DISCRETELY OBSERVED BROWNIAN MOTION GOVERNED BY TELEGRAPH PROCESS: ESTIMATION

VLADIMIR POZDNYAKOV[1], L. MARK ELBROCH[2], ANTHONY LABARGA[1], THOMAS MEYER[3,4], AND JUN YAN[1,3]

Abstract. A Brownian motion whose infinitesimal variance alternates according to a telegraph process is considered. This stochastic process can be employed to model a variety of real-word situations, such as animal movement in ecology and stochastic volatility in mathematical finance. The main goal is to develop an estimation procedure for the underlying model parameters when the process is observed at discrete, possibly irregularly spaced time points. The sequence of observations is not Markov, but the sequence of the state of the telegraph process, if observed, is Markov. The observed sequence is therefore from a hidden Markov model. Likelihood inference is developed via dynamic programming, and is demonstrated to have much higher efficiency than the composite likelihood approach that was applied in an earlier work. The model is applied to model the movement of a mountain lion.

Keywords: Markov process · Dynamic programming · Likelihood estimation

Introduction

Brownian motion (BM) and random walks are often employed by ecologists to model animal movement (Preisler et al., 2004). Horne et al. (2007) introduced the Brownian bridge movement model (BBMM) which characterizes the missing movement path between two sequential positions by a Brownian bridge. The BBMM quickly gained popularity in the ecological community (e.g., Lonergan et al., 2009; Willems and Hill, 2009; Farmer et al., 2010; Takekawa et al., 2010). The standard statistical estimation approach for the BBMM, which is different from the original procedure in Horne et al. (2007) procedure, can be found in Pozdnyakov et al. (2014). The BBMM has been implemented in R packages such as `BBMM` (Nielson et al., 2012) and `adehabitat` (Calenge, 2006). Recently, a dynamic version of the BBMM has been proposed to allow the BM variance to be time-dependent (Kranstauber et al., 2012). An implementation of the dynamic BBMM is available in R package `move` (Kranstauber and Smolla, 2013).

The Brownian Motion governed by telegraph process (BMT process) is a natural generalization of the BBMM that allows two modes of movements (e.g., moving and resting, low speed and high speed). The moving-resting process (Yan et al., 2014) allows an animal to have two states, moving and resting; if in the moving stage, the motion is characterized by a BM; and the duration times in either moving or resting

---

1. Department of Statistics, University of Connecticut.

2. Panthera.

3. Center for Environmental Sciences and Engineering, University of Connecticut.

4. Department of Natural Resources and the Environment, University of Connecticut.

states are exponentially distributed. Another important application is mathematical finance. The BM with alternated trends is used in 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). The BMT process can be employed to model stock price trajectories with stochastic volatility (e.g., Fouque et al., 2011).

The telegraph process, also known as the alternating renewal process or the on-off process, was first studied by Cane (1959) and Page (1960) in the context of animal ethology and maintenance of electronic equipment, respectively. The telegraph process is a non-trivial probabilistic object, and as a result, this area of research is still quite active. In particular, in our analysis we used results from Perry et al. (1999), Stadje and Zacks (2004), Zacks (2004), and Di Crescenzo et al. (2005). For tractability and efficient computation, the time periods of the telegraph process are assumed to be exponentially distributed. What can be done in a more general situation is an open question.

Observations of a continuous-time BMT process are usually only available in discrete, possibly irregularly spaced time points, which make it unnatural to apply the state space model (SSM) (e.g., Jonsen et al., 2005; Patterson et al., 2008). Estimation of BMT parameters is a challenging problem because the on-off states are unobserved, and the observed sequence is not Markov. Yan et al. (2014) estimated the parameters by maximizing a composite likelihood constructed from the marginal distribution of each increment. Here we develop inferences based on the true likelihood after establishing that the BMT process is a hidden Markov model (HMM). Then dynamic programming or the forward algorithm for HMM is used to construct the true likelihood (e.g., Cappé et al., 2005). In our simulation studies, the true likelihood inference is much more efficient than the composite likelihood approach. The model is then applied to model the movement of a mountain lion (*Felis concolor*).

State and BMT processes

The different phases of BM are modeled by an telegraph process with exponentially distributed holding times. More specifically, let $\{M_i\}_{i\geq 1}$ be independent and identically distributed (i.i.d.) random variables with exponential distribution with mean $1/\lambda_1$, and $\{R_i\}_{i\geq 1}$ be i.i.d. random variables with exponential distribution with mean $1/\lambda_0$. Assume that $\{M_i\}_{i\geq 1}$ and $\{R_i\}_{i\geq 1}$ are independent.

Consider a telegraph process that, with probability p_1, $0 \leq p_1 \leq 1$, starts with a *1-cycle* (i.e., we have $M_1, R_1, M_2, R_2, \ldots$) and with probability $p_0 = 1 - p_1$ starts with *0-cycle* (i.e., we have $R_1, M_1, R_2, M_2, \ldots$).

Here we assume that starting probabilities are equal to stationary ones, i.e.,

$$p_1 = \frac{\lambda_0}{\lambda_1 + \lambda_0} \quad \text{and} \quad p_0 = \frac{\lambda_1}{\lambda_1 + \lambda_0}.$$

Let $S(t)$, $t \geq 0$, be the state process; that is, $S(t) = 1$ if the telegraph process is in a 1-cycle and $S(t) = 0$ if the process is in a 0-cycle at time t. Let $X(t)$ be BMT process indexed by time $t > 0$. Conditioning on the state of the underlying renewal process, $S(t)$, the process $X(t)$ is defined by the stochastic differential equation

$$
\text{(1)} \qquad \mathrm{d}X(t) = [(\sigma_1 - \sigma_0)S(t) + \sigma_0]\mathrm{d}B(t) = \begin{cases} \sigma_1 \mathrm{d}B(t), & \text{if } S(t) = 1, \\ \sigma_0 \mathrm{d}B(t), & \text{if } S(t) = 0, \end{cases},
$$

where σ_1 and σ_0 are volatility parameters, and $B(t)$ is the standard Brownian motion.

Here we consider two cases.

Case 1: both $\sigma_1 > 0$ and $\sigma_0 > 0$.

Case 2: $\sigma_1 = \sigma > 0$ and $\sigma_0 = 0$.

One significant mathematical difference between the two cases is that the distribution of the increments of the BMT process in the second case is a mixture of absolutely continuous and discrete distributions. Still the second case can be viewed as a limiting case of the first one. But case 2 plays an important role in analyzing animal movement, so we will provide a separate set of formulas for this situation. A typical realization of BMT process (case 2) is shown in Figure 1.

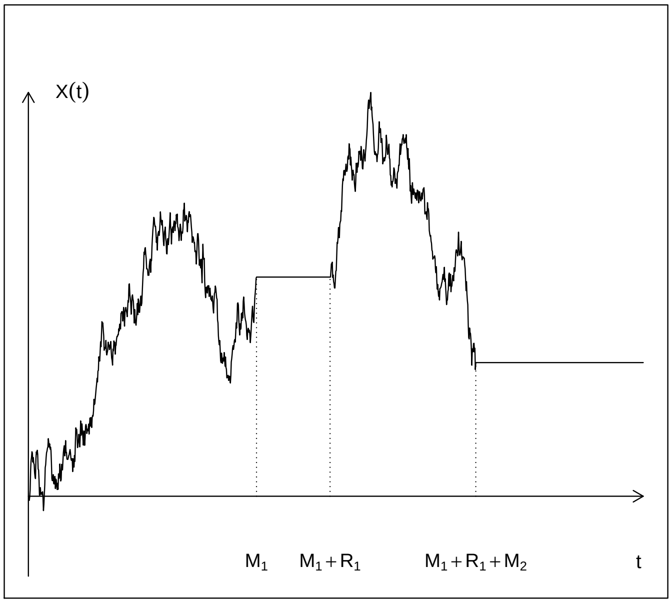

FIGURE 1. BMT Process (Case 2)

Let us note here that $X(t)$ is not Markov even though both $S(t)$ and $B(t)$ are Markov. Indeed, let us consider case 2. If right before t we observe a flat trajectory of $X(t)$ then $S(t) = 0$ and the distribution of the increment $X(t+u) - X(t)$, $u > 0$ will have an atom at 0. If it is not the case, then $X(t+u) - X(t)$ has an absolutely continuous distribution. The bottom line is that

$$\Pr\big(X(t)\in\Gamma|\mathcal{F}_s\big)\neq\Pr\big(X(t)\in\Gamma|X(s)\big),\quad 0\le s\le t,$$

where $\mathcal{F}_s$ is, as usually, a σ-field generated by $\big(S(u),B(u)\big)_{0\le u\le s}$, and Γ is a Borel set.

Nonetheless, because of the memoryless property of the exponential distribution and the Markov property of the Brownian motion, the joint process $\{X(t),S(t):t\ge 0\}$ is a Markov process with stationary increments in $X(t)$. Moreover, we will see that the distribution of the increment $X(t+u)-X(t)$ depends only on $S(t)$.

To derive distribution of increments of the BMT process $X(t)$ we need the joint distributions of the two pairs $\big(M(t),S(t)\big)$ and $\big(R(t),S(t)\big)$ for a given initial state $S(0)$, where $M(t)$ and $R(t)$, $t>0$, are the total time in interval $(0,t]$ spent in the 1-cycles and in the 0-cycles, respectively; consequently, $R(t)=t-M(t)$. It is known that in the case when durations of alternating phases are described by exponential distributions, closed-form expressions for their densities are available (e.g., Zacks, 2004, p. 500). Recent theoretical findings on telegraph processes (Perry et al., 1999; Stadje and Zacks, 2004; Zacks, 2004; Di Crescenzo et al., 2005) make possible finding closed-form formulas for the BMT process.

Following the notations in Yan et al. (2014), define

$$P_1\big[\,\cdot\,\big]=\Pr\big[\,\cdot\,|S(0)=1\big],\qquad\text{and}\qquad P_0\big[\,\cdot\,\big]=\Pr\big[\,\cdot\,|S(0)=0\big].$$

Then, for $0<w<t$, we introduce the following (defective) densities

$$\begin{aligned}p_{11}(w,t)\mathrm{d}w&=P_1\big[M(t)\in\mathrm{d}w,S(t)=1\big],\\p_{10}(w,t)\mathrm{d}w&=P_1\big[M(t)\in\mathrm{d}w,S(t)=0\big],\\p_{01}(w,t)\mathrm{d}w&=P_0\big[R(t)\in\mathrm{d}w,S(t)=1\big],\\p_{00}(w,t)\mathrm{d}w&=P_0\big[R(t)\in\mathrm{d}w,S(t)=0\big].\end{aligned}$$

According to Zacks (2004, p.500), we have that

$$\begin{aligned}p_{11}(w,t)&=e^{-\lambda_1 w-\lambda_0(t-w)}\sum_{n=1}^{\infty}\frac{\lambda_1^n\lambda_0^n}{n!(n-1)!}w^n(t-w)^{n-1},\\p_{10}(w,t)&=\lambda_1 e^{-\lambda_1 w-\lambda_0(t-w)}+\lambda_1\sum_{n=1}^{\infty}p(n,\lambda_1 w)p(n,\lambda_0(t-w)),\end{aligned}$$

where $p(n,\mu)=e^{-\mu}\mu^n/n!$ is the Poisson distribution probability mass with mean μ evaluated at n. Respectively, we have that

$$\begin{aligned}p_{00}(w,t)&=e^{-\lambda_0 w-\lambda_1(t-w)}\sum_{n=1}^{\infty}\frac{\lambda_0^n\lambda_1^n}{n!(n-1)!}w^n(t-w)^{n-1},\\p_{01}(w,t)&=\lambda_0 e^{-\lambda_0 w-\lambda_1(t-w)}+\lambda_0\sum_{n=1}^{\infty}p(n,\lambda_0 w)p(n,\lambda_1(t-w)).\end{aligned}$$

Both $M(t)$ and $R(t)$ have atoms or point masses at $w = t$. Specifically,

$$P_1\big[M(t) = t\big] = e^{-\lambda_1 t} \quad \text{and} \quad P_0\big[R(t) = t\big] = e^{-\lambda_0 t}.$$

The series in the formulas can be evaluated efficiently using the modified Bessel function of the first kind

$$I(z;\alpha) = \sum_{m=0}^{\infty} \frac{\left(\frac{z}{2}\right)^{2m+\alpha}}{m!\Gamma(m+\alpha+1)}.$$

For example, we have

$$p_{11}(w,t) = e^{-\lambda_1 w - \lambda_0(t-w)} \sqrt{\lambda_1 \lambda_0 w/(t-w)} I(z;1)$$

and

$$p_{10}(w,t) = \lambda_1 e^{-\lambda_1 w - \lambda_0(t-w)} I(z;0),$$

where $z = 2\sqrt{\lambda_1 \lambda_0 w(t-w)}$. Similar expressions can be found for p_{00} and p_{01}.

JOINT DISTRIBUTION OF THE INCREMENT OF $X(t)$ AND $S(t)$

Case 1. Without loss of generality, we assume $X(0) = 0$. Given a value of $M(t) = w$, the BMT process X(t) has the normal distribution with mean 0 and variance $\sigma_1^2 w + \sigma_0^2(t-w)$. Combining this observation with formulas from previous section one can get the joint distribution of $\big(X(t), S(t)\big)$. For example, for $0 < w < t$,

$$P_1[X(t) \in \mathrm{d}x, S(t) = 1, M(t) \in \mathrm{d}w] = \phi(x; \sigma_1^2 w + \sigma_0^2(t-w)) p_{11}(w,t) \mathrm{d}x \mathrm{d}w,$$

where $\phi(\cdot;\sigma^2)$ is the density function of a normal variable with mean zero and variance σ^2.

One can get the joint distribution of $\big(X(t), S(t)\big)$ starting from $S(0) = 1$ by integrating w out. Other scenarios can be handled in a similar fashion. For ease of notation in the estimation in the next section, let

$$\begin{aligned}
h_{11}^*(x,t) &= P_1\big[X(t) \in \mathrm{d}x, S(t) = 1\big] \\
&= e^{-\lambda_1 t} \phi(x; \sigma_1^2 t) + \int_0^t \phi(x; \sigma_1^2 w + \sigma_0^2(t-w)) p_{11}(w,t) \mathrm{d}w, \\
h_{10}^*(x,t) &= P_1\big[X(t) \in \mathrm{d}x, S(t) = 0\big] \\
&= \int_0^t \phi(x; \sigma_1^2 w + \sigma_0^2(t-w)) p_{10}(w,t) \mathrm{d}w, \\
h_{00}^*(x,t) &= P_0\big[X(t) \in \mathrm{d}x, S(t) = 0\big] \\
&= e^{-\lambda_0 t} \phi(x; \sigma_0^2 t) + \int_0^t \phi\left(x; \sigma_0^2 w + \sigma_1^2(t-w)\right) p_{00}(w,t) \mathrm{d}w, \\
h_{01}^*(x,t) &= P_0\big[X(t) \in \mathrm{d}x, S(t) = 1\big] \\
&= \int_0^t \phi(x; \sigma_0^2 w + \sigma_1^2(t-w)) p_{01}(w,t) \mathrm{d}w.
\end{aligned}$$

Case 2. The difference here is that in this case $X(t)$ has a singularity at 0. However, the rest of the argument is exactly the same. More specifically, if we introduce the following functions:

$$\begin{aligned}
h_{11}(x,t) &= e^{-\lambda_1 t}\phi(x;\sigma^2 t) + \int_0^t \phi(x;\sigma^2 w)p_{11}(w,t)\mathrm{d}w,\\
h_{10}(x,t) &= \int_0^t \phi(x;\sigma^2 w)p_{10}(w,t)\mathrm{d}w,\\
h_{00}(x,t) &= \int_0^t \phi\left(x;\sigma^2(t-w)\right)p_{00}(w,t)\mathrm{d}w,\\
h_{01}(x,t) &= \int_0^t \phi\left(x;\sigma^2(t-w)\right)p_{01}(w,t)\mathrm{d}w,
\end{aligned}$$

then we have

$$\begin{aligned}
P_1\big[X(t)\in \mathrm{d}x, S(t)=1\big] &= h_{11}(x,t)\mathrm{d}x,\\
P_1\big[X(t)\in \mathrm{d}x, S(t)=0\big] &= h_{10}(x,t)\mathrm{d}x,\\
P_0\big[X(t)\in \mathrm{d}x, S(t)=0\big] &= h_{00}(x,t)\mathrm{d}x + e^{-\lambda_0 t}\delta_0(x),\\
P_0\big[X(t)\in \mathrm{d}x, S(t)=1\big] &= h_{01}(x,t)\mathrm{d}x,
\end{aligned}$$

where $\delta_0(x)$ is the delta function with an atom at 0. The extra part in $P_0\big[X(t)\in \mathrm{d}x, S(t)=0\big]$, $e^{-\lambda_0 t}\delta_0(x)$, is the probability that the entire time period $(0,t]$ is in a 0-phase. The additional term in $h_{11}(x,t)$, $e^{-\lambda_1 t}\phi(x;\sigma^2 t)$, comes from the possibility that the whole $(0,t]$ interval is 1-cycle.

PARAMETER ESTIMATION: CASE 1

Assume that a BMT process $X(t)$ with parameters $\theta=(\lambda_0,\lambda_1,\sigma_0,\sigma_1)$ is observed at times $0=t_0,t_1,\ldots,t_n$. Let $\mathbf{X}=\big(X(0),X(t_1),\ldots,X(t_n)\big)$ be the observed locations. Let $\mathbf{S}=\big(S(0),S(t_1),\ldots,S(t_n)\big)$ be the states of the underlying telegraph process (that are not observable).

Now, if the state process $\mathbf{S}$ is observed, then the full likelihood is available in closed-form because the joint process $\big(X(t),S(t)\big)$ is Markovian. With $X(0)=0$, the transition density of $\big(X(t),S(t)\big)$ is

$$f^*\big(x(t),s(t)|s(0),\theta\big) = \begin{cases} h^*_{11}\big(x(t),t\big) & s(0)=1,\ s(t)=1,\\ h^*_{10}\big(x(t),t\big) & s(0)=1,\ s(t)=0,\\ h^*_{01}\big(x(t),t\big) & s(0)=0,\ s(t)=1,\\ h^*_{00}\big(x(t),t\big) & s(0)=0,\ s(t)=0,\end{cases}$$

where the h^*'s are defined in the previous section. The likelihood function of the observed data is then

$$L(\mathbf{X},\mathbf{S},\theta) = \nu(S(0))\prod_{i=1}^{n} f^*\big(X(t_i)-X(t_{i-1}),S(t_i)|S(t_{i-1}),\theta\big), \tag{2}$$

where $\nu(\cdot)$ is initial distribution (that is, $\nu(0)=p_0$, and $\nu(1)=p_1$).

Maximizing (2) with respect to θ gives us the maximum likelihood estimator (MLE) $\hat{\theta}_n$. The usual properties of $\hat{\theta}_n$ (consistency, asymptotic normality, and asymptotic efficiency) hold for the estimator. The variance of $\hat{\theta}_n$ can be estimated from the inverse of the Fisher information matrix.

In practice, it is more likely that $\mathbf{S}$ is not observed. In this case, we need to work with an $\mathbf{X}$ likelihood function that is given by

$$L(\mathbf{X}, \theta) = \Pr\big(X(t_1) - X(t_0) \in \mathrm{d}x_1, \ldots, X(t_n) - X(t_{n-1}) \in \mathrm{d}x_n\big). \tag{3}$$

Because the observed process $X(t)$ itself is not Markovian, formulas similar to (2) are not available. A naïve approach would be to use

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

Here the summation is taken over all possible trajectories of $\mathbf{S}$. The total number of such trajectories is 2^{n+1} so this approach will not work for any real-world data set.

To address this issue, we notice that $\mathbf{X}$ can be viewed as observations from a hidden Markov model (HMM), and use the dynamic programming method developed in the HMM literature. First, we introduce so-called forward variables:

$$\alpha(\mathbf{X}_k, s_k, \theta) = \sum_{s_0,\ldots,s_{k-1}} \nu(s_0) \prod_{i=1}^{k} f^*\big(X(t_i) - X(t_{i-1}), s_i | s_{i-1}, \theta\big), \tag{4}$$

where $\mathbf{X}_k = \big(X(0), X(t_1), \ldots, X(t_k)\big)$, and $1 \le k \le n$. It is easy to see that the forward variable is the likelihood of observing $\mathbf{X}_k$ and s_k.

To simplify the notation let us denote the increments of $\mathbf{X}$, $X(t_{k+1}) - X(t_k)$, by $Y(t_{k+1})$. Then, using, in essence, the Markov property of $\big(X(t), S(t)\big)$, we get

$$\begin{aligned}
\alpha(\mathbf{X}_{k+1}, s_{k+1}, \theta) &= \sum_{s_0,\ldots,s_k} \nu(s_0) \prod_{i=1}^{k+1} f^*\big(Y(t_i), s_i | s_{i-1}, \theta\big) \\
&= \sum_{s_0,\ldots,s_{k-1}} \sum_{s_k} \nu(s_0) \prod_{i=1}^{k+1} f^*\big(Y(t_i), s_i | s_{i-1}, \theta\big) \\
&= \sum_{s_0,\ldots,s_{k-1}} \sum_{s_k} f^*\big(Y(t_{k+1}), s_{k+1} | s_k, \theta\big) \nu(s_0) \prod_{i=1}^{k} f^*\big(Y(t_i), s_i | s_{i-1}, \theta\big) \\
&= \sum_{s_k} f^*\big(Y(t_{k+1}), s_{k+1} | s_k, \theta\big) \times \sum_{s_0,\ldots,s_{k-1}} \nu(s_0) \prod_{i=1}^{k} f^*\big(Y(t_i), s_i | s_{i-1}, \theta\big) \\
&= \sum_{s_k} f^*\big(Y(t_{k+1}), s_{k+1} | s_k, \theta\big) \alpha(\mathbf{X}_k, s_k, \theta).
\end{aligned}$$

Finally, we obviously have

$$L(\mathbf{X}, \theta) = \sum_{s_n} \alpha(\mathbf{X}_n, s_n, \theta),$$

and

$$\alpha(\mathbf{X}_0, s_0, \theta) = \nu(s_0).$$

The transition from $\alpha(\mathbf{X}_k, s_k, \theta)$ to $\alpha(\mathbf{X}_{k+1}, s_{k+1}, \theta)$ can be performed in $O(1)$ time so we get an algorithm that finds $L(\mathbf{X}, \theta)$ with computational complexity of $O(n)$.

Here is the forward algorithm.

(1) For observed $\mathbf{X}$ and given parameter vector θ, compute $f^*\big(Y(t_{k+1}), s_{k+1}|s_k, \theta\big)$ for all possible pairs (s_k, s_{k+1}), $k = 0, \dots, n-1$. Note that computation for each pair can be done independently, therefore, one can employ here parallel computing.

(2) Base case: $\alpha(\mathbf{X}_0, s_0, \theta) = \nu(s_0)$, where $s_0 = 0, 1$.

(3) Induction: for $s_{k+1} = 0, 1$ compute $\alpha(\mathbf{X}_{k+1}, s_{k+1}, \theta)$ using

$$\alpha(\mathbf{X}_{k+1}, s_{k+1}, \theta) = \sum_{s_k} f^*\big(Y(t_{k+1}), s_{k+1}|s_k, \theta\big)\alpha(\mathbf{X}_k, s_k, \theta).$$

(4) Termination: $L(\mathbf{X}, \theta) = \sum_{s_n} \alpha(\mathbf{X}_n, s_n, \theta)$.

NORMALIZED FORWARD ALGORITHM

One typical computational difficulty with the forward algorithm is the underflow problem. For large k forward variables $\alpha(\mathbf{X}_k, s_k, \theta)$ might be too small and numerically indistinguishable from zero. This issue of underflow is addressed via an appropriate normalization.

Specifically, let us first introduce normalized forward variables by

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

where $L(\mathbf{X}_k, \theta) = \sum_{s_k} \alpha(\mathbf{X}_k, s_k, \theta)$. The normalized forward variables are the conditional probabilities of observing s_k given observed vector $\mathbf{X}_k$. Then the normalized forward variables satisfy the following equation:

$$\bar{\alpha}(\mathbf{X}_{k+1}, s_{k+1}, \theta) = \frac{L(\mathbf{X}_k, \theta)}{L(\mathbf{X}_{k+1}, \theta)} \sum_{s_k} f^*\big(Y(t_{k+1}), s_{k+1}|s_k, \theta\big)\bar{\alpha}(\mathbf{X}_k, s_k, \theta).$$

Next, if for $0 \le k \le n-1$ we define

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

then one can show that

$$d(\mathbf{X}_{k+1}, \theta) = \sum_{s_{k+1}} \sum_{s_k} f^*\big(Y(t_{k+1}), s_{k+1}|s_k, \theta\big)\bar{\alpha}(\mathbf{X}_k, s_k, \theta).$$

Again, note that $d(\mathbf{X}_{k+1}, \theta)$ has a very simple interpretation. It is the conditional density of $Y(t_{k+1})$ given $\mathbf{X}_k$.

This leads us to the normalized version of the forward algorithm.

(1) For observed $\mathbf{X}$ and given parameter vector θ, compute $f^*\big(Y(t_{k+1}), s_{k+1}|s_k, \theta\big)$ for all possible pairs (s_k, s_{k+1}), $k = 0, \dots, n-1$.

(2) Base case: $\bar{\alpha}(\mathbf{X}_0, s_0, \theta) = \nu(s_0)$, where $s_0 = 0, 1$.

(3) Induction: for $s_{k+1} = 0, 1$ compute $\bar{\alpha}(\mathbf{X}_{k+1}, s_{k+1}, \theta)$ using

$$\bar{\alpha}(\mathbf{X}_{k+1}, s_{k+1}, \theta) = \frac{1}{d(\mathbf{X}_{k+1}, \theta)} \sum_{s_k} f^*\big(Y(t_{k+1}), s_{k+1}|s_k, \theta\big)\bar{\alpha}(\mathbf{X}_k, s_k, \theta),$$

and

$$d(\mathbf{X}_{k+1}, \theta) = \sum_{s_{k+1}} \sum_{s_k} f^*\big(Y(t_{k+1}), s_{k+1}|s_k, \theta\big)\bar{\alpha}(\mathbf{X}_k, s_k, \theta).$$

(4) Termination: $\log L(\mathbf{X}, \theta) = \sum_{k=1}^n \log d(\mathbf{X}_k, \theta)$.

Parameter estimation: Case 2

The only difference is that the transitional probability has a discrete probability component in this case. Therefore, one must use the Radon-Nikodym derivative of the probability distribution relative to a dominating measure that includes an atom at $x = 0$. As a result, one should use the following function for computing the likelihood:

$$f\big(x(t), s(t)|s(0), \theta\big) = \begin{cases} 0 & s(0) = 1,\ s(t) = 1,\ x(t) = 0, \\ h_{11}\big(x(t), t\big) & s(0) = 1,\ s(t) = 1,\ x(t) \neq 0, \\ 0 & s(0) = 1,\ s(t) = 0,\ x(t) = 0, \\ h_{10}\big(x(t), t\big) & s(0) = 1,\ s(t) = 0,\ x(t) \neq 0, \\ e^{-\lambda_0 t} & s(0) = 0,\ s(t) = 0,\ x(t) = 0, \\ h_{00}\big(x(t), t\big) & s(0) = 0,\ s(t) = 0,\ x(t) \neq 0, \\ 0 & s(0) = 0,\ s(t) = 1,\ x(t) = 0, \\ h_{01}\big(x(t), t\big) & s(0) = 0,\ s(t) = 1,\ x(t) \neq 0, \end{cases}$$

The rest of derivation is exactly the same.

Simulation

The full likelihood inference made possible by the forward algorithm is expected to be more efficient than the composite likelihood approach suggested by Yan et al. (2014). To assess the efficiency gain in practice, we used the same simulation setting as theirs, and compare the estimators from the two methods. The simulation mimics an animal movement setting with a BMT process. With the time unit in minutes, the parameters were set to be $\lambda_0 = 1/(4 \cdot 60) = 0.004167$, $\lambda_1 = 1/(8 \cdot 60) = 0.002083$, and $\sigma = 25$. The exponential duration parameters mean that the animal on average spends 8 hours in the moving stage and 4 hours in the resting stage. To study the effect of sampling frequency, we generate data on an time interval from zero to 1000×20 minutes, and sample at three frequencies: 250, 500, and 1000. Note that this is different from varying sample size at a fixed sampling frequency.

Two additional scenarios were considered by changing the duration parameters: 1) the average time spent in moving and resting are 4 and 8 hours, respectively (i.e., $\lambda_0 = 0.002083$, $\lambda_1 = 0.004167$); and 2) the average time spent in moving and resting are both 6 hours (i.e., $\lambda_0 = \lambda_1 = 0.002778$). For each dataset, both the composite likelihood estimates and the full likelihood estimates were obtained from R package `smam` (Yan and Pozdnyakov, 2016). For the full likelihood estimator, their standard errors were obtained by inverting the Fisher information matrix. For each configuration, 500 datasets were generated.

Table 1 summarizes the results of the simulation study. For each method, the empirical mean and standard error from the 500 replicates were reported; for the full likelihood method, the average of the standard errors was also reported for each parameter estimate. Both estimators appear unbiased in recovering all three parameters. The full likelihood estimator has much higher efficiency than the composite likelihood estimator in terms of the mean squared error. The relative efficiency of the full likelihood estimator is 1.7 or larger for the volatility parameter σ, and 5.2 or larger for the rate parameters of the exponentially distributed durations in moving or resting states. In scenario 2, where the resting period is twice as long as the moving period on average, the relative efficiency of the estimator of λ_1 is 20.1 for sampling frequency 250. The average of standard errors based on the Fisher information agree well with the empirical ones in most of the configurations, suggesting that the likelihood inferences are valid for inferences.

Increasing the sampling frequency reduces the uncertainty in the parameter estimation, but to quite different degrees on different parameters. From sampling frequency 250 to 1000, the standard errors of the full likelihood estimator of σ almost halved, which is similar to what one would expect when sample size is quadrupled. The standard errors of the estimator of λ_1 and λ_0, however, reduced at a much slower rate. For example, in scenario 3, the empirical standard error of the estimates of λ_1 only decreases from 0.748 to 0.560, only a reduction of 25%. This is expected. Even if we observe the entire trajectory, in contrast to estimation of σ, we cannot estimate λs with certainty. To improve the estimation of λs we need to increase the time horizon, not just the sampling frequency.

Movement of a Mountain Lion

We fit the BMT process to the tracking data on a mountain lion that was analyzed by Yan et al. (2014) with composite likelihood. The data came from a code-only GPS wildlife tracking collar placed on a mature female mountain lion in the Gros Ventre Mountain Range near Jackson Wyoming from 2009 to 2012. The collar was programmed to collect a fix every 8 hours but the sampling times were irregular: the sampling intervals had a standard deviation of 6.45 hours, ranging from 0.5 hours to 120 hours. There were a total of 3917 observations. Temporally, this is a very high-resolution and long-term data set compared to other studies (Schaller and Crawshaw, 1980; Beier et al., 1995). Figure 2 shows the easting and northing offsets (meters, UTM), as well as the increment in distance between successive time intervals, of the lioness's track from her starting position during the first half of 2011 with the location data rounded to the nearest 20m. The figure reveals numerous instances in which the lioness moved very little. This is consistent with previously documented behavior for pumas in which, while consuming a kill or during denning, they move in infrequent

Table 1. Summaries of bias, square root of mean squared error (rMSE), empirical standard error (ESE), and average of standard error (ASE), and relative efficiency in mean squared error for the composite likelihood estimate and the full likelihood estimate. Parameters λ_0 and λ_1 are on the scale of 10^{-3}.

Scenario	Sampling	Parameter	True	Composite Likelihood		Full Likelihood			Relative
	frequency		value	EST	ESE	EST	ESE	ASE	Efficiency
1	250	λ_1	2.083	2.628	2.348	2.242	0.670	0.642	12.2
		λ_0	4.167	4.278	2.695	4.460	1.144	1.082	5.2
		σ	25.000	25.164	1.686	25.037	1.048	1.004	2.6
	500	λ_1	2.083	2.267	1.864	2.094	0.456	0.471	16.8
		λ_0	4.167	4.151	2.878	4.283	0.923	0.918	9.6
		σ	25.000	25.005	1.019	24.977	0.664	0.686	2.4
	1000	λ_1	2.083	2.441	1.985	2.112	0.426	0.430	22.3
		λ_0	4.167	4.604	3.545	4.303	0.929	0.862	14.5
		σ	25.000	25.087	0.642	25.033	0.490	0.485	1.7
2	250	λ_1	4.167	4.660	4.249	4.213	0.953	0.992	20.1
		λ_0	2.083	2.062	1.067	2.155	0.455	0.453	5.4
		σ	25.000	25.068	2.662	24.966	1.405	1.370	3.6
	500	λ_1	4.167	4.620	3.013	4.212	0.857	0.877	12.6
		λ_0	2.083	2.151	1.227	2.145	0.457	0.427	7.1
		σ	25.000	25.113	1.524	25.013	0.971	0.968	2.5
	1000	λ_1	4.167	4.371	3.189	4.161	0.849	0.828	14.2
		λ_0	2.083	2.077	1.362	2.136	0.406	0.415	11.0
		σ	25.000	25.004	0.959	24.984	0.679	0.682	2.0
3	250	λ_1	2.778	3.091	2.392	2.841	0.748	0.704	10.3
		λ_0	2.778	2.804	1.582	2.909	0.664	0.638	5.5
		σ	25.000	25.048	1.828	25.000	1.125	1.130	2.6
	500	λ_1	2.778	2.882	2.215	2.829	0.587	0.600	14.1
		λ_0	2.778	2.755	1.797	2.937	0.582	0.601	8.9
		σ	25.000	24.983	1.137	25.013	0.765	0.787	2.2
	1000	λ_1	2.778	2.982	2.281	2.789	0.560	0.554	16.7
		λ_0	2.778	2.896	2.059	2.896	0.562	0.567	12.9
		σ	25.000	25.019	0.803	24.993	0.564	0.556	2.0

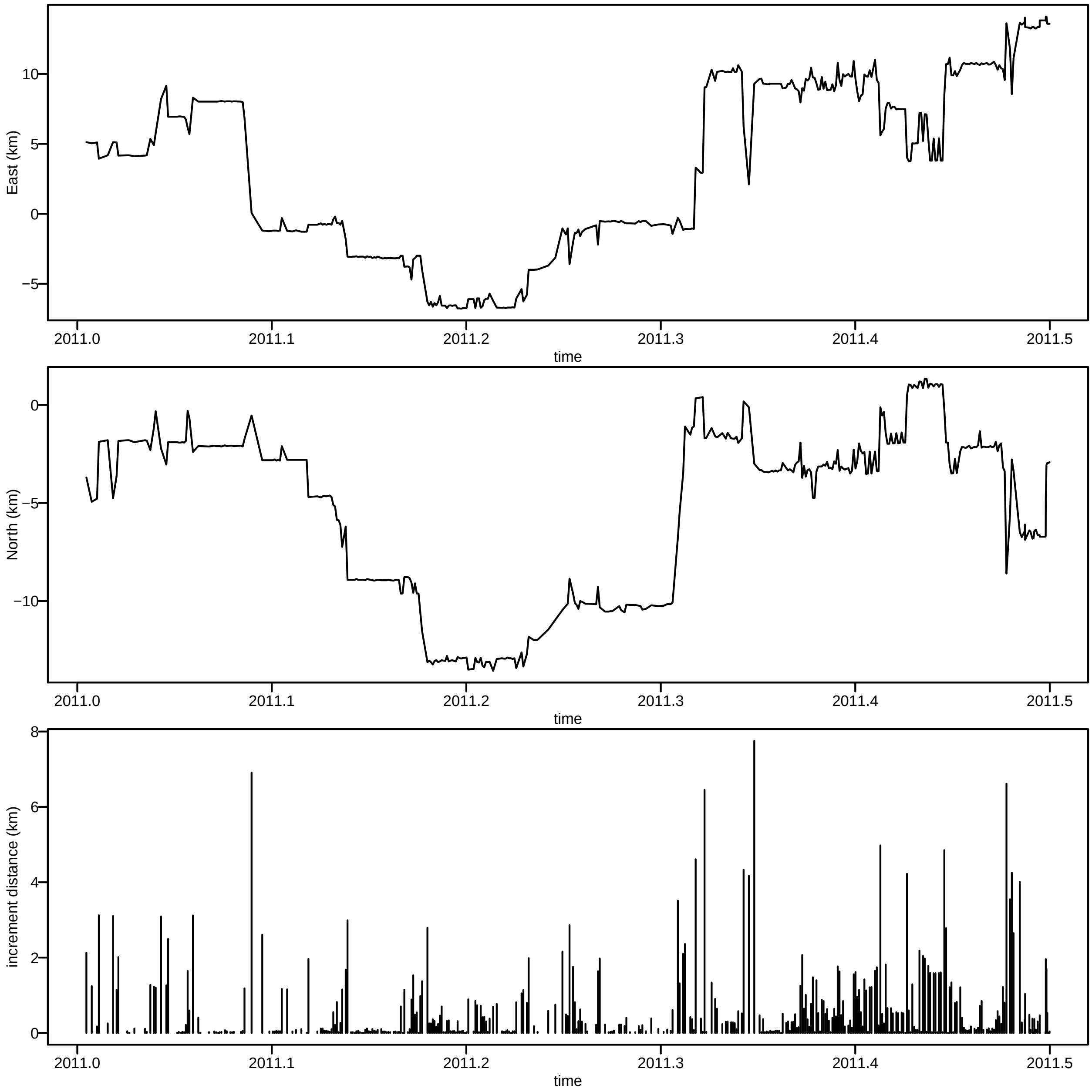

Figure 2. Locations (top and middle) of a mature female mountain lion in the Gros Ventre Mountain Range near Jackson Wyoming tracked with a GPS collar in the first half of 2011 based on locations rounded to the nearest 20m. The bottom plot shows the distance moved since the last sampling time.

short (hundreds of meters) sallies (Hornocker, 1970; Pierce et al., 1998; Pierce and Bleich, 2003). We consider this resting even though the lioness was not always actually stationary.

Table 2 summarizes the fitted parameter values from both the composite likelihood method and the full likelihood method. For the full likelihood method, standard errors are also reported. Since code-only GPS

Table 2. Fitted parameters of the BMT process to the mountain lion data with the composite likelihood (CL) and full likelihood (FL). Parameter estimates (EST) are reported for both the CL and FL methods, while the standard error (SE) is only reported for the FL method.

Parameters	Rounding to nearest 1m			Rounding to nearest 10m			Rounding to nearest 20m		
	CL	FL		CL	FL		CL	FL	
	EST	EST	SE	EST	EST	SE	EST	EST	SE
λ_1	3.904	3.681	0.198	3.415	3.263	0.156	3.043	2.605	0.140
λ_0	0.358	0.336	0.008	0.313	0.296	0.007	0.260	0.245	0.006
σ	979.645	973.195	23.076	989.629	984.545	17.153	1044.946	985.450	16.959

devices are known to have measurement errors of 10m–20m, we also fitted the data with locations rounded to the nearest 10m or 20m. Under all rounding schemes, the fitted BMT process suggests that this lion spent about 10 times as long in resting as in moving. The mean time spent in resting and moving increases when the rounding increases, even though their ratio were relatively stable. The volatility parameter remained quite stable across different rounding precisions.

There is little literature with which to directly compare our biological results. Schaller and Crawshaw (1980) reported activity patterns of a female jaguar (*Panthera onca*) in the Pantanal region of Mato Grosso, Brazil that were based on 97 hours of radiotracking and observations, but excluded time at known kills. Time was portioned into percent spent resting (37%), locally active (22%), and traveling (41%). They defined 'locally active' to mean when an animal was eating or grooming or otherwise making small motions detectable by small variations radio-telemetry signal strength but not so constant to infer resting nor determined travel. We would consider local activity to be captured in our resting periods, which would make the time allocations 59% and 41% for resting and moving, respectively. Beier et al. (1995) studied mountain lions in the Santa Ana Mountains of southern California using radiotracking and *in situ* tracking. They reported that the percentage of time traveling was 9 ± 11 during diurnal periods, 36 ± 21 during nocturnal periods, and 25 ± 13 during diel periods. These values are similar but somewhat higher than ours, but our analysis includes periods of prolonged inactivity, which their analysis did not. Our biological results, which are from a single lioness, are limited; nonetheless they support the findings of other researchers that carnivores tend to spend very little of their time in active behaviors. Jeschke (2007) characterized carnivores as 'lazy' because they are often sufficiently successful in hunting that they have much free time in which they are not engaged in seeking food. We quantify this notion noted that our analysis suggests this lioness spent roughly ten times longer resting than moving over a period of around six months.

References

P. Beier, D. Choate and R. H. Barrett (1995). "Movement patterns of mountain lions during different behaviors." *Journal of Mammalogy* **76**, 1056–1070.

C. Calenge (2006). "The package adehabitat for the R software: Tool for the analysis of space and habitat use by animals." *Ecological Modelling* **197**, 1035.

V. R. Cane (1959). "Behavior sequences as semi-Markov chains." *Journal of the Royal Statistical Society, Series B* **21**, 36–58.

O. Cappé, E. Moulines and T. Rydén (2005). *Inference in Hidden Markov Models*, volume 6. Springer.

A. Di Crescenzo, E. Di Nardo and L. M. Ricciardi (2005). "Simulation of first-passage times for alternating Brownian motions." *Methodology and Computing in Applied Probability* **7**, 161–181.

A. Di Crescenzo, B. Martinucci and S. Zacks (2014). "On the geometric brownian motion with alternating trend." *Mathematical and statistical methods for actuarial sciences and finance, Eds. Perna, C. and Sibillo, M.* pp. 81–85.

A. Di Crescenzo and F. Pellerey (2002). "On prices' evolutions based on geometric telegrapher's process." *Appl. Stoch. Models Bus. Ind.* **18**, 171–184.

A. Di Crescenzo and S. Zacks (2015). "Probability law and flow function of Brownian motion driven by a generalized telegraph process." *Methodol. Comput. Appl. Probab.* **17**, 761–780.

C. Farmer, K. Safi, D. Barber, M. Martel and K. Bildstein (2010). "Efficacy of migration counts for monitoring continental populations of raptors: an example using the osprey (pandion haliaetus)." *The Auk* **127**, 863–970.

J.-P. Fouque, G. Papanicolaou, R. Sircar and K. Sølna (2011). *Multiscale Stochastic Volatility for Equity, Interest Rate, and Credit Derivatives*. Cambridge University Press, Cambridge.

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

M. G. Hornocker (1970). "An analysis of mountain lion predation upon mule deer and elk in the idaho primitive area." Technical Report 21, Wildlife Society.

J. M. Jeschke (2007). "When carnivores are "full and lazy"." *Oecologia* **152**, 357–364.

I. D. Jonsen, J. M. Flemming and R. A. Myers (2005). "Robust state-space modeling of animal movement data." *Ecology* **86**, 2874–2880.

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

B. Kranstauber, R. Kays, S. D. LaPoint, M. Wikelski and K. Safi (2012). "A dynamic brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement." *Journal of Animal Ecology* **81**, 738–746.

B. Kranstauber and M. Smolla (2013). *move: Visualizing and analyzing animal track data*. R package version 1.1.360/r365.

M. Lonergan, M. Fedak and B. McConnell (2009). "The effects of interpolation error and location quality on animal track reconstruction." *Marine Mammal Science* **25**, 275–282.

R. M. Nielson, H. Sawyer and T. L. McDonald (2012). *BBMM: Brownian Bridge Movement Model*. R package version 2.3.

E. S. Page (1960). "Theoretical considerations of routine maintenance." *The Computer Journal* **2**, 199–204.

T. Patterson, L. Thomas, C. Wilcox, O. Ovaskainen and J. Matthiopoulos (2008). "State-space models of individual animal movement." *Trends in Ecology and Evolution* **23**, 87–94.

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

B. M. Pierce and V. C. Bleich (2003). "Mountain lion." In G. A. Feldhamer, B. C. Thompson and J. A. Chapman (eds.), *Wild Mammals of North America, Biology, Management and Conservation*, pp. 744–757. Baltimore, MD: Johns Hopkins University Press, 2 edition.

B. M. Pierce, V. C. Bleich, C.-L. B. Chetkiewicz and J. D. Wehausen (1998). "Timing of feeding bouts of mountain lions." *Journal of Mammalogy* **79**, 222–226.

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.

G. B. Schaller and P. G. Crawshaw, Jr. (1980). "Movement patterns of jaguar." *Biotropica* **12**, 161–168.

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

J. Takekawa, S. Newman, X. Xiao, D. Prosser, K. Spragens, E. Palm, B. Yan, T. Li, F. Lei, D. Zhao, D. Douglas, S. Muzaffar and W. Ji (2010). "Migration of waterfowl in the east asian flyway and spatial relationship to hpai h5n1 outbreaks." *Avian Diseases* **54(s1)**, 466–476.

E. Willems and R. Hill (2009). "Predator-specific landscapes of fear and resource distribution: effects on spatial range use." *Ecology* **90**, 546–555.

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

J. Yan and V. Pozdnyakov (2016). *smam: Statistical Modeling of Animal Movements*. R package version 0.3-0.

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