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

Dr. Christian Asseburg

Centre for Health Economics
University of York, UK

cas05@york.ac.uk

Part 3

Model Selection
Model Selection

- Model selection
  - Bayesian model selection
  - Information criteria: BIC, DIC, ...
- Checking model predictions
- Which model should you choose?
Model selection

- There is no “correct” model. All models are just an approximation.
- It is unknown how likely a model $i$ is. From a Bayesian point of view, this is an unknown quantity and we could model it.

$$P(M_i|data)$$

- Because this is an unknown Bayesian quantity, we also need a prior for it.

$$P(M_i)$$
Model selection

- We can use Bayes' Theorem to calculate the posterior model probability $m_i$.

$$m_i = P(M_i | data) = \frac{P(data | M_i) P(M_i)}{P(data)}$$

If the parameters in model $M_i$ are called $\Theta_i$, we can expand this to:

$$m_i = \frac{\int P(\Theta_i | M_i) P(data | \Theta_i, M_i) d \Theta_i \cdot P(M_i)}{P(data)}$$
Bayes factors

- We may want to compare the (posterior) model probabilities for two models $M_1$ and $M_2$.

\[
\frac{m_1}{m_2} = \frac{P(M_1|\text{data})}{P(M_2|\text{data})} = \frac{P(\text{data}|M_1)P(M_1)}{P(\text{data}|M_2)P(M_2)}
\]

If we assume that a priori both models are equally likely, we end up with the **Bayes factor**:

\[
B_{12} = \frac{m_1}{m_2} = \frac{P(\text{data}|M_1)}{P(\text{data}|M_2)} = \int P(\Theta_1|M_1)P(\text{data} | \Theta_1, M_1) d\Theta_1 \frac{1}{\int P(\Theta_2|M_2)P(\text{data} | \Theta_2, M_2) d\Theta_2}
\]
Bayes factors

How to interpret the value of a Bayes factor:

<table>
<thead>
<tr>
<th>Bayes Factor</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$B_{ij} &lt; 1/10$</td>
<td>Strong Evidence For $M_j$</td>
</tr>
<tr>
<td>$1/10 &lt; B_{ij} &lt; 1/3$</td>
<td>Moderate Evidence For $M_j$</td>
</tr>
<tr>
<td>$1/3 &lt; B_{ij} &lt; 1$</td>
<td>Weak Evidence For $M_j$</td>
</tr>
<tr>
<td>$1 &lt; B_{ij} &lt; 3$</td>
<td>Weak Evidence For $M_i$</td>
</tr>
<tr>
<td>$3 &lt; B_{ij} &lt; 10$</td>
<td>Moderate Evidence For $M_i$</td>
</tr>
<tr>
<td>$B_{ij} &gt; 10$</td>
<td>Strong Evidence For $M_i$</td>
</tr>
</tbody>
</table>

Jeffreys’ Scale of Evidence For Bayes Factors.

$$B_{12} = \frac{m_1}{m_2} = \frac{P(data|M_1)}{P(data|M_2)} = \frac{\int P(\Theta_1|M_1)P(data|\Theta_1,M_1)d\Theta_1}{\int P(\Theta_2|M_2)P(data|\Theta_2,M_2)d\Theta_2}$$
How to compare models

In suitably simple models, you can calculate the integrals in the Bayes factor analytically. However, usually you cannot do that.

There are two methods for comparing models in OpenBUGS.

\[
B_{12} = \frac{m_1}{m_2} = \frac{P\left(\text{data} \mid M_1\right)}{P\left(\text{data} \mid M_2\right)} = \frac{\int P\left(\Theta_1 \mid M_1\right) P\left(\text{data} \mid \Theta_1, M_1\right) d\Theta_1}{\int P\left(\Theta_2 \mid M_2\right) P\left(\text{data} \mid \Theta_2, M_2\right) d\Theta_2}
\]
How to compare models

Method 1: Rewrite your model to include the model indicator $m$. Then you can look at the probabilities $m_i = P(m = i)$ and calculate $B_{ij}$.

Method 2: Fit each model separately and use the DIC summary value.

$$B_{12} = \frac{m_1}{m_2} = \frac{P(data|M_1)}{P(data|M_2)} = \frac{\int P(\Theta_1|M_1)P(data|\Theta_1,M_1)\,d\Theta_1}{\int P(\Theta_2|M_2)P(data|\Theta_2,M_2)\,d\Theta_2}$$
Computing Bayes factors

**Method 1:** Rewrite your model to include the model indicator $m$. Then you can look at the probabilities $m_i = P(m=i)$ and calculate $B_{ij}$.

Here's a simple example. You observe a coin and it falls on “Heads” 20 out of 50 times.

Model 1 is a simple binomial model.

```r
# model 1
model {
  a ~ dbeta(1,1)
  heads ~ dbin(a, n)
}
```
Computing Bayes factors

**Method 1:** Rewrite your model to include the model indicator \( m \). Then you can look at the probabilities \( m_i = P(m = i) \) and calculate \( B_{ij} \).

Here's a simple example. You observe a coin and it falls on “Heads” 20 out of 50 times.

To keep things really simple, model 2 is also a simple binomial model, only the prior is different.

```r
# model 1
model {
  a ~ dbeta(1,1)
  heads ~ dbin(a, n)
}

# model 2
model {
  b ~ dbeta(10,10)
  heads ~ dbin(b, n)
}
```
Computing Bayes factors

**Method 1:** Rewrite your model to include the model indicator \( m \). Then you can look at the probabilities \( m_i = P(m=i) \) and calculate \( B_{ij} \).

Now we combine the two models into one model and add an indicator \( m \) (\( m=1 \) or \( m=2 \)) that picks the model to use.

I use a non-informative prior on \( m \).

```r
model {
# model 1
p[1] ~ dbeta(1,1)
# model 2
p[2] ~ dbeta(10,10)
# m
heads ~ dbin(p[m], n)
m ~ dcat(mprior[1])
mprior[1] <- 0.5
mprior[2] <- 0.5
}
```
Computing Bayes factors

Method 1: Rewrite your model to include the model indicator $m$. Then you can look at the probabilities $m_i = P(m=i)$ and calculate $B_{ij}$.

Running this model in OpenBUGS, the posterior for $m$ is:

```r
model {
   # model 1
   p[1] ~ dbeta(1,1)
   # model 2
   p[2] ~ dbeta(10,10)
   # m
   heads ~ dbin(p[m], n)
   m ~ dcat(mprior[])  
   mprior[1] <- 0.5  
   mprior[2] <- 0.5
}
```
Computing Bayes factors

Method 1: Rewrite your model to include the model indicator $m$. Then you can look at the probabilities $m_i = P(m=i)$ and calculate $B_{ij}$.

Running this model in OpenBUGS, the posterior for $m$ is:

We can conclude that model 2 is about twice as likely as model 1. To calculate the Bayes factor, simply divide $m_1$ by $m_2$. 

![Poster for $m$ sample: 30000]
Model comparison using DIC

**Method 2:** Fit each model separately and use the DIC summary value.

The DIC (Deviance Information Criterion) is defined in analogy with the AIC (Akaike's Information Criterion). So models with small DIC provide a better fit.

\[
B_{12} = \frac{m_1}{m_2} = \frac{P(data|M_1)}{P(data|M_2)} = \frac{\int P(\Theta_1|M_1) P(data|\Theta_1, M_1) d\Theta_1}{\int P(\Theta_2|M_2) P(data|\Theta_2, M_2) d\Theta_2}
\]
Model comparison using DIC

The DIC (Deviance Information Criterion) is defined in analogy with the AIC (Akaike's Information Criterion).

The DIC is unrelated to the posterior model probabilities that we looked at before. Instead, it gives a measure for how well each model fits the data and penalises for the number of parameters, similar to the AIC.

Do not confuse this with the BIC (Bayes Information Criterion), which can be calculated from the Bayes factors.
Model comparison using DIC

You can use OpenBUGS to estimate the DIC.

The approximation built into OpenBUGS requires that
- The posterior is approximately multivariate normal, and
- The effective number of parameters can be calculated.

Notes:
- In a hierarchical model it can be difficult to determine the “effective number of parameters”.
- If your model contains discrete parameters, the posteriors cannot be normally distributed and you cannot use the hard-coded approximation for DIC.
Model comparison using DIC

To calculate the DIC in OpenBUGS, you add a DIC monitor (similar to adding monitors for all the other quantities of interest). Add the DIC monitor after the burn-in! The command is in the “Inference” menu.
Model comparison using DIC

- Note: There's a bug in the OpenBUGS program. If you close the “DIC tool” window, you can open it again but the buttons are greyed out. Do not close the “DIC tool” window.

- After running the posterior sampler for the final iterations, click on “Stats” in the “DIC tool”.

<table>
<thead>
<tr>
<th></th>
<th>Dbar</th>
<th>Dhat</th>
<th>DIC</th>
<th>pD</th>
</tr>
</thead>
<tbody>
<tr>
<td>rA</td>
<td>43.89</td>
<td>41.69</td>
<td>46.09</td>
<td>2.197</td>
</tr>
<tr>
<td>rB</td>
<td>44.39</td>
<td>41.34</td>
<td>47.45</td>
<td>3.055</td>
</tr>
<tr>
<td>total</td>
<td>88.28</td>
<td>83.03</td>
<td>93.53</td>
<td>5.252</td>
</tr>
</tbody>
</table>
Model comparison using DIC

- The relevant DIC number is in the DIC column and in the “total” row.
- To make sense of it, you must compare it to the DIC estimated for a different model.

<table>
<thead>
<tr>
<th></th>
<th>Dbar</th>
<th>Dhat</th>
<th>DIC</th>
<th>pD</th>
</tr>
</thead>
<tbody>
<tr>
<td>rA</td>
<td>43.89</td>
<td>41.69</td>
<td>46.09</td>
<td>2.197</td>
</tr>
<tr>
<td>rB</td>
<td>44.39</td>
<td>41.34</td>
<td>47.45</td>
<td>3.055</td>
</tr>
<tr>
<td>total</td>
<td>88.28</td>
<td>83.03</td>
<td>93.53</td>
<td>5.252</td>
</tr>
</tbody>
</table>

pD is an estimate of the number of parameters in your model.
Model comparison using DIC

- Suppose you have two models, model 1 has a DIC of 93.53 and model 2 has a DIC of 84.02.
- With the DIC you can only compare models that use the same data. A lower DIC value is better.

<table>
<thead>
<tr>
<th></th>
<th>Dbar</th>
<th>Dhat</th>
<th>DIC</th>
<th>pD</th>
</tr>
</thead>
<tbody>
<tr>
<td>rA</td>
<td>43.89</td>
<td>41.69</td>
<td>46.09</td>
<td>2.197</td>
</tr>
<tr>
<td>rB</td>
<td>44.39</td>
<td>41.34</td>
<td>47.45</td>
<td>3.055</td>
</tr>
<tr>
<td>total</td>
<td>88.28</td>
<td>83.03</td>
<td>93.53</td>
<td>5.252</td>
</tr>
</tbody>
</table>
Connection DIC and Bayes Factors

- There is no theoretical connection between DIC and Bayes Factors... They have been developed with different aims.
- So if you choose a model based on Bayes factors (this is Method 1 we saw earlier) or based on DIC, you may end up with different models!
- If you know the BIC values for two models, you can calculate the Bayes factor:

\[ B_{ij} = e^{BIC_i - BIC_j} \]

- I have seen people use the DIC values like this, but there is no theoretical justification for doing so.
Note about priors

- Note that priors are essential parts of Bayesian models. It is not just the models that cause a good model fit, but also the priors.

- The same model with different priors will result in a different posterior model probability (see first example in this presentation).

- If you use “non-informative” priors you should check that they do not influence the model comparisons unduly.
Checking model predictions

- Like with any models, it is a good idea to check the predictions of Bayesian models too.

- For example, if you have spare data, you can test whether your model predicts it appropriately. You could compare models according to how well they predict the spare data.
Checking model predictions

- In health economics there is usually a shortage of data, so this may not be feasible.

- However, you can look at the data that goes into the model, too. For example, you can verify the empirical quantiles of each of your data points and identify outliers.

- If you find many outliers or high variances in the predictive distributions, you may want to revise your model.
Model selection

• From a Bayesian point of view, we can calculate the (posterior) probability that a model is “correct”. Even models with low, non-zero probability are sometimes correct, so we should not discard them.

• Instead, to make Bayesian predictions, you average the predictions from all models, using the (posterior) model probabilities as weights.
Summary

- To do a proper Bayesian model selection, calculate the (posterior) probabilities for each model, $m_i$.

- This calculation is often too complicated, so the DIC has been developed in analogy with the AIC.

- You could select a model with a low DIC, but this is not the same as Bayesian model selection, and it does not give weights for the model averaging.