

Solutions to Homework 3

Shashank Sule

9/10/2019

Problem 1

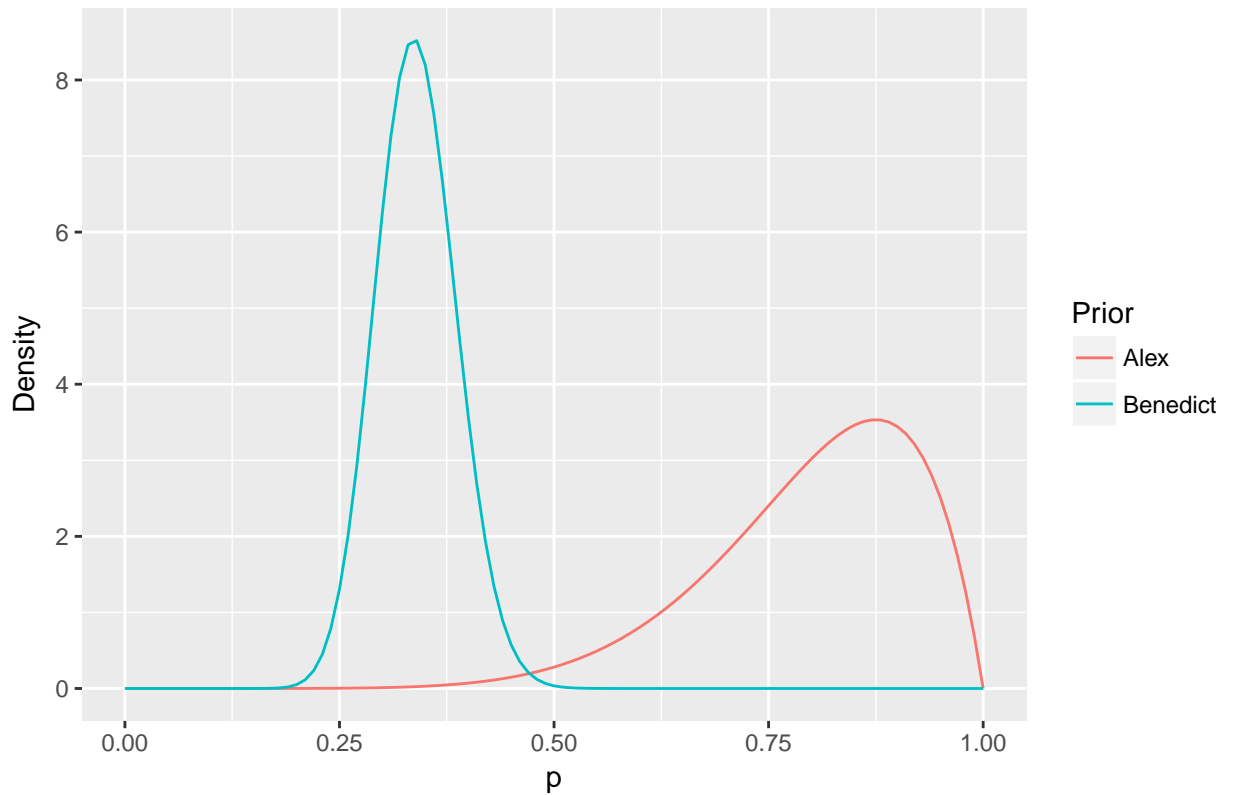
a) In a Beta distribution parametrized as $\text{Beta}(a, b)$, a represents the number of successes and b the number of failures. So if there were 8 successes and 2 failures out of a data of 10 then Alex has a prior $\text{Beta}(a, b)$.

b) We compute the parameters using the `beta.select` function:

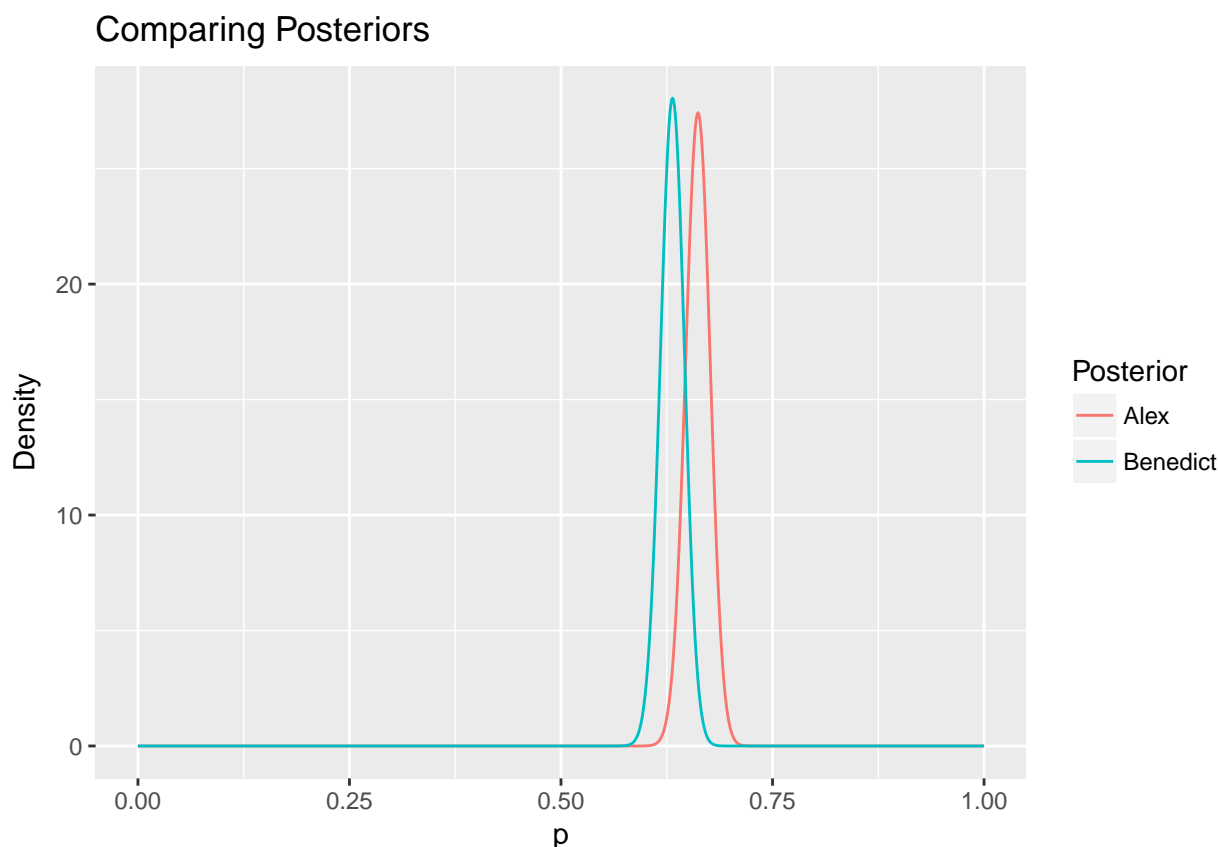
```
## [1] 34.95 67.98
```

c) Note that $y = 692$, $n = 1048$ and the posterior distribution is $\text{Beta}(a + y, b + n - y)$. So Alex has posterior $\text{Beta}(700, 358)$ and Benedict has posterior $\text{Beta}(726.95, 423.98)$.

Comparing priors



The prior:



Credible intervals for Alex:

```
## [1] 0.6328457 0.6898269
```

Credible intervals for Benedict:

```
## [1] 0.6035532 0.6592528
```

- d) To conduct a prior predictive check, we sample values from each prior and then sample a value from the likelihood function. Then we look for clustering of the sampled data around the actual data. I'll do a simulation for 1000 trials.

Alex's probability values:

```
c(sum(alex >= 692)/S, 1 - sum(alex >= 692)/S)
```

```
## [1] 0.865 0.135
```

Benedict's probability values:

```
c(sum(benedict >= 692)/S, 1 - sum(benedict >= 692)/S)
```

```
## [1] 0 1
```

Thus, neither have a good prior distribution for prediction but Benedict's is worse than Alex's because Benedict's probabilities are far more extreme (all of Benedict's data lies on one side of the actual observed value).

Problem 2

To get the odds data, I'll use Monte Carlo sampling and then transform the data. The summarized data is as follows:

```
##           Mean   Median    2.5%    97.5%
## Alex      1.958228 1.956094 1.732098 2.204190
## Benedict  1.719126 1.716058 1.528348 1.935351
```

On average, Alex's odds of kids having a TV in their room are slightly igher than those of Benedict.

Problem 3

```
##      S      5%      95%
## 1    10 0.3703484 0.7249998
## 2   100 0.4514916 0.7401867
## 3   500 0.4152143 0.7546605
## 4  1000 0.4289477 0.7400327
## 5 5000 0.4288951 0.7442373
```

I computed $S = \{10, 100, 500, 1000, 5000\}$ random samples from $\text{Beta}(15.06, 10.56)$ using the `rbeta` function and then computed the quantiles of the data using `quantile`. As the number of samples gets larger, the interval gets closer to the actual credible interval.

Problem 4

- a) Note that given the means μ_A and μ_B , the data follows a normal likelihood. Now if we sample independently from both the accelerated and non-accelerated groups then the joint likelihood of the data is given by the product of their individual likelihood functions. First of all, from the formulas derived in class we set $\phi = 1/16$ and $n = 6$ to get that

$$L_A(\mu_A) = \left(\frac{1}{32\pi}\right)^3 e^{-\frac{1}{32} \sum_{i=1}^6 (y_{A,i} - \mu_A)^2}$$

and

$$L_B(\mu_B) = \left(\frac{1}{32\pi}\right)^3 e^{-\frac{1}{32} \sum_{j=1}^6 (y_{B,j} - \mu_B)^2}$$

The joint likelihood is obtained as a product of the individual likelihoods, so

$$L_{A,B}(\mu_A, \mu_B) = L_A(\mu_A)L_B(\mu_B) = \left(\frac{1}{32\pi}\right)^6 e^{-\frac{1}{32}\sum_{j=1}^6(y_{A,j}-\mu_A)^2 - \frac{1}{32}\sum_{j=1}^6(y_{B,j}-\mu_B)^2}$$

- b) If the joint prior for (μ_A, μ_B) is independent then the density function is a product of the individual priors, i.e $\pi(\mu_A, \mu_B) = \pi(\mu_A)\pi(\mu_B)$. Thus, the joint posterior can be obtained as follows:

$$\pi(\mu_A, \mu_B \mid \mathbf{Y}_A, \mathbf{Y}_B, \sigma) = \frac{L_{A,B}(\mu_A, \mu_B)\pi_{A,B}(\mu_A, \mu_B)}{\int_{\mathbb{R}^2} L_{A,B}(x, y)\pi_{A,B}(x, y) dA}$$

Since both the joint likelihood and prior are products of the marginal likelihood and prior, the integral in the denominator is separable.

$$\begin{aligned} &= \frac{L_A(\mu_A)\pi_A(\mu_A)L_B(\mu_B)\pi_B(\mu_B)}{\int_{\mathbb{R}} L_A(x)\pi_A(x) dx \int_{\mathbb{R}} L_B(y)\pi_B(y) dy} = \left(\frac{L_A(\mu_A)\pi_A(\mu_A)}{\int_{\mathbb{R}} L_A(x)\pi_A(x) dx}\right) \left(\frac{L_B(\mu_B)\pi_B(\mu_B)}{\int_{\mathbb{R}} L_B(y)\pi_B(y) dy}\right) \\ &= \pi_A(\mu_A \mid \mathbf{Y}_A, \sigma)\pi_B(\mu_B \mid \mathbf{Y}_B, \sigma) \end{aligned}$$

Finally, from the result in class, the family of normal distributions forms a conjugate prior under a normal likelihood where the posterior distribution has mean $\frac{\phi_0\mu_0+n\phi\bar{y}}{\phi_0+n\phi}$ and standard deviation $\sqrt{\frac{1}{\phi_0+n\phi}}$. Thus, in this case we have

$$\mu_A \mid \mathbf{Y}_A, \sigma \sim N\left(\frac{\phi_A\gamma_A + (6/16)15.2}{\phi_A + (6/16)}, \sqrt{\frac{1}{\phi_A + (6/16)}}\right)$$

and

$$\mu_B \mid \mathbf{Y}_B, \sigma \sim N\left(\frac{\phi_B\gamma_B + (6/16)6.2}{\phi_B + (6/16)}, \sqrt{\frac{1}{\phi_B + (6/16)}}\right)$$

Problem 5

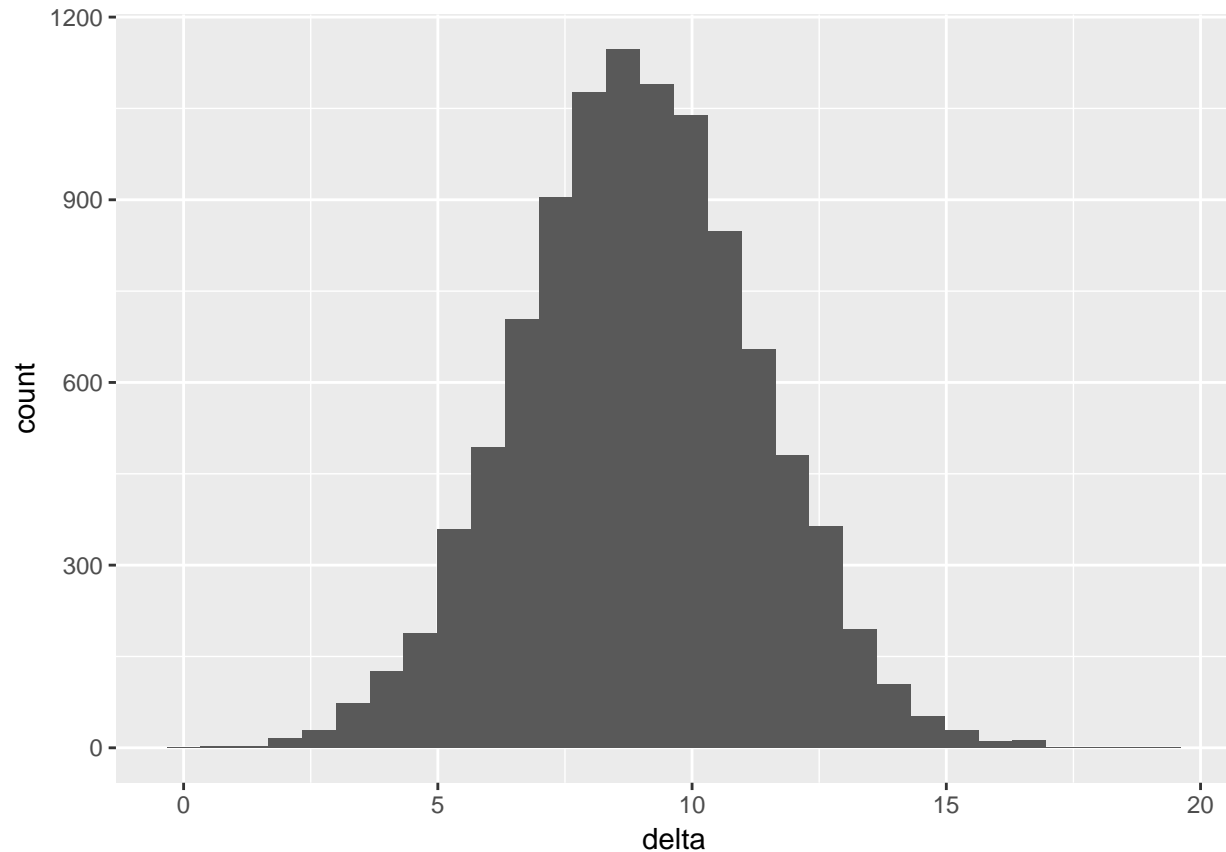
- a) As stated in the hint, I compute $\delta = \mu_A - \mu_B$ and check the proportion using $S = 100000$ monte carlo draws.

```
S <- 10000
n <- 6
phi <- 1/16
phi_A <- (1/(20))^2
mu_A0 <- 0
ybar_A <- 15.2
phi_B <- (1/(20))^2
mu_B0 <- 0
```

```
ybar_B <- 6.2
delta <- rnorm(S, mean = ((phi_A*mu_A0 + n*phi*ybar_A)/(phi_A + n*phi)), sd = sqrt(1/(phi_A + n*phi)))

ggplot(as.data.frame(delta), aes(x=delta)) +
  geom_histogram()
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



```
sum(delta > 0)/S
```

```
## [1] 0.9999
```

Thus, the average improvement for the accelerated group is larger than that for the no growth group since most of δ lies above 0 (using monte carlo simulation).

- b) As suggested in the hint, I'll draw a posterior predictive distribution using the posteriors of μ_A and μ_B and their respective likelihoods. Then I'll find the proportion of positive draws from the $\tilde{Y}_A - \tilde{Y}_B$.

```
A_means <- numeric(length = S)
B_means <- numeric(length = S)

for(i in 1:S){
  mu_A <- rnorm(1, mean = ((phi_A*mu_A0 + n*phi*ybar_A)/(phi_A + n*phi)), sd = sqrt(1/(phi_A + n*phi)))
```

```

mu_B <- rnorm(1, mean = ((phi_B*mu_B0 + n*phi*ybar_B)/(phi_B + n*phi)), sd = sqrt(1/(phi_B + n*phi)))
A_means[i] <- sum(rnorm(6, mean = mu_A, sd = 4))/6
B_means[i] <- sum(rnorm(6, mean = mu_B, sd = 4))/6
}

differences <- A_means - B_means
mean(differences > 0)

## [1] 0.9975

```

Thus, most values of \tilde{Y}_A are higher than those in \tilde{Y}_B a child randomly assigned to the accelerated treatment will show better IQ improvement than a child randomly assigned to the normal treatment.