Session 1

Revision and practical info

Matt Denwood

2024-01-31

Revision

Bayes Rule

Bayes' theorem is at the heart of Bayesian statistics:

$$P(\theta|Y) = \frac{P(\theta) \times P(Y|\theta)}{P(Y)}$$

Bayes Rule

Bayes' theorem is at the heart of Bayesian statistics:

$$P(\theta|Y) = \frac{P(\theta) \times P(Y|\theta)}{P(Y)}$$

Where: θ is our parameter value(s);

Y is the data that we have observed;

$P(\theta|Y)$ is the posterior probability of the parameter value(s);

$P(\theta)$ is the prior probability of the parameters;

$P(Y|\theta)$ is the likelihood of the data given the parameters value(s);

$P(Y)$ is the probability of the data, integrated over parameter space.

- In practice we usually work with the following:

$$P(\theta|Y) \propto P(\theta) \times P(Y|\theta)$$

- In practice we usually work with the following:

$$P(\theta|Y) \propto P(\theta) \times P(Y|\theta)$$

- Our Bayesian posterior is therefore always a combination of the likelihood of the data, and the parameter priors
- But for more complex models the distinction between what is 'data' and 'parameters' can get blurred!

MCMC

- A way of obtaining a numerical approximation of the posterior
- Highly flexible, and easy(ish) using JAGS (or OpenBUGS, or Stan)
- Not inherently Bayesian but most widely used in this context
- Assessing convergence is essential, otherwise we may not be summarising the true posterior
- Our chains are correlated so we need to consider the effective sample size

Hui-Walter models

- A specific class of model for paired diagnostic test data
- Usually (but not necessarily) fit using MCMC
- Requirements are 2 or more tests in 2 or more populations (or 3 tests in 1 population)
- Sensitivity and specificity must be consistent between populations
- Tests must be conditionally independent, although correlation terms can be added
- Easiest to generate using runjags::template_huiwalter

Everyone up to speed?

Any questions so far?

Anything unclear?

...

All OK with GitHub?

Learning outcomes

By the end of the course you should be able to:

- Understand how and why to use simulated data in the context of Hui-Walter models (session 2)
- Use simulation to estimate the expected width of 95% confidence intervals for a given sample size (session 3)
- [Bonus material]: a self-directed “sneak preview” of upcoming features of template_huiwalter including post-hoc estimates of se/sp stratified by population (session 4)