Introduction to Bayesian inference in R

Timothée Bonnet

October 11, 2019

BDSI / RSB

A bird abundance through time
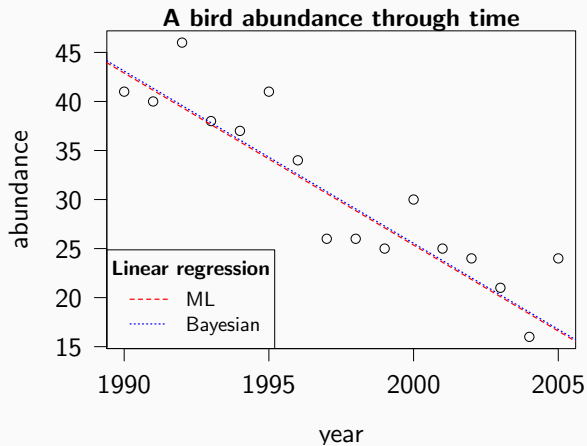
Statistical models exist independently from inference method

- Models are not Bayesian, nor frequentists (ML)
- In principle all can be fitted frequentist or Bayesian
- In general identical / very similar results

So why bother?

Bayesian methods

- More (easily) flexible

So why bother?

Bayesian methods

- More (easily) flexible
- More intuitive interpretation

So why bother?

Bayesian methods

- More (easily) flexible
- More intuitive interpretation
- Uncertainty propagation

So why bother?

Bayesian methods

- More (easily) flexible
- More intuitive interpretation
- Uncertainty propagation
- Easier to include missing values

So why bother?

Bayesian methods

- More (easily) flexible
- More intuitive interpretation
- Uncertainty propagation
- Easier to include missing values
- Incorporate previous knowledge

So why bother?

Bayesian methods

- More (easily) flexible
- More intuitive interpretation
- Uncertainty propagation
- Easier to include missing values
- Incorporate previous knowledge
- The only feasible way for complex models

Today

1. Bayesian philosophy 101
2. Fit your first Bayesian model
3. MCMC in a nutshell
4. The magic of the posterior
5. Using priors

Today

1. Bayesian philosophy 101
2. Fit your first Bayesian model
3. MCMC in a nutshell
4. The magic of the posterior
5. Using priors

Convince you to come back for more

Is it raining outside?

Bayesian philosophy 101

Is it raining outside?
How many heads did I get?

Is it raining outside?
How many heads did I get?

When you think Bayesian

- You need a model

- Prior information may help

- Not one sure answer, only probabilities

Bayesian vs. Frequentist philosophies

Frequentist philosophy
Probability = long run frequencies in fictuous populations

Bayesian philosophy
Probability = plausibility of a proposition

Bayesian vs. Frequentist philosophies

Frequentist philosophy
Probability = long run frequencies in fictuous populations
Uncertainty (p-value, CI, SE) = frequencies of data given parameter values

Bayesian philosophy
Probability = plausibility of a proposition
Uncertainty = probability of parameter values given data

Bayesian vs. Frequentist philosophies

Frequentist philosophy
Probability = long run frequencies in fictuous populations
Uncertainty (p-value, CI, SE) = frequencies of data given parameter values
Unknown parameters have **true, fixed,** values, generating data

Bayesian philosophy
Probability = plausibility of a proposition
Uncertainty = probability of parameter values given data
Unknown parameters have **probability distributions**, data are fixed

Bayesian vs. Frequentist philosophies

Frequentist philosophy
Probability = long run frequencies in fictuous populations
Uncertainty (p-value, CI, SE) = frequencies of data given parameter values
Unknown parameters have **true, fixed,** values, generating data

Bayesian philosophy
Probability = plausibility of a proposition
Uncertainty = probability of parameter values given data
Unknown parameters have **probability distributions**, data are fixed

Enough blabla!

Fit a Bayesian model in MCMCglmm

Fit Bayesian in R

Load metabo data: metabolic rate and mass measured in 3 populations

```
metabo <- read.csv("metabo.csv")
```

Metabolic rate differ among populations?

Fit Bayesian in R

Load metabo data: metabolic rate and mass measured in 3 populations

```
metabo <- read.csv("metabo.csv")
```

Metabolic rate differ among populations?

```
install.packages("MCMCglmm")
```

```
library(MCMCglmm)
m_pop <- MCMCglmm(metabolism ~ 1 + pop, data = metabo)
summary(m_pop)
```

Closer look, and MCMC

The posterior distribution

```
plot(m_pop)
```

The posterior distribution

```
plot(m_pop)
```

```
plot(m_pop$Sol[,"poppop2"])
```
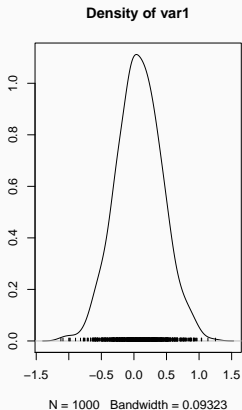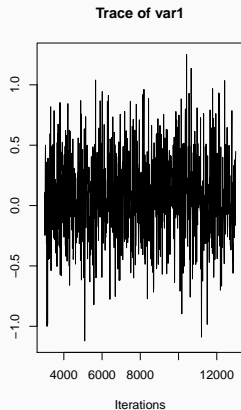
The posterior distribution

Posterior = Distribution of probability(Parameter | Data)

The posterior distribution

Posterior = Distribution of probability(Parameter | Data)

Approximated by Markov Chain Monte Carlo (MCMC) with Metropolis
Hastings:
https://chi-feng.github.io/mcmc-demo/app.html?algorithm=
RandomWalkMH&target=standard

How many MCMC samples to approximate the posterior?

MCMC parameters:

- Number of iterations (nitt)
- Discard N first iterations (burnin)
- Save one iteration every N iterations (thin)

```
m_pop_short <- MCMCglmm(metabolism ~ 1 + pop, data = metabo,
                        nitt = 100, burnin = 0, thin = 10)
plot(m_pop_short)

m_pop_long <- MCMCglmm(metabolism ~ 1 + pop, data = metabo,
                        nitt = 100000, burnin = 0, thin = 10)
plot(m_pop_long)
```

That is fancy, but this model is not smart

```
MCMCglmm(metabolism ~ 1 + pop, data = metabo)
```

That is fancy, but this model is not smart

```r
MCMCglmm(metabolism ~ 1 + pop, data = metabo)
```

```r
m_mpop <- MCMCglmm(metabolism ~ 1 + mass + pop, data = metabo)
summary(m_mpop)
```

Posterior magic

Ask any question, get probabilities

Default pMCMC = parameter estimates different from 0?

```
2*mean(m_mpop$Sol[,"poppop2"]<0)

[1] 0.314
```

Ask any question, get probabilities

Default pMCMC = parameter estimates different from 0?

```
2*mean(m_mpop$Sol[,"poppop2"]<0)

[1] 0.314
```

similar to frequentist p-value:

```
summary(lm(metabolism ~ 1 + mass + pop, data = metabo))$coef

             Estimate Std. Error   t value      Pr(>|t|)
(Intercept) 0.2431512 0.15235074  1.595996 1.225743e-01
mass        0.6521987 0.04121795 15.823173 7.334158e-15
poppop2     0.1053947 0.10333643  1.019918 3.171646e-01
poppop3     0.2175522 0.10940948  1.988422 5.738763e-02
```

In frequentist you are stuck with this, because only point estimate. **In Bayesian, you can change the question, because you know probability of all possible values**

Ask any question, get probabilities

Is the effect pop2 different from 0.1? from -0.2?

```
2*mean(m_mpop$Sol[,"poppop2"]<0.1)

[1] 0.9

2*mean(m_mpop$Sol[,"poppop2"]< - 0.2)

[1] 0.002
```

Ask any question, get probabilities

Is the effect pop2 different from 0.1? from -0.2?

```
2*mean(m_mpop$Sol[,"poppop2"]<0.1)

[1] 0.9

2*mean(m_mpop$Sol[,"poppop2"]< - 0.2)

[1] 0.002
```
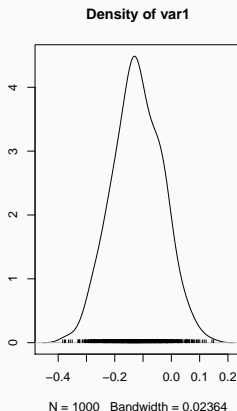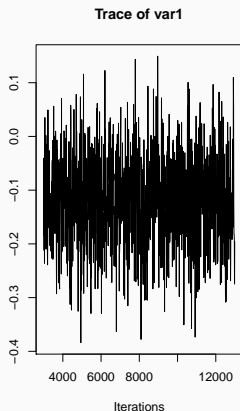
Probability slope of mass different from 0.75? from 1?

Ask any question, get probabilities

Are the effects of pop2 and pop3 different?

```
popdiff <- m_mpop$Sol[,"poppop2"] - m_mpop$Sol[,"poppop3"]
plot(popdiff)
```

Ask any question, get probabilities

Is it likely that the effect of pop3 is more than twice that of pop2? More than half that of pop2?

Ask any question, get probabilities

Is it likely that the effect of pop3 is more than twice that of pop2? More than half that of pop2?

Visualize the posterior of the exponential of the effect of mass plus 1.

```
plot(exp(m_mpop$Sol[,"mass"] + 1))
```

Using the prior

Bayes theorem

Posterior probability $= \dfrac{\text{Likelihood} \times \text{Prior probability}}{\text{Probability of data}}$

Posterior probability is proportional to Likelihood $\times$ Prior probability

Probability of data is unknown, but MCMC can sample in proportion of posterior probability, and probabilities sum to 1, hence we know

Bayes theorem

Posterior probability $= \dfrac{\text{Likelihood} \times \text{Prior probability}}{\text{Probability of data}}$

Posterior probability is proportional to Likelihood $\times$ Prior probability

Probability of data is unknown, but MCMC can sample in proportion of posterior probability, and probabilities sum to 1, hence we know

$$P(\theta|y) = \frac{P(y|\theta) \times P(\theta)}{P(y)}$$
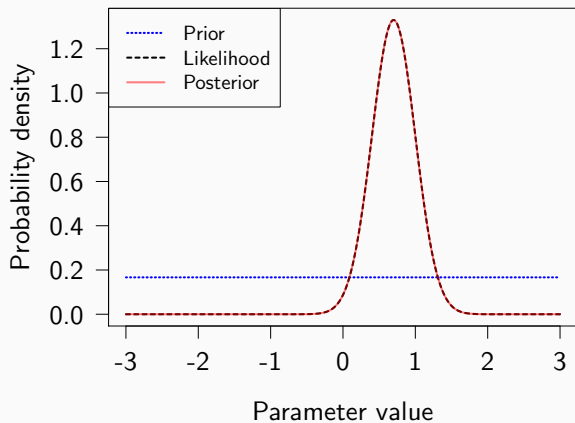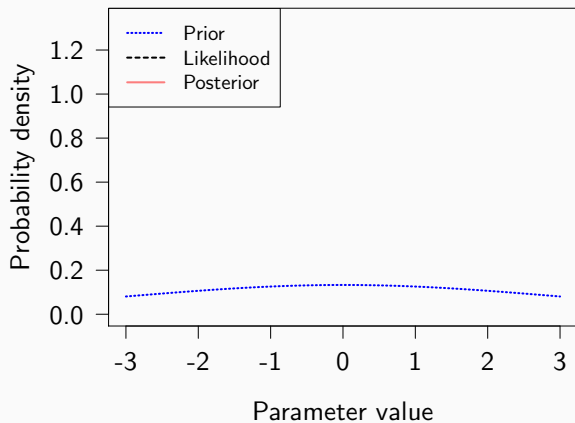$$\propto P(y|\theta) \times P(\theta)$$
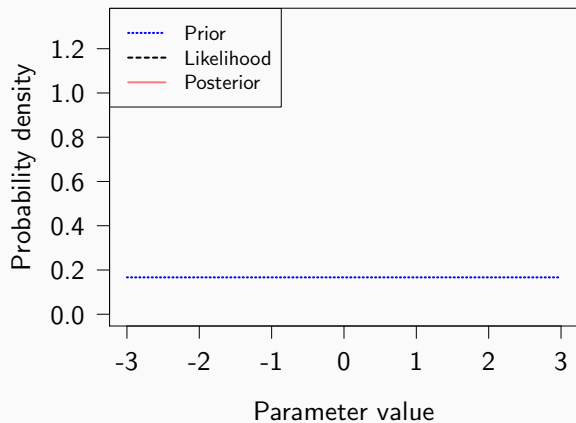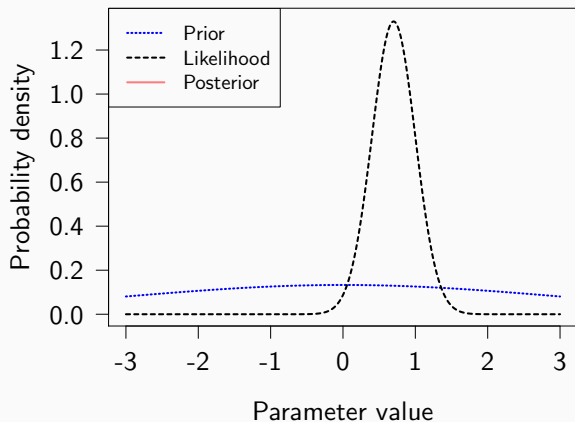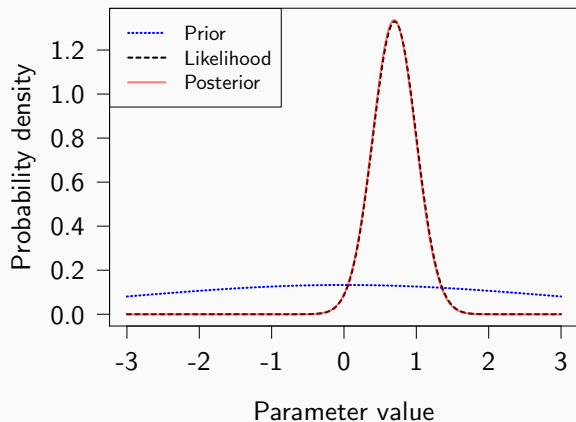
Flat prior
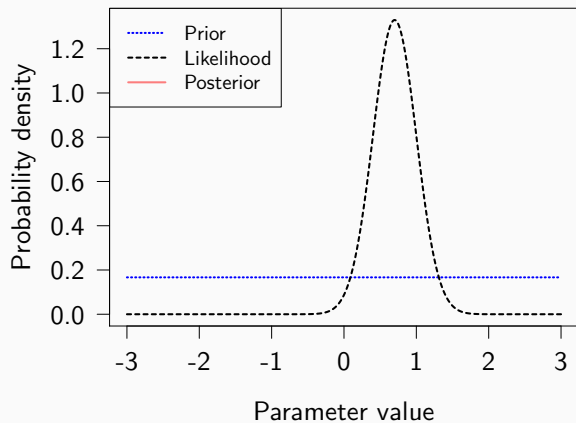
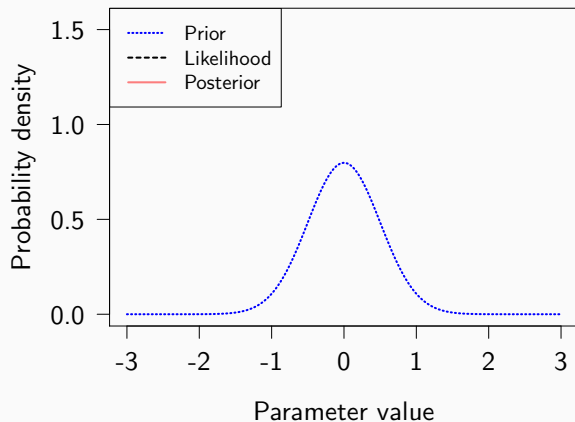# Flat prior

Flat prior

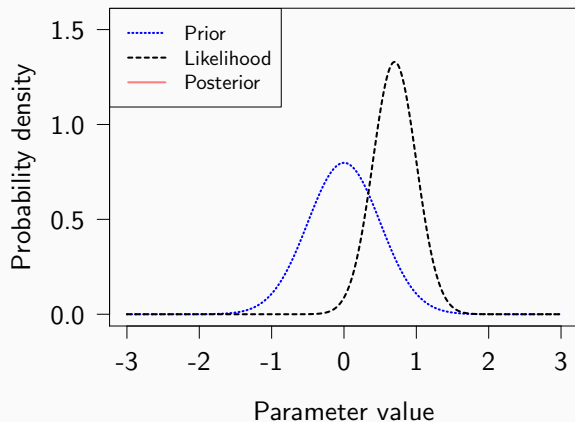## Weakly informative prior

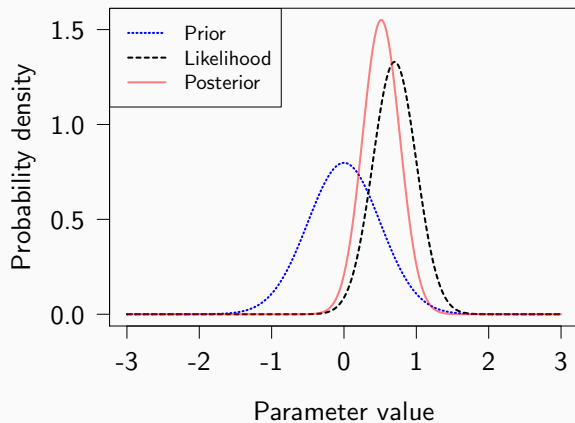# Weakly informative prior

Weakly informative prior

Strong prior

Strong prior
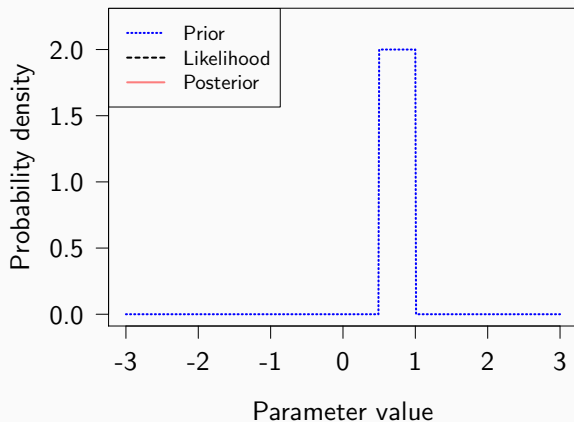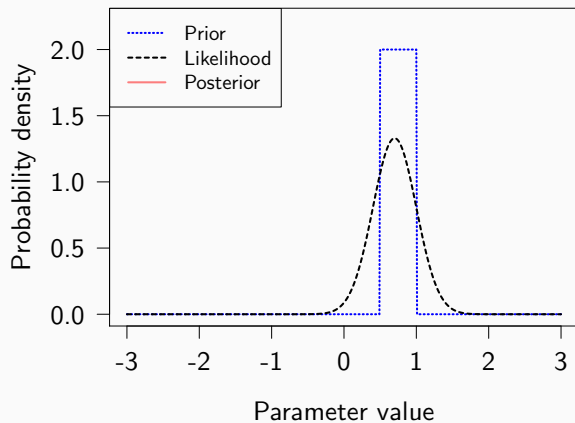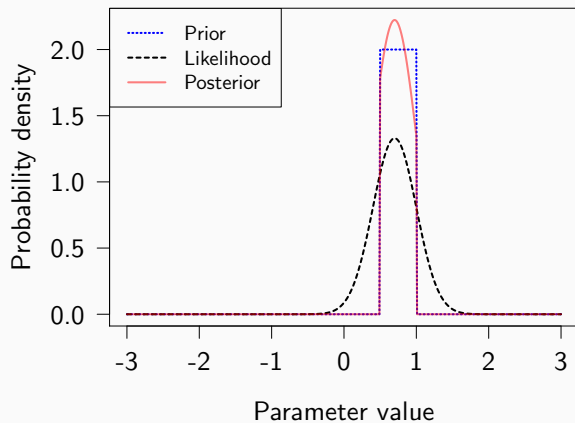
Strong prior

Strong prior

Strong prior

Strong prior

In general you don't want prior to influence results

BUT, sometimes it will anyway. Need to be careful and check alternatives

Sometimes informative priors are useful:

- Correct for a well known nuisance parameter
- Missing data
- Avoid biologically impossible values

Change the prior in MCMCglmm

Fixed effect priors follow normal distribution. Default mean $= 0$, variances $= 10^{10}$, covariances $= 0$; almost flat:

```
priorsamp <- rnorm(n = 10000000, mean = 0, sd = sqrt(10^10))
plot(density(priorsamp), xlim=c(-100, 100))
```


density.default(x = priorsamp)
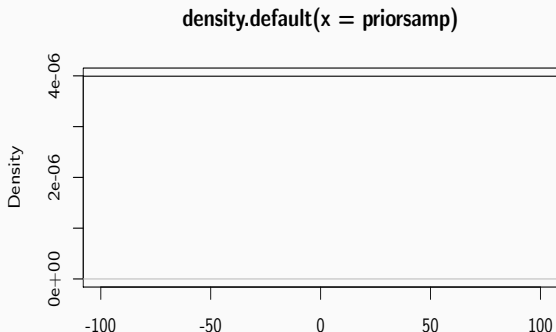
N = 10000000 Bandwidth = 3582

Change the prior in MCMCglmm

We know from other studies the effect of mass is exactly 0.75.

```
varprior <- diag(4)*10000
varprior[2,2] <- 0.000001
prior1 <- list(B= list(mu=c(0,0.75,0,0), V=varprior))
mp <- MCMCglmm(metabolism ~ 1 + mass + pop, data = metabo,
               prior = prior1)
summary(mp)
```

Change the prior in MCMCglmm

We know from other studies the effect of mass is exactly 0.75. Another study has estimated the effect of pop2 to 0.15, with standard error of 0.01

```
varprior <- diag(4)*10000
varprior[2,2] <- 0.000001
varprior[3,3] <- 0.01^2
prior2 <- list(B= list(mu=c(0,0.75,0.15,0), V=varprior))
mp2 <- MCMCglmm(metabolism ~ 1 + mass + pop, data = metabo,
                prior = prior2)
summary(mp2)
```

Some models much easier in Bayesian

Missing data in a predictor with a trend

Some models much easier in Bayesian

Missing data in a predictor with a trend

Missing data in temperarture you want to use to predict Survival at the end of
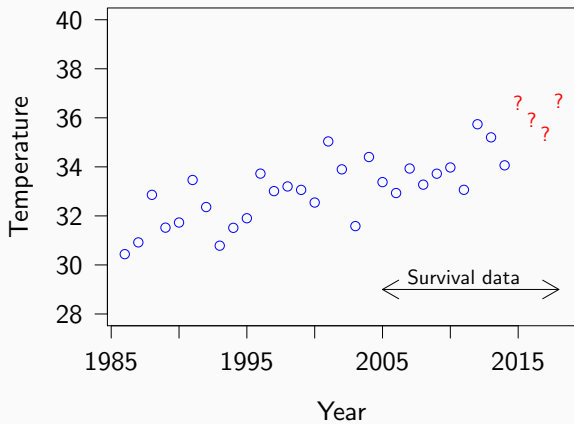
time serie

Some models much easier in Bayesian

Missing data in a predictor with a trend

Combining information from different sources in a single model

Some models much easier in Bayesian

Missing data in a predictor with a trend
Combining information from different sources in a single model

1. Fit individual growth curves to sparse mass data

Some models much easier in Bayesian

Missing data in a predictor with a trend

Combining information from different sources in a single model

1. Fit individual growth curves to sparse mass data
2. Estimate litter birth dates from estimate of mass< birth mass, sibblings', time between litter

Some models much easier in Bayesian

Missing data in a predictor with a trend

Combining information from different sources in a single model

1. Fit individual growth curves to sparse mass data
2. Estimate litter birth dates from estimate of mass$<$ birth mass, sibblings', time between litter
3. Model snow cover

Some models much easier in Bayesian

Missing data in a predictor with a trend
Combining information from different sources in a single model

1. Fit individual growth curves to sparse mass data
2. Estimate litter birth dates from estimate of mass$<$ birth mass, sibblings', time between litter
3. Model snow cover
4. Model survival from mark-recapture data

Some models much easier in Bayesian

Missing data in a predictor with a trend

Combining information from different sources in a single model

1. Fit individual growth curves to sparse mass data
2. Estimate litter birth dates from estimate of mass$<$ birth mass, sibblings', time between litter
3. Model snow cover
4. Model survival from mark-recapture data
5. Estimate interaction snow cover by birth date on survival

Some models much easier in Bayesian

Missing data in a predictor with a trend
Combining information from different sources in a single model

1. Fit individual growth curves to sparse mass data
2. Estimate litter birth dates from estimate of mass$<$ birth mass, sibblings', time between litter
3. Model snow cover
4. Model survival from mark-recapture data
5. Estimate interaction snow cover by birth date on survival
6. (plug in mean survival to a model of the number of individuals in population)

Some models much easier in Bayesian

Missing data in a predictor with a trend
Combining information from different sources in a single model

1. Fit individual growth curves to sparse mass data
2. Estimate litter birth dates from estimate of mass$<$ birth mass, siblings', time between litter
3. Model snow cover
4. Model survival from mark-recapture data
5. Estimate interaction snow cover by birth date on survival
6. (plug in mean survival to a model of the number of individuals in population)
7. (estimate how snow cover and population size impact vegetation growth)

Some models much easier in Bayesian

Missing data in a predictor with a trend
Combining information from different sources in a single model

1. Fit individual growth curves to sparse mass data
2. Estimate litter birth dates from estimate of mass< birth mass, siblings', time between litter
3. Model snow cover
4. Model survival from mark-recapture data
5. Estimate interaction snow cover by birth date on survival
6. (plug in mean survival to a model of the number of individuals in population)
7. (estimate how snow cover and population size impact vegetation growth)
8. . . .

Want more?

Two days on writing your own models in JAGS/STAN

- GLMMs
- Missing data
- Integrating multi-part models
- Priors, diagnostics, cool tools. . .