Discretely Observed Brownian Motion Governed by a Telegraph Process:

Estimation

Chaoran Hu, Mark Elbroch, Tom Meyer, Vladimir Pozdnyakov, and Jun Yan

University of Connecticut

2018

Brownian Bridge Movement Model

The Brownian motion (BM) and random walk are often employed by ecologists to model animal movement. Highly cited Horne et al. (2007) introduced the Brownian bridge movement model (BBMM) which characterizes the missing movement path between two sequential positions (taken with measurement errors) by a Brownian bridge.

Brownian Motion Governed by a Telegraph Process

The Brownian Motion governed by a telegraph process (BMT process) is a natural generalization of the BBMM that allows two or more modes of movements (e.g., moving and resting, low speed and high speed, moving, resting and handling). The moving-resting process 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 states are exponentially distributed. 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.

Telegraph Process

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.
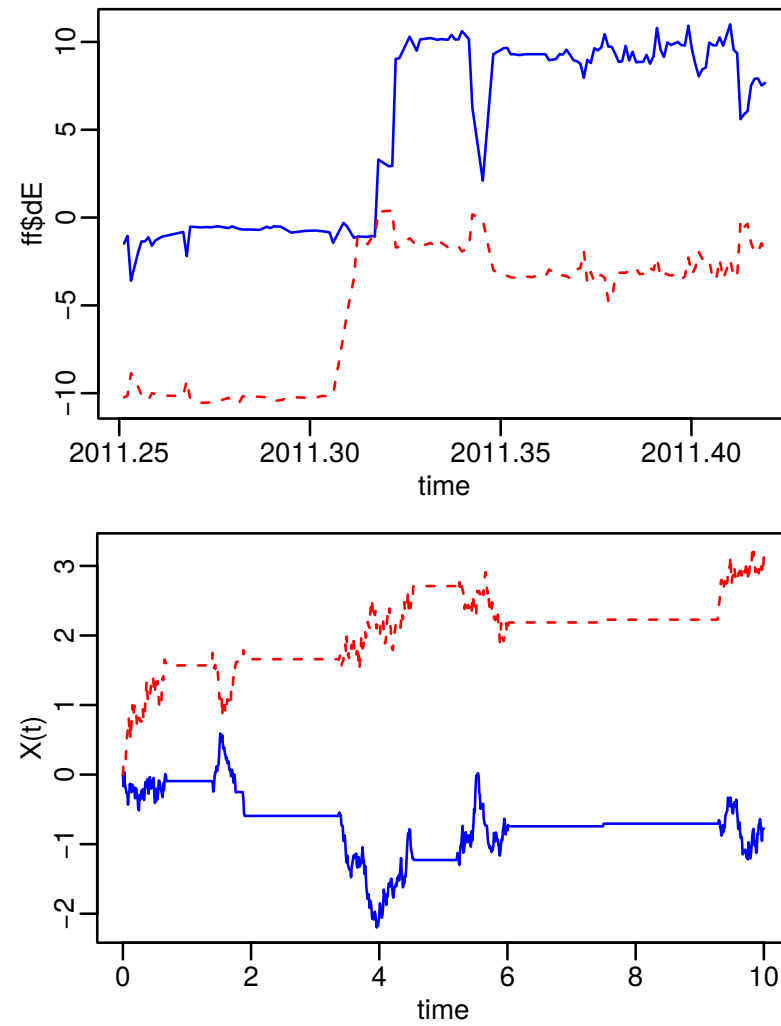
BMT

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

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

where σ is a volatility parameter, and $B(t)$ is the standard Brownian motion independent of $S(t)$.

Real Data vs BMT Process Trajectory

Markov Property

Let us note here that $X(t)$ is not Markov even though both $S(t)$ and $B(t)$ are Markov. 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

$$\text{Pr}\left(X(t) \in \Gamma | \mathcal{F}_s\right) \neq \text{Pr}\left(X(t) \in \Gamma | X(s)\right), \quad 0 \leq s \leq t,$$

where $\mathcal{F}_s$ is, as usually, a σ-field generated by $\left(S(u), B(u)\right)_{0 \leq u \leq s}$, 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 $\{X(t), S(t) : t \geq 0\}$ is a Markov process with stationary increments in $X(t)$. Moreover, the distribution of the increment $X(t+u) - X(t)$ depends only on $S(t)$.

Key Random Variables

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. That is,

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

and $R(t) = t - M(t)$.

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).

Key Random Variables

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

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

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

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

Zacks (2004) provides formulas for these densities.

Joint Distribution of $\big(X(t), S(t)\big)$

Without loss of generality, we assume $X(0) = 0$. Given value of $M(t) = w$ the random variable X(t) has the normal distribution with mean 0 and variance $\sigma^2 w$. Combining this observation with formulas from previous section one can get the joint distribution of $\big(X(t), S(t)\big)$. For example, for $0 < w < t$,

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

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

One can get the joint distribution of $\big(X(t), S(t)\big)$ starting from $S(0) = 1$ by integrating w out.

Joint Distribution of $\big(X(t), S(t)\big)$

If we introduce the following functions:

$$h_{11}(x,t) = e^{-\lambda_1 t}\phi(x;\sigma^2 t) + \int_0^t \phi(x;\sigma^2 w)p_{11}(w,t)\mathrm{d}w,$$

$$h_{10}(x,t) = \int_0^t \phi(x;\sigma^2 w)p_{10}(w,t)\mathrm{d}w,$$

$$h_{00}(x,t) = \int_0^t \phi\big(x;\sigma^2(t-w)\big)p_{00}(w,t)\mathrm{d}w,$$

$$h_{01}(x,t) = \int_0^t \phi\big(x;\sigma^2(t-w)\big)p_{01}(w,t)\mathrm{d}w,$$

then we have

$$P_1\big[X(t) \in \mathrm{d}x, S(t) = 1\big] = h_{11}(x,t)\mathrm{d}x,$$

$$P_1\big[X(t) \in \mathrm{d}x, S(t) = 0\big] = h_{10}(x,t)\mathrm{d}x,$$

$$P_0\big[X(t) \in \mathrm{d}x, S(t) = 0\big] = h_{00}(x,t)\mathrm{d}x + e^{-\lambda_0 t}\delta_0(x),$$

$$P_0\big[X(t) \in \mathrm{d}x, S(t) = 1\big] = h_{01}(x,t)\mathrm{d}x,$$

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.

Estimation

Assume that a BMT process $X(t)$ with parameters $\theta = (\lambda_0, \lambda_1, \sigma)$ 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.

Transitional Probability

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

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

Full Likelihood

The likelihood function of the observed data is then

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

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

No States

In practice, it is more likely that **S** is not observed. In this case, we need to work with **X** likelihood function which is given by

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

Since the observed process $X(t)$ itself is not Markovian, formulas similar to (2) are not available. A naive 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 **S**. Since the total number of such trajectories is 2^{n+1}, this approach will not work for any real world data set.

Forward Algorithm (Dynamic Programming)

First, we introduce so-called forward variables:

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

where $\mathbf{X}_k = \Big(X(0), X(t_1), \ldots, X(t_k)\Big)$, and $1 \leq k \leq n$. Then, using, in essence, the Markov property of $\Big(X(t), S(t)\Big)$, we get

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

where $Y(t_{k+1}) = X(t_{k+1}) - X(t_k)$.

Forward Algorithm (Dynamic Programming)

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}), $k = 0, \ldots, n-1$. 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: $L(\mathbf{X}, \theta) = \sum_{s_n} \alpha(\mathbf{X}_n, s_n, \theta).$

17

Normalized Forward Algorithm

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

Normalized Forward Variables

If we 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 $L(\mathbf{X}_k, \theta) = \sum_{s_k} \alpha(\mathbf{X}_k, s_k, \theta)$. 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

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

Normalized Forward Algorithm

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}), $k = 0, \ldots, n-1$.

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

Hidden Markov Tools

- Forward Algorithm (Likelihood)

- Forward-Backward Algorithm (Marginal distribution of a hidden state given the observations)

- Baum-Welch Algorithm (an EM algorithm, cannot be used in our case)

- Viterbi Algorithm (the most probable sequence of hidden states given the observations)

Current Work

Sometimes it makes sense to use a different continuous time Markov Chain instead of the telegraph process.

A continuous time Markov Chain is fully determined by its initial distribution and transition rate matrix.

The telegraph process is a continuous time Markov Chain (if cycle holding times are exponential) with the following transition rate matrix:

$$\mathbf{Q}_T = \begin{pmatrix} -\lambda_0 & \lambda_0 \\ \lambda_1 & -\lambda_1 \end{pmatrix} \tag{5}$$

For the moving-resting-handling process we use this one:

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 p_1 & \lambda_0 p_2 \\ \lambda_1 & -\lambda_1 & 0 \\ \lambda_2 & 0 & -\lambda_2 \end{pmatrix} \tag{6}$$

References

V.I. Pozdnyakov, T. Meyer, Y. Wang, and J. Yan, On Modeling Animal Movements Using Brownian Motion with Measurement Error, *Ecology*, 95 (2014), 247-253.

J. Yan, Y. Chen, K. Lawrence-Apfel, I.M. Ortega, V.I. Pozdnyakov, S. Williams, and T. Meyer, A Moving-Resting Process with an Embedded Brownian Motion for Animal Movements, *Population Ecology*, 56 (2014), 401-415

V.I. Pozdnyakov, M. Elbroch, A. Labarga, T. Meyer, and J. Yan, Discretely Observed Brownian Motion Governed by Telegraph Process: Estimation, *Methodology and Computing in Applied Probability*, to appear.

V. Pozdnyakov, M. Elbroch, C. Hu, T. Meyer and J. Yan (2018). On estimation for Brownian motion governed by telegraph process with multiple off states, *ArXiv*.