Theoretical Crawling

A guide to the mathematics and statistics of animal movement

Devin S. Johnson

NOAA Fisheries Marine Mammal Laboratory

Seattle, Washington

Email: devin.johnson@noaa.gov

Animal movement workshop

February 14–16, 2017

Outline

Mathematics

1. Discrete-time random walks
2. Correlated random walks
3. Brownian motion
4. Ornstein-Ulenbeck (OU) process
5. Stochastic differential equations
6. Integrated SDEs
7. Continuous-time CRW

Outline

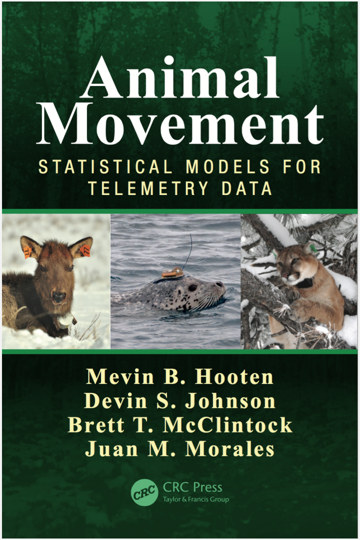
Mathematics

1. Discrete-time random walks
2. Correlated random walks
3. Brownian motion
4. Ornstein-Ulenbeck (OU) process
5. Stochastic differential equations
6. Integrated SDEs
7. Continuous-time CRW

Statistics

1. Maximum likelihood inference
2. Bayesian inference
3. State-space models
4. Kalman filter/smoother (KFS)
5. Practical Bayesian inference
6. Process imputation

For more information...



- Available from Amazon ($86)
- Soon to be available at the MML library
- Borrow from me (maybe Brett)

Part I

Mathematics of animal movement

Discrete-time models

Time series models

Telemetry data are thought of as predominately *spatial*

- We display them on 2d maps
- We want to know something about *space* use
- We want to know which spatial locations are selected over others

Time series models

Telemetry data are thought of as predominately *spatial*

- We display them on 2d maps
- We want to know something about *space* use
- We want to know which spatial locations are selected over others

Most mathematical models and statistical analysis view telemetry data as *time series* processes where a realization occurs in geographical space

Time series models

Telemetry data are thought of as predominately *spatial*

- We display them on 2d maps
- We want to know something about *space* use
- We want to know which spatial locations are selected over others

Most mathematical models and statistical analysis view telemetry data as *time series* processes where a realization occurs in geographical space

Notation (discrete time)

- $\boldsymbol{\mu}_t = (\mu_{x,t}, \mu_{y,t})$ is the location of the animal at time t.
- $\mathbf{s}_t = (s_{x,t}, s_{y,t})$ is the observed location at time t.
- $d\boldsymbol{\mu}_t = \boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1}$ is *movement*!

Random walk

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t$$

$$[\boldsymbol{\epsilon}_t] = N(0, \boldsymbol{\Sigma})$$

- The movements, $d\boldsymbol{\mu}_t = \boldsymbol{\epsilon}_t$ are all independent of each other.
- Typically $\boldsymbol{\Sigma} \equiv \sigma^2 \mathbf{I}$, so, $\mu_{x,t}$ is independent of $\mu_{y,t}$

Random walk

$$\boldsymbol{\mu}_t = \boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t$$

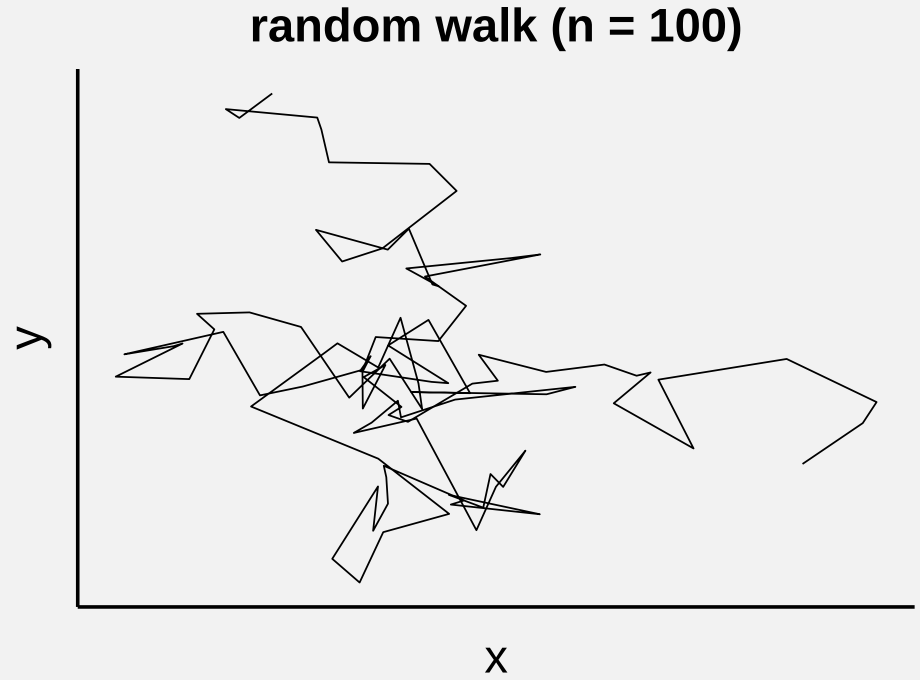
$$[\boldsymbol{\epsilon}_t] = N(0, \boldsymbol{\Sigma})$$

- The movements, $d\boldsymbol{\mu}_t = \boldsymbol{\epsilon}_t$ are all independent of each other.
- Typically $\boldsymbol{\Sigma} \equiv \sigma^2\mathbf{I}$, so, $\mu_{x,t}$ is independent of $\mu_{y,t}$

An unconditional view

We can rewite the RW using the accumulation of movements:

$$\begin{aligned}\boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t \\ &= \boldsymbol{\mu}_{t-2} + \boldsymbol{\epsilon}_{t-1} + \boldsymbol{\epsilon}_t \\ &= \boldsymbol{\mu}_0 + \sum_{i=1}^{t} \boldsymbol{\epsilon}_i\end{aligned}$$


random walk (n = 100)
y
x

Vector autoregressive models (VAR)

Let's suppose that an animal would like to stay close to some focal point, say $\bar{\boldsymbol{\mu}}$.

Vector autoregressive models (VAR)

Let's suppose that an animal would like to stay close to some focal point, say $\bar{\boldsymbol{\mu}}$.

VAR(1) model

$$\begin{aligned}\boldsymbol{\mu}_t &= \bar{\boldsymbol{\mu}} + \mathbf{M}(\boldsymbol{\mu}_{t-1} - \bar{\boldsymbol{\mu}}) + \boldsymbol{\epsilon}_t \\ &= (\mathbf{I} - \mathbf{M})\bar{\boldsymbol{\mu}} + \mathbf{M}\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t\end{aligned}$$

- Weighted mean of last location and point of attraction
- If $\mathbf{M} \equiv \gamma\mathbf{I}$ and $|\gamma| < 1$ then as $t \to \infty$

$$[\boldsymbol{\mu}_t] = N\left(\bar{\boldsymbol{\mu}}, \frac{1}{1-\gamma^2}\boldsymbol{\Sigma}\right)$$

Vector autoregressive models (VAR)

Let's suppose that an animal would like to stay close to some focal point, say $\bar{\boldsymbol{\mu}}$.

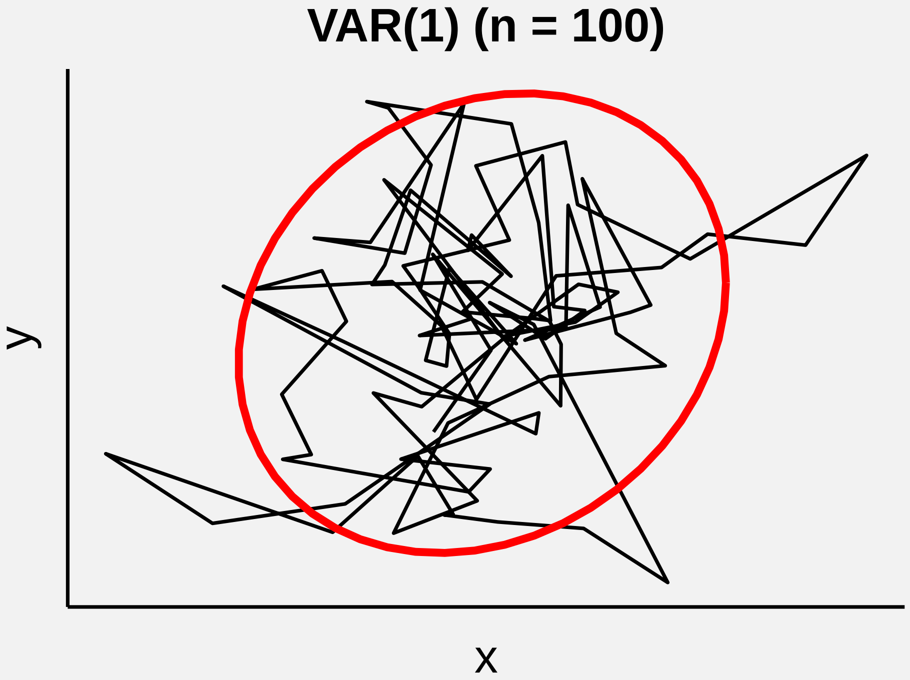
VAR(1) model

$$\begin{aligned}\boldsymbol{\mu}_t &= \bar{\boldsymbol{\mu}} + \mathbf{M}(\boldsymbol{\mu}_{t-1} - \bar{\boldsymbol{\mu}}) + \boldsymbol{\epsilon}_t \\ &= (\mathbf{I} - \mathbf{M})\bar{\boldsymbol{\mu}} + \mathbf{M}\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t\end{aligned}$$

- Weighted mean of last location and point of attraction
- If $\mathbf{M} \equiv \gamma\mathbf{I}$ and $|\gamma| < 1$ then as $t \to \infty$

$$[\boldsymbol{\mu}_t] = N\left(\bar{\boldsymbol{\mu}}, \frac{1}{1-\gamma^2}\boldsymbol{\Sigma}\right)$$

- What happens if $\gamma \to 1$?


VAR(1) (n = 100)
y
x

Modeling velocity

- Rule 1. velocity $\neq$ speed!

Modeling velocity

- Rule 1. velocity $\neq$ speed!
- We are thinking of velocity in the physics sense. Speed is the magnitude of velocity.

$$\text{velocity} = \text{derivative of location process}$$

Modeling velocity

- Rule 1. velocity $\neq$ speed!
- We are thinking of velocity in the physics sense. Speed is the magnitude of velocity.

$$\text{velocity} = \text{derivative of location process}$$

What does that mean for our random walk models in discrete-time?

$$\begin{aligned}\text{velocity} &= d\boldsymbol{\mu}_t \\ &= \boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1}\end{aligned}$$

VAR(1) model for velocity

Instead of directly modeling location as a VAR(1), we can model movement (velocity) with a VAR(1)

$$d\boldsymbol{\mu}_t = \mathbf{M} d\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t$$

Now, the movement steps are *correlated* compared to a simple random walk.

VAR(1) model for velocity

Instead of directly modeling location as a VAR(1), we can model movement (velocity) with a VAR(1)

$$d\boldsymbol{\mu}_t = \mathbf{M} d\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t$$

Now, the movement steps are *correlated* compared to a simple random walk.

We might call this model a *correlated random walk*

VAR(1) model for velocity

Instead of directly modeling location as a VAR(1), we can model movement (velocity) with a VAR(1)

$$d\boldsymbol{\mu}_t = \mathbf{M} d\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t$$

Now, the movement steps are *correlated* compared to a simple random walk.

We might call this model a *correlated random walk*

Why CRW?

- Step at time t, $d\boldsymbol{\mu}_t$, tends to be similar to the previous step, $d\boldsymbol{\mu}_{t-1}$
- The correlation in steps produces a *superdiffusive* process for μ_t
- Why is it useful? Better model of movement in the short term (i.e., $t << \infty$)

Correlated random walk

If we model $d\boldsymbol{\mu}_t = \mathbf{M}\ d\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t$, what does this mean for $\boldsymbol{\mu}_t$?

Correlated random walk

If we model $d\boldsymbol{\mu}_t = \mathbf{M}\ d\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t$, what does this mean for $\boldsymbol{\mu}_t$?

Recall we can always write the position process like this:

$$
\begin{aligned}
\boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + (\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1}) \\
&= \boldsymbol{\mu}_{t-2} + (\boldsymbol{\mu}_{t-1} - \boldsymbol{\mu}_{t-2}) + (\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1}) \\
&\vdots \\
&= \boldsymbol{\mu}_0 + \sum_{i=1}^{t} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i-1}) \\
&= \boldsymbol{\mu}_0 + \sum_{i=1}^{t} d\boldsymbol{\mu}_i
\end{aligned}
$$

Notice that there was no distributional assumption for $d\boldsymbol{\mu}_t$.
It's a simple recursion.

Correlated random walk

The CRW can also be represented as a VAR(2) process

$$\begin{aligned}
\boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \mathbf{M} d\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t \\
&= \boldsymbol{\mu}_{t-1} + \mathbf{M}(\boldsymbol{\mu}_{t-1} - \boldsymbol{\mu}_{t-2}) + \boldsymbol{\epsilon}_t \\
&= (\mathbf{I} + \mathbf{M})\boldsymbol{\mu}_{t-1} - \mathbf{M}\boldsymbol{\mu}_{t-2} + \boldsymbol{\epsilon}_t
\end{aligned}$$

Current location is a weighted sum of previous *two* locations

Correlated random walk

The CRW can also be represented as a VAR(2) process

$$\begin{aligned}\boldsymbol{\mu}_t &= \boldsymbol{\mu}_{t-1} + \mathbf{M} d\boldsymbol{\mu}_{t-1} + \boldsymbol{\epsilon}_t \\ &= \boldsymbol{\mu}_{t-1} + \mathbf{M}(\boldsymbol{\mu}_{t-1} - \boldsymbol{\mu}_{t-2}) + \boldsymbol{\epsilon}_t \\ &= (\mathbf{I} + \mathbf{M})\boldsymbol{\mu}_{t-1} - \mathbf{M}\boldsymbol{\mu}_{t-2} + \boldsymbol{\epsilon}_t\end{aligned}$$

Current location is a weighted sum of previous *two* locations

The **M** matrix

Turning angle specification

$$\mathbf{M} = \gamma \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

- θ represents mean turning angle (probably close to 0)
- $0 < \gamma < 1$ controls the correlation in velocity

Correlated random walk

Let's suppose $\theta = 0$, so,

$$\mathbf{M} = \gamma \mathbf{I}$$

Then $\mathrm{corr}(d\boldsymbol{\mu}_t, d\boldsymbol{\mu}_s) = \gamma^{|t-s|}$

Correlated random walk

Let's suppose $\theta = 0$, so,

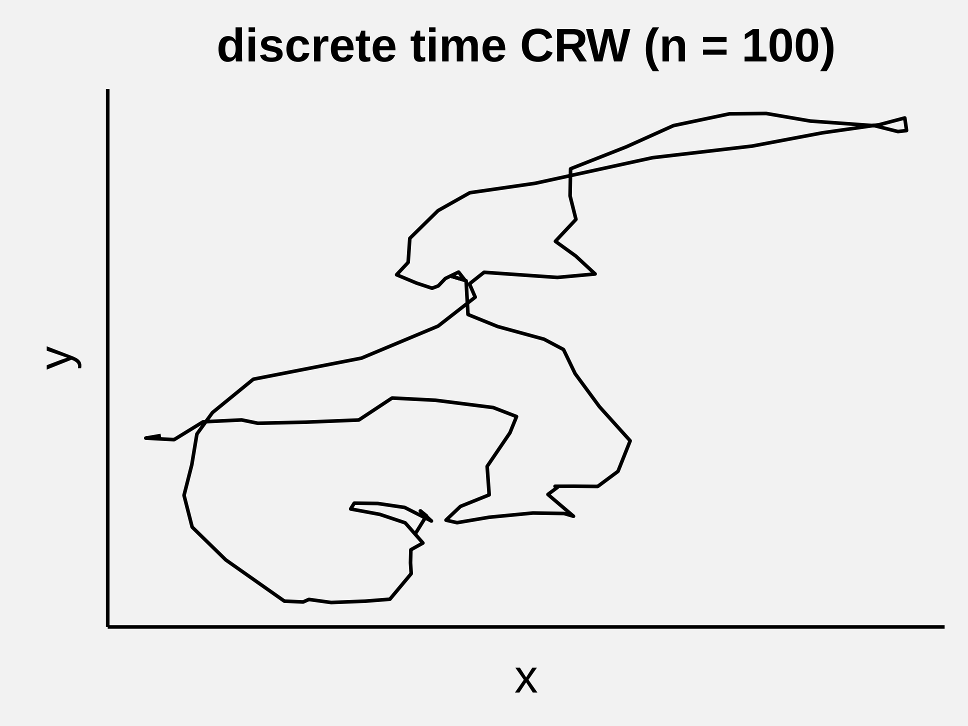
$$\mathbf{M} = \gamma \mathbf{I}$$

Then $\text{corr}(d\boldsymbol{\mu}_t, d\boldsymbol{\mu}_s) = \gamma^{|t-s|}$

But let's reparameterize $\gamma = e^{-\beta}$, then

- $\gamma \approx 1 \implies \beta \approx 0$
- $\gamma \approx 0 \implies \beta$ very large (sort of)

So, $\text{corr}(d\boldsymbol{\mu}_t, d\boldsymbol{\mu}_s) = e^{-\beta|t-s|}$


discrete time CRW (n = 100)
y
x

Continuous-time models

Brownian motion

Forms the basis of all continuous-time models

Brownian motion

Forms the basis of all continuous-time models

Random walk in continuous-time

Brownian motion

Forms the basis of all continuous-time models

Random walk in continuous-time

So, let's start from the beginning...

$$
\begin{aligned}
\mathbf{b}_t &= \mathbf{b}_{t-1} + \boldsymbol{\epsilon}_t \\
&= \mathbf{b}_0 + \sum_{j=1}^{t} [\mathbf{b}_j - \mathbf{b}_{j-1}] \\
&= \mathbf{b}_0 + \sum_{j=1}^{t} d\mathbf{b}_j
\end{aligned}
$$

Brownian motion

Forms the basis of all continuous-time models

Random walk in continuous-time

So, let's start from the beginning...

$$\begin{aligned}\mathbf{b}_t &= \mathbf{b}_{t-1} + \boldsymbol{\epsilon}_t \\ &= \mathbf{b}_0 + \sum_{j=1}^{t} [\mathbf{b}_j - \mathbf{b}_{j-1}] \\ &= \mathbf{b}_0 + \sum_{j=1}^{t} d\mathbf{b}_j\end{aligned}$$

Recall,
$[d\mathbf{b}_j] = N(\mathbf{0}, \sigma^2\mathbf{I})$ and $d\mathbf{b}_j$ is indep. of $d\mathbf{b}_i$
(usually, $\mathbf{b}_0 = \mathbf{0}$ and $\sigma = 1$)

Brownian motion

We can get to continuous-time by making the time gaps, $\delta = \tau_j - \tau_{j-1}$, smaller and smaller

Brownian motion

We can get to continuous-time by making the time gaps, $\delta = \tau_j - \tau_{j-1}$, smaller and smaller

Definition

$$\begin{aligned}
\mathbf{b}_\tau &= \lim_{\delta \to 0} \sum_{j=1}^{\tau} d\mathbf{b}_j \\
&= \int_0^\tau d\mathbf{b}_u
\end{aligned}$$

Brownian motion

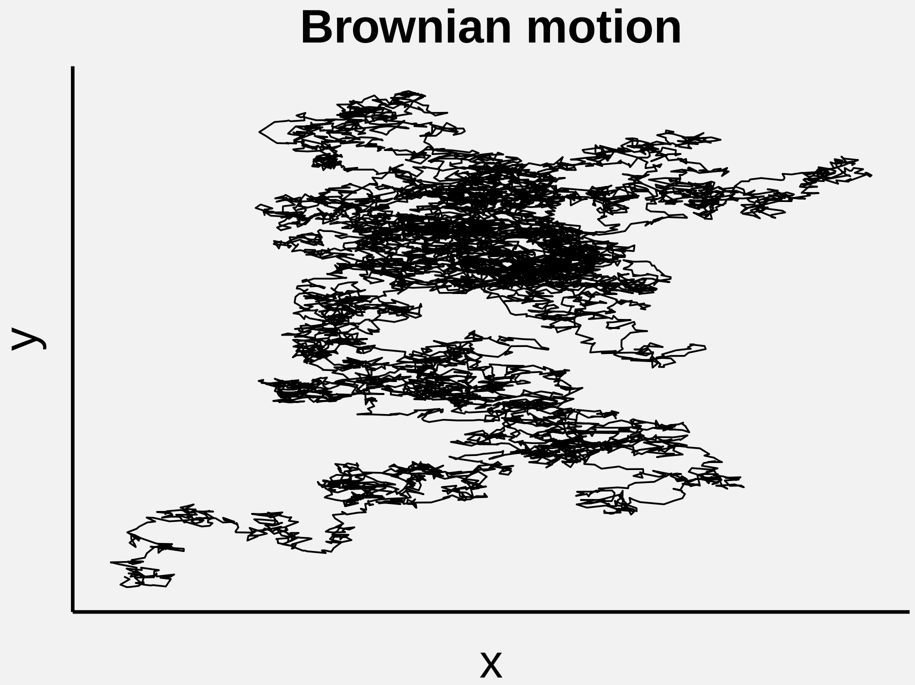
We can get to continuous-time by making the time gaps, $\delta = \tau_j - \tau_{j-1}$, smaller and smaller

Definition

$$\mathbf{b}_\tau = \lim_{\delta \to 0} \sum_{j=1}^{\tau} d\mathbf{b}_j$$

$$= \int_0^\tau d\mathbf{b}_u$$

- $d\mathbf{b}_\tau$ is infinitely rough
- $\mathbf{b}_\tau$ is a continuous function of time with no classical derivative
- Trivia fact for the bar: Here '$\int$' represents an Ito integral


Brownian motion
y
x

Brownian motion

Properties

- $\text{Mean}[\mathbf{b}_\tau] = \mathbf{0}$

- $\text{Var}[\mathbf{b}_\tau] = \tau \mathbf{I}$

- Independent increments ...

  $\mathbf{b}_{\tau_2} - \mathbf{b}_{\tau_1}$ is independent of $\mathbf{b}_{\tau_4} - \mathbf{b}_{\tau_3}$ if $[\tau_1, \tau_2]$ does not overlap $[\tau_3, \tau_4]$.

Ornstein-Uhlenbeck process

What about an AR(1) version of BM?

$$\text{(1)} \quad \boldsymbol{\mu}_t = \gamma(\boldsymbol{\mu}_{t-1} - \bar{\boldsymbol{\mu}}) + \boldsymbol{\epsilon}_t$$

Ornstein-Uhlenbeck process

What about an AR(1) version of BM?

$$(1) \quad \boldsymbol{\mu}_t = \gamma(\boldsymbol{\mu}_{t-1} - \bar{\boldsymbol{\mu}}) + \boldsymbol{\epsilon}_t$$

$$\begin{aligned}(2) \quad \boldsymbol{\mu}_t &= \boldsymbol{\mu}_0 + \sum_{j=1}^{t}[\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j-1}] \\ &= \boldsymbol{\mu}_0 + \sum_{j=1}^{t}(\gamma - 1)(\boldsymbol{\mu}_{j-1}) - \bar{\boldsymbol{\mu}}) + \sum_{j=1}^{t}\boldsymbol{\epsilon}_j\end{aligned}$$

Ornstein-Uhlenbeck process

What about an AR(1) version of BM?

$$(1) \quad \boldsymbol{\mu}_t = \gamma(\boldsymbol{\mu}_{t-1} - \bar{\boldsymbol{\mu}}) + \boldsymbol{\epsilon}_t$$

$$\begin{aligned}(2) \quad \boldsymbol{\mu}_t &= \boldsymbol{\mu}_0 + \sum_{j=1}^{t}[\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j-1}] \\ &= \boldsymbol{\mu}_0 + \sum_{j=1}^{t}(\gamma - 1)(\boldsymbol{\mu}_{j-1}) - \bar{\boldsymbol{\mu}}) + \sum_{j=1}^{t}\boldsymbol{\epsilon}_j\end{aligned}$$

$$(3) \quad \boldsymbol{\mu}_\tau = \boldsymbol{\mu}_0 + \int_0^\tau (\gamma - 1)(\boldsymbol{\mu}_u - \bar{\boldsymbol{\mu}})du + \sigma\mathbf{b}_u$$

Ornstein-Uhlenbeck process

What about an AR(1) version of BM?

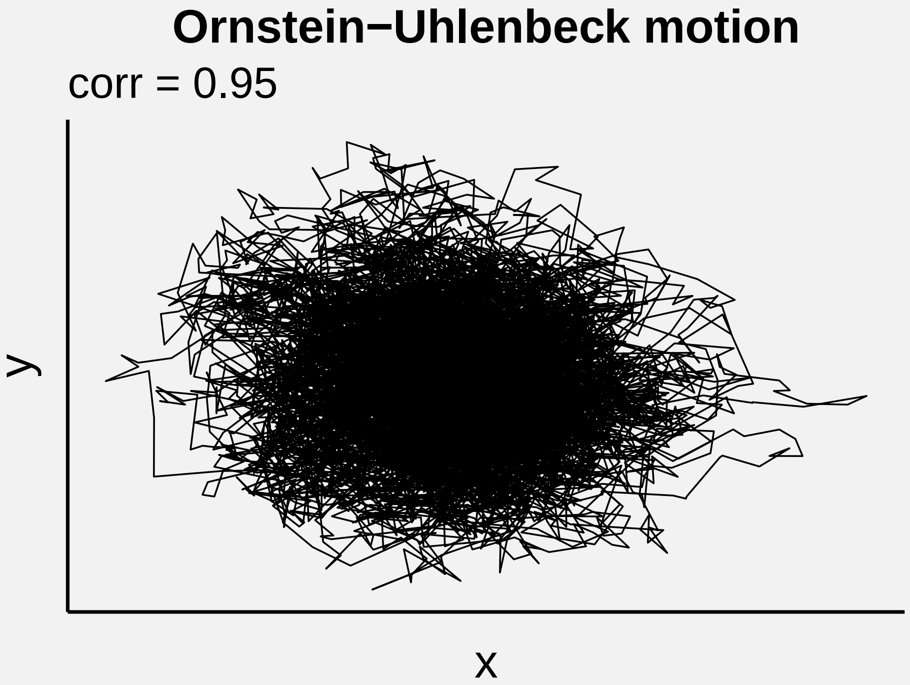
$$(1) \quad \boldsymbol{\mu}_t = \gamma(\boldsymbol{\mu}_{t-1} - \bar{\boldsymbol{\mu}}) + \boldsymbol{\epsilon}_t$$

$$\begin{aligned}(2) \quad \boldsymbol{\mu}_t &= \boldsymbol{\mu}_0 + \sum_{j=1}^{t}[\boldsymbol{\mu}_j - \boldsymbol{\mu}_{j-1}] \\ &= \boldsymbol{\mu}_0 + \sum_{j=1}^{t}(\gamma - 1)(\boldsymbol{\mu}_{j-1}) - \bar{\boldsymbol{\mu}}) + \sum_{j=1}^{t}\boldsymbol{\epsilon}_j\end{aligned}$$

$$(3) \quad \boldsymbol{\mu}_\tau = \boldsymbol{\mu}_0 + \int_0^\tau (\gamma - 1)(\boldsymbol{\mu}_u - \bar{\boldsymbol{\mu}})du + \sigma \mathbf{b}_u$$

which implies ($\beta = 1 - \gamma$)

$$d\boldsymbol{\mu}_\tau = -\beta(\boldsymbol{\mu}_\tau - \bar{\boldsymbol{\mu}})d\tau + \sigma d\mathbf{b}_\tau$$


Ornstein−Uhlenbeck motion
corr = 0.95
y
x

Stochastic differential equations

General form

$$d\boldsymbol{\mu}_\tau = g(\boldsymbol{\mu}_\tau)dt + h(\boldsymbol{\mu}_\tau)d\mathbf{b}_\tau$$

Stochastic differential equations

General form

$$d\boldsymbol{\mu}_\tau = g(\boldsymbol{\mu}_\tau)dt + h(\boldsymbol{\mu}_\tau)d\mathbf{b}_\tau$$

Solution (assume $h \equiv \sigma$)

$$\boldsymbol{\mu}_\tau = \boldsymbol{\mu}_0 + \int_0^\tau g(\boldsymbol{\mu}_u)du + \sigma\mathbf{b}_\tau$$

Stochastic differential equations

General form

$$d\boldsymbol{\mu}_\tau = g(\boldsymbol{\mu}_\tau)dt + h(\boldsymbol{\mu}_\tau)d\mathbf{b}_\tau$$

Solution (assume $h \equiv \sigma$)

$$\boldsymbol{\mu}_\tau = \boldsymbol{\mu}_0 + \int_0^\tau g(\boldsymbol{\mu}_u)du + \sigma\mathbf{b}_\tau$$

For OU model

$$\boldsymbol{\mu}_\tau = e^{-\beta t}\boldsymbol{\mu}_0 + (1 - e^{-\beta\tau})\bar{\boldsymbol{\mu}} + \boldsymbol{\zeta}_\tau$$

where $[\boldsymbol{\zeta}_\tau] = N\left(\mathbf{0}, \frac{\sigma^2(1-e^{-2\beta\tau})}{2\beta}\mathbf{I}\right)$

Integrated SDEs (Velocity modeling)

New notation:

- $\boldsymbol{\nu}_\tau$ = velocity at time τ
- $H(\boldsymbol{\mu}_\tau)$ = potential function to control movement
- $\nabla H(\cdot)$ = spatial gradient of H

Integrated SDEs (Velocity modeling)

New notation:

- $\boldsymbol{\nu}_\tau$ = velocity at time τ
- $H(\boldsymbol{\mu}_\tau)$ = potential function to control movement
- $\nabla H(\cdot)$ = spatial gradient of H

Movement ISDE

$$
\begin{aligned}
&d\boldsymbol{\nu}_\tau = -\beta\{\boldsymbol{\nu}_\tau - \nabla H(\boldsymbol{\mu}_\tau)\} + \sigma d\mathbf{b}_\tau \\
&d\boldsymbol{\mu}_\tau = \boldsymbol{\nu}_\tau \\
&\Downarrow \\
&\boldsymbol{\mu}_\tau = \int_0^\tau \boldsymbol{\nu}_u du
\end{aligned}
$$

Continuous-time CRWs

CTCRW

$$
\begin{aligned}
&d\boldsymbol{\nu}_\tau = -\beta\boldsymbol{\nu}_\tau + \sigma d\mathbf{b}_\tau \\
&d\boldsymbol{\mu}_\tau = \boldsymbol{\nu}_\tau \\
&\Downarrow \\
&\boldsymbol{\nu}_\tau = \mathrm{OU}(\beta, \sigma) \\
&\boldsymbol{\mu}_\tau = \boldsymbol{\mu}_0 + \int_0^\tau \boldsymbol{\nu}_u du
\end{aligned}
$$

- $\boldsymbol{\nu}_\tau$ is an Ornstein-Uhlenbeck (continuous-time AR(1)) process
- $H \equiv 0$
- $\boldsymbol{\mu}_\tau$ accumulates instantaneous changes in location

Continuous-time CRWs

Some useful properties:

- $\boldsymbol{\nu}_{\tau+\delta} = \boldsymbol{\nu}_{\tau} e^{-\beta\delta} + \boldsymbol{\zeta}_{\tau+\delta}$,

- $\boldsymbol{\mu}_{\tau+\delta} = \boldsymbol{\mu}_{\tau} + \boldsymbol{\nu}_{\tau} \left(\frac{1-e^{-\beta\delta}}{\beta} \right) + \boldsymbol{\xi}_{\tau+\delta}$

Continuous-time CRWs

Some useful properties:

- $\boldsymbol{\nu}_{\tau+\delta} = \boldsymbol{\nu}_\tau e^{-\beta\delta} + \boldsymbol{\zeta}_{\tau+\delta}$,

- $\boldsymbol{\mu}_{\tau+\delta} = \boldsymbol{\mu}_\tau + \boldsymbol{\nu}_\tau \left(\frac{1-e^{-\beta\delta}}{\beta}\right) + \boldsymbol{\xi}_{\tau+\delta}$

- $\boldsymbol{\epsilon}_{\tau+\delta} = (\boldsymbol{\zeta}_{\tau+\delta}, \boldsymbol{\xi}_{\tau+\delta})$ are zero mean independent (through time) normal errors that depend only on δ, β, and σ (**Not** τ!)

Continuous-time CRWs

Some useful properties:

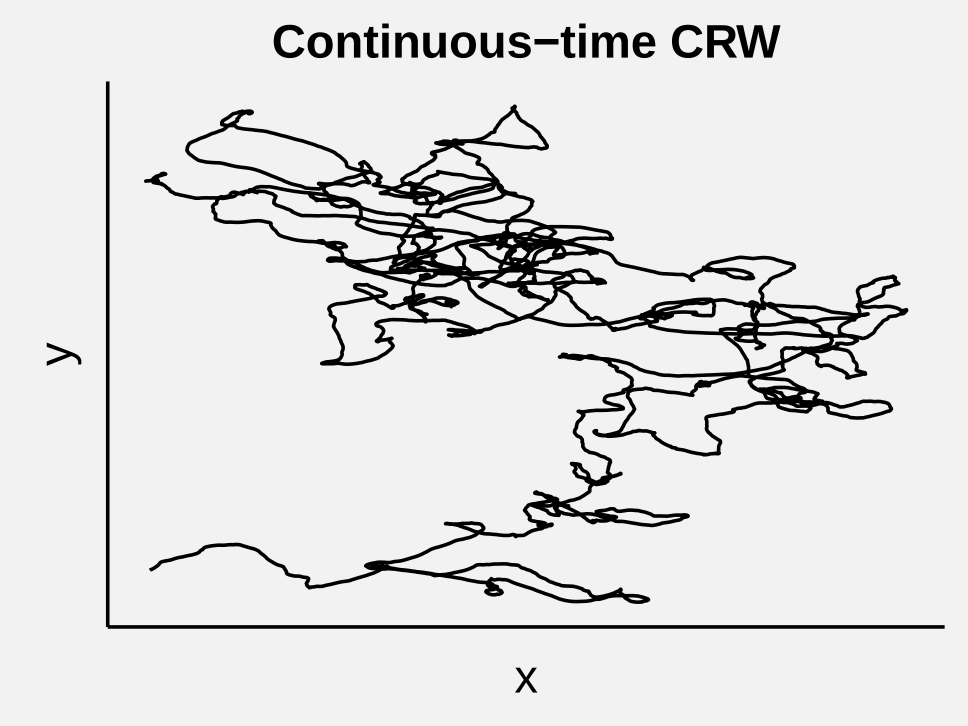
- $\boldsymbol{\nu}_{\tau+\delta} = \boldsymbol{\nu}_\tau e^{-\beta\delta} + \boldsymbol{\zeta}_{\tau+\delta}$,

- $\boldsymbol{\mu}_{\tau+\delta} = \boldsymbol{\mu}_\tau + \boldsymbol{\nu}_\tau \left(\frac{1-e^{-\beta\delta}}{\beta}\right) + \boldsymbol{\xi}_{\tau+\delta}$

- $\boldsymbol{\epsilon}_{\tau+\delta} = (\boldsymbol{\zeta}_{\tau+\delta}, \boldsymbol{\xi}_{\tau+\delta})$ are zero mean independent (through time) normal errors that depend only on δ, β, and σ (**Not** τ!)

- Can we write $\boldsymbol{\mu}_{\tau+\delta}$ just as a function of $\boldsymbol{\mu}_\tau$?

  No. Distribution of $\boldsymbol{\mu}_\tau$ is a function of the whole $\boldsymbol{\mu}_u$, $u < \tau$.


Continuous−time CRW
y
x

Continuous-time CRWs

Some more properties:

- $\text{corr}[\boldsymbol{\nu}_{\tau+\delta}, \boldsymbol{\nu}_\tau] = e^{-\beta\delta}$,
  ≈ 0 for β large
  ≈ 1 for β small

Continuous-time CRWs

Some more properties:

- corr$[\boldsymbol{\nu}_{\tau+\delta}, \boldsymbol{\nu}_\tau] = e^{-\beta\delta}$,
  ≈ 0 for β large
  ≈ 1 for β small

- $\boldsymbol{\mu}_\tau \rightarrow$ Brownian motion as β becomes large
  $\boldsymbol{\mu}_\tau$ becomes very smooth as β becomes small

Continuous-time CRWs

Some more properties:

- corr$[\boldsymbol{\nu}_{\tau+\delta}, \boldsymbol{\nu}_\tau] = e^{-\beta\delta}$,
  ≈ 0 for β large
  ≈ 1 for β small

- $\boldsymbol{\mu}_\tau \to$ Brownian motion as β becomes large
  $\boldsymbol{\mu}_\tau$ becomes very smooth as β becomes small

- $\boldsymbol{\nu}_{\tau+\delta}$ roughly indep. of $\boldsymbol{\nu}_\tau$ at time gap $\delta = 3/\beta$, so, ...

  $\boldsymbol{\mu}_\tau,\ \boldsymbol{\mu}_{\tau+3/\beta},\ \boldsymbol{\mu}_{\tau+6/\beta}, \ldots$ not really distinguishable from Brownian motion

Continuous-time CRWs

What about $H(x, y) \neq 0$? The future?

Continuous-time CRWs

What about $H(x, y) \neq 0$? The future?

Numerical solution to general ISDE model

for small δ

$$\boldsymbol{\nu}_{\tau+\delta} \approx -\beta(\boldsymbol{\nu}_\tau - \nabla H(\boldsymbol{\mu}_\tau))\delta + \boldsymbol{\epsilon}_{\tau+\delta}; \qquad [\boldsymbol{\epsilon}_\tau] = N(\mathbf{0}, \sigma^2\delta\mathbf{I})$$

$$\boldsymbol{\mu}_{\tau+\delta} \approx \boldsymbol{\mu}_\tau + \boldsymbol{\nu}_\tau\delta \implies \boldsymbol{\nu}_\tau \approx (\boldsymbol{\mu}_{\tau+\delta} - \boldsymbol{\mu}_\tau)/\delta$$

Resulting approximation:

$$\boldsymbol{\mu}_{\tau+2\delta} = (2 - \beta\delta)\boldsymbol{\mu}_{\tau+\delta} - (1 - \beta\delta)\boldsymbol{\mu}_\tau + \beta\delta^2\nabla H(\boldsymbol{\mu}_\tau) + \boldsymbol{\epsilon}_{\tau+\delta}$$

Continuous-time CRWs

What about $H(x, y) \neq 0$? The future?

Numerical solution to general ISDE model

for small δ

$$\boldsymbol{\nu}_{\tau+\delta} \approx -\beta(\boldsymbol{\nu}_\tau - \nabla H(\boldsymbol{\mu}_\tau))\delta + \boldsymbol{\epsilon}_{\tau+\delta}; \qquad [\boldsymbol{\epsilon}_\tau] = N(\mathbf{0}, \sigma^2 \delta \mathbf{I})$$

$$\boldsymbol{\mu}_{\tau+\delta} \approx \boldsymbol{\mu}_\tau + \boldsymbol{\nu}_\tau \delta \implies \boldsymbol{\nu}_\tau \approx (\boldsymbol{\mu}_{\tau+\delta} - \boldsymbol{\mu}_\tau)/\delta$$

Resulting approximation:

$$\boldsymbol{\mu}_{\tau+2\delta} = (2 - \beta\delta)\boldsymbol{\mu}_{\tau+\delta} - (1 - \beta\delta)\boldsymbol{\mu}_\tau + \beta\delta^2 \nabla H(\boldsymbol{\mu}_\tau) + \boldsymbol{\epsilon}_{\tau+\delta}$$

Notice that $\boldsymbol{\nu}_\tau$ process disappears and there is a spatial component, $\nabla H(\boldsymbol{\mu}_\tau)$! Something missing from the standard CTCRW model.

Part II

Statistics of animal movement

Inference refresher

Maximum likelihood estimation

Notation

- $\mathbf{d} = (d_1, \ldots, d_n)$ = vector general data
- $\boldsymbol{\theta}$ general set a parameters
- $[d_i|\boldsymbol{\theta}]$ = probability model that generates data
- $L(\boldsymbol{\theta}|\mathbf{d})$ = likelihood function
  typically $L(\boldsymbol{\theta}|\mathbf{d}) = [\mathbf{d}|\boldsymbol{\theta}] = \prod_i [d_i|\boldsymbol{\theta}]$

MLE is very straightforward (in theory) ...

$$\hat{\boldsymbol{\theta}} = \max_{\boldsymbol{\theta}}\{\log L(\boldsymbol{\theta}|\mathbf{d})\}$$

Maximum likelihood estimation

Notation

- $\mathbf{d} = (d_1, \ldots, d_n)$ = vector general data
- $\boldsymbol{\theta}$ general set a parameters
- $[d_i|\boldsymbol{\theta}]$ = probability model that generates data
- $L(\boldsymbol{\theta}|\mathbf{d})$ = likelihood function
  typically $L(\boldsymbol{\theta}|\mathbf{d}) = [\mathbf{d}|\boldsymbol{\theta}] = \prod_i [d_i|\boldsymbol{\theta}]$

MLE is very straightforward (in theory) ...

$$\hat{\boldsymbol{\theta}} = \max_{\boldsymbol{\theta}}\{\log L(\boldsymbol{\theta}|\mathbf{d})\}$$

boom!

MLE details

Large sample theory

If $\mathbf{d} = (d_1 \ldots d_n)$ is 'large' then

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, -\mathbf{H}_{\boldsymbol{\theta}}^{-1}),$$

where $\mathbf{H}_{\boldsymbol{\theta}}$ is the Hessian matrix of $\log L(\boldsymbol{\theta}|\mathbf{d})$.

MLE details

Large sample theory

If $\mathbf{d} = (d_1 \ldots d_n)$ is 'large' then

$$\hat{\boldsymbol{\theta}} \sim N(\boldsymbol{\theta}, -\mathbf{H}_{\boldsymbol{\theta}}^{-1}),$$

where $\mathbf{H}_{\boldsymbol{\theta}}$ is the Hessian matrix of $\log L(\boldsymbol{\theta}|\mathbf{d})$.

Dependent data

If the data are dependent, then $[\mathbf{d}|\boldsymbol{\theta}] \neq \prod_i [d_i|\boldsymbol{\theta}]$.

$$[\mathbf{d}|\boldsymbol{\theta}] = [d_1|\boldsymbol{\theta}] \times [d_2|d_1, \boldsymbol{\theta}] \times [d_3|d_1, d_2, \boldsymbol{\theta}] \times \cdots \times [d_n|d_1, \ldots, d_{n-1}, \boldsymbol{\theta}]$$

If we're lucky, our data are Markov

$$[\mathbf{d}|\boldsymbol{\theta}] = [d_1|\boldsymbol{\theta}] \times [d_2|d_1, \boldsymbol{\theta}] \times [d_3|d_2, \boldsymbol{\theta}] \times \cdots \times [d_n|d_{n-1}, \boldsymbol{\theta}]$$

MLE details

Missing 'data' likelihoods

$$L(\boldsymbol{\theta}|\mathbf{d}_{obs}) = [\mathbf{d}_{obs}|\boldsymbol{\theta}] = \int [\mathbf{d}_{obs}|\mathbf{d}_{mis}, \boldsymbol{\theta}]\ [\mathbf{d}_{mis}|\boldsymbol{\theta}] d\mathbf{d}_{mis}$$

MLE details

Missing ‘data’ likelihoods

$$L(\boldsymbol{\theta}|\mathbf{d}_{obs}) = [\mathbf{d}_{obs}|\boldsymbol{\theta}] = \int [\mathbf{d}_{obs}|\mathbf{d}_{mis}, \boldsymbol{\theta}]\ [\mathbf{d}_{mis}|\boldsymbol{\theta}] d\mathbf{d}_{mis}$$

Penalized likelihood

Sometimes the likelihood is hard to maximize or parameters are not full identifiable. So, a penalty term is added

$$\log L_p(\boldsymbol{\theta}|\mathbf{d}) = \log L(\boldsymbol{\theta}|\mathbf{d}) + \kappa J(\boldsymbol{\theta})$$

We’ll see some examples laster. But this is how spline regressions are fit (e.g., see `mgcv` package).

Bayesian inference

Instead of a fixed quantity, $\boldsymbol{\theta}$, is treated like a random variable itself. Before any data is collected, we might model our uncertainty about the value of $\boldsymbol{\theta}$ with the probability distribution $[\boldsymbol{\theta}]$. This is the 'prior' distribution.

Bayesian inference

Instead of a fixed quantity, $\boldsymbol{\theta}$, is treated like a random variable itself. Before any data is collected, we might model our uncertainty about the value of $\boldsymbol{\theta}$ with the probability distribution $[\boldsymbol{\theta}]$. This is the 'prior' distribution.

We already have the data model $[\mathbf{d}|\boldsymbol{\theta}] = L(\boldsymbol{\theta}|\mathbf{d})$

Bayesian inference

Instead of a fixed quantity, $\boldsymbol{\theta}$, is treated like a random variable itself. Before any data is collected, we might model our uncertainty about the value of $\boldsymbol{\theta}$ with the probability distribution $[\boldsymbol{\theta}]$. This is the 'prior' distribution.

We already have the data model $[\mathbf{d}|\boldsymbol{\theta}] = L(\boldsymbol{\theta}|\mathbf{d})$

Bayes rule and posterior distribution

$$[\boldsymbol{\theta}|\mathbf{d}] = \frac{L(\boldsymbol{\theta}|\mathbf{d})\ [\boldsymbol{\theta}]}{\int [\mathbf{d}|\boldsymbol{\theta}']\ [\boldsymbol{\theta}'] d\boldsymbol{\theta}'}$$

Or, we can look at it on the log scale

$$\log[\boldsymbol{\theta}|\mathbf{d}] = \log L(\boldsymbol{\theta}|\mathbf{d}) + \log[\boldsymbol{\theta}] - const.$$

Bayes inference details

How do we work with a posterior distribution?

- $\hat{\theta}$ = mean, median or mode
- SE of $\hat{\boldsymbol{\theta}}$ = SD of $[\boldsymbol{\theta}|\mathbf{d}]$
- Interval estimates = $(\hat{\boldsymbol{\theta}}_l, \hat{\boldsymbol{\theta}}_u)$ such that $Pr(\hat{\boldsymbol{\theta}}_l < \boldsymbol{\theta}|\mathbf{d} < \hat{\boldsymbol{\theta}}_u) = 0.95$. These are called 'credible intervals'

Bayes inference details

How do we work with a posterior distribution?

- $\hat{\theta}$ = mean, median or mode
- SE of $\hat{\boldsymbol{\theta}} =$ SD of $[\boldsymbol{\theta}|\mathbf{d}]$
- Interval estimates $= (\hat{\boldsymbol{\theta}}_l, \hat{\boldsymbol{\theta}}_u)$ such that $Pr(\hat{\boldsymbol{\theta}}_l < \boldsymbol{\theta}|\mathbf{d} < \hat{\boldsymbol{\theta}}_u) = 0.95$. These are called 'credible intervals'

How do we find these for general posteriors?

- Approximate with a sample $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_m$ from $[\boldsymbol{\theta}|\mathbf{d}]$ and use sample versions
- Numerically (including Monte Carlo) approximate integrals necessary
- Approximate with known distribution that is similar

Telemetry analysis

State-space models

Notation

- $\mathbf{s}_1, \ldots, \mathbf{s}_n$ are observed locations
- $\tau_1, \ldots, \tau_n$ are the observation times
- $\boldsymbol{\mu}_\tau$ is the continuous path of the animal at time τ
- $\boldsymbol{\nu}_\tau$ is the velocity at time τ
- $\boldsymbol{\alpha}_\tau = (\mu_{\tau,x}, \nu_{\tau,x}, \mu_{\tau,y}, \nu_{\tau,y})$,

State-space models

CTCRW model

$$\mathbf{s}_i = \mathbf{z}'\boldsymbol{\alpha}_{\tau_i} + \boldsymbol{\epsilon}_i$$
$$\boldsymbol{\alpha}_{\tau_{i+1}} = \mathbf{T}_i\boldsymbol{\alpha}_{\tau_i} + \boldsymbol{\eta}_i$$

- $[\boldsymbol{\epsilon}_i] = N(\mathbf{0}, \mathbf{V}_i)$; $\mathbf{V}_i$ is the location error variance.
- $\mathbf{z} = (1, 0, 1, 0)$
- $\mathbf{T}_i$ is a function of β and $\delta_i = \tau_{i+1} - \tau_i$
- $[\boldsymbol{\eta}_i] = N(\mathbf{0}, \mathbf{Q}_i)$
- $\mathbf{Q}_i$ depends only on β, δ_i, and σ

Temporally dynamic CTCRW

The parameters do not have to remain constant over time!

- $\tau_1^*, \dots, \tau_m^*$ are known times where β and σ *can* change
- Define model as before based on merged observation and changepoint times, $\tau_1, \dots, \tau_{n+m}$

Temporally dynamic CTCRW

The parameters do not have to remain constant over time!

- $\tau_1^*, \ldots, \tau_m^*$ are known times where β and σ *can* change
- Define model as before based on merged observation and changepoint times, $\tau_1, \ldots, \tau_{n+m}$

Temporally dynamic movement model

$$\mathbf{s}_i = \begin{cases} \mathbf{z}'\boldsymbol{\alpha}_{\tau_i} + \boldsymbol{\epsilon}_i & \text{for } \tau_i \text{ an observed time} \\ \text{NA} & \text{for } \tau_i \text{ in } \tau_1^*, \ldots, \tau_m^* \end{cases}$$

$$\boldsymbol{\alpha}_{\tau_{i+1}} = \mathbf{T}_i \boldsymbol{\alpha}_{\tau_i} + \boldsymbol{\eta}_i$$

Movement model is still continuous in time!

Kalman filter

Method to calculate likelihood, *NOT* a model!

Likelihood for state-space models

$$
\begin{aligned}
L(\boldsymbol{\theta}|\mathbf{s}_{1:n}) &= \prod_i [\mathbf{s}_{i+1}|\mathbf{s}_{1:i}, \boldsymbol{\theta}] \\
&= \int [\mathbf{s}_1|\boldsymbol{\alpha}_{\tau_1}, \theta]\ [\boldsymbol{\alpha}_{\tau_1}|\boldsymbol{\theta}] \ldots [\boldsymbol{\alpha}_{\tau_i}|\boldsymbol{\alpha}_{\tau_{i-1}}, \boldsymbol{\theta}] \times \ldots \\
&\qquad \times [\mathbf{s}_n|\boldsymbol{\alpha}_{\tau_{n+m}}, \theta]\ [\boldsymbol{\alpha}_{\tau_{n+m}}|\boldsymbol{\alpha}_{\tau_{n+m-1}}, \boldsymbol{\theta}] d\boldsymbol{\alpha}_{1:n+m}
\end{aligned}
$$

Kalman filter

Method to calculate likelihood, *NOT* a model!

Likelihood for state-space models

$$L(\boldsymbol{\theta}|\mathbf{s}_{1:n}) = \prod_i [\mathbf{s}_{i+1}|\mathbf{s}_{1:i}, \boldsymbol{\theta}]$$
$$= \int [\mathbf{s}_1|\boldsymbol{\alpha}_{\tau_1}, \theta]\ [\boldsymbol{\alpha}_{\tau_1}|\boldsymbol{\theta}] \dots [\boldsymbol{\alpha}_{\tau_i}|\boldsymbol{\alpha}_{\tau_{i-1}}, \boldsymbol{\theta}] \times \dots$$
$$\times [\mathbf{s}_n|\boldsymbol{\alpha}_{\tau_{n+m}}, \theta]\ [\boldsymbol{\alpha}_{\tau_{n+m}}|\boldsymbol{\alpha}_{\tau_{n+m-1}}, \boldsymbol{\theta}] d\boldsymbol{\alpha}_{1:n+m}$$

Kalman filter is a numerical algorithm that allows calculation of $L(\boldsymbol{\theta}|\mathbf{s}_1, \dots, \mathbf{s}_n)$ in an efficient manner.

- moves forward through the complete likelihood integrating on the way
- requires linear form and normal errors

Kalman smoother

Obtain predictions from model fit

- optimal predictor $\hat{\boldsymbol{\alpha}}_i = E[\boldsymbol{\alpha}_i | \mathbf{s}_{1:n}, \boldsymbol{\theta}]$
- prediction errors $\widehat{\mathrm{var}}(\hat{\boldsymbol{\alpha}}_i) = Var[\boldsymbol{\alpha}_i | \mathbf{s}_{1:n}, \boldsymbol{\theta}]$

Kalman smoother

Obtain predictions from model fit

- optimal predictor $\hat{\boldsymbol{\alpha}}_i = E[\boldsymbol{\alpha}_i | \mathbf{s}_{1:n}, \boldsymbol{\theta}]$
- prediction errors $\widehat{\text{var}}(\hat{\boldsymbol{\alpha}}_i) = Var[\boldsymbol{\alpha}_i | \mathbf{s}_{1:n}, \boldsymbol{\theta}]$

Kalman smoother is an algorithm to calculate mean and variance of $[\boldsymbol{\alpha}_i | \mathbf{s}_{1:n}, \boldsymbol{\theta}]$

- uses output from Kalman filter to go backwards through the model/data to calculate these quantities
- $\boldsymbol{\alpha}_i | \mathbf{s}_{1:n}, \boldsymbol{\theta}$ is normally distributed.

Practical Bayesian inference

Posterior

$$
\begin{aligned}
[\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{s}] &\propto [\mathbf{s} | \boldsymbol{\alpha}, \boldsymbol{\theta}]\ [\boldsymbol{\alpha} | \boldsymbol{\theta}]\ [\boldsymbol{\theta}] \\
&\propto [\boldsymbol{\alpha} | \mathbf{s}, \boldsymbol{\theta}]\ [\boldsymbol{\theta} | \mathbf{s}]
\end{aligned}
$$

Practical Bayesian inference

Posterior

$$
\begin{aligned}
[\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{s}] &\propto [\mathbf{s} | \boldsymbol{\alpha}, \boldsymbol{\theta}] \; [\boldsymbol{\alpha} | \boldsymbol{\theta}] \; [\boldsymbol{\theta}] \\
&\propto [\boldsymbol{\alpha} | \mathbf{s}, \boldsymbol{\theta}] \; [\boldsymbol{\theta} | \mathbf{s}]
\end{aligned}
$$

Approach

1. Approximate $[\boldsymbol{\theta}|\mathbf{s}]$ with something easy to sample from, say $[\boldsymbol{\theta}|\mathbf{s}]^*$

2. Draw $\boldsymbol{\theta}^{(i)} \sim [\boldsymbol{\theta}|\mathbf{s}]^*$ then draw $\boldsymbol{\alpha}^{(i)} \sim [\boldsymbol{\alpha}|\boldsymbol{\theta}^{(i)}, \mathbf{s}]$ ($[\boldsymbol{\alpha}|\boldsymbol{\theta}^{(i)}, \mathbf{s}]$ easy to sample from using KFS algorithms)

3. $(\boldsymbol{\theta}^{(1)}, \boldsymbol{\alpha}^{(1)}), \ldots, (\boldsymbol{\theta}^{(K)}, \boldsymbol{\alpha}^{(K)})$ is a posterior sample

4. $m_i = f(\boldsymbol{\theta}^{(i)}, \boldsymbol{\alpha}^{(i)})$ will be a sample from $[m|\mathbf{s}]$

Approximating $[\boldsymbol{\theta}|\mathbf{s}]$

- Normal approximation
  1. maximize $\log[\mathbf{s}|\boldsymbol{\theta}] + \log[\boldsymbol{\theta}] = L_p(\boldsymbol{\theta}|\mathbf{s})$ (penalized likelihood)
  2. $[\boldsymbol{\theta}|\mathbf{s}]^* = N\left(\hat{\boldsymbol{\theta}}, -\mathbf{H}_{\hat{\boldsymbol{\theta}}}^{-1}\right)$ (possibly truncated)

Approximating $[\boldsymbol{\theta}|\mathbf{s}]$

- Normal approximation
  1. maximize $\log[\mathbf{s}|\boldsymbol{\theta}] + \log[\boldsymbol{\theta}] = L_p(\boldsymbol{\theta}|\mathbf{s})$ (penalized likelihood)
  2. $[\boldsymbol{\theta}|\mathbf{s}]^* = N\left(\hat{\boldsymbol{\theta}}, -\mathbf{H}_{\hat{\boldsymbol{\theta}}}^{-1}\right)$ (possibly truncated)

- Importance sampling (exact sample)
  1. sample $\tilde{\boldsymbol{\theta}}^{(k)} \sim q(\boldsymbol{\theta})$ (maybe normal from last item)
  2. form weights $w_k = [\tilde{\boldsymbol{\theta}}^{(k)}|\mathbf{s}]/q(\tilde{\boldsymbol{\theta}}^{(k)})$
  3. sample $\boldsymbol{\theta}^{(i)}$ from $\tilde{\boldsymbol{\theta}}^{(1)}, \ldots, \tilde{\boldsymbol{\theta}}^{(K)}$ with prob. $\propto w_1, \ldots, w_K$

Approximating $[\boldsymbol{\theta}|\mathbf{s}]$

- Normal approximation
  1. maximize $\log[\mathbf{s}|\boldsymbol{\theta}] + \log[\boldsymbol{\theta}] = L_p(\boldsymbol{\theta}|\mathbf{s})$ (penalized likelihood)
  2. $[\boldsymbol{\theta}|\mathbf{s}]^* = N\left(\hat{\boldsymbol{\theta}}, -\mathbf{H}_{\hat{\boldsymbol{\theta}}}^{-1}\right)$ (possibly truncated)

- Importance sampling (exact sample)
  1. sample $\tilde{\boldsymbol{\theta}}^{(k)} \sim q(\boldsymbol{\theta})$ (maybe normal from last item)
  2. form weights $w_k = [\tilde{\boldsymbol{\theta}}^{(k)}|\mathbf{s}]/q(\tilde{\boldsymbol{\theta}}^{(k)})$
  3. sample $\boldsymbol{\theta}^{(i)}$ from $\tilde{\boldsymbol{\theta}}^{(1)}, \ldots, \tilde{\boldsymbol{\theta}}^{(K)}$ with prob. $\propto w_1, \ldots, w_K$

- Deterministic sample (INLA)
  1. sample $\tilde{\boldsymbol{\theta}}^{(k)}$ from deterministic grid
  2. form weights $w_k = [\tilde{\boldsymbol{\theta}}^{(k)}|\mathbf{s}]$
  3. sample $\boldsymbol{\theta}^{(i)}$ from $\tilde{\boldsymbol{\theta}}^{(1)}, \ldots, \tilde{\boldsymbol{\theta}}^{(K)}$ with prob. $\propto w_1, \ldots, w_K$

Process imputation

- Allows us to account for location uncertainty in other analysis of movement data

- Assume we know $\boldsymbol{\mu}_\tau$ on a sufficiently fine time scale

- Response variable of interest $\mathbf{y} = \mathbf{f}(\boldsymbol{\mu})$, e.g.,
  - Distance traveled,
  - Utilization distribution, or
  - locations of dives

- Model of interest $[\mathbf{y}|\psi]$, e.g.,
  - ANOVA,
  - Spatial regression model, or
  - Point process model

- But,... we don't observe $\boldsymbol{\mu}_\tau$

Process imputation

Solution Average over unknown ‘true’ paths

$$[\psi|\mathbf{s}] = \int [\mathbf{y}_{\boldsymbol{\mu}}|\psi]\ [\psi]\ [\boldsymbol{\mu}|\mathbf{s}] d\boldsymbol{\mu}$$

Process imputation

Solution Average over unknown 'true' paths

$$[\psi|\mathbf{s}] = \int [\mathbf{y}_{\boldsymbol{\mu}}|\psi]\ [\psi]\ [\boldsymbol{\mu}|\mathbf{s}] d\boldsymbol{\mu}$$

Method

1. Simulate $\boldsymbol{\mu}^{(i)} \sim [\boldsymbol{\mu}|\mathbf{s}] \rightarrow$ Calculate $\mathbf{y}^{(i)}$
2. Calculate summaries of $[\psi^{(i)}|\mathbf{y}^{(i)}]$
   - posterior mean
   - posterior predictions
   - UDs
   - MCMC sample
3. Summarize over $\boldsymbol{\mu}^{(i)}$ realizations
   - $E[\psi|\mathbf{s}] = mean(E[\psi|\mathbf{s}])$
   - $Var[\psi|\mathbf{s}] = Var(E[\psi|\mathbf{s}]) + mean(Var[\psi|\mathbf{y}])$

That's all the math folks!

Anyone awake?