Bayesian Analysis

Computer Lab 2 – Fitting models in JAGS with different prior distributions

1 Revision: running JAGS

Remember the steps to run a JAGS analysis:

1. Specify the model in a separate file.
2. Load the data *and* known constants (hyperparameters / parameters for the priors) in `R` in a named list.
3. Run `jags.model()` to import the model into `R`.
4. Run `jags.sample()` to generate the samples.
5. Analyse the output!

When JAGS runs, it produces *samples* from the posterior distribution. We will cover more about how these samples are created in lectures, but for now we can use them to explore the posterior distributions of various models and data sets we might like to fit.

> Task:
>
> - Open the `dating` example again and run it for 10,000 iterations. Look at the sample history, density and summary stats.

2 Files with more than one observation

It is relatively simple to create JAGS code to run models with multiple observations. We simply need to include a loop. In the `chest.model` example (taken from the chest measurements example in Lecture 2), we specify the likelihood using a `for` statement:

```
for(i in 1:N)
{
x[i] ~dnorm(theta,phi.inv)
}
```

This looping statement says that, for each different observation (up to observation `N`), each data point comes from a normal distribution with mean `theta` and precision `phi.inv`. To complete the model we need to give a value for `N` and a list of the data (comma-seperated and surrounded by `c(...)`). These parts can be found in the `chest_data_hypers.R` file.

> Task:
>
> - Run the `chest` example for 10,000 iterations, and confirm the posterior distribution of θ as approximately $N(\bar{x}, \phi/n)$. (note: $\bar{x}$ here is 39.8)

3 Reference and flat prior distributions

JAGS allows for lots of different probability distributions, including the normal, uniform, exponential, gamma, Poisson and Binomial. However it does **not** allow us to specify an improper prior distribution. Three of the closest / most useful options are:

- `dunif(-100,100)` for a truly flat prior, within some set limits. You may need to adjust the limits.
- `dnorm(0,1e-6)` for an almost flat prior, for a parameter that may be positive or negative.
- `dgamma(0.001,0.001)` for an almost flat prior for positive valued parameters. This is a gamma distribution with shape and rate as very small values. Recall that the gamma distribution has pdf: $p(x|\alpha,\beta) \propto x^{\alpha-1}e^{-\beta x}$, thus

$$\lim_{\alpha,\beta\downarrow 0} p(x|\alpha,\beta) \propto x^{-1}$$

  Hence this distribution can be used as the Jeffreys' prior on a variance or precision parameter.

3.1 Comparing different prior distributions

A useful trick for comparing models with differing prior distributions is shown in the files `chest2.model` and `chest2.R`. Here, we compare 2 different prior distributions (an informative $N(38,9)$ prior and a flat prior) on the parameter of interest θ. The script works by looping through the two models (with index `j`) and first creating a separate (but equal) dataset for each. This new data set contains the data in a matrix format such that each row of the matrix comes from a normal distribution with a different mean (`theta`). There are now two prior distributions, one for each element of the vector `theta`.

Task:

- Run the `chest2.model` for 10,000 iterations. Show that the flat prior distribution and the informative prior distribution yield very similar posterior values of θ.

4 Multi-parameter models

JAGS can easily handle models with many parameters. In such circumstances, the only change we need to make is to ensure we specify prior distributions and starting values for each unknown parameter. When sampling from the model, be sure to request all the parameters of interest (i.e. second argument to jags.samples variable.names). JAGS will now produce joint samples of all the parameters which you can summarise individually (i.e. provide their marginal posterior sample distributions).

Task:

- Open the file `rats.model` and check that the formulation of the problem corresponds to slides 8 and 9 of lecture 6. Load the data using `rats_data.R`, then run the model and show that the 95% posterior credible intervals for the mean and variance are approximately (18,24) and (20,74) respectively. **Hint:** use the quantile() function in `R`.

5 Homework exercises

Produce a short report (in either word, open office or pdf format) which answers the following questions and includes all used JAGS code and appropriate plots:

1. Write a JAGS model file to fit the Poisson-Gamma misprints example of Lecture 4. Show that the results obtained are identical to those found in class.
2. Create another script to enable comparison for 3 different prior distributions:
   (a) The $Ga(9,6)$ prior used originally
   (b) A $U(0,10)$ prior distribution

(c) The Jeffreys prior for the Poisson distribution (hint: the Jeffreys prior for the Poisson distribution with parameter λ is $p(\lambda) \propto \sqrt{1/\lambda}$. Use an appropriate gamma distribution to approximate this prior)

Show the different posterior distributions under each of these assumptions. Which of the prior distributions is most informative? Why might two of these distributions be similar?

3. Some more misprints are observed on the following pages as 2,1,6, and 0. Use your posterior distribution created in (1) as a prior distribution for this new data. Show that the new posterior distribution is the same as that if the data had all been observed simultaneously.