ON ESTIMATION FOR BROWNIAN MOTION GOVERNED BY TELEGRAPH PROCESS WITH MULTIPLE OFF STATES

VLADIMIR POZDNYAKOV[1*], L. MARK ELBROCH[2], CHAORAN HU[1], THOMAS MEYER[4], AND JUN YAN[1,3]

Abstract. Brownian motion whose infinitesimal variance changes according to a three-state continuous-time Markov Chain is studied. This Markov Chain can be viewed as a telegraph process with one on state and two off states. We first derive the distribution of occupation time of the on state. Then the result is used to develop a likelihood estimation procedure when the stochastic process at hand is observed at discrete, possibly irregularly spaced time points. The likelihood function is evaluated with the forward algorithm in the general framework of hidden Markov models. The analytic results are confirmed with simulation studies. The estimation procedure is applied to analyze the position data from a mountain lion.

Keywords: Forward algorithm, Likelihood estimation, Markov process, Occupation time

1. Introduction

Random walks on a plane, whether simple, biased, or correlated, have a long history of being employed by ecologists to model the movement of animals, micro-organisms, and cells on a small time scale. By the functional Central Limit Theorem, from an appropriate distance any random walk (under some mild regularity conditions) looks like a Brownian Motion (BM). So, it is not surprising that recently diffusions are often used to model animal movement on a large time scale (e.g., Preisler et al., 2004; Tilles and Petrovskii, 2016). An excellent review on applications of random walks and diffusions in this area of research can be found in Codling et al. (2008).

Horne et al. (2007) introduced the Brownian bridge movement model (BBMM) that, in essence, assumes that animal movement is perpetual and described by a BM. Pauses in animal movement (on a small time scale) were first introduced in Othmer et al. (1988) where the dispersal of cells or organisms is modeled by a process that comprises a sequence of alternating pauses and jumps. The moving-resting (MR) process introduced in Yan et al. (2014) and further investigated in Pozdnyakov et al. (2019) allows an animal to have two states, moving and resting. In the moving state, the motion is characterized by a BM; in the resting state, there is no movement. The duration in either moving or resting states is assumed to be exponentially distributed.

Properties and fitting of the MR model are based on results for telegraph processes (the alternating renewal process or the on-off process) that were obtained in Perry et al. (1999), Di Crescenzo (2001), Stadje and Zacks (2004), and Zacks (2004). The distribution of total time spent in a state plays a critical role in applications driven by a telegraph process (Zacks, 2012). In particular, a BM governed by a telegraph process is an active area of research such as being recently employed in

1. Department of Statistics, University of Connecticut, 215 Glenbrook Road, Storrs, CT 06269-4120
2. Panthera, 8 West 40th Street, 18th Floor, NY, NY 10018
3. Center for Environmental Sciences and Engineering, University of Connecticut, 3107 Horsebarn Hill Road, Storrs, Connecticut 06269-4210
4. Department of Natural Resources and the Environment, University of Connecticut, 1376 Storrs Road, Storrs, Connecticut 06269-4087
* E-mail: vladimir.pozdnyakov@uconn.edu.

continuous-time option pricing theory (e.g., Di Crescenzo and Pellerey, 2002; Kolesnik and Ratanov, 2013; Di Crescenzo et al., 2014; Di Crescenzo and Zacks, 2015).

In animal movement ecology, it is reasonable to assume that there are very different explanations for why a predator is not moving. For example, an animal might spend time resting (as in Yan et al., 2014), consuming a prey item, or denning. Resting can be assumed to not last even a single day. However, some predators that can kill a (relatively) large prey item evolved highly elastic guts, and they consume the kill by repeatedly gorging and digesting over a prolonged period called *handling*. For example, mountain lions (*Puma concolor*) might remain at a kill for days. Both resting and handling are periodic in the time scales of this model but denning is not, and it is inapplicable to male mountain lions in any case. Therefore, this model concerns only two non-moving activities, resting and handling, and it is clear that their durations must be different.

This observation motivates our model. In the new model we have one moving state and two motionless states. From a motionless state one always switches to the moving state. Nonetheless, when moving ends, the motionless state type is chosen randomly. For tractability, all the durations (or holding times) are exponentially distributed. We will call this continuous-time process a *moving-resting-handling process*, or *MRH process*. An extension of the telegraph process to an alternating process with three states is studied in Bshouty et al. (2012). The difference is that in Bshouty et al. (2012) three states alternate deterministically within a renewal cycle. In our case we have only two states within a renewal cycle but one of the motionless states is chosen at random.

In practice, an MRH process is typically observed at discrete, possibly irregularly spaced time points. Estimation of MRH process parameters is challenging because the states are unobserved, and the observed sequence is not Markov. Our estimation procedure uses techniques developed for the hidden Markov model (HMM). More specifically, the dynamic programming, or the forward algorithm, for HMMs is employed to compute the likelihood (e.g., Cappé et al., 2005). As will be seen, the key to this problem is the distribution of the time that the MRH process spends in the moving state. Our methodology differs from the standard approach to occupation time distribution in continuous-time Markov chain (Sericola, 2000). The method is general so that it remains valid when the holding times are not exponentially distributed, in which case, the state process is semi-Markov; see discussions in Section 7. An implementation of the methods in this paper and in our earlier works (Pozdnyakov et al., 2014; Yan et al., 2014; Pozdnyakov et al., 2019), is publicly available in an R package `smam` (Yan et al., 2019).

2. Formal Description of MRH Process

Let $S(t)$, $t \geq 0$, be a continuous-time Markov Chain with the state space $\{0, 1, 2\}$ and the transition rate matrix

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 p_1 & \lambda_0 p_2 \\ \lambda_1 & -\lambda_1 & 0 \\ \lambda_2 & 0 & -\lambda_2 \end{pmatrix} \tag{1}$$

where $p_1, p_2, \lambda_0, \lambda_1, \lambda_2 > 0$ and $p_1 + p_2 = 1$. The zero entries in the matrix means that state 1 or state 2 do not transit between themselves; only a transition to state 0 is allowed from either of them. In animal movement modeling, the mean duration in state 0, 1, and 2 are, respectively, $1/\lambda_0$, $1/\lambda_1$, and $1/\lambda_2$. We assume that the initial distribution ν_0 of $S(0)$ is stationary, that is,

$$\nu_0 = \boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2) = \frac{1}{1/\lambda_0 + p_1/\lambda_1 + p_2/\lambda_2} \left(\frac{1}{\lambda_0}, \frac{p_1}{\lambda_1}, \frac{p_2}{\lambda_2} \right). \tag{2}$$

Recall that $\boldsymbol{\pi}$ has to satisfy $0 = \boldsymbol{\pi}\mathbf{Q}$ (e.g., Norris, 1998, p.120).

Let $B(t)$ be the standard BM independent of $S(t)$. Then the MRH process is given by

$$X(t) = \sigma \int_0^t 1_{\{S(s)=0\}} \mathrm{d}B(s), \tag{3}$$

where $\sigma > 0$ is an infinitesimal standard deviation.

Estimation of the MRH process parameters $\boldsymbol{\theta} = (\lambda_0, \lambda_1, \lambda_2, p_1, \sigma)$ is based on observations at discrete, possibly irregularly spaced time points. The observed data are represented by the vector of observed changes in location

$$\mathbf{X} = \big(X(t_1) - X(0), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})\big),$$

where $0 < t_1 < \cdots < t_n$ are the time points of the observations. As mentioned earlier, the difficulty is that the MRH process itself is not Markov. However, the location-state process $\{X(t), S(t)\}$ is Markov. So, our first objective is to derive formulas for transitional probabilities of the location-state process. The key random variable here is the total time spent in state 0 in the time interval $[0, t]$:

$$M(t) = \int_0^t 1_{\{S(s)=0\}} \mathrm{d}s. \tag{4}$$

We also can call this random variable *0-state occupation time* by time t.

A continuous-time Markov Chain can be alternatively described by representing the process $S(t)$ as a combination of a discrete time Markov Chain, holding times, and initial distribution ν. More specifically, let p_{ij} be the probability of switching to state j at the next jump given that we are currently in state i. The matrix

$$\mathbf{P} = (p_{ij}) = \begin{pmatrix} 0 & p_1 & p_2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

is a stochastic matrix, and it is the transition matrix of the embedded (discrete time) Markov Chain of process $S(t)$. The time spent in a particular state i between two consecutive jumps is called the holding time. The holding time has exponential distribution with rate λ_i. For our task this representation (via an embedded Markov Chain and holding times) is a bit more convenient. Note also that in the case of the standard telegraph process the associated stochastic matrix of the embedded Markov chain is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Our technique is different from the general approach to the distribution of occupation times in homogeneous finite-state Markov processes (e.g., Sericola, 2000). To develop computationally efficient estimation procedure we exploit the specific structure of our Markov chain. More specifically, a telegraph process can be associated with $S(t)$ if we collapse states 1 and 2 into one state. For this new state the holding time is distributed as a mixture of two exponential distributions. As a consequence, the telegraph process is not Markov. This makes computing the likelihood function for $\mathbf{X}$ challenging, because algorithms like the forward algorithm are not applicable. That is, results for telegraph processes can not be directly employed, because we do need to distinguish states 1 and 2. We use a certain periodicity of the Markov Chain and extend the technique developed in Di Crescenzo (2001) for telegraph processes to obtain the joint distribution of $M(t)$ and $S(t)$. An alternative approach can be developed by extending the method presented in Zacks (2012).

3. Joint Distribution of Occupation Time $M(t)$ and $S(t)$

To simulate process $S(t)$ that starts with $S(0) = 0$, we need the following independent sequences of random variables:

(1) $\{M_k\}_{k\geq 1}$ are independent identically distributed (iid) random variables with $\text{Exp}(\lambda_0)$ distribution,
(2) $\{R_k\}_{k\geq 1}$ are iid random variables with $\text{Exp}(\lambda_1)$,
(3) $\{H_k\}_{k\geq 1}$ are iid random variables with $\text{Exp}(\lambda_2)$,
(4) $\{\xi_k\}_{k\geq 1}$ are iid random variables with $\Pr(\xi_k = 1) = p_1$ and $\Pr(\xi_k = 0) = p_2$.

Having these sequences defined we can proceed as follows. To generate a particular realization of $S(t)$, first, generate M_1, the time duration the process spends in state 0. Then generate ξ_1 to decide whether it jumps to state 1 or 2. Depending on ξ_1 generate the duration R_1 or H_1. After that, switch back to state 0, and so on.

Let us introduce some auxiliary random variables. Let $U_k = \xi_k R_k + (1-\xi_k)H_k$, $C_k = M_k + U_k$, and

$$N(t) = \sup\{n \geq 0 : \sum_{k=1}^{n} C_k \leq t\}.$$

Here and everywhere in the text, by convention, a summation over an empty set is 0, for instance, $\sum_{k=1}^{0} C_k = 0$. Random variable $N(t)$ is the number of full cycles C_k by time t.

First, we consider the distribution of occupation time $M(t)$ when $S(t) = 0$. Denote $P_i(\cdot) = \Pr(\cdot|S(0) = i)$, where $i = 0, 1, 2$. With probability 1 the random variable $M(t) \in [0, t]$, and it has an atom at t in the following sense:

$$P_0(M(t) = t, S(t) = 0) = P_0(M(t) = t) = \Pr(M_1 > t) = e^{-\lambda_0 t}.$$

Now, fix $0 < s < t$. Then we have

$$\begin{aligned} P_0(M(t) \in \mathrm{d}s, S(t) = 0) &= \sum_{n=0}^{\infty} P_0(M(t) \in \mathrm{d}s, S(t) = 0, N(t) = n) \\ &= \sum_{n=1}^{\infty} P_0(M(t) \in \mathrm{d}s, S(t) = 0, N(t) = n), \end{aligned}$$

because $S(0) = 0$, $S(t) = 0$ and $N(t) = 0$ implies $M(t) = t$.

Next, for $n \geq 1$ we get that

$$\begin{aligned} &P_0(M(t) \in \mathrm{d}s, S(t) = 0, N(t) = n) \\ &= \Pr\left(\sum_{k=1}^{n} M_k + \sum_{k=1}^{n} U_k \leq t,\ \sum_{k=1}^{n+1} M_k + \sum_{k=1}^{n} U_k > t,\ t - \sum_{k=1}^{n} U_k \in \mathrm{d}s\right) \\ &= \Pr\left(\sum_{k=1}^{n} M_k \leq s,\ \sum_{k=1}^{n+1} M_k > s,\ \sum_{k=1}^{n} U_k \in t - \mathrm{d}s\right) \\ &= \Pr\left(\sum_{k=1}^{n} M_k \leq s,\ \sum_{k=1}^{n+1} M_k > s\right) \Pr\left(\sum_{k=1}^{n} U_k \in t - \mathrm{d}s\right) \\ &= \left[\Pr\left(\sum_{k=1}^{n} M_k \leq s\right) - \Pr\left(\sum_{k=1}^{n+1} M_k \leq s\right)\right] \Pr\left(\sum_{k=1}^{n} U_k \in t - \mathrm{d}s\right). \end{aligned}$$

Here we use independence of $\{M_k\}_{k\geq 1}$ and $\{U_k\}_{k\geq 1}$.

The sums $\sum_{k=1}^{n} M_k$ and $\sum_{k=1}^{n+1} M_k$ have gamma distributions, Gamma(n, λ_0) and Gamma$(n+1, \lambda_0)$, respectively. The distribution of $\sum_{k=1}^{n} U_k$ can be expressed in terms of the convolution of gamma distributions. More specifically, by conditioning on $\{\xi_k\}_{1 \le k \le n}$ one can show that

$$\Pr\left(\sum_{k=1}^{n} U_k \le s\right) = \sum_{k=0}^{n} \Pr\left(\sum_{j=1}^{k} R_j + \sum_{j=1}^{n-k} H_j \le s\right) \binom{n}{k} p_1^k p_2^{n-k}.$$

Random variables $\sum_{j=1}^{k} R_j$ and $\sum_{j=1}^{n-k} H_j$ are independent, and they have Gamma(k, λ_1) and Gamma$(n-k, \lambda_2)$ distributions, respectively. For the convolution of gamma distributions, we refer the reader to Mathai (1982), Moschopoulos (1985) and Hu et al. (2019b). The implementations of these methods are publicly available in R package `coga` (Hu et al., 2019a).

Next, let us work out the case when $S(t) = 1$. Again, the random variable $M(t) \in [0, t]$, but now it has no atoms. For any $0 < s < t$, we have

$$P_0(M(t) \in \mathrm{d}s, S(t) = 1) = \sum_{n=0}^{\infty} P_0(M(t) \in \mathrm{d}s, S(t) = 1, N(t) = n)$$

Then for $n \ge 0$ we now get that

$$\begin{aligned}
&P_0(M(t) \in \mathrm{d}s, S(t) = 1, N(t) = n) \\
&= \Pr\left(\sum_{k=1}^{n+1} M_k + \sum_{k=1}^{n} U_k \le t,\ \sum_{k=1}^{n+1} M_k + \sum_{k=1}^{n} U_k + R_{n+1} > t,\ \sum_{k=1}^{n+1} M_k \in \mathrm{d}s,\ \xi_{n+1} = 1\right) \\
&= \Pr\left(\sum_{k=1}^{n} U_k \le t - s,\ \sum_{k=1}^{n} U_k + R_{n+1} > t - s,\ \sum_{k=1}^{n+1} M_k \in \mathrm{d}s,\ \xi_{n+1} = 1\right) \\
&= p_1 \Pr\left(\sum_{k=1}^{n} U_k \le t - s,\ \sum_{k=1}^{n} U_k + R_{n+1} > t - s\right) \Pr\left(\sum_{k=1}^{n+1} M_k \in \mathrm{d}s\right) \\
&= p_1 \left[\Pr\left(\sum_{k=1}^{n} U_k \le t - s\right) - \Pr\left(\sum_{k=1}^{n} U_k + R_{n+1} \le t - s\right)\right] \Pr\left(\sum_{k=1}^{n+1} M_k \in \mathrm{d}s\right).
\end{aligned}$$

Random variable $\sum_{k=1}^{n+1} M_k$ has Gamma$(n+1, \lambda_0)$ distribution. As before,

$$\Pr\left(\sum_{k=1}^{n} U_k \le s\right) = \sum_{k=0}^{n} \Pr\left(\sum_{j=1}^{k} R_j + \sum_{j=1}^{n-k} H_j \le s\right) \binom{n}{k} p_1^k p_2^{n-k},$$

and

$$\Pr\left(\sum_{k=1}^{n} U_k + R_{n+1} \le s\right) = \sum_{k=0}^{n} \Pr\left(\sum_{j=1}^{k+1} R_j + \sum_{j=1}^{n-k} H_j \le s\right) \binom{n}{k} p_1^k p_2^{n-k}.$$

The formulas for the joint distribution of $M(t)$ and $S(t)$ given $S(0) = 1$ can be derived in a similar way.

To summarize our findings let us first introduce the following notation:

(1) $G(x, \alpha, \beta)$, where $\alpha \ge 0, \beta > 0$, is the cdf of Gamma(α, β) distribution; by convention, Gamma$(0, \beta)$ distribution is the degenerate distribution with atom 1 at 0;
(2) $g(x, \alpha, \beta)$, where $\alpha, \beta > 0$, is the pdf of Gamma(α, β) distribution;
(3) $F(x, \alpha_1, \beta_1, \alpha_2, \beta_2)$, where $\alpha_1, \alpha_2 \ge 0, \beta_1, \beta_2 > 0$, is the cdf of the convolution of Gamma(α_1, β_1) and Gamma(α_2, β_2); note that, for example, $F(x, 0, \beta_1, \alpha_2, \beta_2) \equiv G(x, \alpha_2, \beta_2)$;
(4) $f(x, \alpha_1, \beta_1, \alpha_2, \beta_2)$, where $\beta_1, \beta_2 > 0$, $\alpha_1, \alpha_2 \ge 0$, and $\alpha_1 + \alpha_2 > 0$, is the pdf of $F(x, \alpha_1, \beta_1, \alpha_2, \beta_2)$;

(5) $H(x,\alpha_1,\beta_1,\alpha_2,\beta_2) = F(x,\alpha_1,\beta_1,\alpha_2,\beta_2) - F(x,\alpha_1+1,\beta_1,\alpha_2,\beta_2)$, where $\beta_1,\beta_2 > 0$, $\alpha_1,\alpha_2 \geq 0$, and $\alpha_1+\alpha_2 > 0$, is the difference in cdf with parameters only differing by α_1 versus α_1+1.

Finally, let us denote the (defective) densities of $M(t)$ as

$$p_{ij}(s,t) = P_i(M(t) \in \mathrm{d}s, S(t) = j)/\mathrm{d}s, \tag{5}$$

where $t \geq 0$, $0 < s < t$, $i,j = 0,1,2$.

Here is the main result of the section.

Theorem 1. *Let $t \geq 0$ and $0 < s < t$. Then*

$$P_0(M(t) = t, S(t) = 0) = e^{-\lambda_0 t}, \tag{6}$$

and

$$P_1(M(t) = 0, S(t) = 1) = e^{-\lambda_1 t}. \tag{7}$$

The defective densities are given by

$$p_{00}(s,t) = \sum_{n=1}^{\infty} [G(s,n,\lambda_0) - G(s,n+1,\lambda_0)] \sum_{k=0}^{n} f(t-s,k,\lambda_1,n-k,\lambda_2) \binom{n}{k} p_1^k p_2^{n-k}, \tag{8}$$

$$p_{01}(s,t) = \sum_{n=0}^{\infty} p_1 g(s,n+1,\lambda_0) \sum_{k=0}^{n} H(t-s,k,\lambda_1,n-k,\lambda_2) \binom{n}{k} p_1^k p_2^{n-k}, \tag{9}$$

$$p_{02}(s,t) = \sum_{n=0}^{\infty} p_2 g(s,n+1,\lambda_0) \sum_{k=0}^{n} H(t-s,k,\lambda_2,n-k,\lambda_1) \binom{n}{k} p_2^k p_1^{n-k}, \tag{10}$$

$$p_{10}(s,t) = \sum_{n=0}^{\infty} [G(s,n,\lambda_0) - G(s,n+1,\lambda_0)] \sum_{k=0}^{n} [f(t-s,k+1,\lambda_1,n-k,\lambda_2)] \binom{n}{k} p_1^k p_2^{n-k}, \tag{11}$$

$$p_{11}(s,t) = \sum_{n=1}^{\infty} p_1 g(s,n,\lambda_0) \sum_{k=0}^{n-1} H(t-s,k+1,\lambda_1,n-1-k,\lambda_2) \binom{n-1}{k} p_1^k p_2^{n-1-k}, \tag{12}$$

and

$$p_{12}(s,t) = \sum_{n=1}^{\infty} p_2 g(s,n,\lambda_0) \sum_{k=0}^{n-1} H(t-s,n-1-k,\lambda_2,k+1,\lambda_1) \binom{n-1}{k} p_1^k p_2^{n-1-k}. \tag{13}$$

Note that formula (10) can be obtained from (9) by interchanging state 1 and state 2. Also, in order to get densities $p_{2j}(s,t)$, $j = 0,1,2$ (that are not listed in Theorem 1) we simply need to interchange state 1 and state 2 in all the formulas for $p_{1j}(s,t)$, $j = 0,1,2$ of Theorem 1. Equations (12) and (13) look similar, but they are significantly different. In (12) we start and end in the same off state, but in (13) the starting and ending off states are different.

As an example, Figure 1 presents the defective densities $p_{ij}(s,t)$'s for a MRH model with parameters $\lambda_0 = 4$, $\lambda_1 = .5$, $\lambda_2 = .1$, $p_1 = .8$, and $t = 10$. Applications of the formulas in practice depend on how accurately the infinite sums can be implemented. To check the accuracy of the implementation and to verify that our formulas in Theorem 1 are free of errors or typos, we also simulated 1,000,000 realizations of the Markov chain $S(\cdot)$ for each initial state. The empirical densities follow the theoretical ones computed from our implementation extremely closely (not shown).

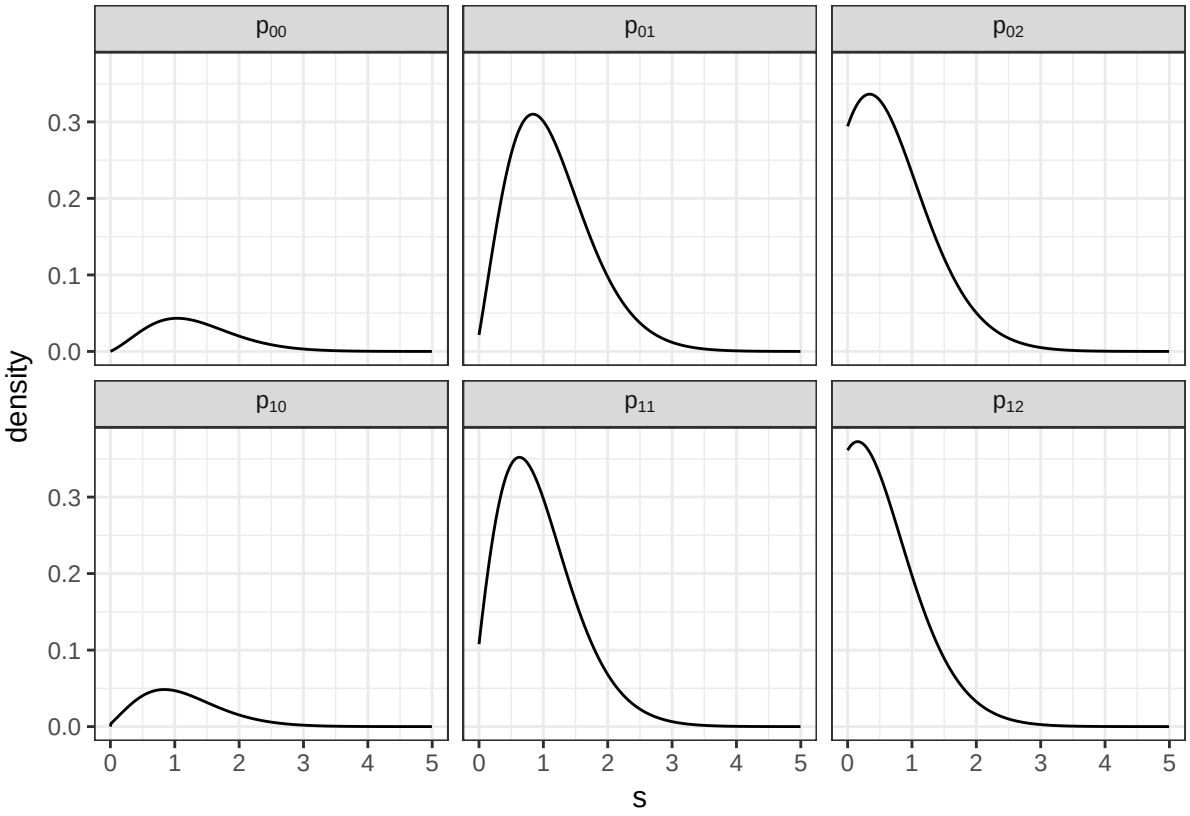

FIGURE 1. Defective densities $p_{ij}(s,t)$: Theorem 1

4. Likelihood Estimation with Forward Algorithm

Estimating the parameters of a discretely observed continuous-time process driven by an on-off process (like $S(t)$) is of practical importance but challenging. Besides our work (Yan et al., 2014; Pozdnyakov et al., 2019), only a few authors dealt with the problem. Iacus and Yoshida (2008) and De Gregorio and Iacus (2008) proposed methods that are not based on the true likelihood, such as pseudo-maximum likelihood and moment based estimators. The estimation procedure in De Gregorio and Iacus (2011) was based on a least squares type of method. There are three key differences between our work and these works. First, on top of the on-off process itself we have an additional source of randomness, the Brownian motion. Second, out data are not assumed to be collected at equidistant times. Equidistance is a very restrictive assumption as the animal movement data contain observations that are often irregularly spaced. Finally, asymptotic normality and consistency of our maximum likelihood estimator (MLE) are not addressed in our work. The aforementioned works do give asymptotic results for their estimators but, in order to prove consistency and normality, they need to assume that the time distance between two consecutive observations goes to 0. This assumption is not realistic in the case of animal tracking data. The number of observations (or the time horizon) might be large, but the time intervals between observations are not getting smaller.

Our technique is of an HMM type, but there are two important distinctions from the classic HMM. First, because of the irregular time spacing, the resulted discrete time Markov process is not time-homogeneous (the transitional probability depends on time). Furthermore, in a classical HMM, given a current hidden state, the corresponding observation is independent of the previous

hidden state; this does not hold in our situation. As a consequence, the Baum–Welch algorithm (e.g., Zucchini et al., 2016), which is a special case of the expectation-maximization (EM) algorithm, does not apply in this case. The forward algorithm, however, can still be constructed. The construction relies heavily on formulas for occupation times (Theorem 1). This underscores a practical importance of the numerous theoretical results on occupation times cited in the introduction.

Assume that we observe the MRH process $X(t)$ at times $0 = t_0 < t_1 < \cdots < t_n$. Let $\mathbf{X} = \big(X_1, X_2, \ldots, X_n\big)$, where $X_i = X(t_i) - X(t_{i-1})$, $i = 1, \ldots, n$ are the observed increments of the MRH process. Let $\mathbf{S} = \big(S(0), S(t_1), \ldots, S(t_n)\big)$ be the corresponding states of the continuous-time Markov Chain, and $\Delta_i = t_i - t_{i-1}$, $i = 1, \ldots, n$. To derive the likelihood of the vector of observed increments $\mathbf{X}$, we need a joint distribution of $X(t)$ and $S(t)$ first.

Let us work out the details the formula for $P_0(X(t) \in \mathrm{d}x, S(t) = 0)$. Fix $0 < s < t$. Given $M(t) = s$, random variable $X(t)$ has a normal distribution with mean 0 and variance $\sigma^2 s$, because Markov Chain $S(\cdot)$ and Brownian Motion $B(\cdot)$ are independent processes. Let $\phi(\cdot, \sigma^2)$ denote the pdf of a normal random variable with mean zero and variance σ^2. Then we get

$$P_0(X(t) \in \mathrm{d}x, S(t) = 0, M(t) \in \mathrm{d}s) = \phi(x, \sigma^2 s) p_{00}(s, t) \mathrm{d}x \mathrm{d}s.$$

Now, recall also that given $S(0) = 0$, random variable $M(t)$ has an atom (with weight $e^{-\lambda_0 t}$ at $s = t$). Therefore, when we integrate s out of the joint distribution of $X(t)$, $S(t)$ and $M(t)$, we get that

$$h_{00}(x, t) = P_0(X(t) \in \mathrm{d}x, S(t) = 0)/\mathrm{d}x = e^{-\lambda_0 t} \phi(x, \sigma^2 t) + \int_0^t \phi(x, \sigma^2 s) p_{00}(s, t) \mathrm{d}s. \tag{14}$$

In a similar fashion, one can show that for $i = 1, 2$

$$h_{0i}(x, t) = P_0(X(t) \in \mathrm{d}x, S(t) = i)/\mathrm{d}x = \int_0^t \phi(x, \sigma^2 s) p_{0i}(s, t) \mathrm{d}s. \tag{15}$$

When $S(0) = 1$, the distribution of random variable $X(t)$ has an atom at $x = 0$ (if $R_0 > t$, that is, the Markov chain stays in state 1 till time t). Taking this into an account we have the following formulas:

$$h_{1i}(x, t) = P_1(X(t) \in \mathrm{d}x, S(t) = i)/\mathrm{d}x = \int_0^t \phi(x, \sigma^2 s) p_{1i}(s, t) \mathrm{d}s, \quad \text{if } x \neq 0 \text{ and } i = 0, 1, 2, \tag{16}$$

and

$$P_1(X(t) = 0, S(t) = 1) = e^{-\lambda_1 t}.$$

Similarly,

$$h_{2i}(x, t) = P_2(X(t) \in \mathrm{d}x, S(t) = i)/\mathrm{d}x = \int_0^t \phi(x, \sigma^2 s) p_{2i}(s, t) \mathrm{d}s, \quad \text{if } x \neq 0 \text{ and } i = 0, 1, 2, \tag{17}$$

and

$$P_2(X(t) = 0, S(t) = 2) = e^{-\lambda_2 t}.$$

It is not essential to use one-dimensional Brownian Motion for these derivations but it simplifies our presentation's notation. If one does want to consider a Brownian Motion of d-dimension, then all we need to do is to substitute the one-dimensional normal pdf in formulas (14)–(17) by the d-dimensional normal density with mean zero and covariance matrix $\sigma^2 I_d$, where I_d is the d-dimensional identity matrix. Of course, in this case x is a vector in the d-dimensional space, not a scalar. In fact, later when we run simulations and analyze real-world data we will use the two-dimensional setup.

Next, the location-state process $\{X(t), S(t)\}$ is Markov, so the likelihood function of $(\mathbf{X}, \mathbf{S})$ is available in closed-form. More specifically, it is given by

$$L(\mathbf{X}, \mathbf{S}, \boldsymbol{\theta}) = \nu(S(0)) \prod_{i=1}^{n} f\big(X_i, S(t_i) | S(t_{i-1}), \Delta_i, \boldsymbol{\theta}\big), \tag{18}$$

where

$$f\big(x, u | v, t, \boldsymbol{\theta}\big) = \begin{cases} 0 & v \neq u,\ x = 0, \\ 0 & v = u = 0,\ x = 0, \\ e^{-\lambda_1 t} & v = u = 1,\ x = 0, \\ e^{-\lambda_2 t} & v = u = 2,\ x = 0, \\ h_{ij}\big(x, t\big) & v = i,\ u = j,\ x \neq 0, \end{cases} \tag{19}$$

$x \in \mathbf{R}$, $u, v = 0, 1, 2$, $t > 0$, and $\boldsymbol{\theta} = (\lambda_0, \lambda_1, \lambda_2, p_1, \sigma)$.

The distribution of the increments of the MRH process is a mixture of absolutely continuous and discrete distributions. Therefore, in order to construct the likelihood function we have to use the Radon–Nikodym derivative of the probability distribution relative to a dominating measure that includes an atom at $x = 0$. That explains the special sets of formulas in the case when $x = 0$.

Now, if the state vector S_t is not observed, then obviously the likelihood of the increment vector $\mathbf{X}$ can be computed using

$$L(\mathbf{X}, \boldsymbol{\theta}) = \sum_{s_0, \ldots, s_n} L(\mathbf{X}, (s_0, \ldots, s_n), \boldsymbol{\theta}),$$

where the summation is taken over all possible trajectories of $\mathbf{S}$. However, this formula is not practical since the number of trajectories grows exponentially as sample size $n \to \infty$. This difficulty is addressed with help of the dynamic programming, namely, the forward algorithm. This algorithm has a linear complexity with respect to n. For a detailed description of the construction of the algorithm in our context, we refer the readers to Pozdnyakov et al. (2019). The adjustments that have to be made are quite straightforward. Here we have three hidden states instead of two, and we have a different transition probability function given by (19).

The forward algorithm can be easily adapted to a situation when some states are completely observed or partially observed. For example, accelerometer data might be used to infer when an animal is moving or not, and direct inspection of a kill-site can confirm handling. If state s_k is known, then first calculate three kth forward variables as usual. Next, set the two forward variables with unobservable states to zero. After that just continue the forward algorithm in the normal fashion until the next location where additional information on the state is available. If at kth location only one state is excluded, then we have to set only one forward variable to zero.

The MLE can be obtained by maximizing the likelihood function facilitated with the forward algorithm. For the MRH model, however, finding the MLE is extremely computing intensive. First, evaluation of the terms in Theorem 1 involves infinite series that are computationally demanding. Second, evaluation of the terms in the likelihood is also very expensive as functions in (19) are numerical integrals of $p_{ij}(s, t)$. In our investigation, evaluating the likelihood once for a dataset of $n = 200$ takes about 30 minutes on a computer with 3.4 GHz CPU, which is why we could only afford small simulation studies in Section 6.

The asymptotic properties the MLE are open questions. Cappé et al. (2005) demonstrates that under certain regularity conditions the MLEs of standard HMM parameters are consistent, asymptotically normal, and asymptotically efficient. As pointed out in Zucchini et al. (2016), however, those conditions do not hold for many practical models, and the asymptotic normality is achieved when the sample size is very large. Similar to what has been in practice, we conjecture that the asymptotic variance matrix of the MLE can be estimated by inverting the Fisher information

matrix. Specifically, we numerically evaluate the observed Fisher information matrix, that is, the Hessian matrix or second derivative of the negative log-likelihood function with respect to the parameters, by the finite difference method at the MLE and use its inverse as the variance estimator. Inverting the observed Fisher information is known to be better than inverting the expected Fisher information, even when the latter can be evaluated, in approximating the variance of the MLE (Efron and Hinkley, 1978). Alternatively, the variance matrix of the MLE can be estimated by parametric bootstrap (Efron and Tibshirani, 1994), where each bootstrap sample is generated on the same time grid as the observed data using the fitted parameters. In Section 6 we will demonstrate (via simulations) in a one-parameter situation that the average of standard error based on the Fisher information are consistent with the empirical one.

5. Prediction of Hidden States: Marginal Probability and Viterbi Path

Another interesting and practical question that can also be addressed by dynamic programming is prediction of hidden states. The prediction question can be formulated in different ways. We consider here two problems.

5.1. **Marginal probability.** First, given parameter set $\boldsymbol{\theta}$ and observations $\mathbf{X}$, what is $\Pr(S(t_k) = i|\mathbf{X})$, the conditional probability that the kth state is equal to i ($i = 0, 1, 2$)? As mentioned near the end of the previous section,

$$L(\mathbf{X}, k, i, \boldsymbol{\theta}) = \sum_{s_0,\ldots,s_n;s_k=i} L(\mathbf{X}, (s_0, \ldots, s_n), \boldsymbol{\theta})$$

can be efficiently computed by the forward algorithm, and then

$$\Pr(S(t_k) = i|\mathbf{X}) = L(\mathbf{X}, k, i, \boldsymbol{\theta})/L(\mathbf{X}, \boldsymbol{\theta})$$

gives us the answer. Nonetheless, if we need a prediction of another state, the forward algorithm must be run again.

This can be avoided by complementing the forward variables with backward ones. More specifically, in addition to the forward variables given by

$$\alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta}) = \sum_{s_0,\ldots,s_{k-1}} \nu(s_0) \prod_{i=1}^{k} f\big(X_i, s_i|s_{i-1}, \Delta_i, \boldsymbol{\theta}\big), \tag{20}$$

where $\mathbf{X}_k = \big(X_1, X_2, \ldots, X_k\big)$, and $1 \le k \le n$, we also introduce the forward variables:

$$\beta(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta}) = \sum_{s_{k+1},\ldots,s_n} \prod_{i=k+1}^{n} f\big(X_i, s_i|s_{i-1}, \Delta_i, \boldsymbol{\theta}\big), \tag{21}$$

where $\tilde{\mathbf{X}}_k = \big(X_{k+1}, X_{k+2}, \ldots, X_n\big)$, and $1 \le k < n$. Note that the forward variable is the likelihood of observing $\mathbf{X}_k$ *and* $S(t_k) = s_k$, and the backward variable is the likelihood of observing $\bar{\mathbf{X}}_k$ *given* that $S(t_k) = s_k$.

Then one can show that that the forward and backward variables satisfy the following recursive formulas:

$$\alpha(\mathbf{X}_{k+1}, s_{k+1}, \boldsymbol{\theta}) = \sum_{s_k} f\big(X_{k+1}, s_{k+1}|s_k, \Delta_{k+1}, \boldsymbol{\theta}\big)\alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta}), \tag{22}$$

and

$$\beta(\tilde{\mathbf{X}}_{k-1}, s_{k-1}, \boldsymbol{\theta}) = \sum_{s_k} f\big(X_k, s_k|s_{k-1}, \Delta_{k-1}, \boldsymbol{\theta}\big)\beta(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta}). \tag{23}$$

The key observation is that the product of forward and backward variables is the likelihood of observing $(\mathbf{X}, S(t_k) = i)$. More specifically, we have that

$$L(\mathbf{X}, k, i, \boldsymbol{\theta}) = \alpha(\mathbf{X}_k, i, \boldsymbol{\theta})\beta(\tilde{\mathbf{X}}_k, i, \boldsymbol{\theta}).$$

But we are not done yet. The problem is that for large k the forward and backward variables might be numerically indistinguishable from zero. To address the underflow issue we need to introduce normalized forward and backward variables. Specifically, the normalized forward variables are defined by

$$\bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}) = \frac{\alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta})}{L(\mathbf{X}_k, \boldsymbol{\theta})}, \tag{24}$$

where $L(\mathbf{X}_k, \boldsymbol{\theta}) = \sum_{s_k} \alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta})$, the likelihood of vector $\mathbf{X}_k$. Then (22) immediately implies that the normalized forward variables satisfy the following equation:

$$\bar{\alpha}(\mathbf{X}_{k+1}, s_{k+1}, \boldsymbol{\theta}) = \frac{L(\mathbf{X}_k, \boldsymbol{\theta})}{L(\mathbf{X}_{k+1}, \boldsymbol{\theta})} \sum_{s_k} f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big)\bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}).$$

If for $0 \le k \le n-1$ we define

$$d(\mathbf{X}_{k+1}, \boldsymbol{\theta}) = \frac{L(\mathbf{X}_{k+1}, \boldsymbol{\theta})}{L(\mathbf{X}_k, \boldsymbol{\theta})},$$

then one can easily verify that

$$d(\mathbf{X}_{k+1}, \boldsymbol{\theta}) = \sum_{s_{k+1}} \sum_{s_k} f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big)\bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}).$$

Now, the normalized backward variables are defined by

$$\bar{\beta}(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta}) = \frac{L(\mathbf{X}_k, \boldsymbol{\theta})}{L(\mathbf{X}, \boldsymbol{\theta})}\beta(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta}). \tag{25}$$

Equation (23) immediately gives us that

$$\bar{\beta}(\tilde{\mathbf{X}}_{k-1}, s_{k-1}, \boldsymbol{\theta}) = \frac{1}{d(\mathbf{X}_k, \boldsymbol{\theta})} \sum_{s_k} f\big(X_k, s_k | s_{k-1}, \Delta_{k-1}, \boldsymbol{\theta}\big)\bar{\beta}(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta}). \tag{26}$$

Here is the procedure for finding $\Pr(S(t_k) = i|\mathbf{X})$. First run the forward algorithm and store all the normalized forward variables, all $d(\mathbf{X}_k, \boldsymbol{\theta})$, and all $f\big(X_k, s_k | s_{k-1}, \Delta_{k-1}, \boldsymbol{\theta}\big)$. Then by employing recursive formula (26) (with starting values $\bar{\beta}(\tilde{\mathbf{X}}_n, s_n, \boldsymbol{\theta}) = 1$), calculate and store the normalized backward variables $\beta(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta})$. Now, one can easily verify that

$$\Pr(S(t_k) = i|\mathbf{X}) = \bar{\alpha}(\mathbf{X}_k, i, \boldsymbol{\theta})\bar{\beta}(\tilde{\mathbf{X}}_k, i, \boldsymbol{\theta}).$$

Note that, given observations $\mathbf{X}$ and a set of parameters $\boldsymbol{\theta}$, we need to run both forward and backward algorithms just once.

5.2. **Viterbi path.** The second prediction problem is finding the so-called Viterbi path (Viterbi, 2006) — the most likely sequence of hidden states given observed $\mathbf{X}$ and a fixed parameter set $\boldsymbol{\theta}$. The naive approach is to calculate the likelihood of every path, and then choose the one with the highest value. The complexity of this procedure is exponential with respect to sample size n. The Viterbi algorithm is a dynamic programming algorithm and is linear in n.

The key idea is to introduce the most likely paths from time 0 to time t_k that end in states $i = 0, 1, 2$, and then by using Markov property to establish a (recursive) relationship between two consecutive likelihoods. More specifically, assume that $(s_0^i, s_1^i, \ldots, s_{k-1}^i, i)$ is the most likely path that ends in the state i given observed vector $\mathbf{X}_k$. Let $L_V(k, i)$ be the likelihood of this path, that is,

$$L_V(k, i) = \Pr(S(0) = s_0^i, S(t_1) = s_1^i, \ldots, S(t_{k-1}) = s_{k-1}^i, S(t_k) = i|\mathbf{X}_k).$$

TABLE 1. Predicted probabilities and relative frequency of the states at $t_{100} = 1000$ based on 1000 replicates of an MRH process with parameters $\lambda_0 = 0.3$, $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, $\sigma = 10$, $p_1 = 0.5$, and $S(0) = 0$ on time grid $(0, 10, 20, \ldots, 2000)$.

	$S(t_{100}) = 0$	$S(t_{100}) = 1$	$S(t_{100}) = 2$
	True state		
Stationary proportion ($\nu(\cdot)$)	0.308	0.231	0.462
Empirical proportion	0.327	0.232	0.441
$\Pr(S(t_{100}) = 0\|\mathbf{X})$	0.484	0.314	0.169
$\Pr(S(t_{100}) = 1\|\mathbf{X})$	0.279	0.286	0.234
$\Pr(S(t_{100}) = 2\|\mathbf{X})$	0.237	0.400	0.597

Then

$$L_V(k+1, i) = \max_{j=0,1,2} \Big\{ L_V(k, j) f\big(X_{k+1}, i | j, \Delta_{k+1}, \boldsymbol{\theta}\big) \Big\},$$

and, respectively, the argument of this maximum gives us the next state of the Viterbi path s_k^i. Once we have all 3 paths $(s_0^i, s_1^i, \ldots, s_{n-1}^i, i)$, $i = 0, 1, 2$, we just choose the one with the highest likelihood.

5.3. **An illustration.** For illustration, consider state prediction based on the marginal probability. We used 1000 realizations of an MRH process with the following parameters: $\lambda_0 = 0.3$, $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, $\sigma = 10$, $p_1 = 0.5$, and $S(0) = 0$. The time grid was $(0, 10, 20, \ldots, 2000)$. Table 1 summarizes the marginal probabilities of the states given the observed data in comparison with the stationary probabilities and the empirical relative frequencies. Note that even though each trajectory was started from a moving state, the empirical distribution of the state process at $t_{100} = 1000$ is very close to the stationary one. This is not surprising because by t_{100} the Markov chain has enough time to enter into a stationary stage. The predicted marginal probabilities of the states at t_{100}, however, are noticeably different depending on what the true states are. For example, the average marginal probabilities $\Pr(S(t_{100}) = 0|\mathbf{X})$ evaluated for those trajectories for which $S(t_{100})$ was actually 0 is 0.484, which is much higher then the stationary probability 0.308. This excess in probability is different for different states, and it looks like it follows a very complicated pattern that depends on model parameters and the time distance between observations.

6. Numerical Studies

We ran a small simulation to demonstrate that the MLE from the forward algorithm successfully recovers the model parameters. The true parameter values were set to be $\lambda_0 = 4$, $\lambda_1 = 0.5$, $\lambda_2 = 0.1$, $p_1 = 0.8$, and $\sigma = 25$. The simulation was small because the computation of the maximum likelihood estimator is very demanding, even with the help of a large beowulf cluster of over 500 CPUs. We generated 49 two-dimensional datasets on a time grid from 0 to 4000, with increment 20, so the resulting series $\mathbf{X}$ is of length 200.

Figure 2 presents the violin plots of the likelihood estimates of the 49 replicates in comparison to the true values of the five parameters. Violin plots are similar to box plots with a rotated kernel density plot on each side, which show more information about the data than box plots. The horizontal bars in the panels are the true parameter values. For each parameter, the true value lies in the bulk part of the violin plot, indicating that the true parameters are recovered well by the likelihood estimates in this small scale simulation study.

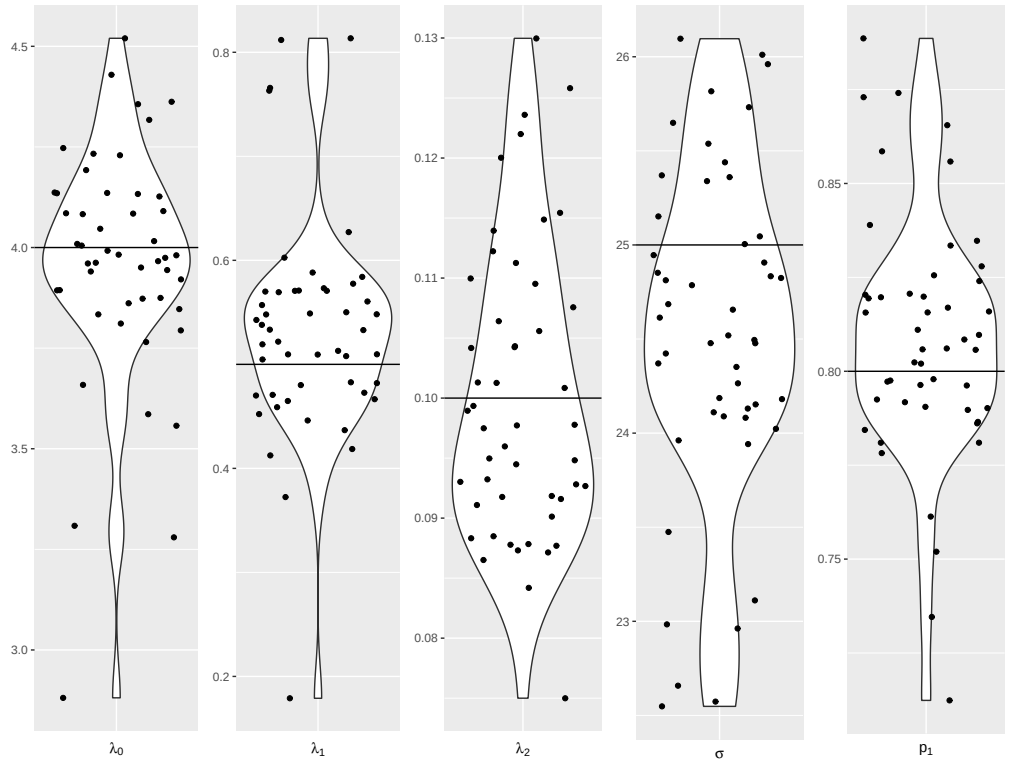

FIGURE 2. Violin plots of the maximum likelihood estimates from 49 replicates using the forward algorithm. The horizontal bar in each panel is the true parameter value.

A separate study focusing on the standard error of the estimator was conducted. In this study, we only estimate p_1 with all other parameters fixed to make the task computationally feasible within a reasonable period of time. The parameters were set the same as in the last study. We generated 100 two-dimensional datasets, each of which was generated on a time grid from 0 to 400 with equal increment 2. The sampling frequency in this study is much higher than in the last study to make the estimation task easier for computational feasibility. For each datasets, standard errors of the MLE were obtained from both Fisher information and parametric bootstrap. In the parametric bootstrap procedure, bootstrap sample size was set to be 25, and time grid remained the same as that in the observed data. Over the 100 replicates, the averaged estimate for p_1 is 0.083 and the empirical standard error is 0.096. The average standard errors from the Fisher information and parametric bootstrap are, respectively, 0.088 and 0.095. The Fisher information approach under-estimates the standard error; a larger sample size is needed for better agreement. On the other hand, the bootstrap approach estimates the standard error very well with steep computing cost. This study, although limited in scope, suggests that inferences based on Fisher information or parametric bootstrap are possible. A fuller investigation merits future research.

We next applied the proposed model to the data from the same mountain lion analyzed by Yan et al. (2014) and Pozdnyakov et al. (2019). This mount lion was a mature female in the Gros Ventre Mountain Range near Jackson Wyoming tracked with a GPS collar from 2009 to 2012. The collar was designed to collect a fix every 8 hours but the actual sampling times were irregular with sampling intervals having standard deviation 5.37 hours, ranging from 0.5 hours to 44 hours. Mountain lions behave differently in the summer and in the winter, so we focused on the summer of 2012, a total of 398 observations spanning from June 1 to August 31. This makes our results

TABLE 2. Parameter estimates and model selection criteria from the mountain lion data analysis with the MRH process and the MR process. The standard errors were obtained from inverting the observed Fisher information matrix.

Parameters	MRH		MR	
	estimate	standard error	estimate	standard error
λ_0	8.402	2.271	3.016	0.320
λ_1	1.324	0.250	0.207	0.013
λ_2	0.130	0.024	—	—
σ	1.563	0.077	1.305	0.064
p	0.685	0.159	—	—
log-likelihood	−982.6		−1021.5	
AIC	1975.2		2049.0	
BIC	1995.1		2061.0	

not directly comparable to existing analyses (Yan et al., 2014; Pozdnyakov et al., 2019). Field personnel determined that some of the sites were places where the mountain lion consumed a prey item. She typically remained within 250 m of a kill site while it was considered to be "handling", which is different from shorter, resting periods. To allow for GPS measurement error, we rounded the locations to the nearest 100 meters.

Table 2 summarizes the MLEs fitted from the MRH process with standard errors obtained from inverting the observed Fisher information matrix. The estimates of the rates of the exponential holding times suggest that, on average, the mountain lions stays for 0.119, 0.755, and 7.692 hours in the moving, resting, and handling states. The mobility parameter estimate means that, if the mountain lion moves without stopping for one hour, the average deviation from the initial position in terms of northing and easting values is 1.563 km. When she stopped moving, she went into resting with probability 0.685 and handling with probability 0.315, respectively. Also reported in the table are the results from fitting the MR process (Yan et al., 2014; Pozdnyakov et al., 2019). Because there is no handling state, the average durations in both moving and resting are estimated longer at 0.332 and 4.831 hours, respectively. The mobility parameter estimate is 17% lower than that in the MRH model, because the animal was assumed to be moving longer. The MRH model has a log-likelihood 139.9 higher than the MR model, which is well worth the two additional parameters. Based on either the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), the MRH model is very strongly preferred to the MR model. We also fitted the BBMM (Horne et al., 2007; Pozdnyakov et al., 2014) with the original, non-rounded data and the GPS measurement error standard deviation fixed at 0.02km. The BM mobility parameter estimate is much lower, 0.553km/hour$^{1/2}$, about 35% of that in the MRH model, as the animal was assumed to be always moving.

7. Concluding Remarks

The results on occupation times obtained in the paper have their own value and can be used for other applications, such as quality control. Indeed, the continuous-time Markov Chain $S(t)$ can be viewed as a telegraph process with two off states. These two states will correspond two different types of breakdown that require different time for repair. The results in Theorem 1 can be easily generalized to cover k motionless states instead of just two. The only difference is that, instead of binomial distribution and convolutions of two gamma distributions, we will have multinomial distribution and convolutions of k gammas.

The methodology developed in Sections 3 works even if the holding times are not exponentially distributed, which is an advantage of our approach. If we want to keep the Markov property, then all holding times *must* have exponential distributions. The memoryless distribution might be not appropriate for some species that follow a cyclic daily routine. Nonetheless, if animals under observation do not exhibit a daily periodic behavior (like mountain lions), then using an exponential distribution is acceptable. The behavior of these animals is subject to interruptions that can cut their time spent in a particular activity. For example, handling might be interrupted by a more dominate predator who drives the lion off her kill before she is finished with it.

A different (from exponential) distribution should be used for species with a periodic routine. One interesting possibility is to employ stable distributions (for example, Lévy distribution). Because a linear combination of two independent random variables with a stable distribution has the same distribution, up to location and scale parameters, the formulas in Theorem 1 will be even nicer. The drawback is that the state process is then semi-Markov, and, as a result, the likelihood inferences from standard HMM tools are not available. Nevertheless, this still might be of interest for practitioners in ecological science, because estimation can be done via alternative methods such as the composite likelihood estimation (Lindsay, 1988).

References

D. Bshouty, A. Di Crescenzo, B. Martinucci and S. Zacks (2012). "Generalized telegraph process with random delays." *Journal of Applied Probability* **49**, 850–865.

O. Cappé, E. Moulines and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer.

E. A. Codling, M. J. Plank and S. Benhamou (2008). "Random walk models in biology." *Journal of The Royal Society Interface* **5**, 813–834.

A. De Gregorio and S. M. Iacus (2008). "Parametric estimation for the standard and geometric telegraph process observed at discrete times." *Statistical Inference for Stochastic Processes* **11**, 249–263.

A. De Gregorio and S. M. Iacus (2011). "Least-squares change-point estimation for the telegraph process observed at discrete times." *Statistics* **45**, 349–359.

A. Di Crescenzo (2001). "On random motions with velocities alternating at Erlang-distributed random times." *Advances in Applied Probability* **33**, 690–701.

A. Di Crescenzo, B. Martinucci and S. Zacks (2014). "On the geometric brownian motion with alternating trend." In C. Perna and M. Sibillo (eds.), *Mathematical and Statistical Methods for Actuarial Sciences and Finance*, pp. 81–85. Dordrecht: Springer.

A. Di Crescenzo and F. Pellerey (2002). "On prices' evolutions based on geometric telegrapher's process." *Applied Stochastic Models in Business and Industry* **18**, 171–184.

A. Di Crescenzo and S. Zacks (2015). "Probability law and flow function of Brownian motion driven by a generalized telegraph process." *Methodology and Computing in Applied Probability* **17**, 761–780.

B. Efron and D. Hinkley (1978). "Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information." *Biometrika* **65**, 457–487.

B. Efron and R. Tibshirani (1994). *An Introduction to the Bootstrap*. CRC Press.

J. S. Horne, E. O. Garton, S. M. Krone and S. Lewis, J (2007). "Analyzing animal movements using Brownian bridges." *Ecology* **88**, 2354–2363.

C. Hu, V. Pozdnyakov and J. Yan (2019a). *coga: Convolution of Gamma Distributions*. R package version 1.0.0.

C. Hu, V. Pozdnyakov and J. Yan (2019b). "Density and distribution evaluation for convolution of independent gamma variables." *Computational Statistics* In press.

S. M. Iacus and N. Yoshida (2008). "Estimation for the discretely observed telegraph process." *Teorīya Ĭmovīrnosteĭ ta Matematichna Statistika* **78**, 32–42.

A. D. Kolesnik and N. Ratanov (2013). *Telegraph processes and option pricing*. Springer Briefs in Statistics. Springer, Heidelberg.

B. G. Lindsay (1988). "Composite likelihood methods." *Contemporary Mathematics* **80**, 221–239.

A. Mathai (1982). "The storage capacity of a dam with gamma type inputs." *Annals of Institute of Statistical Mathematics* **34**, 591–597.

P. Moschopoulos (1985). "The distribution of the sum of independent gamma random variables." *Annals of Institute of Statistical Mathematics* **37**, 541–544.

J. R. Norris (1998). *Markov Chains*, volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. Reprint of 1997 original.

H. G. Othmer, S. R. Dunbar and W. Alt (1988). "Models of dispersal in biological systems." *Journal of Mathematical Biology* **26**, 263–298.

D. Perry, W. Stadje and S. Zacks (1999). "First-exit times for increasing compound processes." *Communications in Statistics: Stochastic Models* **15**, 977–992.

V. Pozdnyakov, L. Elbroch, A. Labarga, T. Meyer and J. Yan (2019). "Discretely observed Brownian motion governed by telegraph process: estimation." *Methodology and Computing in Applied Probability* **21**, 907–920.

V. Pozdnyakov, T. H. Meyer, Y.-B. Wang and J. Yan (2014). "On modeling animal movements using brownian motion with measurement error." *Ecology* **95**, 247–253.

H. K. Preisler, A. A. Ager, B. K. Johnson and J. G. Kie (2004). "Modeling animal movements using stochastic differential equations." *Environmetrics* **15**, 643–657.

B. Sericola (2000). "Occupation times in markov processes." *Communications in Statistics. Stochastic Models* **16**, 479–510.

W. Stadje and S. Zacks (2004). "Telegraph processes with random velocities." *Journal of Applied Probability* **41**, 665–678.

P. F. C. Tilles and S. V. Petrovskii (2016). "How animals move along? exactly solvable model of superdiffusive spread resulting from animal's decision making." *Journal of Mathematical Biology* **73**, 227–55.

A. J. Viterbi (2006). "A personal history of the Viterbi algorithm." *IEEE Signal Processing Magazine* **23**, 120–142.

J. Yan, Y.-W. Chen, K. Lawrence-Apfel, I. Ortega, V. Pozdnyakov, S. Williams and T. Meyer (2014). "A moving-resting process with an embedded Brownian motion for animal movements." *Population Ecology* **56**, 401–415.

J. Yan, V. Pozdnyakov and C. Hu (2019). *smam: Statistical Modeling of Animal Movements*. R package version 0.4.0.

S. Zacks (2004). "Generalized integrated telegraph processes and the distribution of related stopping times." *Journal of Applied Probability* **41**, 497–507.

S. Zacks (2012). "Distribution of the total time in a mode of an alternating renewal process with applications." *Sequential Analysis* **31**, 397–408.

W. Zucchini, MacDonald, I. L. and R. Langrock (2016). *Hidden Markov Models for Time Series: An Introduction Using R*. Chapman and Hall/CRC, 2 edition.