Statistica Sinica **31** (2021), 29-51
doi:https://doi.org/10.5705/ss.202018.0211

LEVEL SET AND DRIFT ESTIMATION FROM A REFLECTED BROWNIAN MOTION WITH DRIFT

Alejandro Cholaquidis, Ricardo Fraiman, Ernesto Mordecki
and Cecilia Papalardo

CABIDA, Universidad de la República

Abstract: We estimate the drift and the level sets of the stationary distribution of a Brownian motion with drift, reflected in the boundary of a compact set $S \subset \mathbb{R}^d$, departing from the observation of a trajectory of this process. We obtain the uniform consistency and rates of convergence for the proposed kernel-based estimators. This problem has relevant applications in ecology, for example, when estimating the home range and the core area of an animal based on tracking data. Recent attempts to estimate the domain of a reflected Brownian motion have considered a uniform stationary distribution; however in this case the estimation of the core area, defined as a level set of the stationary distribution, is meaningless. We also give an estimator of the drift function, based on the increments of the process. In order to prove our results, we obtained several new theoretical properties of the reflected Brownian motion with drift, under fairly general assumptions. These properties allow us to perform the estimation for flexible regions close to reality. Lastly, the theoretical findings are illustrated using simulated and real-data examples.

Key words and phrases: Core-area, drift estimation, home-range estimation, reflected Brownian motion with drift, stationary distribution.

1. Introduction

Given a reflected Brownian motion with drift (RMBD) inside a (smooth enough) compact domain S, we consider three statistical problems: 1) estimating the density of the stationary distribution; 2) estimating the level sets of this density, with and without shape restrictions; and 3) estimating the drift function. The practical motivation for these problems is made explicit in the following paragraphs.

Level-set estimation falls within the field of nonparametric set estimation, where the goal is to reconstruct (in the statistical sense) an unknown set S from random data related to S. Usually, such random information comes from a

Corresponding author: Alejandro Cholaquidis, Iguá 4225, piso 16. 11400, Montevideo, Uruguay. E-mail: acholaquidis@cmat.edu.uy.

sample of independent points drawn from an absolutely continuous distribution with density f. In addition, the target set is either the support of f or a level set of the $\{x : f(x) > \lambda\}$ (which, depending on λ, can be viewed as a sort of "substantial support" of the underlying distribution).

There are, however, two important practical applications of set estimation techniques in which the assumption of independence is clearly unsuitable; in this case, the above-mentioned RBMD approach might be particularly useful. The first application estimates a so-called home range (the region where an individual of an animal species develops its activities; Burt (1943)), and the second estimates a core area (the sub-region of the home range where the individual spends most of its time; see Hayne (1949), Worton (1987)). Recent advances in animal tracking technology allow an almost continuous record of the movement. Therefore, in both cases, it might be reasonable to assume that the sample information comes from a grid of points along the (random) trajectory followed by the animal during its activities. In our setup, core areas can be modeled by the level sets of the stationary distribution, whereas the drift function provides information about the dynamics of the movement of the animal.

This perspective was followed by Cholaquidis et al. (2016), where the home range S is identified with the support of the stationary distribution of a reflected Brownian motion (with no drift), and a suitable set estimator is proposed and analyzed. However, under the (quite natural) regularity conditions on S considered in the aforementioned study, the stationary distribution of the reflected Brownian motion is necessarily uniform (Burdzy, Chen and Marshall (2006)), which can be somewhat restrictive. In particular, the problem of estimating a core area (which should be addressed in terms of a level-set estimation) becomes meaningless or uninteresting.

Thus, the main contribution of this study is to extend the approach of Cholaquidis et al. (2016) to the case of an RBMD. Such an extension is beyond a simple technical generalization, because it enables us to estimate a core area in terms of an appropriate level set of the stationary distribution of an RBMD in the home range S. Note that this can be done because the stationary distribution is non-uniform, in general.

As a by-product, we provide explicit conditions for the existence and geometric ergodicity of an RBMD on the domain. Lastly, the drift estimation problem is also addressed.

Using a different approach, an exponential rate has been obtained in the estimation of the stationary distribution of ergodic diffusions in unbounded domains (Dalalyan (2005)). See also Cattiaux, León and Prieur (2017), where a similar problem is considered. The estimation of the stationary distribution of a stochastic differential equation with drift, without reflection, has been studied by several authors, including Veretnnikov (1999). More recently, Dalalyan and Reiss (2007) estimated the drift and stationary distribution for the same model, but without reflection, whereas Gobet, Hoffmann and Reiβ (2004) consider estimation problems for one-dimensional diffusions, with and without reflection.

Before introducing the formal framework, we briefly discuss the application of the proposed method to estimating the core area and drift from animal tracking data. For a description of home-range estimation, see, for instance, Cholaquidis et al. (2016), and the references therein.

1.1. Roadmap

This paper is organized as follows. In Section 2, we discuss the conditions necessary for the existence, uniqueness, and geometric ergodicity of the RBMD. The main results in this section are given in Propositions 1 and 2. Proposition 1 gives sufficient conditions for Harris recurrence and for the the domain to be non-trap for the RBMD process $\{X_t\}$ (a condition introduced in Burdzy, Chen and Marshall (2006), which we describe in Section 2). In Proposition 2, we show that if the domain is non-trap, we have an exponential rate of convergence to the stationary distribution for the total variation norm. All proofs for this section are given in Appendix B of the Supplementary Material. In Section 3, Theorem 1 derives strong uniform convergence rates for the kernel estimators of the stationary distribution, based on a trajectory of the RBMD. In Corollary 1 and Theorem 3, we prove the strong consistency of two families of level-set estimators with respect to the Hausdorff distance. Theorem 4 estimates level sets with given content, and Theorem 5 derives consistent estimators of the drift function. Lastly, in Section 4, we consider simulated and real-data examples to illustrate the behavior of the proposed estimation methods.

2. RBMD

In this section, we establish the conditions necessary for the existence of an RBMD and its stationary distribution, and study the connections between these conditions and several geometric constraints on its support. All proofs for this section are given in Appendix B of the Supplementary Material.

2.1. Notation

Given a set $S \subset \mathbb{R}^d$, we denote by ∂S, $\mathrm{int}(S)$, and $\overline{S}$ the boundary, interior, and closure, respectively, of S. If S is a finite set, we denote its cardinal by $\#S$. The Borel sigma algebra in S is denoted by $\mathcal{B}(S)$. We denote by $\langle \cdot, \cdot \rangle$ the usual inner product in $\mathbb{R}^d$, and by $\|\cdot\|$ the Euclidean norm. A closed ball of radius ε centred at x is denoted by $\mathcal{B}(x, \varepsilon)$, and an open ball is denoted by $\mathring{\mathcal{B}}(x, \varepsilon)$. Given $\epsilon > 0$ and a bounded set $A \subset \mathbb{R}^d$, $B(A, \epsilon)$ denotes the parallel set $B(A, \epsilon) = \{x \in \mathbb{R}^d : \ d(x, A) \leq \epsilon\}$, where $d(x, A) = \inf\{\|x - a\| : \ a \in A\}$. The d-dimensional Lebesgue measure on $\mathbb{R}^d$ is denoted by μ_L.

2.2. Definition of the reflected Brownian motion

Let D be a bounded domain in $\mathbb{R}^d$ (i.e, a bounded, connected open set), such that ∂D is C^2. Given a d-dimensional Brownian motion $\{B_t\}_{t \geq 0}$, departing from $B_0 = 0$ and defined on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P}_x)$, we examine the existence and uniqueness of the solution to a reflected stochastic differential equation on $\overline{D}$, given by

$$X_t = X_0 + B_t + \int_0^t \mu(X_s) ds + \int_0^t \mathbf{n}(X_s) \xi(ds), \text{ where } X_t \in \overline{D}, \ \forall t \geq 0, \quad (2.1)$$

where the drift, $\mu(x)$, is assumed to be Lipschitz, and $\mathbf{n}(x)$ denotes the inner unit vector at the boundary point $x \in \partial D$; this boundary satisfies some regularity conditions (to be specified later). This equation is called a *Skorokhod stochastic differential equation*. Its solution is a pair of stochastic processes $\{X_t, \xi_t\}_{t \geq 0}$, where the first coordinate $\{X_t\}_{t \geq 0}$ is a *reflected diffusion*, called an RBMD, and $\{\xi_t\}_{t \geq 0}$ is the corresponding *local time*, that is, a one-dimensional continuous nondecreasing process with $\xi_0 = 0$ that satisfies $\xi_t = \int_0^t \mathbb{I}_{\{X_s \in \partial D\}} d\xi_s$. Because we have assumed that ∂D is C^2, we know that a ball of positive radius rolls freely inside and outside $\overline{D}$ (see Walther (1999)). Then, using the same arguments as those used to prove Proposition 3 in Cholaquidis et al. (2016), we can ensure that the geometric shape conditions for the existence and uniqueness of a solution to equation (2.1), as required in Saisho (1987), are satisfied. From Theorem 5.1 in Saisho (1987), it follows that there exists a unique strong solution to the Skorokhod stochastic differential equation given in (2.1). The solution is a strong solution in the sense of definition 1.6 in Ikeda and Watanabe (1981).

Remark 1. There exists a unique positive function $p(s, x, t, y)$ satisfying $\mathbb{P}(X_t \in$

$\Gamma|X_s = x) = P(s,x,t,\Gamma) = \int_\Gamma p(s,x,t,y)dy$ and, from Theorem 3.2.1 of Stroock and Varadhan (1997), the function p satisfies the forward equation $\partial_s p + \mathcal{L}^* p = 0$ and $\lim_{s\to t^-} p(s,.,t,y) = \delta_y$, where δ_y is the point-mass at y, and $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$; that is, $\mathcal{L}^* h = (1/2)\Delta h - \langle \mu, \nabla h\rangle$.

2.3. Ergodic properties

We now introduce the concepts of an invariant measure and an ergodic process, following Meyn and Tweedie (1993).

Definition 1. A probability measure π on S is said to be an *invariant measure* for a time-homogeneous Markov process $\{Z_t\}_{t\geq 0}$ if $\int_S \mathbb{P}_x(Z_t \in A)\pi(dx) = \pi(A)$, for all $t > 0$ and all $A \in \mathcal{B}(S)$.

Definition 2. A Markov process $\{Z_t\}_{t\geq 0}$ with state space S is *ergodic* if there exists an invariant probability measure π, such that $\lim_{t\to+\infty} \left\|\mathbb{P}_x(Z_t \in \cdot) - \pi(\cdot)\right\|_{TV} = 0, \quad \forall x \in S$. Here, $\|\mu\|_{TV}$ denotes the total variation norm of the measure μ. In this case, π is called a *stationary distribution*.

Remark 2. If the drift is given by the gradient of some function f, that is $\mu(x) = (1/2)\nabla f(x)$, then by Green's formula, there exists a unique stationary distribution, given by $\pi(dx) = ce^{-f(x)}\mathbb{I}_D dx = g(x)dx$, where c is the normalization constant.

Definition 3. A Markov process $\{Z_n\}_{n\in\mathbb{N}}$ with state space S is geometrically ergodic if there exists an invariant probability π and real numbers $0 < \rho < 1$ and $\gamma > 0$, such that

$$\left|\mathbb{P}_x(Z_n \in B) - \pi(B)\right| \leq \gamma\rho^n \text{ for all } x \in S \text{ and all } B \in \mathcal{B}(S). \tag{2.2}$$

2.4. Harris recurrence and the trap condition

Let $D \subset \mathbb{R}^d$ be an open, bounded set, and $\mathcal{B} \subset D$. Define $T_\mathcal{B} = \inf\{t > 0\colon Z_t \in \mathcal{B}\}$ as the first hitting time of $\mathcal{B}$ by a stochastic process $\{Z_t\}_{t\geq 0}$.

Definition 4. A Markov process $\{Z_t\}_{t\geq 0}$ is *Harris recurrent* if for some σ-finite measure μ, we have $\mathbb{P}_x(T_A < \infty) = 1$ whenever $\mu(A) > 0$, for $A \in \mathcal{B}(\overline{D})$.

Under Harris recurrence, there exists a unique (up to a multiplicative constant) invariant measure (see Azéma, Kaplan-Duflo and Revuz (1967)). For the RBMD, Proposition 1 proves a sufficient condition for Harris recurrence (where, in Definition 4, μ is the Lebesgue measure restricted to D), that is slightly stronger than the *non-trap* condition introduced in Burdzy, Chen and Marshall (2006)).

Definition 5. We say that D is a *trap domain* for the stochastic process $\{Z_t\}_{t\geq 0}$ if there exists a closed ball $\mathcal{B} \subset D$ with positive radius, such that $\sup_{x\in D} \mathbb{E}_x T_{\mathcal{B}} = \infty$, where $\mathbb{E}_x$ denotes the expectation w.r.t. $\mathbb{P}_x$. Otherwise, D is called a *non-trap domain.*

The non-trap condition is mandatory when estimating the stationary distribution and the drift function in order to visit infinitely many often a small ball at each point x.

It is proved in Lemma 3.2 of Burdzy, Chen and Marshall (2006) that if $\{X_t\}_{t\geq 0}$ is a reflected Brownian motion (without drift) in a connected open set D with finite volume, and $\mathcal{B}_1$ and $\mathcal{B}_2$ are closed nondegenerate balls in D, then $\sup_{x\in D} \mathbb{E}_x T_{\mathcal{B}_1} < \infty$ if and only if $\sup_{x\in D} \mathbb{E}_x T_{\mathcal{B}_2} < \infty$.

Proposition 1. *Let $D \subset \mathbb{R}^d$ be a bounded domain, such that ∂D is C^2. Let $\{X_t\}_{t\geq 0}$ be the solution to* (2.1)*. Then, for all Borel sets A, such that $\mu_L(A\cap D) > 0$, we have that $\sup_{x\in D} \mathbb{E}_x(T_A) < \infty$, where $\mathbb{E}_x$ denotes the expectation w.r.t. $\mathbb{P}_x$, which implies Harris recurrence.*

The next proposition states that under the non-trap condition, the process is geometrically ergodic. This result can also be derived using functional inequalities, as proposed in Cattiaux, León and Prieur (2017); see Section 3.1.

Proposition 2. *Let $D \subset \mathbb{R}^d$ be a bounded domain, such that ∂D is C^2. Denote by π the invariant distribution of $\{X_t\}_{t\geq 0}$. If D is a non-trap domain for $\{X_t\}_{t\geq 0}$, then there exist positive constants α and β, such that $\sup_{x\in D} \big\|\mathbb{P}_x(X_t \in \cdot) - \pi(\cdot)\big\|_{TV} \leq \beta e^{-\alpha t}$.*

3. Drift and Level-Set Estimation

In this section, we first obtain strong uniform convergence rates for the classical kernel density estimator $\hat{g}_n$ of the density g of the stationary distribution of a geometrically ergodic Markov chain (see Theorem 1). This allows us to estimate the density, g, of the stationary distribution of the RBMD $\{X_t\}_{t\geq 0}$ by considering a sequence $\{X_{kn_1}\}_{k\in\mathbb{N}}$ (the choice of n_1 is given explicitly in the proof of Proposition 2). As is well known, uniform convergence is crucial to obtaining the convergence of level sets (see Theorem 3). Next, in Corollary 1 and Theorem 3, we show the convergence of two families of estimators of the level sets. In Theorem 4, we estimate level sets with given content.

The proof of Theorem 1 is based on some of the ideas proposed in Campos and Dorea (2005); however, we aim to obtain uniform convergence, in order to estimate the level sets. Before doing so, we need to introduce some notation.

Let $\{X_n\}_{n\in\mathbb{N}}$ be a Markov process with state space $S \subset \mathbb{R}^d$, and let $\mu_0(dy)$ be an arbitrary initial distribution. Let $\mu_n(dy)$ denote the distribution of X_n, that is, $\mathbb{P}_{\mu_0}(X_n \in A) = \int_A \mu_n(dy)$, for $A \in \mathcal{B}(S)$, where $\mathbb{P}_{\mu_0}$ indicates that the initial distribution is μ_0. Similarly, $\mathbb{E}_{\mu_0}$ indicates the corresponding expectation.

Let $K\colon \mathbb{R}^d \to \mathbb{R}$ be a bounded function, such that $K \geq 0$ and $\int K(t)dt = 1$. Consider the classical kernel estimator $\hat{g}_n$, based on $\{X_1, \dots, X_n\}$, given by

$$\hat{g}_n(x) = \frac{1}{nh_n^d}\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) = \frac{1}{n}\sum_{i=1}^n K_h(x-y),$$

where $h = h_n \to 0$ and $K_h(x) = K(x/h)/h^d$.
The following generalization of the Bernstein inequality, obtained in Collomb (1984), is used throughout this discussion. Recall that a stochastic process $\{X_k\}_{k\in\mathbb{Z}}$ is φ-mixing if $\sup_{j\in\mathbb{Z}} \sigma(\mathcal{F}_{-\infty}^j, \mathcal{F}_{j+n}^\infty) \to 0$ as $n \to \infty$, where $\mathcal{F}_j^k = \sigma(X_s, j \leq s \leq k)$.

Lemma 1. *(Bernstein inequality for φ-mixing processes). Let Y_i be a sequence of φ-mixing random variables, such that $\mathbb{E}(Y_i) = 0$, $|Y_i| \leq C_1$, $\mathbb{E}|Y_i| \leq \eta$, and $\mathbb{E}(Y_i^2) \leq D$. Write $\tilde{\varphi}(m) = \varphi(1) + \cdots + \varphi(m)$, for each $m \in \mathbb{N}$. Then, for each $\varepsilon > 0$ and $n \in \mathbb{N}$, we have*

$$\mathbb{P}\left(\Big|\sum_{k=1}^n Y_k\Big| > \varepsilon\right) \leq 2\exp\left(3e^{1/2}n\frac{\varphi(m)}{m} - \alpha\varepsilon + \alpha^2 nC_2\right), \tag{3.1}$$

where $C_2 = 6(D+4\eta C_1\tilde{\varphi}(m))$, and α and m are any positive real number and any positive integer less than or equal to n and satisfying $\alpha m C_1 \leq 1/4$, respectively. The numbers α and m may also depend on n.

Theorem 1. *Let $S \subset \mathbb{R}^d$ be a compact set and $\{X_n\}_{n\in\mathbb{N}}$ be a geometrically ergodic Markov chain, with state space S and constants γ and ρ given by (2.2), with a stationary distribution, π, that has a Lipschitz density g w.r.t. the Lebesgue measure. Define $g_1 = \max_{x\in S} g(x)$, and denote by C_g the Lipschitz constant of g. Let $\hat{g}_n(x) = (1/n)\sum_{i=1}^n K_h(x - X_i)$, where $K : \mathbb{R}^d \to \mathbb{R}$ is a nonnegative bounded Lipschitz function, such that $\int K(t)dt = 1$. Define $\kappa = \int |u|K(u)du < \infty$ and $k_1 = \max K(x)$. Let $h = h_n \to 0$, $\alpha_n \to 0$, and $\beta_n \to \infty$, such that $\beta_n h_n \to 0$, $\alpha_n = o(1/\beta_n)$, and $\log(n)/\beta_n \to 0$. Then, for all $\epsilon > 0$ and all $n > n_1$ (n_1 is* *given in the proof), we have*

$$\mathbb{P}\left(\beta_n \sup_{x\in S}\big|\hat{g}_n(x)-g(x)\big|>\epsilon\right)\leq C\Gamma_n + C'h\beta_n + \frac{3c}{h^{d(d+2)}}\exp\Big(-\frac{\epsilon\alpha_n}{4\Gamma_n}\Big), \quad (3.2)$$

where $\Gamma_n=\beta_n/(nh^d)$, $C=2k_1\gamma\sum_{n=1}^{\infty}\rho^n$, $C'=\kappa C_g$, *and* c *is a constant depending only on* d *and* $\mu_L(S)$.
Moreover, if β_n *and* h_n *satisfy* $\alpha_n nh^d/(\beta_n\log(n))\to\infty$, *then* $\beta_n\sup_{x\in S}\big|\hat{g}_n(x)-g(x)\big|\to 0\quad a.s.$

Remark 3.

i) Taking $h=n^{-1/\nu}$ and $\beta_n=n^{\gamma}$, the best attainable rate that can be derived from Theorem 1 is for $\gamma=1/(d+2)$; that is, $\beta_n=\mathcal{O}(n^{1/(d+2)})$.

ii) If we need uniform convergence only, we can relax the conditions in h_n, and replaced them with $h=\mathcal{O}((1/n)^{1/(d+1)})$.

Using Theorem 2 of Cuevas, Gonzalez-Manteiga and Rodríguez-Casal (2006), we obtain the following direct corollary, which establishes the rate of consistency for the Hausdorff distance of the boundary of the estimated level sets $\partial G_{\hat{g}_n}(\lambda)$ (where $G_g(\beta)=\{x\colon g(x)>\beta\}$). Recall that $d(a,C)=\inf_{c\in C}d(a,c)$ and, given two nonempty compact sets $A,C\subset\mathbb{R}^d$, the Hausdorff distance between A and C is defined as

$$d_H(A,C)=\max\left\{\max_{a\in A}d(a,C),\ \max_{c\in C}d(c,A)\right\}.$$

Corollary 1. *Under the hypotheses of Theorem* 1*, suppose in addition that there exists* $\lambda>0$*, such that* $\partial G_g(\lambda)\neq\emptyset$*, and there exist* $\gamma>0$ *and* $A>0$*, such that if* $|t-c|\leq\gamma$*, then* $d_H(\{g=c\},\{g=t\})\leq A|t-c|$*. Then,* $d_H\big(\partial G_g(\lambda),\partial G_{\hat{g}_n}(\lambda)\big)=o(1/\beta_n)$ *a.s.*

Remark 4. As pointed out in Cuevas, Gonzalez-Manteiga and Rodríguez-Casal (2006) in Section 2.4 point 1, the hypotheses of Corollary 1 are satisfied if g is C^2 on a neighborhood E of the level set λ, and the gradient of g is strictly positive on E.

3.1. Level-set estimation under shape restrictions

In this subsection, we propose another estimator of the level sets, under a quite general shape condition. We assume that there exists an $r>0$, such that

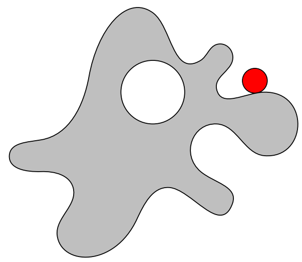

Figure 1. A general r-convex set. The small ball has radius r.

$\overline{G_g(\lambda)}$ is compact and r-convex; that is, $\overline{G_g(\lambda)} = C_r\big(\overline{G_g(\lambda)}\big)$, where

$$C_r\big(\overline{G_g(\lambda)}\big) = \bigcap_{\left\{\mathring{\mathcal{B}}(x,r)\,:\ \mathring{\mathcal{B}}(x,r)\cap\overline{G_g(\lambda)}=\emptyset\right\}} \big(\mathring{\mathcal{B}}(x,r)\big)^c$$

is the r-convex hull of $G_g(\lambda)$.

This condition has been studied extensively in the context of set estimation; see, for instance, Cuevas, Fraiman and Pateiro-López (2012), and Rodríguez-Casal (2007). It is also related to the level-set estimation problem; see Walther (1997). Although r-convexity is much less restrictive than convexity, inlets that are too sharp are not allowed; see Figure 1.

Following the notation in Federer (1959), let $\mathrm{Unp}(S)$ be the set of points $x \in \mathbb{R}^d$ with a unique projection on S, denoted by $\xi_S(x)$. That is, for $x \in \mathrm{Unp}(S)$, $\xi_S(x)$ is the unique point that attains the minimum of $\|x-y\|$, for $y \in S$.

Definition 6. For $x \in S$, let $reach(S,x) = \sup\big\{r > 0 : \mathring{\mathcal{B}}(x,r) \subset \mathrm{Unp}(S)\big\}$. The reach of S is defined by $reach(S) = \inf\big\{reach(S,x) : x \in S\big\}$, and S is said to be of positive reach if $reach(S) > 0$.

The relation between r-convexity and reach has been studied in Cuevas, Fraiman and Pateiro-López (2012).

Definition 7. The outer Minkowski content of $S \subset \mathbb{R}^d$ is given by $L_0(\partial S) = \lim_{\epsilon\to 0} \mu_L(B(S,\epsilon)\setminus S)/\epsilon$, provided that the limit exists and is finite.

Definition 8. Let $S \subset \mathbb{R}^d$ be a closed set. A ball of radius r is said to roll freely in S if, for each boundary point $s \in \partial S$, there exists some $x \in S$, such that $s \in \mathcal{B}(x,r) \subset S$. The set S is said to satisfy the outside r-rolling condition if a ball of radius r rolls freely in $\overline{S^c}$.

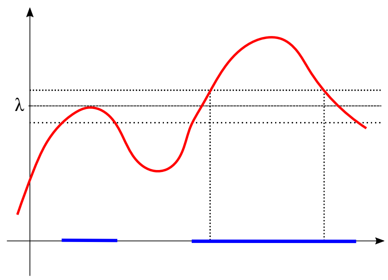

Figure 2. If $g'(x) = 0$ for $x \in G_g(\lambda)$, it is not necessarily true that $d_H(G_g(\lambda+\varepsilon), G_g(\lambda-\varepsilon)) \to 0$.

We also assume the following condition.

HR: A level set $G_g(\lambda)$ satisfies **HR** if there exist $\delta_0 > 0$ and $r > 0$, such that $\overline{G_g(\lambda+\varepsilon)}$ is r-convex for all $-\delta_0 < \varepsilon < \delta_0$.

Theorem 2 in Walther (1997) gives sufficient conditions for **HR** to hold, expressed in terms of the gradient of g. Thus, we have the following result.

Theorem 2. *Let $g\colon \mathbb{R}^d \to \mathbb{R}$ and $-\infty < l \leq u < \sup g$. Assume that $g \in C^1(U)$, where U is a bounded open set that contains $\overline{G_g(l-\eta)} \setminus G_g(u+\eta)$, for some $\eta > 0$; ∇g satisfies $\|\nabla g\| \geq m > 0$ on U, as well as, the following Lipschitz condition on U (or on $\partial G_g(\lambda)$): for all $\lambda \in (l, u)$, $\|\nabla g(x) - \nabla g(y)\| \leq k\|x - y\|$, for $x, y \in U$ (or in $\partial G_g(\lambda)$). Then, for each $\lambda \in (l, u)$, $\overline{G_g(\lambda)}$ and $G_g(\lambda)^c$ are r_0-convex, with $r_0 = m/k$.*

Lemma 2. *Let $g\colon S \to \mathbb{R}$, where $S \subset \mathbb{R}^d$ is a compact set. Assume that $g \in C^2(S)$, and that λ is such that there exists $0 < \delta_1 < \lambda$, for which $\nabla g(x) \neq 0$ for all $x \in \overline{G_g(\lambda-\delta_1)} \setminus G_g(\lambda+\delta_1) := \mathcal{G}_g(\lambda, \delta_1)$. Then, for all $\varepsilon < \delta_1$, $d_H\big(G_g(\lambda-\varepsilon), G_g(\lambda+\varepsilon)\big) \leq 3M\varepsilon/m^2$, where $M = \max_{\{x \in \mathcal{G}_g(\lambda,\delta_1)\}} \|\nabla g(x)\|$ and $m = \min_{\{x \in \mathcal{G}_g(\lambda,\delta_1)\}} \|\nabla g(x)\|$.*

Consider $\hat{g}_n$, as before. We study the convergence in the Hausdorff distance of the following estimator:

$$A_n(\lambda) = C_r\big(\{X_i \colon \hat{g}_n(X_i) > \lambda\}\big), \tag{3.3}$$

that is, the r-convex hull of the sample points belonging to the λ level set of $\hat{g}_n$. The rates of convergence for (3.3) in the independent case are provided in Saavedra-Nieves, González-Manteiga and Rodríguez-Casal (2016), along with an estimator for the parameter r. Note that it is not necessary to compute the whole set $G_{\hat{g}_n}(\lambda)$ (which, in practice, is not feasible in most cases), because the estimator proposed in Corollary 1 is based only on sample points that belong to the set $G_{\hat{g}_n}(\lambda)$. Moreover, for the two-dimensional case, the r-convex hull can be computed easily using the R software package `alphahull` (see Pateiro-López and Rodríguez-Casal (2010)).

Theorem 3. *Under the hypothesis of Theorem* 1*, assume further that g and λ are defined as in Lemma* 2*, and that condition **HR** holds. In addition, assume that $0 < g_0 < g(x)$, for all $x \in S$. Denote as Ψ_n the right-hand side of* (3.2)*, with $\epsilon = 1$. Let $\varepsilon_n \to 0$, such that $\varepsilon_n \beta_n > 1$ for all n, and assume that $\varepsilon_n < min\{\delta_0, \delta_1\}$ for all n, where δ_0 is defined in condition **HR**, and δ_1 is defined in Lemma 2, Then, for all $n > n_2$ (n_2 is given in the proof),*

$$\mathbb{P}\left(d_H\big(\overline{A_n(\lambda)}, \overline{G_g(\lambda)}\big) \leq \frac{3M}{m^2}\varepsilon_n\right) > 1 - 3\Psi_n.$$

The following corollary follows directly from condition **HR** and Theorem 3 in Cuevas, Fraiman and Pateiro-López (2012).

Corollary 2. *Under the hypotheses of Theorem* 3*, with probability one,* $\lim_{n\to\infty} d_H\big(\partial\overline{A_n(\lambda)}, \partial\overline{G_g(\lambda)}\big) = 0$.

3.2. Estimation of level sets with a fixed content

Theorem 4. *Let $S \subset \mathbb{R}^d$ be a compact set and $\{X_n\}_{n\in\mathbb{N}}$ be a geometrically ergodic Markov chain with state space S. For $\tau \in (0,1)$, define $l_\tau = \inf\{\lambda > 0 : \pi(G_g(\lambda)) \leq 1 - \tau\}$, where π denotes the stationary distribution. Assume that π has a C^2 density g, such that $\|\nabla g(x)\| \neq 0$ for all $x \in U$, where U is an open set containing $\overline{G_g(l_\tau - \varepsilon_0)} \setminus G_g(l_\tau + \varepsilon_0)$, for some $\tau > 0$ and $0 < \varepsilon_0 < l_\tau$. Let $\hat{g}_n(x) = (1/n)\sum_{i=1}^n K_h(x - X_i)$, with K a bounded Lipschitz density. Let $h = h_n$ be such that $h = \mathcal{O}((1/n)^{1/(d+1)})$. If we define*

$$\hat{l}_\tau = \inf\left\{\lambda > 0 : \frac{1}{n}\#\big\{i : X_i \in G_{\hat{g}_n}(\lambda)\big\} \leq 1 - \tau\right\},$$

then, with probability one, $d_H\big(G_{\hat{g}_n}(\hat{l}_\tau), G_g(l_\tau)\big) \to 0$.

3.3. Drift estimation

In what follows, we propose an estimator of the drift function. Assume that $\{X_t : t \geq 0\}$ is uniformly sampled at times $\{t = t_1, \dots, t_n\}$ in the interval $[0, T]$, where $T > 0$; that is, a sample of size n of the process X_t, $\{X_{\Delta_{n,T}}, \dots, X_{n\Delta_{n,T}}\}$ is observed at $t_i = i\Delta_{n,T}$ and $\Delta_{n,T} = T/n$. To simplify the notation, we refer to Δ rather than $\Delta_{n,T}$. Fix $x \in int(S)$. Define $N_x = \#\{1 \leq i \leq n : X_{t_i} \in \mathcal{B}(x, h_n)\}$, for some $h_n \to 0$. Then, we define the estimator, $\hat{\mu}_{n,T}(x) = (1/\Delta N_x) \sum_{i=1}^{n} (X_{t_{i+1}} - X_{t_i}) \mathbb{I}_{\{X_{t_i} \in \mathcal{B}(x,h_n)\}}$.

Theorem 5. *Assume that $T \to \infty$, $\Delta \to 0$, $h_n \to 0$, $\Delta n h_n^2 \to \infty$, and $\Delta n h_n^3 \to 0$. Then, for all $x \in int(S)$, $\hat{\mu}_{n,T}(x) \to \mu(x)$ in probability.*

The proof is given in Appendix C of the Supplementary Material. According to Remark 2, in the gradient case, the drift estimator can be derived easily from the stationary density estimator using the plug-in rule

$$\hat{\mu}_1(x) = \frac{1}{2} \nabla \log(\hat{g}_n(x)). \tag{3.4}$$

4. Examples

In this section, we first use a simulation study to assess the performance of the r-convex hull of the sample points belonging to the level set of the estimator, as proposed in (3.3). Then, we present the results of applying this method to real data.

4.1. Simulations

The discrete version of the RBMD (2.1) is produced using the Euler scheme proposed in Bossy, Gobet and Talay (2004), as follows. We first choose a step $\delta > 0$, and denote by $\text{sym}(z)$ the symmetrization of the point z with respect to ∂S. Start with $X_0 = x$, and suppose that we have obtained $X_i \in S$. To produce the following point, set $Y_{i+1} = X_i + Z_i + \delta\mu(X_i)$, where Z_i is a centered Gaussian random vector, independent w.r.t. $Z_1, \ldots, Z_{i-1}$, with covariance matrix $\delta(I_d)_{\mathbb{R}^2}$. Then: 1) if $Y_{i+1} \in S$, set $X_{i+1} = Y_{i+1}$; 2) if $Y_{i+1} \notin S$ and $\text{sym}(Y_{i+1}) \in S$, set $X_{i+1} = \text{sym}(Y_{i+1})$; and 3) if $Y_{i+1} \notin S$ and $\text{sym}(Y_{i+1}) \notin S$, set $X_{i+1} = X_i$. In our example, we consider an RBMD in the set $S = E \setminus \mathcal{B}((4/5, 0), 1/2)$, where $E = \{(x, y) \in \mathbb{R}^2 \colon 4x^2/9 + y^2 \leq 1\}$, with drift function given by $\mu(x, y) = -(x, y)$. The trajectory is shown in Figure 3 for $\delta = 0.001$ in the first row, and $\delta = 0.003$ in the second row. The values for N are 10,000; 50,000; and 100,000 in the first, second, and third columns, respectively. The stationary density is shown in Figure 4 a), the estimated density using a Gaussian kernel with bandwidth $h = 0.2$ is shown in Figure 4 b), and the estimated density using an Epanechnikov kernel with bandwidth $h = 0.4$ is shown in Figure 4 c). In all cases, we used the trajectory shown in Figure 3, with $\delta = 0.003$ and $N = 100{,}000$. Because we can estimate the support, we have forced the estimation to be zero outside the

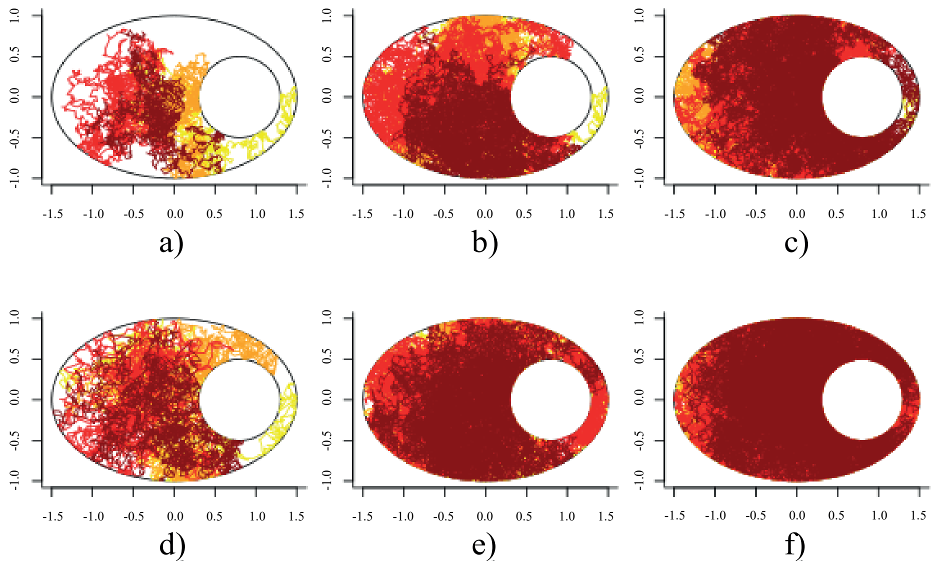

Figure 3. The trajectory of the RBMD, for different values of δ and N, in a), b), and c), $\delta = 0.001$ and $N = 10{,}000$, $N = 50{,}000$, and $N = 100{,}000$, respectively. In d), e), and f), $\delta = 0.003$ and $N = 10{,}000$, $N = 50{,}000$, and $N = 100{,}000$, respectively.

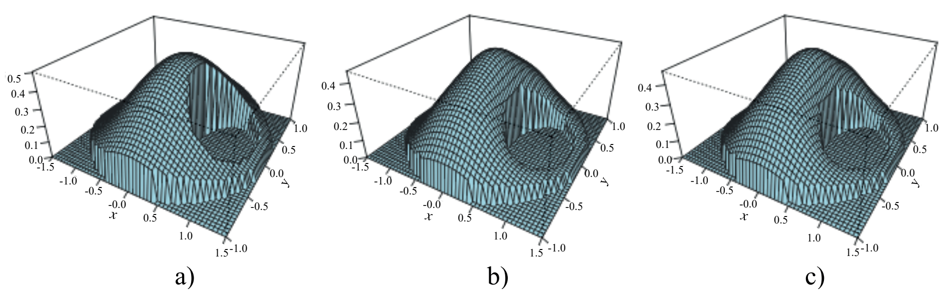

Figure 4. a) Real density, b) estimated using Gaussian Kernel with $h = 0.2$, c) estimated using Epanechnikov kernel with $h = 0.4$.

estimation of the support.

For the level sets, we consider the levels $\lambda = 0.44$, 0.41, 0.34, 0.27, and 0.03. Figure 5 a) shows the theoretical level sets for the considered values of λ, and b) shows the corresponding estimated level sets. The estimation is based on a trajectory with $\delta = 0.003$ and $N = 500{,}000$, using (3.3) with $r = 0.4$. We use the Gaussian kernel with $h = 0.1$. The choice of an optimal bandwidth for level-set estimation has been studied recently for the independent and identically

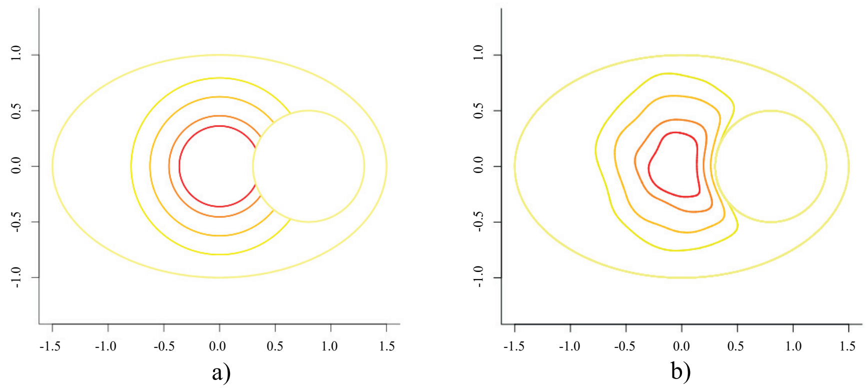

Figure 5. a) Theoretical level sets. b) Estimation using (3.3) for $r = 0.4$, with a Gaussian kernel and bandwidth $h = 0.1$. Red indicates a core area.

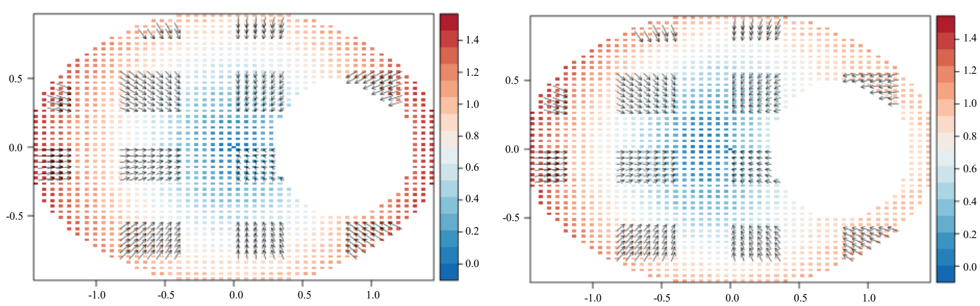

Figure 6. (Left) Theoretical drift. (Right) Estimation using (3.4).

distributed (i.i.d.) case; see Qiao (2018). Although it should behave similarly for geometric mixing processes, extending the results of Qiao (2018) is beyond the scope of this study. It is clear that the hole in the domain will produce border effects for the density estimation, and therefore for the level sets. A way to overcome this problem (which is computationally very expensive) is to first estimate the support using the r-convex hull of the trajectory, and then to use a variable bandwidth kernel estimate, where the bandwidth is given by the lesser of a fixed h and the distance from the point x to the boundary of the support.

The left side of Figure 6 shows the theoretical vector field corresponding to the drift, and the right shows the estimator (3.4) based on the trajectory given in Figure 3 f), using the Gaussian kernel and bandwidth $h = 0.45$.

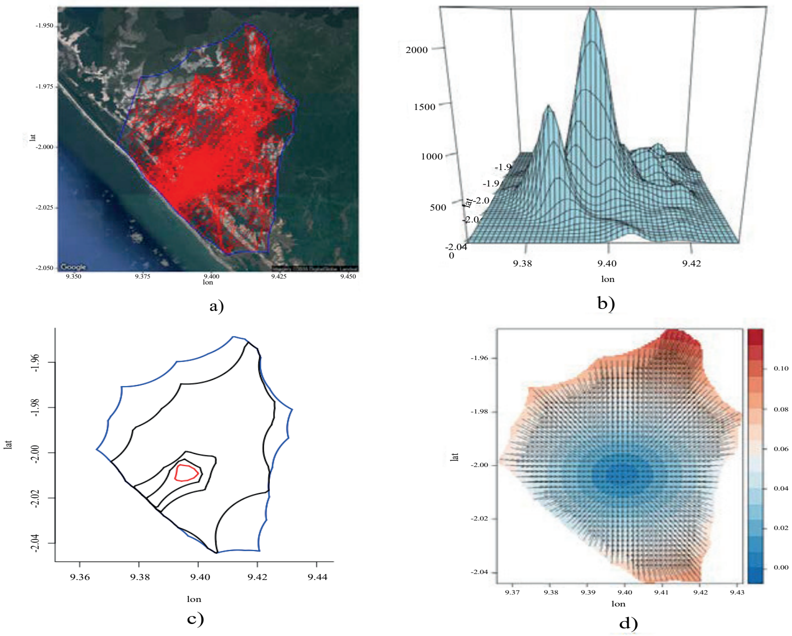

Figure 7. a) Trajectory and 0.02-convex hull, b) density estimator using Gaussian Kernel with $h = 0.01$, c) r-convex hull of level sets, d) Estimation of the drift.

4.2. Real-data examples

We considered a data set from the Movebank database, where a natural barrier acts as a boundary on an animal's movement. GPS collars were placed on elephants in Loango National Park in western Gabon. The area is protected by the Atlantic Ocean on the west and by Lagoon Iguéla to the east. Figure 7 a) shows the movement of an elephant (in red) using an estimator $N = 1{,}633$ for recorded positions. The blue lines represent the boundary of the r-convex hull estimator for $r = 0.02$. The estimated density is shown in b), using the Gaussian kernel with bandwidth $h = 0.01$. The r-convex hulls of the level sets are shown in c) for $\lambda_1 = 100, \lambda_2 = 600$, $\lambda_3 = 1{,}100$, $\lambda_4 = 1{,}600$, and $r = 0.02$. In d) we show the estimation of the drift, using (3.4) with $h = 0.5$.

Supplementary Material

The proofs for lemma 2 in Section 3.1 and Lemmas 3 and 4 given in Appendix A are provided in Appendix A1 of the online Supplementary Material. In addition, the proofs for results in Sections 2 and 3.3 are given in Appendix B and C, respectively, in the Supplementary Material.

Acknowledgments

We thank Professor K. Burdzy for his helpful comments during the preparation of this manuscript, as well as two referees for their constructive comments and suggestions. We also thank Dr. Stephen Blake of the Max Planck Institute for Ornithology for facilitating access to the data set used in this study. All remaining errors are the responsibility of the authors.

A. Appendix

Here we include the proofs of the propositions stated in Section 3.

Proof of Theorem 1. We will deal separately with each term on the right hand side of the following inequality:

$$\beta_n \sup_x \big|g(x) - \hat{g}_n(x)\big| \leq \beta_n \sup_x \big|g(x) - \mathbb{E}_\pi(\hat{g}_n(x))\big| + \beta_n \sup_x \big|\hat{g}_n(x) - \mathbb{E}_\pi(\hat{g}_n(x))\big|. \tag{A.1}$$

First we bound the bias term.

$$\begin{aligned} \big|\mathbb{E}_\pi(K_h(x-X_k)) - g(x)\big| &\leq \left|\int_S K_h(x-y)g(y)dy - g(x)\right| \\ &\leq \int_S K_h(x-y)|g(y)-g(x)|dy \qquad\qquad (A.2) \\ &\leq hC_g \int_{\mathbb{R}^d} \|u\|K(u)du = \kappa C_g h. \end{aligned}$$

Observe that $\mathbb{E}_{\mu_0}(K_h(x{-}X_k)) = \int_S K_h(x{-}y)\mu_k(dy)$. Recall that $k_1 = \max_x K(x)$. Now, by (2.2), since $\|K_h(x-y)\|_\infty \|\mu_k - \pi\|_{TV} \leq (k_1/h^d)\gamma\rho^k$.

$$\left|\int_S K_h(x-y)\mu_k(dy) - \int_S K_h(x-y)d\pi(y)\right| \leq \frac{k_1}{h^d}\gamma\rho^k. \tag{A.3}$$

Observe that $\mathbb{E}(\hat{g}_n(x)) = (1/n)\sum_{k=1}^n \mathbb{E}_{\mu_0}(K_h(x-X_k))$, and $\mathbb{E}_\pi(K_h(x-X_k)) = \int_S K_h(x-y)g(y)dy$. Hence (A.3) implies

$$\beta_n \sup_x \left|\frac{1}{n}\sum_{k=1}^n \big[\mathbb{E}_{\mu_0}(K_h(x-X_k)) - \mathbb{E}_\pi(K_h(x-X_k))\big]\right| \leq C\Gamma_n \tag{A.4}$$

which, together with (A.2), entails

$$\beta_n \sup_x \big|\mathbb{E}(\hat{g}_n(x)) - g(x)\big| \leq C\frac{\beta_n}{nh^d} + \kappa C_g h\beta_n. \tag{A.5}$$

Since S is compact, we can cover S with $\nu \le c/h^{d(d+2)}$ balls of radius h^{d+2} centred at some fixed points $\{x_1, \dots, x_\nu\} \subset S$, c being a positive constant depending only on d and $\mu_L(S)$. For $i = 1, \dots, \nu$, $\mathbb{P}\big(\big|\hat{g}_n(x_i) - \mathbb{E}(\hat{g}_n(x_i))\big| > \varepsilon\big)$ equals

$$\mathbb{P}\left(\Big|\sum_{j=1}^{n}\big[K_h(x_i - X_j) - \mathbb{E}_{\mu_0}(K_h(x_i - X_j))\big]\Big| > n\varepsilon\right).$$

By proposition 4.1 of Campos and Dorea (2005), the sequence $\{X_n\}_{n\in\mathbb{N}}$ is φ mixing with $\varphi(n) = 2\gamma\rho^n$ (γ as in (2.2)). Let $x \in S$ and x_i be such that $\|x - x_i\| < h^{d+2}$. Since K is Lipschitz, denote by R the Lipschitz constant of K, then $|\hat{g}_n(x) - \hat{g}_n(x_i)| \le R\|x - x_i\|/h^{d+1} \le Rh$. Hence, $\big|\hat{g}_n(x) - \mathbb{E}(\hat{g}_n(x))\big| \le \big|\hat{g}_n(x_i) - \mathbb{E}(\hat{g}_n(x_i))\big| + 2Rh$. If we take n so large that $2Rh < \varepsilon/(2\beta_n)$, we get

$$\mathbb{P}\left(\sup_{x\in\mathcal{B}(x_i,h^{d+2})}\big|\hat{g}_n(x) - \mathbb{E}(\hat{g}_n(x))\big| > \frac{\varepsilon}{\beta_n}\right) \le \mathbb{P}\left(|\hat{g}_n(x_i) - \mathbb{E}(\hat{g}_n(x_i))| > \frac{\varepsilon}{2\beta_n}\right).$$

Now use the Bernstein inequality (3.1) with $Y_j = K((x - X_j)/h) - \mathbb{E}_{\mu_0}(K(x - X_j)/h)$ and $C_1 = 2k_1$. Recall that $g_1 = \max_{x\in S} g(x)$. Let us take n_0 such that for all $n > n_0$ $\rho^n/(h_n^d g_1\mu_L(S)) < 1$ and $2Rh < \varepsilon/(2\beta_n)$. Denote by $C'' = 2k_1\gamma g_1\mu_L(S)$, then for $n > n_0$, by (A.3), $\mathbb{E}_{\mu_0}(K((x - X_j)/h)) \le k_1\gamma\rho^n + \int_S K((x - y)/h)\, g(y)dy \le k_1\gamma\rho^n + k_1h^d g_1\mu_L(S) \le C''h^d$.

Hence $\eta = 2C''h^d$, $D \le 2k_1C''h^d$, and $\tilde{\varphi}(m) \le \sum_{i=1}^{\infty} 2\gamma\rho^i < 2\gamma$, so $C_2 = 12k_1C''h^d(1 + 16\gamma)$. Since $\alpha_n = o(1/\beta_n)$, if $m = \lfloor\beta_n\rfloor$, then $\alpha_nC_2h^{-d} < \varepsilon/(4\beta_n)$ and $\alpha_n mC_1 < 1/4$ for n large enough. On the other hand, since $\log(n)/\beta_n \to 0$, $3e^{1/2}n(\varphi(m)/m) \to 0$, as $n \to \infty$. Let us take $n_1 > n_0$, such that for all $n > n_1$, $\beta_n\alpha_nC_2h^{-d} < \epsilon/4$, $\alpha_n mC_1 < 1/4$ and $2\exp(3e^{1/2}n(\varphi(m)/m)) < 3$. Now Bernstein's inequality implies that, for all $n > n_1$

$$\begin{aligned}\mathbb{P}\left(\big|\hat{g}_n(x_i) - \mathbb{E}(\hat{g}_n(x_i))\big| > \frac{\varepsilon}{2\beta_n}\right) &\le 2\exp\left(3e^{1/2}n\frac{\varphi(m)}{m} - \frac{\alpha_n\varepsilon}{2\Gamma_n} + \alpha_n^2C_2n\right)\\ &\le 3\exp\left(-\frac{\varepsilon\alpha_n}{4\Gamma_n}\right).\end{aligned}$$

Lastly, for $n > n_1$,

$$\mathbb{P}\left(\sup_x\big|\hat{g}_n(x) - \mathbb{E}(\hat{g}_n(x))\big| > \frac{\varepsilon}{\beta_n}\right) \le \sum_{i=1}^{\nu}\mathbb{P}\left(\big|\hat{g}_n(x_i) - \mathbb{E}(\hat{g}_n(x_i))\big| > \frac{\varepsilon}{2\beta_n}\right)$$

$$\leq \frac{3c}{h^{d(d+2)}} \exp\left(-\frac{\epsilon\alpha_n}{4\Gamma_n}\right)$$

which, together with (A.1) and (A.5), implies (3.2).

To prove the almost surely convergence, observe that $\alpha_n/(4\Gamma_n \log(n)) \to \infty$ and $nh \to \infty$, imply that $(1/\log(n))\big[\epsilon\alpha_n/(4\Gamma_n) + (d(d+2))\log(h)\big] \to \infty$, and then we can apply Borel-Cantelli's Lemma.

Proof of Theorem 3. Let us consider $n > n_1$ where n_1 is given in Theorem 1. Let us denote $\mathcal{A}_n = \{\beta_n \sup_x |\hat{g}_n(x) - g(x)| < 1\}$, we now that $\mathbb{P}(\mathcal{A}_n) > 1 - \Psi_n$ for all $n > n_1$. Since $\varepsilon_n\beta_n > 1$ and $\varepsilon_n < \delta_0$, by condition **HR**, $\mathbb{P}(A_n(\lambda) \subset G_g(\lambda-\varepsilon_n)) > 1-\Psi_n$ for all $n > n_1$. By Lemma 2 it is enough to prove that there exists n_2 such that for all $n > n_2$, $\mathbb{P}(G_g(\lambda+\varepsilon_n) \subset A_n(\lambda)) > 1-2\Psi_n$ or what is the same we have to prove that for $n > n_1$, $\mathbb{P}(\exists x_n \in G_g(\lambda+\varepsilon_n) : x_n \notin A_n(\lambda)) \leq 2\Psi_n$. Let us denote $\mathcal{X}_n = \{X_1, \dots, X_n\}$ and

$$\mathcal{C}_n = \Big\{\exists x_n \in G_g(\lambda + \varepsilon_n) \text{ such that } \exists y_n : x_n \in \mathcal{B}(y_n, r), \#\big\{\mathcal{X}_n \cap \mathcal{B}(y_n, r)\big\} = 0\Big\}.$$

Then,

$$\begin{aligned}
&\big\{\exists x_n \in G_g(\lambda + \varepsilon_n) : x_n \notin A_n(\lambda)\big\} \subset \mathcal{C}_n \cup \\
&\Big\{\{\exists x_n \in G_g(\lambda + \varepsilon_n) \text{ such that } \exists y_n : x_n \in \mathcal{B}(y_n, r), \#\big\{\mathcal{X}_n \cap \mathcal{B}(y_n, r)\big\} > 0\big\} \\
&\cap \big\{\forall X_i \in \mathcal{B}(y_n, r), \hat{g}_n(X_i) \leq \lambda\big\}\Big\} = \mathcal{C}_n \cup \mathcal{F}_n.
\end{aligned}$$

Since g is Lipschitz if $x_n \in G_g(\lambda+\varepsilon_n)$, $g(z) > \lambda+\varepsilon_n/2$ for all $z \in \mathcal{B}(x_n, \nu_n)$ where $\nu_n = \varepsilon_n/(2C_g)$. Then on $\mathcal{A}_n$, for all $n > n_1$ $\hat{g}_n(z) > \lambda$, for all $z \in \mathcal{B}(x_n, \nu_n)$. The set $\mathcal{E}_n$ defined as follows fulfills,

$$\begin{aligned}
&\Big\{\exists x_n \in G_g(\lambda + \varepsilon_n) \text{ such that } \exists y_n : x_n \in \mathcal{B}(y_n, r), \#\{\mathcal{X}_n \cap \mathcal{B}(y_n, r)\} > 0\Big\} \\
&\cap \big\{\forall X_i \in \mathcal{B}(y_n, r) \cap \mathcal{B}(x_n, \nu_n), \hat{g}_n(X_i) \leq \lambda\big\} \subset \mathcal{A}_n^c.
\end{aligned}$$

And then, $\mathbb{P}(\mathcal{F}_n) \leq \mathbb{P}(\mathcal{E}_n) \leq \Psi_n$.

Let us bound $\mathbb{P}(\mathcal{C}_n) \leq \Psi_n$. To do that, let us introduce, for each fixed $n > n_1$, the random variables $Z_k(y) = K\big(\|X_k - y\|/r\big) \quad k = 1, \dots, n$, where K is a Lipschitz function such that $\mathbb{I}_{[0,1/2]}(x) \leq K(x) \leq \mathbb{I}_{[0,1]}(x)$ and $K(x) > 0$ for all $x \in (1/2, 1)$, then $\mathbb{P}(\mathcal{C}_n) \leq \mathbb{P}\big(\inf_{y\in S}(1/n)\sum_{k=1}^n Z_k(y) = 0\big)$. Proceeding as in (A.4),

$$\sup_{y\in S}\left|\frac{1}{n}\sum_{k=1}^{n}\left[\mathbb{E}_{\mu_0}(Z_k(y))-\mathbb{E}_{\pi}(Z_k(y))\right]\right|\leq\frac{2\gamma}{n}\sum_{k=1}^{\infty}\rho^k. \tag{A.6}$$

Since $g(x)>g_0>0$ for all $z\in S$,

$$\mathbb{E}_{\pi}(Z_k(y))=\int_{\mathcal{B}(y,r)}K\left(\frac{\|t-y\|}{r}\right)g(t)dt\geq g_0\left(\frac{r}{2}\right)^d\omega_d>0. \tag{A.7}$$

Let us fix $0<\epsilon<g_0(r/2)^d\omega_d/3$, from (A.6) and (A.7), if we take n large enough such that $(2\gamma/n)\sum_{k=1}^{\infty}\rho^k<g_0(r/2)^d\omega_d/3$, it is enough to prove that there exists n_2 such that for all $n>n_2$,

$$\mathbb{P}\left(\sup_{y\in S}\Big|\frac{1}{n}\sum_{k=1}^{n}[Z_k(y)-\mathbb{E}_{\mu_0}(Z_k(y))]\Big|>\epsilon\right)<\Psi_n.$$

As before, since S is compact, we can cover it with $\zeta\leq c/\iota^d$ balls of radius ι centred at some fixed points $\{x_1,\dots,x_\zeta\}$ where c is a constant which depends only on d and $\mu_L(S)$. First, observe that if $(2\gamma/n)\sum_{k=1}^{\infty}\rho^k<\epsilon/5$, then, $\sup_{y_i}\left|(1/n)\sum_{k=1}^{n}[\mathbb{E}_{\mu_0}(Z_k(y_i))-\mathbb{E}_{\pi}(Z_k(y_i))]\right|\leq\epsilon/5$.
If $\mu_L(\mathcal{B}(y_i,r)\triangle\mathcal{B}(y,r))g_1<\epsilon/5$, where g_1 is the maximum of g,

$$\sup_{y_i\in\mathcal{B}(y,\iota)}\left|\frac{1}{n}\sum_{k=1}^{n}[\mathbb{E}_{\pi}(Z_k(y_i))-\mathbb{E}_{\pi}(Z_k(y))]\right|\leq\frac{\epsilon}{5}$$
$$\sup_{y\in S}\left|\frac{1}{n}\sum_{k=1}^{n}[\mathbb{E}_{\pi}(Z_k(y))-\mathbb{E}_{\mu_0}(Z_k(y))]\right|\leq\frac{\epsilon}{5}.$$

Using Berstein inequality, we can bound, for a fixed y,

$$\mathbb{P}\left(\Big|\frac{1}{n}\sum_{k=1}^{n}[Z_k(y)-\mathbb{E}_{\mu_0}(Z_k(y))]\Big|>\frac{\epsilon}{\beta_n}\right)\leq3\exp\left(-\frac{\epsilon\alpha_n n}{4\beta_n}\right). \tag{A.8}$$

where $\alpha_n=o(1/\beta_n)$ and $\log(n)/\beta_n\to0$. Then

$$\mathbb{P}\left(\sup_{y\in S}\Big|\frac{1}{n}\sum_{k=1}^{n}[Z_k(y)-\mathbb{E}_{\mu_0}(Z_k(y))]\Big|>\epsilon\right)\leq I_1+I_2,$$

where $I_1=\mathbb{P}\big(\sup_{y_i\in\mathcal{B}(y,\iota)}|1/n\sum_{k=1}^{n}[Z_k(y)-Z_k(y_i)]|>\epsilon/5\big)$ and $I_2=\mathbb{P}\big(\sup_{y_i}$ $|1/n\sum_{k=1}^{n}[Z_k(y_i)-\mathbb{E}_{\mu_0}(Z_k(y_i))]|>\epsilon/5\big)$.

Since K is Lipschitz (let us denote C_K the Lipschitz constant of K) we can bound $I_1 \leq C_K \iota/r$ and from (A.8), $I_2 \leq (3c/\iota^d)\exp\big(-\epsilon\alpha_n n/(4\beta_n)\big)$. Now take $\iota = h_n$ (being h_n as in Theorem 1), then $I_1 + I_2 \leq \Gamma_n$.

In order to prove Theorem 4 we will need two lemmas. For the first, recall that given a probability distribution P, $\mathcal{A}$ is a P-uniformity class if $\sup_{A\in\mathcal{A}} |P_n(A) - P(A)| \to 0$ whenever $P_n \to P$ weakly. Theorem 5 in Cuevas, Fraiman and Pateiro-López (2012) proves that the class of sets with reach bounded from below by a positive constant included in a compact set is a P-uniformity class.

Lemma 3. *Let $S \subset \mathbb{R}^d$ be a compact set and $g : S \to \mathbb{R}$ a C^2 function such that that there exists an $\varepsilon_0 > 0$ and a $c > 0$ such that $\|\nabla g(x)\| > m$ for all $x \in U$, where U is an open set containing $\overline{G_g(l_\tau - \varepsilon_0)} \setminus G_g(l_\tau + \varepsilon_0)$. Then $\{G_g(\lambda) : l_\tau - \varepsilon_0/2 \leq \lambda \leq l_\tau + \varepsilon_0/2\}$ is a P-uniformity class for all probability distributions P on S absolutely continuous w.r.t. Lebesgue measure.*

Lemma 4. *Under the hypotheses of Lemma 3, for all $0 \leq \varepsilon < \varepsilon_0/2$ and all $l_\tau - \varepsilon < \lambda < l_\tau + \varepsilon$, $G_g(\lambda-\varepsilon) \setminus G_g(\lambda+\varepsilon) \subset B\big(\partial G_g(\lambda), 3\varepsilon M/m^2\big)$ where $M = \max_{\{x\in\overline{G_g(l_\tau-\varepsilon_0)}\setminus G_g(l_\tau+\varepsilon_0)\}} \|\nabla g(x)\|$ and $m = \min_{\{x\in\overline{G_g(\lambda-\delta_1)}\setminus G_g(\lambda+\delta_1)\}} \|\nabla g(x)\|$.*

Proof of Theorem 4. By Remark 3 ii) we have that $\sup_{x\in S} |\hat{g}_n(x) - g(x)| \to 0$ a.s. We will prove that $\hat{l}_\tau \to l_\tau$ a.s. Define $L(\lambda) = \pi(G_g(\lambda))$, $\hat{L}(\lambda) = (1/n)\#\{i : X_i \in G_{\hat{g}_n}(\lambda)\}$ and $\tilde{L}(\lambda) = (1/n)\#\{i : X_i \in G_g(\lambda)\}$ and $I(\epsilon_0, l_\tau) = [l_\tau - \varepsilon_0/2, l_\tau + \varepsilon_0/2]$,

$$\sup_{\lambda\in I(\epsilon_0,l_\tau)} |L(\lambda) - \hat{L}(\lambda)| \leq \sup_{\lambda\in I(\epsilon_0,l_\tau)} |L(\lambda) - \tilde{L}(\lambda)| + \sup_{\lambda\in I(\epsilon_0,l_\tau)} |\tilde{L}(\lambda) - \hat{L}(\lambda)|.$$

$$\begin{aligned}|\tilde{L}(\lambda) - \hat{L}(\lambda)| &= \frac{1}{n}\Big|\#\{i : X_i \in G_g(\lambda)\} - \#\{i : X_i \in G_{\hat{g}_n}(\lambda)\}\Big| \\ &= \frac{1}{n}\Big(\#\{i : X_i \in G_g(\lambda) \setminus G_{\hat{g}_n}(\lambda)\} + \#\{i : X_i \in G_{\hat{g}_n}(\lambda) \setminus G_g(\lambda)\}\Big).\end{aligned}$$

Since $\sup_x |\hat{g}_n(x) - g(x)| \to 0$ a.s., we have that for all λ and ε, $G_g(\lambda+\varepsilon) \subset G_{\hat{g}_n}(\lambda) \subset G_g(\lambda-\varepsilon)$ with probability one, for n large enough. Then, with probability one, for n large enough, for all $0 \leq \varepsilon < \varepsilon_0/2$,

$$\sup_{\lambda\in I(\epsilon_0,l_\tau)} |\tilde{L}(\lambda) - \hat{L}(\lambda)| \leq \sup_{\lambda\in I(\epsilon_0,l_\tau)} \frac{2}{n}\#\Big\{i : X_i \in G_g(\lambda-\varepsilon) \setminus G_g(\lambda+\varepsilon)\Big\}.$$

By Lemma 3, $G_g(\lambda)$ is a P-uniformity class. Hence,

$$\sup_{\lambda\in I(\epsilon_0,l_\tau)} \left|\frac{1}{n}\#\Big\{i : X_i \in G_g(\lambda-\varepsilon) \setminus G_g(\lambda+\varepsilon)\Big\} - \pi\big(G_g(\lambda-\varepsilon) \setminus G_g(\lambda+\varepsilon)\big)\right| \to 0$$

and $\pi\big(G_g(\lambda-\varepsilon)\setminus G_g(\lambda+\varepsilon)\big) \le g_1\mu_L\big(G_g(\lambda-\varepsilon)\setminus G_g(\lambda+\varepsilon)\big)$, where $g_1 = \max_{x\in S} g(x)$. By Lemma 4,

$$\sup_{\lambda\in I(\epsilon_0,l_\tau)} \mu_L\big(G_g(\lambda-\varepsilon)\setminus G_g(\lambda+\varepsilon)\big) \le \sup_{\lambda\in I(\epsilon_0,l_\tau)} \mu_L\left(B\left(\partial G_g(\lambda), \frac{3\varepsilon M}{m^2}\right)\right).$$

For a fixed $\varepsilon > 0$, $\mu_L\big(B\big(\partial G_g(\lambda), 3\varepsilon M/m^2\big)\big)$ is a continuous function of λ, and so its maximum is attained in some $\lambda_0 \in I(\epsilon_0, l_\tau)$. Since $reach(\partial G_g(\lambda_0)) > 0$, the outer Minkowski content of $G_g(\lambda_0)$ and $G_g(\lambda_0)^c$ exist, and so by corollary 3 of Ambrosio, Colesanti and Villa (2008), $\sup_{\lambda\in I(\epsilon_0,l_\tau)} \mu_L\big(G_g(\lambda-\varepsilon)\setminus G_g(\lambda+\varepsilon)\big) = \mathcal{O}(\varepsilon)$, from which it follows that $\sup_{\lambda\in I(\epsilon_0,l_\tau)} |\tilde{L}(\lambda)-\hat{L}(\lambda)| \to 0$. Using Lemma 3 it follows that $\sup_{\lambda\in I(\epsilon_0,l_\tau)} |L(\lambda)-\tilde{L}(\lambda)| \to 0$, then $\sup_{\lambda\in I(\epsilon_0,l_\tau)} |L(\lambda)-\hat{L}(\lambda)| \to 0$. To prove that $\hat{l}_\tau \to l_\tau$ a.s., let $0 < \varepsilon < \varepsilon_0/2$ and $\gamma = \min\big\{1-\tau-L(l_\tau+\varepsilon/2), L(l_\tau-\varepsilon/2)-(1-\tau)\big\}$. Now observe that $\gamma > 0$ since L is decreasing in $l_\tau-\varepsilon_0 \le \lambda \le l_\tau+\varepsilon_0$. Let n be so large that $\sup_{\lambda\in I(\epsilon_0,l_\tau)} |L(\lambda)-\hat{L}(\lambda)| < \gamma/2$. Then $l_\tau-\varepsilon/2 < \hat{l}_\tau < l_\tau+\varepsilon/2$. To conclude the proof, observe that since $\|\nabla g(x)\| > m$ for all $x \in U$, where U is an open set containing $\overline{G_g(l_\tau-\varepsilon_0)}\setminus G_g(l_\tau+\varepsilon_0)$, it follows that $\overline{\{x : g(x) < \lambda\}} = \{x : g(x) \le \lambda\}$ for all $l_\tau-\varepsilon_0 < \lambda < l_\tau+\varepsilon_0$. Now we apply theorem 2.1 of Molchanov (1998), which implies that, with probability one, $\sup_{\lambda\in I(\epsilon_0,l_\tau)} d_H\big(G_{\hat{g}_n}(\lambda), G_g(\lambda)\big) \to 0$. Lastly, the result follows since $d_H\big(G_{\hat{g}_n}(\hat{l}_\tau), G_g(l_\tau)\big) \le d_H\big(G_{\hat{g}_n}(\hat{l}_\tau), G_g(\hat{l}_\tau)\big) + d_H\big(G_g(\hat{l}_\tau), G_g(l_\tau)\big)$, and the second one converges to zero by Lemma 4.

References

Ambrosio, L., Colesanti, A. and Villa, E. (2008). Outer Minkowski content for some classes of closed sets. *Math. Ann.* **342**, 727–748.

Azéma, J., Kaplan-Duflo, M. and Revuz, D. (1967). Mesure invariante sur les classes récurrentes des processus de Markov *Z. Wahrscheinlichkeits* **8**, 157–181.

Bossy, M., Gobet, E. and Talay, D. (2004). Symmetrized Euler scheme for an efficient approximation of reflected diffusions *J. Appl. Probab.* **41**, 877–889.

Burdzy, K., Chen, Z.-Q. and Marshall, D. E. (2006). Traps for reflected Brownian motion. *Math. Z.* **252**, 103–132.

Burt, W. H. (1943). Territoriality and home range concepts as applied to mammals. *J. Mammal.* **24**, 346-352.

Campos, V. S. M. and Dorea, C. C. Y. (2005). Kernel estimation for stationary density of Markov chains with general state space. *Ann. Inst. Statist. Math.* **57**, 443–453.

Cattiaux, P., León, J. R., Prieur, C. (2017). Invariant density estimation for a reflected diffusion using an Euler scheme. *Monte Carlo Methods and Applications* **23**.

Cholaquidis, A., Fraiman, R., Lugosi, G. and Pateiro-López, B. (2016). Set estimation from reflected Brownian motion. *J. R. Stat. Soc. Ser. B. Stat. Methodol* **78**, 1057–1078.

Collomb, G. (1984). Uniform complete convergence of the kernel predictor. *Wahrsch. Verw. Gebiete* **57**, 441–460.

Cuevas, A., González-Manteiga, W. and Rodríguez-Casal, A. (2006). Plug-in estimation of general level sets. *Aust. N. Z. J. Stat.* **48**, 7–19.

Cuevas, A., Fraiman, R. and Pateiro-López, B. (2012). On statistical properties of sets fullfilling rolling-type conditions. *Adv. in Appl. Probab.* **44**, 311–239.

Dalalyan, A. (2005). Sharp adaptative estimation of the drift function for ergodic diffusions. *Ann. Statist.* **33**, 2507–2528.

Dalalyan, A. and Reiss, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probability Theory and Related Fields* **137**, 25–47.

Federer, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93**, 418–491.

Gobet, E., Hoffmann, M. and Reiβ (2004). Nonparametric estimation of scalar diffusions based on low-frequency data. *Ann. Stat.* **32**, 2223–2253.

Hayne, D. W. (1949). Calculation of size of home range. *J. Mammal.* **30**, 1–18.

Ikeda, N. and Watanabe, S. (1981). *Stochastic Differential Equations and Diffusion Processes.* Amsterdam: North-Holland.

Qiao, W. (2018). Asymptotics and optimal bandwidth selection for nonparametric estimation of density level sets. *arXiv:* `https://arxiv.org/pdf/1707.09697.pdf`.

Meyn, S. P. and Tweedie, R. L. (1993). Stability of Markovian processes II: Continuous-time processes and sampled chains *Adv. Appl. Prob.* **25**, 487–517.

Molchanov, I. (1998). A limit theorem for solutions of inequalities. *Scandinavian Journal of Statistics* **25**, 235–242.

Pateiro-López, B. and Rodríguez-Casal, A. (2010). Generalizing the Convex Hull of a Sample: The R Package alphahull. *J. Stat. Softw.* **34**, 1–28.

Rodríguez-Casal, A. (2007). Set estimation under convexity-type assumptions. *Ann. Inst. Henri Poincaré Probab. Stat.* **43**, 763–774.

Saavedra-Nieves, P., González-Manteiga, W., Rodríguez-Casal, A. (2016). A comparative simulation study of data-driven methods for estimating density level sets. *Journal of Statistical Computation and Simulation* **86**, 236-251.

Stroock, D. W. and Varadhan, S. R. S. (1997). *Multidimensional Diffusion Processes.* Springer-Verlag, Berlin.

Saisho, Y. (1987). Stochastic differential equations for multi-dimensional domain with reflecting boundary. *Probab. Theory Related Fields* **74**, 455–477.

Veretnnikov, A. Yu. (1999). On Castellana–Leadbetter's condition for diffusion density estimation. *Statistical Inference for Stochastic Processes* **2**, 1–9.

Walther, G. (1997). Granulometric smoothing. *Ann. Statist.* **25**, 2273–2299.

Walther, G. (1999). On a generalization of Blaschke's Rolling Theorem and the Smoothing of Surfaces. *Math. Methods Appl. Sci.* **22**, 301–316.

Worton, B. J. (1987). A review of models of home range for animal movement. *Ecol. Model.* **38**, 277–298.

Alejandro Cholaquidis

Iguá 4225, piso 16. 11400, Montevideo, Uruguay.

E-mail: acholaquidis@cmat.edu.uy
Ricardo Fraiman
Iguá 4225, piso 16. 11400, Montevideo, Uruguay.
E-mail: rfraiman@cmat.edu.uy
Ernesto Mordecki
Iguá 4225, piso 16. 11400, Montevideo, Uruguay.
E-mail: mordecki@cmat.edu.uy
Cecilia Papalardo
Iguá 4225, piso 16. 11400, Montevideo, Uruguay.
E-mail: cecipap@gmail.com

(Received May 2018; accepted February 2019)