat{D}_n = \frac{2}{n} \cdot D^2 \tag{17}$$

With $n = 50$ observation points, our relative error is 20 %. To get an estimate whose standard (relative) error is 10 %, one needs 200 observation points!

"Sufficiency" of $\hat{D}_n$

We've derived the estimator $\hat{D}_n$ as the maximum likelihood estimator for the diffusion constant D, given n observations of the diffuser's position. But can we be sure that this is really the best possible estimator given those data? In statistics there is a concept of a "sufficient statistic." Our estimator

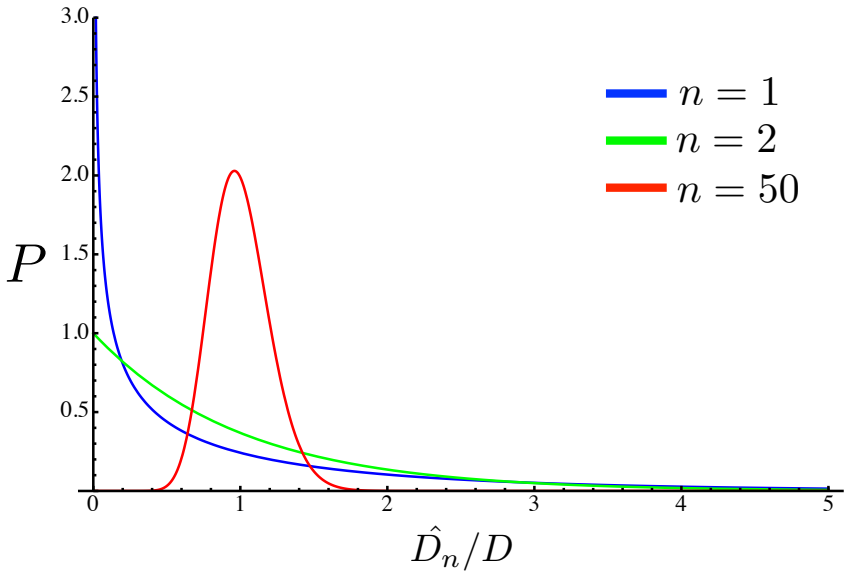

Figure 2: Probability density for the estimator $\hat{D}_n$ for $n = 1, 2$, and 50.

$\hat{D}_n$ is said to be a sufficient statistic for D if the conditional probability of observing the values $x_1, x_2, \ldots, x_n$, assuming a fixed value of $\hat{D}_n$, does not depend on the underlying parameter D. That is:

$$\text{Prob}_D\left(x_1, x_2, \ldots, x_n | \hat{D}_n = D_0\right) \qquad \text{independent of } D \tag{18}$$

This is rather a tricky concept. Intuitively, it means that once the value of the estimator $\hat{D}_n$ has been specified, there is no further information, in the observed data, concerning the value of the parameter D to be estimated. Now, it is possible (a little tricky) to verify that this equation holds for our case. Fortunately though, there is a theorem called the Fisher-Neyman factorization theorem that gives an easy characterization of sufficient statistics. It tells us that $\hat{D}_n$ is a sufficient statistic for D if and only if the probability of the observations can be factored as follows:

$$\text{Prob}_D(x_1, x_2, \ldots) = h(x_1, x_2, \ldots) \cdot g(D, \hat{D}_n) \tag{19}$$

That is, the likelihood function can be written as the product of a factor which is independent of the parameter to be estimated (D), times a factor that depends on the observed data only through the estimator $\hat{D}_n$. In our

case, looking at the likelihood function in Eqn. (11), we can just take $h = 1$. Our likelihood function only depends on the observed data through the estimator! Therefore the estimator is a sufficient statistic for D.