Guanlin Zhang
Fall '18

Question 1: Chapter 2 Question 11 on page 59:

Solution 1. *For part (a):*

The general equation to compute the unnormalized posterior density is:

$$\begin{aligned} p(\theta|y) &\propto p(y|\theta)\cdot p(\theta) \\ &= \Big(\prod_{i=1}^{5} p(y_i|\theta)\Big)\cdot p(\theta) = \Big(\prod_{i=1}^{5}\frac{1}{1+(y_i-\theta)^2}\Big)\cdot\frac{1}{100} \end{aligned}$$

Since we intend to evaluate θ *on the grid* $0, \frac{1}{m}, \frac{2}{m}, \ldots, 100$ *for some large integer* m*, without loss of generality, we could let* $m = 100$*, and we can define sequence*

$$\theta_j = \frac{j-1}{m}, j = 1, 2, \ldots, 10001$$

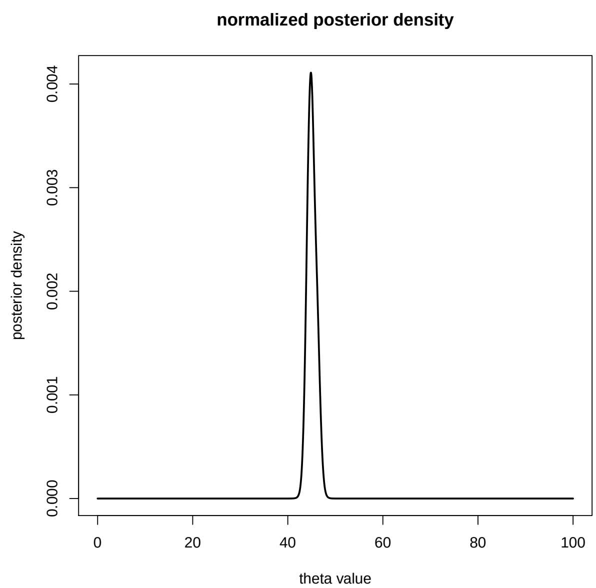
so $\theta_1 = 0, \theta_2 = \frac{1}{100}, \ldots, \theta_{10001} = 100$ *and evaluate* $p(\theta_j|y)$ *according to the equation above.*

To compute the normalized posterior density, we just need to do the following:

$$p(\theta|y) = \frac{\Big(\prod_{i=1}^{5}\frac{1}{1+(y_i-\theta)^2}\Big)\cdot\frac{1}{100}}{\sum_{j=1}^{10001}\Big(\prod_{i=1}^{5}\frac{1}{1+(y_i-\theta_j)^2}\Big)\cdot\frac{1}{100}} = \frac{\Big(\prod_{i=1}^{5}\frac{1}{1+(y_i-\theta)^2}\Big)}{\sum_{j=1}^{10001}\Big(\prod_{i=1}^{5}\frac{1}{1+(y_i-\theta_j)^2}\Big)}$$

The following R code fulfill the computations above.

```r
#input data
y <- c(43, 44, 45, 46.5, 47.5)
#define grid
theta <- seq(from = 0, to = 100, by = 0.01)
#create vector for unnoramlized posterior
post_unnorm <- rep(NA, 10001)
#compute the unnoramlized posterior density
for (j in 1:10001){
post_unnorm[j] <- prod(1/(1 + (y - theta[j])^2))*0.01
}
#compute the normalizing constant
C <- sum(post_unnorm)
#compute the normalized posterior density
post_norm <- post_unnorm/C
#plot the normalized posterior density
plot(theta, post_norm, type="l", lty = 1, lwd = 2, xlab = "theta value",
     ylab = "posterior density", main = "normalized posterior density")
```

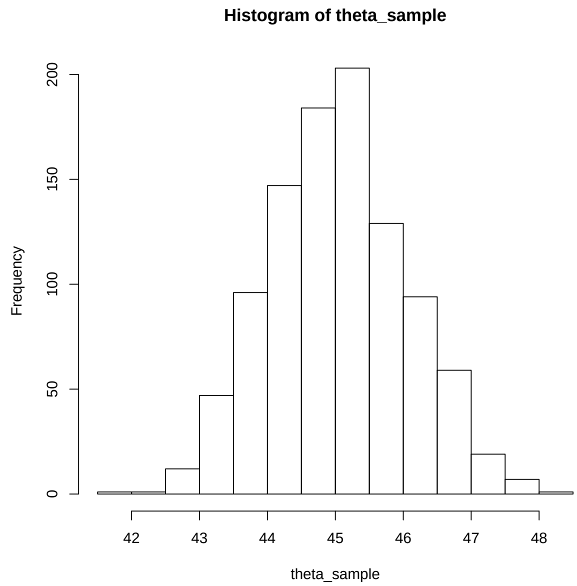



```
#output the high density section:
post_norm[4491:4501]

##  [1] 0.004106013 0.004103606 0.004100638 0.004097111 0.004093027
##  [6] 0.004088391 0.004083205 0.004077474 0.004071203 0.004064397
## [11] 0.004057062
```

Since there are over 10^4 *grid points, we could not output all of the values for posterior density. However based on the plot, we gave the output above* 10 *representative high posterior density values(normalized).*

For part (b): the following R code sample 1000θ *values from the normalized posterior density we have computed from part (a) and plot a histogram as well:*

```
#draw 1000 samples from posterior density
set.seed(1)
theta_sample <- sample(theta, size = 1000, replace = TRUE, prob = post_norm)
hist(theta_sample)
```



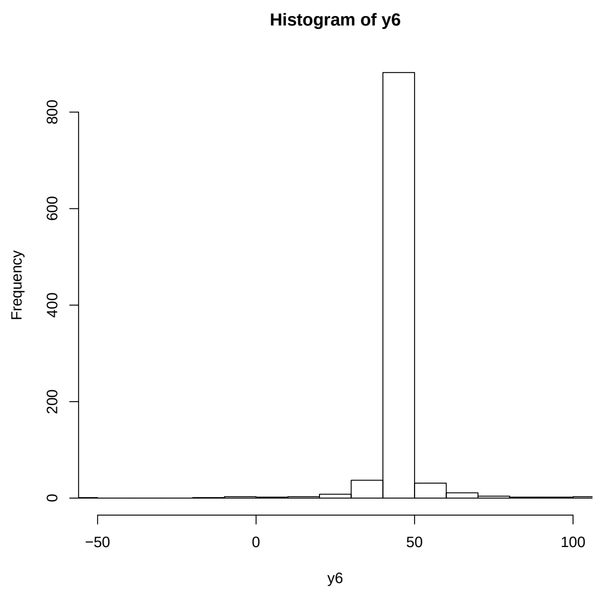
For part (c):

We are really considering the posterior predictive distribution here:

$$p(\tilde{y}|y) = \int p(\tilde{y}|\theta)p(\theta|y)d\theta$$

with $\tilde{y} = y_6$ *and* $y = (y_1, y_2, y_3, y_4, y_5)$. *The following R code draw* 1000 *predictive samples for* y_6*(each corresponding to one sample of* θ*) and plot a histogram:*

```
set.seed(1)
y6 <- rcauchy(1000, location = theta_sample, scale = 1)
hist(y6, nclass = 100, xlim = c(-50, 100))
```



we can see that the prediction for y_6 is with high chance slightly less than 50.

Question 2:

An experiment was performed to estimate the effect of beta-blockers on mortality of cardiac patients. A group of patients were randomly assigned to treatment and control groups: out of 674 patients receiving the control, 39 died, and out of 680 receiving the treatment, 22 died. Assume that the outcomes are independent and binomially distributed, with probabilities of death of p_0 and p_1 under the control and treatment, respectively. Set up a noninformative or weakly informative prior distribution on (p_0, p_1).

(a) Summarize the posterior distribution for the odds ratio, $\Big(p_1/(1-p_1)\Big)/\Big(p_0/(1-p_0)\Big)$.

(b) Discuss the sensitivity of your inference to your choice of prior density.

Solution 2. *We have the following data:*

	death(y)	*total*(n)
Control(0)	39	674
Treatment(1)	22	680

For part (a):

Since the two samples are independent, we can just look at the marginal prior of p_0 and p_1 separately, as well as the posterior. Indeed we have

$$\begin{aligned} p(p_0, p_1) &= p(p_0)\cdot p(p_1) \\ p(p_0, p_1|y_0, y_1) &= p(p_0|y_0)\cdot p(p_1|y_1) \propto p(p_0)p(y_0|p_0)\cdot p(p_1)\cdot p(y_1|p_1) \end{aligned}$$

A natural choice of noninformative prior for (p_0, p_1) could be Uniform$[0,1]\times[0,1]$ with p_0 and p_1 each has a marginal prior density Uniform$[0,1]$.

Hence we have the following posterior distributions:

$$
\begin{aligned}
p(p_0|y_0) &\propto p_0^{y_0}(1-p_0)^{n_0-y_0} \\
p(p_1|y_1) &\propto p_1^{y_1}(1-p_1)^{n_1-y_1} \\
p\Big((p_0,p_1)|y_0,y_1\Big) &\propto p_0^{y_0}(1-p_0)^{n_0-y_0}\cdot p_1^{y_1}(1-p_1)^{n_1-y_1}
\end{aligned}
$$

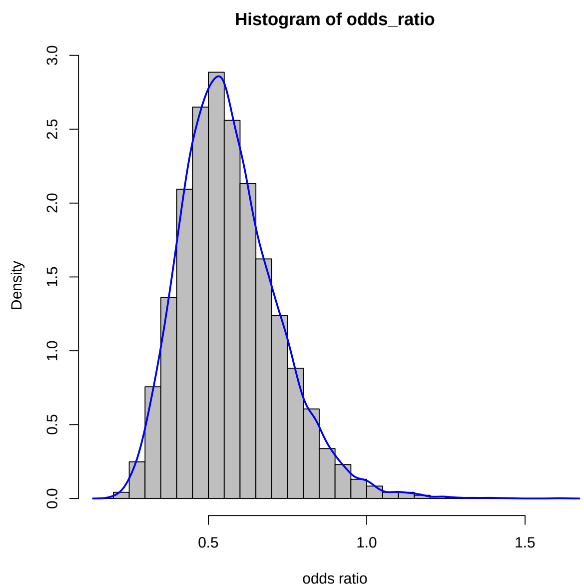
which implies that

$$
\begin{aligned}
p_0|y_0 &\sim Beta(p_0|y_0+1, n_0-y_0+1) = Beta(p_0|40, 636) \\
p_1|y_1 &\sim Beta(p_1|y_1+1, n_1-y_1+1) = Beta(p_1|23, 659)
\end{aligned}
$$

The following R code draws the posterior values of p_0 and p_1 and then compute the odds ratio. We also make Bayesian inference for the odds ratio based on the simulation by looking at the mean, variation and 95% posterior interval:

```
set.seed(35)
#input the data
y_0 <- 39; n_0 <- 674; y_1 <- 22; n_1 <- 680

#draw posterior sample for p0, p1
p_0 <- rbeta(10000, shape1 = 40, shape2 = 636)
p_1 <- rbeta(10000, shape1 = 23, shape2 = 659)

#compute posterior sample for odds ratio
odds_ratio <- p_1*(1 - p_0)/(p_0*(1 - p_1))

#make Bayesian inference
post_mean <- mean(odds_ratio)
post_var <- var(odds_ratio)
sort_oddsratio <- sort(odds_ratio)
post_interval <- c(quantile(sort_oddsratio, probs = 0.025),
                   quantile(sort_oddsratio, probs = 0.975))

#print outcomes
cat("the posterior mean of odds ratio is", post_mean, "\n")

## the posterior mean of odds ratio is 0.5682376

cat("the posterior variance of odds ratio is", post_var, "\n")

## the posterior variance of odds ratio is 0.02376768

cat("the posterior interval of odds ratio is", post_interval)

## the posterior interval of odds ratio is 0.3174496 0.9173986
```

```
hist(odds_ratio, prob = TRUE, col = "grey",
   xlab = "odds ratio", nclass = 40)# prob=TRUE for probabilities not counts
lines(density(odds_ratio), col="blue",
    lwd=2) # add a density estimate with defaults
```



According to simulation, the point estimate for the posterior odds ratio is 0.57*, we say that based on the data the posterior odds of death in the treatment group is about half the odds of death in the control group.*

The 95% *posterior interval is* $(0.32, 0.92)$*, so we say that a posteriori the odds ratio has* 95% *probability in the range between* 0.32 *and* 0.92*.*

We also give a histogram and an approximate density curve of the odds ratio.

For part (b):

As we know that the binomial distribution has conjugate prior as $Beta(\theta|\alpha, \beta)$*, and the posterior distribution under the conjugate prior is* $Beta(\theta|\alpha + y, \beta + n - y)$*, so we have*

$$E(\theta|y) = \frac{\alpha + y}{\alpha + \beta + n}$$
$$var(\theta|y) = \frac{E(\theta|y)[1 - E(\theta|y)]}{\alpha + \beta + 1}$$

We can say that when α *and* β *are much smaller than* n *and* y *the data is dominant and the posterior distribution for* θ *is less sensitive to the choice of prior. But when* α *and* β *are getting larger our prior makes more impact on the posterior distribution.*

Of course the above is only talking about the sensitivity of posterior distribution for p_0 *and* p_1 *when we look at the equations. But since the odds ratio is a function of* (p_0, p_1)*, we figure its posterior distribution should have the same kind of sensitivity towards the choice of prior, as of* p_0 *and* p_1*.*

Question 3:

Consider a case where the same factory has two production lines for manufacturing car windshields. Independent samples from the two production lines were tested for hardness. The hardness measurements for the two samples y_1 and y_2 are shown below.

windshield y_1	windshield y_2
13.357	15.98
14.928	14.206
14.896	16.011
15.297	17.25
14.82	15.993
12.067	15.722
14.824	17.143
13.865	15.23
17.447	15.125
	16.609
	14.735
	15.881
	15.789

(a) What can you say about $\mu_d = \mu_1 - \mu_2$? Sumarize your results using Bayesian point and interval estimates.

(b) Are the means the same?

Solution 3. *With the unknown values of σ_1 and σ_2, we could choose the noninformative prior distribution as*

$$p(\mu_j, \sigma_j^2) \propto (\sigma^2)^{-1}$$

It has been known that with the chosen prior above the posterior marginal distribution for σ_j^2 follows a scaled inverse-χ^2 distribution:

$$\sigma_j^2 | y_j \sim Inv - \chi^2(n_j - 1, s_j^2), \;\; j = 1, 2$$

and the conditional posterior distribution for μ_j given σ_j^2 follows normal distributions:

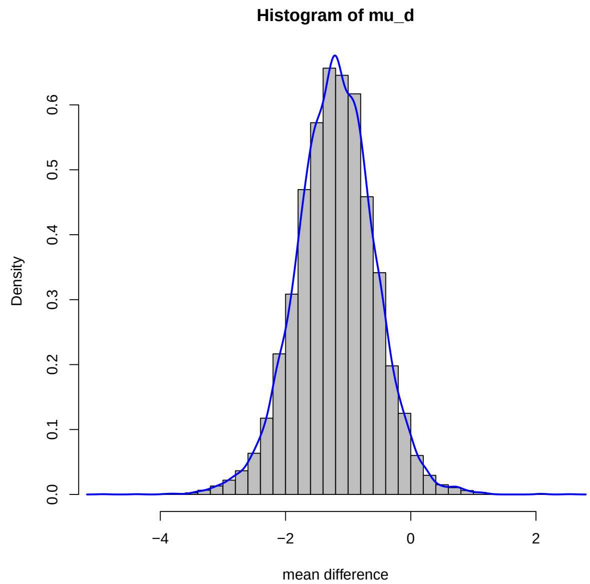
$$\mu_j | \sigma_j^2, y_j \sim N(\overline{y}_j, \sigma_j^2 / n_j)$$

To make Bayesian inference on μ_j, we can first draw σ_j^2, then draw μ_j based on the above distributions.

The following R code does the simulation:

```
#input the data
y1 <- c(13.357, 14.928, 14.896, 15.297, 14.82, 12.067, 14.824, 13.865
      , 17.447)
```

```
y2 <- c(15.98, 14.206, 16.011, 17.25, 15.993, 15.722, 17.143
      , 15.23, 15.125, 16.609, 14.735, 15.881, 15.789)
n1 <- length(y1)
n2 <- length(y2)

set.seed(35)
#draw samples for sigma^2
var1 <- (n1 - 1)*var(y1)/rchisq(10000, df = n1- 1)
var2 <- (n2 - 1)*var(y2)/rchisq(10000, df = n2- 1)
#draw samples for mu
mu1 <- rnorm(10000, mean = mean(y1), sd = sqrt(var1/n1))
mu2 <- rnorm(10000, mean = mean(y2), sd = sqrt(var2/n2))
#compute mu_d
mu_d <- mu1 - mu2
#summarize Bayesian inference results

mu_d_mean <- mean(mu_d)
mu_d_var <- var(mu_d)
mu_d_sort <-sort(mu_d)
mu_d_interval <- c(quantile(mu_d_sort, probs = 0.025),
                 quantile(mu_d_sort, probs = 0.975))

#print outcome
cat("the posterior mean of mean difference is", mu_d_mean, "\n")

## the posterior mean of mean difference is -1.203291

cat("the posterior variance of  mean difference is", mu_d_var, "\n")

## the posterior variance of  mean difference is 0.3938078

cat("the posterior interval of mean difference is", mu_d_interval)

## the posterior interval of mean difference is -2.46591 0.002683758

hist(mu_d, prob = TRUE, col = "grey", xlab = "mean difference",
   nclass = 40)# prob=TRUE for probabilities not counts
lines(density(mu_d), col="blue",
    lwd=2) # add a density estimate with defaults
```



Based on the simulation the a posteriori estimate for μ_d is -1.20, which says windshields hardness from product line 1 is -1.20 less than those from product line 2.

The 95% posterior interval estimate is $(-2.47, 0.003)$, which says that the a posteriori there is 95% chance the hardness difference of the windshields between line 1 and line 2 is in this interval. Since almost the whole interval lies on the left side of 0, again it is consistent with the conclusion of point estimate.

For part (b):

With the noninformative prior distribution we used here, we can conclude that the means are not the same, and the one from product line 1 is smaller a posteriori.

Question 4: Chapter 3 Question 10 on page 82:

Solution 4. *We first mention a simple probability derivation before jumping into our problem.*

Suppose there are two random variables X and Y such that $Y = kX$ where k is a constant, then if we look at the relationship of their CDFs, there is:

$$F_Y(y) = P(Y \leq y) = P(kX \leq y) = P(X \leq \frac{y}{k}) = F_X(\frac{y}{k})$$

Now if we take derivatives on both sides, then we get the relationship of their densities:

$$p_Y(y) = \frac{1}{k} p_X(\frac{y}{k})$$

Back to the problem:

Given the two normal samples with parameters (μ_j, σ_j^2) with prior $p(\mu_j, \sigma_j^2) \propto \sigma_j^{-2}$, we know the marginal posterior distribution for σ_i^2 in each sample is:

$$\sigma_j^2 | y_j \sim Inv-\chi^2(n_j - 1, s_j^2), \;\; j = 1, 2$$

and hence

$$p_{\sigma_j^2|y_j} \propto (s_j^2)^{\frac{n_j-1}{2}} \cdot \theta^{-(\frac{n_j-1}{2}+1)} \cdot \exp\Big\{ -(n_j-1)s_j^2/(2\theta)\Big\}$$

Using the probability derivation we mentioned above, then there is

$$p_{\frac{\sigma_j^2}{(n_j-1)s_j^2}\Big|y_j}(\theta) = (n_j-1)s_j^2 p_{\sigma^2|y_j}\Big((n_j-1)s_j^2\theta\Big)$$

thus we have:

$$\begin{aligned} p_{\frac{\sigma_j^2}{(n_j-1)s_j^2}\Big|y_j}(\theta) &\propto s_j^2 \cdot \Big(s_j^2\Big)^{\frac{n_j-1}{2}} \cdot \Big(s_j^2\theta\Big)^{-(\frac{n_j-1}{2}+1)} \cdot \exp\Big\{ -\frac{(n_j-1)s_j^2}{2(n_j-1)s_j^2\theta}\Big\} \\ &= s_j^{2+(n_j-1)-(n_j-1+2)} \cdot \theta^{-(\frac{n_j-1}{2}+1)} \cdot e^{-1/(2\theta)} \\ &= \theta^{-(\frac{n_j-1}{2}+1)} \cdot e^{-1/(2\theta)} \end{aligned}$$

which is the kernel of the density of Inv-$\chi^2_{n_j-1}(\theta)$*, so*

$$\begin{aligned} &\frac{\sigma_j^2}{(n_j-1)s_j^2}|y_j \sim Inv-\chi^2_{n_j-1}(\theta) \\ \Longrightarrow &\frac{(n_j-1)S_j^2}{\sigma_j^2}|y_j \sim \chi^2_{n_j-1}(\theta) \end{aligned}$$

So we have:

$$\frac{s_1^2/s_2^2}{\sigma_1^2/\sigma_2^2}\Big|y_1,y_2 = \frac{\frac{(n_1-1)s_1^2}{\sigma_1^2}/(n_1-1)}{\frac{(n_2-2)s_2^2}{\sigma_2^2}/(n_2-1)}\Big|y_1,y_2 = \frac{\frac{(n_1-1)s_1^2}{\sigma_1^2}/(n_1-1)|y_1}{\frac{(n_2-2)s_2^2}{\sigma_2^2}/(n_2-1)|y_2}$$

The last " = " above in our equation holds because of independence between both samples and parameters of samples. Notice that this is really the following:

$$\frac{\chi^2_{n_1-1}/(n_1-1)}{\chi^2_{n_2-1}/(n_2-1)} \sim F(n_1-1, n_2-1)$$

Again the two χ^2 *distributions are independent due to the independence of samples and parameters of samples, which lead to the F distribution.*