

Inference Tutorial 9

https://people.maths.bris.ac.uk/~sw15190/TOI/jags_user_manual.pdf is a link to the JAGS manual. The `str` function in R is useful for examining the structure of an R object that you are unsure of. To use JAGS you have to produce a text file containing your model specification. On the university windows computers the editors 'notepad' or 'wordpad' can be used for this ('word' tends to create problems with hidden characters). It is probably best to create the file wherever you usually store your R script files, for example on the network drive accessed by opening the file manager (3rd icon from lower left in windows), and clicking on 'Computer' followed by 'MyFiles'. If you use this location then `setwd("O:")` (that's the letter, not zero) within R will cause R to look for jags (and other) files in this location (If you stored your files in a directory called, e.g., 'foo' within 'Myfiles' then use `setwd("O:/foo")`). Don't forget `library(rjags)`.

1. As a simple example of the use of JAGS, consider the grasshopper survival data and model from tutorials 7 and 8, one more time, but this time allowing the α parameter to take any positive value. Recall that the data are observations of x , the length of time in excess of 14 days that it took for a random sample of 10 grasshoppers infected with a fungal disease to die, where the times are measured from infection.

1.6 12.1 5.2 6.2 6.8 9.6 17.8 24.4 4.1 7.8

A gamma distribution may be a good model here, so that the p.d.f. of x is

$$f(x) = \frac{\beta^\alpha}{(\alpha - 1)!} x^{\alpha-1} e^{-x\beta} \quad x > 0$$

in JAGS the gamma distribution (`dgamma`) is parameterized in terms of α and β , in that order. Use a gamma prior for β (`dgamma(1, .01)`) and a log normal prior for α (`dlnorm(1, 1)`).

- (a) Use JAGS to sample 10000 values from the posterior density of α, β .
- (b) Produce trace plots and ACFs for α and β to check the chain mixing, convergence and correlation.
- (c) Compute the correlation lengths and effective sample sizes for the two chains.
- (d) Produce histograms illustrating the marginal posterior distributions of α and β and a scatter plot illustrating the posterior correlation.
- (e) Find 95% credible intervals for α and β .

Solution

- (a) The JAGS model specification file ('hopper.jags', say) contains

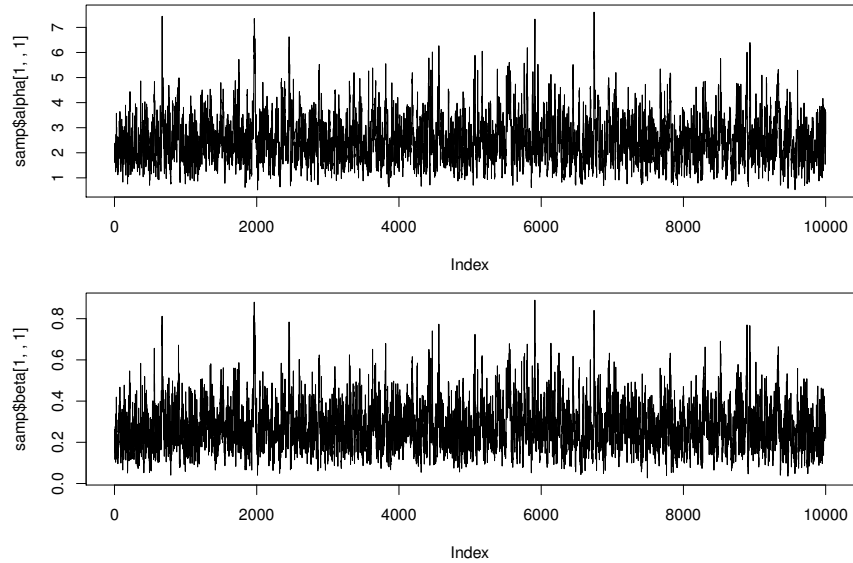
```
model {
  for (i in 1:10) { x[i] ~ dgamma(alpha,beta) }
  alpha ~ dlnorm(1,1)
  beta ~ dgamma(1,.01)
}
```

(or equivalent code). To use this from R requires something like

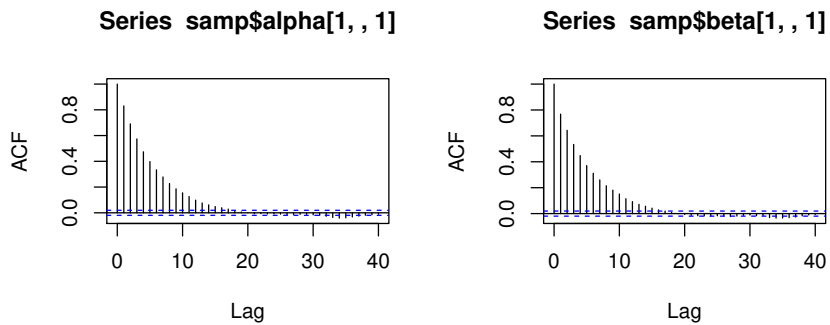
```
x <- c(1.6,12.1,5.2,6.2,6.8,9.6,17.8,24.4,4.1,7.8)
library(rjags)
setwd("where/hopper.jags/is/located") ## replace this with path you are using
dat <- list(x=x)
jmod <- jags.model("hopper.jags",data=dat,init=list(alpha=2,beta=.2),
                  n.adapt=1000)
samp <- jags.samples(jmod,c("alpha","beta"),n.iter=10000)
```

- (b)

```
str(samp) ## check out the structure of the returned object
par(mfrow=c(2,1))
plot(samp$alpha[1,,1],type="l")
plot(samp$beta[1,,1],type="l")
```



```
par(mfrow=c(1,2))
cla <- sum(acf(samp$alpha[1,,1])[[1]][,1])*2 - 1
clb <- sum(acf(samp$beta[1,,1])[[1]][,1])*2 - 1
```



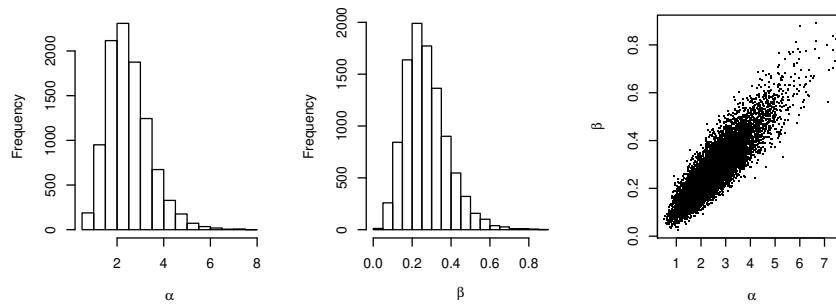
So there is some visible autocorrelation, but the chains are mixing reasonably well, and the 1000 steps for adaptation seems to have been enough to allow convergence.

(c) The correlation lengths were computed above

```
> cla;clb
[1] 9.549154
[1] 8.82942
> neff.a <- 10000/cla
> neff.b <- 10000/clb
> neff.a;neff.b
[1] 1047.213
[1] 1132.577
```

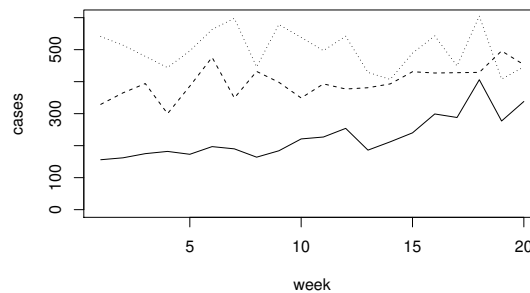
(d) `par(mfrow=c(1,3))`

```
hist(samp$alpha,xlab=expression(alpha),main="")
hist(samp$beta,xlab=expression(beta),main="")
plot(samp$alpha,samp$beta,xlab=expression(alpha),ylab=expression(beta),pch=".")
```



```
(e) > quantile(samp$alpha, c(.025, .975))
      2.5%      97.5%
1.048985 4.648240
> quantile(samp$beta, c(.025, .975))
      2.5%      97.5%
0.09830003 0.53002320
```

2. 20 weeks of new case data for an emergent disease are observed for three neighbouring towns in a remote region. The data are shown here (towns 1, 2 and 3 are shown solid, dashed and dotted, respectively).



The public health department trying to control the disease needs to establish whether the underlying case rate is increasing or decreasing in each town, using the JAGS model specification and R code given at the end of this question. The case data for the 3 towns are stored in vectors Y1, Y2 and Y3 of list dat. Study the code and answer the following questions as concisely as possible.

- Briefly describe the key principles of Gibbs sampling.
- What is the key feature of a directed acyclic graph (DAG), and why is the DAG of a model useful in the automatic construction of Gibbs samplers?
- Write out the model specified in the JAGS code, mathematically.
- What do the r parameters and the α parameters mean in terms of the disease?
- Comment on the final plots from the point of view of chain convergence, and describe two further diagnostics that might be examined.
- Comment on the implications of the results in terms of the disease outbreak, and suggest a further quantitative summary that might be useful to compute.

```
model { ## JAGS model specification
  for (i in 1:M) {
    Y1[i] ~ dpois(mu1[i]+alpha[1]*mu2[i]+alpha[2]*mu3[i])
    d1[i] ~ dnorm(0,tau1)
    mu1[i] <- N1*exp(r1*i+d1[i])
    Y2[i] ~ dpois(mu2[i]+alpha[3]*mu1[i]+alpha[4]*mu3[i])
    d2[i] ~ dnorm(0,tau2)
    mu2[i] <- N2*exp(r2*i+d2[i])
  }
}
```

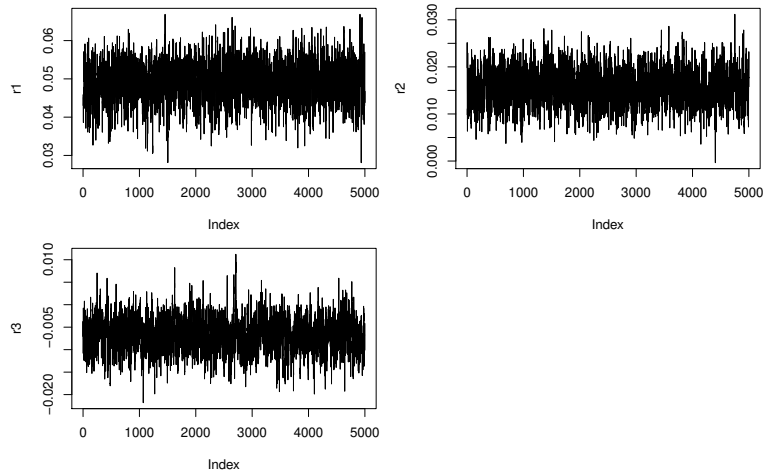
```

Y3[i] ~ dpois(mu3[i]+alpha[5]*mu1[i]+alpha[6]*mu2[i])
d3[i] ~ dnorm(0,tau3)
mu3[i] <- N3*exp(r3*i+d3[i])
}
for (i in 1:6) {
  alpha[i] ~ dunif(0,0.1)
}
tau1 ~ dgamma(1.0,0.1)
r1 ~ dnorm(0,.1)
N1 ~ dnorm(100,0.01)
tau2 ~ dgamma(1.0,0.1)
r2 ~ dnorm(0,.1)
N2 ~ dnorm(300,0.01)
tau3 ~ dgamma(1.0,0.1)
r3 ~ dnorm(0,.1)
N3 ~ dnorm(500,0.01)
} ## end of JAGS code

## R code...
library(rjags)
load("disease.rda") ## load 'dat' list
dat$M <- length(dat$Y1) ## include number of time points for JAGS
mod <- jags.model("jags.jags",data=dat)
um <- jags.samples(mod,c("r1","r2","r3"),n.iter=50000,thin=10)

par(mfrow=c(2,2),mar=c(4,4,1,1))
plot(um$r1,type="l",ylab="r1")
plot(um$r2,type="l",ylab="r2")
plot(um$r3,type="l",ylab="r3")

```



Solution

1. Gibbs sampling cycles through the simulation variables, simulating a draw from the conditional distribution of the current variable, given the most recently simulated values of all the rest. In this way we end up simulating from the joint density of all the variables.
2. The graph edges have directions. Starting from any node there is no path through the graph that follows the directed edges to arrive back at the starting node. The DAG encodes the conditional independence structure of the model, so that once we have the DAG we can immediately identify which other nodes play a part in simulating from the conditional distribution of any node of interest.
3. The structure for explaining the data is...

$$Y_1(t) \sim \text{Poi}\{\mu_1(t) + \alpha_1\mu_2(t) + \alpha_2\mu_3(t)\}, \quad \mu_1(t) = N_1 \exp\{r_1 t + d_1(t)\}, \quad d_1(t) \sim N(0, 1/\tau_1) \quad Y_2(t) \sim \text{Poi}\{\mu_2(t) +$$

$\alpha_3\mu_1(t) + \alpha_4\mu_3(t)\}$, $\mu_2(t) = N_2 \exp\{r_2t + d_2(t)\}$, $d_2(t) \sim N(0, 1/\tau_2)$ $Y_3(t) \sim \text{Poi}\{\mu_3(t) + \alpha_5\mu_1(t) + \alpha_6\mu_2(t)\}$, $\mu_3(t) = N_3 \exp\{r_3t + d_3(t)\}$, $d_3(t) \sim N(0, 1/\tau_3)$
 and the priors are $\alpha_j \sim U(0, 0.1)$, $\tau_j \sim \Gamma(1, 0.1)$, $r_1 \sim N(0, 10)$, $N_1 \sim N(100, 100)$, $N_2 \sim (300, 100)$, $N_3 \sim N(500, 100)$.

4. r_j is the underlying case rate growth rate in town j . The α parameters control how many of the cases seen in a particular town are related to the rate in a neighbouring town (i.e they relate to travel related cases).
5. Convergence looks reasonable here, but it would be sensible to compute effective sample sizes and to look at the ACFs, and to do the same for the other model components.
6. The number of cases is clearly growing in towns 1 and 2. For town 3 the posterior mode growth rate is negative, but it would be useful to compute a credible interval to see just how confident we could be about a conclusion of underlying rate decline in this town.