

**SUPPLEMENTARY MATERIALS : A STATISTICAL
ANALYSIS OF PARTICLE TRAJECTORIES IN LIVING
CELLS**

BY VINCENT BRIANE^{*‡}, CHARLES KERVRANN^{*} AND MYRIAM VIMOND[‡]

INRIA, Centre de Rennes Bretagne Atlantique, Serpico Project-Team^{}
CREST (Ensaï, Université Bretagne Loire)[‡]*

1. Introduction. In this supplementary materials, we first recall briefly the quantities of interest and the models of the article (see Section 2). In Section 3, we give the proofs of Theorem 3.1, Proposition 1 and 2 of the article and some results which justify the choice of the parameters in the power analysis in Sections 5.1 and 5.2. Finally in Section 4, we give some supplementary figures on both simulations and real data.

2. Context. The test statistic of our different test procedures is given by :

$$(2.1) \quad T_n = \frac{\max_{i=1, \dots, n} \|X_{t_i} - X_{t_0}\|_2}{\sqrt{(t_n - t_0)\hat{\sigma}_n^2}}$$

In the sequel, we use the following estimate of the diffusion coefficient :

$$(2.2) \quad \hat{\sigma}_{1,n}^2 = \frac{1}{2n\Delta} \sum_{j=1}^n \|X_{t_j} - X_{t_{j-1}}\|_2^2$$

First we study the asymptotic of the test statistic (2.1) under the null hypothesis of Brownian motion (Theorem 3.1). Then we analyse the asymptotic of $\hat{\sigma}_{1,n}^2$ under the following parametric diffusion processes (Proposition 1) :

$$(2.3) \quad dX_t^i = -\lambda(X_t^i - \theta_i)dt + \sigma dB_t^{1/2,i},$$

$$(2.4) \quad dX_t^i = \sigma dB_t^{h,i},$$

$$(2.5) \quad dX_t^i = v_i dt + \sigma dB_t^{1/2,i},$$

with $i = 1, 2$ such that $X_t = (X_t^1, X_t^2)$.

We study the asymptotic of the test statistic (2.1) under the processes (2.3), (2.4) and (2.5). Finally we analyse the dependence over the parameters of the test statistic (2.1) in the purpose of carrying the power analysis of our procedure (Sections 5.1 and 5.2).

3. Proofs.

PROOF OF THEOREM 3.1. Under the null hypothesis, $X_t/\sigma = B_t$ is a standard Brownian Motion. Let us introduce the following random variable,

$$(3.1) \quad \tilde{T}_n = \max_{k=1\dots n} \left\| \frac{1}{\sqrt{n}} R_k \right\|_2,$$

where $R_k = \sum_{j=1}^k (B_{j\Delta} - B_{(j-1)\Delta})/\sqrt{\Delta}$. Since $\hat{\sigma}_n$ is a consistent estimator of σ and using the Slutsky Lemma, it remains to prove that \tilde{T}_n converges in distribution to S_0 . Using the fact that the increments of the Brownian process are independent and Gaussian, R_k is the sum of j independent identically $\mathcal{N}(0, 1)$ -distributed random variables. We define the following process,

$$W_t^{(n)} = \frac{1}{\sqrt{n}} R_{[nt]}, \quad t \in [0, 1],$$

where $[x]$ denotes the integer part of $x \in \mathbb{R}$. Then we get:

$$(3.2) \quad \tilde{T}_n = \sup_{t \in [0, 1]} \left\| W_t^{(n)} \right\|_2.$$

Due to Donsker's Theorem (Billingsley, 2013, Theorem 8.2), $(W_t^{(n)})$ converges in distribution to the Wiener measure as $n \rightarrow \infty$ over the space of continuous function on $[0, 1]$. Since $x \rightarrow \sup_{t \in [0, 1]} \|x(t)\|$ is a continuous function on the space of continuous functions from $[0, 1]$ to \mathbb{R} , \tilde{T}_n converges in distribution to S_0 . \square

3.1. *Proof of Proposition 1 : the convergence of the estimator (2.2) of the diffusion coefficient.* Notice that $\hat{\sigma}_n = \hat{\sigma}_{1,n}$ is strongly consistent under the null hypothesis due to the strong law of large numbers and the independence of the increments of the Brownian motion.

We focus now on the three alternatives. According to the alternative, we denote by \mathbb{E} the expectation associated to the measure P of the solution of the related SDE ((2.4) or (2.3) or (2.5)).

BROWNIAN WITH DRIFT. We may rewrite the strong solution of the SDE (2.5) as,

$$X_{t_k} = X_{t_{k-1}} + v\Delta + \sigma\sqrt{\Delta}\epsilon_k, \quad k = 1 \dots n,$$

where $\sqrt{\Delta}\epsilon_k = B_{t_k} - B_{t_{k-1}}$, and (B_t) is a standard Brownian motion. Then the random variables $Z_k = \|v\Delta + \sigma\sqrt{\Delta}\epsilon_k\|^2$, $k = 1 \dots n$, are positive independent identically distributed random variables, and admit a moment of order 1,

$$\mathbb{E}(Z_k) = \Delta^2 \|v\|^2 + 2\Delta\sigma^2.$$

Then according to the strong law of large numbers, $\hat{\sigma}_n$ converges almost surely to $\Delta\|v\|^2/2 + \sigma^2$. \square

ORNSTEIN-UHLENBECK PROCESS. Let (X_t) be an Ornstein-Uhlenbeck process (2.3). The SDE (2.3) admits a unique solution (Bressloff, 2014, Section 2.2.3)

$$(3.3) \quad X_t - X_s = (X_s - \theta)(e^{-\lambda(t-s)} - 1) + \sigma \int_s^t e^{-\lambda(t-u)} dB_u^{1/2}.$$

Then (X_t) is a stationary Gaussian process where transition density $p(s, x, t, y)$ is the density of

$$\mathcal{N}\left(x + (x - \theta)(e^{-\lambda(t-s)} - 1), \sigma^2(1 - e^{-2\lambda(t-s)})/(2\lambda)\mathbf{I}_2\right).$$

Then we get that,

$$\begin{aligned} \mathbb{E}(\|X_{t+\Delta} - X_t\|^2 \mid X_t = x) &= \int \|x - y\|^2 p(t, x, t + \Delta, y) dy, \\ &= \|x - \theta\|^2(e^{-\lambda\Delta} - 1)^2 + \sigma^2(1 - e^{-2\lambda\Delta})/\lambda. \end{aligned}$$

Moreover the density μ of the stationary distribution of (X_t) is the Gaussian variable $\mathcal{N}(\theta, (\sigma^2\mathbf{I}_d)/(2\lambda))$. Then we obtain that,

$$\begin{aligned} \mathbb{E}(\|X_{t+\Delta} - X_t\|^2) &= \int \mathbb{E}(\|X_{t+\Delta} - X_t\|^2 \mid X_t = x)\mu(x)dx, \\ &= \sigma^2(e^{-\lambda\Delta} - 1)^2/\lambda + \sigma^2(1 - e^{-2\lambda\Delta})/\lambda, \\ &= 2\sigma^2(1 - e^{-\lambda\Delta})/\lambda. \end{aligned}$$

Now, according to Bibby and Sørensen (1995, Lemma 3.1), if (X_t) is a stationary diffusion, $\hat{\sigma}_n^2$ converges in probability to $\mathbb{E}(\|X_{t+\Delta} - X_t\|^2)/(2\Delta)$. We deduce the result. \square

FRACTIONAL BROWNIAN MOTION. Let (X_t) be a fractional Brownian motion (2.4). Due to the self-similarity property and the stationary increments of the fractional Brownian motion, the following process,

$$W_t^{(n)} = \frac{X_{t_0+n\Delta t} - X_{t_0}}{(n\Delta)^h\sigma}, \quad t \in [0, 1],$$

is a standard fractional Brownian motion. The statistic associated to the quadratic variation of the process $(W_t^{(n)})$ may be defined as,

$$\begin{aligned} V_n &= \frac{1}{n} \sum_{i=1}^n \frac{\|W_{i/n}^{(n)} - W_{(i-1)/n}^{(n)}\|^2}{\mathbb{E} \|W_{i/n}^{(n)} - W_{(i-1)/n}^{(n)}\|^2} - 1, \\ &= \frac{\hat{\sigma}_n^2}{\sigma^2 \Delta^{2h-1}} - 1. \end{aligned}$$

According to [Coeurjolly \(2001, Proposition 1\)](#), V_n converges almost surely to 0. Then we deduce that $\hat{\sigma}_n^2/\sigma^2$ tends to Δ^{2h-1} almost surely. \square

3.2. Proof of Proposition 2 : the asymptotic behaviour of the test statistic under parametric alternatives. Since the diffusion parameter σ is unknown, the test statistic (2.1) is normalized by an estimator of σ . Proposition 1 states that $\hat{\sigma}_n/\sigma$ converges in probability to a constant. Therefore, it is sufficient to study the asymptotic behaviour of the test statistic as if σ was known. Then, in this subsection, we consider the test statistic T_n as :

$$(3.4) \quad T_n = \frac{\max_{i=1, \dots, n} \|X_{t_i} - X_{t_0}\|_2}{\sigma \sqrt{t_n - t_0}}.$$

BROWNIAN MOTION WITH DRIFT (H_2). The process (X_t) is a Brownian motion with drift (2.5) and may be rewritten as,

$$X_{t_n} - X_{t_0} = v(t_n - t_0) + \sigma(B_{t_n} - B_{t_0}).$$

Using that (B_t) is a Brownian motion, the distribution of $B_{t_n} - B_{t_0}$ is $\mathcal{N}(\mathbf{0}_2, (t_n - t_0)\mathbf{I}_2)$. Then we have :

$$(3.5) \quad \mathbb{E} \left(\left\| \frac{X_{t_n} - X_{t_0}}{\sigma(t_n - t_0)} - \frac{v}{\sigma} \right\|^2 \right) = \frac{2}{t_n - t_0}.$$

As $t_n - t_0 = n\Delta$, we deduce that $V_n = (X_{t_n} - X_{t_0})/(\sigma(t_n - t_0))$ converges in probability to v/σ . As the euclidean norm is a continuous function, the variable $\|V_n\|$ converges in probability to $\|v\|/\sigma > 0$. Then $\sqrt{n\Delta}V_n$ converges in probability to $+\infty$. Since T_n is lower bounded by $\sqrt{n\Delta}V_n = \|(X_{t_n} - X_{t_0})\|/(\sigma\sqrt{t_n - t_0})$, the proof is complete. \square

THE ORNSTEIN-UHLENBECK PROCESS (H_1). The process (X_t) is an Ornstein-Uhlenbeck process (2.3). We assume that the process is in its stationary regime, that means X_{t_0} is drawn from the stationary distribution that is

$X_{t_0} \sim \mathcal{N}(\theta, \sigma^2/(2\lambda)\mathbf{I}_2)$. The SDE (2.3) admits an unique solution (Bressloff, 2014, Section 2.2.3)

$$(3.6) \quad X_t - \theta = (X_{t_0} - \theta)e^{-\lambda(t-t_0)} + \sigma \int_{t_0}^t e^{-\lambda(t-u)} dB_u^{1/2}.$$

Then we may bound the test statistic T_n by,

$$\frac{\|X_{t_0} - \theta\|}{\sigma\sqrt{n\Delta}} + \sum_{i=1}^2 \max_{k=1\dots n} \frac{|X_{t_k}^i - \theta_i|}{\sigma\sqrt{n\Delta}}.$$

Since X_{t_0} is drawn from the stationary distribution, the term $\|X_{t_0} - \theta\|/\sqrt{n\Delta}$ converges in probability to zero.

Now we show that the second term in the previous equation tends to zero in probability as well. We introduce the variables (ξ_k^1, ξ_k^2) defined as,

$$\xi_k^i = (X_{t_k}^i - \theta_i)\sqrt{2\lambda}/\sigma, \quad k = 1 \dots n, \quad i = 1, 2.$$

Then for $i = 1, 2$, the sequence $(\xi_k^i)_k$ is a standardized stationary normal sequence with covariance function,

$$r_k = \mathbb{E}(\xi_\ell^i \xi_{\ell+k}^i) = e^{-k\Delta}, \quad k \geq -\ell.$$

Let i be in $\{1, 2\}$. Then $(a_n(\max_{k=1\dots n}(\xi_k^i) - b_n))_n$ converges in distribution according to (Leadbetter, Lindgren and Rootzén, 1983, Theorem 4.3.3), where $a_n = \sqrt{2\log(n)}$ and $b_n = a_n - (2a_n)^{-1}(\log \log(n) + \log(4\pi))$. We deduce that $\max_{k=1\dots n}(\xi_k^i)/\sqrt{n\Delta}$ converges in probability to 0. Moreover, since $(\xi_k^i)_k$ is a centred Gaussian process, then $\max_{k=1\dots n}(-\xi_k^i)/\sqrt{n\Delta}$ converges in probability to 0 by symmetry. Then we conclude that $\max_{k=1\dots n} |X_{t_k}^i - \theta_i|/\sqrt{n\Delta}$ converges in probability to 0. \square

THE FRACTIONAL BROWNIAN MOTION (H_1). The process (X_t) is a fractional Brownian motion with $\mathfrak{h} \in (0, 1/2)$. From the property of self-similarity and stationarity of increments of the fractional Brownian motion, the following process,

$$(3.7) \quad Z_t^{(n)} = \frac{X_{tn\Delta+t_0} - X_{t_0}}{\sigma(n\Delta)^\mathfrak{h}}, \quad t \in [0, 1],$$

is a fractional Brownian motion. We rewrite the test statistic as,

$$T_n = \frac{1}{(n\Delta)^{1/2-\mathfrak{h}}} \max_{k=1\dots n} \|Z_{k/n}^{(n)}\|$$

Then T_n is bounded by,

$$\frac{1}{(n\Delta)^{1/2-h}} \sum_{i=1}^2 \max_{k=1\dots n} \left| Z_{k/n}^{i,(n)} \right|,$$

where $Z_t^{(n)} = (Z_t^{1,(n)}, Z_t^{2,(n)})$. The process $Z^{(n)}$ has a version with continuous path as a result of being γ -Holder continuous for any $\gamma < \mathfrak{h}$. Let $i \in \{1, 2\}$ be fixed. Then the random variable $\max_{k=1\dots n} \left| Z_{k/n}^{i,(n)} \right|$ is bounded by,

$$M_i^{(n)} = \sup_{t \in [0,1]} \left| Z_t^{i,(n)} \right|,$$

which possesses an absolutely continuous density on \mathbb{R}_+^* according to [Zaïdi et al. \(2003\)](#). That means the sequence $\left(\max_{k=1\dots n} \left\| Z_{k/n}^{(n)} \right\| \right)_n$ is tight. Since $\mathfrak{h} < 1/2$, we deduce that T_n converges in probability to 0. \square

THE FRACTIONAL BROWNIAN MOTION (H_2). The process (X_t) is a fractional Brownian motion with $\mathfrak{h} \in (1/2, 1)$. From the property of self-similarity we get that:

$$(3.8) \quad Y_n = \frac{\|X_{t_n} - X_{t_0}\|_2^2}{\sigma^2(t - t_0)^{2\mathfrak{h}}} \sim \chi^2(2).$$

We observe that $T_n^2 \geq Y_n(n\Delta)^{2\mathfrak{h}-1}$. Let x be a positive constant. We have :

$$(3.9) \quad \begin{aligned} P(T_n < x) &\leq P\left(Y_n(n\Delta)^{2\mathfrak{h}-1} < x^2\right) \\ &\leq P\left(Y_n < x^2/(n\Delta)^{2\mathfrak{h}-1}\right). \end{aligned}$$

Since $\mathfrak{h} > 1/2$, $x^2/(n\Delta)^{2\mathfrak{h}-1}$ converges to 0 as $n \rightarrow \infty$. Then the right hand side of (3.9) converges to 0. That means $P(T_n < x)$ converges to 0 as $n \rightarrow \infty$: T_n converges to $+\infty$ in probability. \square

3.3. *Dependency of the power on the parameters of the parametric alternatives.*

LEMMA 3.1. *Let (X_t) be a Brownian motion with drift (2.5). Let $\hat{\sigma}_n$ be the estimator of the diffusion coefficient defined in Equation (2.2). The distribution of T_n (2.1) depends only on the parameter $v\sqrt{\Delta}/\sigma$ and the trajectory size n .*

PROOF OF LEMMA 3.1. We may rewrite the strong solution of the SDE (2.5) as,

$$X_{t_k} = X_{t_{k-1}} + v\Delta + \sigma\sqrt{\Delta}\epsilon_k, \quad k = 1 \dots n,$$

where $\sqrt{\Delta}\epsilon_k = B_{t_k} - B_{t_{k-1}}$, and (B_t) is a standard Brownian motion. Then (ϵ_k) is a sequence of independent Gaussian variables $\mathcal{N}(0, 1)$. Furthermore, we have immediately :

$$X_{t_k} - X_{t_0} = vk\Delta + \sigma\sqrt{\Delta} \sum_{i=1}^k \epsilon_i, \quad k = 1 \dots n.$$

Finally the test statistic T_n may be rewritten as,

$$T_n = \frac{\max_{k=1, \dots, n} \left\| k \frac{v\sqrt{\Delta}}{\sigma} + \sum_{i=1}^k \epsilon_i \right\|}{\sqrt{\frac{1}{2} \sum_{i=1}^n \left\| \frac{v\sqrt{\Delta}}{\sigma} + \epsilon_i \right\|^2}}.$$

As the distribution of (ϵ_k) is free of the parameters the distribution of T_n depends only on $v\sqrt{\Delta}/\sigma$. \square

LEMMA 3.2. Let (X_t) be a fractional Brownian motion (2.4). Let $\hat{\sigma}_n$ be the estimator of the diffusion coefficient defined in Equation (2.2). The distribution of T_n (2.1) depends only on the parameter \mathfrak{h} and the trajectory size n .

PROOF OF LEMMA 3.2. The fractional Brownian motion may be described by its incremental process Taqqu (2003) :

$$(3.10) \quad \epsilon_k = (X_{t_k} - X_{t_{k-1}})/(\sigma\Delta^{\mathfrak{h}}), \quad k \geq 1,$$

where (ϵ_k) is a fractional Gaussian noise which is a stationary standardized Gaussian process with autocovariance function $\mathbb{E}(\epsilon_k\epsilon_{k+i}) = (1/2)(|i+1|^{2\mathfrak{h}} - 2|i|^{2\mathfrak{h}} + |i-1|^{2\mathfrak{h}})$. Finally the test statistic T_n may be rewritten as,

$$T_n = \frac{\max_{k=1, \dots, n} \left\| \sum_{i=1}^k \epsilon_i \right\|}{\sqrt{\frac{1}{2} \sum_{i=1}^n \|\epsilon_i\|^2}}.$$

Then the distribution of T_n depends only on the trajectory size n and on \mathfrak{h} through the distribution of (ϵ_k) . \square

4. Supplementary figures.

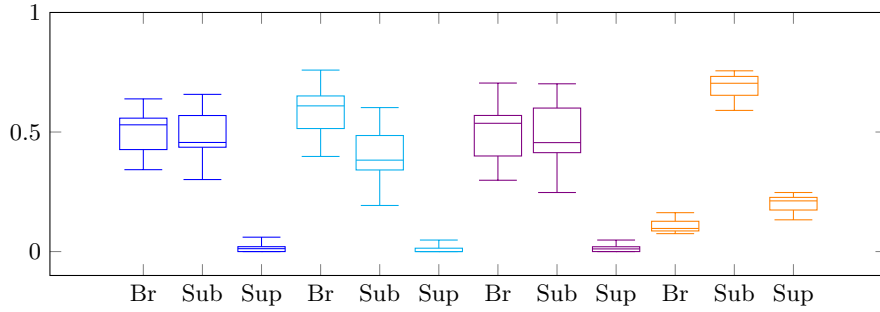


FIG 1. *Boxplots of the proportions of Brownian, subdiffusion and superdiffusion computed from 12 Rab11a sequences. In blue proportions obtained with the single test procedure, in cyan with the Procedure 1, in violet with the adaptive Procedure 1 and in orange with the MSD method. Br stand for Brownian, Sub for subdiffusion and Sup for superdiffusion.*

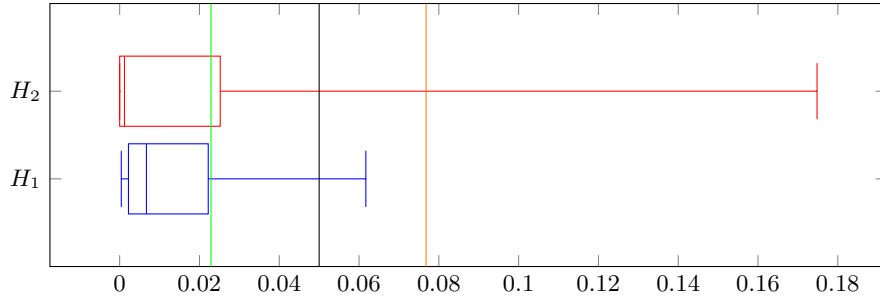


FIG 2. *Boxplots of the p-value p_{30} (Equation 3.12 of the article) under H_1 and H_2 . We simulate a set of trajectories \mathcal{X}_m with $m = 100$ and $m_0 = 20$ according to the simulation scheme described in Section 5.2 of the article. We plot the boxplot of the p-values $p_{30}^{(i;m)}$ corresponding to each true alternative hypotheses H_1 and H_2 . The green (respectively orange) line is the threshold $h = p^{(k^*)}$ obtained by the first step of Procedure 1 (respectively Procedure 1). The null hypothesis is rejected if the p-value is lower than h . The black line is the level $\alpha = 5\%$.*

References.

- BIBBY, B. M. and SØRENSEN, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* 17–39.
- BILLINGSLEY, P. (2013). *Convergence of probability measures*. John Wiley & Sons.
- BRESSLOFF, P. C. (2014). *Stochastic Processes in Cell Biology* 41. Springer.
- COEURJOLLY, J.-F. (2001). Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Statistical Inference for stochastic processes* 4 199–227.
- LEADBETTER, M. R., LINDGREN, G. and ROOTZÉN, H. (1983). *Extremes and related properties of random sequences and processes*. Springer.
- TAQQU, M. S. (2003). Fractional Brownian motion and long-range dependence. *Theory and applications of long-range dependence* 5–38.

ZAIÏDI, N. L., NUALART, D. et al. (2003). Smoothness of the law of the supremum of the fractional Brownian motion. *Elect. Comm. Probab* **8** 102–111.

INRIA RENNES - BRETAGNE ATLANTIQUE
CAMPUS UNIVERSITAIRE DE BEAULIEU
35042 RENNES CEDEX - FRANCE
E-MAIL: vincent.briane@inria.fr
charles.kervrann@inria.fr

ENSAI - CAMPUS DE KER-LANN
RUE BLAISE PASCAL - BP 37203
35712 BRUZ CEDEX - FRANCE
E-MAIL: myriam.vimond@ensai.fr