Assignment 3 with Solutions

STAT314/STAT461

Set: Tue, Aug-24. Due: Fri Sept-17

Problem 1.

(a)

If a person has reported drinking at least one drink in the past week, then they are an alcohol drinker by definition. So,

$$Pr(z_i = 1|x_i > 0) = 1$$

For $x_i = 0$, let's use the Bayes' formula:

$$Pr(z_i = 1|x_i = 0, \omega) = \frac{Pr(x_i = 0|z_i = 1, p)Pr(z_i = 1|\omega)}{Pr(x_i = 0|z_i = 1, p)Pr(z_i = 1|\omega) + Pr(x_i = 0|z_i = 0, p)Pr(z_i = 0|\omega)}.$$

Substituting the Bernoulli likelihoods:

$$Pr(z_i = 1|x_i = 0, \omega) = \frac{(1-p)^7\omega}{(1-p)^7\omega + 1(1-\omega)}.$$

(b)

We have already run the ZiB in class. So, the code should be familiar. It is, of course, simpler for the simple Binomial model.

```
rm(list=ls())
library(R2WinBUGS)

# the zero-inflated model
my.model <- function(){

# likelihood
for(i in 1:k){
x[i] ~ dbin(p.x[i],7)
p.x[i] <- p*z[i]
z[i] ~ dbern(omega)
}

# priors
p ~ dbeta(1,1)
omega ~ dbeta(1,1)
}
```

```r
# the file I will temporarily save it too
my.file <- "p:\\temp.txt"

## write model file:
write.model(my.model,my.file)

## and let's take a look at that file:
#file.show(my.file)

# the drinkers vs. teetotallers data
x <- rep(0:6,c(22,6,18,23,18,10,3))

# running the model in WinBUGS:

system.time(
m.bugs.ZiB <- bugs(data=list(k=length(x),x=x),
      inits=list(list(p=.5,omega=.5,z=rep(1,length(x)))),
      parameters.to.save=c("p","omega","z"),
      ### PATH TO THE MODEL FILE
      model.file=my.file,
      n.chains=1,
      n.iter=10000,
      n.sim=5000,
      n.burnin=5000,
      n.thin=1,
      DIC=T,
      bugs.directory=paste0(Sys.getenv(c("USERPROFILE")), "\\WinBUGS14"),debug=F
      #bugs.directory="c:\\Program Files\\WinBUGS14\\WinBUGS14",debug=F
      )
)
```

```
##    user  system elapsed
##    0.68    0.09  241.19
```

```r
## the simple beta-binomial model:
my.model <- function(){

# likelihood
for(i in 1:k){
x[i] ~ dbin(p,7)
}

# priors
p ~ dbeta(1,1)
}

# the file I will temporarily save it too
my.file <- "p:\\temp.txt"

## write model file:
write.model(my.model,my.file)

# running the model in WinBUGS:
```

```
system.time(
m.bugs.BB <- bugs(data=list(k=length(x),x=x),
      inits=list(list(p=.5)),
      parameters.to.save=c("p"),
      ### PATH TO THE MODEL FILE
      model.file=my.file,
      n.chains=1,
      n.iter=10000,
      n.sim=5000,
      n.burnin=5000,
      n.thin=1,
      DIC=T,
      bugs.directory=paste0(Sys.getenv(c("USERPROFILE")), "\\WinBUGS14"),debug=F
      #bugs.directory="c:\\Program Files\\WinBUGS14\\WinBUGS14",debug=F
      )
)
```

```
##    user  system elapsed
##    0.02    0.03  240.17
```

And getting the DICs:

```
m.bugs.ZiB$DIC
```

```
## [1] 316.1404
```

```
m.bugs.BB$DIC
```

```
## [1] 418.617
```

The zero-inflated model has a much lower DIC, so it is a better fit statistically, then the simple beta-binomial model. (NB. your numbers may be slightly different from mine due to random number generators)

Problem 2.

(a) The probability that you have a two-sided coin given the observation of two heads can be evaluated using Bayes' formula as follows:

$$\begin{aligned} Pr(\text{Two-sided}|HH) &= \frac{Pr(HH|\text{Two-sided})Pr(\text{Two-sided})}{Pr(HH|\text{Two-sided})Pr(\text{Two-sided}) + Pr(HH|\text{Fair})Pr(\text{Fair})} \\ &= \frac{1 * 1/2}{1 * 1/2 + 1/4 * 1/2} \\ &= \frac{1/2}{5/8} = 4/5 = 0.80 \end{aligned} \tag{1}$$

(b) How many consecutive "heads" do you need to observe to be at least 90% certain that the coin is two-sided? **1pt**

Let's generalise the above to k consecutive heads:

$$
\begin{aligned}
Pr(\text{Two-sided}|k\text{ Heads}) &= \frac{Pr(HH|k\text{ Heads})Pr(\text{Two-sided})}{Pr(k\text{ Heads}|\text{Two-sided})Pr(\text{Two-sided}) + Pr(k\text{ Heads}|\text{Fair})Pr(\text{Fair})} \\
&= \frac{1*1/2}{1*1/2+(1/2)^k*1/2} \\
&= \frac{1}{1+(1/2)^k} = \frac{2^k}{2^k+1}
\end{aligned} \tag{2}
$$

Now, let's solve for k the following:

$$\frac{2^k}{2^k+1} \geq 0.90$$

Multiplying both sides by 2^k+1 and doing some rearranging:

$$
\begin{aligned}
2^k &\geq (2^k+1)0.90 \\
2^k * 0.10 &\geq 0.90 \\
2^k &\geq 9 \\
k\log 2 &\geq \log 9 \\
k &\geq \frac{\log 9}{\log 2} = 3.17
\end{aligned}
$$

So, we need at least four consecutive heads to conclude with at least 90% certainty that the coin is one-sided.

(c) Consider the following two hypotheses:

- H_0: the coin if fair.

and

- H_A: the coin is not fair (automatically assumed as the opposite of H_0 by default)

Let $y = 2$ denote the observed number of heads. Given that the coin is fair, 2 heads is as extreme as 2 tails. The $p - value$ is defined as the probability of observing the data at least as extreme as the actual observation, given that the null hypothesis is correct. In other words,

$$p = Pr(\text{HH or TT}|\text{Fair}) = 1/4 + 1/4 = 1/2 = 0.5.$$

Not enough to reject the null hypothesis of "fair coin". Note, that this set-up does not take into account the alternative hypothesis at all, but I accepted it as a default solution.