Bayesian statistics, WinBUGS , and R

Russell Millar

Department of Statistics

University of Auckland

Outline

1. Bayesian versus frequentist
2. Bayes theorem
3. Bayesian inference
4. Prior distributions
5. Bayesian computation
6. The Gibbs sampler
7. Introduction to WinBUGS
8. Hierarchical models
9. BUGS and R
10. Bayesian diagnostics
11. Other topics

Bayes vs frequentist

Bayesian statistics is not simply another statistical tool to be added to ones analytical toolbox. It is a different way of thinking about, and doing, statistical analyses of all types.

Words of Caution

from B. Dennis, Should ecologists become Bayesians? (1996, Eco. Apps., p. 1095-1103)

"Ecologists should be aware that Bayesian methods constitute a radically different way of doing science. Bayesian statistics is not just another tool to be added into ecologists' repertoire of statistical methods. Instead, Bayesians categorically reject various tenets of statistics and the scientific method that are currently widely accepted in ecology and other sciences. The Bayesian approach has split the statistics world into warring factions (ecologists' "density independence" vs "density dependence" debates of the 1950s pale by comparison), and it is fair to say that the Bayesian approach is growing rapidly in influence."

More words to ponder

from M. Maunder, Paradigm shifts in fisheries stock assessment: from integrated analysis to Bayesian analysis and back again. (2003, Nat. Res. Mod., p. 465-475)

"…recent advances in integrated analysis, including the special case of meta-analysis, have made Bayesian analysis somewhat redundant. I describe how data used to create priors for use in Bayesian analysis can be integrated directly into the analyses. This provides a much more convenient way of accurately including the information and associated uncertainty into the analyses."

Very brief history of statistics

- Bayesian conceptual framework was developed by the Reverend Thomas Bayes (1702-1761), and published posthumously in 1764.
- Classical philosophy formalized in early 20th century (Karl Pearson, Ronald Fisher et al.) and quickly became dominant.
- Revival of Bayesian statistics in late 20th century due largely to computational advances (MCMC, BUGS software, etc.).

Classical/Frequentist Statistics

- Fixed-effects parameters are fixed and unknown.
- Probabilities are defined to be the long term average under repetition of the experiment.
  - If a balanced coin is tossed many times then, on average, it will be Heads half the time.
  - 95% confidence intervals are constructed so that they will contain the parameter 95 times out of 100 under repetition of the experiment.
  - No probabilistic interpretation for the particular experiment performed.

Classical/Frequentist Statistics

Pre-experiment interpretation:

- What is the probability of snow tomorrow?
- What is the probability of overfishing this stock?

Classical/Frequentist Statistics

Post-experiment interpretation:

- At U. Auckland, stage 2 stats students are used as replicates of the experiment that observes 100 values from a standard normal distribution, N(0,1). They each compute 95% confidence intervals for μ (=0).
  - In a large class there will typically be at least one CI that lies entirely below zero, and another that lies entirely above.
- Suppose that I toss a balanced coin, but do not reveal the outcome. What is the probability of Heads?

Bayesian Statistics

There is a common misperception that Bayesians regard parameters as random (e.g, Dixon and Ellison, 1996, Eco. Apps).

- Fixed-effects parameters are fixed and unknown (same as Classical).
  - We express our uncertainty about the parameter using a probability distribution. Probability is regarded as an expression of belief.
  - Similarly, knowledge/uncertainty about any other unobserved quantities (e.g. probability of collapse), or preference for competing hypotheses, etc., is expressed probabilistically.
  - Very natural to model random-effects.

Bayesian Statistics

Requires specification of prior knowledge

- Experimental observation is used to update the prior knowledge to obtain posterior knowledge.
- Uses Bayes formula

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

In the next Chapter:

$$\pi(\theta \mid y) = \frac{f(y,\theta)}{f(y)}$$

Bayesian Statistics

Post-experiment interpretation:

- At U. Auckland, stage 2 stats students are used as replicates of the experiment that observes 100 values from a standard normal distribution, N(0,1). They each compute 95% credible intervals for μ (=0).
  - In a large class there will typically be at least one CI that lies entirely below zero, and another that lies entirely above.
- Suppose that I toss a balanced coin, but do not reveal the outcome. What is the probability of Heads?

Example: Future biomass

Objective: Determine the future biomass, *B*, of some fishery, under various management strategies.

Assume that we have some data to which an appropriate population dynamics model can be fitted.

Frequentist: Future biomass

Frequentist – might be able to produce a confidence interval for *B*.

How is this used to determine the consequences of various management strategies?

Bayesian: Future biomass

The posterior distribution for *B* incorporates prior information/uncertainty, information provided by the data.

Moreover, it quantifies the probability of *B* taking certain values.

Bayesian vs Frequentist

Who's right?

- Jury still divided after nearly a century.
- Bayesians have moral high ground.
  - Frequentists rely on the conditionality principle (CP) to define "replication" of the experiment. However, CP implies the likelihood principle (LP). LP is consistent with Bayesian method, but not frequentist.
- Frequentists claim objectivity.
  - Specification of priors is seldom unequivocal and "non-informative" priors can only be rigorously defended in the simplest of cases. Frequentists view priors as subjective.

Bayesian vs Frequentist

Classical: some good features

- Regarded as objective.
- Works well in many cases.
- Well developed methodology
  - Unbiasedness of estimators
  - Efficient estimators (e.g., minimum variance)
  - Very general large sample (asymptotic) properties based on maximum likelihood.
  - Model checking.

Classical: some bad features

- Technical problems re conditionality
  - Numerous examples demonstrate the problems, notwithstanding that these examples are all rather contrived.
- Not doing what we want
  - The null hypothesis is always false, so why bother testing it?
  - p-values *do not* quantify how well the data support the hypothesis.
  - Doesn't permit probabilistic interpretations, e.g., what is the probability of extinction?

Classical: another bad feature?

- Restrepo et al. (1992, Fish. Bull. 90: 736-748) present a Monte-Carlo approach (MCA) for incorporating uncertainty in nuisance parameters that are assumed known in stock assessment models.
  - Randomly simulate a value of the nuisance parameter(s) from some appropriately estimated or specified distributed.
  - Bootstrap the stock assessment data.
  - Re-fit the stock assessment model conditional on the simulated value(s) of the nuisance parameters, and save the MLE.
  - Repeat the above steps a bunch of times.
  - Use the distribution of these simulated MLE's to approximate the sampling distribution of the actual MLE.

Classical: another bad feature?

- MCA was subsequently used by Punt and Butterworth (1997, Rep. Int. Whal. Comm. 47: 603-618) in an assessment of the Bering-Chukchi-Beaufort Seas bowhead whale.
- Poole et al. (1999, Fish. Bull. 97: 144-152) took a closer look at MCA and identified serious concerns
  - MCA simulates the nuisance parameter(s) from a specified distribution, much like using a prior.
  - However, a Bayesian analysis would update the prior using the stock assessment data - MCA does not do this.
  - For example, a Bayesian analysis would "down-weight" nuisance parameter values that were not well supported by the stock assessment model. MCA does not do this.

Classical: another bad feature?

However, some folk would argue that the problem with the MCA example does not establish an inherent weakness in the classical approach.

Rather, the nuisance parameter uncertainty should be incorporated in a different way, e.g., using "integrated" likelihood analysis.

Bayesian: some good features

- Provides probabilistic interpretations.
- Logically consistent method to update prior knowledge and uncertainty, using information contained in the data.
- Uses prior knowledge, which often exists.
- MCMC (Markov Chain Monte Carlo) provides very general purpose software for fitting complex models using off-the-shelf software.

Bayesian: some bad features

- Needs prior knowledge, which is seldom unequivocal.
- “Non-informative” priors are mythical in all but the simplest models.
- “Reference” priors may be complicated and improper.
- Posterior will depend strongly on prior if data poor.
- MCMC is dangerous.
- Techniques for model evaluation, diagnostics, sensitivity etc., are less well developed.

Bayesian vs Frequentist inference

Point Estimation

- Frequentists rely on the well developed theory of minimum variance unbiased estimators (MVUE).
  - Sample mean and least squares estimators are MVUE for linear models with normal data.
  - Maximum likelihood estimators are asymptotically MVUE.
- Bayesians typically use posterior mean, which is known as the “Bayes estimator”.
  - Posterior mean is optimal under squared error loss.

Bayesian vs Frequentist inference

Hypothesis testing

- Frequentists attempt to reject a null hypothesis, H_0, using a test which has small probability of falsely rejecting (though H_0 is usually known to be false!).
  - Methodology is based on computing most powerful tests – that is, tests with greatest probability of rejecting H_0 when it is false.
- Bayesians specify prior probabilities on two (or more) hypotheses and obtain the posterior probabilities.
  - Optimal choice is the hypothesis with highest posterior probability.

Bayesian vs Frequentist inference

Interval Estimation

- Frequentists compute confidence intervals with a given "coverage" probability.
  - "coverage" probability is interpreted as the proportion of such intervals expected to contain unknown parameter under repetition of the experiment.

- Bayesians use intervals of highest posterior density.
  - The HPD interval is the minimum width interval containing 95% (say) posterior probability.