An Introduction to Using WinBUGS for Cost-Effectiveness Analyses in Health Economics

Dr. Christian Asseburg

Centre for Health Economics
University of York, UK

ca505@york.ac.uk

Part 1

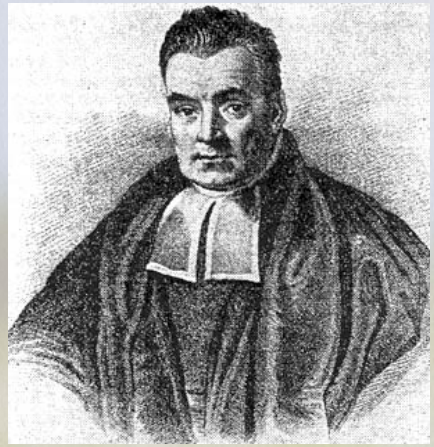
Bayesian Statistics: An Introduction

Talk overview

- Foundations of Bayesian statistics
- Comparison between Frequentist and Bayesian approaches
- Calculations and computer implementations
- Example from health economics
- Questions and discussion

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

History of Bayesian Statistics


Revd. Thomas Bayes

The reverend Thomas Bayes (1702-1761) proved a special case of what is now known as *Bayes' Theorem*.

Pierre-Simon Laplace (1749-1827) proved a more general version of *Bayes' Theorem* and used it for various applications.

The relevance of *Bayes' Theorem* to statistics, however, was not appreciated until the 20th century.

The *Frequentist paradigm* has been the mainstay of probability theory during the 19th and 20th century, with important contributions by e.g. Jerzy Neyman, Egon Pearson, John Venn, R.A. Fisher, and Richard von Mises.

Frequentist tools such as hypothesis testing and confidence intervals have allowed many advances in statistics. *Bayesian* equivalents exist, but they often require more computations – it was during the last two decades of increasing availability of computing resources that Bayesian statistics gained ground.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Bayes' Theorem

Bayes' Theorem can be derived easily from the expression of the joint probability of two events A and B:

Let *p(A)* denote the probability that event A will occur, let *p(B)* denote the probability that event B will occur, and let *p(A,B)* denote the probability that both of the events occur.

Then $$p(A,B) = p(A) \cdot p(B|A) = p(B) \cdot p(A|B)$$

Bayes' Theorem states simply that

$$p(B|A) = \frac{p(B) \cdot p(A|B)}{p(A)}$$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Priors and Posteriors (1)

Of course, *Bayes' Theorem* as a way to relate the conditional probabilities of two events is valid both in Frequentist as well as in Bayesian statistics.

However, in Bayesian statistics it is also applied to unknown parameters *x* directly:

$$p(B|A) = \frac{p(B) \cdot p(A|B)}{p(A)}$$

$$p(x|data) = \frac{p(x) \cdot p(data|x)}{p(data)}$$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Priors and Posteriors (2)

Unknown parameter(s):	*x*
Data (known):	*data*
Probability of *data* given *x*:	*p(data|x)*
"Prior" probability of *x*:	*p(x)*
"Posterior" probability of *x*:	*p(x|data)*

$$p(x|data) = \frac{p(x) \cdot p(data|x)}{p(data)}$$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Priors and Posteriors (2)

Unknown parameter(s):	*x*
Data (known):	*data*
Probability of *data* given *x*:	*p(data|x)*
"Prior" probability of *x*:	*p(x)*
"Posterior" probability of *x*:	*p(x|data)*

"Likelihood"

$$p(x|data) = \frac{p(x) \cdot p(data|x)}{p(data)}$$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Priors and Posteriors (2)

Unknown parameter(s):	*x*
Data (known):	*data*
Probability of *data* given *x*:	*p(data|x)*
"Prior" probability of *x*:	*p(x)*
"Posterior" probability of *x*:	*p(x|data)*

$$p(x|data) = \frac{p(x) \cdot p(data|x)}{p(data)}$$

The denominator is a constant and can usually be ignored.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Priors and Posteriors (3)

Bayes' Theorem is thus used to combine data with a prior belief on an unknown quantity, resulting in a posterior belief on the unknown quantity.

This approach has been compared to the task of *learning* in humans, where experience supports a constant updating of a person's belief system.

"Prior" probability of *x*: $p(x)$

"Posterior" probability of *x*: $p(x|data)$

$$p(x|data) = \frac{p(x) \cdot p(data|x)}{p(data)}$$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Definition of "Probability"

F R E Q U E N T I S T

B A Y E S I A N

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Definition of "Probability"

FREQUENTIST

The *"probability"* of an event A occurring (or of a quantity taking a value in a given interval) is a **frequency**. Imagine many (hypothetical or actual) circumstances in which the data have been observed. The proportion of circumstances in which event A occurs (out of all circumstances) is the *"probability"* of A. This probability is *objective*.

BAYESIAN

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Definition of "Probability"

FREQUENTIST

The *"probability"* of an event A occurring (or of a quantity taking a value in a given interval) is a **frequency**. Imagine many (hypothetical or actual) circumstances in which the data have been observed. The proportion of circumstances in which event A occurs (out of all circumstances) is the *"probability"* of A. This probability is *objective*.

BAYESIAN

The *"probability"* of an event A occurring (or of a quantity taking a value in a given interval) is a **degree of belief**. The degree of belief in A may change when we are confronted with new data. The *"probability"* of A is a numerical representation of this degree of belief.
If you and I (and everyone else) agree on the belief in event A, we define an *objective* probability, otherwise we define a *subjective* probability.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

What is fixed, what is random? (1)

FREQUENTIST

BAYESIAN

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

What is fixed, what is random? (1)

FREQUENTIST

There is a **fixed, but unknown value** for each parameter. The data are an instance of many **possible data** that could have been collected. A Frequentist statistician evaluates *how likely the given data* are according to different hypothetical values for the unknown quantities. Thus, statements about the probability of observing the data given different hypothetical parameter values are summarised in a **confidence interval**.

BAYESIAN

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

What is fixed, what is random? (1)

FREQUENTIST

There is a **fixed, but unknown value** for each parameter. The data are an instance of many **possible data** that could have been collected. A Frequentist statistician evaluates *how likely the given data* are according to different hypothetical values for the unknown quantities. Thus, statements about the probability of observing the data given different hypothetical parameter values are summarised in a **confidence interval**.

BAYESIAN

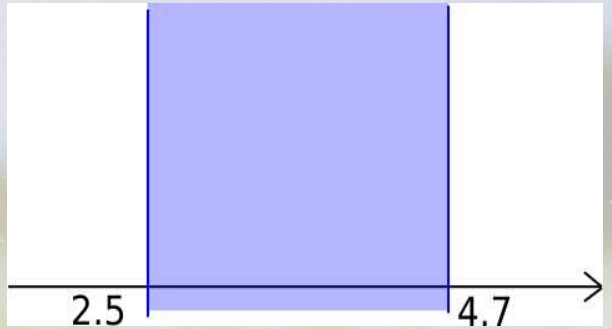
The value for **each parameter is unknown**. The **data are known**, they have been observed. A Bayesian statistician evaluates *how likely different values* for the underlying quantities are, given the observed data. Thus, statements can be made about the probability of the unknown quantity taking a value in a certain **credibility interval**.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

What is fixed, what is random? (2)

FREQUENTIST

A 95% **confidence interval** for a quantity *x*:



BAYESIAN

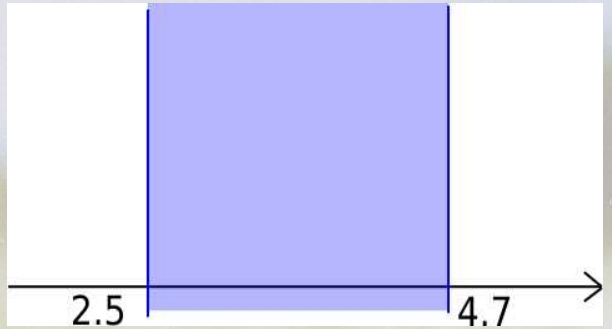
Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE
Centre For Health Economics

What is fixed, what is random? (2)

FREQUENTIST

A 95% **confidence interval** for a quantity *x*:

"If new data are collected many times and confidence intervals are calculated, then 95% of these confidence intervals contain the true value of *x*."



BAYESIAN

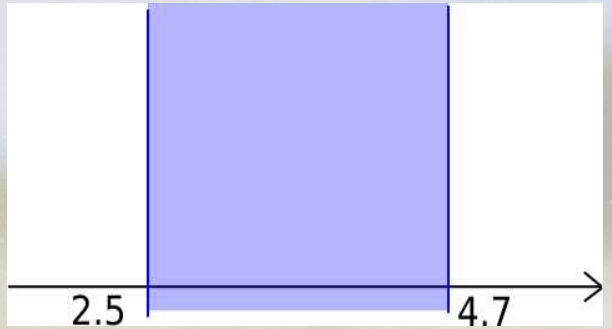
Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE
Centre For Health Economics

What is fixed, what is random? (2)

FREQUENTIST

A 95% **confidence interval** for a quantity *x*:

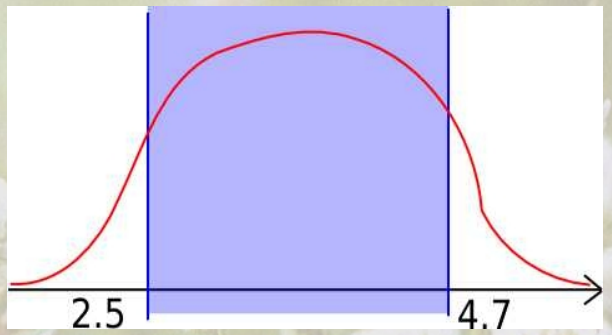
"If new data are collected many times and confidence intervals are calculated, then 95% of these confidence intervals contain the true value of *x*."



BAYESIAN

A 95% **credibility interval** for a quantity *x*:

"The probability that the value of *x* lies between 2.5 and 4.7 is 95%, given the observed data and the prior belief."



Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE
Centre For Health Economics

Hypothesis testing

FREQUENTIST

Given two hypotheses, H_0 and H_1, ...

BAYESIAN

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Hypothesis testing

FREQUENTIST

Given two hypotheses, H_0 and H_1, calculate the probability of observing the data (or more extreme data) if H_0 is true. If this probability is low (*p-value*), reject H_0.

BAYESIAN

Given two hypotheses, H_0 and H_1, ...

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Hypothesis testing

FREQUENTIST

Given two hypotheses, H_0 and H_1, calculate the probability of observing the data (or more extreme data) if H_0 is true. If this probability is low (*p-value*), reject H_0.

BAYESIAN

Given two hypotheses, H_0 and H_1, calculate the probability of each of them, given the data and the priors. Favour the hypothesis that has the higher probability.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Hypothesis testing

FREQUENTIST

Given two hypotheses, H_0 and H_1, calculate the probability of observing the data (or more extreme data) if H_0 is true. If this probability is low (*p-value*), reject H_0.

→ Because a hypothesis is either true or false (this is just not known) and only the likelihood of observing the data is calculated, a Frequentist cannot assign a probability to each hypothesis.

BAYESIAN

Given two hypotheses, H_0 and H_1, calculate the probability of each of them, given the data and the priors. Favour the hypothesis that has the higher probability.

→ The probability of each of the hypotheses being true can be calculated. Relative statements (e.g. "H_0 is twice as likely as H_1") can be made.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

A Simple Example (1)

In roulette, a spin of the wheel results in a red or a black number (or 0). In one hour, the roulette wheel resulted in 25 red and 15 black numbers. What is the probability z that this wheel gives a red number?

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

A Simple Example (1)

In roulette, a spin of the wheel results in a red or a black number (or 0). In one hour, the roulette wheel resulted in 25 red and 15 black numbers. What is the probability *z* that this wheel gives a red number?

FREQUENTIST

The probability of observing 25 red and 15 black numbers can be described by a Binomial distribution with 25 successes out of 40.

The sample proportion of success is 25/40, or 0.625. Using the central limit theorem, an approximate confidence interval for a proportion can be found. The sampling distribution is summarised by its mean (0.625) and standard deviation (0.0765), and these are used to obtain a 95% confidence interval for the mean of a normal distribution. After correcting for the discrete nature of the data, the confidence interval for *z* is found: [0.46, 0.79].

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

A Simple Example (2)

In roulette, a spin of the wheel results in a red or a black number (or 0). In one hour, the roulette wheel resulted in 25 red and 15 black numbers. What is the probability *z* that this wheel gives a red number?

B A Y E S I A N

The probability of observing 25 red and 15 black numbers can be described by a Binomial distribution with 25 successes out of 40.

The prior probability for *z* is assumed to be Beta(1,1).

Bayes' Theorem is used to calculate the posterior probability of *z*. (See next slide)

The 95% credibility interval for *z* is [0.47, 0.76].

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

A Simple Example (3)

In roulette, a spin of the wheel results in a red or a black number (or 0). In one hour, the roulette wheel resulted in 25 red and 15 black numbers. What is the probability *z* that this wheel gives a red number?

B A Y E S I A N

Bayes' theorem $$p(z|data) \propto p(z) \cdot p(data|z)$$

(The denominator of Bayes' Theorem, *p(data)*, is a constant and can usually be ignored.)

p(data|z) = Binomial (25 out of 40 with prob. *z*)

p(z) = Beta(1, 1)

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

A Simple Example (4)

In roulette, a spin of the wheel results in a red or a black number (or 0). In one hour, the roulette wheel resulted in 25 red and 15 black numbers. What is the probability *z* that this wheel gives a red number?

BAYESIAN

Bayes' theorem $p(z|data) \propto p(z) \cdot p(data|z)$

$$p(z|data) \propto \frac{1}{B(1,1)} z^{1-1}(1-z)^{1-1} \frac{40!}{25!(40-25)!} z^{25}(1-z)^{40-25}$$

$$p(z|data) \propto z^{25}(1-z)^{15}$$

So *p(z|data)* = Beta(26,16), and the credibility interval can be calculated easily by looking up the cumulative probabilities.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

A Simple Example (5)

In roulette, a spin of the wheel results in a red or a black number (or 0). In one hour, the roulette wheel resulted in 25 red and 15 black numbers. What is the probability z that this wheel gives a red number?

B A Y E S I A N

In this simple example, when the prior is from a particular family (Beta) and the likelihood of the data is also from a particular family (Binomial), the posterior likelihood also belongs to a particular family of distributions (Beta). The Beta prior and Binomial likelihood distribution are called **conjugate**.

This is a special case – usually the Bayesian posterior distributions cannot be calculated analytically, and numerical methods are required to approximate the posterior distribution.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

A Simple Example (6)

In roulette, a spin of the wheel results in a red or a black number (or 0). In one hour, the roulette wheel resulted in 25 red and 15 black numbers. What is the probability *z* that this wheel gives a red number?

B A Y E S I A N

Different choices of prior distributions lead to different posterior distributions and thus to different credibility intervals.

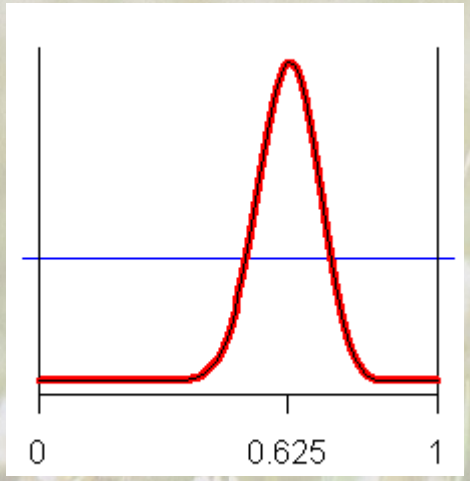
Prior	*Data*	*Posterior*	*95% credibility interval*
Beta(1,1)	25 out of 40	Beta(26,16)	[0.47, 0.76]
Beta(50,50)	25 out of 40	Beta(75,65)	[0.45, 0.62]
Beta(26,16)	25 out of 40	Beta(51,31)	[0.52, 0.72]

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

A Simple Example (7)

Different choices of prior distributions lead to different posterior distributions and thus to different credibility intervals.

Prior	*Data*	*Posterior*	*95% credibility interval*
Beta(1,1)	25 out of 40	Beta(26,16)	[0.47, 0.76]

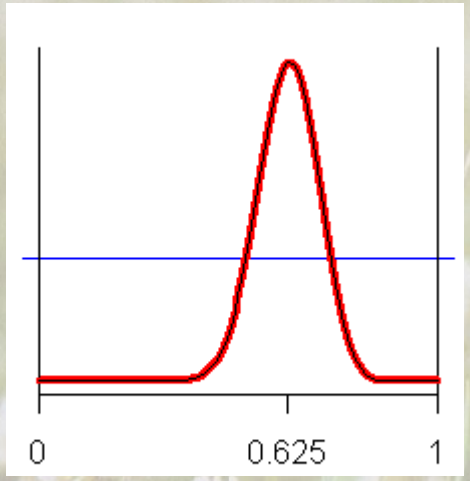


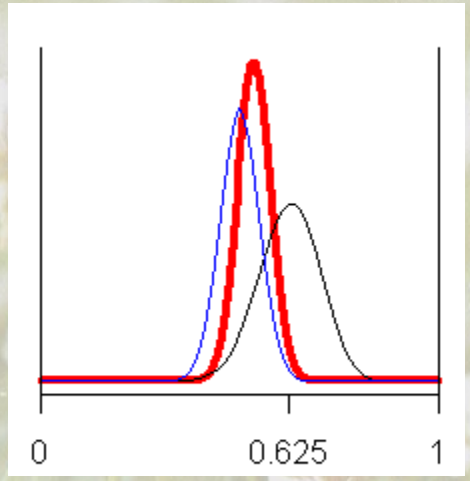
Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

A Simple Example (7)

Different choices of prior distributions lead to different posterior distributions and thus to different credibility intervals.

Prior	*Data*	*Posterior*	*95% credibility interval*
Beta(1,1)	25 out of 40	Beta(26,16)	[0.47, 0.76]
Beta(50,50)	25 out of 40	Beta(75,65)	[0.45, 0.62]



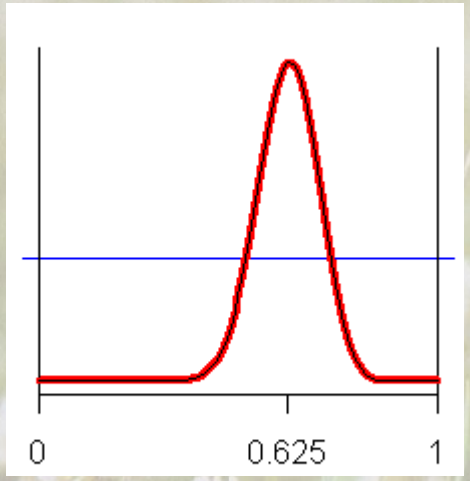


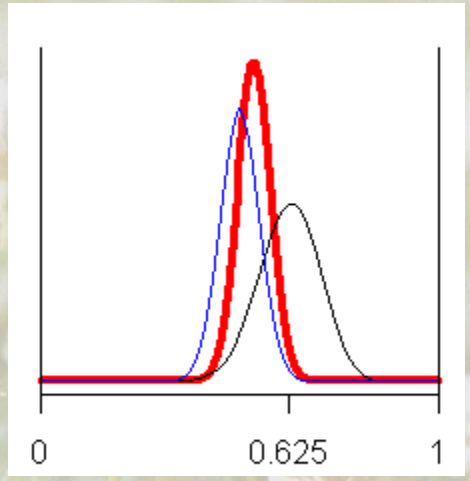
Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE
Centre For Health Economics

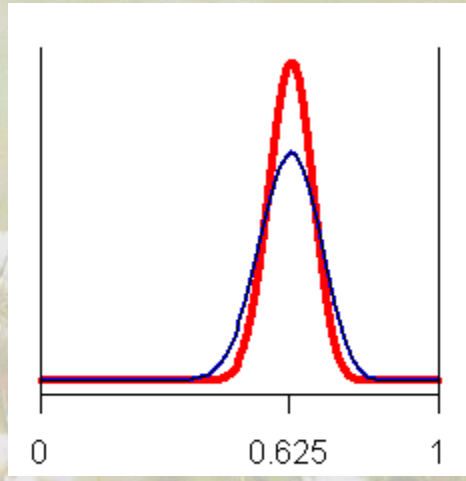
A Simple Example (7)

Different choices of prior distributions lead to different posterior distributions and thus to different credibility intervals.

Prior	*Data*	*Posterior*	*95% credibility interval*
Beta(1,1)	25 out of 40	Beta(26,16)	[0.47, 0.76]
Beta(50,50)	25 out of 40	Beta(75,65)	[0.45, 0.62]
Beta(26,16)	25 out of 40	Beta(51,31)	[0.52, 0.72]







Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE
Centre For Health Economics

Priors - again...

Different choices of prior distributions lead to different posterior distributions and thus to different credibility intervals.

Prior	*Data*	*Posterior*	*95% credibility interval*
Beta(1,1)	25 out of 40	Beta(26,16)	[0.47, 0.76]
Beta(50,50)	25 out of 40	Beta(75,65)	[0.45, 0.62]
Beta(26,16)	25 out of 40	Beta(51,31)	[0.52, 0.72]

So how does one choose **the "right" prior**?

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Controversy regarding priors

There is no “right” prior.

- A good prior choice may be obvious, for example when **earlier studies** on a model quantity can be used.
- The influence of the prior on the model output can be minimised by choosing an “**uninformative**” prior or a “**reference prior**”.
- If different stakeholders are involved, whose prior opinions on a model quantity differ, each of them may propose a prior. The model can then be run in turn for each prior. Afterwards, it may be possible to reconcile the different posterior opinions.

In general, if the prior choice makes a difference to the model's output, then more data should be collected. A good modelling application should either have an **informative** prior or be **robust** to prior choice.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Model Selection

In Bayesian statistics, it is relatively straightforward to evaluate different explanations for a data-set (nested models or totally different models). The models are all evaluated simultaneously, together with additional parameters m_i for the probabilities of each of the models.

The posteriors for the parameters m_i summarise how well each of the competing models fits the data. Depending on the model application, one most suitable model may be found, or predictions can be made from all models simultaneously, using the posterior values for m_i as weights (**model averaging**).

In Frequentist statistics, it is relatively easy to evaluate nested models – but the evaluation of other competing models is not straightforward.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Summary

	FREQUENTIST	BAYESIAN
Probability	Frequency	Belief
Statements	Probability of observing data	Probability of model quantity
Objectivity	Result depends only on data	Result depends on prior and data – subjective
Computation	Often feasible	Often complicated
Flexibility	Some applications require normal or other simplifying assumptions	no intrinsic limitations
Model selection	Sometimes possible	Straightforward

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Interval.....

..... any questions? .....

Good Bayesian text book that starts with a comparison of Bayesian and Frequentist methods:

D'Agostini, G: Bayesian Reasoning in Data Analysis.
World Scientific Publishing, Singapore, 2003.

Why use Bayesian methods in health economics? E.g. **B Luce, Y Shih, K Claxton:** *International Journal of Technology Assessment in Health Care 17/1, 2001, pp 1-5*.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Calculating Bayesian Posteriors

Models that can be solved **analytically** (such as the simple example before) are rare and require **conjugacy**. Most multi-dimensional models do not fall in this class.

The problem is that, for each possible set of parameter values, Bayes' Theorem gives the posterior probability, but if the parametric form of the distribution cannot be recognised, there is no obvious method for calculating e.g. its mean value, or for sampling from it.

Therefore, a Bayesian model usually requires **numerical methods** for calculating the posteriors of interest. Any algorithm that generates samples from a distribution that is defined by its probability density function could be used.

$$p(\Theta|data) \propto p(\Theta) \cdot p(data|\Theta)$$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Calculating Bayesian Posteriors

$$p(\Theta|data) \propto p(\Theta) \cdot p(data|\Theta)$$

The most commonly used algorithms for sampling from the Bayesian posterior fall in two groups:

- **Metropolis-Hastings**: some algorithms in this class are Markov chain Monte Carlo (MCMC), e.g. Gibbs sampling or Reversible Jump.

  → These algorithms work well when the posterior model space can be written as a product, such that factors correspond to subspaces.

- **Sequential Importance Sampling**.

  → Very suitable for posteriors that can be written as products, such that factors correspond to individual data.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Calculating Bayesian Posteriors

$$p(\Theta|data) \propto p(\Theta) \cdot p(data|\Theta)$$

The most commonly used algorithms for sampling from the Bayesian posterior fall in two groups:

- **Metropolis-Hastings**: some algorithms in this class are Markov chain Monte Carlo (MCMC), e.g. Gibbs sampling or Reversible Jump.

$$\rightarrow \quad p(\Theta|data) \propto p(\Theta_a)\, p(data|\Theta_a) \cdot p(\Theta_b)\, p(data|\Theta_b)$$

- **Sequential Importance Sampling**.

$$\rightarrow \quad p(\Theta|data) \propto p(\Theta) \cdot p(data_1|\Theta) \cdot p(data_2|\Theta) \cdot \ldots$$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Markov chain Monte Carlo (1)

MCMC generates samples from the posterior space **M** by defining a chain $C=\{x_1, x_2, x_3, ...\}$ in **M**.

At each step *i* in the chain, candidate values x^* are generated randomly for each of the parameters. (The **proposal distribution** may depend on the current values, x_i.)

The posterior probabilities are calculated for both x_i and x^*. Depending on the likelihood of x^* relative to x_i, an **acceptance probability** is calculated, and the chain either moves to x^* ($x_{i+1}=x^*$) or stays at its current value ($x_{i+1}=x_i$).

Ergodic theory ensures that, in the limit, the distribution of the values of C converges to the posterior distribution of interest. The beginning of the chain is discarded because the initial values dominate it ("**burn-in**").

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Markov chain Monte Carlo (2)

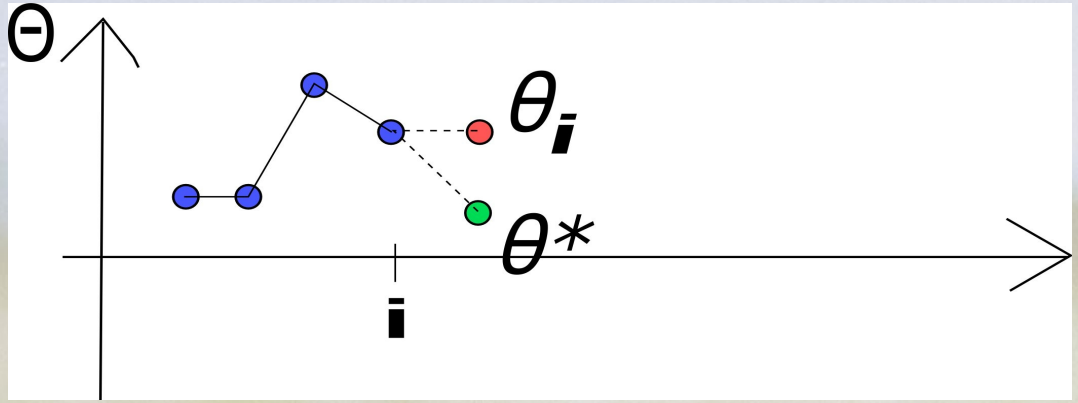


Illustration:

At step *i* in the MCMC, the chain may jump to the candidate value θ^* or stay at the current value θ_i.

This depends on the posterior probabilities for these two points in parameter space.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Markov chain Monte Carlo (2)

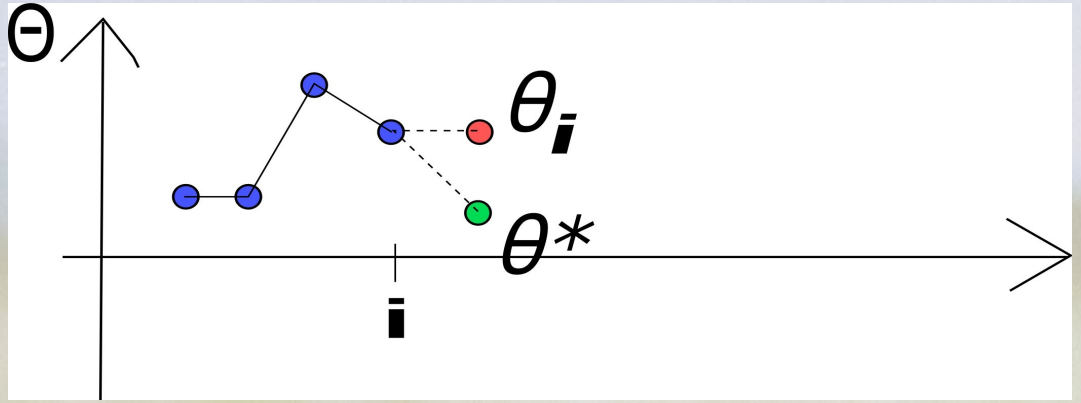
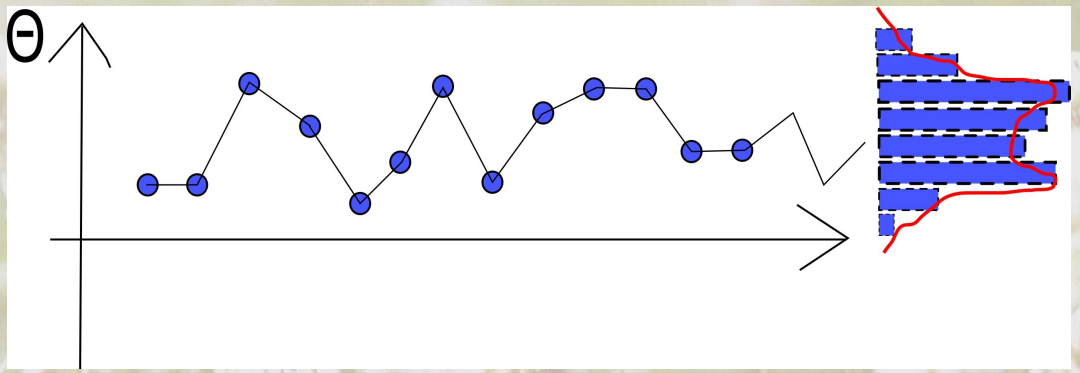


Illustration:

At step *i* in the MCMC, the chain may jump to the candidate value θ^* or stay at the current value θ_i.

This depends on the posterior probabilities for these two points in parameter space.

In the long run, the distribution of points in the chain approximates the posterior distribution.

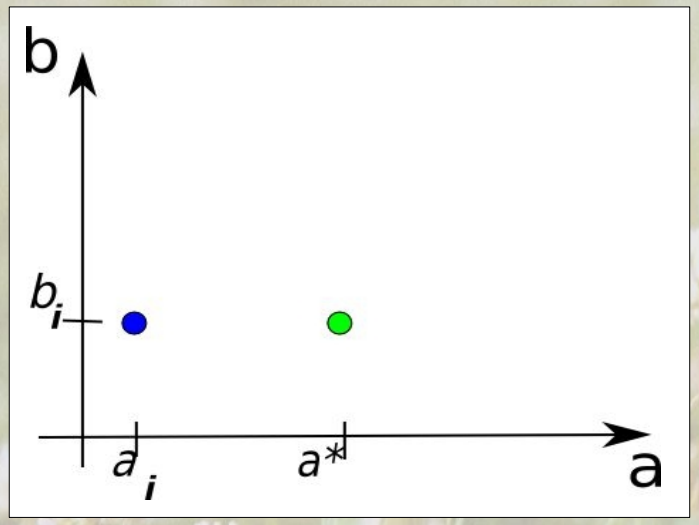


Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Markov chain Monte Carlo (3)

To sample from a multi-dimensional posterior (e.g. the posterior of a model with several unknown parameters), parameters can be grouped together (*block sampling*). Blocks are chosen such that calculations can be simplified.

At each iteration, a new candidate is suggested for one block (and parameters in the other blocks retain their current value). The candidate values for that block are either accepted or rejected. Then the same is done for the next parameter block, etc.



Example

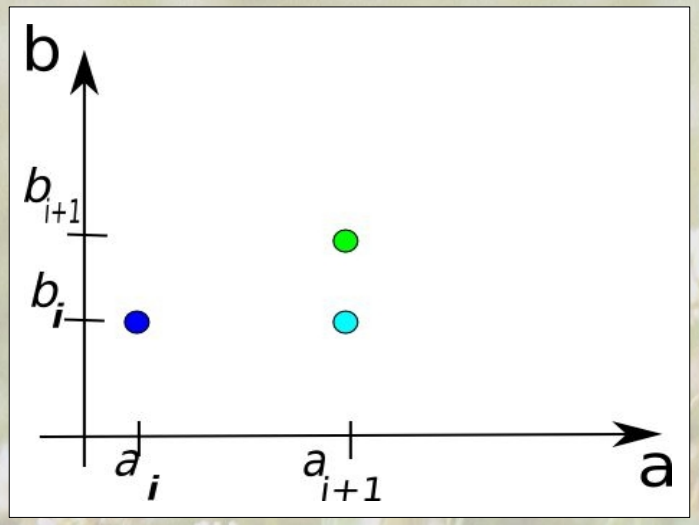
1. Suggest a candidate for *a*.
   (In this example, *a** is accepted.)
2.
3.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Markov chain Monte Carlo (3)

To sample from a multi-dimensional posterior (e.g. the posterior of a model with several unknown parameters), parameters can be grouped together (*block sampling*). Blocks are chosen such that calculations can be simplified.

At each iteration, a new candidate is suggested for one block (and parameters in the other blocks retain their current value). The candidate values for that block are either accepted or rejected. Then the same is done for the next parameter block, etc.



Example

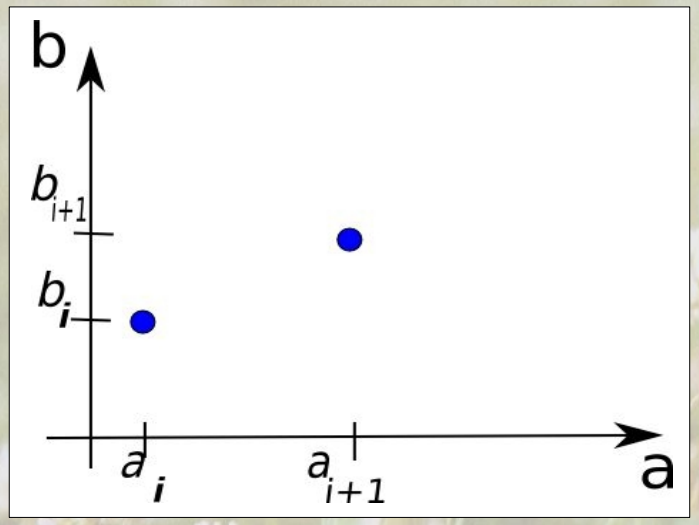
1. Suggest a candidate for *a*.
   (In this example, *a** is accepted.)
2. Suggest a candidate for *b*.
   (In this example, *b** is accepted.)
3.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Markov chain Monte Carlo (3)

To sample from a multi-dimensional posterior (e.g. the posterior of a model with several unknown parameters), parameters can be grouped together (*block sampling*). Blocks are chosen such that calculations can be simplified.

At each iteration, a new candidate is suggested for one block (and parameters in the other blocks retain their current value). The candidate values for that block are either accepted or rejected. Then the same is done for the next parameter block, etc.



Example

1. Suggest a candidate for *a*.
   (In this example, *a** is accepted.)
2. Suggest a candidate for *b*.
   (In this example, *b** is accepted.)
3. The chain moves to [*a**, *b**].

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Gibbs sampling

Gibbs sampling is a special case of MCMC. Here, the posterior parameter space is divided into blocks of parameters, such that for each block, the conditional posterior probabilities are known.

Then, at each step *i* in the chain, candidate values *x** are generated randomly from the conditional posterior probability for each parameter block, given the current values of the other parameters in the model.

Because *x** is a draw from the conditional posterior probability, the calculation of the MCMC acceptance probability always gives 1. Thus, the chain always moves to *x** (x_{i+1}=*x**). The sampler thus converges more quickly.

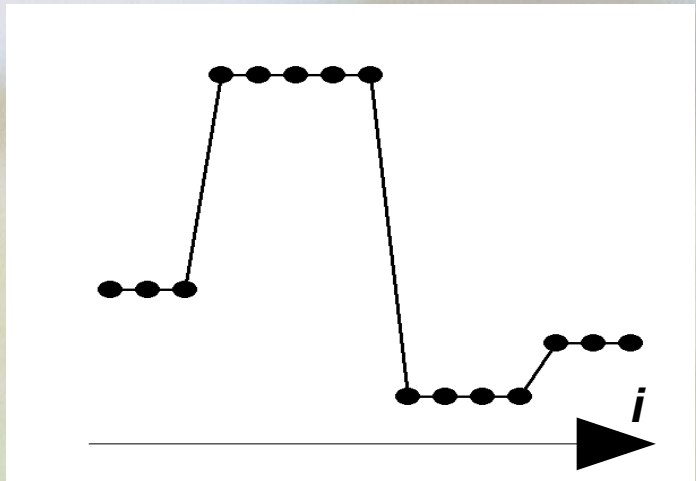
Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Difficulties with MCMC

Unfortunately, the Markov chain Monte Carlo algorithms do not always work well, and some care is needed when checking for convergence to the posterior distribution of interest.

The most common problems are:

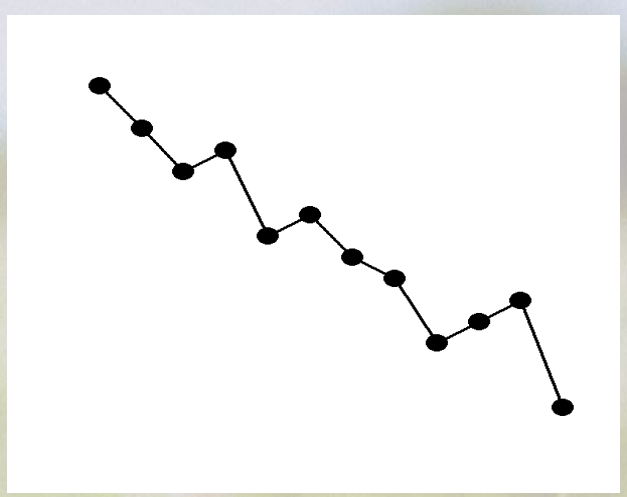
- **Bad "mixing"**: The chain does not move well because the candidate acceptance rate is too low.
  Cause: The candidate generator often suggests candidates that are too unlikely compared to the current value.



Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Difficulties with MCMC

Unfortunately, the Markov chain Monte Carlo algorithms do not always work well, and some care is needed when checking for convergence to the posterior distribution of interest.



The most common problems are:

- **Bad "mixing"**
- **Trends in the chain**: The exploration of posterior model space is slow and the chain seems to have a direction.
  Cause: The candidate generator suggests candidates too close to the current values.

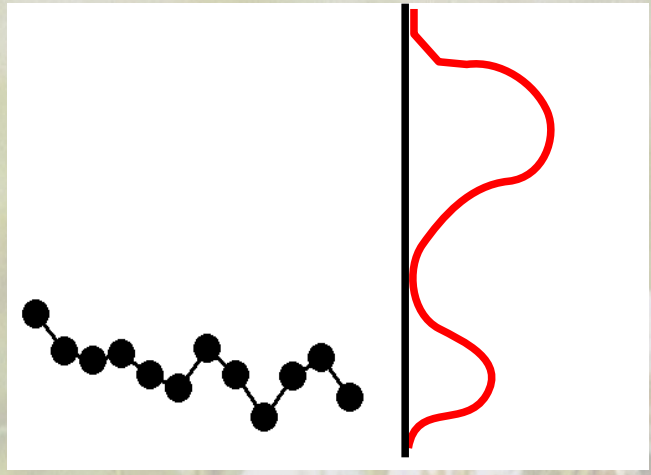
Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Difficulties with MCMC

Unfortunately, the Markov chain Monte Carlo algorithms do not always work well, and some care is needed when checking for convergence to the posterior distribution of interest.

The most common problems are:

- **Bad "mixing"**
- **Trends in the chain**
- **Poor coverage of posterior probability**:
  The chain seems to mix well, but it is stuck at a local maximum of posterior probability. The samples thus do not exhaust the posterior model space.
  Cause: Inappropriate candidate generator.



Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Difficulties with MCMC

Unfortunately, the Markov chain Monte Carlo algorithms do not always work well, and some care is needed when checking for convergence to the posterior distribution of interest.

The most common problems are:

- **Bad "mixing"**
- **Trends in the chain**
- **Poor coverage of posterior probability**

Because of these difficulties, generating samples from a Bayesian posterior requires a lot of attention to detail and can often not be fully automated.
Diagnostic criteria exist to aid in detecting convergence and good mixing of the MCMC sampler.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Sequential Importance Sampling

This algorithm is very suitable for data that can be obtained sequentially, for example to monitor an industrial process. SIS (usually implemented as a **particle filter**) can also be applied to more general problems.

In SIS, the posterior distribution of interest is approximated by a "swarm" of particles, where each particle is one possible realisation of the model. For example, in a model with two parameter values, a and b, a particle could be the pair (a=4.5, b=-2).

The posterior density function is split into factors, and at each step in the algorithm, all particles are resampled based on weights. These weights are derived from the factors that make up the pdf. For example, the first step might weight the particle sample according to the Bayesian prior. The second step might weight the updated set of particles according to the factor that corresponds to the first datum. The next resampling may take into account the next datum, etc, until the data are used up.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Sequential Importance Sampling

This algorithm is very suitable for data that can be obtained sequentially, for example to monitor an industrial process. SIS (usually implemented as a **particle filter**) can also be applied to more general problems.

$$p(\Theta|data) \propto p(\Theta) \cdot p(data_1|\Theta) \cdot p(data_2|\Theta) \cdot \ldots$$

Prior

- a=2.5
- a=3.1
- a=-1
- a=4
- a=2.7
- a=1.7
- ...

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Sequential Importance Sampling

This algorithm is very suitable for data that can be obtained sequentially, for example to monitor an industrial process. SIS (usually implemented as a **particle filter**) can also be applied to more general problems.

$$p(\Theta|data) \propto p(\Theta) \cdot p(data_1|\Theta) \cdot p(data_2|\Theta) \cdot \ldots$$

- a=2.5
- a=3.1
- a=-1
- a=4
- a=2.7
- a=1.7
- ...

1.5
0.1
5.5
0.0
1.3
3.4

Weights due to the first datum

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Sequential Importance Sampling

This algorithm is very suitable for data that can be obtained sequentially, for example to monitor an industrial process. SIS (usually implemented as a **particle filter**) can also be applied to more general problems.

$$p(\Theta|data) \propto p(\Theta) \cdot p(data_1|\Theta) \cdot p(data_2|\Theta) \cdot \ldots$$

Weighted resampling...

- a=2.5 1.5 a=-1
- a=3.1 0.1 a=-1
- a=-1 5.5 a=-1
- a=4 0.0 a=-1
- a=2.7 1.3 a=1.7
- a=1.7 3.4 a=1.7
- ...

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Sequential Importance Sampling

This algorithm is very suitable for data that can be obtained sequentially, for example to monitor an industrial process. SIS (usually implemented as a **particle filter**) can also be applied to more general problems.

$$p(\Theta|data) \propto p(\Theta) \cdot p(data_1|\Theta) \cdot p(data_2|\Theta) \cdot ...$$

again, calculate weights and resample...

- a=2.5 1.5 a=-1 2.5 a=-1
- a=3.1 0.1 a=-1 2.5 a=-1
- a=-1 5.5 a=-1 2.5 a=-1
- a=4 0.0 a=-1 2.5 a=-1
- a=2.7 1.3 a=1.7 0.1 a=-1
- a=1.7 3.4 a=1.7 0.1 a=1.7
- ...

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Sequential Importance Sampling

This algorithm is very suitable for data that can be obtained sequentially, for example to monitor an industrial process. SIS (usually implemented as a **particle filter**) can also be applied to more general problems.

$$p(\Theta|data) \propto p(\Theta) \cdot p(data_1|\Theta) \cdot p(data_2|\Theta) \cdot \ldots$$

a=2.5	1.5	a=-1	2.5	a=-1
a=3.1	0.1	a=-1	2.5	a=-1
a=-1	5.5	a=-1	2.5	a=-1
a=4	0.0	a=-1	2.5	a=-1
a=2.7	1.3	a=1.7	0.1	a=-1
a=1.7	3.4	a=1.7	0.1	a=1.7
...				

When all the data are used up, the final swarm of particles is a sample from the posterior distribution.

Because of its sequential structure, SIS is often used with time-series data.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Difficulties with SIS

The main problem with SIS is **particle depletion**: At each resampling step, the number of different particles is reduced, and no new particles are created. Because the number of particles is finite, eventually there are many identical particles.

Different solutions have been suggested, usually based on randomly generating new particles at each step that are slightly different from the existing particles but not too different to break the ergodic properties of the sampler. Other methods are being explored – this is an area of active research.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Comparison MCMC and SIS

	MCMC	SIS
Sampling	Chain generates samples one by one	All samples are generated at once
Data	Required from the start	Can be added sequentially
Computational cost	10,000's of iterations	10,000's of particles
Challenges	Convergence and mixing	Particle depletion
Uses	Very versatile	"Live" time-series

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Implementations

For **MCMC**, many ready-made implementations exist. A good place to start is the package **OpenBUGS** (ongoing development of **WinBUGS**), which implements the Gibbs and other samplers. With a familiar Windows interface and a very general symbolic language to specify models, **OpenBUGS** can solve most classes of Bayesian models.

R offers several add-on packages with MCMC capabilities, as well as an interface to **OpenBUGS**, called **BRugs**.

In terms of speed and efficiency, it may be best to hand-code the MCMC sampler directly in **Fortran**, **C** or another suitable language.

For SIS, I am not aware of any ready-made packages, but there are ongoing developments.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Hands-on Example

Here I demonstrate the use of OpenBUGS. I've made up this example – but the basic approach carries through to real applications in health economics.

8 RCTs have been carried out to investigate the effectiveness of treatments A and B (observing the number of symptom-free patients after 1 year). Treatment A costs SEK 10,000, whereas treatment B costs SEK 14,000. QALY values are given by a probability distribution.

The trial data is summarised as follows:

n^A	120	15	84	398	80	40	97	121
r^A	65	9	39	202	45	17	48	63
n^B	120	16	45	402	77	20	100	115
r^B	81	15	29	270	52	12	68	80

QALY symptom-free: Beta(9,1) QALY with symptoms: Beta(5,5)

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Statistical model

An ***evidence synthesis*** model is required to combine the information from the 8 RCTs.

- Let's choose a **random-baselines, random-effects** model. We model trial outcomes on the log-odds probability scale, with the treatment effect being additive on the log-odds scale.

Letting *i* denote a trial, we have:

Probability with treatment A (baseline): $logit(p_i^A) = \mu_i$

Log-odds treatment effect: t_i

Probability with treatment B: $logit(p_i^B) = \mu_i + t_i$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Statistical model (2)

Random-baselines, random-effects model on the log-odds scale

Probability with treatment A (baseline): $logit(p_i^A) = \mu_i$

Log-odds treatment effect: t_i

Probability with treatment B: $logit(p_i^B) = \mu_i + t_i$

We need to relate the trial-specific parameters μ_i and t_i to their underlying values *M* and *T*. On the log-odds scale, these are usually assumed to be normally distributed.

Random baseline: $\mu_i \sim \text{Norm}(M, \sigma_M^2)$

Random treatment effect: $t_i \sim \text{Norm}(T, \sigma_T^2)$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Statistical model (3)

Next, the model requires a sampling distribution: Given a set of values for the unknown parameters, how likely is an observed datum?

- The model yields a probability and we have binomial data, so the only sensible choice is

$$r_i^A \sim \text{Binom}(p_i^A, n_i^A) \qquad r_i^B \sim \text{Binom}(p_i^B, n_i^B)$$

By now we have 20 unknown parameters (8 t_i, 8 μ_i, M, σ_M, T and σ_T).

So far we have made arbitrary choices in model design – we could just as well have chosen a fixed-effects model (with fewer parameters) or designed something more complicated.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Statistical model (4)

Now, because this is a Bayesian model, we need priors for the unknown parameters M, σ_M, T and σ_T.

- If we already know something about the parameters we could add this knowledge as prior information.

For example, there may be further information on the baseline probability of symptom-free days – we could express this through the prior on M, if we consider the information relevant.

- Otherwise, we choose "sensible" priors that have little information, along the lines of: M and T lie in the real line, and we know little about it, so let's pick a Normal prior with mean 0 and large variance.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Statistical model (5)

Model equations: $logit(p_i^A) = \mu_i \qquad logit(p_i^B) = \mu_i + t_i$

$$\mu_i \sim \text{Norm}(M, \sigma_M^2) \qquad t_i \sim \text{Norm}(T, \sigma_T^2)$$

Sampling distribution: $r_i^A \sim \text{Binom}(p_i^A, n_i^A) \qquad r_i^B \sim \text{Binom}(p_i^B, n_i^B)$

Priors:

$$M \sim \text{Norm}(0, 10000)$$
$$T \sim \text{Norm}(0, 10000)$$

(log-odds probabilities)

$$\sigma_M \sim \text{Unif}(0, 2)$$
$$\sigma_T \sim \text{Unif}(0, 2)$$

(log-odds standard deviations)

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Statistical model (6)

Model equations: $logit(p_i^A) = \mu_i \qquad logit(p_i^B) = \mu_i + t_i$

$$\mu_i \sim \text{Norm}(M, \sigma_M^2) \qquad t_i \sim \text{Norm}(T, \sigma_T^2)$$

Sampling distribution: $r_i^A \sim \text{Binom}(p_i^A, n_i^A) \qquad r_i^B \sim \text{Binom}(p_i^B, n_i^B)$

The model equations and the sampling distribution are common to the Frequentist and the Bayesian approaches. If you already have a Frequentist model, then you (should) already have specified these.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Statistical model (7)

Priors do not occur in the Frequentist setting, so you probably have to make them up.
In this example, the priors are meant to be ***uninformative***, i.e. they are supposed to add no information to the result. It is good practice to test this by changing the priors a little bit and observing the impact on the results of your model.

Priors:

$$M \sim \text{Norm}(0,10000) \qquad \text{(log-odds probabilities)}$$
$$T \sim \text{Norm}(0,10000)$$

$$\sigma_M \sim \text{Unif}(0,2) \qquad \text{(log-odds standard deviation)}$$
$$\sigma_T \sim \text{Unif}(0,2)$$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

OpenBUGS

Let us fit this Bayesian model using OpenBUGS.

The OpenBUGS syntax is relatively straightforward and similar to R.

```
model {
    for (i in 1:N) {
        logit(pA[i])<-mu[i]
        logit(pB[i])<-mu[i]+t[i]
        rA[i]~dbin(pA[i],nA[i])
        rB[i]~dbin(pB[i],nB[i])
        mu[i]~dnorm(M,precM)
        t[i] ~dnorm(T,precT)
    }
    M~dnorm(0,0.0001)
    T~dnorm(0,0.0001)
    precM<-1/pow(sigmaM,2)
    precT<-1/pow(sigmaT,2)
    sigmaM~dunif(0,2)
    sigmaT~dunif(0,2)
}
```

$$logit(p_i^A) = \mu_i$$
$$logit(p_i^B) = \mu_i + t_i$$
$$r_i^A \sim \text{Binom}(p_i^A, n_i^A)$$
$$r_i^B \sim \text{Binom}(p_i^B, n_i^B)$$
$$\mu_i \sim \text{Norm}(M, \sigma_M^2)$$
$$t_i \sim \text{Norm}(T, \sigma_T^2)$$

Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE Centre For Health Economics

OpenBUGS (2)

The data are specified in a separate section so that they can be entered or changed easily.

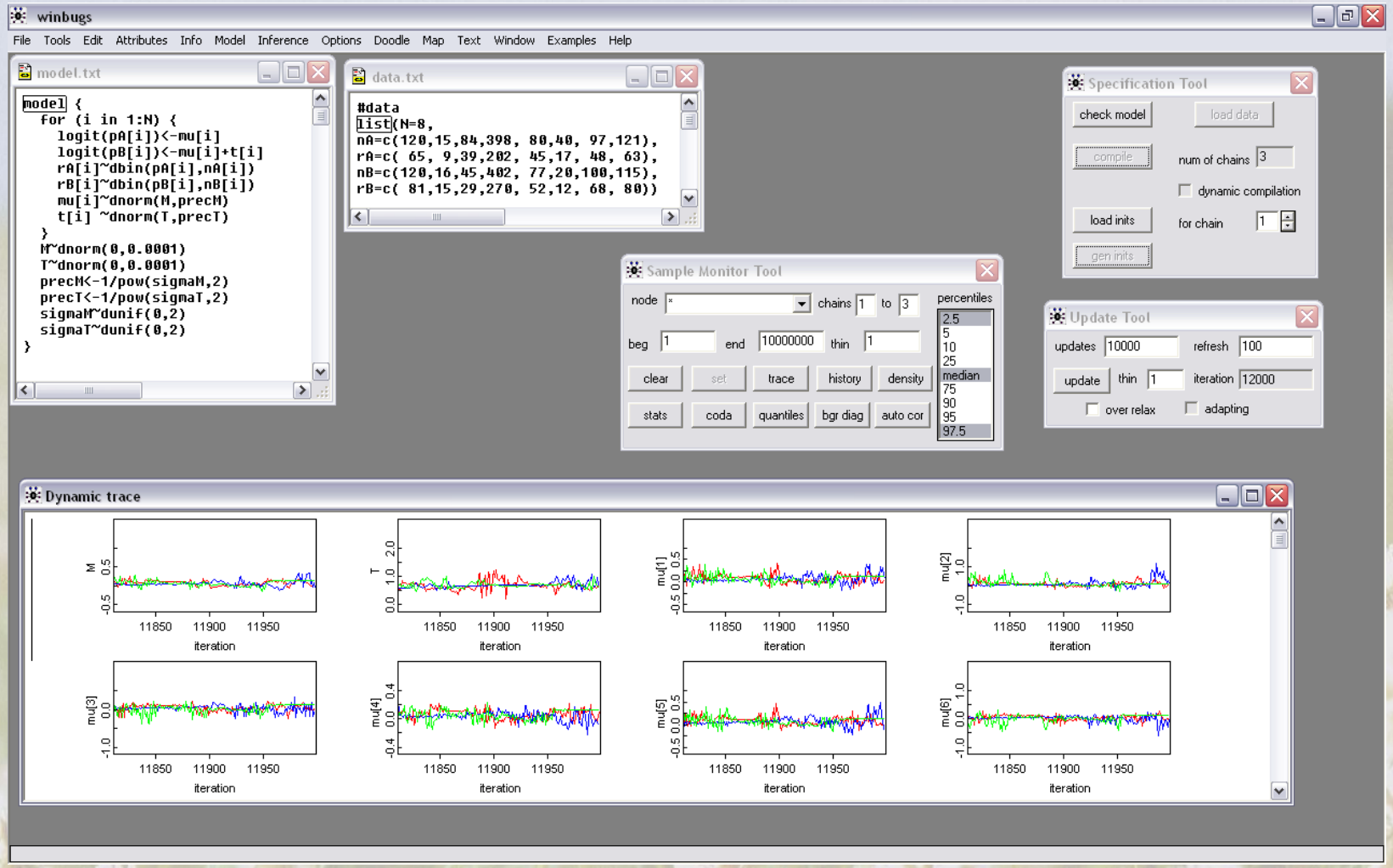
```
model {
    for (i in 1:N) {
        logit(pA[i])<-mu[i]
        logit(pB[i])<-mu[i]+t[i]
        rA[i]~dbin(pA[i],nA[i])
        rB[i]~dbin(pB[i],nB[i])
        mu[i]~dnorm(M,precM)
        t[i] ~dnorm(
    }
    M~dnorm(0,0.000
    T~dnorm(0,0.000
    precM<-1/pow(si
    precT<-1/pow(si
    sigmaM~dunif(0,
    sigmaT~dunif(0,2)
}
```

```
#data
list(N=8,
nA=c(120,15,84,398, 80,40, 97,121),
rA=c( 65, 9,39,202, 45,17, 48, 63),
nB=c(120,16,45,402, 77,20,100,115),
rB=c( 81,15,29,270, 52,12, 68, 80))
```


Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

OpenBUGS (3)

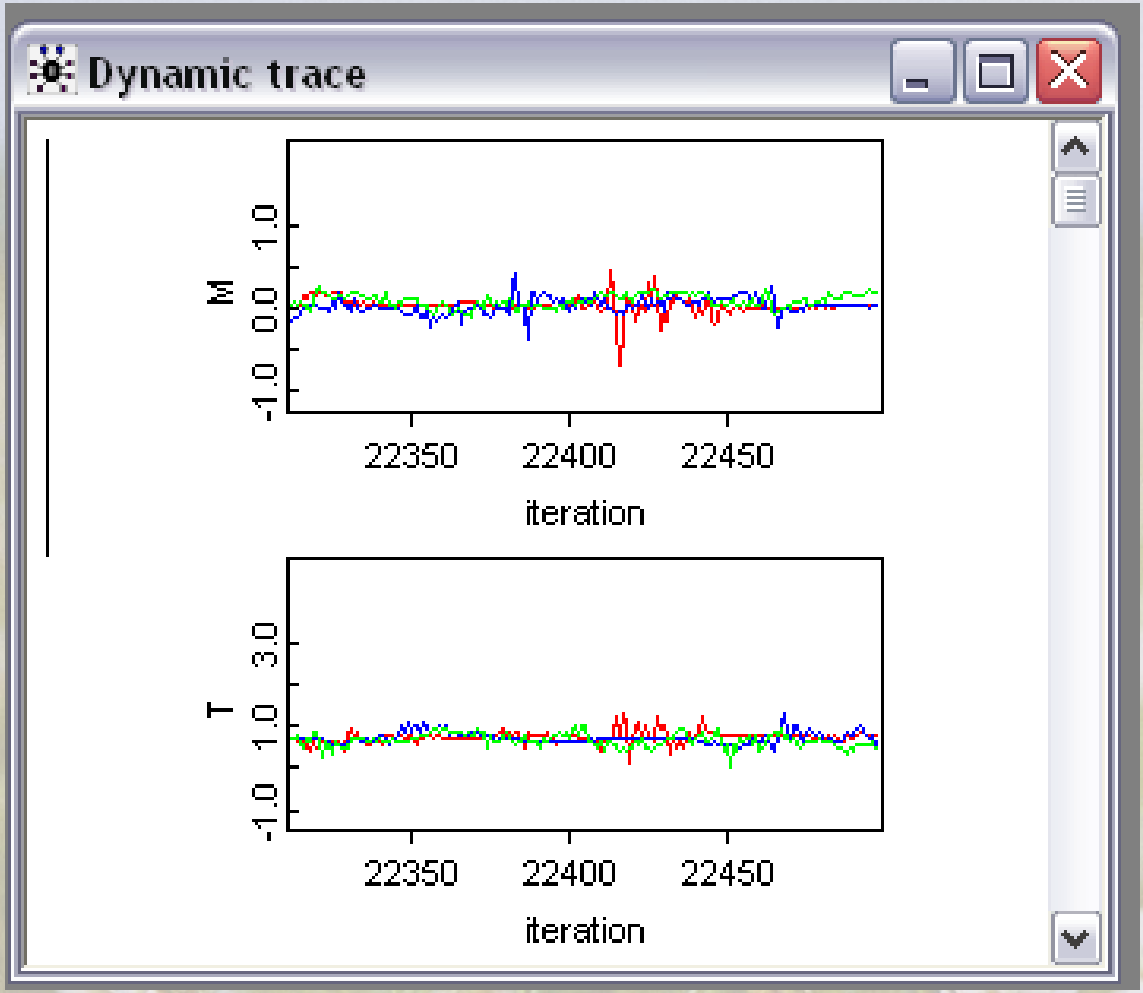
The OpenBUGS window can look like this.



Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE
Centre For Health Economics

OpenBUGS (4)

In this example, OpenBUGS explores the model's posterior reasonably well.



The three colours denote three chains that are run in parallel.

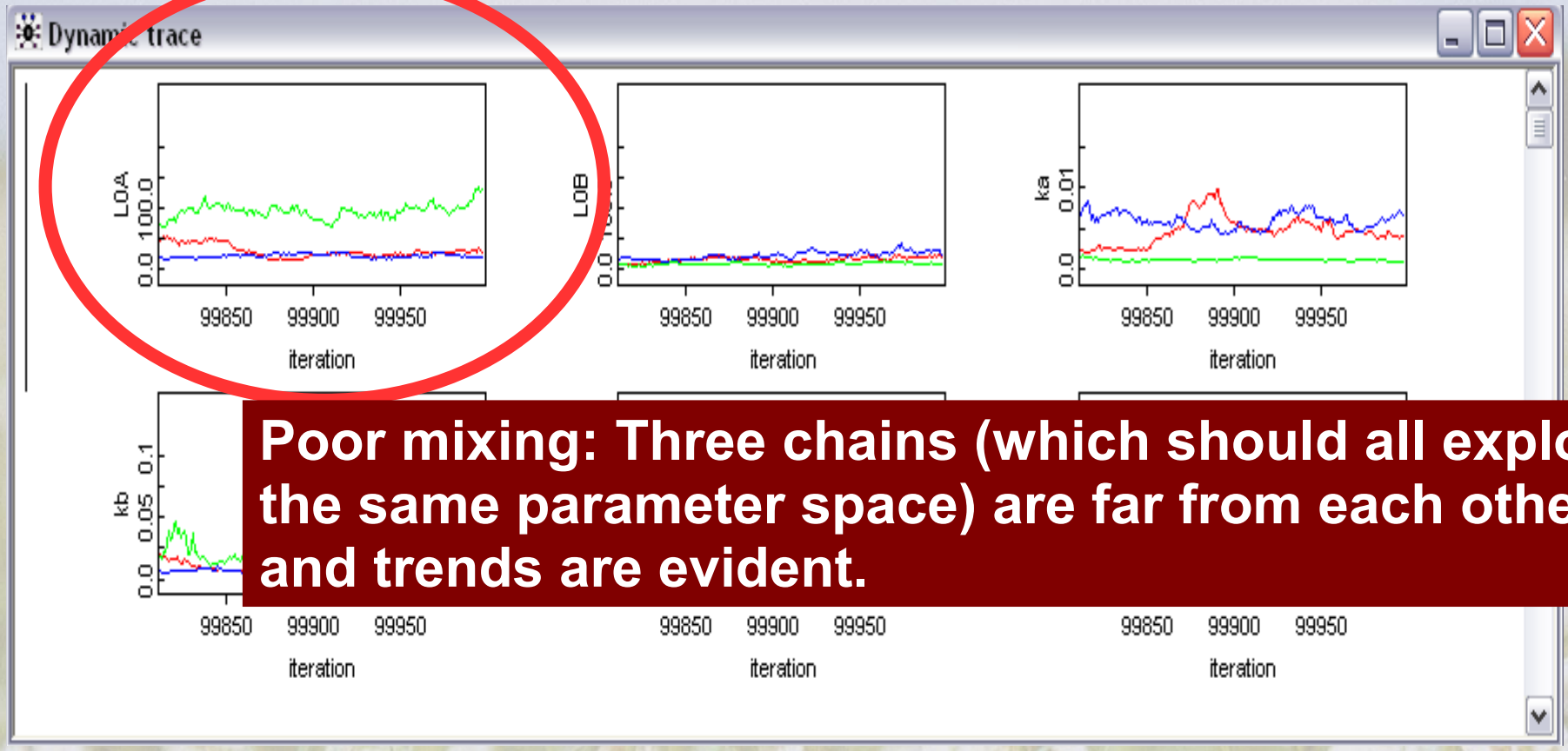
Note that there is no evidence that initial values are influencing the chains.

Also, each chain appears to "wiggle" quite well and the three chains overlap, indicating that they are exploring the same posterior space (as they should).

Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE Centre For Health Economics

OpenBUGS (5)

Here's a screenshot from another model, in which the sampler did not converge (just to give you an idea of what to look for...).



Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE
Centre For Health Economics

Model convergence

WinBUGS and OpenBUGS provide a few formal diagnostics to check for convergence and performance of the sampler, for example the ***Brooks-Gelman-Rubin diagram*** and plots of ***within-chain autocorrelation***. High "***MC_error***" can also indicate convergence problems.

To avoid convergence problems, bear in mind the following.

1. It is difficult to fit many unknown parameters to little data.

2. If you can exploit conjugate distributions, do so.

3. Sometimes you can get around convergence problems by re-writing your model equations without changing the underlying model.

4. If high within-chain autocorrelation is the only problem, you can ***thin*** the posterior samples and only keep every nth draw.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Example continued

When you are satisfied with the posterior sampling, you can generate any desired summary statistic for your posterior.

For example, here's an overview of the posterior means and credibility intervals.

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
M	0.04836	0.08293	0.002087	-0.1076	0.04576	0.2129	10001	30000
T	0.7127	0.1249	0.00317	0.481	0.7105	0.9601	10001	30000
sigmaM	0.09838	0.09447	0.002391	0.004115	0.07186	0.3512	10001	30000
sigmaT	0.135	0.1373	0.003368	0.005719	0.09571	0.4989	10001	30000

In this example, the treatment effect *T* is positive (on the log-odds scale), i.e. the treatment B has a higher probability of symptom-free.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Example continued

When you are satisfied with the posterior sampling, you can generate any desired summary statistic for your posterior.

Node statistics

	mean	sd	MC_error	val2.5pc	median
M	0.04836	0.08293	0.002087	-0.1076	0.0457
T	0.7127	0.1249	0.00317	0.481	0.7105
sigmaM	0.09838	0.09447	0.002391	0.004115	0.0718
sigmaT	0.135	0.1373	0.003368	0.005719	0.0957

MC_error is another indication for how well the sampler performed.

The sampling error should be much smaller than the estimated posterior standard deviation.

Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE Centre For Health Economics

Example continued

When you are satisfied with the posterior sampling, you can generate any desired summary statistic for your posterior.

The 95% credibility interval for the treatment efficacy parameter *T* is [0.481, 0.9601], i.e. treatment B has a very significant effect on the probability of symptom-freeness after 1 year.

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
M	0.04836	0.08293	0.002087	-0.1076	0.04576	0.2129	10001	30000
T	0.7127	0.1249	0.00317	0.481	0.7105	0.9601	10001	30000
sigmaM	0.09838	0.09447	0.002391	0.004115	0.07139	0.3512	10001	30000
sigmaT	0.135	0.1373	0.003368	0.005719	0.09571	0.4989	10001	30000

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Example continued

But what is the probability *P* that treatment B is the cost-effective choice at a willingness-to-pay (WTP) of λ=SEK 50,000?

This probability *P* can be calculated from the model parameters as follows.

Let's assume that the underlying baseline and treatment effect would apply to the target population, i.e. in the target population $logit(p^A)=M$ and $logit(p^B)=M+T$. We calculate the net benefit of the treatments (NB), using the costs C and utilities U.

$$NB^A=[p^A\cdot U_{\text{free}}+(1-p^A)\cdot U_{\text{symptoms}}]\cdot\lambda-C^A$$

$$NB^B=[p^B\cdot U_{\text{free}}+(1-p^B)\cdot U_{\text{symptoms}}]\cdot\lambda-C^B$$

The probability *P* is given by $P=\Pr(NB^B>NB^A)$

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Example continued

But what is the probability *P* that treatment B is the cost-effective choice at a willingness-to-pay (WTP) of λ=SEK 50,000?

We can calculate this directly in WinBUGS, adding a few more lines of code.

```
model {
    ...
    logit(PA)<-M
    logit(PB)<-M+T
    Uf~dbeta(9,1)
    Us~dbeta(5,5)
    NBA<-(PA*Uf+(1-PA)*Us)*WTP-CA
    NBB<-(PB*Uf+(1-PB)*Us)*WTP-CB
    P<-step(NBB-NBA)
}
```

Numerically, we simply look at all the draws from the posterior and check which of them fulfill the condition. This proportion is the posterior probability *P*. There is no need for any further tests.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Example continued

But what is the probability *P* that treatment B is the cost-effective choice at a willingness-to-pay (WTP) of λ=SEK 50,000?

Here's the posterior summary for our new quantities.

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
NBA	25210.0	4364.0	27.08	16240.0	25330.0	33200.0	10002	30000
NBB	24620.0	3942.0	25.09	15580.0	25040.0	31090.0	10002	30000
P	0.3498	0.4769	0.00473	0.0	0.0	1.0	10002	30000

There is a lot of overlap between the net benefits for treatment A (95%-CI [SEK 16240, 33200]) and B (95%-CI [SEK 15580, 31090]).

Accordingly, the probability that treatment B is cost-effective is estimated at 0.3498.

Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE Centre For Health Economics

Example continued

But what is the probability *P* that treatment B is the cost-effective choice at a willingness-to-pay (WTP) of λ=SEK 50,000?

Here's the posterior summary for our new quantities.

Node statistics

	mean	sd	MC_error	val2.5pc	median	val97.5pc	start	sample
NBA	25210.0	4364.0	27.08	16240.0	25330.0	33200.0	10002	30000
NBB	24620.0	3942.0	25.09	15580.0	25040.0	31090.0	10002	30000
P	0.3498	0.4769	0.00473	0.0	0.0	1.0	10002	30000

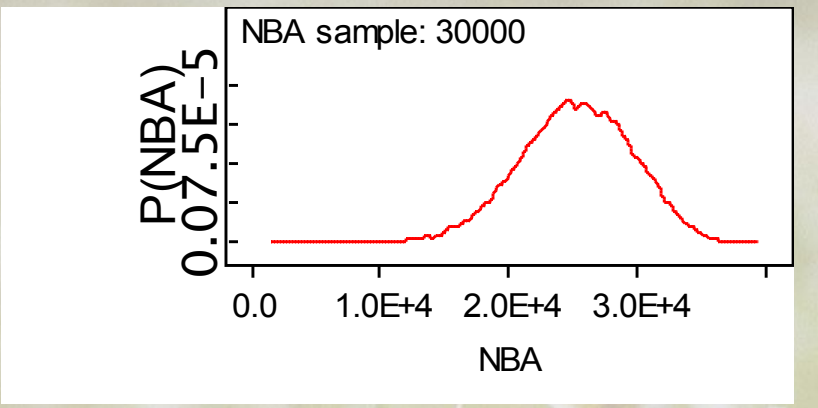
In this example, I defined the probability *P* by checking a condition, rather than as a stochastic variable with its own prior and probability density. This is why the confidence intervals for *P* do not mean anything.

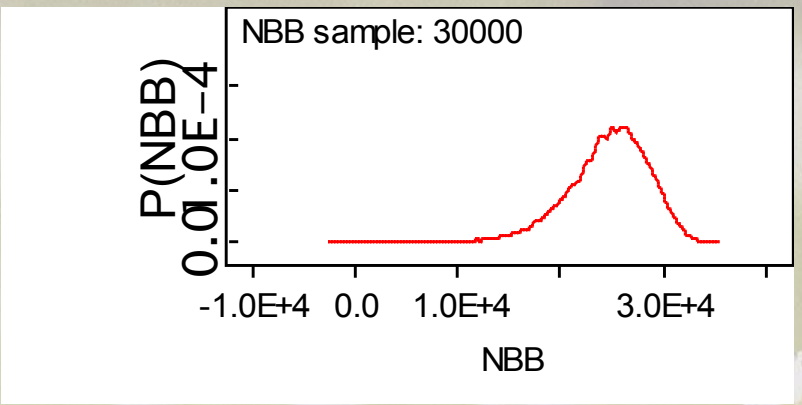
Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE Centre For Health Economics

Example continued

OpenBUGS can also produce graphical output for all quantities of interest.

For example, here are the posterior densities for the net benefits NB^A and NB^B.





They both show quite wide distributions and their support on the *x*-axes overlaps substantially.

Dr Christian Asseburg
University of York, UK
ca505@york.ac.uk
CHE Centre For Health Economics

Bayesian decision theory

With a probabilistic net benefit function, Bayesian decision theory can be applied to optimise management decisions.

Bayesian models can thus directly feed in to management processes.

In a Frequentist model, it is not generally possible to find a probability distribution for an unknown parameter – because a Frequentist calculates the likelihood of observing the data and uses this to make inferences on the model parameters.

Frequentist models do not offer an obvious way for calculating quantities that are derived from individual parameter values (such as the probability *P* or a net benefit) – this makes them less amenable to management processes.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics

Summary

- Frequentist modelling centres on the likelihood function, i.e. how likely are the data given a particular model.
- Bayesian modelling centres on the probabilities of models and model parameters, by combining the likelihood of the data with prior probabilities of the unknown parameters.
- Both can be used equally well to fit models and to make inferences on model parameters.
- However, only Bayesian statistics is capable of assigning probabilities to model quantities. This makes it possible to calculate derived quantities and their uncertainties.
- Numerical methods for fitting Bayesian models require some care and experience.

Dr Christian Asseburg
University of York, UK ca505@york.ac.uk
CHE Centre For Health Economics