Module 2: Introduction to Bayesian Statistics

Andrew Parnell, School of Mathematics and Statistics,
University College Dublin

Learning outcomes

- Know the difference between Frequentist and Bayesian statistics
- Understand the terms posterior, likelihood and prior. Be able to suggest suitable probability distributions for these terms
- Be able to interpret the posterior distribution through plots, summaries, and credible intervals

A bigger aim, either:

1. Stop using SIAR (for dietary proportions) and start writing your own JAGS code
2. Stop using SIAR and start using MixSIAR/simmr/SiBER instead

Who was Bayes?

An essay towards solving a problem on the doctrine of chances (1763)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

What is Bayesian statistics?

- Bayesian statistics is based on an interpretation of Bayes' theorem
- All quantities are divided up into *data* (i.e. things which have been observed) and *parameters* (i.e. things which haven't been observed)
- We use Bayes' interpretation of the theorem to get the *posterior probability distribution*, the probability of the unobserved given the observed
- Used now in almost all areas of statistical application (finance, medicine, environmetrics, gambling, etc, etc)

Why is this relevant to SIMMs?

- Easy to specify Bayesian models hierarchically in layers so that the data depend on some parameters, which then depend on further parameters, and so on. This allows us to create richer statistical models which will better match reality
- Almost all the modern Stable Isotope Mixing Models (SIMMs) use Bayesian statistics
- MixSIR, SIAR, MixSIAR, simmr, IsotopeR, ...

What is Bayes' theorem?

Bayes' theorem can be written in words as:

$$\text{posterior is proportional to likelihood times prior}$$

. . . or . . .

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

Each of the three terms *posterior*, *likelihood*, and *prior* are *probability distributions* (pdfs).
In a Bayesian model, every item of interest is either data (which we will write as x) or parameters (which we will write as θ). Often the parameters are divided up into those of interest, and other *nuisance parameters*

Bayes' theorem in more detail

Bayes' equation is usually written mathematically as:

$$p(\theta|x) \propto p(x|\theta) \times p(\theta)$$

or, more fully:

$$p(\theta|x) = \frac{p(x|\theta) \times p(\theta)}{p(x)}$$

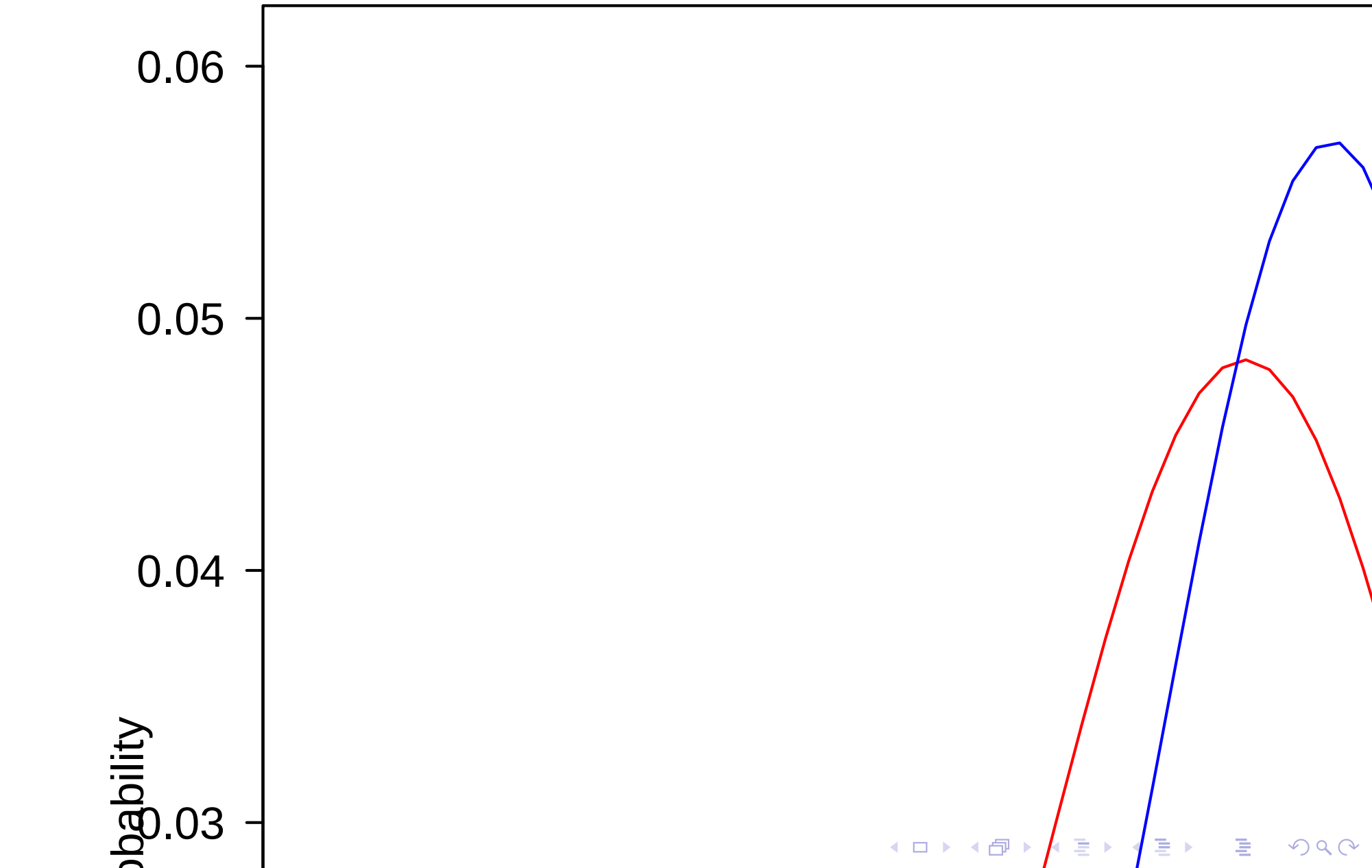
- The *posterior* is the probability of the parameters given the data
- The *likelihood* is the probability of observing the data given the parameters (unknowns)
- The *prior* represents external knowledge about the parameters

A simple example

- An ecologist listens for the calls of the southern brown tree frog (*Litoria ewingi*)
- She wants to know the mean length of the calls
- We will assume that the standard deviation of the calls is known to be 0.8 seconds
- She hears a call of length 3.1 seconds
- A study conducted the previous year estimated the mean to be 2.3 seconds with standard error 0.5 seconds

If we assume that the calls she hears are normally distributed then x follows a normal distribution with mean θ and standard deviation 0.8s, written $x|\theta \sim N(\theta, 0.8^2)$. The prior distribution is $\theta \sim N(2.3, 0.5^2)$.

Simple example (continued)



Simple example (continued)

Code used to produce previous plot:

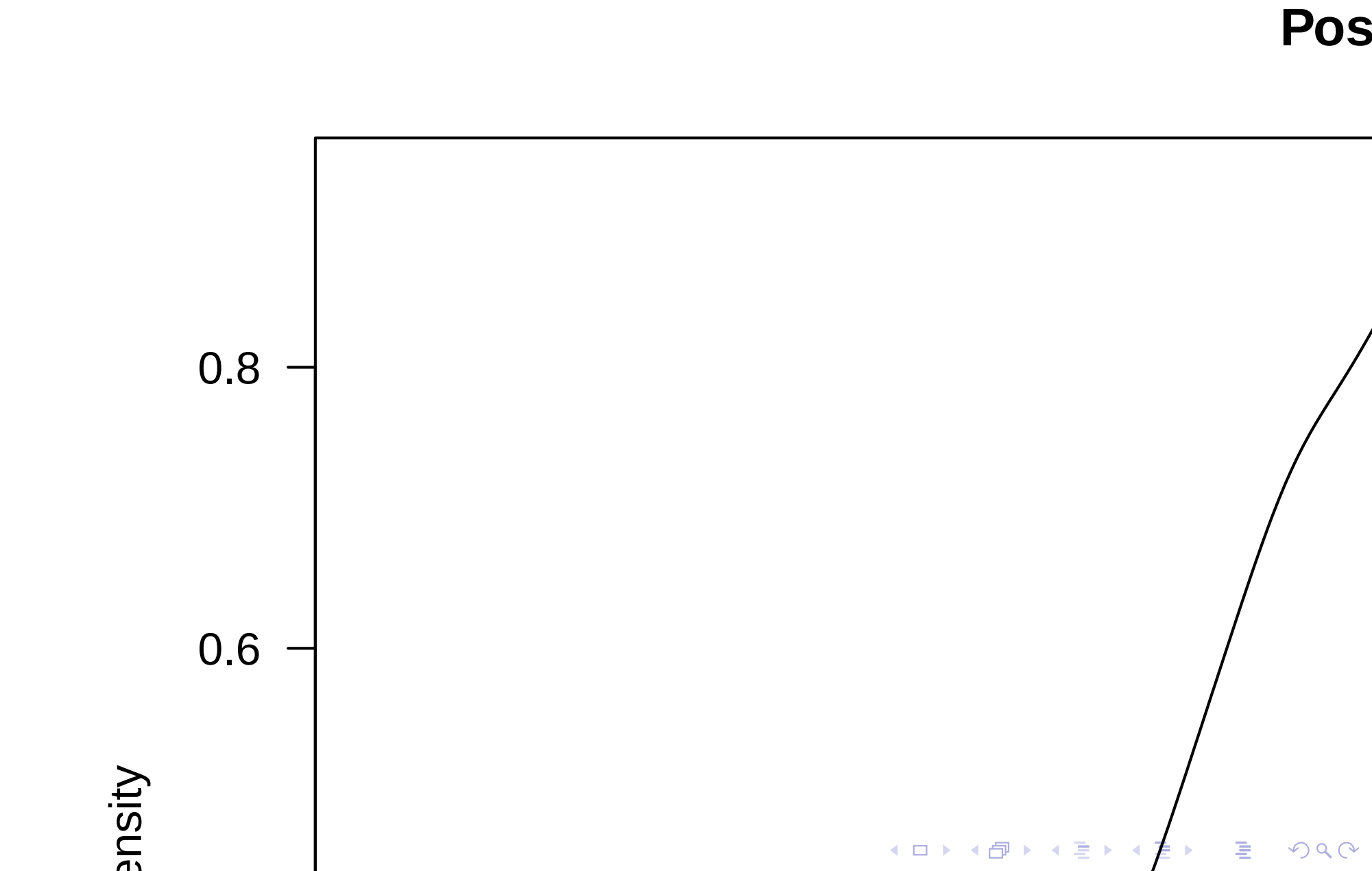
```
# Create grid for theta
theta = seq(0,6,length=100)
# Evalutate prior, likelihood and posterior
prior = dnorm(theta,mean=2.3,sd=0.5)
likelihood = dnorm(3.1,mean=theta,sd=0.8)
posterior = prior*likelihood
# Produce plot
plot(theta,likelihood/sum(likelihood),type='l',
     ylab='Probability',ylim=c(0,0.06))
lines(theta,prior/sum(prior),col='red')
lines(theta,posterior/sum(posterior),col='blue')
legend('topright',legend=c('Likelihood','Prior',
                           'Posterior'),
       col=c('black','red','blue'),lty=1)
```

Simple example in JAGS

In later modules we will start using JAGS to fit models like this. The code is much simpler than the previous R version:

```
library(rjags)
modelstring ='
  model {
    # Likelihood
    x ~ dnorm(theta,1/pow(0.8,2))
    # Prior
    theta ~ dnorm(2.3,1/pow(0.5,2))
  }
'
# Set up data
data=list(x=3.1)
# Run jags
model=jags.model(textConnection(modelstring), data=data)
output=coda.samples(model=model,variable.names=c("theta"),
# Plot output
plot(density(output[[1]]))
```

Plot from JAGS



What are the assumptions involved in this example?

- We've assumed that the normal distribution is appropriate for the likelihood and the prior
- We've only observed one data point. What if we observed many?
- We've assumed that the likelihood standard deviation is fixed at 0.8

A more complicated JAGS version

```
modelstring ='
  model {
    # Likelihood
    for(i in 1:n) { x[i] ~ dnorm(theta,1/pow(sd,2)) }
    # Prior
    theta ~ dnorm(2.3,1/pow(0.5,2))
    sd ~ dunif(0,100)
  }
'
# Set up data
data=list(x=c(3.1,2.7,4.2,3.6),n=4)
# Run jags
model=jags.model(textConnection(modelstring), data=data)
output=coda.samples(model=model,variable.names=
                      c("theta","sd"), n.iter=1000)
```

Now have four data points, two parameters and a prior for each

What if the observations aren't normal?

No problem! We just choose distributions which are appropriate for the type of data

```
modelstring ='
  model {
    # Likelihood
    for(i in 1:n) {
      x[i] ~ dgamma(alpha,beta)
    }
    # Prior
    alpha ~ dunif(0,100)
    beta ~ dunif(0,100)
    # Mean
    mean <- alpha/beta
  }
'
data=list(x=c(3.1,2.7,4.2,3.6),n=4)
model=jags.model(textConnection(modelstring), data=data)
output=coda.samples(model=model,variable.names=c("mean"),n
```

How do I specify the prior distribution?

There are several choices when it comes to specifying prior distributions:

- *Informative*, when there is information from a previous study, or other good external source, e.g $\theta \sim N(2.3, 0.5^2)$
- *Vague*, when there is only weak information, perhaps as to the likely range of the parameter e.g. $\theta \sim U(0, 100)$
- *Flat*, when there is no information at all about a parameter (very rare). In JAGS, write `theta ~ dflat()`

Choosing the prior and choosing the likelihood are very similar problems

Choosing likelihoods and priors

When creating Bayesian models it's helpful to know a lot of probability distributions. The ones we will use most are:

Distribution	Range	Useful for:
Normal, $N(\mu, \sigma^2)$	$(-\infty, \infty)$	A good default choice
Uniform, $U(a, b)$	(a, b)	Vague priors when we only know the range of the parameter
Binomial, $Bin(k, \theta)$	$[0, k]$	Count or binary data restricted to have an upper value
Poisson, $Po(\lambda)$	$[0, \infty)$	Count data with no upper limit
Gamma, $Ga(\alpha, \beta)$	$(0, \infty)$	Continuous data with a lower bound of zero
Multivariate	$(-\infty, \infty)$	Multivariate

Creating the posterior distribution

- In the very simple example, I was able to calculate the posterior distribution in just a couple of lines of `R` code
- When we have lots of parameters, and complicated prior distributions, we have to resort to *simulation*
- This means that we obtain *samples* from the posterior distribution rather than creating the probability distribution directly
- JAGS uses Markov chain Monte Carlo (MCMC) to create these samples. We will talk about this a bit more in later lectures/discussion

Summarising the posterior distribution

- Because we obtain samples from the posterior distribution, we can create any quantity we like from them
- e.g. we can obtain the mean or standard deviation simply from combining the samples together
- We can create quantiles e.g. 50% for the median
- We can create a Bayesian *credible interval* (CI) by calculating lower and upper quantiles
- When the posterior distribution is messy (e.g. multi-modal) we can use a *highest posterior density* (HPD) region

Example:

From the earlier simple example. First 5 posterior samples

```
output[[1]][1:5]
```

```
## [1] 2.597655 1.819589 2.700412 2.993008 2.507259
```

The mean and standard deviation:

```
c(mean(output[[1]]),sd(output[[1]]))
```

```
## [1] 2.5175496 0.4227122
```

A 95% credible interval

```
quantile(output[[1]],probs=c(0.025,0.975))
```

```
##      2.5%     97.5%
## 1.689061 3.345698
```

Why is this better?

The Bayesian approach has numerous advantages:

- It's easier to build complex models and to analyse the parameters you want directly
- We automatically obtain the best parameter estimates and their uncertainty from the posterior samples
- It allows us to get away from (terrible) null hypothesis testing and p-values

Some further reading

- The Bayesian bible: Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*, Third Edition. CRC Press.
- The MCMC bible: Brooks, S., Gelman, A., Jones, G., & Meng, X. (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press.
- Something simpler: McCarthy, M. A. (2007). *Bayesian Methods for Ecology*. Cambridge University Press.

Summary

- Bayesian statistical models involve a likelihood and a prior. These both need to be carefully chosen. From these we create a posterior distribution
- The likelihood represents the information about the data generating process, the prior represents information about the unknown parameters
- We usually create and analyse samples from the posterior probability distribution of the unknowns (the parameters) given the knowns (the data)
- From the posterior distribution we can create means, medians, standard deviations, credible intervals, etc