

TEST STATISTIQUE POUR DÉTECTER LES DIFFUSIONS NON BROWNIENNES

Vincent Briane ^{1 2} Charles Kervrann ¹ & Myriam Vimond ²

¹ *Inria, Centre Rennes – Bretagne Atlantique, Campus Universitaire de Beaulieu, 35042 Rennes Cedex, France*

² *CREST-ENSAI, ENSAI, Ker-Lann, 35172 Bruz, France*

Résumé. La modélisation de la dynamique intracellulaire permet de répondre à des questions fondamentales en biologie. Dans ce travail, nous supposons que le mouvement des molécules/protéines est décrit à l'aide de processus stochastiques particuliers: les processus de diffusion. Nous développons ici un test statistique permettant de classer les trajectoires observées en trois groupes distincts: la diffusion confinée, dirigée et libre (ou mouvement Brownien). Cette méthode est une alternative à l'analyse du déplacement carré moyen ("*Mean Square Displacement*") utilisée bien souvent en biophysique. Notre procédure est évaluée sur des données simulées et sur des séquences d'images réelles de microscopie par fluorescence.

Mots-clés. Test statistique, diffusion, mouvement Brownien, déplacement carré moyen, classification de trajectoires.

Abstract. Assessing the dynamics of particles inside live cell is of paramount interest to understand cell mechanisms. In this paper, we assume that the motions of particles follow a certain class of random process: the diffusion processes. Our contribution is to propose a statistical method able to classify the motion of the observed trajectories into three groups: confined, directed and free diffusion (namely Brownian motion). This method is an alternative to Mean Square Displacement (MSD) analysis. We assess our procedure on both simulations and real cases.

Keywords. statistical test, diffusion, Brownian motion, mean square displacement, trajectory classification.

1 Introduction

Une cellule est composée d'organites interagissant les unes avec les autres. Elles échangent de la matière organique, comme des protéines, en la libérant directement dans le cytoplasme ou en l'envoyant via des vésicules sur des réseaux de communication intracellulaires (filaments d'actine, microtubules). L'étude des particules mobiles à l'intérieur de la cellule est d'un grand intérêt biologique. Par exemple la modélisation du mouvement de virus permet de calculer le temps que met le virus à atteindre le noyau cellulaire où il va se répliquer [7]. Par ailleurs, l'environnement intracellulaire est soumis à

de nombreuses perturbations, ainsi les particules qui s’y déplacent ont un mouvement erratique. Un tel mouvement résulte des nombreux chocs entre particules mais aussi d’autres types d’interactions avec l’environnement intracellulaire. Pour représenter ces dynamiques, nous modélisons les trajectoires des particules par une certaine classe de processus stochastiques: les processus de diffusion. Il s’agit de processus stochastiques définis sur un intervalle de temps continu et possédant des trajectoires continues. Notre objectif est de classer les trajectoires selon trois catégories:

- *diffusion dirigée*: la particule est transporté activement le long des réseaux intracellulaires par des moteurs moléculaires [4] ;
- *diffusion libre ou mouvement Brownien*: la particule évolue librement dans le cytoplasme[4] ;
- *diffusion confinée*: la particule se déplace dans un petit domaine ou évolue dans une région encombrée par des molécules [4, 2].

En biophysique [11], pour classer les types de diffusion mentionnées ci-dessus, on recourt au déplacement carré moyen (MSD "*Mean Square Displacement*")

$$\text{MSD}(t) = \mathbb{E}(\|X_t - X_0\|_2)^2 \tag{1}$$

en notant $\mathbb{E}(\cdot)$ l’espérance mathématique, $\|\cdot\|_2$ la norme Euclidienne et t le temps. La mesure du MSD est estimée et ajustée à la fonction $t \rightarrow t^\beta$ avec une méthode des moindres carrés. On choisit le type de mouvement selon certaines règles empiriques basées sur la valeur de β [5, 8]. Cependant les estimations des courbes du MSD sont très variables dès que t augmente [9, 10]. Ainsi l’estimation de β est peu sûre, ce qui implique aussi une classification du type de mouvement peu fiable. Nous proposons donc une procédure de test statistique permettant de classer les trajectoires des particules selon les trois types de diffusion mentionnés plus haut. Les processus qui nous intéressent seront présentés en section 2. Dans la section 3 nous présenterons notre procédure de test. Dans la dernière section nous montrerons des résultats sur des données simulées et sur des séquences d’images réelles de microscopie par fluorescence.

2 Equations différentielles stochastiques

Un processus de diffusion est un processus stochastique Markovien défini sur un intervalle de temps continu et possédant des trajectoires continues [6, chapitre 14]. On peut définir un processus de diffusion comme la solution d’une équation différentielle stochastique (EDS) [6, chapitre 15,section 4]. Heuristiquement, une EDS modélise le mouvement d’une particule dans un fluide soumis à une force déterministe due au fluide et à une force aléatoire issue des collisions (elles-mêmes aléatoires) avec d’autres particules. En dimension 2, le déplacement d’une particule ente t et $t + \Delta t$ s’écrit:

$$X_{t+\Delta t} - X_t \approx \mu(X_t, t)\Delta t + \sigma(X_t, t)(B_{t+\Delta t} - B_t), \tag{2}$$

où $\mu(x, t)$ est un vecteur de taille 2 appelé la dérive et modélisant la force déterministe, $\sigma(x, t)$ est une matrice de taille 2×2 appelée le coefficient de diffusion et modélisant la force aléatoire et B_t est un mouvement brownien 2D. Par définition les composantes de B_t sont des mouvements browniens 1D indépendants. La propriété de Markov provient de l'indépendance des incréments Browniens. Lorsque $\Delta t \rightarrow 0$ on associe l'équation (2) à l'EDS suivante (en notations infinitésimales):

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t. \quad (3)$$

Dans la suite nous ferons l'hypothèse suivante sur le coefficient de diffusion:

Hypothèse 1. *On suppose que le coefficient de diffusion est une matrice diagonale 2×2 constante selon les variables x et t :*

$$\sigma(x, t) = \sigma \mathbf{I}_2 \quad (4)$$

où σ est un réel positif inconnu et \mathbf{I}_2 est la matrice identité en dimension 2.

Nous devons estimer le coefficient de diffusion dans notre procédure. L'invariance en temps et en espace permet de l'estimer facilement. Le choix d'un tel coefficient de diffusion suppose la propriété d'isotropie des trajectoires.

3 Procédure statistique

Notre procédure s'apparente à un test statistique:

H_0 "la trajectoire observée est un mouvement Brownien anisotrope, $X_t = \sigma B_t$ "

versus

H_1 "la trajectoire observée est issue d'une diffusion confinée ou dirigée."

Statistique de test Le mouvement Brownien en 2D est récurrent: avec une probabilité 1, il revient dans le voisinage de n'importe quel point une infinité de fois. C'est pourquoi on l'appelle aussi diffusion libre. Ce n'est ni le cas pour une diffusion confinée (la particule est contrainte à rester dans une petite région), ni pour une diffusion dirigée (la particule est propulsée par un moteur moléculaire dans une certaine direction). Ainsi, nous proposons de discriminer les trois types de diffusion à l'aide de la statistique mesurant l'éloignement maximal du processus à son point d'origine,

$$S_X(t) = \sup_{0 \leq s \leq t} \|X_s - X_0\|_2. \quad (5)$$

Ce critère est simple à interpréter: si $S_X(t)$ est petit, la trajectoire est restée proche de son point de départ ; par contre, si $S_X(t)$ est grand, la trajectoire s'est beaucoup éloignée de sa position initiale. Sous l'hypothèse nulle, on connaît la distribution de probabilité de $S_X(t)$ [3, formule 1.1.4,p280]. Dans le cas où σ est connu (par exemple lorsqu'on observe le processus de manière continue dans le temps [1, lemme 4.2,p212]) on peut montrer que la variable:

$$T = S_X(t)/(\sigma\sqrt{t}) \quad (6)$$

est une statistique dont la loi est indépendante de (σ, t) . On définit alors la région d'acceptation de H_0 au niveau $\alpha \in]0, 1[$ par:

$$\left\{ q\left(\frac{\alpha}{2}\right) \leq T \leq q\left(1 - \frac{\alpha}{2}\right) \right\}, \quad (7)$$

où $q(x)$ désigne le quantile d'ordre x de (6) sous H_0 , c'est à dire dans le cas Brownien. Finalement nous adoptons la règle de décision suivante pour classer les trajectoires selon le type de diffusion:

- si T est dans la région (7) on déclare que $(X_t)_{t>0}$ est un mouvement Brownien ;
- si $T < q\left(\frac{\alpha}{2}\right)$ on déclare que $(X_t)_{t>0}$ est une diffusion confinée ;
- si $T > q\left(1 - \frac{\alpha}{2}\right)$ on déclare que $(X_t)_{t>0}$ est une diffusion dirigée.

Cependant le paramètre σ est inconnu dans notre modèle. Il faut donc l'estimer selon la procédure décrite ci-après.

Implémentation On suppose que l'on observe $(X_t)_{t>0}$ de manière discrète à intervalle de temps Δt . On observe donc le vecteur $(X_0, X_{\Delta t}, \dots, X_{n\Delta t})$. Dans ce contexte, on redéfinit $S_X(t)$ de la manière suivante: $S_X^n(n\Delta t) = \max_{i=0, \dots, n} \|X_{i\Delta t} - X_0\|_2$. La statistique de test en temps discret est analogue à (6). Avec l'hypothèse 1 sur le coefficient de diffusion, on peut estimer σ^2 par:

$$\hat{\sigma}^2 = \frac{1}{2n\Delta t} \sum_{j=1}^n \|X_{j\Delta t} - X_{(j-1)\Delta t}\|_2^2. \quad (8)$$

D'après [1, Lemma 4.2, p 212], cet estimateur converge presque sûrement vers σ^2 sous H_0 et sous H_1 . La statistique de test devient alors:

$$T_n = \frac{S_X^n(n\Delta t)}{\sqrt{n\Delta t\hat{\sigma}^2}}. \quad (9)$$

Les quantiles $q_n(x)$ de la loi de T_n dépendent de la taille de la trajectoire n puisque S_X^n et $\hat{\sigma}^2$ dépendent de n . On estime ces nouveaux quantiles $q_n(x)$ à l'aide de simulations de Monte-Carlo. On simule un grand nombre de T_n sous H_0 puis on estime les quantiles $q_n(x)$ par les quantiles empiriques.

4 Etudes de Monte Carlo et cas réels

Nous évaluons notre procédure de test en simulant deux processus paramétriques illustrant chaque type de diffusion. Pour illustrer la diffusion dirigée nous utiliserons le mouvement Brownien dirigé:

$$dX_t^i = v_i dt + \sigma dB_t^i \quad i = 1, 2, \quad (10)$$

où v_i est la dérive constante. Pour illustrer la diffusion confinée nous utiliserons le processus Ornstein-Uhlenbeck [6, 2,]:

$$dX_t^i = -\lambda(X_t - m_i)dt + \sigma dB_t^i \quad i = 1, 2, \quad (11)$$

Label	BR	BRD	OU	BR	BRD	OU	IMM	NC
Vérité terrain	Notre test			Méthode du MSD				
BR	94.3	3.0	2.7	20.0	26.9	51.9	0.9	0.3
BRD	12.7	87.3	0	1.4	97.9	0.7	0	0
OU	20.9	0	79.1	0	0	47.8	31.2	21

Table 1: Matrices de confusion de notre test au niveau 5% et de la méthode du MSD (résultat exprimé en %). BR,OU,BRD,IMM,NC désignent respectivement Brownien Ornstein-Uhlenbeck , Brownien dirigé, immobile et non classée. On lit 3.0% des trajectoires Browniennes simulées sont détectées comme confiné (Ornstein-Uhlenbeck) par notre test. Dans la catégorie NC on a mis les trajectoires pour lesquelles $\beta < 0$ ce qui n'est pas considéré dans [5].

où $\lambda > 0$ modélise une force de rappel vers la moyenne à long terme $m = (m_1, m_2)^T$. On simule 1 000 trajectoires de taille 30 pour chaque type de processus (Brownien, Ornstein-Uhlenbeck et Brownien dirigé) avec $\Delta t = 1$ et $\sigma = 1$. Pour le Brownien dirigé on choisit $v_i = 0.5$, pour le processus Ornstein-Uhlenbeck on choisit $m = (0, 0)^T$ et $\lambda = 0.7$. On constate que la méthode du MSD confond largement le mouvement Brownien avec les deux autres types de diffusion (ligne 1 colonne MSD du tableau 1). Notre test contrôle cette erreur puisqu'il s'agit de l'erreur de première espèce pour ce test. Il la contrôle donc au niveau $\alpha\%$. On détecte 94.3% de trajectoires browniennes parmi celles qui sont réellement browniennes ce qui est attendu pour un test de niveau 5%. Pour ce choix de paramètre on a une bonne puissance pour notre test: 87.3% pour le Brownien dirigé 79.1% pour le processus Ornstein-Uhlenbeck . Par ailleurs, notre méthode a été validée sur des séquences d'images fluorescentes (microscopie TIRF) montrant le trafic de la protéine Langerin (figure 1). Sur cette séquence comprenant 1 600 trajectoires (pas toutes représentée sur la figure 1), d'après notre test 68% des trajectoires sont browniennes, 27% des trajectoires sont confinées et 5% des trajectoires sont dirigées. Avec la méthode du MSD, on détecte sur cette séquence 12% de trajectoires browniennes, 66% de diffusion confinée, 14% de diffusion dirigée et 8% de particules immobiles. Sur les séquences réelles comme sur les données simulées, on obtient des résultats très différents entre les deux méthodes. Au vu des résultats sur données simulées, notre méthode est plus pertinente que la méthode du MSD.

5 Conclusion

Nous avons proposé une procédure de test pour classer les différents types de diffusion des particules intracellulaires qui est statistiquement consistant. Il s'agit d'une alternative à la méthode du MSD qui donne de meilleurs résultats d'après notre étude de Monte-Carlo. La méthode a été évaluée sur des séquences simulées et réelles en microscopie par fluorescence.

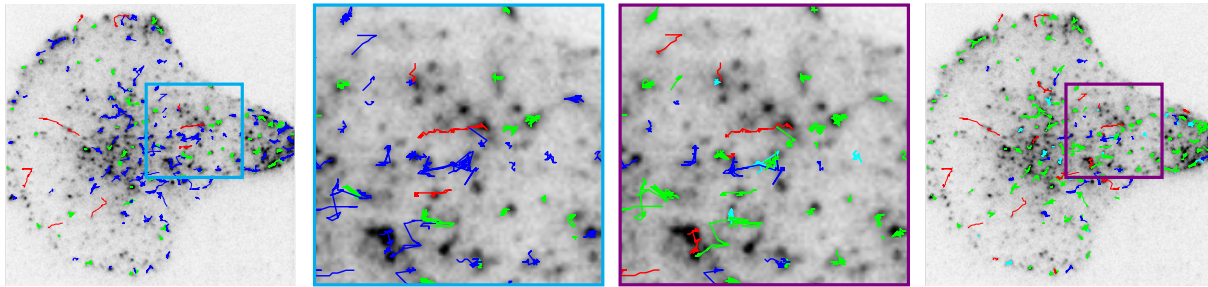


Figure 1: Classification des types de diffusion sur une séquence de la protéine Langerin avec notre test au niveau 5% (double image à gauche) et la méthode du MSD (double image à droite). On code en rouge la diffusion dirigée, en vert la diffusion confinée, en bleu le mouvement Brownien et en bleu clair l'immobilité (catégorie "immobilité" définie seulement pour la méthode du MSD) (remerciement à l'UMR 144 CNRS Institut Curie - PICT IBiSA).

Bibliographie

- [1] I.V. BASAWA et B.L.S. PRAKASA RAO : *Statistical Inferences for Stochastic Processes*. Academic, 1980.
- [2] H. BERRY et H. CHATÉ : Anomalous diffusion due to hindering by mobile obstacles undergoing brownian motion or ornstein-ulhenbeck processes. *Phys. Rev. E*, 89(2):022708, 2014.
- [3] A.N. BORODIN et P. SALMINEN : *Handbook of Brownian Motion-Facts and Formulae*. Birkhäuser, 1996.
- [4] P.C. BRESSLOFF et J.M. NEWBY : Stochastic models of intracellular transport. *Reviews of Modern Physics*, 85(1):135, 2013.
- [5] T.J. FEDER *et al.* : Constrained diffusion or immobile fraction on cell surfaces: a new interpretation. *Biophysical J.*, 70(6):2767–2773, 1996.
- [6] S. KARLIN : *A Second Course in Stochastic Processes*. Academic, 1981.
- [7] T. LAGACHE, E. DAUTY et David HOLCMAN : Quantitative analysis of virus and plasmid trafficking in cells. *Physical Review E*, 79(1):011921, 2009.
- [8] F.W. LUND *et al.* : Spatrack: An imaging toolbox for analysis of vesicle motility and distribution in living cells. *Traffic*, 15(12):1406–1429, 2014.
- [9] X. MICHALET : Mean square displacement analysis of single-particle trajectories with localization error: Brownian motion in an isotropic medium. *Phys. Rev. E*, 82(4):041914, 2010.
- [10] H. QIAN, M.P. SHEETZ et E.L. ELSON : Single particle tracking. analysis of diffusion and flow in two-dimensional systems. *Biophysical J.*, 60(4):910, 1991.
- [11] Michael J SAXTON et Ken JACOBSON : Single-particle tracking: applications to membrane dynamics. *Annual review of biophysics and biomolecular structure*, 26(1):373–399, 1997.