lecture 2 exercises (ESSLLI)

Shravan Vasishth

August 10, 2015

Solutions are given at the end of this document.

Problem 1

Let's say there is a hormone measurement test that yields a numerical value that can be positive or negative. We know the following:

- The doctor's prior: 75% interval ("patient healthy") is [-0.3,0.3].
- Data from patient: $x = 0.2$, known $\sigma = 0.15$.

Compute posterior N(m*,v*).

Problem 2

We are given five measurements of a rat's weight, in grams, as a function of some x (say some nutrition-based variable). Note that here we will center the predictor in the model code. The goal is to understand how the amount of nutrition x affects the rats' weights.

First we load/enter the data:

```
data<-list(x=c(8,15,22,29,36),
           y=c(177,236,285,350,376))
```

Determine the coefficient estimates and sigma estimate using a JAGS model that corresponds to the linear model:

```
fm<-lm(y~x,data)
```

Also find out the 95% credible intervals for each estimate:

Problem 3

This uses the rats data from problem 2.

Fit the JAGS model:

```
model
   {
    ## specify model for data:
    for(i in 1:5){
    y[i] ~ dnorm(mu[i],tau)
    mu[i] <- beta0 + beta1 * (x[i]-mean(x[]))
    }
    # priors:
    beta0 ~ dunif(-500,500)
    beta1 ~ dunif(-500,500)
    tau <- 1/sigma2
    sigma2 <-pow(sigma,2)
    sigma ~ dunif(0,200)
   }
```

Here, we have centered the predictor so that the average value of x is now represented as 0.

What has changed in the model compared to problem 2? Write down the estimates and their uncertainty intervals and compare with Problem 2's estimates.

Problem 4

In the beetle data, we fit uniform priors to the coefficients β:

```
    # priors:
    beta0 ~ dunif(0,100)
    beta1 ~ dunif(0,100)
```

Fit the beetle data again, using suitable normal distribution priors for the coefficients beta0 and beta1. Does the posterior distribution depend on the prior?

Solution to problem 1

1. We know the prior mean m=0 but **we need to figure out prior sd s** from the 75% credible interval given.
2. We know the sample mean $\bar{x} = 0.2$ and the sample sd σ=0.15.
3. We know how to find the posterior mean and sd given the above information.

Finding the prior sd s:

1. Recall that if a 75% interval in a normal distribution is [-0.3,0.3], then $Prob(\theta > 0.3) = 0.125$.
2. We know that the z-score in N(0,1) which has area 0.125 to its right is 1.15 (`qnorm(.125,lower.tail=FALSE)`).
3. From the relation (slide **??**) $z = 1.15 = \frac{0.3-0}{s}$, we can compute s given z: $0.3/1.15 = 0.2609$ is the standard deviation of the prior distribution.

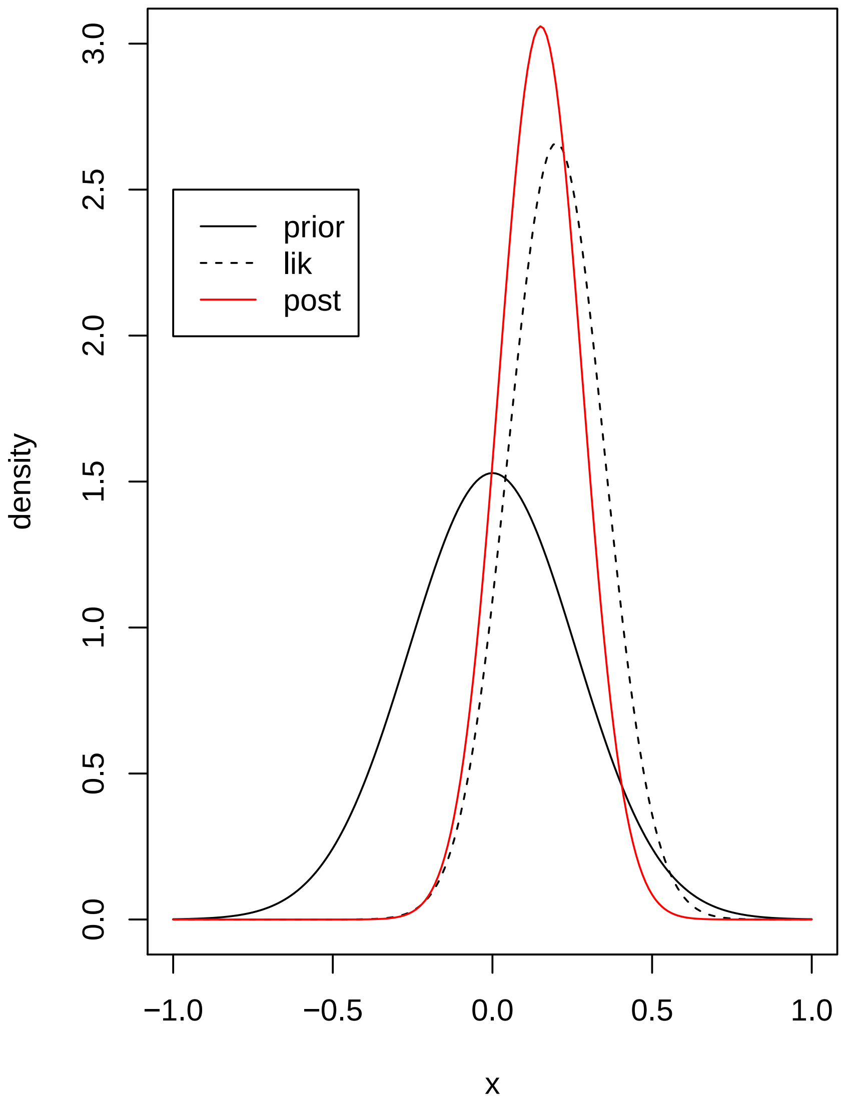
Posterior variance (v*):

$$v^* = \frac{1}{\frac{1}{v} + \frac{n}{\sigma^2}} = \frac{1}{\frac{1}{0.2609^2} + \frac{1}{0.15^2}} = 0.02 \quad (1)$$

Posterior mean (m*):

$$m^* = v^* \left(\frac{m}{v} + \frac{n\bar{x}}{\sigma^2} \right) = 0.017 \times \left(\frac{0}{0.2609^2} + \frac{1 \times 0.2}{0.15^2} \right) = 0.15 \quad (2)$$

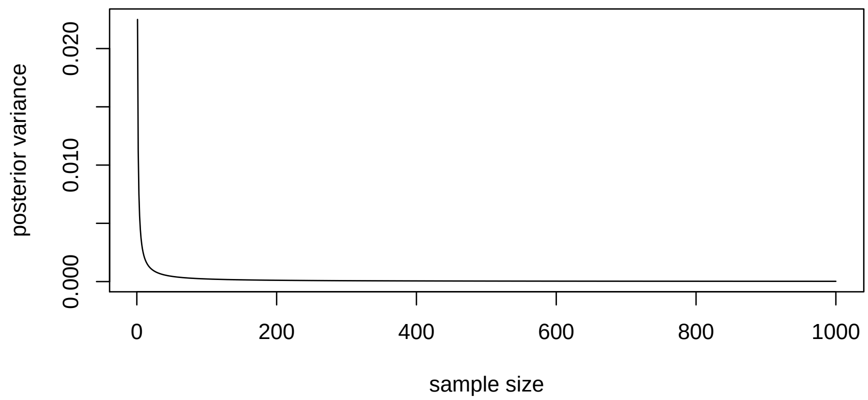
Optional plot:



Comment on this problem: Think about what happens to posterior variance as sample size n increases, for given v and σ:

$v^* = \frac{1}{\frac{1}{v}+\frac{1}{\sigma^2/n}}$

```
vpost<-function(v=0.2609^2,n=1,s=0.15^2){
  return(1/((1/v)+n/s))
}
n<-seq(1,1000,by=1)
plot(n,vpost(v=2600,n=n),type="l",ylab="posterior variance",xlab="sample size")
```



Solution to problem 2

Load data:

```
data<-list(x=c(8,15,22,29,36),
           y=c(177,236,285,350,376))

track.variables<-c("beta0","beta1",
                   "sigma")
library(rjags)

## Linked to JAGS 3.4.0
## Loaded modules:  basemod,bugs

cat("
model
   {
    ## specify model for data:
    for(i in 1:5){
    y[i] ~ dnorm(mu[i],tau)
    mu[i] <- beta0 + beta1 * (x[i])
    }
```

```
    # priors:
    beta0 ~ dunif(-500,500)
    beta1 ~ dunif(-500,500)
    tau <- 1/sigma2
    sigma2 <-pow(sigma,2)
    sigma ~ dunif(0,200)
   }",
     file="ratsexample1.jag" )

rats.mod <- jags.model(
  file = "ratsexample1.jag",
                    data=data,
                    n.chains = 2,
                    n.adapt =2000,
                     quiet=T)

rats.res <- coda.samples( rats.mod,
                               var = track.variables,
                            n.iter = 10000)
```

```
## mean:
summary(rats.res)$statistics[,1:2]

##        Mean   SD
## beta0 124.5 35.0
## beta1   7.3  1.5
## sigma  22.5 20.7

## lower, median, upper credible interval:
summary(rats.res)$quantiles[,c(1,3,5)]

##       2.5%   50% 98%
## beta0 58.4 124.1 188
## beta1  4.6   7.3  10
## sigma  7.0  16.1  80
```

Results should match lm fit (check this).

Solution to Problem 3

```
cat("
model
   {
```

```
    ## specify model for data:
    for(i in 1:5){
    y[i] ~ dnorm(mu[i],tau)
    mu[i] <- beta0 + beta1 * (x[i]-mean(x[]))
    }
    # priors:
    beta0 ~ dunif(-500,500)
    beta1 ~ dunif(-500,500)
    tau <- 1/sigma2
    sigma <-pow(sigma2,1/2)
    #sigma ~ dunif(0,200)
    log(sigma2) <- 2* log.sigma
    log.sigma ~ dunif(0,8)
   }",
     file="ratsexample2llogsigma.jag" )
```

```
track.variables<-c("beta0","beta1","sigma")

## define model:
rat.mod <- jags.model(
  file = "ratsexample2llogsigma.jag",
                    data=data,
                    n.chains = 4,
                    n.adapt =2000,
                    quiet=T)

## sample from posterior:
rat.res <- coda.samples(rat.mod,
                         var = track.variables,
                         n.iter = 2000,
                         thin = 1 )

summary(rat.res)$statistics[,1:2]

##        Mean    SD
## beta0 285.0  7.74
## beta1   7.3  0.92
## sigma  15.2 10.98
```

Solution to Problem 4

Try a normal distribution prior with mean 0 and increasing precision. For low precisions, not much should change, but for high precision we should see the prior dominating.