STAT431_Homework4

Giang Le

10/21/2021

Problem 1

(a) Write all the mathematical formulas for the integrals I will compute:

$$Var(\theta|y) = E(\theta^2|y) - E(\theta|y)^2 = \int_0^1 \theta^2 p(\theta|y)\, d\theta - (\int_0^1 \theta p(\theta|y)\, d\theta)^2 = \int_0^1 \theta^2 \frac{f(y|\theta)\pi(\theta)}{m(y)}\, d\theta - (\int_0^1 \theta \frac{f(y|\theta)\pi(\theta)}{m(y)}\, d\theta)^2$$

where m(y) is

$$\int_0^1 f(y|\theta)\pi(\theta)\, d\theta$$

(b) I use integrate() in R to approximate the integrals above numerically. The Jeffreys' prior for a binomial model is Beta(.5, .5)

```
n <- 70

y <- 12
### Define likelihood and prior (up to proportionality)

like <- function(theta) theta^y * (1-theta)^(n-y)

prior <- function(theta) dbeta(theta, 0.5, 0.5)

### Compute posterior expectation of theta^2

numerator2 <- integrate(function(theta) theta^2 * prior(theta) * like(theta), 0, 1)
numerator2
```

```
## 3.696948e-17 with absolute error < 4.5e-18
```

```
denominator <- integrate(function(theta) prior(theta) * like(theta), 0, 1)
denominator
```

```
## 1.119929e-15 with absolute error < 6.2e-17
```

```
expect2 <- (numerator2$value / denominator$value)

### Compute the posterior expectation of theta

numerator <- integrate(function(theta) theta * prior(theta) * like(theta), 0, 1)
numerator
```

```
## 1.971706e-16 with absolute error < 6.9e-18
```

```r
denominator <- integrate(function(theta) prior(theta) * like(theta), 0, 1)
denominator
```

```
## 1.119929e-15 with absolute error < 6.2e-17
```

```r
expect <- (numerator$value / denominator$value)

### Compute the posterior variance of theta

expect2 - expect^2
```

```
## [1] 0.002014729
```

(c) To estimate the posterior variance of theta analytically, I use the formula of variance of the Beta posterior distribution where $\alpha = 12.5$ and $\beta = 58.5$

$$Var(\theta|y) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

```r
alpha <- 12.5
beta <- 58.5

### Applying the formula
var <- (alpha * beta)/((alpha + beta + 1)*(alpha + beta)^2)
var
```

```
## [1] 0.002014729
```

The result up to 9 decimal places is the same as what we got from using numerical integration in R.

Problem 2

(a) Entering the data for the small cars and SUV categories to compute their sample means and standard deviations.

```r
small <- c(133, 114, 113, 113, 111, 108, 108, 104)
suv <- c(129, 125, 120, 111, 105, 99,  97, 97, 97, 86)
n1 <- length(small)
n2 <- length(suv)
ybar1 <- mean(small)
ybar2 <- mean(suv)
s1 <- sd(small)
s2 <- sd(suv)

ybar1
```

```
## [1] 113
```

```r
ybar2
```

```
## [1] 106.6
```

```r
s1
```

```
## [1] 8.75051
```

```r
s2
```

```
## [1] 14.14371
```

(b) Compute an approximate 95% equal-tailed credible interval for the difference between the mean for small cars and the mean for SUVs. Do the means appear to differ?

```
### Independent draws from posteriors
set.seed(777)
Nsim <- 100000

sigma1.2s <- 1 / rgamma(Nsim, (n1-1)/2, (n1-1)*s1^2/2)
sigma2.2s <- 1 / rgamma(Nsim, (n2-1)/2, (n2-1)*s2^2/2)

mu1s <- rnorm(Nsim, ybar1, sqrt(sigma1.2s/n1))
mu2s <- rnorm(Nsim, ybar2, sqrt(sigma2.2s/n2))

### 95% equal-tailed credible interval for mu1 - mu2
quantile(mu1s - mu2s, c(0.025, 0.975))
```

```
##      2.5%     97.5%
## -6.078377 18.894078
```

So there is a 95% probability that the true mu1 - mu2 falls within this interval. Because the interval contains 0, the means don't appear to differ.

(c)

```
# approx. posterior probability that mu1 does not exceed mu2
mean(mu1s <= mu2s)
```

```
## [1] 0.14327
```

The posterior probability that mu1 does not exceed mu2 is roughly 14%, high enough to say that we do not reject the hypothesis that mu1 does not exceed mu2.

(d) Compute the (frequentist) Welch two sample t-test one-sided p-value for testing the null hypothesis that the mean for small cars does not exceed the mean for SUVs. (Use R function t.test(. . . , . . . , alternative=. . . , var.equal=FALSE), making sure to select the correct alternative.) Also compute the usual (frequentist) two sample t-test one-sided p-value that assumes equal variances (t.test(. . . , . . . , alternative=. . . , var.equal=TRUE)). Compare with the Bayesian probability of the previous part.

```
t.test(small, suv, alternative="greater", var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  small and suv
## t = 1.1768, df = 15.199, p-value = 0.1287
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.125544       Inf
## sample estimates:
## mean of x mean of y
##     113.0     106.6
```

```
t.test(small, suv, alternative="greater", var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  small and suv
## t = 1.1165, df = 16, p-value = 0.1403
```

```
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.607375       Inf
## sample estimates:
## mean of x mean of y
##     113.0     106.6
```

The p-values given by the Welch two sample t-test one-sided p-value for testing the null hypothesis that the mean for small cars does not exceed the mean for SUVs are about the same as the Bayesian probability obtained from the posterior probability when the variances are assumed to be equal and less than the Bayesian probability when variances are assumed not to be equal. Under the t-test (whether or not variances are equal), we fails to reject the null hypothesis that the mean for small cars does not exceed the mean for SUVs. We have evidence that the mean for small cars does not exceed the mean for SUVs. This is consistent with we have above under Bayesian inference based on posterior distribution.

(e) Compute an approximate 95% equal-tailed credible interval for the ratio of the variance for small cars to the variance for SUVs. Do the variances appear to differ?

```r
### approx. 95% credible interval for sigma1^2/sigma2^2
quantile(sigma1.2s / sigma2.2s, c(0.025, 0.975))
```

```
##       2.5%      97.5%
## 0.09166248 1.83718076
```

The variances do not appear to differ. The quantile contains 1.

Problem 3

(a)

With known ν

$$f(y|\beta) = \frac{\beta^{\nu+1}}{\Gamma(\nu+1)} y^{\nu} e^{-\beta y} \propto \beta^{\nu+1} e^{-\beta y}$$

So the full conditional distribution of β can be expressed as follows.

$$p(\beta|y) \propto f(y|\beta)\pi(\beta) \propto \beta^{\nu+1} e^{-\beta y} e^{-\beta} \propto \beta^{\nu+1} e^{-\beta(y+1)}$$

This is kernel of a Gamma distribution with $\alpha = \nu + 2$ and $\beta = y + 1$

(b) With known β

$$f(y|\nu) = \frac{\beta^{\nu+1}}{\Gamma(\nu+1)} y^{\nu}$$

So the full conditional distribution of ν is proportional to

$$p(\nu|y) \propto f(y|\nu)\pi(\nu) \propto \frac{\beta^{\nu+1}}{\Gamma(\nu+1)} y^{\nu} \frac{1}{2^{\nu}} \propto \frac{1}{\Gamma(\nu+1)} (\frac{\beta y}{2})^{\nu}$$

This is kernel of a Poisson distribution, λ is $\frac{\beta y}{2}$

(c) I found that the full conditional distribution of β follows a Gamma distribution (ν + 2, y + 1) and the full conditional distribution of ν follows a Poisson distribution where λ is $\beta * y/2$.

I run the following Gibbs sampler to calculate the probability asked.

```r
### Run Gibbs Sampler ...

  # set up
set.seed(777)
Nsim <- 100000

betas <- numeric(Nsim)
nus <- numeric(Nsim)

  # initialize

betas[1] <- 3

nus[1] <- 1

  # iterate

for(k in 2:Nsim){
  betas[k] <- rgamma(1, 0+2, 10+1)
  nus[k] <- rpois(1, betas[k]*10/2)
}

### Probability of upsilon being 0 when y = 10
length(which(nus==0))/Nsim
```

```
## [1] 0.47193
```