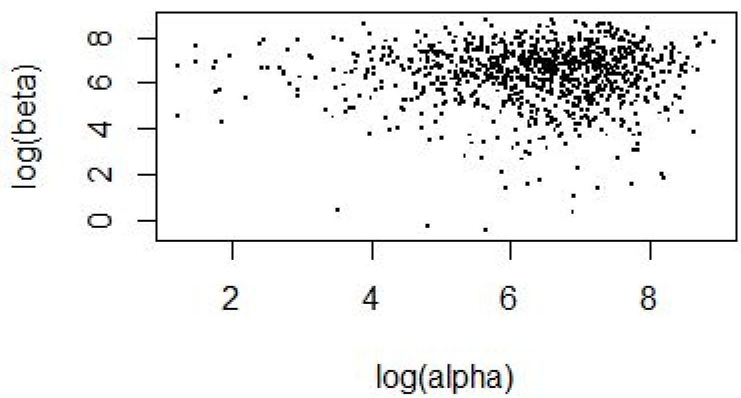
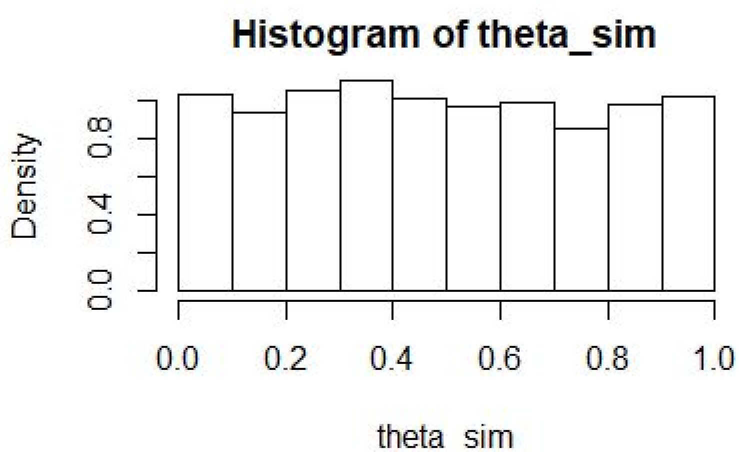
amana4@illinois.edu

Answer 1:

a)
i)R-code
```
alpha <- rexp(1000,.001)
beta <- rexp(1000,.001)
plot(log(alpha),log(beta),pch=".",cex=2)
```



ii) R-code
```
theta_sim = rbeta(1000,alpha,beta)
hist(theta_sim,freq = FALSE)
```



amana4@illinois.edu

b)

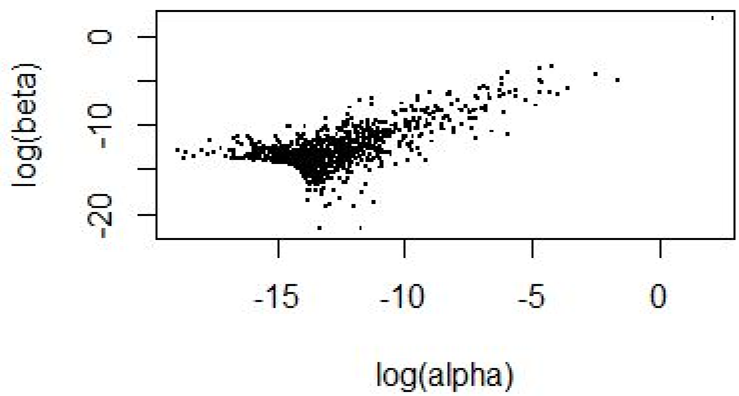
i) phi1 <- runif(1000,0,1)

phi2 <- runif(1000,0,1000)

alpha <- phi1/(phi2^2)

beta <- (1-phi1)/(phi2^2)

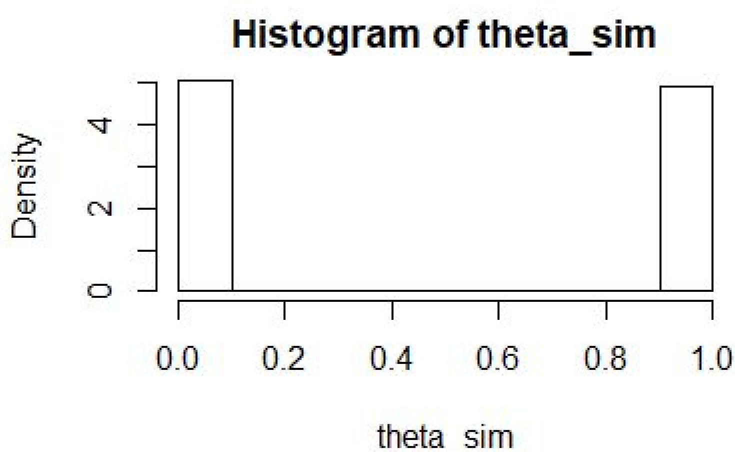
plot(log(alpha),log(beta),pch=".",cex=2)



ii)

theta_sim = rbeta(1000,alpha,beta)

hist(theta_sim, freq = FALSE)



`

amana4@illinois.edu

Answer 2:

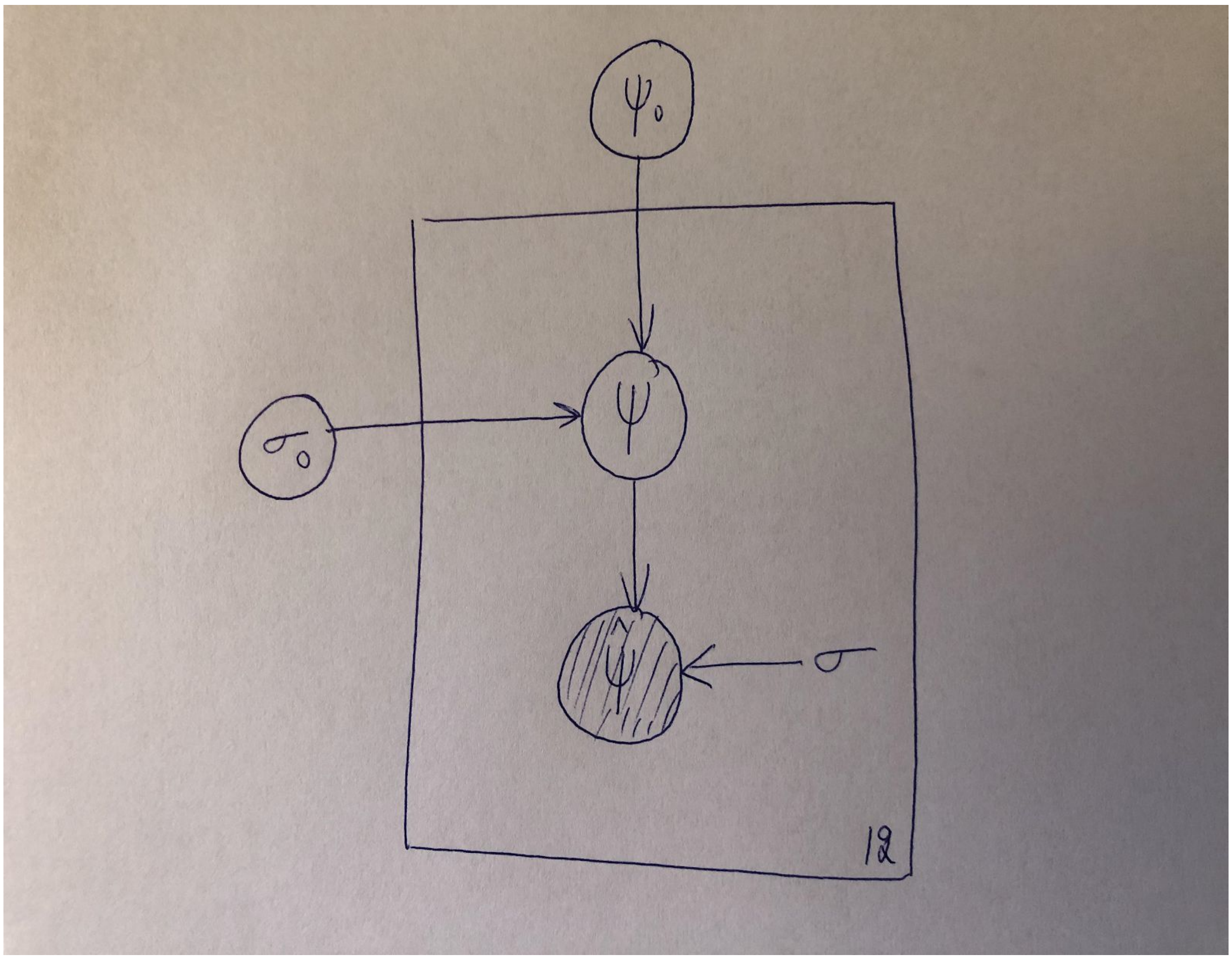
a)

i) $\sigma 0$ and $\psi 0$ are hyperparameters

$\sigma 0$: intended to approximate a uniform distribution on $(0,\infty)$

$\psi 0$: intended to approximate a diffuse normal distribution with mean 0 and infinite variance

b) DAG



c) **R setup code:**

```
#Read and format data for JAGS model
df <- read.csv("numbers.txt",sep = "",header = FALSE)
df <- df[2:3]
names(df) <- c("psihat","sigma")

#Run jags model and set the seed for reproducibility
library(rjags)
m <- jags.model("asgn2template.bug", df,inits = list(.RNG.name = "base::Wichmann-Hill",
.RNG.seed = 1989))
```

Jags model (asgn2template.bug):

```
model {
  for (j in 1:length(psihat)) {
    psihat[j] ~ dnorm(psi[j],1/sigma[j]^2)
    psi[j] ~ dnorm(psi0,1/sigmasq0)
  }

  psi0 ~ dnorm(0,1000^2)
  sigma0 ~ dunif(0,1000)

  sigmasq0 <- sigma0^2
}
```

d)
```
#Burn in 10000 iterations
 update(m,10000)

# Run the model for  100,000 iterations
x <- coda.samples(m, c("psi0","sigmasq0"), n.iter=100000)
summary(x)
```

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
psi0	1.661e-05	0.001	3.163e-06	3.203e-06
sigmasq0	1.689e-01	0.145	4.586e-04	1.106e-03

2. Quantiles for each variable:

	2.5%	25%	50%	75%	97.5%
psi0	-0.00195	-0.0006553	1.626e-05	0.0006943	0.001978
sigmasq0	0.01899	0.0776716	1.319e-01	0.2147620	0.534149

Posterior mean

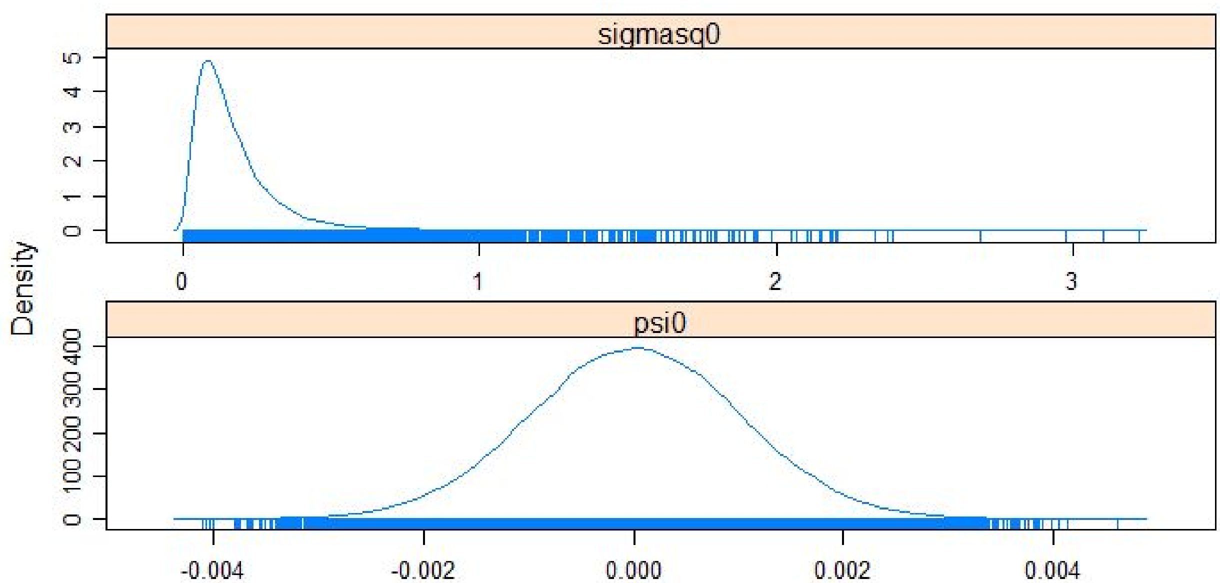
E(psi0|**psihat**) = 1.661e-05
E(sigmasq0 |**psihat**) = 1.689e-01

Posterior SD

Posterior standard deviation of psi0 = 0.001
Posterior standard deviation of sigmasq0 = 0.145

95% central posterior interval

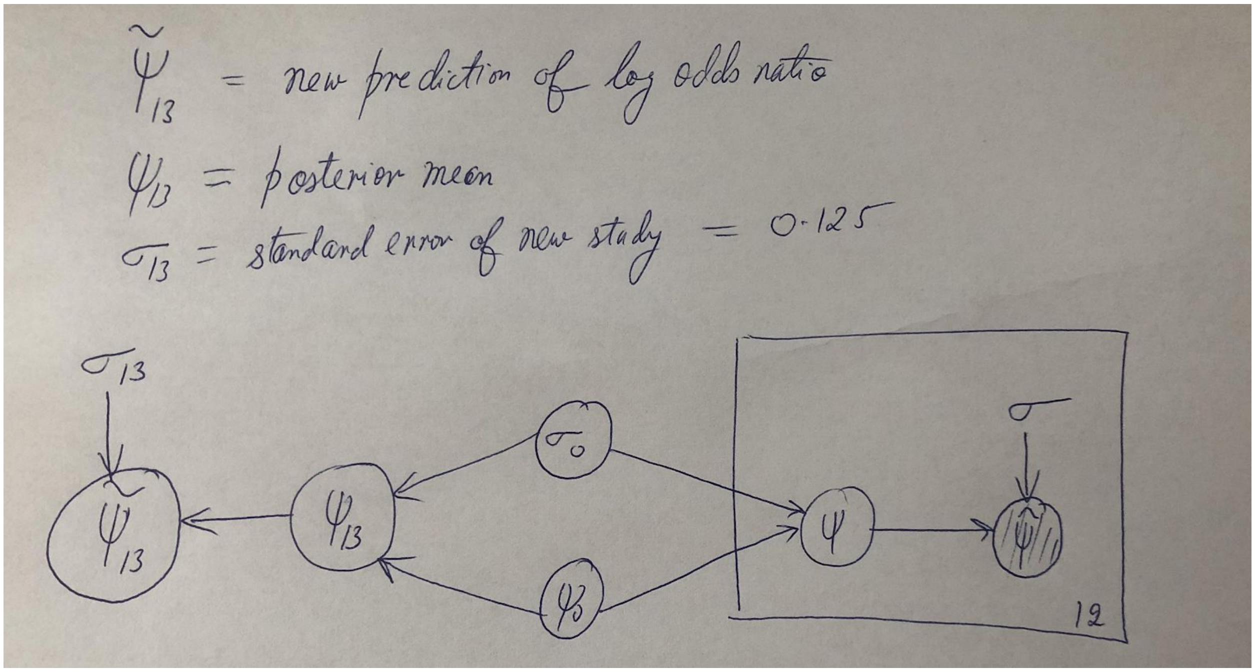
psi0 = (-0.00195, 0.001978)
sigmasq0 = (0.01899, 0.534149)



e)

amana4@illinois.edu

i)DAG for new prediction



Ii)

<u>R setup code:</u>

#Read and format data for JAGS model

```
df <- read.csv("numbers.txt",sep = "",header = FALSE)
df <- df[2:3]
names(df) <- c("psihat","sigma")
```

##Run jags model and set the seed for reproducibility

```
library(rjags)
m <- jags.model("asgn2template1.bug", c(as.list(df), sigma.13 =.125),inits = list(.RNG.name =
"base::Wichmann-Hill", .RNG.seed = 1989))
```

#Burn in 10000 iterations

```
 update(m,10000)
```

 # Run the model for 100,000 iterations and save the stats for prediction and indicator variable

```
x <- coda.samples(m, c(”psi.tilde.13”,”ind”), n.iter=100000)
```

amana4@illinois.edu

JAGS model

```
model {
  for (j in 1:length(psihat)) {
    psihat[j] ~ dnorm(psi[j],1/sigma[j]^2)
    psi[j] ~ dnorm(psi0,1/sigmasq0)
  }

  psi0 ~ dnorm(0,1000^2)
  sigma0 ~ dunif(0,1000)
  sigmasq0 <- sigma0^2

  psi.tilde.13 ~ dnorm(psi.13, 1/sigma.13^2)
  psi.13 ~ dnorm(psi0, 1/sigma0^2)
  ind <- psi.tilde.13 >2*sigma.13
}
```

iii)
Summary(x)

```
Iterations = 11001:111000
Thinning interval = 1
Number of chains = 1
Sample size per chain = 1e+05

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

                  Mean     SD Naive SE Time-series SE
ind          0.2480400 0.4319 0.001366       0.001547
psi.tilde.13 -0.0008291 0.4276 0.001352       0.001352

2. Quantiles for each variable:

               2.5%     25%       50%    75%  97.5%
ind          0.0000  0.0000 0.0000000 0.0000 1.0000
psi.tilde.13 -0.8754 -0.2471 0.0007902 0.2477 0.8646
```

Posterior mean of psi.tilde.13 = -0.0008291
Posterior SD of psi.tilde.13 =0.4276
95% posterior central interval of psi.tilde.13 = (-0.8754, 0.8646)

iv)

1. Empirical mean and standard deviation for each variable,
   plus standard error of the mean:

	Mean	SD	Naive SE	Time-series SE
ind	0.2480400	0.4319	0.001366	0.001547

Mean of indicator variable is the posterior probability that the predicted log odds ratio is greater than twice its standard error (i.e. will find statistically significant result) **=** 0.2480400