Introduction to WinBUGS for Ecologists

Marc Kéry

Swiss Ornithological Institute,
Sempach/Switzerland

USGS
Patuxent Wildlife Research Center,
Laurel/MD,
2-6 November 2009

Outline Intro

- Role of models in science
- Bayesian and frequentist analyses of statistical models
- Bayesian computation
- WinBUGS
- Advantages of Bayesian analysis
- Some disadvantages of Bayesian analysis
- Structure and special features of this workshop

Role of models

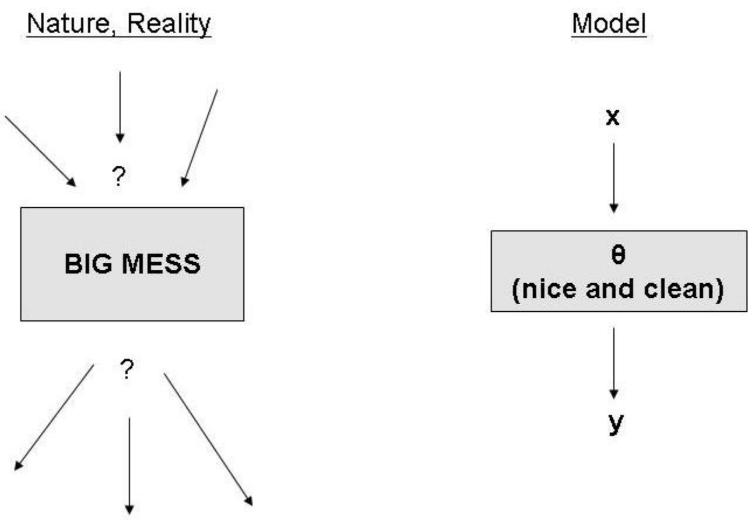
- Science: explain Nature, so you can better understand and predict her
- Nature too high-dimensional, so have to reduce complexity when trying to understand her
- Model: greatly simplified version of Nature, should help understand/predict
- Every model has an objective

Role of models

- Some of my favourite sayings about models:
  - All models are wrong, but some are useful (Box)
  - „There has never been a straight line nor a Normal distribution in history, and yet, using assumptions of linearity and Normality allows, to a good approximation, to understand and predict a huge number of observations." (some famous statistician)
  - „Nothing is gained if you replace a world that you don't understand with a model that you don't understand" (??)

    -> so keep it simple ... or become a better modeller
  - The usefulness of a model depends entirely on its goal: can be good for one and not good for another

Role of models



- Mathematical model: Replace reality of infinite dimension with much smaller set of system descriptors: parameters
- Response (y) = Function (θ) of input/explanatory variables (x)

Role of models

- The role of chance:
  - NOT: no reason for happening
  - But: Don't know the reasons
  - Combined effect of unmeasured/unknown factors
- Stochastic systems
- EVERYWHERE -- >> stochastic i.e., statistical models
- Only predictable up to a certain degree
- Describe act of chance by probability distributions
- Statistical models:

  Response (y) = systematic/deterministic + stochastic/random

Role of models

- Objectives of modeling:
  - various
  - Broadly two:
    - Explain/understand: which parameters are important ?
    - Predict: given θ and x, what is y ?
  - Common to both:
    - Estimation of parameters
    - Analysis of models

Role of models

- **Everybody has a model, whether he knows it or not.**
- Everybody is a modeller, whether he knows it or not.
- Interpretation of data without a model is impossible

  (model: broadly, set of assumptions)
- Implicit models vs. explicit models
- Implicit models may be more or less appropriate
- ex. Model underlying CR interpretation of animal counts:
  - closed population, no unmodelled heterogeneity in p
- Model underlying interpretation of raw counts of animals
  - same plus p=1 or E(p) = constant over dimensions of interest

Analysis of models

Analysis of models

- Inference: from observed data (x, y) find best guess at θ (what's inside the box that produces the observations ?)
- How should we deal with the uncertainty about θ ?
- Difference between Bayesian and classical analysis of models !
- (there are no "Bayesian models")

$$x \longrightarrow \boxed{\theta} \longrightarrow y$$

Analysis of models

Classical analysis

- Data result of stochastic process
- Parameters fixed and unknown
- Probability defined as long-run frequency of events in hypothetical replicates -> Frequentist statistics

$$x \longrightarrow \boxed{\theta} \longrightarrow y$$

Analysis of models

Bayesian analysis

- Data result of stochastic process
- Parameters fixed and unknown ...
- ... however: Probability defined as degree of belief in, or plausibility of, an event or the magnitude of a parameter
- no hypothetical replicate data sets
- Uncertainty about parameters described by probability
- Hence, parameters ***treated*** as if they were random variables

$$x \longrightarrow \boxed{\theta} \longrightarrow y$$

Analysis of models

<u>Classical analysis</u>

- Sampling distribution of data: p(y | θ)

- Likelihood function central: sampling distribution read in reverse, i.e., p(y | θ) as function of θ **(-> sketch)**

- choose that value which maximises function as most probable value of unknown parameter

- long history (Fisher, 1920s), desirable features (e.g., **asymptotic** unbiasedness, consistency, invariance to transformation)

- probability statements about data sets, not about params

Analysis of models

<u>Bayesian analysis</u>

- Data, once collected, fixed
- What is known ? The data
- What is unknown ? θ
- What to do ? Calculate $p(\theta \mid y)$

"Probability of paramater, given data"

- How should that be done ?
- Bayes rule !

Analysis of models

- Bayes rule
- $p(A|B) = p(B|A) * p(A) / p(B)$
- fact of conditional probability
- provable from three axioms of probability
- Application of Bayes rule to observable quantities undisputed, e.g., diagnostic testing, e.g. p(disease | positive test result)
- What Bayes did (1760s): apply rule to unobservables (parameters)

Analysis of models

- Bayes rule

$$p(\theta|y) = p(\theta|y) * p(\theta) / p(y)$$

$p(\theta|y)$: posterior distribution

$p(y|\theta)$: likelihood

$p(\theta)$: prior distribution

$p(y)$: normalising constant

- $p(\theta|y)$ proportional to $p(\theta|y) * p(\theta)$
- "posterior proportional to likelihood x prior"
- Prior: Need express uncertainty about params by probability
- Subjective !?

Analysis of models

- Hence, Bayesian statistics:
  - adopt degree-of-belief concept of probability
  - use Bayes rule to update your knowledge (prior) before inspecting your data set to the posterior (what you know afterwards)
  - ALL inference (probabilistic conclusions about unknowns) based on posterior distribution (**-> sketch that**)
  - all unkowns have a distribution
  - can summarise distribution (e.g., mean=point estimate, sd=SE)
  - make direct probability statement about parameter (θ) !

Analysis of models

- Bayes rule

$$p(\theta|y) = p(\theta|y) * p(\theta) / p(y)$$

- "posterior proportional to likelihood x prior"
- like human learning:
  - Conclusions = experience plus new information
  - weigh evidence in the data by experience (e.g., turkey 2 m high ?)

Analysis of models

- Big problem with application of Bayes rule: denominator !

$$p(\theta|y) = p(\theta|y) * p(\theta) / \mathbf{p(y)}$$

- high-dimensional integral

- unsolvable for most except for simple problems

- Hence, in theory better stats, but unfortunately cannot be applied to practical problems

- until 1980s: development simulation-based methods for estimating the posterior distribution

- Markov chain Monte Carlo (MCMC)

Bayesian computation

- Key idea: don't solve Bayes rule analytically, but use simulation to draw random samples from posterior distribution
- Metropolis-Hastings algorithm (1953, 1970)
- Gibbs-Sampling (1984)
- More and more powerful computers
- --> „Bayesian statistics without tears" (Smith & Gelfand 1993)

Bayesian computation

- Markov chain Monte Carlo (MCMC)
- Markov chain: sequence of numbers where there is a dependency only among immediate neighbours
- Has stationary distribution
- Choose Markov chain so that stationary distribution equals the desired posterior distribution
- Have to start somewhere, arbitrary, should converge eventually on target (though in practice can get lost)
- How does one assess convergence ?
- Have to discard starting bit of chain
- **-> Sketch**

Bayesian computation

- MCMC: Black box (but compare Newton-Raphson algorithm, iteratively reweighted least-squares)
- simple and intuitive understanding of key features:
  - Initial values
  - Autocorrelation
  - Convergence
  - Target is distribution (posterior distribution), not a single point
  - Effectively gives us left-hand-side of Bayes rule

Bayesian computation

- Want to know more about MCMC ?
- see rich literature, e.g., Link & Barker (2010) and many others

- Great thing about MCMC: allows one to get solutions (parameter estimates) for in principle arbitrarily complex statistical models
- Problem with MCMC: required special computational knowledge and custom-written code, e.g. in Fortran or C (or R)
- Said to be quite doable, and yet, ...
- out of reach for most ecologists !
- But then came ...

WinBUGS !

- Windows implementation of a computer program that does **B**ayesian **A**nalysis **u**sing **G**ibbs **S**ampling
- Place and Date of Birth: UK (Cambridge), 1989; see Lunn et al. (2009)
- Can be great pain in the neck, and yet, ONLY currently available program accessible to somewhat numerate ecologists
- Groundbreaking software
- **Frees the modeller in you !**

WinBUGS

- **B**ayesian **A**nalysis using **G**ibbs **S**ampling
- What does WinBUGS do ?
  - Gibbs sampling and a few other MCMC flavours
  - Use its own S-like model description language to translate our model and construct recipe for Markov chains that should converge onto desired posterior distribution
  - Let the chains evolve for desired, arbitrary number of steps (=iterations, updates, draws)
  - Some facilities to check chains (convergence) and to summarise them

WinBUGS

- Standalone application under Windows

- Almost always best run from R (e.g., via interface R2WinBUGS package or BRugs)

- Active developmental branch of WinBUGS has moved over to OpenBUGS

Advantages of Bayesian statistics

- Useful: Get solutions for complex models
- Exact: inference exact for YOUR sample size (not asymptotic, interesting for small ecological data sets)
- Ease of error propagation: using posterior samples from MCMC analysis allows derived quantities to be computed trivially easily (forget the delta rule)
- Formal framework for combining information: via priors, hence can avoid feigning being stupid if you want to do
- Intuitive: Definition of probability and uncertainty assessment very intuitive („I'm 100% certain ...")
- Coherent: internally logical foundations, three axioms explain it all

Advantages of Bayesian statistics

- Mathematical formalisation of learning
- Possible to specify (largely) absence of prior knowledge -> then get parameter estimates that greatly resemble classical MLEs

Disadvantages of Bayesian statistics

- Need a prior, and results (i.e., posterior distribution) depend on the prior
- Especially for „small" sample sizes, choice of prior makes a difference
- „Small" defined by ratio of information in data and complexity of the model -> usually, we always push modeling so hard until all data sets become small
- can be difficult to avoid injecting information unknowingly
- Computations (MCMC) hard to understand
- Take a long time to run

This workshop

- Gentle introduction to applied Bayesian modeling using WinBUGS
- Tutorial of sample analysis
- Progression of trivially simple to moderately complex

	Single random process	Two or more random processes
Normal response	Linear model (LM)	Linear mixed model (LMM)
Exponential family response	Generalized linear model (GLM)	Generalized linear mixed model (GLMM)

Table 1-1: Classification of some core models

This workshop

- Comprehensive and non-mathematical overview of linear models, GLMs, linear mixed models and GLMMs
- Follow sample analyses step-by-step
- Learning by example best for many/most ecologists
- Do all analysis using classical stats an R function such as lm(), glm() and lmer() as well using Bayesian stats using WinBUGS
- Observe great numerical similarity between them
- Gain confidence in „new" approach
- And especially: Only simulated data !

This workshop

- Why simulated data ?
    - Truth is known, i.e., know what solutions should look like
    - Check that have not made coding error in WinBUGS
    - Get a grasp for what sampling error is (normally only ever see a single replicate of the number-generating stochastic machine)
    - Check frequentist characteristics of estimates, e.g., bias, precision
    - Decide on required sample sizes
    - Check whether params estimable/identifiable

This workshop

- Why simulated data ? (ctd.)
  - Check effects of assumption violations
  - And especially: Prove to yourself that you have understood a model; build up a data set and then break it down again using the analysis (data analysis is like fixing a motor-bike ...)
  - and finally, imagine we studied really nice and attractive organisms (rather than pupils in schools in town districts of London)

This workshop

- **Caveat !**
- Hardly any theory -> go to books like Link & Barker (2010) or an increasingly large number of other sources
- Really nothing on Bayesian computation (MCMC): ditto

This workshop

- Some further reading:
- Bayesian inference and computation:
    - Link & Barker (10), Ntzoufras (09), Gelman & Hill (07), McCarthy (07), Gelman *et al.* (04); also Lindley (06)
- Program R:
    - Dalgaard (01), Crawley (05), Venables & Ripley (02), Bolker (08)
- Ecological statistics:
    - Buckland et al. (01), Borchers et al. (02), Williams et al. (02), Royle & Dorazio (08), King et al. (09), Link & Barker (10)

This workshop

Main goals:

1. Demystify Bayesian analyses in most widely used, general-purpose modeling program WinBUGS
2. Enhance understanding of core of applied statistics: linear, generalised linear, linear mixed and generalised linear mixed models
3. Demonstrate great value of simulation
4. Free the modeller in you