Homework 1

Peter Tran

BST 5610

Spatial Epidemiology and Disease Mapping

Homework 1

Due 11:59 PM, October 17

1. Suppose that a test for a disease has three possible outcomes: positive, negative and inconclusive. If a person has the disease, the probability of testing positive is 0.90, the probability of getting an inconclusive test is 0.06, and the probability of testing negative is 0.04. On the other hand, if a person does not have the disease, the probability of testing positive is 0.05, the probability of the test being inconclusive is 0.15, and the probability of testing negative is 0.80. The base rate of the disease in the population is 0.025. A person is selected at random from the population and tested for the disease

$$
\begin{aligned}
D &= \text{Has disease} \\
H &= \text{Does not have disease} \\
P &= \text{Tests positive} \\
N &= \text{Tests negative} \\
I &= \text{Tests inconclusive}
\end{aligned}
$$

$$
\begin{aligned}
P(P|D) &= 0.90 \\
P(I|D) &= 0.06 \\
P(N|D) &= 0.04 \\
P(P|H) &= 0.05 \\
P(I|H) &= 0.15 \\
P(N|H) &= 0.80 \\
P(D) &= 0.025
\end{aligned}
$$

a. What is the probability that the person tests positive.

$$
\begin{aligned}
P(H) &= 1 - P(D) \\
&= 1 - 0.025 \\
&= 0.975
\end{aligned}
$$

By the law of total probability

$$P(P) = P(P|D)P(D) + P(P|H)P(H)$$
$$= 0.90 \cdot 0.025 + 0.05 \cdot 0.975$$
$$= 0.07125$$

b. What is the probability that a person who tests positive actually has the disease?

By Baye's theorem

$$P(D|P) = \frac{P(P|D)P(D)}{P(P)}$$
$$= \frac{0.90 \cdot 0.025}{0.07125}$$
$$= \frac{0.0225}{0.07125}$$
$$\approx 0.316$$

c. What is the probability that a person who has an inconclusive test actually has the disease?

$$P(D|I) = \frac{P(I|D)P(D)}{P(I)}$$
$$= \frac{0.06 \cdot 0.025}{P(I|D)P(D) + P(I|H)P(H)}$$
$$= \frac{0.0015}{0.06 \cdot 0.025 + 0.15 \cdot 0.975}$$
$$= \frac{0.0015}{14775}$$
$$\approx 0.010$$

d. What is the probability that a person who tests negative does not have the disease?

$$P(H|N) = \frac{P(N|H)P(H)}{P(N)}$$
$$= \frac{0.80 \cdot 0.975}{P(N|D)P(D) + P(N|H)P(H)}$$
$$= \frac{0.78}{0.04 \cdot 0.025 + 0.80 \cdot 0.975}$$
$$= \frac{0.78}{0.781}$$
$$\approx 0.999$$

2. A sample of size 15 is selected from some population. The BMI of the person is determined for each of these 15 people. Assume that the measurements are

28, 32, 21, 23, 22, 25, 28, 19, 26, 27, 34, 22, 18, 26, 28

and that they represent a sample from the $N(\mu, \sigma^2)$ distribution. Use the posterior mean to estimate the population mean μ. Use the HDI (highest density interval) to obtain a 95% credible interval for μ.

a. Suppose the prior is the normal-inverse chi-square distribution with

mu0 = 29

sigma0 = 6

phi0 = sigma0^2

kappa0 = 3

nu0 = 2

Use the analytic method we called Bayesian updating to obtain the point estimate (the posterior mean) and the 95% HDI credible interval for μ.

$$\bar{X} = 25.26667$$
$$n = 15$$

$$
\begin{aligned}
\hat{\mu} &= \frac{\kappa_0}{\kappa_0 + n}\mu_0 + \frac{n}{\kappa_0 + n}\bar{X} \\
&= \frac{3}{3 + 15}29 + \frac{15}{3 + 15}25.26667 \\
&\approx 25.889
\end{aligned}
$$

I have no clue how to compute the HDI to obtain a credible interval analytically.

b. Suppose we choose independent priors for μ and $\phi = 2$ where $\mu \sim \mathrm{N}(25, 5^2)$ and $\phi \sim \mathrm{GAM}(3, 0.075)$. Here 3 is the shape parameter and 0.075 is the rate parameter. Use nimble to perform the Markov chain Monte Carlo simulations. Justify your choice of a burn-in period. Give a point estimate for and give an approximate 95% HDI credible interval for μ.

A prior of 26.55 for μ was selected as that is roughly the mean BMI of the US population.

A burn-in period of 2000 was selected by examining the posterior sampling distribution for steady state and reconfiguring the burn-in period.

```
library(nimble)
```

```
## nimble version 0.12.2 is loaded.
## For more information on NIMBLE and a User Manual,
## please visit https://R-nimble.org.

##
## Attaching package: 'nimble'

## The following object is masked from 'package:stats':
##
##     simulate
```

```
bmi_code <- nimbleCode({
  for (i in 1:N) {
    x[i] ~ dnorm(mu, phi)
  }
  mu ~ dnorm(25, 5^2)
  phi ~ dgamma(3, 0.075)
})

bmi_data <- list(x = c(28, 32, 21, 23, 22, 25, 28, 19, 26, 27, 34, 22, 18, 26, 28))
```

```r
bmi_consts <- list(N = length(bmi_data[[1]]))

bmi_inits <- list(mu = 26.55, phi = 2)

bmi_model <- nimbleModel(bmi_code, data = bmi_data, constants = bmi_consts,
                         inits = bmi_inits)
```

Defining model

Building model

Setting data and initial values

Running calculate on model
[Note] Any error reports that follow may simply reflect missing values in model variables.

Checking model sizes and dimensions

```r
bmi_compile <- compileNimble(bmi_model)
```

Compiling
[Note] This may take a minute.
[Note] Use 'showCompilerOutput = TRUE' to see C++ compilation details.

```r
bmi_conf <- configureMCMC(bmi_compile, print = T)
```

===== Monitors =====
thin = 1: mu, phi
===== Samplers =====
conjugate sampler (2)
- mu
- phi

```r
bmi_conf$addMonitors(c("mu", "phi"))
```

thin = 1: mu, phi

```r
bmi_mcmc <- buildMCMC(bmi_conf)
bmi_mcmc_compile <- compileNimble(bmi_mcmc, project = bmi_model)
```

Compiling
[Note] This may take a minute.
[Note] Use 'showCompilerOutput = TRUE' to see C++ compilation details.

```r
bmi_samples <- runMCMC(bmi_mcmc_compile, niter = 10000, nburnin = 2000,
                       inits = bmi_inits, nchains = 1, samplesAsCodaMCMC = T)
```
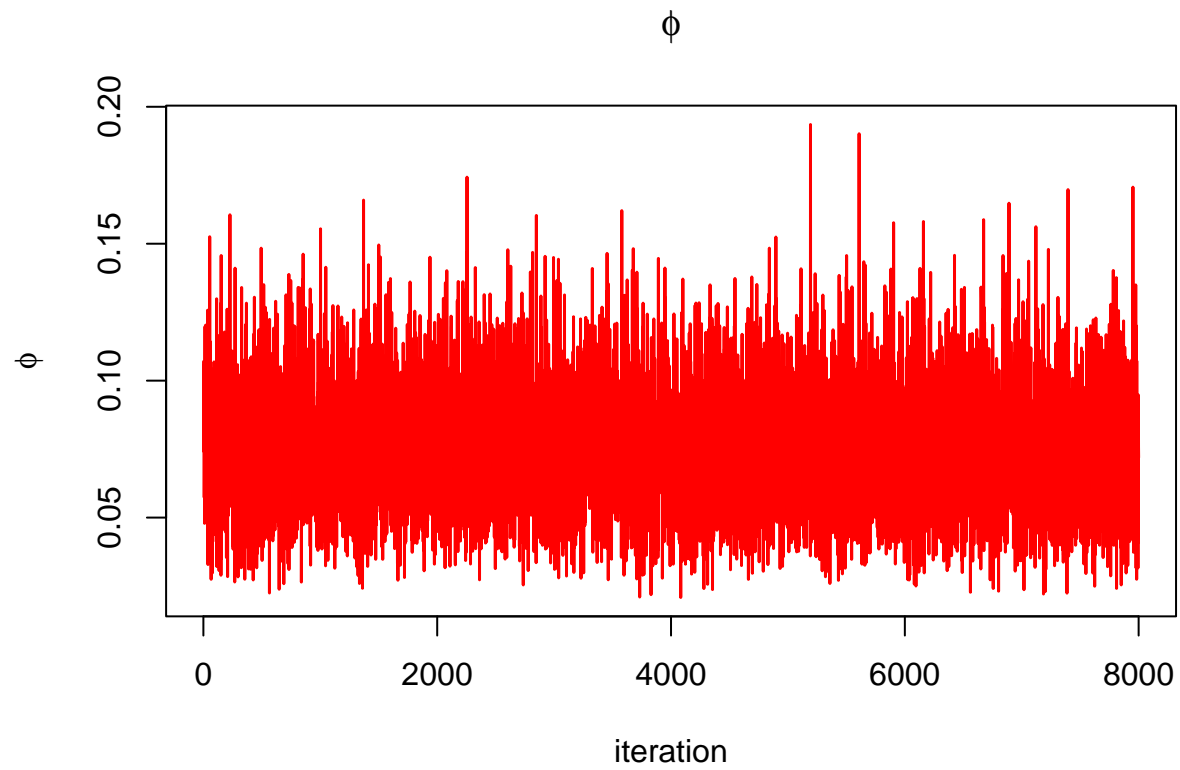
Running chain 1 ...

|-------------|-------------|-------------|-------------|
|---|

Posterior sampling distribution for μ.

```r
ts.plot(bmi_samples[,"mu"], xlab = "iteration", col = "red", lwd = 1.5,
        ylab = expression(mu), main = expression(mu))
```

μ
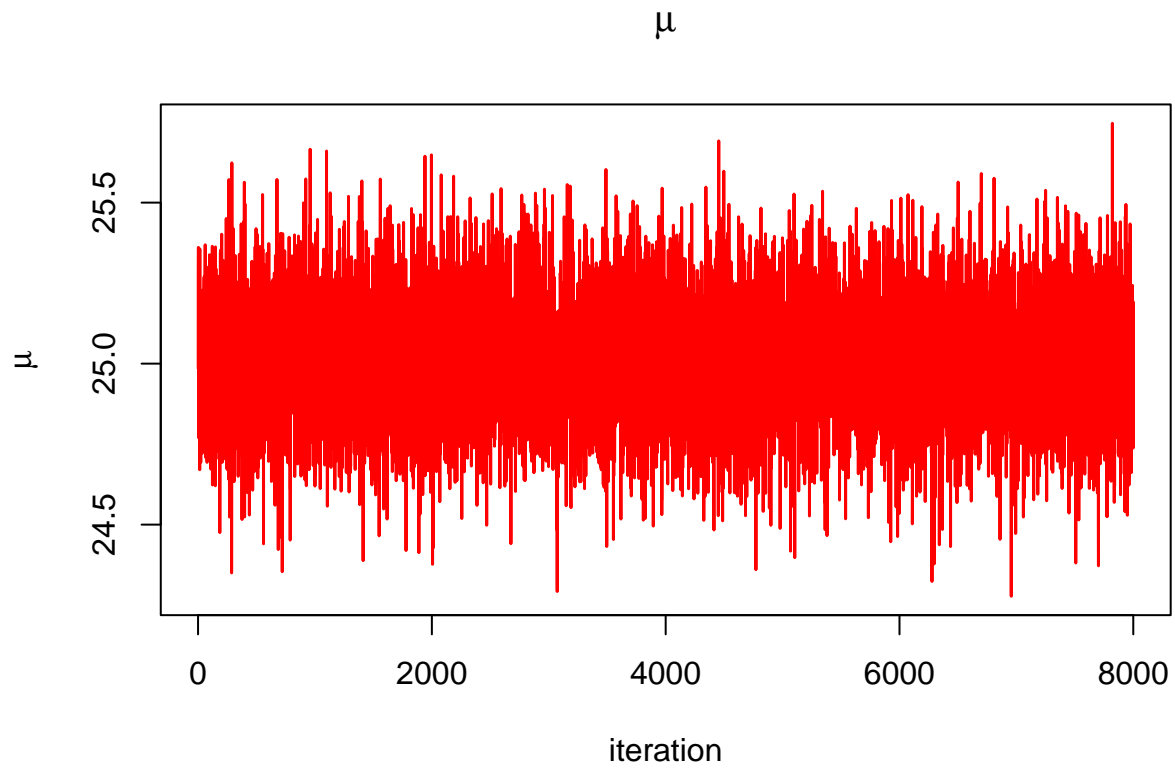
Posterior sampling distribution for ϕ.

```
ts.plot(bmi_samples[,"phi"], xlab = "iteration", col = "red", lwd = 1.5,
        ylab = expression(phi), main = expression(phi))
```


ϕ

Mean point estimate for μ.

```r
mean(bmi_samples[,"mu"])
```

```
## [1] 25.00903
```

95% HDI credible interval for μ.

```r
source("./class_code/HDIofMCMC.R")

HDIofMCMC(bmi_samples[,"mu"])
```

```
## [1] 24.63320 25.39084
```