Stat 641 Bayesian Statistics

Homework 9, due Monday, November 23, 2020.

1. Sampling from prior distributions. If you specify in your JAGS model statement a model that doesn't have a likelihood (that is, it refers to no observations), you can obtain samples from the probability distribution of whatever else is stated in your model. (In such situations, it doesn't matter whether we call it a "prior distribution" or simply a "distribution"; we're effectively using software that's designed for Bayesian models to do something that wasn't necessarily intended.) We might wish to do this – fit a model with no likelihood – to make sure we understand what the distributions are that JAGS is using: is the 2nd parameter the variance or the precision? Is the mean α/β or is it $\alpha\beta$? Etc.

   (a) Use JAGS to simulate 10,000 samples from what-it-calls a Gamma(3,5) distribution using this model statement:

   ```
   model{
     theta ~ dgamma(3,5)
   }
   ```

   Use the JAGS output (`summary(my.coda.samples)`) to answer these questions: Are the mean and variance of this distribution α/β and α/β^2? Or are they $\alpha\beta$ and α/β^2?

   (b) Use JAGS to simulate 10,000 samples from what-it-calls an InverseGamma(3,5) distribution using this model statement:

   ```
   model{
    foo ~ dgamma(3,5)
    theta <- 1/foo
   }
   ```

   Use the JAGS output to answer these questions: Are the mean and variance of this distribution $\beta/(\alpha-1)$ and $\beta^2/[(\alpha-1)^2(\alpha-2)]$? Or are they $1/(\beta(\alpha-1))$ and $1/[\beta^2(\alpha-1)^2(\alpha-2)]$?

   (c) Use JAGS code to simulate 10,000 values from the distribution specified by the following model statement:

   ```
   model{
    k ~ dcat( probs[] )
   }
   ```

   Here's some R code for fitting this model. Be sure to run it one line at a time so you can figure out what it's doing. (The Latex file is posted along with this assignment, so you can copy and paste the R code.)

   ```
   library(rjags)
   my.data <- list( probs = 0.1*rep(1,10) )
   my.inits1 <- list( k = 1 )
   my.inits <- list( my.inits1 )
   my.fname <- "dcat-model.txt"
   my.jags.model <- jags.model(
     file = my.fname, data = my.data, inits = my.inits,
     n.chains = 1, n.adapt = 1000, quiet = FALSE )
   my.variables <- c("k")
   ```

```
my.coda.samples <- coda.samples(
  my.jags.model,my.variables,100000,thin=1)
summary(my.coda.samples)
k.samples <- my.coda.samples[[1]][,"k"]
unique( k.samples )
hist(k.samples, prob=TRUE, main="")
lines(density(k.samples))
#plot(my.coda.samples)
```

i. What is the distribution of k? What values does k take, and what are the associated probabilities?
ii. Be sure to include the histogram so you can discuss it.
iii. Use the statement below rather than the earlier code for my.data, and refit the model. What is the distribution of k? What values does k take, and what are the associated probabilities?

```
sum(1:8)
my.data <- list( probs = (1:8)/sum(1:8) )
```

(d) Use JAGS code to simulate 10,000 values from the distribution specified by the following model statement:

```
model{
 foo ~ dcat( probs[] )
 k <- values[foo]
}
```

Here's some R code for fitting this model.

```
library(rjags)
probs <- c(1,2,3)/6
values <- c( 1, 42, 17 )
my.data <- list( probs = probs, values = values )
my.inits1 <- list( foo = 1 )
my.inits <- list( my.inits1 )
my.fname <- "dcat-model.txt"
my.jags.model <- jags.model(
  file = my.fname, data = my.data, inits = my.inits,
  n.chains = 1, n.adapt = 1000, quiet = FALSE )
my.variables <- c("k")
my.coda.samples <- coda.samples(
  my.jags.model,my.variables,100000,thin=1)
summary(my.coda.samples)
k.samples <- my.coda.samples[[1]][,"k"]
unique( k.samples )
hist(k.samples, prob=TRUE, main="")
lines(density(k.samples))
#plot(my.coda.samples)
```

i. What is the distribution of k? What values does k take, and what are the associated probabilities?
ii. Be sure to include the histogram so you can discuss it.

2. A changepoint model. Changepoint models are models for data that are measured over time, where the behavior of the data appears to change at some point in time. For example, $Y_i =$ number of meals eaten at a restaurant in week i, where the values were fairly steady pre-covid-19, and fairly steady post-covid-19, but at a different level. The data set for this problem consists of counts of coal mining disasters in Great Britain, by year, from 1851 to 1962, which is 112 years worth of data. (Here, "disaster" is defined as an accident resulting in the deaths of 10 or more miners.) The data is posted on Blackboard in the file, `coal_mining_disasters.txt`.

   (a) Plot the data; discuss briefly. Does it appear to you that the rate of coal mining disasters changed appreciably during this time interval? If so, in approximately what year do you think the change began? Here's some R code. Be sure to include the plot in your solutions.

```
my.data <- read.table("coal-mining-disasters.txt",header=TRUE)
names(my.data)
disasters <- my.data$disasters
year <- my.data$year
N <- length(disasters); N
plot(year,disasters, type='l')
```

   (b) We consider the following changepoint model for these data.

$$Y_i \sim \begin{cases} \text{Poisson}(\lambda_1), & i = 1, 2, \ldots, k \\ \text{Poisson}(\lambda_2), & i = k+1, k+2, \ldots, n \end{cases}$$

$$\lambda_i \overset{ind}{\sim} \text{Exponential}(1), \; i = 1, 2$$

   We initially assume $k = 40$. What year is the coal mining disaster rate assumed to have changed?

   (c) State the posterior distribution: $p(\lambda_1, \lambda_2 | \mathbf{y}) \propto L(\lambda_1, \lambda_2)\pi(\lambda_1, \lambda_2) \propto \cdots$

   (d) This will be our model statement.

```
model {
  for( i in 1:N ) {
    disasters[i] ~ dpois( lambda[ idx[i] ] )
    idx[i] <- 1+step( i-k-0.5 )
  }
  lambda[1] ~ dexp(1)
  lambda[2] ~ dexp(1)
}
```

   Explain carefully these two lines, which specify the likelihood:

```
    disasters[i] ~ dpois( lambda[ idx[i] ] )
    idx[i] <- 1+step( i-k-0.5 )
```

   For which values of i will idx[i] be equal to 1? For which values of i will it be equal to 2? Are we sure that idx[i] will always be either 1 or 2, which it needs to be in the line, `disasters[i] ~ dpois( lambda[ idx[i] ] )`? Explain briefly.

(e) Fit the model using JAGS. Here's some R code for fitting the model:

```
library(rjags)
my.data <- list( disasters = disasters, N = N, k = 40 )

my.inits1 <- list( lambda = c(40,40) )
my.inits2 <- list( lambda = c(20,20) )
my.inits3 <- list( lambda = c(100,10) )
my.inits <- list( my.inits1, my.inits2, my.inits3 )

my.model.1.fname <- "coal-mining-disasters-model-1.txt"
my.jags.model <- jags.model(
  file = my.model.1.fname, data = my.data, inits = my.inits,
  n.chains = 3, n.adapt = 1000, quiet = FALSE )

dic.samples(my.jags.model, n.iter=10000, thin=1, type="pD")
my.variables <- c("lambda" )

my.coda.samples <- coda.samples(my.jags.model,
                                my.variables,10000, thin=1)
lambda1.samples <- my.coda.samples[[1]][,"lambda[1]"]
lambda2.samples <- my.coda.samples[[1]][,"lambda[2]"]
n.samples <- length(lambda1.samples); n.samples
which.ones <- seq(1,n.samples, length=500)
some.lam1.samples <- lambda1.samples[which.ones]
some.lam2.samples <- lambda2.samples[which.ones]

plot( which.ones, some.lam1.samples, type='l',
      xlab="iteration", ylab="lambda[1]")
plot( which.ones, some.lam2.samples, type='l',
      xlab="iteration", ylab="lambda[2]")
rejectionRate(my.coda.samples)

hist(lambda1.samples, prob=TRUE, main="")
lines(density(lambda1.samples))
hist(lambda2.samples, prob=TRUE, main="")
lines(density(lambda2.samples))
summary(my.coda.samples)
effectiveSize(my.coda.samples)
```

(f) Report your results: traceplots, density plots, summary statistics (including MC error!), and credible intervals. What is the effective sample size for each parameter? What is the rejection rate? Can you tell whether the steps are Gibbs steps (or something other than Gibbs steps, such as perhaps Metropolis steps)?

(g) Next you'll fit a model in which k, the changepoint year, is a parameter to be estimated:

$$Y_i \sim \begin{cases} \text{Poisson}(\lambda_1), & i = 1, 2, \ldots, k \\ \text{Poisson}(\lambda_2), & i = k+1, k+2, \ldots, n \end{cases}$$

$$\lambda_1 \sim \text{Gamma}(0.5, 3.5), \lambda_2 \sim \text{Gamma}(0.5, 1.6)$$

$$k \sim \text{Discrete uniform}\{1, 2, \ldots, n\}$$

Note: This problem is a more-than-slightly modified version of Problems 3.9 and 3.10 of Carlin & Louis, 3rd ed.; I've chosen simplified priors for λ_1 and λ_2, which is why the values may seem a bit odd.

Please state the posterior distribution,

$$p(\lambda_1, \lambda_2, k) \propto L(\lambda_1, \lambda_2, k)\pi(\lambda_1, \lambda_2, k) \ = ?$$

Note that k will simply appear in the posterior distribution as part of an "indicator function", e.g. $\frac{1}{n} I(1 \leq k \leq n)$, to make sure that k stays between 1 and n.

(h) Here's the model statement:

```
model{
  for( i in 1:n ) {
    disasters[i] ~ dpois( lambda[ idx[i] ] )
    idx[i] <- 1 + step( i-k-0.5 )
    punif[i] <- 1/n
  }
  k ~ dcat( punif[] )
  lambda[1] ~ dgamma(0.5,3.5)
  lambda[2] ~ dgamma(0.5,1.6)
}
```

Fit this model using BUGS and report your results – traceplots, density plots, summary statistics (including MC error!), and credible intervals.

Note that k is now a parameter rather than data, so you need to remove it from the list of data and add it to the list of values to initialize.

Be sure to include plots for k as well as λ_1 and λ_2.

(i) Comment on what this second model tells us. Was our initial guess of $k = 40$ justified? Does the data unequivocally support a changepoint at $k = 40$?

(j) Based on DIC, which statistical model is preferred for these data? The model with $k = 40$ or the model where k is sampled as a parameter. (I'm a bit puzzled as to how DIC is calculated for the second model, since the posterior mean of k is not an integer; perhaps JAGS uses the (rounded value of the) posterior median for its DIC calculations.)

(k) Refit your model after including an additional calculation in your model statement: `R <- lambda[2] / lambda[1]`

What does `R` represent? Please provide summary statistics for it, and discuss briefly.

3. How does the use of `dcat` in problems 1(c), 1(d), and 2(h) differ from the use of `dcat` on page 146 of the lecture notes?