Biophysical Journal, Volume 98

**Supporting Material**

# Quantifying biomolecule diffusivity using an optimal Bayesian method

Guillaume Voisinne, Antigoni Alexandrou, and Jean-Baptiste Masson

# Supplementary Material to :

# Quantifying biomolecule diffusivity using an optimal Bayesian method.

Guillaume Voisinne[a], Antigoni Alexandrou[b] , Jean-Baptiste Masson[a]

[a] Institut Pasteur, CNRS URA 2171, Unit In Silico Genetics, 75724 Paris Cedex 15, France
[b] Laboratoire d'Optique et Biosciences, Ecole Polytechnique, CNRS, INSERM, 91128 Palaiseau Cedex, France

# 1 Transition probabilities in different geometries

We expose here the transition probabilities corresponding to most of the geometries that may be encountered experimentally. For rectangular domains of length $L$ and width $W$, the problem can be further simplified since $\Delta$ is the sum of $\partial_x^2$ and $\partial_y^2$. The transition probability can thus be factorized, $P(\vec{r},t|\vec{r}_0,t_0) = P(x,t|x_0,t_0) \times P(y,t|y_0,t_0)$, ($x$ and $y$ being the cartesian coordinates), and we find that

$$P(x,t|x_0,t_0) = \sum_{k=0}^{\infty} a_k \cos(k\frac{\pi}{L}x_0)\cos(k\frac{\pi}{L}x)e^{-k^2\frac{\pi^2}{L^2}D(t-t_0)}. \tag{1}$$

with $a_0 = \frac{1}{L}$ and $a_{k\geq 1} = \frac{2}{L}$. $P(y,t|y_0,t_0)$ is deduced by substituting the appropriate variables. For circular and elliptic domains, no such factorization is possible. In circular geometry, eigenfunctions of the laplacian operator involve Bessel functions [1] and the transition probability expressed in polar coordinates for a circular domain of radius $a$ reads

$$P(r,\theta,t|r_0,\theta_0,t_0) = \sum_{m=0}^{\infty}\sum_{i=1}^{\infty} A_{m,i}\cos\left(m\left(\theta-\theta_0\right)\right) J_m\left(k_{m,i}\frac{r}{a}\right) e^{-\frac{k_{m,i}^2}{r^2}D(t-t_0)} \tag{2}$$

with $J_m$ the m$^{th}$ Bessel function, $k_{m,i}$ the i$^{th}$ zero of the first derivative of the m$^{th}$ Bessel function defined as $J_m'\left(k_{m,i}\right) = 0$, and if $m \neq 0$ and $i \neq 1$,

$$A_{m,i} = \frac{2J_m\left(k_{m,i}\frac{r_0}{a}\right)}{\pi a^2\left(1+\delta_{0,m}\right)\left(1-\frac{m^2}{k_{m,i}^2}\right)J_m^2\left(k_{m,i}\right)}, \tag{3}$$

else $A_{0,1} = \frac{1}{\pi a^2}$. The case of elliptic geometry leads to the Mathieu's and modified Mathieu's functions [2, 1]. For an ellipse of major axis $a$ and minor axis $b$, the general solution for the transition probability in elliptic coordinates $(\xi,\eta)$ is

$$P\left(\xi,\eta,t|\xi_0,\eta_0,t_0\right) = \sum_{l\in\{c,s\};r,m=0}^{\infty} A_{r,m,l}(\xi_0,\eta_0)\phi_{r,m,l}(\xi,\eta)e^{-\frac{4k_{r,m,l}^{\xi_a}}{f^2}D(t-t_0)}, \tag{4}$$

where

$$\phi_{r,m,c}(\xi,\eta) = Ce_r\left(k_{r,m,c}^{\xi_a},\xi\right)ce_r\left(k_{r,m,c}^{\xi_a},\eta\right),$$
$$\phi_{r,m,s}(\xi,\eta) = Se_{r+1}\left(k_{r,m,s}^{\xi_a},\xi\right)se_{r+1}\left(k_{r,m,s}^{\xi_a},\eta\right),$$

and $f^2 = a^2 - b^2$. $ce_r$ and $se_r$ are the even and the odd r$^{th}$ Mathieu Functions, and $Ce_r$ and $Se_r$ the even and odd r$^{th}$ Modified Mathieu Functions. $k_{r,m,c}^{\xi_a}$ and $k_{r,m,s}^{\xi_a}$ are respectively defined as

$$\frac{dCe_r\left(k_{r,m,c}^{\xi_a},\xi\right)}{d\xi}|_{\xi=\xi_a} = 0\,\text{and}\,\frac{dSe_{r+1}\left(k_{r,m,s}^{\xi_a},\xi\right)}{d\xi}|_{\xi=\xi_a} = 0. \tag{5}$$

From the initial conditions, we can deduce the coefficients $A_{r,m,l}(\xi_0,\eta_0)$,

$$A_{r,m,l}(\xi_0,\eta_0) = \frac{\phi_{r,m,l}(\xi_0,\eta_0)}{\rho_{r,m,l}}, \tag{6}$$

with

$$\rho_{r,m,l} = \int_0^{\xi_a}\int_0^{2\pi} d\xi d\eta\left(\cosh\left(2\xi\right)-\cos\left(2\eta\right)\right)\phi_{r,m,l}^2(\xi,\eta). \tag{7}$$

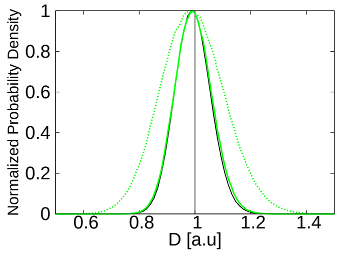



Figure 1: Normalized distributions of MAP estimates of the diffusivity for free Brownian motion, for two different noise levels: $\sigma = 0.3\,a.u$(solid green line) and $\sigma = 1\,a.u.$(dashed green line). The trajectories were generated with a diffusivity $\tilde{D} = 1\,a.u.$, a time step $\Delta t = 1\,a.u.$ and $N = 500$ steps. Thus, these noise levels correspond respectively to a relative uncertainty of 21% and 70% on the position of the particle with respect to the mean 1D displacement $\sqrt{2D\Delta t}$. The distribution of the estimates without noise, common to MAP and MSD estimators, is represented with the solid black line. The mean values of the distributions are equal to the actual value of the diffusivity ($\tilde{D} = 1\,a.u.$) at which a vertical black line is positioned. The MAP estimator thus remains unbiased as the noise on the position of the particle increases. Note that each distribution is normalized so that its maximum value equals one.

# 2 Transition probability in a square domain with a Gaussian noise on the position

In square domains, the transition probability can still be factorized when each coordinate is affected by an independent noise and one can write

$$P(\vec{r}', t|\vec{r}'_0, t_0) = P(x', t|x'_0, t_0) \times P(y', t|y'_0, t_0). \tag{8}$$

As in the case of free diffusion, we write,

$$P(x', t|x'_0, t_0) = \int dx dx_0 P(x'|x) P(x, t|x_0, t_0) P(x'_0|x_0), \tag{9}$$

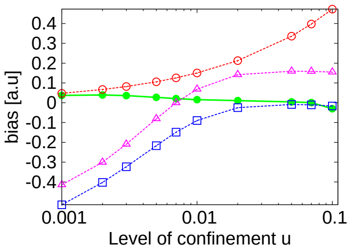
where $P(x'|x)$ is given by

$$P(x'|x) = \frac{1}{(2\pi\sigma^2)^{\frac{d}{2}}} e^{-\frac{(x'-x)^2}{2\sigma^2}}, \tag{10}$$

with $d = 1$. $P(x'_0|x_0)$ is obtained in the same way. Integration should be performed over all the possible values of $x$ and $x_0$. We assume that $P(x'_0|x_0)$ and $P(x'|x)$ are sufficiently localized around positions $x'_0$ and $x'$ respectively so that integration can be extended to $\mathbb{R}^2$, thus yielding the following approximated expression of $P(x', t|x'_0, t_0)$ :

$$P(x', t|x'_0, t_0) = \sum_{k=0}^{\infty} a_k \cos(k\frac{\pi}{L}x'_0) \cos(k\frac{\pi}{L}x') e^{-k^2 \frac{\pi^2}{L^2}(D(t-t_0)+\sigma^2)}. \tag{11}$$

$P(y', t|y'_0, t_0)$ is obtained in the same way and one gets the expression of the transition probability in the presence of a Gaussian noise on the position.



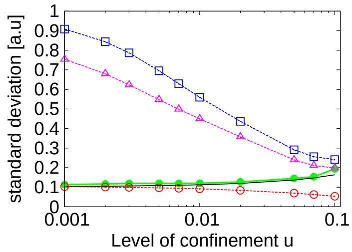




Figure 2: Plots of the bias and standard deviation of estimations of the diffusivity by the $MAP$(green curve), $MSD^{(1)}$(red curve), $MSD^{(2)}$(blue curve) and $MSD^{(3)}$(purple curve) estimators in the case where both the length of the confining domain and the diffusivity are parameters that have to be determined. The solid black line is the Cramér-Rao lower bound for the standard deviation of unbiased estimators.

# 3 Distribution of MAP estimates for free Brownian motion in presence of noise

In the case of free Brownian motion with a Gaussian noise on the position of the particle, the distributions of MAP estimates of the diffusivity remain peaked on the actual value of the diffusivity as the noise level increases as demonstrated for two different noise levels ($\sigma = 0.3\,a.u.$ and $\sigma = 1\,a.u.$) on Fig. 1. The trajectories were generated with a diffusivity $\tilde{D} = 1\,a.u.$, a time step $\Delta t = 1\,a.u.$ and $N = 500$ steps. The MAP estimator remains unbiased as the noise level increases.

# 4 Comparison of MAP and MSD estimators when the domain length $L$ is estimated along with $D$

We consider here the case where both the length of the confining domain and the diffusivity are parameters that have to be determined. The results concerning the estimation of the diffusivity are summarized in Fig. 2. They don't differ significantly from the previous case where the length of the confining domain is considered as known. Thus, for the sake of simplicity, we only focus in the core of the paper on the case where the only parameter to be inferred is the diffusivity.

# 5 Influence of the geometry of the confining domain

The Bayesian inference method makes use of transition probabilities that depend on the geometry of the confining domain. We investigate here the effect of the geometry of the confining domain by generating trajectories inside circular domains and inferring the diffusivity using both the square estimator (the one used for square domains) and the circular estimator (the one using the transition probability derived for circular domains) denoted respectively as $MAP_{square}$ and $MAP_{circle}$. The results are shown in Fig. 3. It appears that the square estimator is rather accurate when the level of confinement is sufficiently low. As the level of confinement increases the effect of the geometry becomes strong and the square estimator is largely biased and has a standard deviation lower that the Cramér-Rao limit. This reflects the fact that the model of diffusion within a square domain is no longer adequate to describe the diffusive motion within a circular domain just as the model of free diffusion was no longer adequate to describe the diffusive motion within a square domain when the level of confinement is high (cf. the behavior of $MSD^{(1)}$ in the core of the paper).

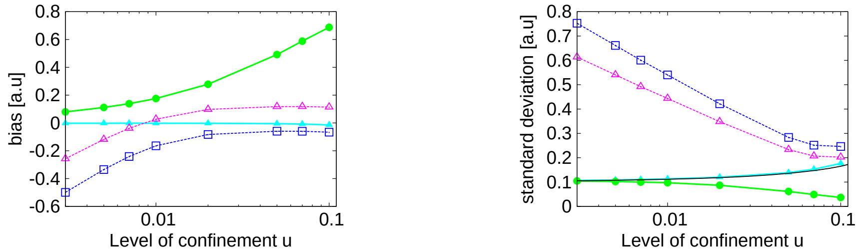



Figure 3: Plots of the bias and standard deviation of the estimations of the diffusivity by the $MAP_{circle}$(light blue curve), $MAP_{square}$(green curve), $MSD^{(1)}$(red curve), $MSD^{(2)}$(blue curve) and $MSD^{(3)}$(purple curve) estimators for 100-time step trajectories generated inside a circular domain. The solid black line is the Cramér-Rao lower bound for the standard deviation of unbiased estimators for the model of diffusion within a circular domain.

The sensitivity of the method to the geometry of the confining domain can be used to compare the accuracy of different models of motion. Indeed, the Bayesian approach allows one to quantitatively compare the validity of models by calculating their evidence from the posterior probability function. The evidence of a model is a well defined object of Bayesian analysis [3]. It is defined as the probability of the considered model given the collected data. Given one trajectory, one can calculate the evidence of a given model, choose the model with the maximum evidence and evaluate the corresponding parameters. In order to compare the validity of models of diffusion within square and circular domains in the case where trajectories are generated inside circular domains, we have calculated for each 100-step realization, the evidence for each model and found that when the square estimator becomes largely biased the evidence for the model of diffusion within a circular domain is almost always (in 98% of the cases for $u = 0.05$) larger than the evidence for the model of diffusion within a square domain. One should then choose to estimate the diffusivity using the maximum evidence estimator.

# References


[1] Abramowitz, M., and I. A. Stegun. 1964. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.

[2] Gutiérrez-Vega, J. C. 2000. Formal analysis of the propagation of invariant optical fields in elliptic coordinates. *Ph. D. Thesis, INAOE, Mexico.*

[3] Mackay, D.J.C. 2003. Information Theory, Inference,and Learning Algorithms. *Cambridge University Press.*