

Homework 3

Megan Robertson

Monday, September 14, 2015

Consider the setting of the Perchlorate Example.

1. Write down simpler expressions for $\Pr(M=1|data)$ and Bayes factor.

$$\mathbb{L}(data|M=0) = \frac{c(a,b)}{c(\tilde{a},\tilde{b})} = \frac{\frac{1}{\beta(a,b)}}{\frac{1}{\beta(\tilde{a},\tilde{b})}} = \frac{\beta(\tilde{a},\tilde{b})}{\beta(a,b)} = \frac{\frac{\Gamma(\tilde{a})\Gamma(\tilde{b})}{\Gamma(\tilde{a}+\tilde{b})}}{\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}} = \frac{\Gamma(\tilde{a})\Gamma(\tilde{b})\Gamma(a+b)}{\Gamma(\tilde{a}+\tilde{b})\Gamma(a)\Gamma(b)} = \frac{\Gamma(a+2)\Gamma(b+64)\Gamma(a+b)}{\Gamma(a+b+64)\Gamma(a)\Gamma(b)} = A$$

$$\begin{aligned} \mathbb{L}(data|M=1) &= \left(\frac{c(c,d)}{c(\tilde{c},\tilde{d})}\right)\left(\frac{c(e,f)}{c(\tilde{e},\tilde{f})}\right) = \left(\frac{\frac{1}{\beta(c,d)}}{\frac{1}{\beta(\tilde{c},\tilde{d})}}\right)\left(\frac{\frac{1}{\beta(e,f)}}{\frac{1}{\beta(\tilde{e},\tilde{f})}}\right) = \left(\frac{\beta(\tilde{c},\tilde{d})}{\beta(c,d)}\right)\left(\frac{\beta(\tilde{e},\tilde{f})}{\beta(e,f)}\right) = \left(\frac{\frac{\Gamma(\tilde{c})\Gamma(\tilde{d})}{\Gamma(\tilde{c}+\tilde{d})}}{\frac{\Gamma(c)\Gamma(d)}{\Gamma(c+d)}}\right)\left(\frac{\frac{\Gamma(\tilde{e})\Gamma(\tilde{f})}{\Gamma(\tilde{e}+\tilde{f})}}{\frac{\Gamma(e)\Gamma(f)}{\Gamma(e+f)}}\right) = \\ &= \frac{\Gamma(\tilde{c})\Gamma(\tilde{d})\Gamma(c+d)\Gamma(\tilde{e})\Gamma(\tilde{f})\Gamma(e+f)}{\Gamma(\tilde{c}+\tilde{d})\Gamma(c)\Gamma(d)\Gamma(\tilde{e}+\tilde{f})\Gamma(e)\Gamma(f)} = \frac{\Gamma(c)\Gamma(d+30)\Gamma(c+d)\Gamma(e+2)\Gamma(f+28)\Gamma(e+f)}{\Gamma(c+d+30)\Gamma(c)\Gamma(d)\Gamma(e+f+30)\Gamma(e)\Gamma(f)} = B \end{aligned}$$

Define F to be the Bayes factor. Therefore,

$$F = \frac{A}{B}$$

$$\Pr(M=1|data) = \frac{1}{1+F} = \frac{1}{1+\frac{A}{B}}$$

2. Trying different priors report the $\Pr(M=1|data)$ and Bayes factor for perchlorate example compared with Fisher's exact test. How do the priors make a difference in the probability and the Bayes factor? Try at least two different priors

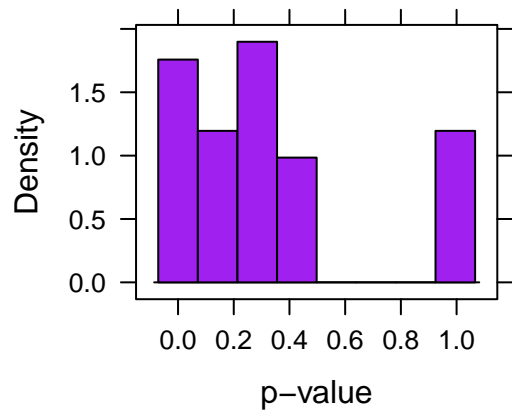
The first set of priors take into account the historical data that is available in this experiment. These priors were $\theta \sim \beta(1, 450)$, $\theta_1 \sim \beta(1, 450)$ and $\theta_2 \sim \beta(1, 50)$, where θ is the parameter under the null hypothesis, θ_1 is the parameter for the control group under the alternative hypothesis, and θ_2 is the parameter for the treatment group under the alternative hypothesis. The Bayes factor calculated using these priors is 0.03588651 and the $\Pr(M=1|data) = 0.9653567$.

The second set of priors chosen were uninformative. The priors for all three parameters were set to be $\beta(1/2, 1/2)$. The Bayes factor calculated using these priors is 1.66851 and the $\Pr(M=1|data) = 0.374741$. The choice of priors greatly affects the calculation of the Bayes factor and thus the $\Pr(M=1|data)$. The priors chosen using the historical data resulted in a much higher probability than the calculations made using the uninformative priors. This occurs because the historical priors assume that there is a greater chance that a rat exposed to perchlorate develops a tumor.

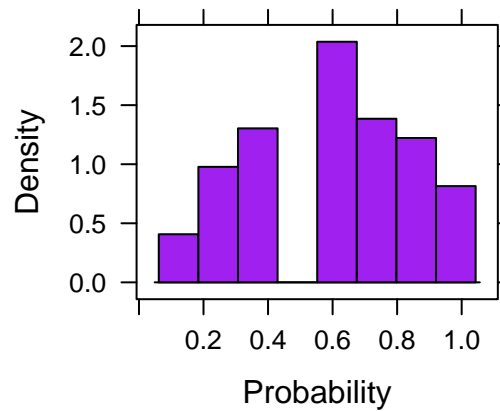
Fisher's exact test resulted in a p-value of 0.4915. Therefore, according to this test there is not evidence in support of the alternative hypothesis that being exposed to perchlorate causes tumors in rats. This matches the conclusion of the Bayes hypothesis testing (assuming there is a 0-1 loss function).

3. Run a simulation study under H_0 and H_1 . Compare the Bayes and Frequentist results. The first simulation run assumes that the alternative hypothesis is true and thus that being exposed to perchlorate causes tumors in rats. The frequentist test correctly rejected the null hypothesis ten times. The Bayes factors calculated for this scenario ranged from 0.01455 to 6.90800 and the mean was 1.32000. The values for $\Pr(M=1|data)$ ranged from 0.1265 to 0.9857 and the mean was 0.5786. Assuming a 0-1 loss function, the Bayes test correctly rejected the null hypothesis 67 times. The Bayes test has a power of 0.67 compared to 0.1 for Fisher's exact test. Thus, the Bayes hypothesis testing performed better than the frequentist test when the alternative hypothesis was true.

**Fisher's Exact Test p-values
Alternative True**

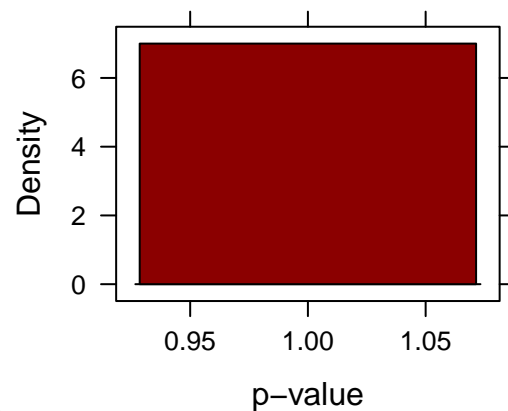


**$\Pr(M=1|\text{data})$
Alternative True**

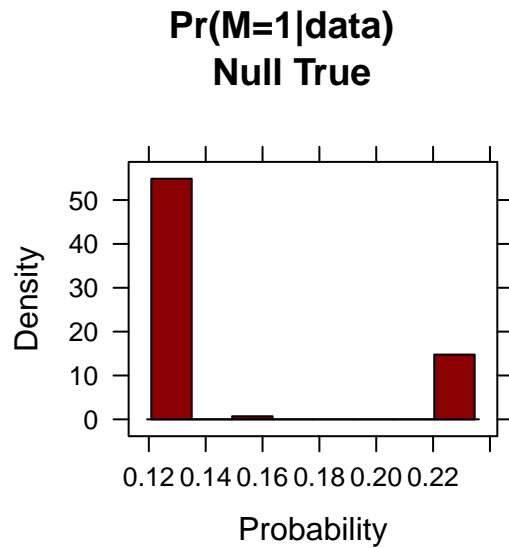


The second simulation assumes that the null hypothesis is true. Each p-value calculated using Fisher's exact test for this simulation was greater than 0.05, so each time the test correctly showed that there was not evidence in favor of the alternative hypothesis. There was no Type I error in this situation. The range of the Bayes Factors calculated for this scenario was from 3.425 to 6.908 and had an average of 6.159. The probabilities ranged from 0.1265 to 0.2260 and averaged 0.1477. Thus, assuming a 0-1 loss function, the test does not reject

**Fisher's Exact Test P-values
Null True**



the null hypothesis since all of the probabilities are less than 0.5.



I worked with multiple peers on this assignment and looked up things online.

Code Section

Code for Problem 2

```
#priors based on historical data
a=1; b=450; c=1; d=450; e=1; f=50
likelihood.null= beta(a+2, b+60-2)/beta(a,b)
likelihood.alt.one = beta(c+0, d+30)/beta(c,d); likelihood.alt.two = beta(e+2, f+28)/beta(e,f)
likelihood.alt = likelihood.alt.one*likelihood.alt.two

Bayes.factor.1 = likelihood.null/likelihood.alt
Prob.1 = 1/(1+Bayes.factor.1)

#uninformative priors
a=1/2; b=1/2; c=1/2; d=1/2; e=1/2; f=1/2
likelihood.null.2= beta(a+2, b+60-2)/beta(a,b)
likelihood.alt.one.2 = beta(c+0, d+30)/beta(c,d); likelihood.alt.two.2 = beta(e+2, f+28)/beta(e,f)
likelihood.alt.2 = likelihood.alt.one.2*likelihood.alt.two.2

Bayes.factor.2 = likelihood.null.2/likelihood.alt.2
Prob.2 = 1/(1+Bayes.factor.2)

#Fisher's Exact Test
control.vector = c(0, 30); treatment.vector = c(2, 28); freq.tab =
  rbind(control.vector, treatment.vector)
fisher.test(freq.tab)

##
## Fisher's Exact Test for Count Data
##
## data: freq.tab
## p-value = 0.4915
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
```

```
## 0.000000 5.295628
## sample estimates:
## odds ratio
## 0
```

Code for Problem 3

```
#simulation assuming the alternative is true
set.seed(92)
null.sim.control = rbinom(n=100, size=30, p=1/500) #vector with the number of successes for control
null.sim.treatment = rbinom(n=100, size=30, p=1/10) #vector with the number of successes for treatment

#frequentist
pvalue.vectors.freq.alt.true=c() #create vector for the p-values
for (i in 1:length(null.sim.control)){
  control.row = cbind(null.sim.control[i], 30-null.sim.control[i])
  exposed.row = cbind(null.sim.treatment[i], 30-null.sim.treatment[i])
  contingency.table = rbind(control.row, exposed.row)
  pvalue.vectors.freq.alt.true = append(pvalue.vectors.freq.alt.true,
    fisher.test(contingency.table)$p.value)
}
length(which(pvalue.vectors.freq.alt.true<=0.05)) #how many correctly rejected
```

```
## [1] 10

a=1/2; b=1/2; c=1/2; d=1/2; e=1/2; f=1/2
BF.alt.true=c(); probabilities.Bayes.alt.true=c()
for (i in 1:length(null.sim.control)){
  likelihood.null = beta(a+null.sim.control[i]+null.sim.treatment[i], b+30+30-
    (null.sim.control[i]+null.sim.treatment[i]))/beta(a,b)
  likelihood.alt.one = beta(c+null.sim.control[i], d+30-(null.sim.control[i]))/beta(c,d)
  likelihood.alt.two = beta(e+null.sim.treatment[i], f+30-null.sim.treatment[i])/beta(e,f)
  likelihood.alt = likelihood.alt.one*likelihood.alt.two
  next.Bayes = likelihood.null/likelihood.alt
  BF.alt.true = append(BF.alt.true, next.Bayes)
  next.prob = 1/(1+next.Bayes)
  probabilities.Bayes.alt.true = append(probabilities.Bayes.alt.true, next.prob)
}
summary(BF.alt.true); summary(probabilities.Bayes.alt.true)
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.01455 0.32370 0.79798 1.32029 1.66851 6.90772

## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 0.1265 0.3747 0.5562 0.5786 0.7591 0.9857

length(which(probabilities.Bayes.alt.true>=0.5))
```

```
## [1] 67

#simulation assuming that the null is true
set.seed(5)
null.sim.control = rbinom(n=100, size=30, p=1/300) #vector with the number of successes
null.sim.treatment = rbinom(n=100, size=30, p=1/300)

#frequentist
pvalue.vectors=c()
for (i in 1:length(null.sim.control)){
```

```

control.row = cbind(null.sim.control[i], 30-null.sim.control[i])
exposed.row = cbind(null.sim.treatment[i], 30-null.sim.treatment[i])
contingency.table = rbind(control.row, exposed.row)
pvalue.vectors = append(pvalue.vectors, fisher.test(contingency.table)$p.value)
}
which(pvalue.vectors<=0.05)

```

```
## integer(0)
```

```
#Bayesian
```

```
a=1/2; b=1/2; c=1/2; d=1/2; e=1/2; f=1/2
```

```
BF=c(); probabilities=c()
```

```
for (i in 1:length(null.sim.control)){
```

```
  likelihood.null = beta(a+null.sim.control[i]+null.sim.treatment[i], b+30+30-
    (null.sim.control[i]+null.sim.treatment[i]))/beta(a,b)
```

```
  likelihood.alt.one = beta(c+null.sim.control[i], d+30-(null.sim.control[i]))/beta(c,d)
```

```
  likelihood.alt.two = beta(e+null.sim.treatment[i], f+30-null.sim.treatment[i])/beta(e,f)
```

```
  likelihood.alt = likelihood.alt.one*likelihood.alt.two
```

```
  next.Bayes = likelihood.null/likelihood.alt
```

```
  BF = append(BF, next.Bayes)
```

```
  next.prob = 1/(1+next.Bayes)
```

```
  probabilities = append(probabilities, next.prob)
```

```
}
```

```
summary(BF); summary(probabilities)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      3.425   6.908   6.908   6.159   6.908   6.908

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.1265  0.1265  0.1265  0.1477  0.1265  0.2260
```