Bayesian methods for ecological and environmental modelling

Trainers:

Lindsay Flynn Banin, David Cameron, Pete Henrys & Peter Levy

UK Centre for Ecology & Hydrology

© UKCEH 2024

1

Session 1b

An overview and introduction to Bayesian statistics

Pete Henrys

UK Centre for Ecology & Hydrology

2

What we will cover in this session:

- Some probability basics
- Conditional probability
- Bayes theorem

- Fundamentals of Bayesian inference
- Concept of updating
- Priors

- Why be Bayesian???

3

Statistics

Statistics can be considered as the science of uncertainty

In every modelling study:
Model parameters, inputs and structure are uncertain

↓

How to deal with uncertainties optimally?

4

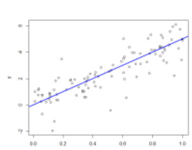
Uncertainty

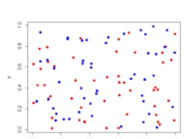
Uncertainties are everywhere:

Models (environmental inputs, parameters, structure)

Data (detectability, measurement error)

We need methods that:
- Quantify uncertainties
- Show how to reduce them
- Efficiently transfer information:
  data → models → model application





5

Probability

In statistics we quantify uncertainty based on probability theory

- A random event describes an act where the outcome is uncertain
- A sample space is the set of all possible outcomes
- Each outcome is assigned a probability between 0 and 1

0 ———————————————— 1

Impossible ———————— Certain

Image © Pixaby CC01

6

Example

In my sock draw, I have 4 yellow socks, 6 blue socks and 2 black socks.

Without looking, I pull a sock out of the draw.

What is the probability that I pull a yellow sock out?

I then pull another, again without looking, what is the probability that is also yellow?

Image adapted from © Pixaby CC01

7

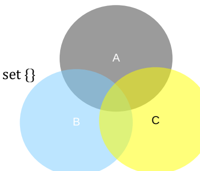
Some basic notation

Write P to denote probability

$P(E)$ to denote the probability of an event E

Sample space is often denoted by Ω , and the sample set $\{\}$

e.g. $\Omega = \{\text{A, B, C}\}$



8

Probability Rules

Rule 1

The range of all possible probabilities for an event is bounded by 0 (impossible) to 1 (definite)

$$0 \leq P(E) \leq 1$$

Image © Pixaby CC01

9

Probability Rules

Rule 2:

The sum of probabilities across all possible events is equal to 1

That is, there is a guarantee that something has to happen by the definition of "all possible events"

Image © Pixaby CC01

10

Probability Rules

Rule 3:

The complement of any event is equal to 1 minus the probability of the event.

$$P(E^c) = 1 - P(E)$$

That is, the probability of an event not happening

Image © Pixaby CC01

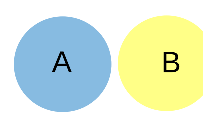
11

Probability Rules

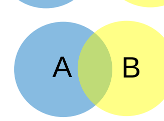
Rule 4:

Addition Rule

What is the probability of either event A or event B happening?



If two events A and B are independent, then the probability of A or B is equal to $P(A) + P(B)$



If the events are not independent, then the probability of A or B is equal to the sum of the two minus the intersection

Image © Pixaby CC01

12

Example

A large survey of adults finds that:

- 58% enjoy going for a long (>3mile) walk at least once per week
- 35% own a dog
- And 22% both own a dog AND enjoy a long walk once a week

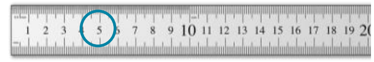
What is the probability that a randomly selected individual from the population is either a dog owner (D) or enjoys a long walk once a week (W)

$$P(D \text{ or } W) = P(D \cup W)$$

$$= P(D) + P(W) - P(D \cap W)$$

13

Probability Rules



Rule 5:

Multiplication Rule

What is the probability of event A AND event B happening?

If two events A and B are independent, then the probability of A and B is equal to $P(A)\ x\ P(B)$

If the events are not independent, then the probability of A and B is equal to the probability of A multiplied by the probability of B conditional on A

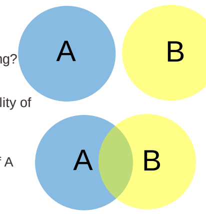


Image © Pixaby CC01

14

Example

With a fair coin, what is the probability of tossing 5 heads in a row?

$$P(H) = 0.5$$

Individual coin flips are independent – the outcome of one does not depend on what has happened previously.

So,

$$P(HHHHH) = P(H) \times P(H) \times P(H) \times P(H) \times P(H)$$

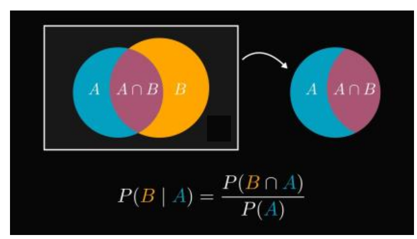
$$= 0.5 \times 0.5 \times 0.5 \times 0.5 \times 0.5$$

= 0.03

Image under free licence from © Adobe Stock

15

Conditional Probability



16

A familiar example

P(Covid) 0.01

P(Test positive) 0.69 **Sensitivity**
P(Test positive | Covid)

P(Test negative) 0.31
P(Test negative | Covid)

0.99 P(Not got Covid)

P(Test positive) 0.007
P(Test positive | Not Covid)

P(Test negative) 0.993 **Specificity**
P(Test negative | Not Covid)

Please note, this an illustrative example and numbers are not to be considered as truth

17

Bayes' Theorem

$$P(A \mid B) = \frac{P(B \mid A)\, P(A)}{P(B)}$$

$P(A \text{ and } B)$

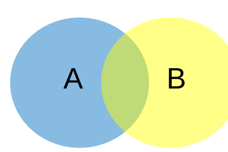


Image © Wikipedia CC 4.0 BY-SA

18

A familiar example

P(Test positive) 0.69 **Sensitivity**
P(Test positive | Covid)

P(Covid)

P(Covid | Test positive) = P(Test positive | Covid) x P(Covid) / P(Test positive)

| Covid)

= 0.69 x 0.01 / 0.01383

= 0.4989

| Not Covid)

0.99
P(Not got Covid)

P(Test negative) 0.993 **Specificity**
P(Test negative | Not Covid)

Please note, this an illustrative example and numbers are not to be considered as truth

19

Probability Distributions

A mathematical function to describe the probabilities across the entire event space

Distinguish between **discrete** and **continuous** random variables.

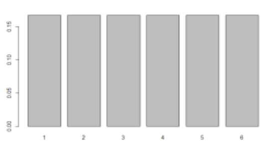
In the discrete case, it is enough to specify a probability mass function assigning a probability to each possible outcome.

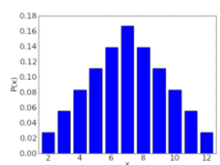
The probability density function describes the probability of any given value. You can compute the probability that the outcome lies in a given interval by integrating the probability density function over that interval.

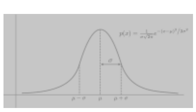
An alternative description is given by the cumulative distribution function

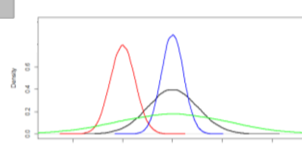
20

Common pdfs









21

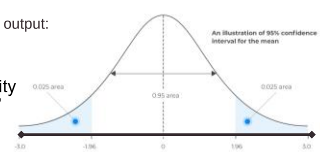
Beyond Bayes' Theorem

Bayesian inference uses probability concepts to describe what is known about parameters in a given "model"

Contrast that with the frequentist view (using confidence intervals and p-vales) which does not tell us about parameters.

Have you ever had the following output:

"There is a 95% probability that β lies in this interval"



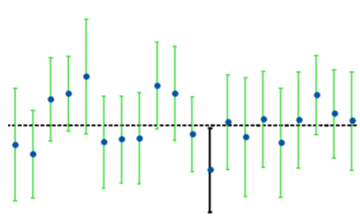
22

Frequentists' view of uncertainty

In a frequentists world, parameters are FIXED.

Uncertainty is entirely due to sampling

So what is a 95% confidence interval again?



23

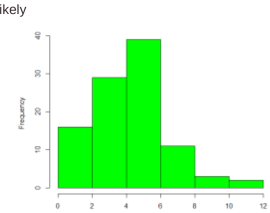
Likelihood

Given the data we have observed, what is the likely value of parameters?

This terminology is important for later, so let's consider an example in detail

Suppose we count the number of different bird species observed in our garden

Question: what is our best guess at the typical number of species???



24

Start with the probability density function (pdf) for the Poisson distribution

$P(X=x)=\frac{\lambda^x e^{-\lambda}}{x!}$ The likelihood is therefore given by $L(\lambda)=\prod_{i=1}^{n}\frac{\lambda^{x_i}e^{-\lambda}}{x_i!}$

And the log-likelihood is given by

$$l(\lambda)=\sum_{i=1}^{n}(x_i\log(\lambda)-\lambda-\log(x_i!)) = \left(\log(\lambda)\sum_{i=1}^{n}x_i\right)-n\lambda-\left(\sum_{i=1}^{n}\log(x_i!)\right)$$

To find the mle we need to maximise this function with respect to λ (i.e. find the most likely value of λ). To find the maximum of a function, we differentiate

$$\frac{\partial l}{\partial \lambda}=\frac{1}{\lambda}\sum_{i=1}^{n}x_i-n=0$$

Rearranging this gives

$$\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}x_i$$

i.e., the sample mean.

25

Bayesian Inference

- For Bayesians parameters are random variables
- Each parameter therefore has a corresponding probability distribution

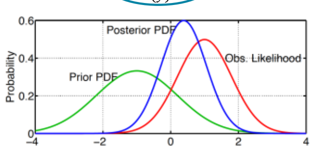
- Rather than estimating FIXED parameters, in a Bayesian setting parameters are random so you predict a distribution of likelihood for the true parameter
- Data y are observations from a random process
- Parameters θ are random quantities of that process
- Joint distribution of all these random quantities – P(y, θ)

26

Bayes rule for probability densities

What we saw earlier

$$P(\theta|y)=\frac{P(y|\theta)\cdot P(\theta)}{P(y)}\propto P(y|\theta)\cdot P(\theta)$$



Note that if prior is uninformative, the posterior returns the likelihood

27

A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes they have seen a mule.

♀ Horse X ♂ Donkey = Mule

Images © Pixaby CC01

28

P(ϑ) = *Prior probability distribution for the parameters ϑ of model f. This represents our uncertainty about the parameter values, before seeing the data.*

P(D|ϑ) = *Probability of getting the data D if the true value is f(ϑ). This probability is a function of the parameter values ϑ, i.e. the so-called 'likelihood'-function'.*

Prior pdf for the parameters

Likelihood of the data

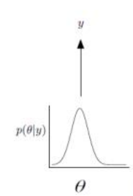
Bayes' Theorem **P(ϑ|D) = P(ϑ) P(D|ϑ) / P(D)**

Posterior pdf for the parameters

Scaling constant (= ∫ P(ϑ) P(D|ϑ) d ϑ)

29

Bayesian statistical inference



$$p(\theta|y)=\frac{p(y|\theta)p(\theta)}{p(y)}$$

θ and y before it is observed are *random variables*. The data in hand are the true outcome of the process governed by θ. We use the observations to understand the probability distribution of θ using Bayes theorem.

All unobserved quantities are treated the same way:
- Model parameters
- Missing data
- Predictions/forecasts
- Latent states
- Data (before they are observed)

30

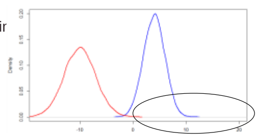
Priors

Isn't this a huge assumption? - we make assumptions all the time in science and statistics

Priors come from all data and information external to the current study.

Can be based on expert opinion or data

Capture all known information without constrictir the posterior



31

Types of prior

Informative priors – expert opinion; previous studies; intervals; expressions of uncertainty

Proper priors (i.e. area under curve integrates to 1)

Vague priors (a.k.a. flat, diffuse; at the beginning of your analysis, plot your priors to check they make sense and genuinely are vague)

Shape of prior becomes less influential and likelihood more influential with increasing sample size

Conjugacy – allows more straight forward solution

Support – what values are logical given the nature of the data?

32

So, why all the fuss?

Because we can sample to approximate the posterior distribution rather than having to solve mathematically.

In complex models, we can evaluate the posterior distribution for a given set of parameters but can often not solve it.

So, follow algorithms (MCMC) where we sample the parameter space with probabilities proportional to the posterior distribution

33

Why be a Bayesian?

- Philosophical reasons – viewing parameters as random variables
- Model complexity – unable to use standard approaches

- Combining data from different sources, indirect information (e.g. primary data, prior knowledge)
- Data collected at different scales, sampling methods
- Modelling temporal/spatial dependence
- Forecasting, projecting uncertainty
- Identifying thresholds

34

Intro to Practical Session

What is the proportion of adults who drink at least one cup of tea per day?

We are going to look at a simple Bayesian evaluation using a prior and likelihood function to determine this proportion.

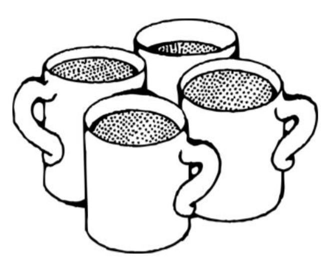
We will also see how to update this as new information becomes available.

But first………

© openclipart (CC01 licence)

35

Time for a break



© openclipart (CC01 licence)

36

Practical Session

See the practical session link for session 1b

37

Any Questions?

38

This concludes

An overview and introduction to Bayesian statistics

Pete Henrys

UK Centre for Ecology & Hydrology

39

Conclusion
What we covered today:

- Some probability basics
- Conditional probability
- Bayes theorem

- Fundamentals of Bayesian inference
- Concept of updating
- Priors

- Why be Bayesian???

40

Your Feedback

What Went Well?

Even Better If!

41

Tomorrow (Tuesday)

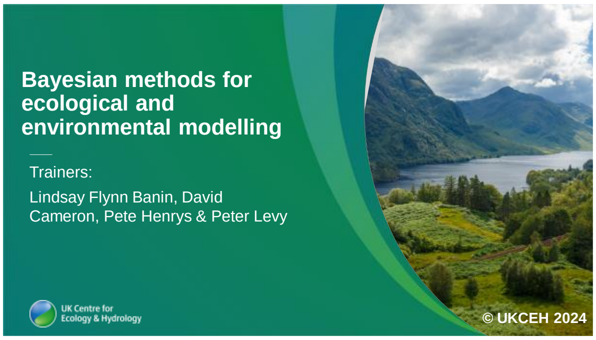
Start time: 9:30 am

We will cover

2a More basics (Pete H.)
2b Monto Carlo Markov Chains (David)

3a&b Linear modelling (Peter L.)

42


Bayesian methods for ecological and environmental modelling
Trainers:
Lindsay Flynn Banin, David Cameron, Pete Henrys & Peter Levy
UK Centre for Ecology & Hydrology
© UKCEH 2024

43