Homework 3

Kent Codding

2025-03-25

```
load("C:/Users/khcod/Downloads/HW3-1 (1).RData")
```

1)

Using the JAGS model for the relative risk, estimate the probability that study 1 has the highest relative risk among all studies of TABB dataset. You can assume the studies are independent.

```
library(rjags)
```

```
## Warning: package 'rjags' was built under R version 4.4.2
```

```
## Loading required package: coda
```

```
## Warning: package 'coda' was built under R version 4.4.2
```

```
## Linked to JAGS 4.3.1
```

```
## Loaded modules: basemod,bugs
```

```
RRmodel ="model {
for( i in 1 : Num ){
  x0[i] ~ dbin(theta0[i],n0[i]);
  x1[i] ~ dbin(theta1[i],n1[i]);
  theta0[i] ~ dunif(0,1);
  log(theta1[i]) <- log(theta0[i])+lambda[i];
  lambda[i] ~ dnorm(mu,precision.tau);
}
mu ~ dnorm(0.0, 0.01);
# tau ~ dunif(0,2);
tau.int.prior <- rep(.03,25)
tau.int ~ dcat(tau.int.prior)
tau <- tau.int * .01
maxrelrisk <- exp(max(lambda))
relrisk <- exp(mu);
precision.tau <- 1/(tau*tau);
}
"
```

```
tamBayesRR = jags.model(textConnection(RRmodel),
                    data = list( x1 = TABB$x1, n1 = TABB$n1, x0 = TABB$x0, n0 = TABB$n0, Num=nrow(TABB)),
                    n.chains = 2,
                    n.adapt = 100)
```

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
```

```
## Graph information:
##    Observed stochastic nodes: 14
##    Unobserved stochastic nodes: 16
##    Total graph size: 79
##
## Initializing model
```

```
# add study-specific lambda for all seven studies
mcmc.out = coda.samples(tamBayesRR,c("relrisk","maxrelrisk","tau", "lambda"),n.iter = 40000,thin=10)
save(mcmc.out,file="mcmc.out.Rdata")
dic.out = dic.samples(tamBayesRR,n.iter = 9000)
Srelrisk = as.vector(mcmc.out[[1]][,"relrisk"])
Smaxrelrisk = as.vector(mcmc.out[[1]][,"maxrelrisk"])
Stau = as.vector(mcmc.out[[1]][,"tau"])
```

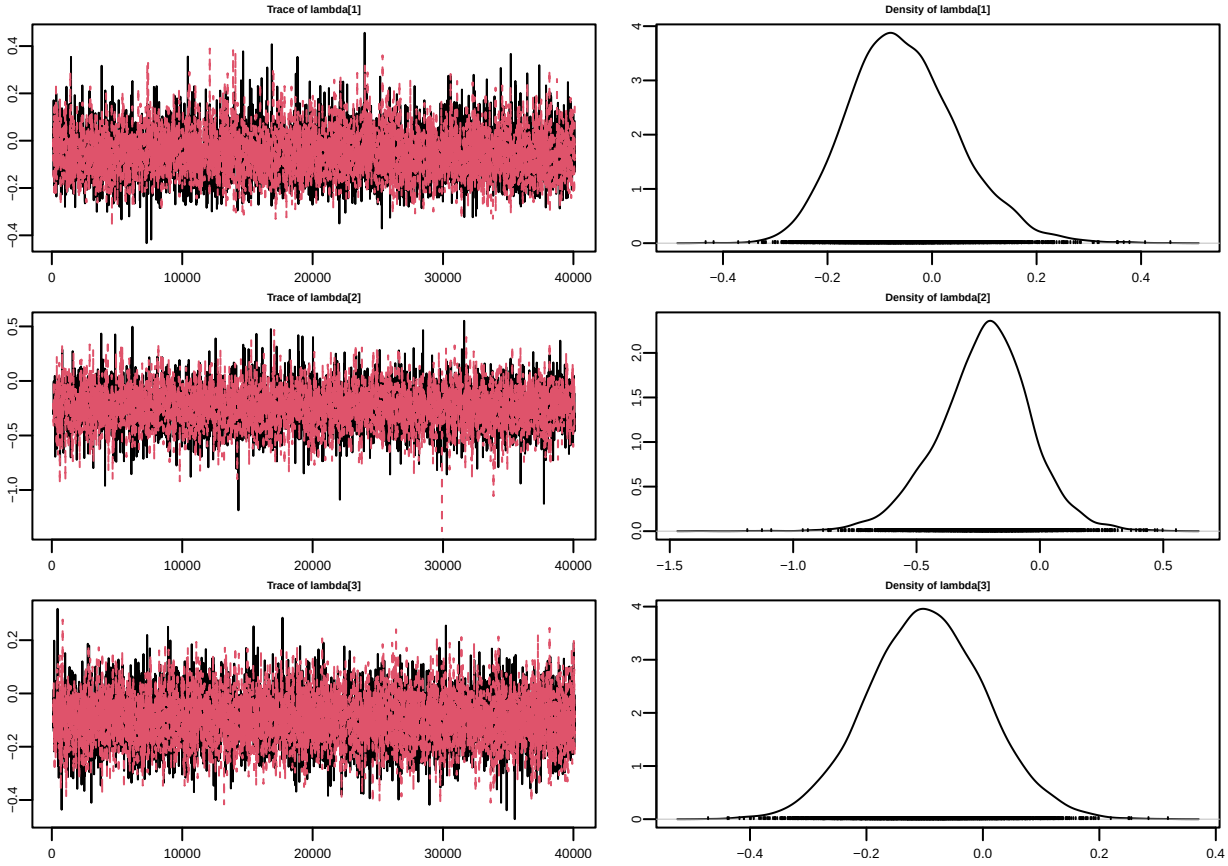
```
summary(Srelrisk)
```

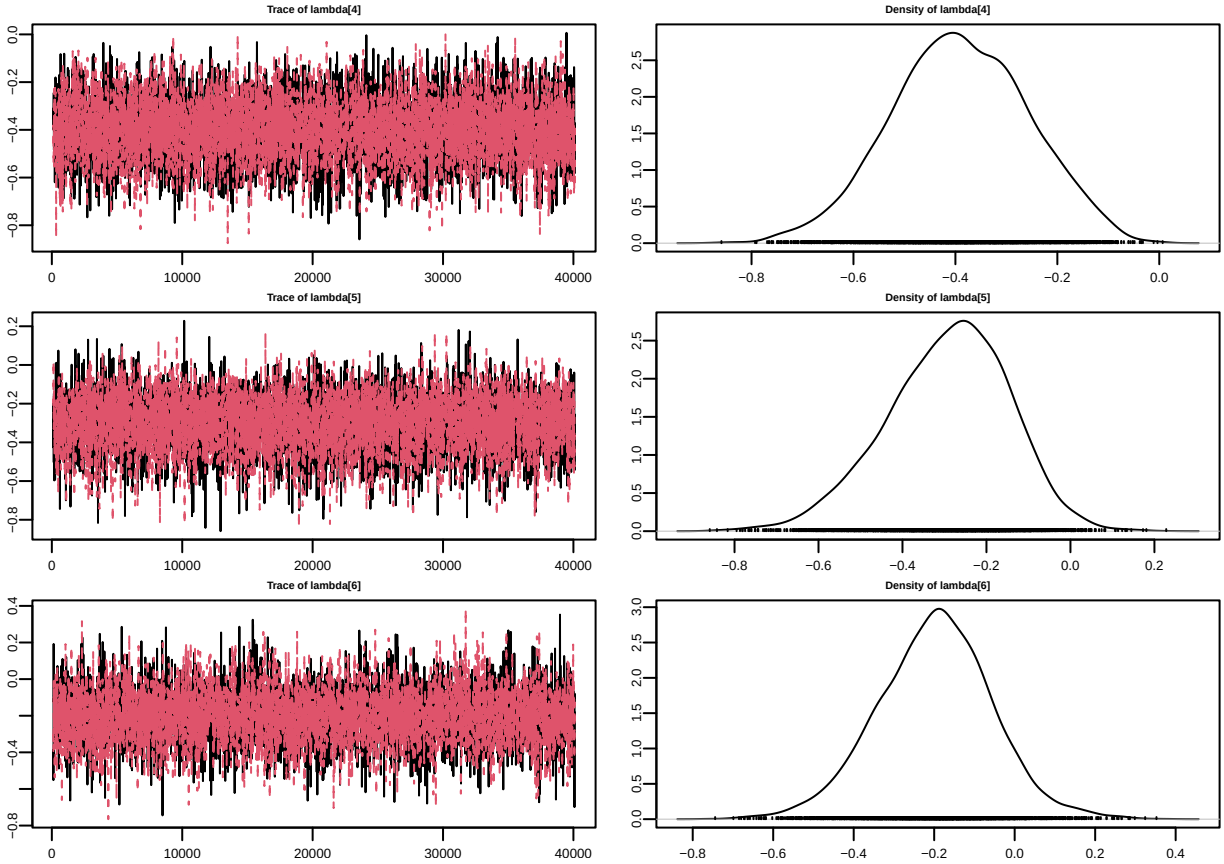
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.5687  0.7775  0.8271  0.8280  0.8758  1.1811
```

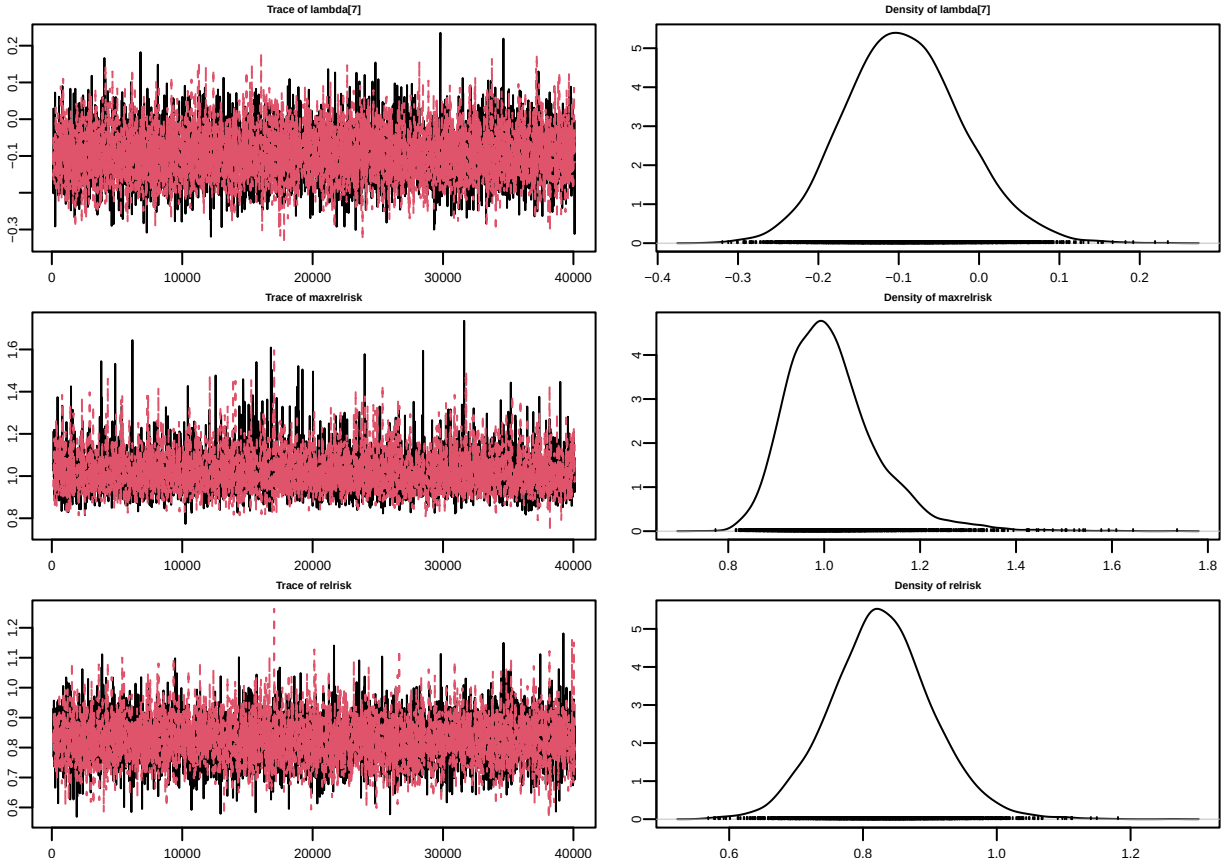
```
summary(Stau)
```

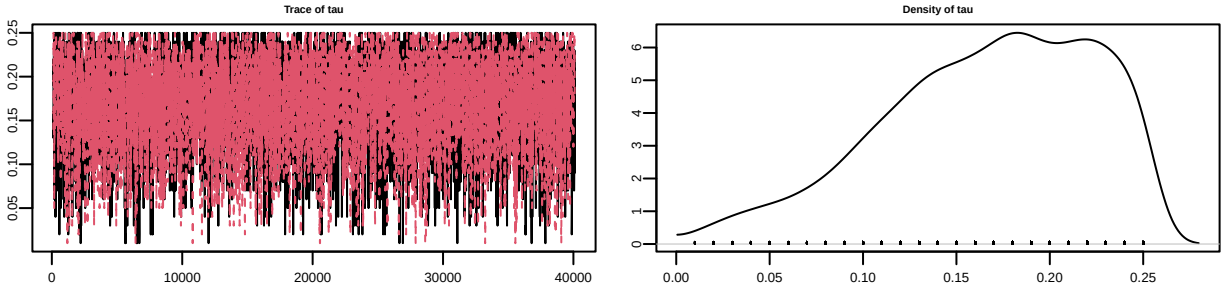
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.010   0.130   0.170   0.166   0.210   0.250
```

```
cex <- 0.6
par(cex.lab=cex, cex.axis=cex, cex.main=0.55)
par(mgp=c(1.5, 0.4, 0))
par(oma=c(0,0,0,0))
par(mar=rep(1.2, 4))
plot(mcmc.out,smooth=FALSE,auto.layout = T)
```


Trace of lambda[1]
Density of lambda[1]
Trace of lambda[2]
Density of lambda[2]
Trace of lambda[3]
Density of lambda[3]


Trace of lambda[4]
Density of lambda[4]
Trace of lambda[5]
Density of lambda[5]
Trace of lambda[6]
Density of lambda[6]


Trace of lambda[7]
Density of lambda[7]
Trace of maxrelrisk
Density of maxrelrisk
Trace of relrisk
Density of relrisk




```
#mcmc.out into matrix, iterate over each MCMC entry and get number of study with max
all.lambda <- as.matrix(mcmc.out)[, grep("lambda\\[", colnames(as.matrix(mcmc.out)))]
lambda1 <- all.lambda[, "lambda[1]"]
max.study <- apply(all.lambda, 1, which.max)
prob_study1 <- mean(max.study == 1)

cat('probability that study 1 has the highest lambda (relative risk) among the 7 studies: ', prob_study
```

```
## probability that study 1 has the highest lambda (relative risk) among the 7 studies:  0.375
```

2)

Now, modify the JAGS model you have so that the effect size is the risk difference instead. Then, please:

```
library(rjags)
# instead of log difference, measure difference in thetas
# theta1 - theta0
# log(theta1[i] <- logtheta0[i] + lambda[i])
# resulting lambdas are risk dif

RDmodel ="model {
for( i in 1 : Num ){
  x0[i] ~ dbin(theta0[i],n0[i]);
  x1[i] ~ dbin(theta1[i],n1[i]);
  theta0[i] ~ dunif(0,1);
  theta1[i] <- theta0[i]+lambda[i]; #here, lambda[i] is rel dif because lambda[i] = theta1 - theta0
```

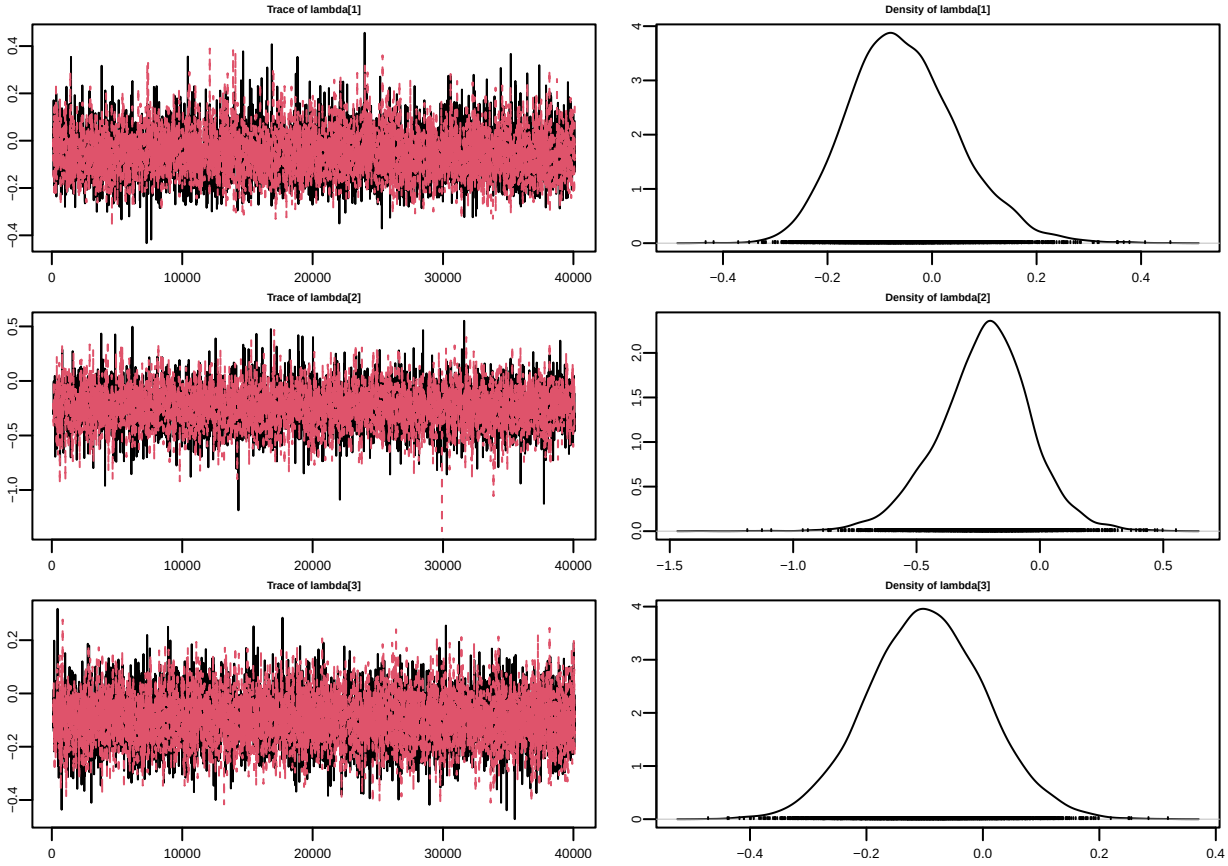
```
  lambda[i] ~ dnorm(mu,precision.tau);
}
mu ~ dnorm(0.0, 0.01);
tau ~ dgamma(2.0, 1.0);
precision.tau <- 1/(tau*tau);
maxreldiff <- max(lambda)
reldiff <- lambda;
}
"
```

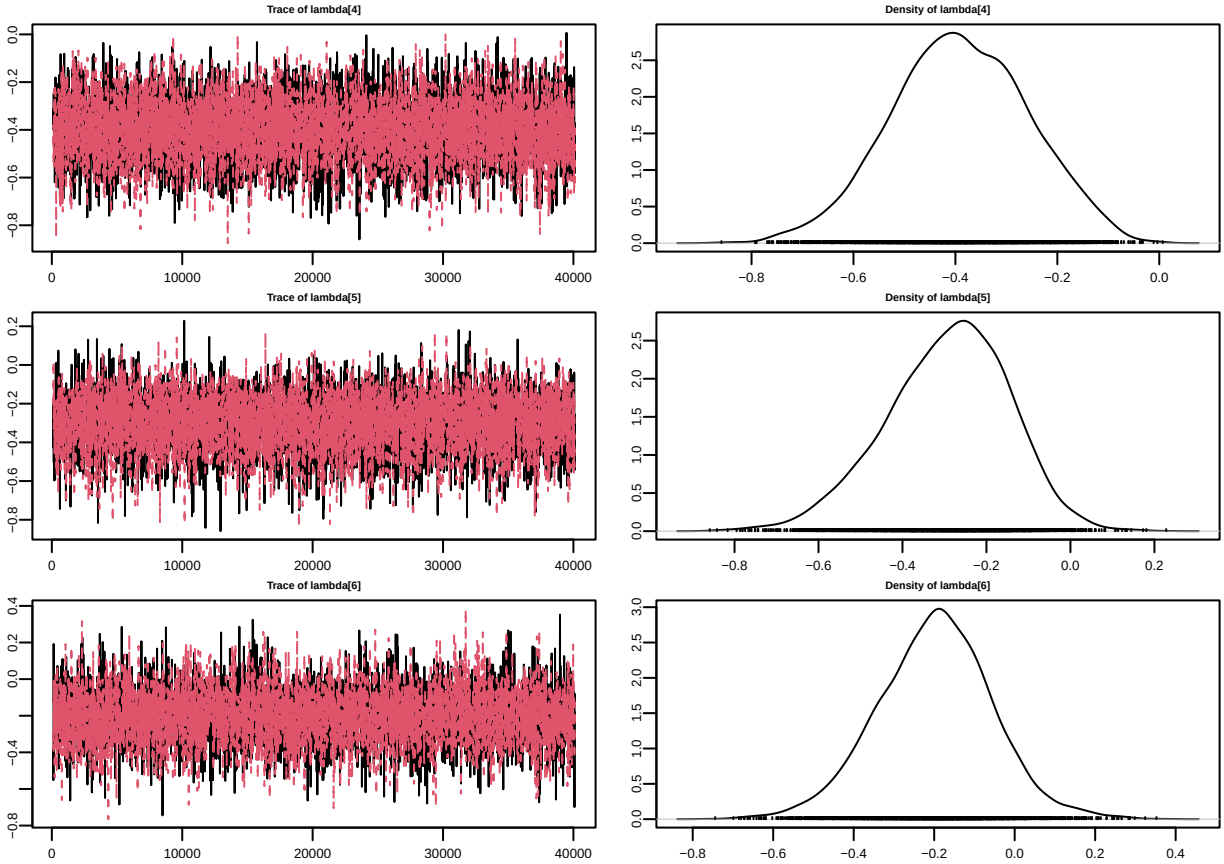
```
num_chains <- 2
init_vals <- vector("list",length=num_chains)
for(i in 1:num_chains){
  init_vals[[i]]$.RNG.name="base::Mersenne-Twister"
  init_vals[[i]]$.RNG.seed=117+i
}

tamBayesRD = jags.model(textConnection(RDmodel),
                   data = list( x1 = TABB$x1, n1 = TABB$n1, x0 = TABB$x0, n0 = TABB$n0, Num=nrow(TABB)),
                   n.chains = 2,
                   n.adapt = 100)
```

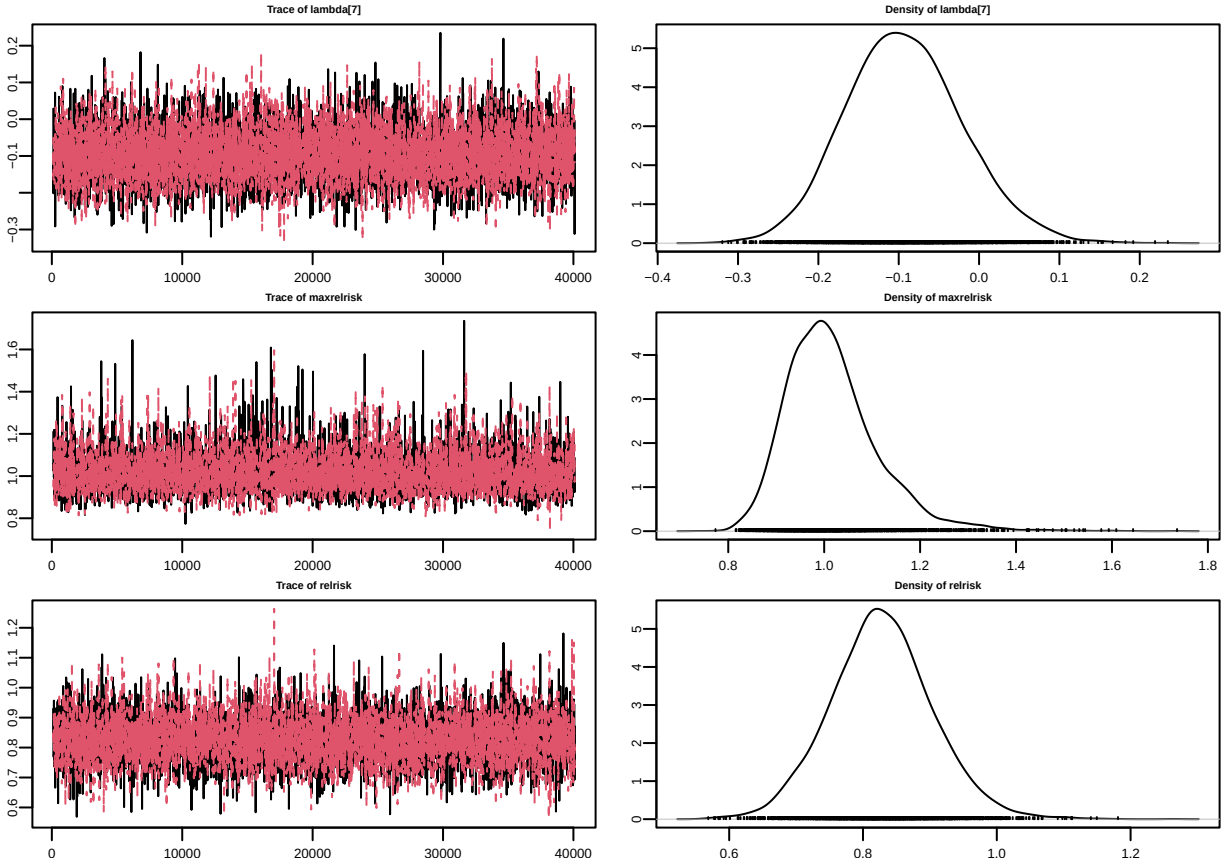
```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 14
##    Unobserved stochastic nodes: 16
##    Total graph size: 60
##
## Initializing model
```

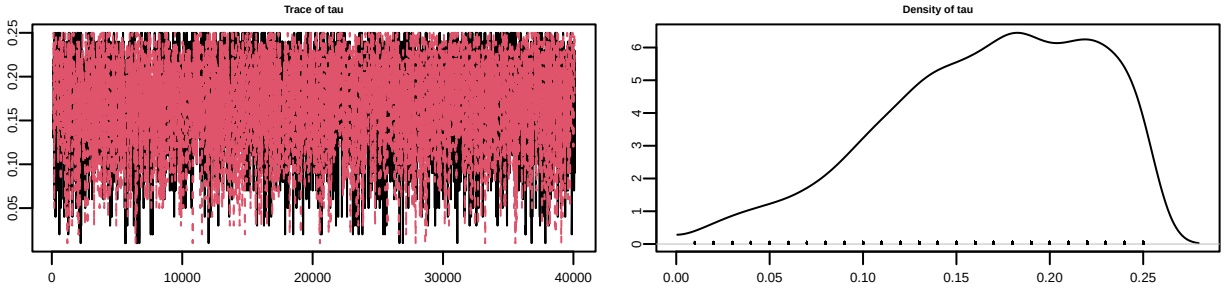
```
mcmc.out2 = coda.samples(tamBayesRD,c("reldiff","maxreldiff","tau", "lambda"),n.iter = 9000,thin=10)
save(mcmc.out,file="mcmc.out.Rdata")
dic.out = dic.samples(tamBayesRD,n.iter = 40000)
```

```
cex <- 0.6
par(cex.lab=cex, cex.axis=cex, cex.main=0.55)
par(mgp=c(1.5, 0.4, 0))
par(oma=c(0,0,0,0))
par(mar=rep(1.2, 4))
plot(mcmc.out,smooth=FALSE,auto.layout = T)
```


Trace of lambda[1]
Density of lambda[1]
Trace of lambda[2]
Density of lambda[2]
Trace of lambda[3]
Density of lambda[3]


Trace of lambda[4]
Density of lambda[4]
Trace of lambda[5]
Density of lambda[5]
Trace of lambda[6]
Density of lambda[6]


Trace of lambda[7]
Density of lambda[7]
Trace of maxrelrisk
Density of maxrelrisk
Trace of relrisk
Density of relrisk



a) Estimate the probability that the risk difference is greater than 0.

```
#put 2 chains into matrix, get col names, use same methodology as rel risk

all.lambda.RD <- as.matrix(mcmc.out2)[, grep("lambda\\[", colnames(as.matrix(mcmc.out2)))]

prob_lambda_over_0 <- apply(all.lambda.RD, 2, function(x) mean(x > 0))
cat('the probability that the risk difference is greater than 0: ', mean(prob_lambda_over_0))
```

```
## the probability that the risk difference is greater than 0:  0.147619
```

b) Estimate the probability that study 1 has the highest risk difference among all studies.

```
lambda1 <- all.lambda.RD[, "lambda[1]"]
max.study.RD <- apply(all.lambda.RD, 1, which.max)
prob_study1.RD <- mean(max.study.RD == 1)

cat('probability that study 1 has highest risk difference among all 7 studies: ', prob_study1.RD)
```

```
## probability that study 1 has highest risk difference among all 7 studies:  0.345
```

Finally, summarize in your own words what the estimates above tell you.

The above estimates of effect size measure a) the probability that the risk difference RD deviates from the null hypothesis (RD = 0) - that `BRCA1` has no positive or negative effect on debulking status. In the above measurement, there is roughly a 14% chance that the true risk of suboptimal debulking is higher for patients with higher levels of `BRCA1`. Given the low value for the probability that risk difference is greater than zero, the above estimate tells us that `BRCA1` is likely not a viable biomarker to predict tumor debulking status. b) the probability that out of all the studies, study 1 has the largest effect size based on simulated samples from

the markov chain monte carlo, which is risk difference, the difference in probability of optimal debulking in dichotomized levels of `BRCA1` expression (high versus low) in this question. There is about a 37% chance that study 1 has the highest risk difference according to the above output from the posterior distributions for each λ.

NOTE: Generative AI was utilized to assist syntax of grep() commands for part 1 and 2.