

## Inference Tutorial 10

[https://people.maths.bris.ac.uk/~sw15190/TOI/jags\\_user\\_manual.pdf](https://people.maths.bris.ac.uk/~sw15190/TOI/jags_user_manual.pdf) links 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)`.

The data loaded with the command

```
h1n1 <- read.table("https://people.maths.bris.ac.uk/~sw15190/TOI/sh-H1N1.dat")
```

give the number of laboratory confirmed cases of H1N1 influenza in the Southern Hemisphere each week starting in the 17th week of 2009. The data cover the period of the southern hemisphere winter and are from the WHO. It may be reasonable to assume that the expected number of lab confirmed cases is proportional to the actual number of cases each week, but of course most cases will not be lab tested. One possible model for these data is that they result from a S-I-R (susceptible-infected-recovered) model, as follows. Let  $C_t$  denote the observed number of new cases in week  $t$  of the data, while  $S_t$  and  $I_t$  are the number of individuals susceptible to infection and actually infected, respectively. Then the model is

$$\begin{aligned}S_{t+1} &= S_t e^{-a_t I_t S_t}, & a_t &\sim \text{lognorm}(\alpha, \tau_a) \\I_{t+1} &= (I_t - S_{t+1} + S_t) e^{-d_t I_t}, & d_t &\sim \text{lognorm}(\delta, \tau_d)\end{aligned}$$

The idea is that the susceptible population at  $t + 1$  is made up the susceptible population at  $t$  not infected between  $t$  and  $t + 1$ , while the infected population is made up of the infected population at  $t$  plus those infected between  $t$  and  $t + 1$ , who have not recovered (or died) between  $t$  and  $t + 1$ .

The data contain no information to determine the scale of  $S$  alongside the other parameters, so w.l.o.g. we set  $S_1 = 1$ . For the sake of simplicity also use the prior  $I_1 \sim \text{lognormal}(-5, 1)$  (-5 is the expectation of  $\log I_1$ , here). The observed cases are then related to the infected population via

$$C_t \sim \text{Poi}(\phi I_t).$$

Use `dgamma(1, .01)` priors for  $\tau_a, \tau_d$  and  $\phi$  and `N(0, 102)` priors for  $\alpha$  and  $\delta$  (noting that in JAGS the normal is parameterized using the precision, rather than the variance, and `dlnorm` is the lognormal).

1. Implement the model in JAGS. `jags.model` has difficulty initialising the model. To help it along use the argument `init=list(phi=1000, alpha=1, delta=.1)`.
2. Generate 10000 samples from the posterior distribution of the model unknowns (remember to tell `jags.samples` which variables to output).
3. Produce trace plots for  $\alpha$ ,  $\delta$  and  $\phi$ , and an ACF for the most problematic (you'll probably need to use the argument `max.lag=320` to get a sensible ACF plot). Estimate how many samples you would need to get an effective sample size around 500.
4. Produce plots showing 20 (approximately independent) random trajectories for  $S\phi$  and  $I\phi$  overlaid on the case data.
5. Produce a plot to visualize the average of the infection rate  $a_t$  at each time, according to the posterior, and interpret it.

## Solution

1. The JAGS model specification file ('sir.jags', say) should contain something like

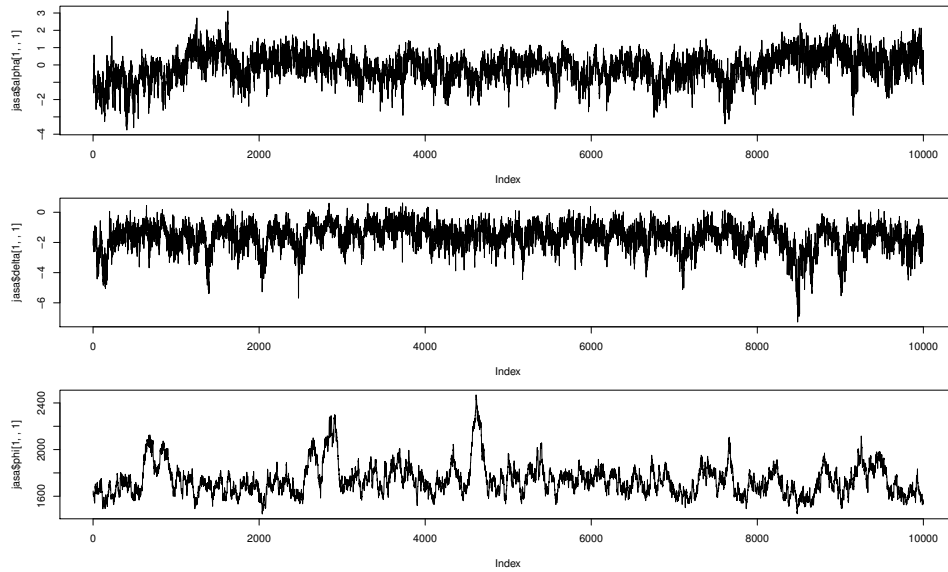
```
model {
  S[1] <- 1
  I[1] ~ dlnorm(-5,1)
  for (i in 1:(N-1)) {
    a[i] ~ dlnorm(alpha,taud)
    S[i+1] <- S[i]*exp(-a[i]*I[i]*S[i])
    d[i] ~ dlnorm(delta,taua)
    I[i+1] <- (I[i] + S[i] - S[i+1])*exp(-d[i]*I[i])
  }
  for (i in 1:N) {
    case[i] ~ dpois(phi*I[i])
  }
  phi ~ dgamma(1,.01)
  delta ~ dnorm(0,.01)#dgamma(1,.01) ## mean is product
  alpha ~ dnorm(0,.01)#dgamma(1,.01)
  taud ~ dgamma(1,.01)
  taua ~ dgamma(1,.01)
}
```

To run it from R...

```
library(rjags)
setwd("where/your/jags/file/is") ## modify this!!
h1n1 <- read.table("https://people.maths.bris.ac.uk/~sw15190/TOI/sh-H1N1.dat")
N <- length(h1n1$cases)
dat <- list(case=h1n1$cases[1:N],N=N)
jmod <- jags.model("sir.jags",data=dat,n.adapt=1000,
                  init=list(phi=1000,alpha=1,delta=.1))
jasa <- jags.samples(jmod,c("S","I","a","alpha","delta","phi"),n.iter=10000)
```

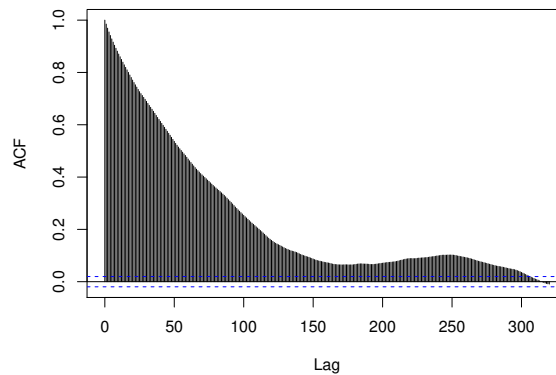
2. 

```
par(mfrow=c(3,1),mar=c(4,4,1,1))
plot(jasa$alpha[1,,1],type="l")
plot(jasa$delta[1,,1],type="l")
plot(jasa$phi[1,,1],type="l")
```



$\phi$  is mixing rather slowly...

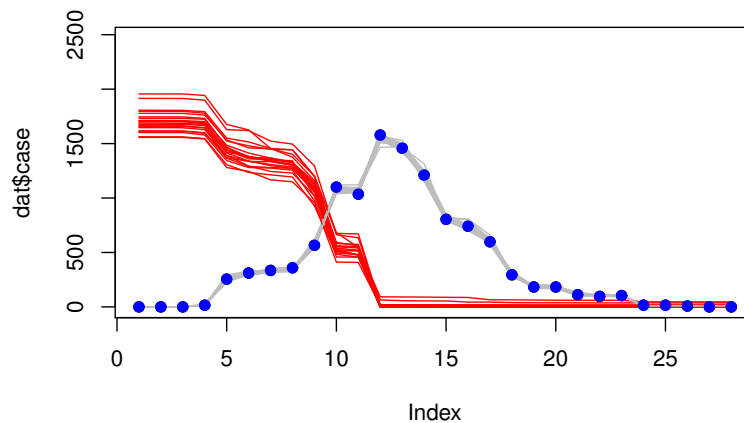
```
par(mfrow=c(1,1))
pcl <- 2*sum(acf(jasa$phi[1,,1],lag.max=320)[[1]][,1])-1
```



Using the correlation length computed above we can estimate the run length needed for an effective sample size of 500:

```
> pcl*500
[1] 75735.82
```

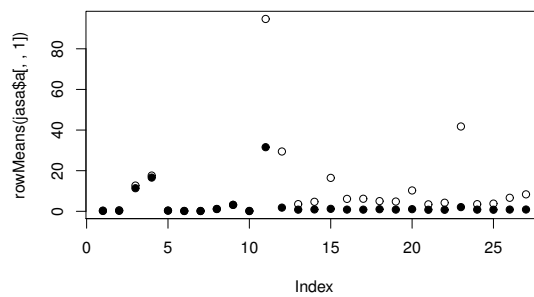
```
3. par(mfrow=c(1,1))
plot(dat$case,ylim=c(0,max(jasa$phi)))
for(k in 1:20*500) {
  lines(jasa$phi[,k,]*jasa$S[,k,1],col="red")
  lines(jasa$phi[,k,]*jasa$I[,k,1],col="grey")
}
points(dat$case,pch=19,col=4)
```



So the model seems to imply that everyone who could have been infected has been, which actually seems rather unlikely.

- Let's look at the mean and median of  $a_t$  through time (really we should use a longer run for this).

```
plot(rowMeans(jasa$a[, , 1])) ## mean open circles
points(apply(jasa$a[, , 1], 1, median), pch=19) ## median solid
```



So peak transmission seems to be at about week 11, which is mid-July - the middle of the Southern winter.