Statistical Parameter Estimation and Inference for Dynamical Models

Jennifer Hoeting
Colorado State University
www.stat.colostate.edu/∼jah

UQ18 Minitutorial

April 2018

Minitutorial: Overview

1. Overview
2. Motivating example
3. Methods for statistical parameter estimation and inference for dynamical models

   Part 1: Parameter estimation for SDEs

   Part 2: Model selection to choose between different types of dynamical models
4. Conclusions

Minitutorial: Statistics topics

1. Expectation
2. Monte Carlo Integration
3. Importance sampling

4. Maximum likelihood estimation
5. Penalized maximum likelihood estimation
6. Simulated maximum likelihood estimation

7. Bayesian statistics
8. Markov chain Monte Carlo
9. Approximate Bayesian Computation

Statistics: Three inference paradigms in statistics

1. Frequentist inference
   - Procedures evaluated based on repeated sampling
   - Parameters: assumed fixed, but unknown
   - Pre-experimental set-up
2. Likelihood inference
   - Assume a sampling model (the likelihood)
   - Parameters: assumed fixed, but unknown
   - Post-experimental set-up
3. Bayesian inference
   - Assume a sampling model (the likelihood) and a prior distribution for the parameters
   - Parameters: assumed to be random variables with distributions
   - Post-experimental set-up

Motivation

Classification of modeling approaches

Mathematical	Statistical
Deterministic	Stochastic
Mechanistic	Phenomenological
Process	Pattern

Traditional approach:

choose between a mathematical <u>or</u> a statistical approach

Modern approach:

combine the advantages of mathematical and statistical models

History

There is a long history of combining the advantages of mathematical and statistical models.

Many of the speakers at UQ18 exemplify these ideas.

Some useful references:

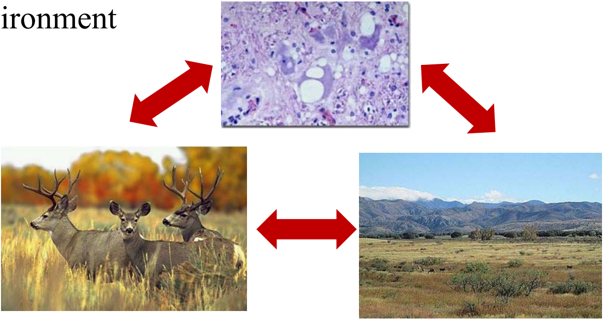
- Classic references by L. M. Berliner:
  - ”Statistics, Probability and Chaos,” *Statistical Science* (1992)
  - “Likelihood and Bayesian Prediction of Chaotic Systems” *JASA* (1991).
- Recent review:
  “Statistical inference for dynamical systems: A review”
  K. McGoff, S. Mukherjee, N. Pillai, *Statistics Surveys* (2015)
- Statistical methods: *Computational Statistics*
  G. Givens and J. Hoeting, 2013, Wiley, 2nd edition.

Motivating example

Ecology of infectious diseases

Explores the relationships between

1. Diseases: Parasitic, bacterial, viral infectious
2. Hosts: animal and human
3. Their environment



Motivating example

Challenges in the ecology of infectious diseases

- Data can be messy and sparse
- You have to find the animals who are sick, they don't visit the nearest health clinic
- You need knowledge of mathematical biology
- You need to able to synthesize a broad range of statistical methods

Sounds like fun!

Motivating example:

Chronic Wasting Disease

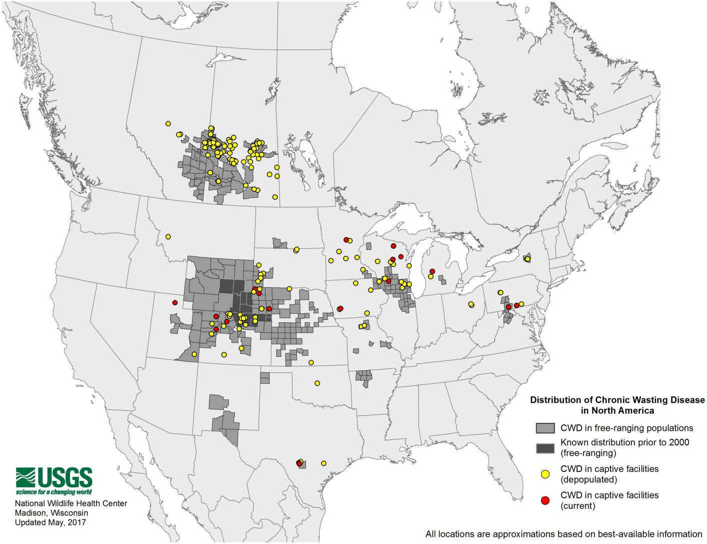
Deer (female) with
Chronic Wasting Disease

Healthy deer (male)

Motivating example: CWD Transmission

- Chronic wasting disease (CWD) is a 100% fatal contagious disease that affects cervids: deer, elk and moose.
- It is important to understand the transmission mechanisms of CWD.
- Several deterministic epidemic models were proposed by Miller, Hobbs & Tavener (2006) in an effort to understand the transmission of CWD.

Motivating example: CWD Transmission



Distribution of Chronic Wasting Disease in North America

- ▨ CWD in free-ranging populations
- ■ Known distribution prior to 2000 (free-ranging)
- ● CWD in captive facilities (depopulated)
- ● CWD in captive facilities (current)

USGS science for a changing world

National Wildlife Health Center
Madison, Wisconsin
Updated May, 2017

All locations are approximations based on best-available information

Motivating example: CWD Transmission

The observed data:

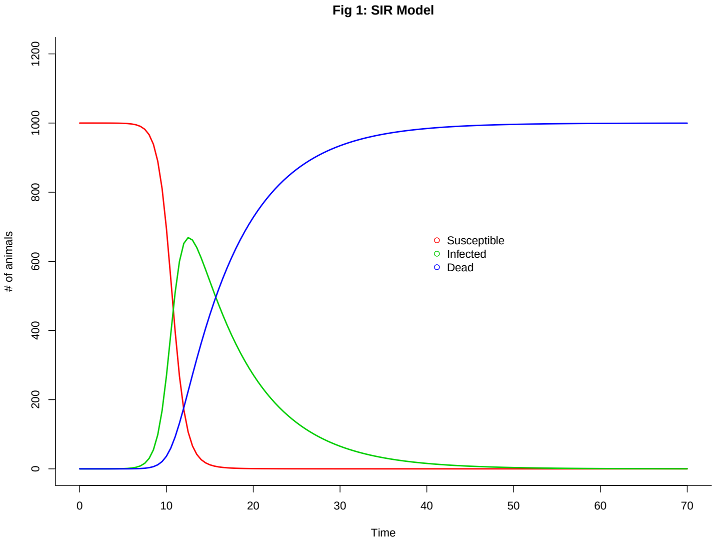
- Deer were held at the Colorado Division of Wildlife in Fort Collins
- Annual observations of cumulative mortality from two CWD epidemics in captive mule deer
- No live-animal test, vaccine, or treatment for CWD existed prior to 2008.
- Epidemic 1: 1974 to 1985
  Epidemic 2: 1992 to 2001 (in a new deer herd)
- 21 observations over time
- The dataset also includes
  - annual number of new deer added to the herd
  - per capita losses due to natural deaths and removals

Motivating example
CWD Transmission

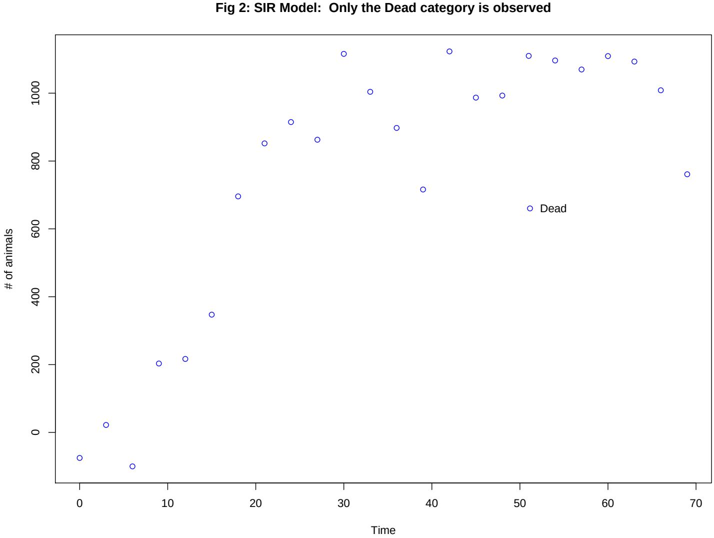
Mathematical model for Chronic Wasting Disease

SIR model (Susceptible-Infected-Recovered)

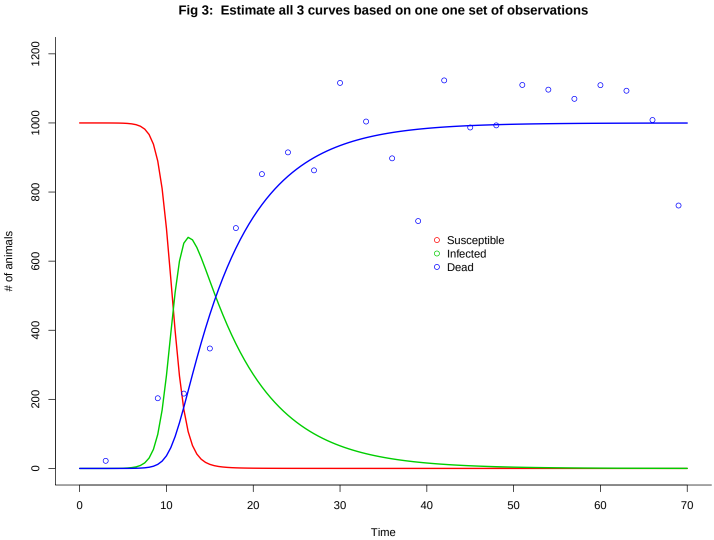
Motivating example
CWD Transmission



Motivating example
CWD Transmission



Motivating example
CWD Transmission



Motivating example: CWD Transmission

We develop a type of Susceptible-Infected-Recovered (SIR) model for disease transmission where the state variables are described by a set of differential equations.

Consider the state vector $\mathbf{X}(t) = (S(t), I(t), C(t))^T$, where

- S is the number of susceptible animals,
- I is the number of infected animals,
- C is the cumulative number of deaths from CWD over time.

Only C is observed and the other two state variables, S and I, are unobserved.

Motivating example:
Direct transmission ODE model for CWD

$$
\begin{aligned}
dS &= [a - S(\beta I + m)]\, dt \\
dI &= [\beta S I - I(\mu + m)]\, dt \\
dC &= \mu I\, dt
\end{aligned}
$$

where

- β is the transmission coefficient
- μ is the per capita CWD mortality rate

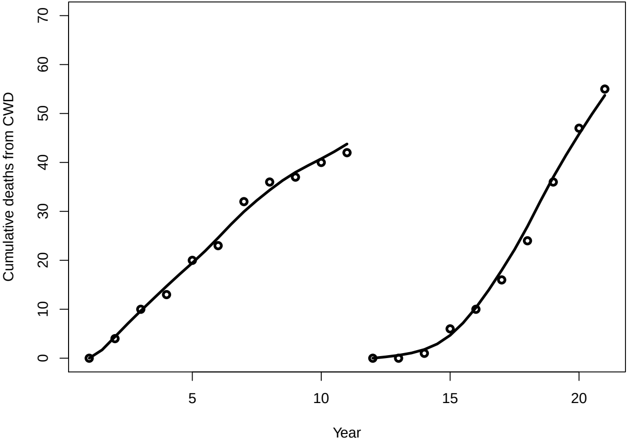
unknown

- a is the number of susceptible animals annually added to the population via births or importation
- m is the per capita natural mortality rate

known

We assume $\mathbf{X}(0) = (S(0), I(0), C(0))^T$ are known initial conditions.

Motivating example: Ordinary Differential Equation Model for CWD



Deterministic dynamical models can be used to determine whether or not transmission will occur.

Motivating example: Direct transmission SDE Model for CWD

A SDE model for direct transmission of CWD is given by

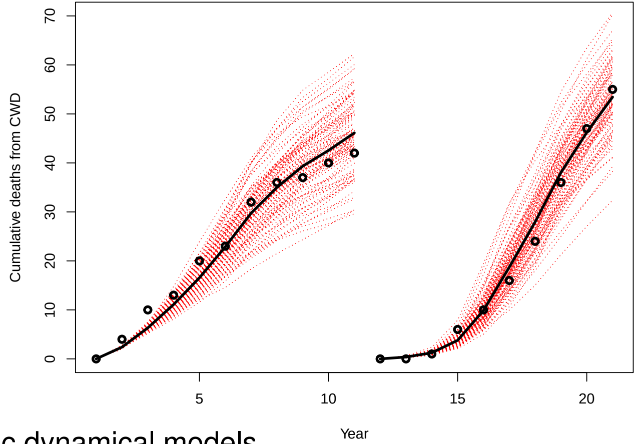
$$\begin{aligned} dS =& [a - S(\beta I + m)]dt + B_{11}dW_1 + B_{12}dW_2 + B_{13}dW_3, \\ dI =& [\beta SI - I(\mu + m)]dt + B_{21}dW_1 + B_{22}dW_2 + B_{23}dW_3, \\ dC =& \mu I dt + B_{31}dW_1 + B_{32}dW_2 + B_{33}dW_3, \end{aligned}$$

where

- initial condition $\mathbf{X}(0) = (S(0), I(0), C(0))^T$ assumed known
- **W** is a *k*-dimensional standard Wiener process.
- $B = (B_{ij}) = \sqrt{\Sigma}$ with

$$\Sigma = \begin{bmatrix} a + S(\beta I + m) & -\beta SI & 0 \\ -\beta SI & \beta SI + I(\mu + m) & -\mu I \\ 0 & -\mu I & \mu I \end{bmatrix}.$$

Motivating example:
Stochastic Differential Equation Model for CWD



Stochastic dynamical models

- Can be used to determine the probability of disease transmission between two individuals
- Allow more realistic description of the transmission of disease.

Part 1: PSML Parameter Estimation for SDEs

Goal: Estimate the parameters of SDE model from the data.

Challenges:

- Data are partially observed, discrete, sparse.
- The transition density between two observations is typically unknown.

Key ideas:

- We propose a new importance sampling approach with an auxiliary parameter which improves the approximation of the transition density.
- We embed the auxiliary importance sampler in a penalized simulated maximum likelihood (PSML) framework which produces more accurate and computationally efficient parameter estimates.

PSML: General Multivariate SDE Model

Consider a general multivariate stochastic differential equation model,

$$d\mathbf{X}(t) = f(\mathbf{X}(t), \theta)dt + g(\mathbf{X}(t), \theta)d\mathbf{W}(t)$$

where

- $\mathbf{X}(t) = \{X_1(t), \ldots, X_k(t)\}^T$ denotes a *k*-dimensional state variable vector at time *t*,
- Initial conditions $\mathbf{X}(0) = \mathbf{x}_0$ are assumed known,
- $\boldsymbol{\theta} \in \Theta \subseteq \mathbb{R}^p$ is an unknown *p*-dimensional parameter vector,
- **W** is a *k*-dimensional standard Wiener process.
- the functions *f* and *g* are known

We also assume that this SDE has a unique weak solution.

PSML: Partially observed state process

We assume that only a subset of the states can be observed at each discrete time point.

Partition $\boldsymbol{X}$ so at time t, $\boldsymbol{X}(t) = \{\boldsymbol{X}_{-\text{obs}}(t), \boldsymbol{X}_{\text{obs}}(t)\}$

- unobservable state variables $\boldsymbol{X}_{-\text{obs}}(t) = \{X_1(t), \ldots, X_{j-1}(t)\}$
- observable states variables $\boldsymbol{X}_{\text{obs}}(t) = \{X_j(t), \ldots, X_k(t)\}$

Note that the time intervals do not have to be equidistant.

PSML: Likelihood

The discrete-time likelihood of the general multivariate SDE model is given by

$$L(\theta) = p(\boldsymbol{X}_{\text{obs}}(t_1)|\boldsymbol{X}(t_0), \theta) \prod_{i=2}^{n} p(\boldsymbol{X}_{\text{obs}}(t_i)|\boldsymbol{X}(t_0), \boldsymbol{X}_{\text{obs}}(t_1 : t_{i-1}), \theta)$$

where $\boldsymbol{X}_{\text{obs}}(t_1 : t_{i-1})$ denotes all observations of $\boldsymbol{X}_{\text{obs}}$ from time t_1 to t_{i-1}.

PSML: Approximate likelihood

Likelihood is based on

1. Transition density $p(\mathbf{X}(t_i)|\mathbf{X}(t_{i-1}))$ which typically has no closed form
2. Approximation of $p(\mathbf{X}(t_i)|\mathbf{X}(t_{i-1}))$ based on multivariate integral

Approximate the likelihood

1. Use Euler-Maruyama scheme which allows us to approximate $p(\mathbf{X}(t_i)|\mathbf{X}(t_{i-1}))$
2. Use importance sampling to approximate integral expression in the likelihood

PSML: Euler-Maruyama

Goal: approximate the transition probability density $p(\mathbf{X}(t_i)|\mathbf{X}(t_{i-1}))$ which has no closed form in most cases.

If the time interval between two observations is small enough, we can approximate $p(\mathbf{X}(t_i)|\mathbf{X}(t_{i-1}))$ using a multivariate normal density.

The Euler-Maruyama scheme:

$$\mathbf{X}(t+\delta) - \mathbf{X}(t) \approx f(\mathbf{X}(t), \boldsymbol{\theta})\delta + g(\mathbf{X}(t), \boldsymbol{\theta})(\mathbf{W}(t+\delta) - \mathbf{W}(t)),$$

where

- δ is called step size
- $\mathbf{W}(t+\delta) - \mathbf{W}(t) \sim N(0, \delta\mathcal{I}_{k\times k})$

PSML: What if your time intervals are large?

If the time interval between observations is large, the Euler-Maruyama approximation will introduce bias.

We can partition the interval t_{i-1} to t_i to M subintervals such that $\delta = (t_i - t_{i-1})/M$ is small enough for the Euler-Maruyama scheme.

By the Markov property, $p(\mathbf{X}(t_i)|\mathbf{X}(t_{i-1}))$ can be estimated by

$$\int \prod_{m=1}^{M} p(\mathbf{X}(t_{i-1} + m\delta)|\mathbf{X}(t_{i-1} + (m-1)\delta))d\mathbf{X}((t_{i-1} + \delta) : (t_i - \delta)),$$

(Pedersen 1995)

PSML

We have a formula for the approximate transition density.

Our next goal is to estimate parameters of the density.

How will we do that? We will generate samples from the approximate transition density using importance sampling.

So, first, what is importance sampling?

Statistics

1. Expectation
2. Monte Carlo Integration
3. Importance sampling

4. Maximum likelihood estimation
5. Penalized maximum likelihood estimation
6. Simulated maximum likelihood estimation

7. Bayesian statistics
8. Markov chain Monte Carlo
9. Approximate Bayesian Computation

Statistics: Expectation

Expectation is the backbone of much of the field of statistics.

The expected value is the average value of a random variable.

Let $\mathbf{X}$ be a random variable with distribution $p(\mathbf{x})$. Then

$$E(\mathbf{X}) = \int \mathbf{x} p(\mathbf{x})\, d\mathbf{x}$$

Many quantities of interest in inferential statistical analyses can be expressed as the expectation of a function of a random variable, $E(h(X))$, where

$$\mu = E(h(\mathbf{X})) = \int h(\mathbf{X}) p(\mathbf{x})\, d\mathbf{x}$$

Statistics: Monte Carlo integration

$E\left(h(\mathbf{X})\right)$ isn't always available in closed form, so we use alternative approaches to approximate it.

Monte Carlo Integration

When an i.i.d. random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ is obtained from p, we can approximate μ by a sample average:

$$\hat{\mu}_{\mathrm{MC}} = \frac{1}{n}\sum_{i=1}^{n} h(\mathbf{X}_i) \quad \longrightarrow \quad \mu = E\left(h(\mathbf{X})\right) = \int h(\mathbf{x})p(\mathbf{x})\, d\mathbf{x}$$

as $n \to \infty$ by the strong law of large numbers.

Statistics: Importance Sampling

Importance sampling is based upon the principle that the expectation of $h(\mathbf{X})$ with respect to its density p can be written in the alternative form

$$\mu = \int h(\mathbf{x})p(\mathbf{x})\, d\mathbf{x} = \int h(\mathbf{x})\frac{p(\mathbf{x})}{q(\mathbf{x})}q(\mathbf{x})\, d\mathbf{x}$$

where q is another density function called the *importance sampling function* or *envelope*.

Statistics: Importance Sampling

This suggests that a Monte Carlo approach to estimating $\mathrm{E}\{h(\mathbf{X})\}$ is to draw $\mathbf{X}_1, \ldots, \mathbf{X}_n$ i.i.d. from q and use the estimator

$$\hat{\mu}^*_{\mathrm{IS}} = \frac{1}{n} \sum_{i=1}^{n} h(\mathbf{X}_i) w^*(\mathbf{X}_i)$$

where $w^*(\mathbf{X}_i) = p(\mathbf{X}_i)/q(\mathbf{X}_i)$ are unstandardized weights, also called *importance ratios*.

For this strategy to be convenient, it must be easy to sample from q and to evaluate p, even when it is not easy to sample from p.

PSML: Importance sampling approximation

Back to our problem:
Using importance sampling, we draw J samples, $\{\mathbf{X}^{(j)}((t_{i-1}+\delta):(t_i-\delta)), j = 1, \cdots, J\}$, from an importance sampling density q.

Then $p(\mathbf{X}(t_i)|\mathbf{X}(t_{i-1}))$ can be approximated by

$$\frac{1}{J}\sum_{j=1}^{J}\frac{\prod_{m=1}^{M} p(\mathbf{X}^{(j)}(t_{i-1}+m\delta)\,|\,\mathbf{X}^{(j)}(t_{i-1}+(m-1)\delta))}{q(\mathbf{X}^{(j)}((t_{i-1}+\delta):(t_i-\delta)))} \tag{1}$$

To decrease the importance sampler variance and reduce the sample size J, we want to choose an importance sampling function q that is as close as possible to the integrand in the likelihood

This is very challenging for our SDE problem!

Statistics

1. Expectation
2. Monte Carlo Integration
3. Importance sampling

4. Maximum likelihood estimation
5. Penalized maximum likelihood estimation
6. Simulated maximum likelihood estimation

7. Bayesian statistics
8. Markov chain Monte Carlo
9. Approximate Bayesian Computation

Statistics: Maximum Likelihood Estimation

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sequence of random variables with joint density $p(\mathbf{x}|\boldsymbol{\theta})$ viewed as a function of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)$.

The likelihood function is the density viewed as a function of $\boldsymbol{\theta}$ so

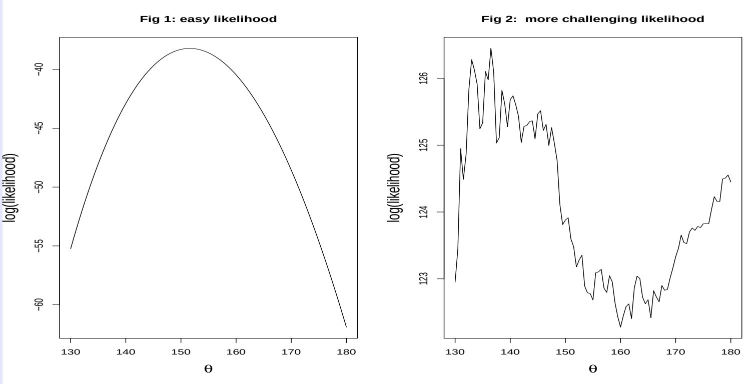
$$L(\boldsymbol{\theta}|\mathbf{x}) = p(x_1, \ldots, x_n|\theta_1, \ldots, \theta_p)$$

- The observed data, $x_1, \ldots, x_n$, might have been realized under many different values for $\boldsymbol{\theta}$.
- The parameters for which observing $x_1, \ldots, x_n$ would be most likely constitute the maximum likelihood estimate (MLE) of $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}}_{MLE} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\max}\, L(\boldsymbol{\theta}|\mathbf{x})$$

- We often consider the log likelihood, $\ell(\boldsymbol{\theta}|\mathbf{x})$.

Statistics: Maximum Likelihood Estimation



- For simple univariate problems, it is easy to find the maximum likelihood estimate (MLE).
- For many real-world problems finding the MLE can be challenging (multivariate, multiple maxima, etc).

Statistics: Penalized MLE

One drawback with likelihood inference across multiple models is that the likelihood increases with model complexity.

Penalized MLE has been used in many contexts, particularly for simultaneous model selection and parameter estimation in regression-type models (e.g., LASSO).

The general idea is to maximize a penalized likelihood of the form

$$\hat{\boldsymbol{\theta}}_{PMLE} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\arg\max} \; \{\ell(\boldsymbol{\theta}|\mathbf{x}) - \lambda P(\boldsymbol{\theta})\}$$

where

- λ controls the tradeoff between the likelihood and the penalty
- P is a function that penalizes less attractive models (e.g., penalty for model complexity or higher variance)

Statistics: Simulated MLE

- Simulated Maximum Likelihood Estimation is the idea of maximizing an approximate likelihood
- Also called: Maximum Simulated Likelihood Estimation
- Basic idea: replace the integral in the likelihood with the Monte Carlo approximation to the integral.
- Example: See slide 35, equation (1)

PSML: Maximum simulated Likelihood Estimation

- SMLE is the idea of maximizing an approximate likelihood
- For our problem: SMLE uses uses the Euler-Maruyama approximation conditional only on $\mathbf{X}(t_{i-1})$, so the first $M - 1$ terms in the approximation are canceled.

Hence, the importance sampler reduces to

$$\frac{1}{J} \sum_{j=1}^{J} p \left(\mathbf{X}(t_i) \mid \mathbf{X}^{(j)}(t_i - \delta) \right) .$$

SMLE constructs an importance sampler q by simulating J paths on each subinterval.

And then you find the parameters that maximize this quantity.

PSML: Maximum simulated Likelihood Estimation

SMLE requires simulation of J trajectories of all states in $\mathbf{X}$ from time $t_i - \delta$ to time t_i using the Euler-Maruyama scheme with the step size δ.

Challenges

- **Computationally intensive:** SMLE is computationally intensive in practice, especially for a multivariate SDE model.
- **High variance:** The SML estimator can have variance.

PSML: Penalized maximum simulated likelihood

- **Auxiliary importance sampling**
  Augment the likelihood with an auxiliary parameter ρ which tunes the importance sampler
- **Penalized maximum simulated likelihood**
  We maximize the log likelihood subject to a constraint that the sum of the variance of the Monte Carlo approximation of the transition density is less than a prespecified level.

PSML: Penalized maximum simulated likelihood (PSML)

We seek to simultaneously estimate $\boldsymbol{\theta}$ and find ρ which minimizes the variation of the Monte Carlo approximation of the transition density.

The PSML estimator $(\hat{\boldsymbol{\theta}}, \hat{\rho})$ is defined by

$$(\hat{\boldsymbol{\theta}}, \hat{\rho}) = \text{arg max} \sum_{i=1}^{n} \log\left(\frac{1}{J}\sum_{j=1}^{J} h_\rho\right) - \lambda \sum_{i=1}^{n} \widehat{\text{cv}}\,(h_\rho)$$

where

- $\lambda \geq 0$ controls the tradeoff between the likelihood and the penalty
- h_ρ is the importance sampling ratio where ρ is an auxiliary parameter used to tune the importance sampling function.
- $\widehat{\text{cv}}(h_\rho)$ is the sample coefficient of variation of h_ρ

PSML: Selecting optimal tuning parameter

- The tuning parameter λ controls how close the importance sampler density q is to the target probability density.
- We choose the optimal tuning parameter λ that minimizes the estimated prediction error,

$$\frac{1}{nL}\sum_{\ell=1}^{L}\sum_{i=1}^{n}\left|\left|\widehat{\mathbf{X}}_{\text{obs}}^{(\ell)}(t_i)-\mathbf{X}_{\text{obs}}(t_i)\right|\right|,$$

where $\widehat{\mathbf{X}}_{\text{obs}}^{(\ell)}(t_i)$ is the ℓth simulated $\mathbf{X}_{\text{obs}}$ at observation time t_i by the Euler-Maruyama scheme with θ set to $\widehat{\theta}(s)$.

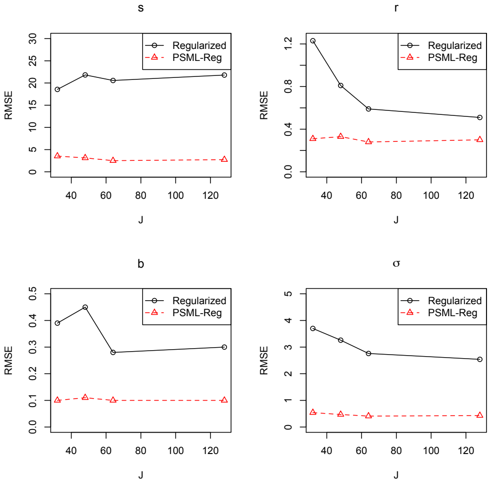
PSML: Summary of Simulation Results

We have shown via simulations and analyses of real data that our **penalized simulated maximum likelihood** (PSML) method has superior performance as compared to previously proposed methods:

- Modified Brownian Bridge (Durham and Gallant, 2002)
- Regularized bridge sampler (Lindstrom 2012)

Stochastic version of the Lorenz model

RMSE using our approach (PSML-Reg) is much lower than RMSE of current best approach (Regularized)



PSML: CWD transmission

Recall the direct transmission SDE model

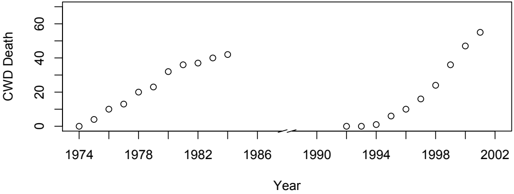
$$\begin{aligned} dS &= [a - S(\beta I + m)]\, dt + B_{11} dW_1 + B_{12} dW_2 + B_{13} dW_3, \\ dI &= [\beta SI - I(\mu + m)]\, dt + B_{21} dW_1 + B_{22} dW_2 + B_{23} dW_3, \\ dC &= \mu I dt + B_{31} dW_1 + B_{32} dW_2 + B_{33} dW_3, \end{aligned}$$

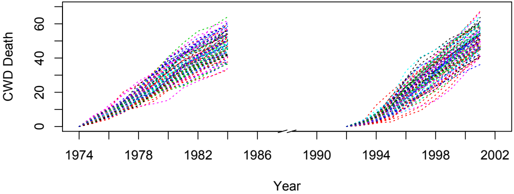
where the unknown parameters are:

- β is the direct transmission coefficient
- μ is the per capita CWD mortality rate
- a is the known number of susceptible animals
- m is the known per capita natural mortality rate
- $\mathbf{W} = (W_1, W_2, W_3)^T$ is a standard Wiener process
- $B = (B_{ij})$ is the positive definite square root of the covariance matrix, $B = \sqrt{\Sigma}$

PSML: Modeling two CWD epidemics

Upper: observed cumulative number of deaths for CWD.
Lower: 100 simulated trajectories of SDE model estimated with parameters via PSML.





Our approach successfully captures the pattern of the CWD epidemic, especially for such a small sample size.

PSML: Modeling two CWD epidemics

Parameter	Method	Estimate	95% CI
β	MBB	0.03	(0.027,0.186)
	Our method	0.03	(0.027,0.120)
μ	MBB	0.27	(0.148,0.599)
	Our method	0.21	(0.143,0.388)

95% Confidence intervals computed using Bootstrapping (another statistics topic....)

Part 2: Model selection for dynamical models

A hierarchical model is one approach to constructing models for complex problems.

Hierarchical model consists of

Stage 1: Data model
Stage 2: Process model
Stage 3: Parameter model

Typically, we use the Bayesian paradigm for inference for hierarchical models.

Statistics

1. Expectation
2. Monte Carlo Integration
3. Importance sampling

4. Maximum likelihood estimation
5. Penalized maximum likelihood estimation
6. Simulated maximum likelihood estimation

7. Bayesian statistics
8. Markov chain Monte Carlo
9. Approximate Bayesian Computation

Statistics: Bayes Theorem

$$p(\theta|\mathbf{x}) = \frac{p(\theta,\mathbf{x})}{p(\mathbf{x})} \tag{2}$$

$$\overbrace{p(\theta|\mathbf{x})}^{\textit{posterior}} = \frac{\overbrace{p(\mathbf{x}|\theta)}^{\textit{likelihood}}\ \overbrace{p(\theta)}^{\textit{prior}}}{\int p(\mathbf{x}|\theta)p(\theta)d\theta} \tag{3}$$

$$p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta) \tag{4}$$

Statistics: Markov chain Monte Carlo (MCMC)

- The most common use of MCMC is to obtain draws from the posterior distribution, $p(\theta|\mathbf{x})$, to enable Bayesian inference.
- The basic idea is that we can't obtain $p(\theta|\mathbf{x})$ in closed form or we can't simulate easily from $p(\theta|\mathbf{x})$.
- MCMC methods can be used to generate a draw from a distribution that approximates some target distribution $p(\cdot)$, but they are more properly viewed as methods for generating a sample from which $E(h(\mathbf{X}))$ can reliably be estimated.
- MCMC is a type of Monte Carlo integration (see page 32), but now you simulate from a Markov chain
- Common MCMC algorithms: Metropolis-Hastings algorithm and Gibbs sampling

Statistics: ABC

Approximate Bayesian Computation (ABC):

- Method to estimate the model parameters when the likelihood is difficult to compute
- Basic idea:
    1. Simulate data from the model given a set of parameters
    2. Compute a distance function between simulated data and the observed data
    3. If the simulated data are similar to the observed data, accept this set of parameters

ABC-SMC (sequential Monte Carlo):

Improve ABC by simulating data through a sequence of intermediate distributions

Approximate Bayesian Computation (ABC)

For more on
Approximate Bayesian Computation (ABC)

Attend David Notts minitutorial 2- 4 pm today
Grand Ballroom, G

Model selection: model set-up

For our chronic wasting disease example:

Stage 1: Data Model

At time t let $\tilde{C}(t)$ = observed cumulative number of deaths from CWD where

$$\tilde{C}(t) \sim \text{Binomial}\left(N(t); \frac{C(t)}{N(t)}\right)$$

where

- $N(t) = S(t) + I(t) + C(t)$ is the total # of animals at time t
- Only $\tilde{C}(t)$ and $N(t)$ observed at discrete time $t = t_0, t_1, \ldots, t_n$,

Model selection: model set-up

Stage 2: Process Model

$$d \begin{pmatrix} S \\ I \\ C \end{pmatrix} = \begin{pmatrix} a - S(\beta I + m) \\ \beta SI - I(\mu + m) \\ \mu I \end{pmatrix} dt$$

Stage 3: Parameter Model

Prior distributions for all model parameters

Inference: We can't write the likelihood in closed form so we certainly don't have the posterior distribution in closed form

Model selection: multiple models

I have multiple reasonable differential equation models. Which one should I use when analyzing my data?

Case 1: Choose between 3 dynamical model types

1. Ordinary differential equation (ODE) model
2. Stochastic differential equation (SDE) model
3. Continuous time Markov chain (CTMC) model

Case 2: Choose between 3 forms of the same type of ODE

1. Direct disease transmission (basic SIR)
2. Indirect disease transmission (environmental transmission)
3. Both direct and indirect disease transmission

Model Selection: ODE model for indirect transmission of CWD

An ODE model for the indirect transmission of CWD (Miller et al. 2006)

$$d\begin{pmatrix} S \\ I \\ E \\ C \end{pmatrix} = \begin{pmatrix} a - S(\gamma E + m) \\ \gamma SE - I(\mu + m) \\ \epsilon I - \tau E \\ \mu I \end{pmatrix} dt,$$

where

- γ is the indirect transmission coefficient
- ϵ is the per capita rate of excretion of infectious material by infected animals
- τ is the mass-specific rate of loss of infectious material from the environment

The unknown quantities to be estimated are $(\gamma, \mu, \epsilon, \tau, S(t_0), I(t_0), E(t_0))$.

Model Selection: Model for CWD

Models considered for CWD problem

Model	Data Model	Process Model CWD transmission	 Dynamical model
$\mathcal{M}_1$	Binom	Direct	ODE
$\mathcal{M}_2$	Binom	Indirect	ODE
$\mathcal{M}_3$	Binom	Direct	SDE
$\mathcal{M}_4$	Binom	Indirect	SDE
$\mathcal{M}_5$	Binom	Direct	CTMC
$\mathcal{M}_6$	Pois	Direct	ODE
$\mathcal{M}_7$	Pois	Indirect	ODE
$\mathcal{M}_8$	Pois	Direct	SDE
$\mathcal{M}_9$	Pois	Indirect	SDE
$\mathcal{M}_{10}$	Pois	Direct	CTMC

No CTMC model for indirect CWD transmission was considered as excreta left in the environment is not a discrete variable.

Model selection: Approaches for model selection

Model selection:

1. Compare models via their posterior model probabilities.
   - For model $\mathcal{M}_k$ the posterior model probability is given by $P(\mathcal{M}_k|D)$.
   - Related to Bayesian model averaging (BMA).
   - See Hoeting, Madigan, Raftery, Volinsky (1999)
   - More info: Merlise Clyde's talk on Tuesday
2. Compare models using Bayes factors (Kass & Raftery 1995)

Model Selection: Results for CWD

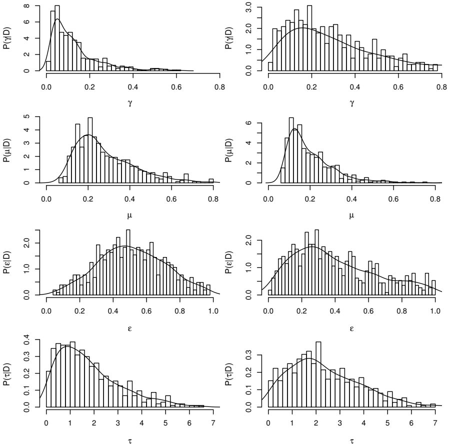
Goal: compare models for Chronic wasting disease.

Posterior model probabilities for each model $P(\mathcal{M}|D)$

Data Model	Process Model	Informative prior set $P(\mathcal{M}|D)$	Bayes factor
Binom	Indirect SDE	0.21	1.00
Binom	Direct SDE	0.18	1.15
Binom	Direct ODE	0.13	1.55
Binom	Direct CTMC	0.11	1.87
Binom	Indirect ODE	0.09	2.43
Pois	Indirect SDE	0.09	2.27
Pois	Direct ODE	0.06	3.48
Pois	Direct SDE	0.05	3.87
Pois	Indirect ODE	0.04	4.63
Pois	Direct CTMC	0.03	6.17

Model selection: Parameter estimates

The marginal posterior distribution for the parameters of the indirect transmission SDE model based on the CWD epidemic data.



Model selection: Parameter estimates

The marginal posterior modes and 95% highest posterior density (HPD) intervals of the parameters of the indirect transmission SDE process model with the Binomial data model based on the CWD epidemic data.

	Informative prior set	
Parameter	Mode	95% HPD
γ = Indirect transmission rate (mass^{-1}yr^{-1})	0.05	$(0.01, 0.36)$
μ = CWD mortality rate (yr^{-1})	0.20	$(0.10, 0.59)$
ϵ = Per capita rate of excretion of infectious agent (yr^{-1})	0.47	$(0.15, 0.91)$
τ = Rate of loss of infectious agent (yr^{-1})	0.88	$(0.01, 4.52)$
$S(0)$ of the first epidemic	18	(10,26)
$I(0)$	10	(5,18)
$E(0)$	1.73	(0.97,5.84)
$S(0)$ of the second epidemic	48	(24,50)
$I(0)$	2	(0,5)
$E(0)$	3.47	(0.24,4.85)

Conclusions

We have covered two approaches for statistical parameter estimation and inference for dynamical models.

Many other approaches have been proposed.

Hopefully this minitutorial will tempt you to learn more statistics and/or dynamical models as there is much work to be done! We need

- More computationally feasible methods for parameter inference and estimating uncertainty
- More understanding about which methods to use when
- Advances on many other open problems

Key References

Textbook:
Computational Statistics
G. Givens and J. A. Hoeting (2013), 2nd edition, Wiley.

Topics include:

- Optimization and solving nonlinear equations
  Optimization methods for discrete and continuous-valued functions, EM Algorithm
- Integration and Simulation
  Numerical integration, simulation, Monte Carlo integration, MCMC (basic and advanced)
- Bootstrapping
- Density Estimation and Smoothing

Book webpage including R code:
www.stat.colostate.edu/computationalstatistics

Key References, continued

Part 1: L. Sun, C. Lee, and J. A. Hoeting (2015)
A penalized simulated maximum likelihood approach in parameter estimation for stochastic differential equations
Computational Statistics and Data Analysis, 84: 54–67

Part 2: L. Sun, C. Lee, and J. A. Hoeting (2015)
Parameter inference and model selection in deterministic and stochastic dynamical models via approximate Bayesian computation: modeling a wildlife epidemic
Environmetrics, 26: 451–462.

Acknowledgments

- Special thank you to my main collaborators on this work:
    - Geof Givens, Givens Statistical Solutions
    - Libo Sun, Johnson and Johnson
    - Chihoon Lee, Stevens Institute of Technology

- N. Thompson Hobbs, Colorado State University
  for introducing us to this problem
- Mike Miller and the Colorado Division of Parks and Wildlife
  for sharing the CWD data

Acknowledgments

This material is based upon work supported by

- Sun & Hoeting: National Science Foundation (EF-0914489)
- Hoeting: National Science Foundation (NSF AGS - 1419558)
- Lee: Army Research Office (W911NF-12-1-0494) & National Security Agency (H98230-12-1-0250)
- We used the CSU ISTeC Cray HPS System, which is supported by NSF Grant CSN-0923386
- Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the funding agencies.