Assignment 6

AUTHOR
Joaquin Peraza

Data model:

$$[y_{ij}|z_i, a_j] = \begin{cases} 0 \text{ if } z_i = 0, \\ Bern(a_j) \text{ if } z_i = 1, \end{cases}$$

Process model:

$$[Zi|p] = Bern(p)$$

1.

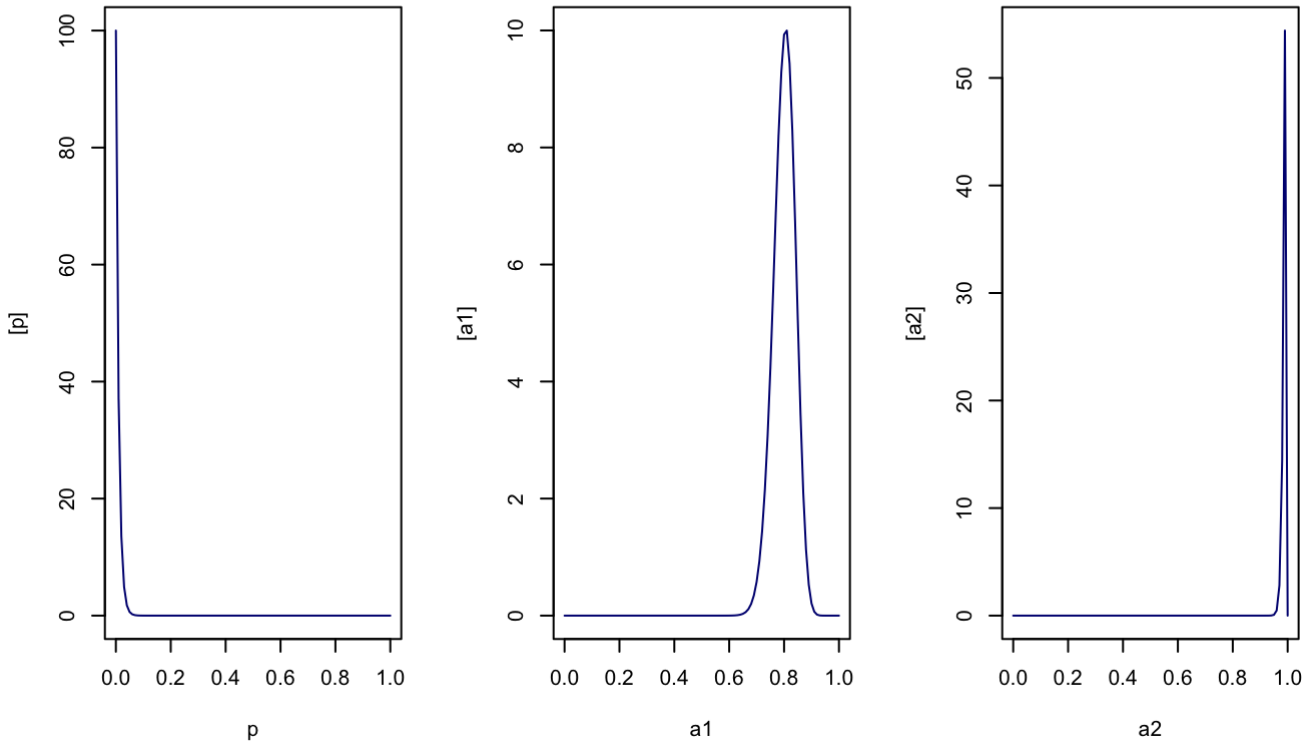
Priors

$$[p] = Beta(\alpha_p, \beta_p) \;\; \alpha_p = 1, \beta_p = 100$$

$$[a1] = Beta(\alpha_{a1}, \beta_{a1}) \;\; \alpha_p = 80, \beta_p = 20$$

$$[a2] = Beta(\alpha_{a2}, \beta_{a2}) \;\; \alpha_p = 22, \beta_p = 200$$

```
# Hyperparameters for priors
alpha.p <- 1
beta.p <- 100

alpha.a1 <- 80
beta.a1 <- 20

alpha.a2 <- 200
beta.a2 <- 2

par(mfrow = c(1, 3), mar = c(5, 5, 8, 1))
plot(seq(0,1,0.01), dbeta(seq(0,1,0.01),alpha.p,beta.p), type = "l",
     col = "navy", xlab = "p", ylab = "[p]")
plot(seq(0,1,0.01), dbeta(seq(0,1,0.01),alpha.a1,beta.a1), type = "l",
     col = "navy", xlab = "a1", ylab = "[a1]")
plot(seq(0,1,0.01), dbeta(seq(0,1,0.01),alpha.a2,beta.a2), type = "l",
     col = "navy", xlab = "a2", ylab = "[a2]")
```




```
# Test how reasonable is [a1] (Rapid test true positive rate) prior
a1.prior <- rbeta(1e5,alpha.a1,beta.a1)
mean(a1.prior)
```

```
[1] 0.7997552
```

```
HDInterval::hdi(a1.prior, credMass = 0.95)
```

```
   lower    upper
0.721404 0.876135
attr(,"credMass")
[1] 0.95
```

```
# Test how reasonable is [a2] (PCR true positive rate) prior
a2.prior <- rbeta(1e5,alpha.a2,beta.a2)
mean(a2.prior)
```

```
[1] 0.9900829
```

```
HDInterval::hdi(a2.prior, credMass = 0.95)
```

```
    lower     upper
0.9765010 0.9997431
attr(,"credMass")
[1] 0.95
```

```r
# Test how reasonable is [p] prior
p.prior <- rbeta(1e5,alpha.p,beta.p)
mean(p.prior)
```

```
[1] 0.009904964
```

```r
HDInterval::hdi(p.prior, credMass = 0.95)
```

```
       lower        upper
2.975577e-09 2.964842e-02
attr(,"credMass")
[1] 0.95
```

The priors used for the true positive rates of the rapid test and PCR test seem reasonable for the rabies data example. The expected values of the proposed beta distributions are very close to the true positive rates given by the manufacturer, indicating that our prior beliefs are consistent with external information. The 95% credible intervals also encompass the known true positive rates, suggesting that the priors are not too narrow or too wide. Also the expected value and HDI of the prevalence rate matches with the literature review done in the previous assignment.

2.

1. Choose starting values for the parameters: p, z1, a1, and a2.
2. For each iteration of the algorithm:
    - Sample a new value for a1 from the posterior distribution $[a1|p, z1, a2, y1, y2]$ using Metropolis-Hastings.
    - Sample a new value for a2 from the posterior distribution $[a2|p, z1, a1, y1, y2]$ using Metropolis-Hastings
    - Sample a new value for p from the full conditional $[p|z1, a1, a2, y1, y2]$.
    - Take one draw from the full conditional of z1 $[z1|p, a1, a2, y1, y2]$.
    - Store the sampled values.
3. Repeat step 2 for a large number of iterations (e.g., 1e6).
4. Discard the first (20% of K) as burn-in interval.
5. Store the remaining iterations as samples from the posterior distribution.

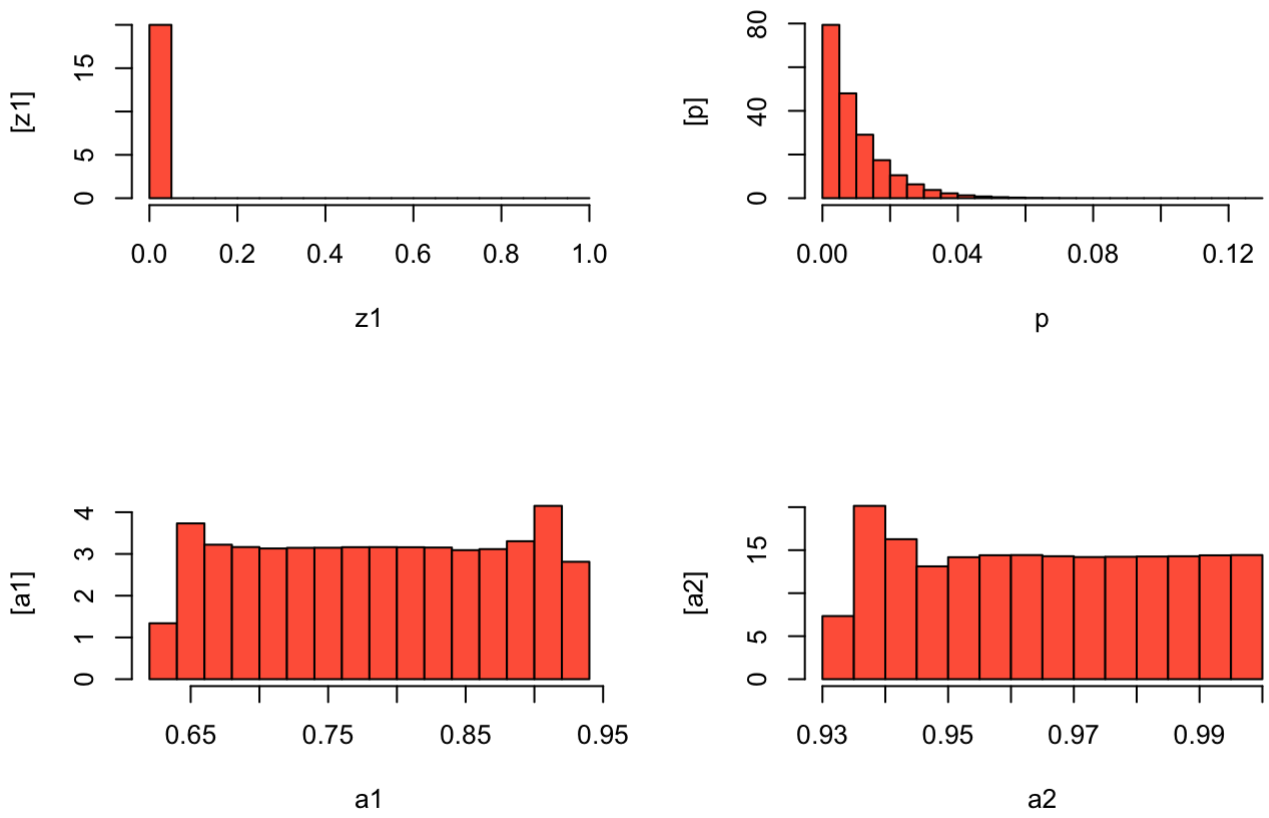
3.

```r
set.seed(1)

# Data
y1 <- 0 # Results from rapid test
y2 <- 0 # Results from PCR test


# Preliminary MCMC stuff
K <- 1e6
sigma.tune <- 0.01
samples <- matrix(NA,K,4)
colnames(samples) <- c("z1", "p", "a1", "a2")

p.initial <- rbeta(1,alpha.p, beta.p)
a1.initial <- rbeta(1,alpha.a1, beta.a1)
a2.initial <- rbeta(1,alpha.a2, beta.a2)

a1 <- a1.initial
a2 <- a2.initial
p <- p.initial
z1 <- 0
samples[1,] <- c(NA, p, a1.initial, a2.initial)

# Gibbs sampler
for(k in 2:K){

  #Metropolis-Hasting for [a1]
  a1.try <- rnorm(1,a1, sigma.tune)
  mh <- ifelse(a1.try > 0 & a1.try < 1,
               ifelse(z1==0, 1, dbinom(y1, 1, a1.try)) *
                 dbeta(a1.try, alpha.a1, beta.a1) /
                 ifelse(z1==0, 1, dbinom(y1, 1, a1)) *
                 dbeta(a1, alpha.a1, beta.a1)
               ,0)
  keep <- ifelse(mh>1,1,rbinom(1,1,mh))
  a1 <- ifelse(keep==1, a1.try, a1)

  #Metropolis-Hasting for [a2]
  a2.try <- rnorm(1,a2, sigma.tune)
  mh <- ifelse(a2.try > 0 & a2.try < 1,
               ifelse(z1==0, 1, dbinom(y2, 1, a2.try)) *
                 dbeta(a2.try, alpha.a2, beta.a2) /
                 ifelse(z1==0, 1, dbinom(y2, 1, a2)) *
                 dbeta(a2, alpha.a2, beta.a2)
               ,0)
  keep <- ifelse(mh>1,1,rbinom(1,1,mh))
  a2 <- ifelse(keep==1, a2.try, a2)

  #Gibbs Sampler
```

```
  p.tilde <- (p*(1-a1)*(1-a2))/((p*(1-a1)*(1-a2))+1-p)
  z1 <- ifelse(min(c(y1,y2))==0,rbinom(1,1,p.tilde),1)

  # Take one draw from the full conditional of p (i.e., [p|z1,y1,y2])
  p <- rbeta(1,alpha.p+z1,beta.p+1-z1)

  # Save samples
  samples[k,] <- c(z1,p,a1,a2)
}

# Discard burn-in interval
samples <- samples[round(K*0.2):K,]
```

4-7.

```
par(mfrow = c(2,2)) # Set up 2x2 grid for plots

hist(samples[, "z1"], breaks = 20, col = "tomato", xlab = "z1",
     ylab="[z1]", freq = FALSE, main="")
hist(samples[, "p"],  breaks = 20, col = "tomato", xlab = "p",
     ylab="[p]",  freq = FALSE, main="")
hist(samples[, "a1"], breaks = 20, col = "tomato", xlab = "a1",
     ylab="[a1]", freq = FALSE, main="")
hist(samples[, "a2"], breaks = 20, col = "tomato", xlab = "a2",
     ylab="[a2]", freq = FALSE, main="")
```



8.

```
# Prevalence rate posterior summaries
mean(samples[,"p"])
```

```
[1] 0.009812956
```

```
HDInterval::hdi(samples[,"p"], credMass = 0.95)
```

```
       lower        upper
5.216792e-09 2.925968e-02
attr(,"credMass")
[1] 0.95
```

```
# Bat rabies state posterior summaries
mean(samples[,"z1"])
```

```
[1] 6.374992e-05
```

```
HDInterval::hdi(samples[,"z1"], credMass = 0.95)
```

```
lower upper
    0     0
attr(,"credMass")
[1] 0.95
```

The prevalence rate (p) shows an expected value of p given all the rest of the parameters is 0.0098, indicating that the estimated proportion of infected bats in the population is low. The 95% highest density interval (HDI) for the prevalence rate is [5.22e-09, 0.029], showing a 95% credible interval within extremely low values.

For the bat rabies state, the expected value (E) of $[z1|p, a1, a2, y1, y2]$ is 6.37e-05, indicating that the probability of the actual bat rabies state being positive is very low. The 95% HDI for the bat rabies state is [0, 0], indicating less than 5% of the credible interval is outside 0.

When we have priors for a1 (accuracy of rapid tests) and a2 (accuracy of PCR tests) in addition to p, it means that we have some prior beliefs or knowledge about the possible values of a1 and a2 (true positive rate), for example, past studies, while still accounting for possibles uncertainties in this parameters, allowing more accurate estimations of [p] and [z1], thus, improve the probability of a correct numeric estimation of z1 (actual state of the tested bat).