

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

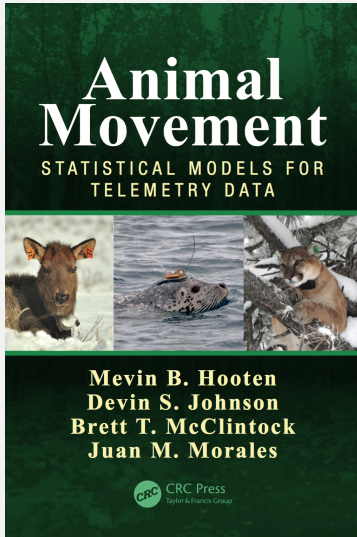
Mathematics

- 1 Discrete-time random walks
- 2 Correlated random walks
- 3 Brownian motion
- 4 Ornstein-Uhlenbeck (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/smoothing (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

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

Notation (discrete time)

- $\mu_t = (\mu_{x,t}, \mu_{y,t})$ 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 .
- $d\mu_t = \mu_t - \mu_{t-1}$ is *movement*!

Random walk

$$\mu_t = \mu_{t-1} + \epsilon_t$$

$$[\epsilon_t] = N(0, \Sigma)$$

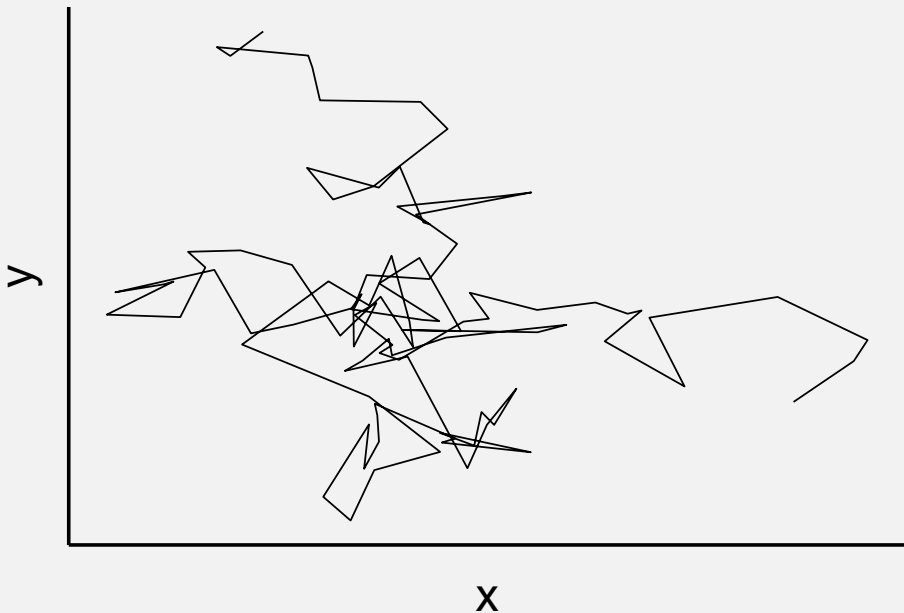
- The movements, $d\mu_t = \epsilon_t$ are all independent of each other.
- Typically $\Sigma \equiv \sigma^2 \mathbf{I}$, so, $\mu_{x,t}$ is independent of $\mu_{y,t}$

An unconditional view

We can rewrite the RW using the accumulation of movements:

$$\begin{aligned}\mu_t &= \mu_{t-1} + \epsilon_t \\ &= \mu_{t-2} + \epsilon_{t-1} + \epsilon_t \\ &= \mu_0 + \sum_{i=1}^t \epsilon_i\end{aligned}$$

random walk (n = 100)



Vector autoregressive models (VAR)

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

VAR(1) model

$$\begin{aligned}\mu_t &= \bar{\mu} + \mathbf{M}(\mu_{t-1} - \bar{\mu}) + \epsilon_t \\ &= (\mathbf{I} - \mathbf{M})\bar{\mu} + \mathbf{M}\mu_{t-1} + \epsilon_t\end{aligned}$$

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

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

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

VAR(1) (n = 100)



Modeling velocity

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

velocity = derivative of location process

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

$$\begin{aligned}\text{velocity} &= d\mu_t \\ &= \mu_t - \mu_{t-1}\end{aligned}$$

VAR(1) model for velocity

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

$$d\mu_t = \mathbf{M}d\mu_{t-1} + \epsilon_t$$

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 , $d\mu_t$, tends to be similar to the previous step, $d\mu_{t-1}$
- 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., $t \ll \infty$)

Correlated random walk

If we model $d\mu_t = \mathbf{M} d\mu_{t-1} + \epsilon_t$, what does this mean for μ_t ?

Recall we can always write the position process like this:

$$\begin{aligned}\mu_t &= \mu_{t-1} + (\mu_t - \mu_{t-1}) \\ &= \mu_{t-2} + (\mu_{t-1} - \mu_{t-2}) + (\mu_t - \mu_{t-1}) \\ &\vdots \\ &= \mu_0 + \sum_{i=1}^t (\mu_i - \mu_{i-1}) \\ &= \mu_0 + \sum_{i=1}^t d\mu_i\end{aligned}$$

Notice that there was no distributional assumption for $d\mu_t$.
It's a simple recursion.

Correlated random walk

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

$$\begin{aligned}\mu_t &= \mu_{t-1} + \mathbf{M}d\mu_{t-1} + \epsilon_t \\ &= \mu_{t-1} + \mathbf{M}(\mu_{t-1} - \mu_{t-2}) + \epsilon_t \\ &= (\mathbf{I} + \mathbf{M})\mu_{t-1} - \mathbf{M}\mu_{t-2} + \epsilon_t\end{aligned}$$

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

The \mathbf{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,

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

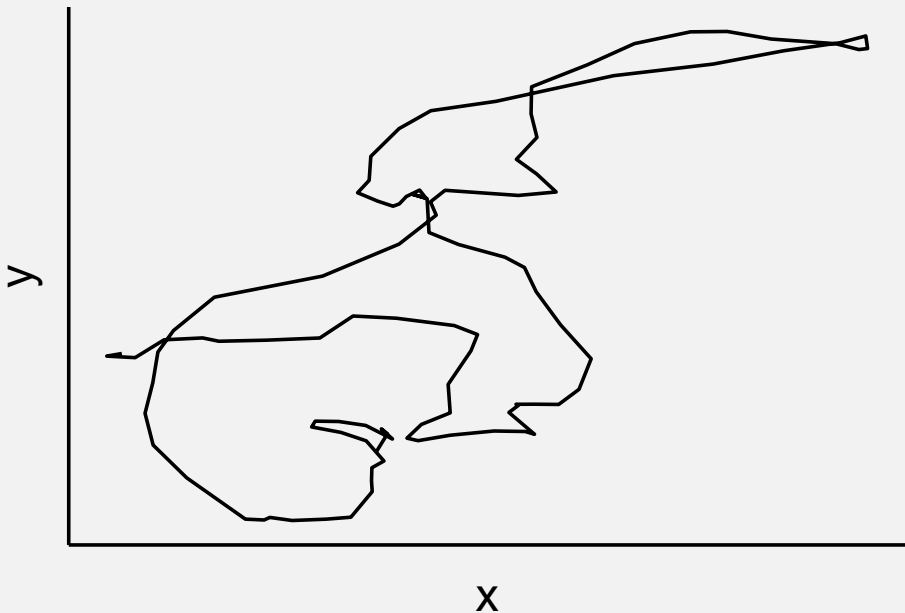
Then $\text{corr}(d\boldsymbol{\mu}_t, d\boldsymbol{\mu}_s) = \gamma^{|t-s|}$

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)



Continuous-time models

Brownian motion

Forms the basis of all continuous-time models

Random walk in continuous-time

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

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

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

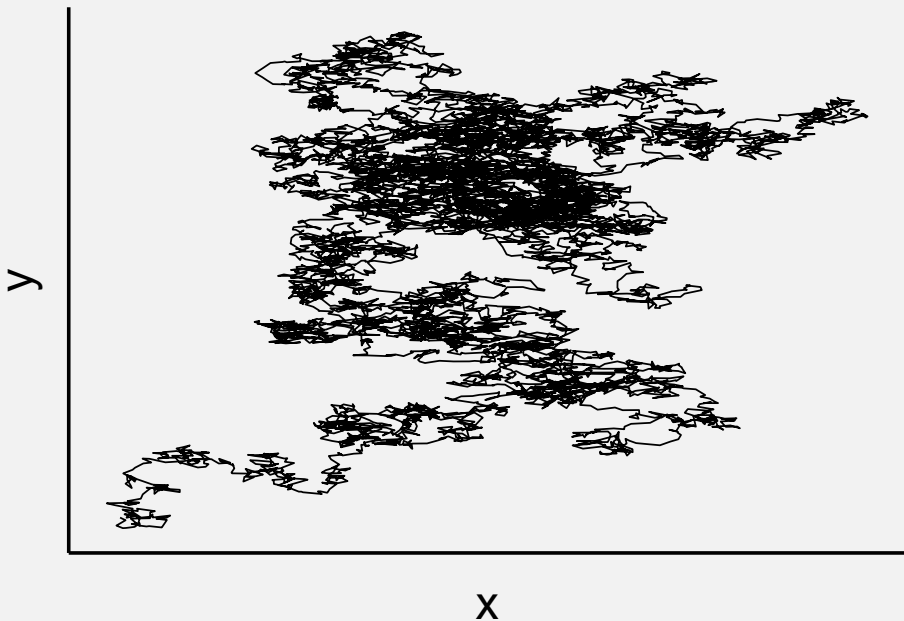
We can get to continuous-time by making the time gaps, $\delta = \tau_j - \tau_{j-1}$, smaller and smaller

Definition

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

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

Brownian motion



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?

$$(1) \quad \mu_t = \gamma(\mu_{t-1} - \bar{\mu}) + \epsilon_t$$

$$\begin{aligned}(2) \quad \mu_t &= \mu_0 + \sum_{j=1}^t [\mu_j - \mu_{j-1}] \\ &= \mu_0 + \sum_{j=1}^t (\gamma - 1)(\mu_{j-1} - \bar{\mu}) + \sum_{j=1}^t \epsilon_j\end{aligned}$$

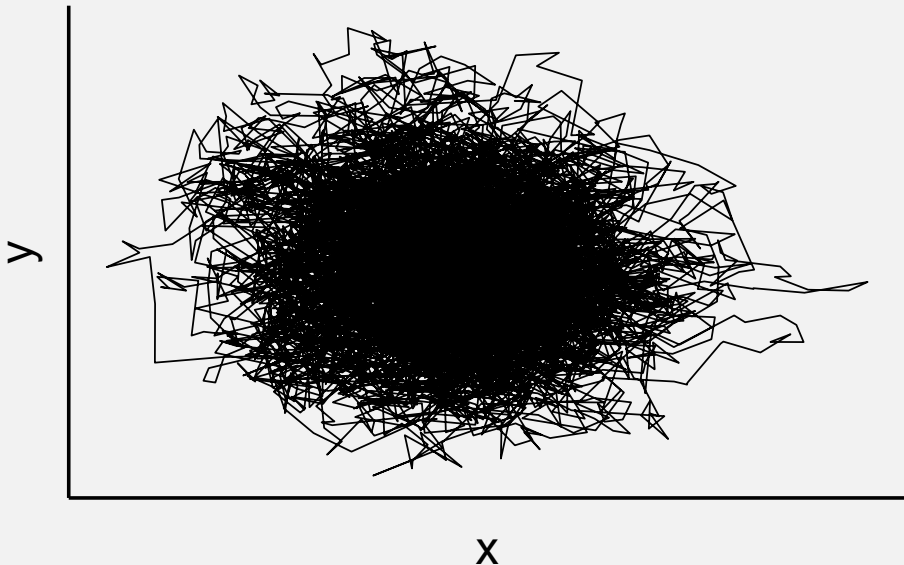
$$(3) \quad \mu_\tau = \mu_0 + \int_0^\tau (\gamma - 1)(\mu_u - \bar{\mu}) du + \sigma \mathbf{b}_\tau$$

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

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

Ornstein–Uhlenbeck motion

corr = 0.95



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

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

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

Movement ISDE

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

$$d\boldsymbol{\mu}_\tau = \boldsymbol{\nu}_\tau$$

\Downarrow

$$\boldsymbol{\mu}_\tau = \int_0^\tau \boldsymbol{\nu}_u du$$

Continuous-time CRWs

CTCRW

$$d\boldsymbol{\nu}_\tau = -\beta\boldsymbol{\nu}_\tau + \sigma d\mathbf{b}_\tau$$

$$d\boldsymbol{\mu}_\tau = \boldsymbol{\nu}_\tau$$

$$\Downarrow$$

$$\boldsymbol{\nu}_\tau = \text{OU}(\beta, \sigma)$$

$$\boldsymbol{\mu}_\tau = \boldsymbol{\mu}_0 + \int_0^\tau \boldsymbol{\nu}_u du$$

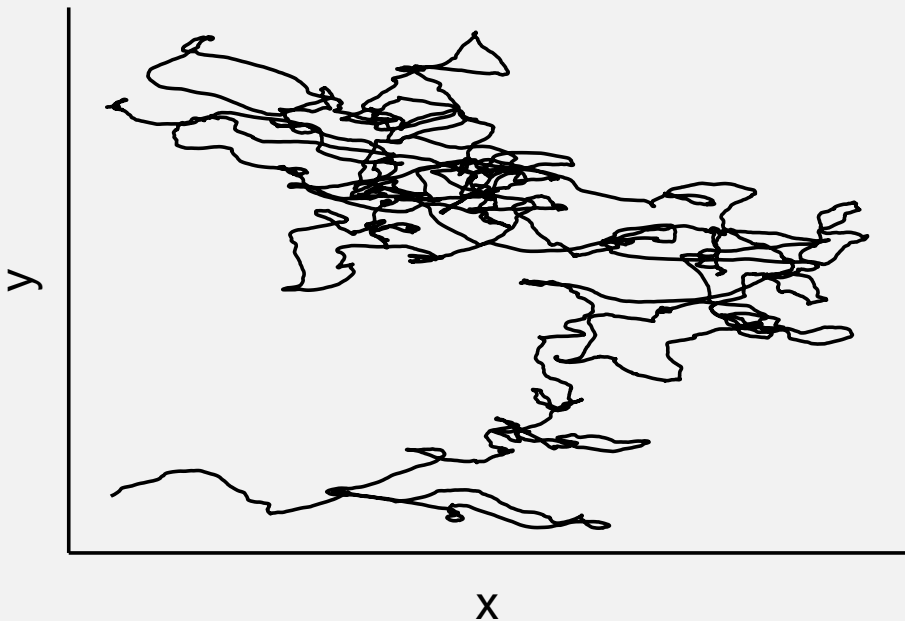
- $\boldsymbol{\nu}_\tau$ is an Ornstein-Uhlenbeck (continuous-time AR(1)) process
- $H \equiv 0$
- $\boldsymbol{\mu}_\tau$ accumulates instantaneous changes in location

Continuous-time CRWs

Some useful properties:

- $\nu_{\tau+\delta} = \nu_{\tau} e^{-\beta\delta} + \zeta_{\tau+\delta},$
- $\mu_{\tau+\delta} = \mu_{\tau} + \nu_{\tau} \left(\frac{1-e^{-\beta\delta}}{\beta} \right) + \xi_{\tau+\delta}$
- $\epsilon_{\tau+\delta} = (\zeta_{\tau+\delta}, \xi_{\tau+\delta})$ are zero mean independent (through time) normal errors that depend only on δ , β , and σ (**Not** τ !)
- Can we write $\mu_{\tau+\delta}$ just as a function of μ_{τ} ?
No. Distribution of μ_{τ} is a function of the whole $\mu_u, u < \tau$.

Continuous-time CRW



Continuous-time CRWs

Some more properties:

- $\text{corr}[\nu_{\tau+\delta}, \nu_{\tau}] = e^{-\beta\delta},$
 ≈ 0 for β large
 ≈ 1 for β small
- $\mu_{\tau} \rightarrow$ Brownian motion as β becomes large
 μ_{τ} becomes very smooth as β becomes small
- $\nu_{\tau+\delta}$ roughly indep. of ν_{τ} at time gap $\delta = 3/\beta$, so, ...
 $\mu_{\tau}, \mu_{\tau+3/\beta}, \mu_{\tau+6/\beta}, \dots$ not really distinguishable from Brownian motion

Continuous-time CRWs

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

Numerical solution to general ISDE model

for small δ

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

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

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

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

MLE is very straightforward (in theory) ...

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

boom!

MLE details

Large sample theory

If $\mathbf{d} = (d_1 \dots d_n)$ 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 $[\mathbf{d}|\boldsymbol{\theta}] \neq \prod_i [d_i|\boldsymbol{\theta}]$.

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

If we're lucky, our data are Markov

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

MLE details

Missing 'data' likelihoods

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

Penalized likelihood

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

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

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

Bayesian inference

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

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

Bayes rule and posterior distribution

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

Or, we can look at it on the log scale

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

Bayes inference details

How do we work with a posterior distribution?

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

How do we find these for general posteriors?

- Approximate with a sample $\theta_1, \dots, \theta_m$ from $[\theta|\mathbf{d}]$ and use sample versions
- Numerically (including Monte Carlo) approximate integrals necessary
- Approximate with known distribution that is similar

Telemetry analysis

State-space models

Notation

- $\mathbf{s}_1, \dots, \mathbf{s}_n$ are observed locations
- τ_1, \dots, τ_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 movement model

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

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

Movement model is still continuous in time!

Kalman filter

Method to calculate likelihood, *NOT* a model!

Likelihood for state-space models

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

Kalman filter is a numerical algorithm that allows calculation of $L(\theta | \mathbf{s}_1, \dots, \mathbf{s}_n)$ 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 $\hat{\alpha}_i = E[\alpha_i | \mathbf{s}_{1:n}, \theta]$
- prediction errors $\widehat{\text{var}}(\hat{\alpha}_i) = \text{Var}[\alpha_i | \mathbf{s}_{1:n}, \theta]$

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

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

Practical Bayesian inference

Posterior

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

Approach

- 1 Approximate $[\theta | \mathbf{s}]$ with something easy to sample from, say $[\theta | \mathbf{s}]^*$
- 2 Draw $\theta^{(i)} \sim [\theta | \mathbf{s}]^*$ then draw $\alpha^{(i)} \sim [\alpha | \theta^{(i)}, \mathbf{s}]$
($[\alpha | \theta^{(i)}, \mathbf{s}]$ easy to sample from using KFS algorithms)
- 3 $(\theta^{(1)}, \alpha^{(1)}), \dots, (\theta^{(K)}, \alpha^{(K)})$ is a posterior sample
- 4 $m_i = f(\theta^{(i)}, \alpha^{(i)})$ will be a sample from $[m | \mathbf{s}]$

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

- Normal approximation
 - ① maximize $\log[\mathbf{s}|\theta] + \log[\theta] = L_p(\theta|\mathbf{s})$ (penalized likelihood)
 - ② $[\theta|\mathbf{s}]^* = N\left(\hat{\theta}, -\mathbf{H}_{\hat{\theta}}^{-1}\right)$ (possibly truncated)
- Importance sampling (exact sample)
 - ① sample $\tilde{\theta}^{(k)} \sim q(\theta)$ (maybe normal from last item)
 - ② form weights $w_k = [\tilde{\theta}^{(k)}|\mathbf{s}]/q(\tilde{\theta}^{(k)})$
 - ③ sample $\theta^{(i)}$ from $\tilde{\theta}^{(1)}, \dots, \tilde{\theta}^{(K)}$ with prob. $\propto w_1, \dots, w_K$
- Deterministic sample (INLA)
 - ① sample $\tilde{\theta}^{(k)}$ from deterministic grid
 - ② form weights $w_k = [\tilde{\theta}^{(k)}|\mathbf{s}]$
 - ③ sample $\theta^{(i)}$ from $\tilde{\theta}^{(1)}, \dots, \tilde{\theta}^{(K)}$ with prob. $\propto w_1, \dots, w_K$

Process imputation

- Allows us to account for location uncertainty in other analysis of movement data
- Assume we know μ_τ on a sufficiently fine time scale
- Response variable of interest $\mathbf{y} = \mathbf{f}(\mu)$, 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 μ_τ

Process imputation

Solution Average over unknown 'true' paths

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

Method

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

That's all the math folks!

Anyone awake?

